

UNIVERSITE DE MONTPELLIER



Mémoire d'Habilitation à Diriger des Recherches

Ecole Doctorale : GAIA

**PHYLOGENOMIQUE : LA RECONSTRUCTION PHYLOGENETIQUE
ET SES APPLICATIONS A L'ERE DE LA GENOMIQUE**

par

Frédéric DELSUC

Soutenu le 7 Novembre 2016 à Montpellier
devant le jury composé de :

MME. Céline BROCHIER-ARMANET, Professeur, Université de Lyon
MME. Simonetta GRIBALDO, Directeur de Recherche, Institut Pasteur
MR. Ludovic ORLANDO, Professeur, Université de Copenhague
MR. Manolo GOUY, Directeur de Recherche CNRS, Lyon
MR. David MOREIRA, Directeur de Recherche CNRS, Orsay
MR. Jean-Yves RASPLUS, Directeur de Recherche INRA, Montpellier

Rapporteur
Rapporteur
Rapporteur
Examineur
Examineur
Président

TABLE DES MATIERES

1. THEMATIQUES DE RECHERCHE.....	3
2. ACTIVITES DE RECHERCHE	3
2.1. Evolution des Tuniciers	3
2.1.1 La Place des Tuniciers au Sein des Chordés	3
2.1.2 Mitogénomique des Tuniciers	4
2.1.3 Phylogénomique et Evolution Moléculaire des Tuniciers	4
2.2. Phylogénie Moléculaire et Evolution des Mammifères Placentaires	6
2.2.1 Phylogénie des Xénarthres Actuels	6
2.2.2 ADN ancien et Xénarthres Récemment Eteints	8
2.2.3 Mitogénomique des Chauve-Souris Phyllostomidés	13
2.3. Phylogénomique pour la Résolution de Grandes Questions Evolutives	14
2.3.1 La Position Phylogénétique des Tortues	14
2.3.2 La Racine des Placentaires	16
2.4. Exemples d'Applications des Phylogénies	17
2.4.1 Origine des Gènes Hox chez les Animaux	17
2.4.2 Evolution Moléculaire des Facteurs de Restriction des Virus HIV	18
2.4.3 Evolution des gènes CCT chez les Parasites Apicomplexes	19
2.4.4 Origine et Evolution des Régulateurs du Protéasome chez les Eukaryotes	21
2.5. Outils Bioinformatiques pour la Phylogénie et l'Evolution Moléculaire	22
2.5.1 La Base de Données OrthoMaM	22
2.5.2 Le Portail Web PhyloExplorer	23
2.5.3 Le Programme d'Alignement MACSE	24
3. PROJET DE RECHERCHE	25
3.1. ConvergeAnt : Une Approche Intégrative de l'Evolution Convergente.....	26
3.1.1 La Convergence Evolutive	26
3.1.2 Le Modèle des Mammifères Myrmécophages	27
3.1.3 Les Modalités de la Convergence Morphologique	28
3.1.4 Les Traces de la Convergence Moléculaire	29
3.1.5 Le Rôle du Microbiome	30
3.2. Mise en Œuvre du Projet ConvergeAnt	33
3.2.1 Tâche 1 : Convergence Morphologique et Simplification Dentaire	33
3.2.2 Tâche 2 : Convergence Moléculaire à l'Echelle Génomique	36
3.2.3 Tâche 3 : Evolution et des Microbiomes Oraux et Intestinaux	38
4. CURRICULUM VITAE	42
4.1. Informations personnelles	42
4.2. Cours Universitaire et Professionnel.....	43
4.3. Distinctions	44
4.4. Production Scientifique.....	44
4.4.1 Bibliométrie	44
4.4.2 Publications dans Revues à Comité de Lecture	45
4.4.3 Chapitres de Livres	51
4.4.4 Actes de Colloques à Comité de Lecture	51
4.4.5 Autres Publications	51
4.4.6 Communications Invitées	52
4.4.7 Communications dans des Congrès	53
4.4.8 Séminaires et Workshops	57
4.5. Encadrement d'Etudiants.....	59
4.5.1 Post-Doctorants	59
4.5.2 Doctorants	59
4.5.3 Etudiants de Master 2	59
4.5.4 Etudiants de Master 1	60
4.6. Contrats de Recherche	61
4.6.1 Porteur de Contrats	61
4.6.2 Participation à des Contrats	63
4.7. Evaluation de la Recherche	64
4.7.1 Jurys d'Habilitation à Diriger des Recherches	64
4.7.2 Jurys de Thèse	64

4.7.3	Jurys de Master 2	65
4.7.4	Comités de Sélection	65
4.7.5	Comités d'Evaluation	66
4.7.6	Expertise	66
4.7.7	Animation de la Recherche	66
4.7.8	Activités Editoriales	66
4.8.	Enseignement, Formation et Diffusion	67
4.8.1	Enseignements	67
4.8.2	Formations	68
4.8.3	Diffusion Scientifique	69
5.	BIBLIOGRAPHIE	70
6.	ANNEXES.....	77
6.1.	Annexe 1 : Publications Choisies.....	77

1. THEMATIQUES DE RECHERCHE

Depuis Octobre 2009, je suis Chargé de Recherche 1^{ère} classe au sein de l'équipe Phylogénie et Evolution Moléculaire qui appartient au département Génome de l'Institut des Sciences de l'Évolution (ISEM, UMR5554-CNRS) à l'Université Montpellier 2. Mon intégration dans cette équipe lors de mon recrutement en 2005 a été l'occasion de mettre en place une nouvelle thématique sur la phylogénomique et l'évolution des Tuniciers. Parallèlement, j'ai continué à contribuer activement à la thématique générale de l'équipe en m'intéressant à la méthodologie de la reconstruction phylogénétique et à l'évolution des mammifères avec un intérêt particulier pour les Xénarthres (tatous, fourmiliers et paresseux). Je développe également depuis quelques années une nouvelle thématique propre sur les mécanismes génomiques qui sous-tendent les phénomènes de convergence évolutive en utilisant comme modèle les mammifères myrmécophages (voir Projet de Recherche).

2. ACTIVITES DE RECHERCHE

2.1. EVOLUTION DES TUNICIERS

Cet axe de recherche initié à mon arrivée dans l'équipe a été l'objet de la thèse de Georgia Tsagkogeorga soutenue en Décembre 2009 et constitue toujours une des thématiques que nous développons notamment au travers de collaborations avec l'équipe de Patrick Lemaire du Centre de Recherches de Biochimie Macromoléculaire (CRBM-UMR5237) de Montpellier.

2.1.1 LA PLACE DES TUNICIERS AU SEIN DES CHORDES

Cette thématique est la continuité logique de mes travaux post-doctoraux réalisés en avec Hervé Philippe de l'Université de Montréal (Canada). Ainsi, appliquant une approche phylogénomique pour reconstruire l'histoire évolutive des Deutérostomiens, nous avons montré contre toute attente que les Tuniciers, et non pas les Céphalocordés, représentent les plus proches parents des Vertébrés (**Delsuc et al. 2006**). Ce résultat inattendu a eu des implications majeures pour l'interprétation des données paléontologiques, morphologiques et développementales. Cependant, cette étude phylogénomique révélait une autre surprise majeure, les Chordés n'apparaissant pas monophylétiques car l'amphioxus était regroupé avec l'oursin. Bien que ne bénéficiant pas d'un soutien statistique significatif, ce résultat préliminaire constituait néanmoins une hypothèse de travail très intéressante de par ses implications sur notre vision de l'évolution des Deutérostomiens (Gee 2006).

L'accumulation continue des données de séquences nous a ensuite permis d'assembler un jeu de données phylogénomique mis à jour à partir des séquences disponibles en banque pour un échantillonnage plus représentatif comportant 25 deutérostomiens. Ces analyses phylogénomiques utilisant les modèles d'évolution des séquences les plus réalistes à disposition ont conforté le résultat surprenant d'une proche parenté entre Tuniciers et Vertébrés, mais elles ont également retrouvé la monophylie des chordés dont l'amphioxus représente la branche la plus basale (**Delsuc et al. 2008**). Ces résultats, confirmé par d'autres études phylogénomiques (ex. Putnam et al. 2008), soulignent encore une fois l'importance de l'échantillonnage taxonomique et du modèle d'évolution

des séquences pour les reconstructions phylogénétiques basées sur des données génomiques (**Delsuc et al. 2005**).

2.1.2 MITOGENOMIQUE DES TUNICIERS

Dans la continuité de notre collaboration avec Daniel Chourrout du SARS International Centre for Marine Biology de Bergen (Norvège), nous avons participé à l'analyse du génome du tunicier pélagique modèle *Oikopleura dioica* qui possède l'un des plus petits génomes de Chordés (**Denoeud et al. 2010**). La compaction extrême de ce génome a notamment permis d'éclairer les mécanismes à l'origine de l'évolution des introns dans les gènes protéiques et le rôle joué par les éléments transposables. Nous avons plus particulièrement contribué à cette étude en étudiant le génome mitochondrial d'*Oikopleura dioica* qui s'est avéré être particulièrement difficile à séquencer à cause de l'insertion dans la plupart des séquences codantes de régions polyT qui sont probablement excisées par un mécanisme post-transcriptionnel. L'analyse phylogénétique de ce génome mitochondrial a révélé un taux de substitution extrêmement élevé par rapport aux autres tuniciers et aux autres chordés. De façon plus générale, l'analyse des génomes mitochondriaux de tuniciers a confirmé l'extrême plasticité de l'ordre des gènes sur la molécule d'ADN mitochondrial au sein de ce groupe, ce qui en fait un modèle intéressant pour comprendre les mécanismes d'évolution moléculaire à l'origine de ces changements (**Singh et al. 2009**). La mitogénomique des tuniciers constitue une thématique que nous continuons donc de développer en séquençant des génomes mitochondriaux complets par séquençage shotgun Illumina d'ADN total, ce qui permet d'obtenir facilement le génome mitochondrial indépendamment de l'ordre des gènes sur la molécule (**Rubinstein et al. 2013**), lequel est extrêmement variable même entre espèces proches phylogénétiquement (Figure 1).

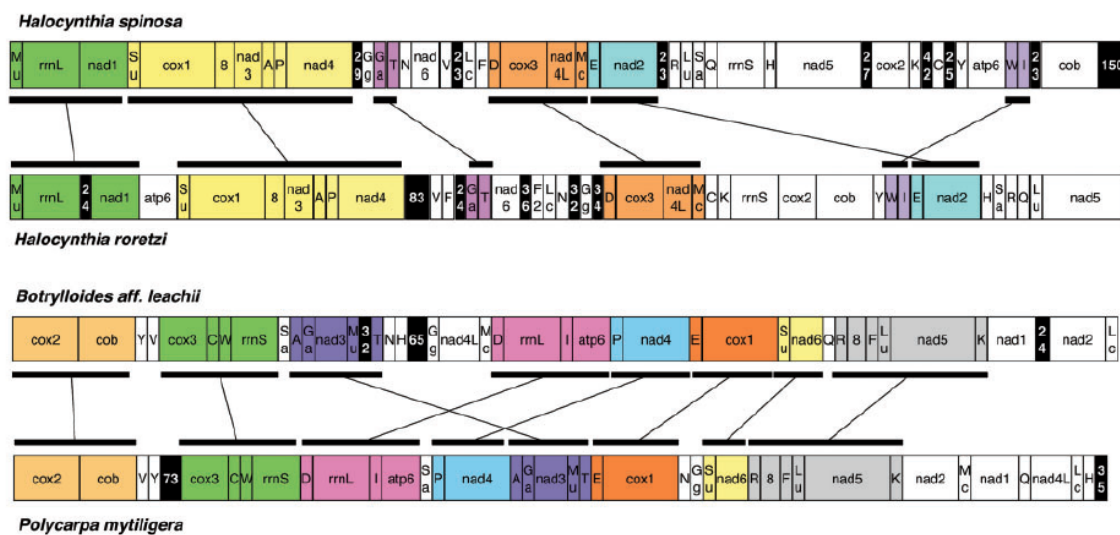


Figure 1 | Remaniements de l'ordre des gènes entre mitogénomes de tuniciers.
D'après Rubinstein et al. (2013).

2.1.3 PHYLOGENOMIQUE ET EVOLUTION MOLECULAIRE DES TUNICIERS

La Thèse de Georgia Tsagkogeorga enfin a permis une mise à jour de la phylogénie des tuniciers sur la base de l'ARNr 18S (**Tsagkogeorga et al. 2009**). Cette étude nous a fourni un cadre phylogénétique de référence pour choisir des espèces représentatives pour l'acquisition de données génomiques.

Par une approche en séquençage 454, nous avons d'abord obtenu des données transcriptomiques pour l'espèce invasive *Microcosmus squamiger*. Celles-ci ont permis de mener une étude détaillée des taux d'évolution entre les trois phylums de chordés sur la base de la comparaison des séquences de 35 gènes de ménage incluant sept espèces de tuniciers. Les résultats obtenus montrent qu'en moyenne les Tuniciers évoluent environ deux fois plus vite que les autres Chordés (**Tsagkogeorga et al. 2010**). Cependant, des analyses détaillées suggèrent une image plus complexe avec différences entre gènes et des accélérations taxon-spécifiques au sein des Tuniciers. Par ailleurs, l'analyse des patrons de sélection sur ces 35 gènes suggère que la cause de l'évolution rapide des Tuniciers pourrait être un taux de mutation élevé, une hypothèse confirmée par des données de génomique des populations (Tsagkogeorga et al. 2012).

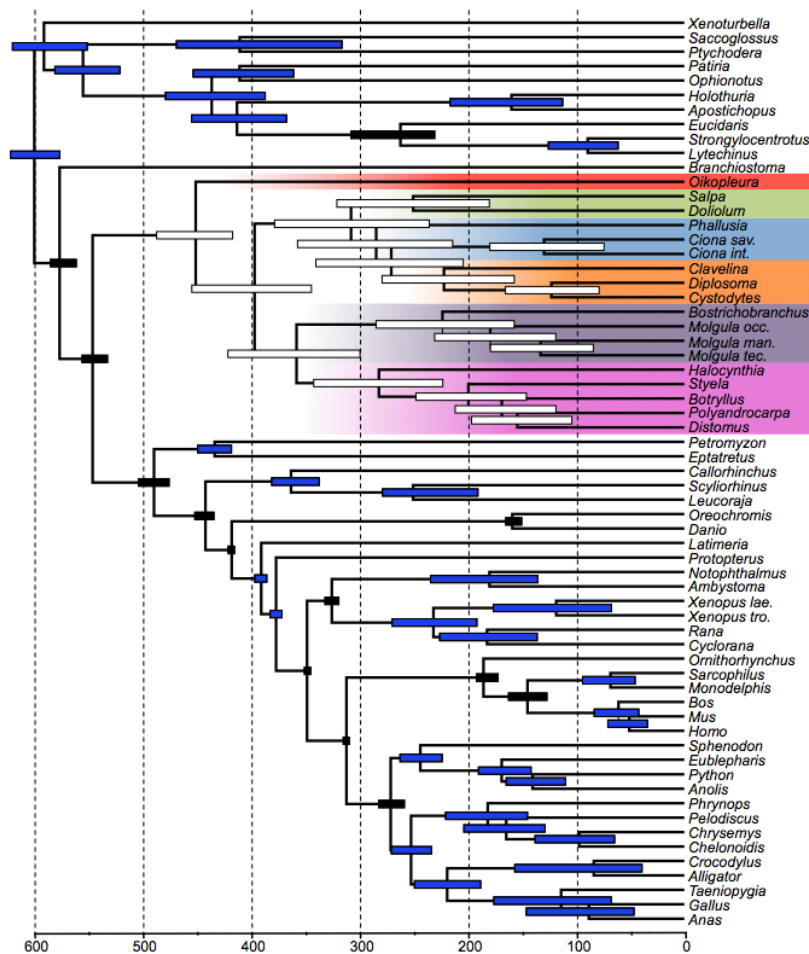


Figure 2 | Un cadre temporel pour l'évolution des tuniciers. Le chronogramme a été obtenu par l'analyse d'une supermatrice de 258 gènes sous un modèle de mélange CAT-GTR+G et un modèle log-normal de relaxation du taux avec le logiciel PhyloBayes.

Nous avons ensuite produits des données transcriptomiques pour onze nouvelles espèces représentant les groupes majeurs de Tuniciers y compris les Thaliacés pélagiques. L'analyse phylogénétique de ces données permet de reconstruire de façon fiable l'histoire évolutive des rameaux majeurs de ce groupe (**Delsuc et al. Soumis**). Elles permettent également de proposer pour la première fois un cadre temporel pour l'évolution de ce groupe dont on ignore totalement les dates de divergence (Figure 2). La proposition d'un cadre phylogénétique et temporel

fiable pour l'évolution des tuniciers au sein des chordés est indispensable aux études de génomique comparatives menées sur les nombreuses espèces modèles que contiennent ces groupes.

Enfin, l'obtention de données génomiques supplémentaires pour des groupes d'intérêt particuliers comme les Thaliacés pélagiques (en cours de séquençage) devrait aboutir à une compréhension accrue des modalités de l'évolution de ces génomes en les contrastant avec celles gouvernant les génomes de vertébrés. C'est cette problématique qui est abordée dans le projet ANR TED (Tunicate Evo-Devo) mené en collaboration avec l'équipe de Patrick Lemaire. Dans le cadre de ce projet, j'encadre actuellement Paul Simion dont le post-doc porte sur la compréhension de l'évolution génomique rapide des Tuniciers.

2.2. PHYLOGENIE MOLECULAIRE ET EVOLUTION DES MAMMIFERES PLACENTAIRES

2.2.1 PHYLOGENIE DES XENARTHRES ACTUELS

C'est dans le cadre de cette problématique, qui est le thème de recherche historique de l'équipe d'accueil, que j'ai réalisé ma thèse au cours de laquelle je m'étais intéressé à reconstruire l'histoire évolutive des Xénarthres (tatous, fourmiliers et paresseux) sur la base de l'étude de gènes nucléaires et mitochondriaux. Aujourd'hui, je continue de développer cette problématique dans le but d'obtenir un cadre phylogénétique de référence incluant les 31 espèces de ce groupe qui constitue l'un des quatre grands clades de Mammifères Placentaires (**Delsuc & Douzery 2008**). L'un des buts de ces recherches est d'utiliser ce cadre phylogénétique pour définir des priorités de conservation pour les différentes espèces menacées. J'ai en effet été nommé en 2009 membre du groupe de spécialistes de l'IUCN pour la conservation des fourmiliers, paresseux et tatous (ASASG) (<http://www.xenarthrans.org/>). Ainsi, nous avons pu obtenir au fil des années des données supplémentaires qui nous ont permis finalement d'inclure tous les genres de xénarthres existants (**Möller-Krull et al. 2007 ; Delsuc et al. 2012**). L'étape suivante était d'inclure l'ensemble de la diversité spécifique du groupe. Ceci a été réalisé dans le cadre d'une collaboration avec Hendrik Poinar du McMaster Ancient DNA Centre de l'Université de McMaster (Hamilton, Canada) soutenue par un projet PEPS CNRS et par le conseil scientifique de l'Université Montpellier 2. Nous avons ainsi pu obtenir le financement du post-doc de Gillian C. Gibb dont le projet a abouti au séquençage Illumina des génomes mitochondriaux complets de toutes les espèces actuelles (**Gibb et al. 2016**). Ce travail établit les Xénarthres comme le premier grand clade de Placentaires être entièrement séquencé au niveau spécifique pour le mitogénomes.

L'analyse de ces données mitogénomiques, nous a permis la reconstruction d'un cadre phylogénétique robuste (Figure 3) qui apparaît compatible avec les études antérieures menées au niveau générique en utilisant des gènes nucléaires (**Delsuc et al. 2002, 2003, 2012**). L'intégration de l'ensemble de la diversité spécifique du groupe a pointé néanmoins un certain nombre d'incohérences dans la systématique des Xénarthres. Ainsi, nous avons proposé par exemple de scinder les tatous en deux familles distinctes Dasypodidae (Dasypodinsés) et Chlamyphoridae (Euphractinés, Chlamyphorinés et Tolypeutinés) afin de mieux refléter leur divergence ancienne, estimée à environ 42 millions d'années. La délimitation des espèces chez les tatous à long museau (genre *Dasypus*) est apparue également plus complexe que prévue, avec notamment la mise en

évidence d'une lignée divergente en Guyane. Les analyses de diversification ont par ailleurs révélé une corrélation négative entre le taux de spéciation et les fluctuations de la température globale avec une augmentation du taux de spéciation correspondant au refroidissement général observé au cours des 15 derniers millions d'années. Finalement, les reconstructions biogéographiques ont identifié la forêt tropicale l'Amazonienne et le bouclier Guyanais comme les berceaux de l'histoire évolutive des Xénarthres avec des dispersions ultérieures vers des habitats plus ouverts et plus secs tels que les pampas argentines et les savanes brésiliennes du Cerrado et du Caatinga.

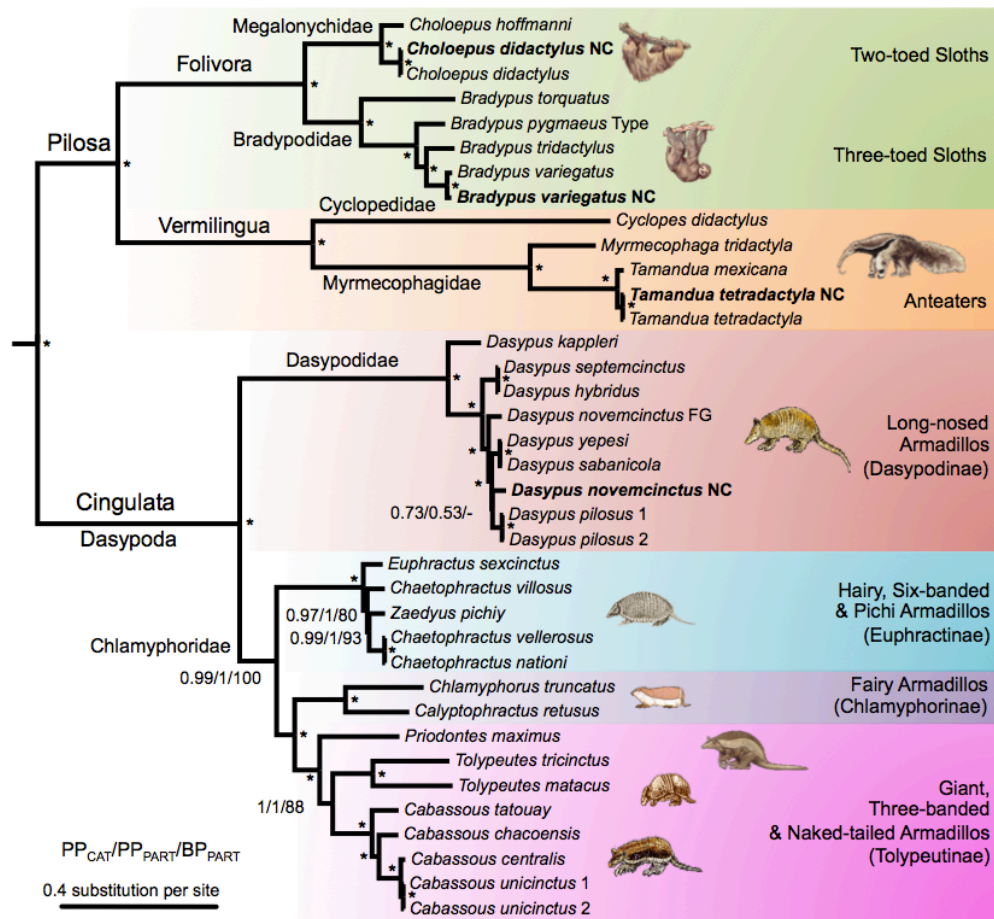


Figure 3 | Phylogénie de référence pour les 31 espèces de Xénarthres actuels.

Ce phylogramme a été obtenu à partir de l'analyse des mitogénomomes complets sous un modèle de mélange CAT-GTR+G avec le logiciel PhyloBayes.

Nous continuons d'appliquer cette méthode de mitogénomique par séquençage shotgun Illumina à des genres dont la taxonomie et les délimitations d'espèces demeurent incertaines tels que le tatou à long museau (genre *Dasypus*) et les paresseux à trois doigts (genre *Bradypus*). Cette technique fait partie intégrante d'une approche de taxonomie intégrative incluant également des données de capture d'exons nucléaires et de géométrie morphométrique sur les crânes. Ces projets sont soutenus financièrement par le Labex CEBA (Centre d'Etude de la Biodiversité Amazonienne) dont je suis le responsable à l'ISEM.

2.2.2 ADN ANCIEN ET XENARTHRES RECEMMENT ETEINTS

Les Xénarthres représentent un groupe de Placentaires dont la majeure partie de l'histoire évolutive s'est déroulée en isolement quasi total pendant la majorité du Cénozoïque. Au cours de cette période d'isolement d'environ 60 millions d'années, les Xénarthres ont réalisé une radiation évolutive dont les espèces actuelles n'offrent qu'un aperçu très fragmentaire (Patterson & Pascual 1972). En effet, l'une des particularités de ce groupe est de comporter un certain nombre de lignées majeures qui se sont éteintes il y a seulement 10 000 ans environ, à la fin des dernières glaciations du Pléistocène (Figure 4). Ces lignées éteintes comportent des taxons emblématiques dont on connaît de nombreux restes subfossiles tels que les paresseux géants (dont les fameux *Megatherium* et *Mylodon*), les glyptodons (*Glyptodon*, *Doedicurus*) et les Pamphathères (*Pampatherium*) dont les affinités phylogénétiques avec les espèces actuelles demeurent pour la plupart énigmatiques.

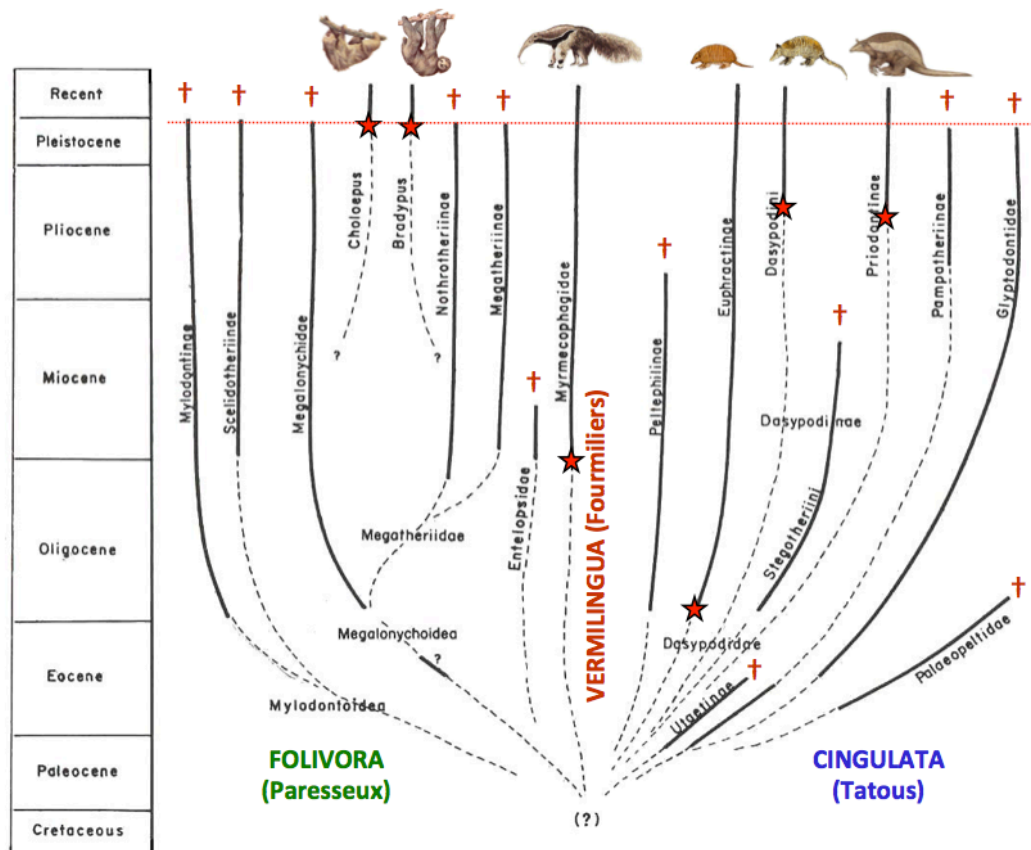


Figure 4 | Radiation des lignées de Xénarthres. La figure représente les principales lignées actuelles et fossiles (+) replacées dans le contexte stratigraphique du Cénozoïque d'Amérique du Sud. Les étoiles représentent la première apparition dans le registre fossile des lignées conduisant aux espèces actuelles.

Modifié d'après Patterson & Pascual (1972).

L'essor des nouvelles technologies de séquençage à haut débit telles que la technique Illumina basée sur le séquençage de courts fragments d'ADN, a entraîné un renouveau des études d'ADN ancien en facilitant dans un premier temps le séquençage de mitogénomes (Paijmans et al. 2013) et désormais de génomes complets (Orlando et al. 2015) d'espèces éteintes dans le dernier million d'années. Nous avons ainsi utilisé le séquençage shotgun Illumina sur un

fragment d'os daté d'environ 13 000 ans appartenant au paresseux terrestre *Mylodon darwinii*. Cet échantillon provient de la grotte du Mylodon en Patagonie chilienne, célèbre pour avoir livrée de nombreux restes de ce paresseux nommé en l'honneur de Charles Darwin (Borrero & Martin 2012). A partir des données de séquences générées nous avons pu reconstruire un mitogénome complet de haute qualité avec une profondeur de couverture moyenne de 120x. Les analyses phylogénétiques et de datations moléculaires sur un jeu de données mitogénomiques incluant toutes les espèces de xénarthres actuelles nous ont permis de résoudre la position du Mylodon qui apparaît comme groupe-frère des paresseux à deux doigts du genre *Choloepus* dont il a divergé il y a plus de 20 millions d'années (Figure 5). En outre, les analyses des substitutions post-mortem montrent que cet échantillon est exceptionnellement bien préservé avec un maximum de 15% de déaminations (C->T ou A->G) observées à l'extrémité des lectures de séquençage. Pour un subfossile, ce fragment d'os contient également une grande quantité d'ADN endogène avec 44% des contigs de plus 150pb qui matchent sur le génome de *Choloepus hoffmanni*. Cet exceptionnel état de conservation provient vraisemblablement des conditions sèches et froides qui règnent dans la grotte du Mylodon, ce qui offre des perspectives encourageantes pour le futur séquençage du génome complet de ce taxon emblématique.

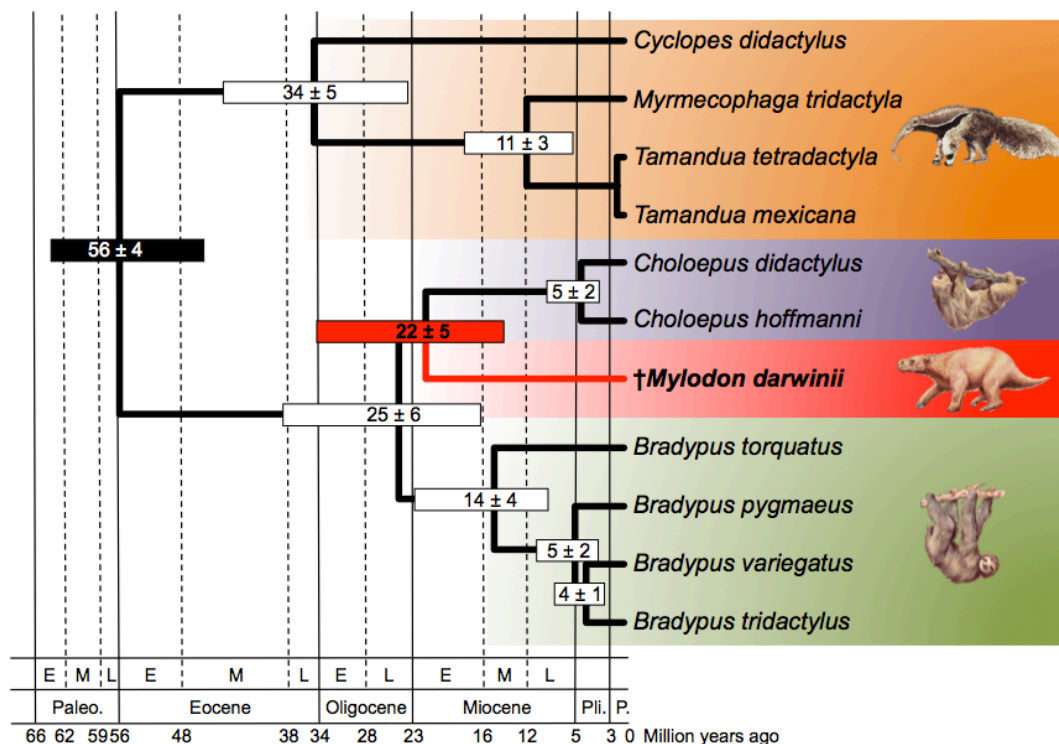


Figure 5 | Phylogénie et cadre temporel pour l'histoire évolutive des Pilosa incluant le paresseux terrestre éteint *Mylodon darwinii*. La phylogénie et les datations moléculaires ont été inférées à partir de mitogénomes complets analysés sous un modèle de mélange CAT-GTR+G avec le logiciel PhyloBayes.

Néanmoins, si l'ADN ancien a le potentiel pour résoudre un grand nombre de questions évolutives, il reste souvent extrêmement difficile d'obtenir assez d'ADN endogène à partir de spécimens fossiles contenant une très grande majorité d'ADN contaminant provenant de l'environnement. Ainsi, dans l'immense

majorité des cas, l'approche par séquençage shotgun est économiquement prohibitive puisque seule une infime minorité des lectures produites correspondra à l'organisme ciblé. Dans la plupart des cas, il faut donc utiliser des méthodes de capture d'ADN permettant de cibler spécifiquement l'ADN endogène au sein de l'extrait initial grâce des sondes d'ADN ou d'ARN complémentaire marquées à la biotine (Briggs et al. 2009). Ces sondes qui permettent d'enrichir l'extrait à séquencer en séquences d'intérêt sont en général définies à partir de séquences d'espèces actuelles supposées proches du subfossile à séquencer. La température d'hybridation des sondes laisse une certaine marge de manœuvre quant au degré de divergence autorisé entre les sondes et les séquences ciblées, mais il est toujours avantageux de définir des sondes les plus spécifiques possibles afin d'accrocher un maximum de fragments cibles tout en évitant les séquences contaminantes. Dans le cas de taxons fossiles d'affinités phylogénétiques incertaines, il peut être intéressant de définir les sondes non pas sur des séquences actuelles mais sur des séquences ancestrales reconstruites bioinformatiquement. C'est cette stratégie que nous avons appliquée au cas du glyptodon *Doedicurus*.

Depuis Charles Darwin, qui a été l'un des premiers scientifiques à en collecter les restes fossiles en Amérique du Sud, les glyptodons n'ont cessé de fasciner les naturalistes. Leur caractéristique distinctive, au delà de leur taille imposante pouvant aller jusqu'à plus d'une tonne, est leur carapace composée d'ostéodermes fusionnés (Figure 6). Cette particularité les rapproche des tatous qui sont les seuls mammifères dans la nature actuelle à posséder une carapace. Néanmoins chez ces derniers, la carapace est articulée en bandes mobiles, alors qu'elle est constituée d'un seul bloc rigide chez les glyptodons. En raison de nombreuses différences morphologiques, les glyptodons ont longtemps été considérés comme un groupe distinct constituant le groupe-frère des tatous au sein des xénarthres (Hoffstetter 1958; Engelman 1985). Néanmoins, une étude phylogénétique récente basée sur l'analyse cladistique de caractères cranio-dentaires a plus récemment suggéré une position phylogénétique plus interne au sein des tatous (Gaudin & Wible 2006). Afin de résoudre la position phylogénétique des glyptodons, nous avons étudié un fragment de carapace attribué au genre *Doedicurus* conservé au Musée Argentin des Sciences Naturelles à Buenos Aires (Argentine) et daté à environ 12 000 ans par le carbone 14. *Doedicurus* était l'un des plus gros glyptodons avec une masse corporelle estimée à 1,5 tonne. L'une de ses particularités était sa queue en forme de massue équipée de pointes, susceptible d'être utilisée dans la défense contre ses prédateurs comme les tigres à dents de sabre, mais aussi probablement dans des combats intraspécifiques.



Figure 6 | Spécimens de glyptodons exposés au Musée de La Plata (Argentine).
De gauche à droite: *Glyptodon*, *Panochthus* et *Doedicurus*. © Museo de La Plata.

Pour enrichir sélectivement notre extrait initial obtenu à partir d'un ostéoderme en ADN mitochondrial endogène de glyptodon, nous avons défini des sondes à partir de séquences ancestrales inférées par maximum de vraisemblance en utilisant l'arbre phylogénétique et les mitogénomes des espèces de Xénarthres actuels précédemment obtenus (**Gibb et al. 2016**). Ces séquences ancestrales ont ensuite servi à synthétiser des sondes d'ARN permettant de cibler plus spécifiquement les fragments d'ADN mitochondrial de Xénarthres dans l'extrait d'ADN ancien. Le séquençage des dizaines de milliers de fragments ainsi récoltés a ensuite permis de reconstruire le génome mitochondrial quasiment complet de ce taxon éteint (**Delsuc et al. 2016**). Les résultats obtenus avec les sondes définies à partir de séquences ancestrales montrent que notre procédure permet la capture d'ADN mitochondrial endogène de *Doedicurus* avec une spécificité accrue par rapport aux sondes conçues à partir des espèces actuelles (Figure 7).

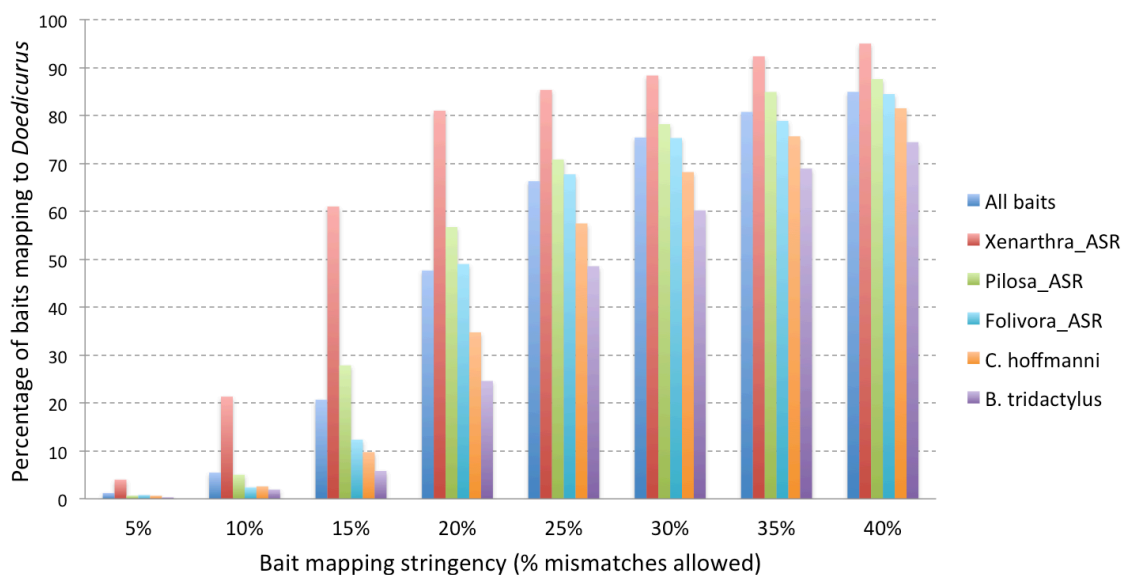


Figure 7 | Spécificité de la capture pour différents jeux de sondes. La spécificité est exprimée en pourcentage des sondes qui mappent sur le mitogénome de *Doedicurus* à différents seuils (de 5% à 40% de mésappariements autorisés).
ASR : Reconstruction de séquences ancestrales.

Comme on peut le voir, le pourcentage de sondes mappant sur le mitogénome de *Doedicurus* à des seuils tolérant de 5% à 40% de mésappariements est toujours plus élevé pour les sondes issues de séquences ancestrales. Cela est particulièrement évident pour conditions de mapping les plus strictes (5% à 20%) avec lesquelles les pourcentages de mapping les plus élevés sont obtenus avec des sondes conçues à partir du mitogénome ancestral des Xénarthres. Par exemple, lorsque 15% de mésappariements sont autorisés, 61% des sondes conçues à partir du mitogénome ancestral des Xénarthres mappent, contre seulement 6% des sondes définies à partir du mitogénome de *Bradypus tridactylus*. Même s'il a été démontré que les sondes de spécificité faibles (jusqu'à 40% de divergence) pouvaient être relativement efficaces pour capturer des régions cibles (Li et al. 2013), notre procédure basée sur la reconstruction de séquences ancestrales semble améliorer la spécificité de capture.

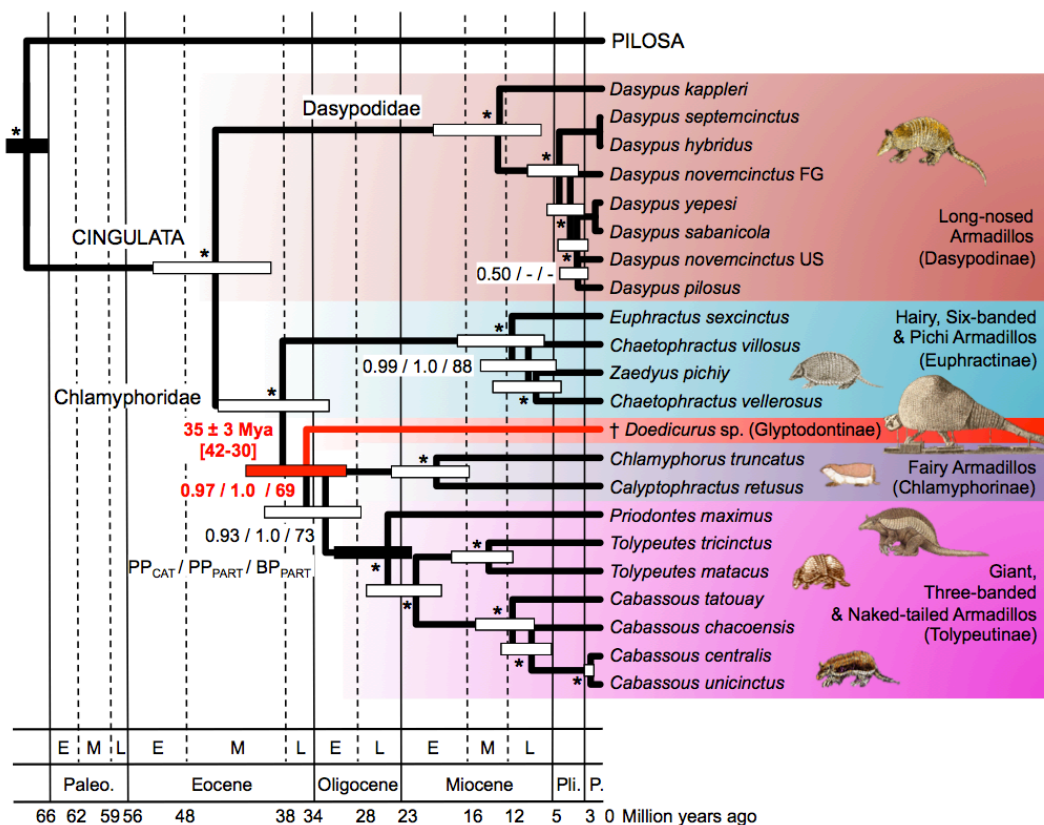


Figure 8 | Phylogénie et cadre temporel pour l'histoire évolutive des tatous incluant le glyptodon éteint *Doedicurus* (Delsuc et al. 2016). La phylogénie et les datations moléculaires ont été inférées à partir de mitogénomes complets analysés sous un modèle de mélange CAT-GTR+G avec le logiciel PhyloBayes.

Les analyses phylogénétiques incluant le mitogénome de *Doedicurus* avec toutes les espèces de xénarthres actuelles ont permis de placer de façon non ambiguë les glyptodons au sein des tatous (Figure 8), les rapprochant d'un groupe constitué des Chlamyphorinés (tatous nains) et des Tolypeutinés (tatous à trois bandes, tatous à queue-nue, et tatous géants) dont ils ont divergé il y a environ 35 millions d'années (Delsuc et al. 2016). Ces résultats démontrent que les glyptodons représentent une lignée éteinte de Tatou qui a subi une augmentation spectaculaire de la taille jusqu'à son extinction il y a environ 10 000 ans. Leur

position phylogénétique à l'intérieur des tatous actuels implique également que leur carapace rigide est en fait un caractère dérivé plutôt qu'une caractéristique ancestrale comme supposé précédemment. Du point de vue la classification, les glyptodons devraient donc être plutôt considérés comme une sous-famille éteinte de tatous (les Glyptodontinae), et non pas une super-famille distincte comme c'est le cas actuellement. Cette étude illustre le potentiel des méthodes de capture d'ADN ancien basées sur des sondes définies à partir de séquences ancestrales pour reconstituer l'histoire évolutive des nombreuses espèces éteintes récemment et dont on dispose d'un grand nombre de restes sub-fossiles.

2.2.3 MITOGENOMIQUE DES CHAUVES-SOURIS PHYLLOSTOMIDES

La mitogénomique s'impose aujourd'hui grâce aux nouvelles technologies de séquençage comme une approche simple permettant d'obtenir un échantillonnage taxonomique exhaustif de différents groupes de mammifères. Nous avons ainsi mis en place au laboratoire un protocole « low-cost » pour la préparation de bibliothèques Illumina pour séquençage shotgun (**Tilak et al. 2015**). Ces développements techniques nous permettent désormais d'obtenir en routine et à coût réduit de nombreux mitogénomes de mammifères. C'est à partir de cette approche qu'a été menée la thèse de Fidel Botero-Castro soutenue en 2014 consacrée à la diversification des chauves-souris de la famille des Phyllostomidae qui comprend notamment les vampires. Cette famille de chauves-souris néotropicales très diversifiée, avec plus de 160 espèces décrites couvrant une diversité de traits d'histoire de vie, est un modèle de radiation évolutive. L'approche mitogénomique mise en place (**Botero-Castro et al. 2013, 2016**), nous a permis d'obtenir plus d'une centaine de nouveaux mitogénomes sur la base desquels, une phylogénie de référence a été reconstruite au niveau des sous-familles et des genres afin de permettre d'étude des processus de diversification sous-jacents. Ces données ont également révélé l'évolution particulière des mitogénomes de vampires qui présentent un taux d'évolution élevé résultant à la fois de biais mutationnels et d'épisodes de sélection positive, possiblement liés à leur régime strictement hématoophage (Figure 9).

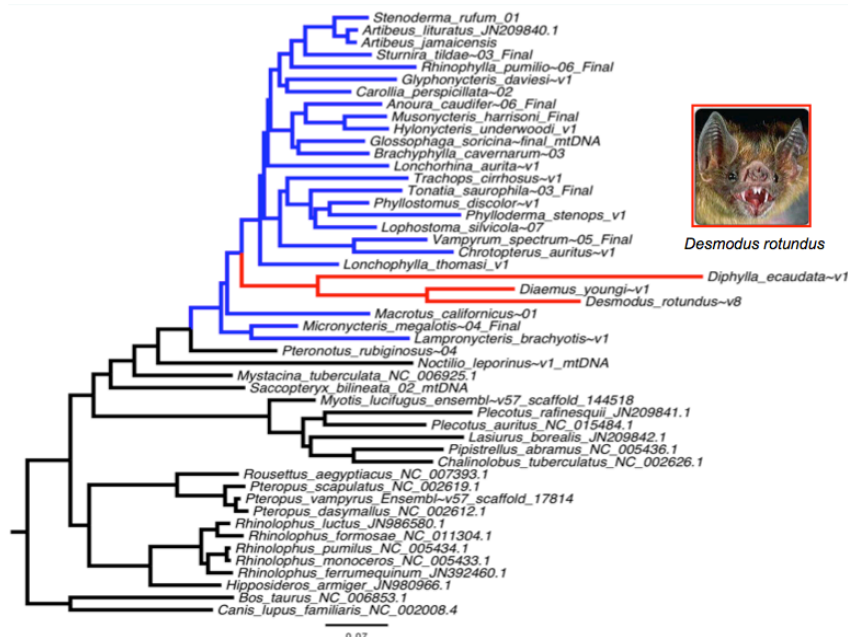


Figure 9 | Phylogénie des mitogénomes de chauves-souris. L'accélération du taux d'évolution observée dans les séquences protéiques des vampires est figurée en rouge.

2.3. PHYLOGENOMIQUE POUR LA RESOLUTION DE GRANDES QUESTIONS EVOLUTIVES

2.3.1 LA POSITION PHYLOGENETIQUE DES TORTUES

La position phylogénétique des tortues est une des grandes questions évolutives qui restait à résoudre dans le contexte de l'évolution des vertébrés et plus précisément des amniotes. Traditionnellement considérées comme le groupe-frère des autres lignées de reptiles (incluant les oiseaux) sur la base de l'absence de fenêtre temporelle considérée comme un caractère ancestral des Amniotes, les tortues se sont avérées particulièrement difficiles à placer dans la phylogénie. Ainsi, les analyses des données morphologiques (Harris et al. 2007) et moléculaires (Zardoya & Meyer 2001) ont abouti à des hypothèses phylogénétiques contradictoires proposant tour à tour les tortues comme : 1) le groupe-frère de l'ensemble des autres reptiles (squamates, sphénoïdons, oiseaux, et crocodiles), 2) le groupe-frère des lépidosaures (sphénoïdons et squamates), 3) le groupe-frère des archosaures (oiseaux et crocodiles), et enfin 4) le groupe-frère des crocodiles. Pour répondre à cette question, nous avons adoptée une approche phylogénomique basée sur le séquençage 454 des transcriptomes de quatre espèces de tortues (*Emys orbicularis*, *Chelonoidis nigra*, *Phrynops hillarii*, et *Caretta caretta*), d'une espèce de lézard (*Podarcis* sp.), d'un caïman (*Caiman crocodylus*), et d'un dipneuste (*Protopterus annectens*). Ces données nous ont permis d'assembler un jeu de données constitué de 248 gènes protéiques pour 16 taxons de vertébrés (**Chiari et al. 2012**).

Les analyses phylogénétiques de cette supermatrice ont abouti à des résultats contrastés selon les caractères (acides aminés versus nucléotides) et les modèles utilisés (site-homogènes versus site-hétérogènes). La Table 1 montre les différentes hypothèses soutenues selon les modèles et les données utilisées. Les analyses des acides aminés avec différents modèles soutiennent toujours très fortement l'hypothèse Tortues + Archosaures. Au contraire, les analyses des séquences nucléotidiques incluant les 3 positions du codon sous un modèle simple ou un modèle partitionné par gène soutiennent plutôt l'hypothèse Tortues + Crocodiles. En fait, dans le cas des nucléotides incluant les trois positions du codon, seuls les modèles partitionnés par codon et le modèle de mélange CAT-GTR+G retrouvent l'hypothèse Tortues + Archosaures avec un soutien statistique maximal. Lorsque les troisièmes positions du codon, qui sont très saturées dans ce jeu de données basés sur des gènes conservés, sont exclues des analyses, tous les modèles y compris les modèles simples soutiennent l'hypothèse Tortues + Archosaures. Au contraire, les analyses basées sur ces seules troisièmes positions retournent un soutien maximal pour l'hypothèse Tortues + Crocodiles, sauf le modèle CAT-GTR+G qui soutient à nouveau l'hypothèse Tortues + Archosaures avec une probabilité postérieure maximale. Ces observations suggèrent fortement que le soutien observé pour l'hypothèse Tortues + Crocodiles est un artéfact de reconstruction dû principalement à l'incapacité des modèles site-homogènes de gérer l'hétérogénéité introduite par la saturation substitutionnelle des troisièmes positions. Ces résultats soulignent l'importance d'utiliser le modèle site-hétérogène CAT-GTR+G (Lartillot et al. 2007) qui accomode beaucoup mieux les substitutions multiples que les modèles site-homogènes ou même qu'un modèle partitionné par gènes dans ce cas précis.

Table 1 | Résultats des analyses de différentes supermatrices sous différents modèles.

	Amino acids	Nucleotides		
	All positions	All positions	Positions 1+2	Positions 3
Total sites	62,342	187,026	124,684	62,342
Constant sites	41,170 (66.0%)	99,638 (53.3%)	92,128 (73.9%)	7,510 (11.2%)
Informative sites	8,749 (14.0%)	54,880 (29.3%)	14,009 (11.2%)	40,871 (65.6%)
RaxML LG+G / GTR+G	Turtles + Archosaurs BP _{ML} = 100	Turtles + Crocodiles BP _{ML} = 76	Turtles + Archosaurs BP _{ML} = 100	Turtles + Crocodiles BP _{ML} = 100
RaxML GTR+G partitioned by gene	Turtles + Archosaurs BP _{PARTG} = 100	Turtles + Crocodiles BP _{PARTG} = 54	-	-
RaxML GTR+G partitioned by codon	-	Turtles + Archosaurs BP _{PARTC} = 100	-	-
MrBayes WAG+G / GTR+G	Turtles + Archosaurs PP _{BAY} = 1.0	Turtles + Crocodiles PP _{BAY} = 1.0	Turtles + Archosaurs PP _{BAY} = 1.0	Turtles + Crocodiles PP _{BAY} = 1.0
MrBayes GTR+G partitioned by codon	-	Turtles + Archosaurs PP _{PARTC} = 1.0	-	-
PhyloBayes CAT-GTR+G	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0

Nos analyses phylogénomiques ont donc montré avec robustesse que les tortues constituent le groupe-frère des Archosaures (crocodiles + oiseaux) au sein des amniotes (**Chiari et al. 2012**). Ces résultats, qui ont été confirmés indépendamment par l'étude des éléments ultra-conservés (UCEs ; Crawford et al. 2012), ont une importance capitale pour l'interprétation de l'évolution des caractères morphologiques chez les amniotes tels que l'absence de fenêtrage temporel qui résulte très probablement d'une perte secondaire chez les tortues (Figure 10). Ce nouveau cadre phylogénétique de référence établi pour les Amniotes a enfin des implications pour comprendre les origines évolutives et développementales de la carapace des tortues (Wang et al. 2013).

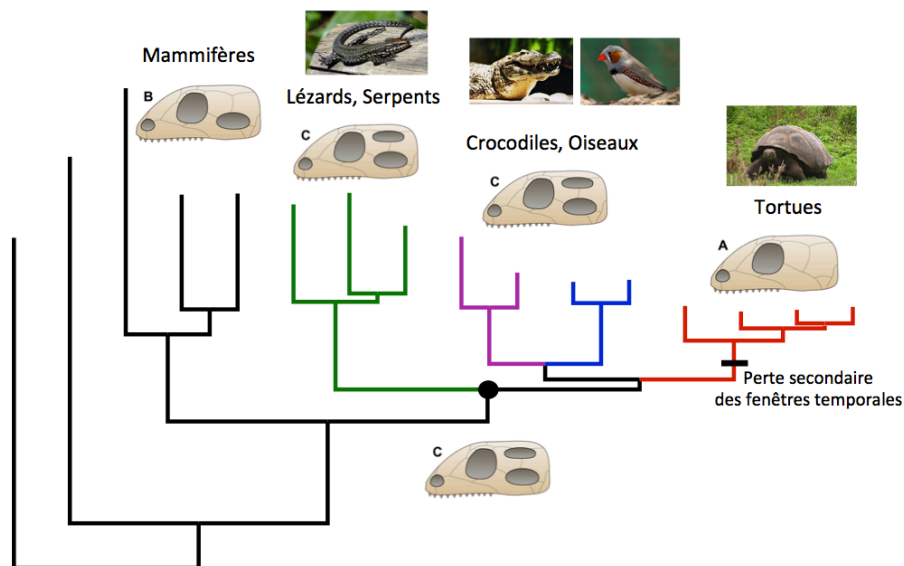


Figure 10 | Phylogénie des Amniotes et évolution des types crâniens : A) Anapside, B) Synapside et C) Diapside. Le phylogramme présenté a été obtenu par l'analyse Bayésienne des séquences protéiques d'une supermatrice de 248 gènes sous le modèle CAT-GTR+G avec PhyloBayes.

2.3.2 LA RACINE DES PLACENTAIRES

La phylogénie des Mammifères placentaires a été révolutionnée par l'arrivée des données moléculaires au début des années 2000 qui ont permis de reconnaître quatre clades majeurs : Afrothériens, Xénarthres, Laurasiathériens et Euarchontoglires (Murphy et al. 2001 ; **Delsuc et al. 2002**). De plus l'existence d'un super-groupe originaire de l'hémisphère Nord regroupant Laurasiathériens et Euarchontoglires en Boreoethériens a été mise en évidence par le séquençage de gènes nucléaires. Cependant, la question de la position de la racine de l'arbre des mammifères placentaires dont dépendent les relations entre ces trois groupes demeure un point problématique. En effet, les données issues de l'analyse des patrons d'insertion d'éléments transposables suggèrent même l'existence d'une trifurcation due à l'occurrence probable d'événements de tri de lignées incomplet suite à la séparation rapide des trois branches descendantes à cet endroit de l'arbre (Churakov et al. 2009 ; Nishihara et al. 2009).

Cette question phylogénétique majeure peut désormais être testée en utilisant des jeux de données issus de génomes complets. Cependant, les études basées sur la concaténation des gènes orthologues simple-copies ont abouti à des résultats contradictoires soit au regroupement des Xénarthres et Afrothériens, l'hypothèse Atlantogenata (Hallström et al. 2007 ; Wildman et al. 2007) soit à l'impossibilité de conclure à cause du tri incomplet des lignées (Hallström & Janke 2010). Par ailleurs, l'application de méthodes modélisant explicitement l'incongruence entre les arbres de gènes due au tri incomplet des lignées, aux gènes orthologues simple-copies (Song et al. 2012) et aux UCEs (McCormack et al. 2012) ont également soutenu des conclusions opposées.

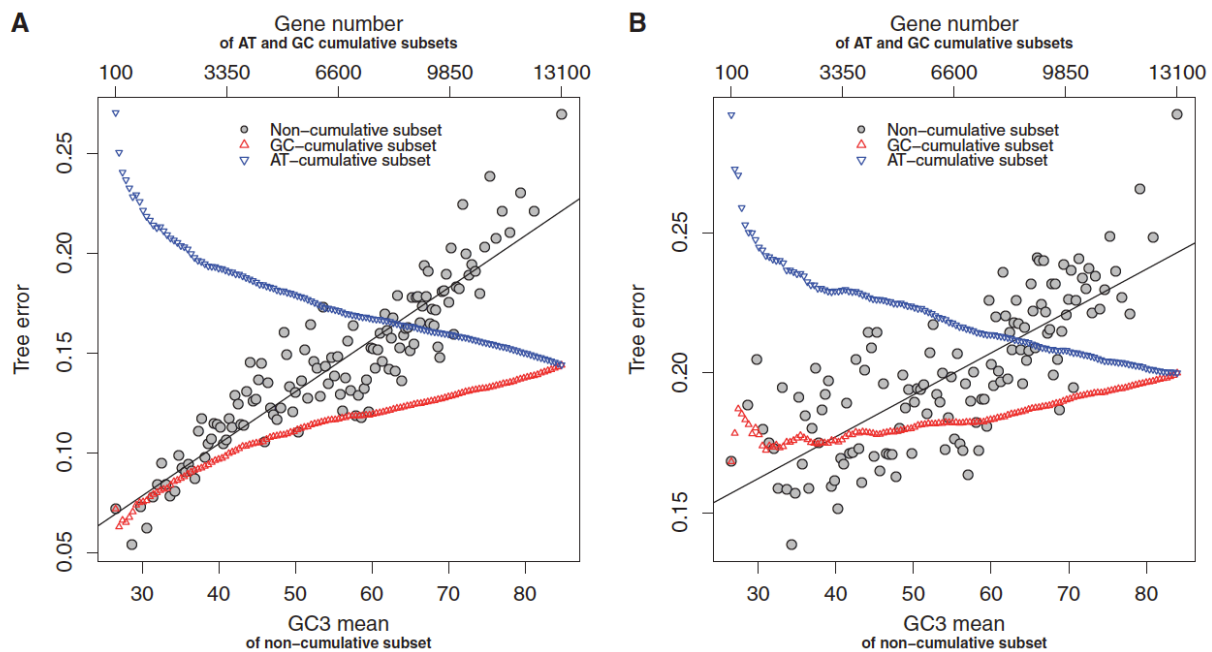


Figure 1 | Effet du GC sur l'erreur de reconstruction des arbres de gènes. A)

Données brutes. B) Données modifiées pour contrôler les données manquantes. Chaque cercle représente 100 alignements positionnés en fonction de leur moyenne en GC3 (axe X) et leurs erreurs topologiques moyennes (axe Y). Les triangles représentent des sous-ensembles cumulatifs, où l'axe X est le nombre de gènes du sous-ensemble. Les sous-ensembles cumulatifs GC contiennent un nombre croissant de gènes riches en GC, les sous-ensembles cumulatifs AT comptent un nombre croissant de gènes AT-riches.

Nous avons ainsi exploré la quarantaine de génomes de mammifères disponibles en développant une approche permettant d'identifier des gènes les plus pertinents pour répondre à la question. Nous avons ainsi pu montrer qu'en raison de phénomènes liés à la dynamique de recombinaison dans les génomes de mammifères, les gènes riches en GC sont particulièrement susceptibles aux artéfacts de reconstruction phylogénétiques (Figure 11). Les analyses phylogénétiques basées sur les gènes riches en AT nous ont ainsi permis de mettre en évidence la position basale des Afrotheria par rapport aux autres groupes (**Romiguier et al. 2013**). Ce résultat, en accord avec l'analyse des UCEs (McCormack et al. 2012), est très intéressant de par son explication biologique et a des conséquences importantes pour l'interprétation des scénarii à l'origine de la diversification des Placentaires et la nature même de l'ancêtre de ce groupe.

2.4. EXEMPLES D'APPLICATIONS DES PHYLOGENIES

En tant que phylogénéticien, je suis fréquemment sollicité pour analyser les séquences obtenues par des collègues travaillant dans des domaines connexes telles que la biologie du développement, la biologie moléculaire et cellulaire, ou encore la virologie. Certaines de ses sollicitations se sont transformées en collaborations fructueuses pour lesquelles l'approche évolutive s'est avérée cruciale.

2.4.1 ORIGINE DES GENES HOX CHEZ LES ANIMAUX

Ce projet est issu d'une collaboration avec les Pr. Daniel Chourrout du SARS International Center for Marine Molecular Biology de l'Université de Bergen (Norvège) et Ulrich Technau, aujourd'hui à l'Université de Vienne (Autriche), qui m'ont proposé d'analyser des données issues du séquençage des génomes de deux Cnidaires modèles, l'hydre (*Hydra magnipapillata*) et l'anémone de mer (*Nematostella vectensis*). Le répertoire de gènes Hox de ces deux génomes a ainsi été identifié en couplant les approches bioinformatiques et expérimentales. Nous avons ensuite analysé ces données du point de vue phylogénétique dans la perspective d'éclairer les origines des gènes Hox chez les animaux sur la base des séquences conservées de leur Homeodomain. En effet, il existe plusieurs hypothèses concernant l'origine évolutive des gènes Hox basées sur des scénarios de duplications à partir de gènes ancestraux, le cluster ProtoHox (Garcia-Fernandez 2005). Ainsi, la caractérisation et l'analyse phylogénétique du répertoire de gènes Hox n'avaient encore jamais été réalisées chez les Cnidaires qui sont des organismes à deux feuillet embryonnaires (diploblastes), alors que l'on connaît bien leur organisation et leur évolution chez les animaux bilatériens à trois feuillet embryonnaires (triploblastes).

Les analyses phylogénétiques, notamment basées sur des réseaux reconstruits à partir des distances estimées en maximum de vraisemblance (Figure 12), nous ont permis de montrer que les répertoires de gènes Hox des Cnidaires ont suivi leur propre trajectoire évolutive au cours de l'évolution des Cnidaires avec des expansions notamment chez *Nematostella*, et des pertes chez *Hydra*, qui sont significatives. En effet, les deux génomes de Cnidaires présentent des gènes Hox clairement homologues des gènes antérieurs des Bilatériens, mais également des Gènes Hox qui ne présentent pas d'affinités claires avec les gènes centraux ou postérieurs des Bilatériens. De plus, des gènes homologues des gènes ParaHox Gsx et Xlox/Cdx ont été identifiés dans le génome de *Nematostella*. Ces résultats nous ont permis de proposer un scénario d'évolution simple pour l'origine du

système Hox à partir d'un ensemble ProtoHox ne contenant que deux gènes ancestraux, alors que les scénarios proposés jusque là en inféraient au moins trois (**Chourrout et al. 2006**). Les relations d'orthologie des différentes familles de gènes Hox entre Cnidaires et Bilatériens restent cependant difficiles à établir (Thomas-Chollier et al. 2010 ; DuBuc et al. 2012). Le séquençage de nouveaux génomes a depuis montré que de façon surprenante les gènes Hox étaient absents des autres phylums d'animaux diploblastes tels que les Spongiaires, les Cténophores et les Placozoaires (Holland 2013). Ceci complexifie encore l'élaboration du scénario à l'origine des gènes Hox chez les Métazoaires, d'autant plus que les relations phylogénétiques entre ces différents Phylums restent empreintes d'incertitude (Telford et al. 2016).

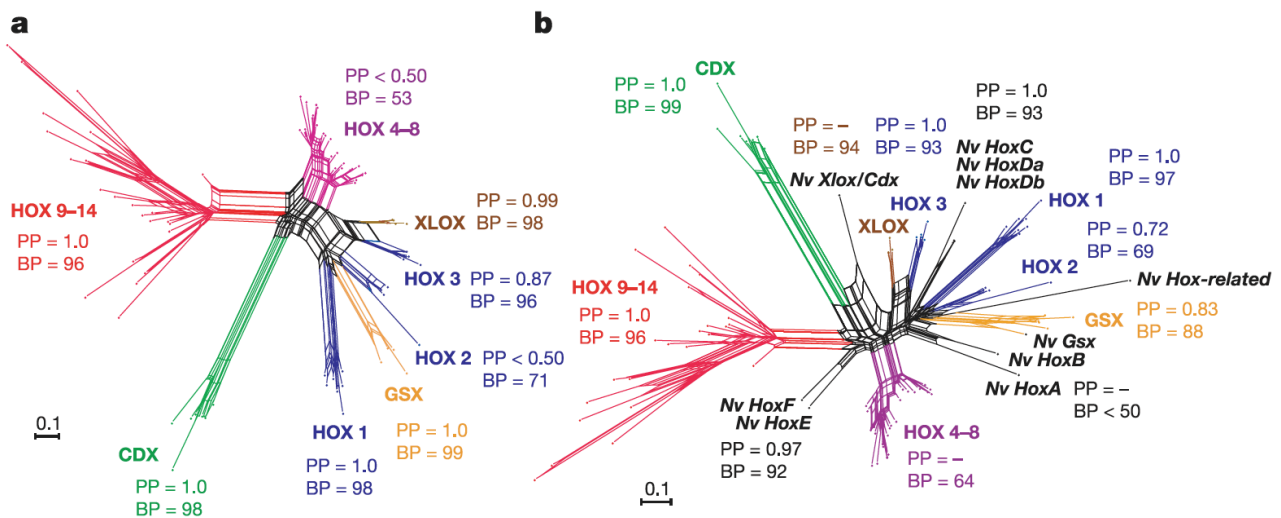


Figure 2 | Analyses phylogénétiques des gènes HOX des Métazoaires basées sur les séquences de leurs Homéodomains. Réseaux phylogénétiques reconstruits par la méthode Neighbour-net à partir des distances ML estimées sous un modèle JTT+G8 pour : a) 82 gènes Hox et ParaHox de bilatériens, et b) 92 gènes Hox et ParaHox de bilatériens incluant ceux du Cnidaire modèle *Nematostella vectensis* (Nv). Les probabilités postérieures Bayésiennes (PP) et les pourcentages de bootstrap (BP) sont indiqués pour les groupes majeurs.

2.4.2 EVOLUTION MOLECULAIRE DES FACTEURS DE RESTRICTION DES VIRUS HIV

L'identification de gènes de l'immunité chez l'homme capables de contrecarrer l'infection par des virus comme le VIH (appelés facteurs de restriction) est un domaine de recherche particulièrement actif avec de nombreuses perspectives thérapeutiques (Harris et al. 2012). Jusqu'en 2011, seuls quatre facteurs de restriction du VIH avaient été identifiés. Un cinquième facteur codé par le gène SAMHD1 (Laguet et al. 2011) a été découvert par une équipe de l'Institut de Génétique Humaine (IGH) de Montpellier dirigée par Moncef Benkirane. Les analyses évolutives comparatives d'évolution moléculaire pouvant apporter une aide précieuse à la caractérisation de ces facteurs et leur mode d'action moléculaire, une nouvelle collaboration m'a été proposée par cette équipe. En effet, l'évolution moléculaire des facteurs de restriction de virus est caractérisée par une coévolution entre ces gènes et les gènes viraux antagonistes. Cette véritable course aux armements entre hôte et virus laisse des signatures détectables par des analyses d'évolution moléculaire de ces gènes qui portent les traces d'épisodes de sélection positive récurrents (Duggal & Emerman 2012).

Nous avons ainsi montré que l'évolution de SAMHD1 a été dictée par son interaction avec la protéine virale antagoniste *vpx* que l'on retrouve chez le virus HIV2 humain et chez certains virus SIV de primates (**Laguette et al. 2012**). Notre analyse détaillée des pressions de sélection ayant affectées SAMHD1 au cours de l'évolution des primates a permis de déterminer l'âge probable d'apparition des premières pressions sélectives correspondant à l'origine de l'infection des primates par les lentivirus et l'identification de sites dans la molécule SAMHD1 qui lui confèrent son activité de restriction du virus. Nos résultats ont été confirmés de manière indépendante par une publication conjointe parue dans la même issue de *Cell Host & Microbe* (Lim et al. 2012). Suite à cette publication, l'application d'une nouvelle méthode Bayésienne de caractérisation des pressions de sélection le long des branches d'un arbre phylogénétique développée par Nicolas Lartillot, nous a permis de comparer les détails des pressions de sélection qui ont gouverné l'évolution des principaux facteurs de restriction du HIV au cours de l'évolution des primates (Figure 13). Ces résultats ont des implications potentielles pour le développement de futurs vaccins contre le VIH basés sur ces facteurs de restriction. Nous étendons désormais ces analyses évolutives à d'autres gènes potentiellement impliqués dans la détection des infections virales chez les mammifères tels que les gènes de la voie RNAi comme DICER1 et DROSHA. Ce projet est mené en collaboration entre l'ISEM et l'IGH au travers du co-encadrement d'une post-doctorante (Nadia Rahm).

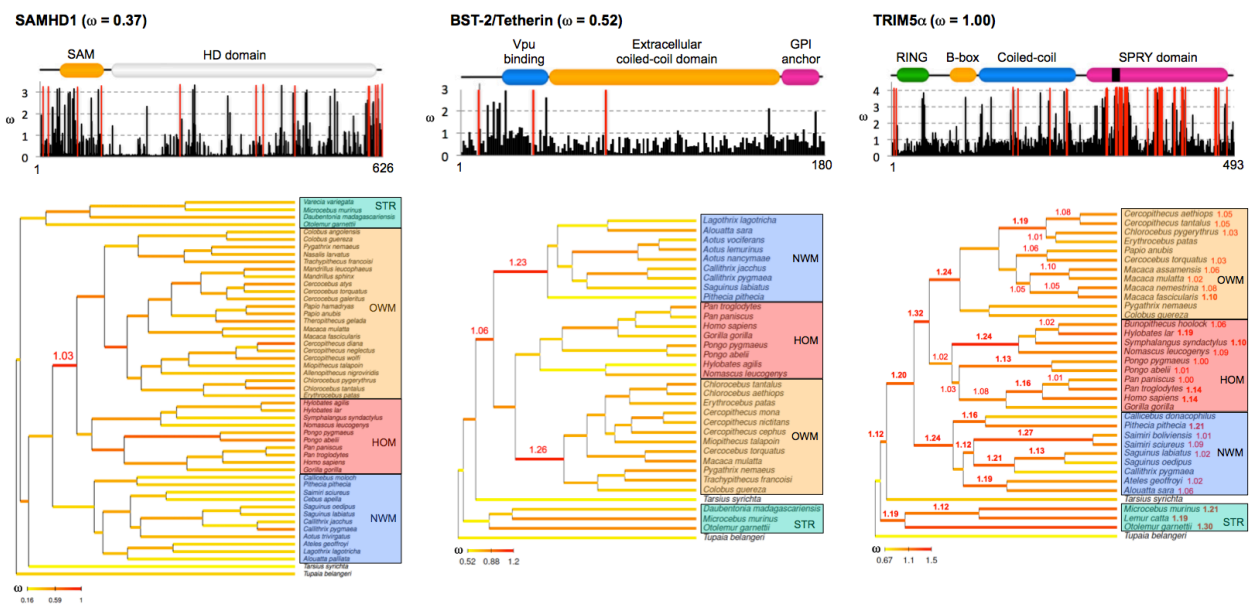


Figure 3 | Evolution moléculaire des facteurs de restriction des virus chez les Primates. Haut : Distribution estimée des valeurs de dN/dS le long des séquences des gènes SAMHD1, BST-2/Tetherin, et TRIM5 α chez les primates. Les positions sous sélection positive ($\omega > 1$) sont indiquées en rouge et les domaines structuraux sont figurés. Bas : Reconstruction de la variation du dN/dS le long de la phylogénie sous un modèle Bayésien utilisant une prior hiérarchique (i.e. valeurs de ω par branche distribuées selon une loi gamma de moyenne et variance estimées sur les données).

2.4.3 EVOLUTION DES GENES CCT CHEZ LES PARASITES APICOMPLEXES

La phosphatidylcholine est le composant lipidique majeur des membranes du parasite responsable du paludisme (*Plasmodium falciparum*). Ce composé est

essentiel pour la multiplication du parasite dans les globules rouges humains en phase infectieuse. La CTP: phosphocholine cytidyltransferase (CCT) qui est l'enzyme limitante de la voie de biosynthèse de la phosphatidylcholine, constitue donc une cible potentielle pour le développement de molécules antipaludiques. C'est sur cette problématique que travaille Rachel Cerdan du laboratoire Dynamique des Interactions Membranaires Normales et Pathologiques à l'Université de Montpellier qui m'a proposé d'étudier l'évolution de ce gène qui possède une structure particulière chez *Plasmodium falciparum*.

En effet, contrairement à ses orthologues chez les mammifères, le gène codant pour l'enzyme CCT chez *Plasmodium falciparum* contient un domaine cytidyltransferase dupliqué avec des domaines de liaison à la membrane. En reconstruisant la phylogénie des séquences de CCT disponibles pour les autres parasites Apicomplexes (**Nagy et al. 2013**), nous avons pu montrer que cette duplication de domaines est en fait propre aux genres *Plasmodium*, *Babesia* et *Theileria* (Aconoidasida), alors que les genres *Cryptosporidium*, *Toxoplasma* et *Neospora* (Coccidia) comportent un seul domaine cytidyltransferase, comme les autres Eucaryotes (Figure 14a). De plus, en analysant les domaines dupliqués entre eux, nous avons pu inférer l'existence de deux événements de duplication indépendants ayant abouti à la présence de ces deux domaines chez l'ancêtre de *Plasmodium* d'une part, et chez l'ancêtre commun à *Babesia* et *Theileria* d'autre part (Figure 14b). Une étude fonctionnelle subséquente a permis de démontrer que les deux domaines dupliqués sont catalytiquement actifs chez *Plasmodium falciparum* (Contet et al. 2015). Ce travail a également débouché sur l'identification d'une molécule entraînant l'inactivation de l'enzyme CCT chez le parasite, ouvrant ainsi la porte à la conception de futurs agents antipaludiques.

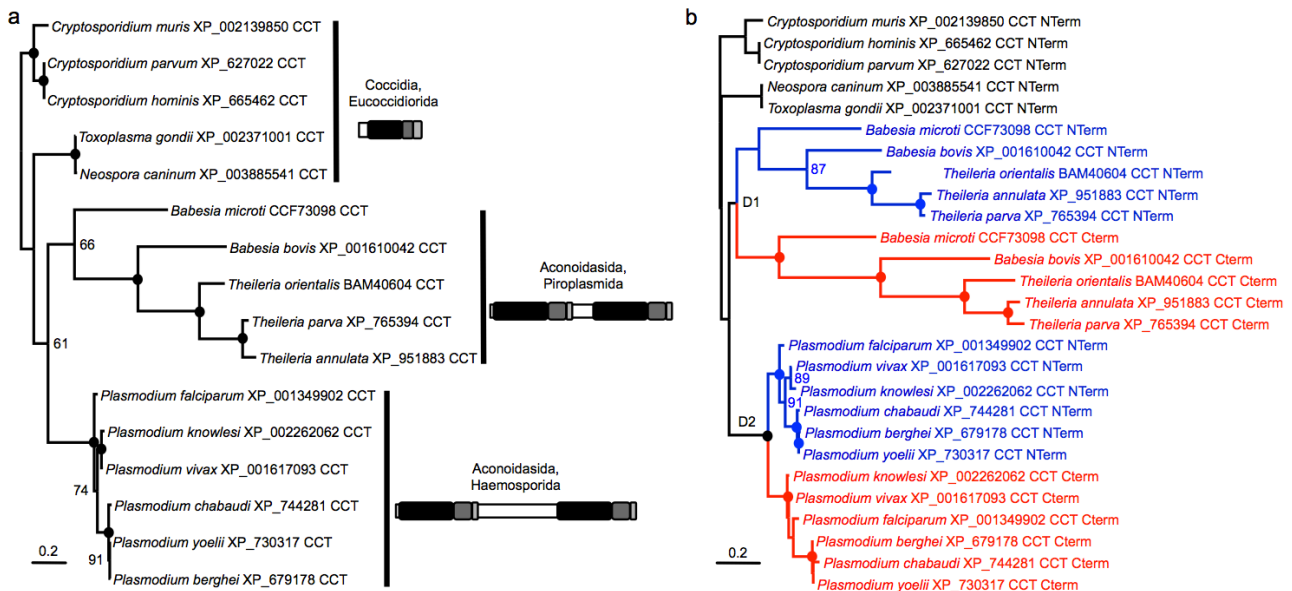


Figure 4 | Analyse phylogénétique des gènes CCT et de leurs domaines structuraux chez les Apicomplexes. a) Arbre de maximum de vraisemblance des séquences protéiques complètes de CCT pour 16 espèces d'apicomplexes parasites. La structure des gènes CCT avec l'emplacement des segments CMP contenant les domaines dupliqués est représentée pour les groupes majeurs d'apicomplexes. b) Arbre de maximum de vraisemblance de 27 segments protéiques CMP du gène CCT. D1 et D2 indiquent les deux événements de duplication indépendants des segments C-terminaux et N-terminaux inférés à partir de cette topologie. Les valeurs aux nœuds représentent les pourcentages de bootstrap (BP), les cercles indiquant les nœuds statistiquement soutenus (BP>95).

2.4.4 ORIGINE ET EVOLUTION DES REGULATEURS DU PROTEASOME CHEZ LES EUKARYOTES

Tous les organismes vivants ont besoin de dégrader des protéines afin de mettre fin à des processus biologiques et d'éliminer les protéines endommagées dans les cellules. Une telle machinerie est représentée par le protéasome 20S, une protéase multi-catalytique et compartimentée en forme de « tonneau » (Finley et al. 2016). L'activation de ce système nécessite généralement la liaison de régulateurs/activateurs du protéasome qui contrôlent l'entrée des substrats. Ces derniers incluent PA700 (ou complexe 19S) qui s'assemble avec le protéasome 20S pour former le protéasome 26S permettant la dégradation de protéines marquées par ubiquitinylation, PA200 and PA28 qui sont impliqués dans la protéolyse ubiquitine-indépendante, et PI31 qui a été initialement identifié comme un inhibiteur du protéasome 20S in vitro (Voges et al. 1999). Contrairement au protéasome 20S, présent chez tous les Eucaryotes et les Archées, l'histoire évolutive des activateurs du protéasome est mal connue.

A l'invitation de Philippe Fort et Olivier Coux du Centre de Recherche en Biochimie Macromoléculaire (CRBM) du CNRS à Montpellier, j'ai ainsi participé à l'analyse phylogénétique des quatre types de régulateurs du protéasome dont les séquences sont disponibles dans les génomes Eucaryotes (PA700, PA200, PA28 et PI31). Nos résultats montrent que les régulateurs de protéasome actuellement connus sont largement répartis dans les principaux groupes d'Eucaryotes, ce qui permet d'établir qu'ils étaient vraisemblablement déjà tous présents chez le dernier ancêtre commun Eucaryote (LECA ; **Fort et al. 2015**). Nous avons révélé une remarquable conservation des sous-unités de PA700 chez les Eucaryotes étudiés, ce qui suggère que le complexe actuel était déjà en place chez le LECA. Bien que probablement présents également chez le LECA, PA200, PA28 et PI31 montrent une évolution plus dynamique marquée par de nombreuses pertes secondaires dans plusieurs lignées majeures (Figure 15).

La présence vraisemblable de PA700, PA200, PA28 et PI31 dans tous les groupes Eucaryotes suggère que ces régulateurs remplissent les fonctions de base de la physiologie cellulaire, telles que la réparation de l'ADN, le contrôle du cycle cellulaire et l'apoptose. Compte tenu de l'importance primordiale de ces fonctions physiologiques, les pertes sélectives de PA200 chez les insectes Brachycères et de PA28 chez champignons Ascomycètes soulève la question de savoir si la levure et la drosophile sont des modèles adaptés pour étudier les fonctions de ces régulateurs chez les mammifères. Ainsi, en dépit de l'origine vraisemblablement très ancienne de tous ces régulateurs chez les Eucaryotes, la « boîte à outils » du protéasome apparaît comme un mécanisme d'adaptation dynamique dont la fonction dans la physiologie des cellules Eucaryotes a considérablement varié en fonction de la biologie des espèces.

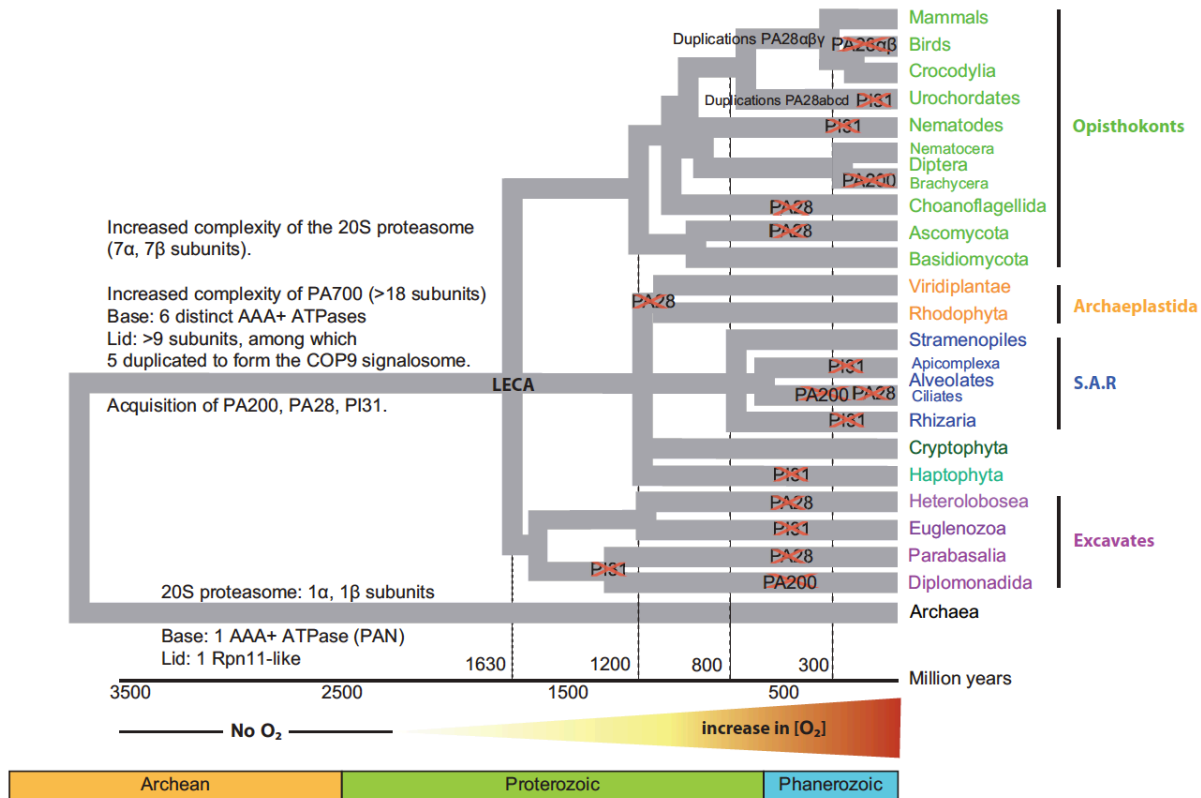


Figure 5 | Evolution des sous-unités du protéasome 20S et de ses régulateurs chez les Eucaryotes. Le chronogramme résume les différents événements de gains et pertes inférés au cours de l'histoire évolutive des Eucaryotes qui ont façonné le répertoire actuel des régulateurs du protéasome.

2.5. OUTILS BIOINFORMATIQUES POUR LA PHYLOGENIE ET L'ÉVOLUTION MOLECULAIRE

La pratique de la reconstruction phylogénétique au quotidien requiert l'utilisation d'un véritable arsenal de programmes bioinformatiques pour manipuler les séquences et les arbres. J'ai ainsi participé au développement de plusieurs outils notamment au travers de collaborations fertiles avec Vincent Ranwez, aujourd'hui à SupAgro Montpellier, et toujours à l'affût de programmes pouvant faciliter la tâche aux biologistes.

2.5.1 LA BASE DE DONNEES ORTHOMAM

L'approche génomique en biologie a révolutionné notre compréhension de l'évolution des organismes grâce à l'énorme quantité d'information moléculaire disponible dans les bases de données publiques telles que GenBank (Benson et al. 2013) ou Ensembl (Flicek et al. 2014). L'activité de recherche de notre équipe étant axée en grande partie sur l'histoire évolutive des Mammifères, nous avons recensé de manière exhaustive les séquences codantes (CDS) orthologues présentes sous forme de copies uniques dans Ensembl pour construire la base de données OrthoMaM (Ranwez et al. 2007; <http://www.orthomam.univ-montp2.fr/>). Cette dernière fournit des descripteurs évolutifs — taux d'évolution, composition en bases, variabilité relative en première, deuxième et troisième positions du codon — permettant d'apprécier le pouvoir résolutif phylogénétique de CDS complets et d'exons orthologues (Figure 16). Cette base de données qui a acquis une renommée internationale, est mise à jour régulièrement (i.e. la

version courante est la 8^{ème} ; **Douzery et al. 2014**) et sert de base à de nombreux travaux de génomique évolutive au sein de l'équipe (ex. Galtier et al. 2009 ; **Romiguiet et al. 2013**).



Figure 16 | Captures d'écran du site web d'OrthoMaM. Cet exemple illustre le résultat d'une recherche des CDS présents chez 15–40 espèces, ayant un taux d'évolution relatif entre 0.5 et 3, un paramètre alpha d'hétérogénéité des taux entre sites entre 1 et 1.5, et une composition en GC3 située entre 22% et 35%. Cette requête retourne 23 CDS parmi lesquels le gène *LRR63* dont les descripteurs évolutifs, l'alignement, et l'arbre de maximum de vraisemblance sont affichés.

2.5.2 LE PORTAIL WEB PHYLOEXPLORER

En réponse au nombre grandissant d'arbres phylogénétiques à manipuler dans les analyses phylogénomiques, nous avons créé le portail web interactif PhyloExplorer (<http://www.ncbi.orthomam.univ-montp2.fr/phyloexplorer>) dédié à l'exploration taxonomique et à la manipulation de larges collections d'arbres phylogénétiques (**Ranwez et al. 2009**). Pour une collection d'arbres donnée, PhyloExplorer fournit ainsi des statistiques descriptives permet de relier et corriger les noms des taxons en fonction d'une taxonomie de référence (NCBI ou ITIS). Le programme recherche aussi automatiquement des informations sur les taxons contenus dans les arbres en faisant des liens vers les pages Wikipedia et iSpecies et extrait les images correspondantes de WikiSpecies lorsqu'elles sont disponibles. Grâce à la possibilité de faire des requêtes taxonomiques complexes sur des collections d'arbres personnelles (Figure 17) ou directement sur la base de données TreeBASE (Sanderson et al. 1994), PhyloExplorer permet de construire facilement des collections d'arbres contenant des taxons d'intérêt en amont de l'application de méthodes de construction de superarbres (Bininda-Emonds et al. 2002), ou plus fréquemment aujourd'hui de méthodes d'analyse des arbres de gènes / arbres d'espèces (Liu et al. 2009). Il peut également permettre la construction de supermatrices si les arbres sont reliés aux séquences sous-jacentes (ex. **Chiari et al. 2012**).

PhyloExplorer
(Phylogenetic, Tree), Explorer

Upload collection > Statistics

Go to the bottom of the page

Search trees Prune trees

Here you can obtain a new tree collection by searching trees satisfying a taxonomic query in either your collection or in Treebase.

{EUARCHONTOGLIRES}>0 and {LAURASIATHERIA}>0 and {A} query against Treebase Search

query example: {muridae} > 25 and {(mus)=1 or {rattus}=1}

You can download your result after your query

Results for -- {EUARCHONTOGLIRES}>0 and {LAURASIATHERIA}>0 and {AFROTHERIA}>0 and {XENARTHRA}>0 and {{METATHERIA}>0 or {MONOTREMATA}>0} and {EUTHERIA} > 6. -- | [Back to original collection](#)

General information

25 different label(s) occur(s) in 23 tree(s)

- 25 are successfully mapped to 25 different taxa of the reference taxonomy

Download corrections

Tree size distribution

[8-9]	1 tree with 8-9 distinct leaf labels
[12-13]	1 tree with 12-13 distinct leaf labels
[14-15]	4 trees with 14-15 distinct leaf labels
[16-17]	6 trees with 16-17 distinct leaf labels
[18-19]	4 trees with 18-19 distinct leaf labels
[20-21]	4 trees with 20-21 distinct leaf labels
[22-23]	3 trees with 22-23 distinct leaf labels

Taxon frequency distribution

[8-9]	1 leaf label is present in 8-9 tree
[10-11]	5 leaf labels are present in 10-11 tree
[12-13]	2 leaf labels are present in 12-13 tree
[14-15]	3 leaf labels are present in 14-15 tree
[16-17]	4 leaf labels are present in 16-17 tree
[18-19]	3 leaf labels are present in 18-19 tree
[20-21]	2 leaf labels are present in 20-21 tree
[22-23]	5 leaf labels are present in 22-23 tree

Figure 17 | Capture d'écran du portail web PhyloExplorer. Cet exemple montre le résultat et les statistiques associées d'une requête taxonomique complexe effectuée sur une collection d'arbres issus de la base de données OrthoMaM pour lesquels les taxons ont été reliés à la taxonomie sur la base de leurs noms. Cette requête permet de restreindre la collection initiale de 79 arbres aux 23 arbres comprenant les taxons d'intérêt.

2.5.3 LE PROGRAMME D'ALIGNEMENT MACSE

Lors de la construction d'OrthoMaM, nous avons constaté que la majorité des programmes d'alignement disponibles traitaient les séquences nucléotidiques codantes sans prendre en compte leur structure en codons. Des solutions indirectes passant par l'alignement des séquences protéiques traduites ont bien été proposées et implémentées dans des programmes comme TransAlign (Bininda-Emonds 2005) ou TranslatorX (Abascal et al. 2010), mais celles-ci se heurtent à la présence d'indels créant des décalages du cadre de lecture. Nous avons donc développé une solution algorithmique spécialement conçue pour l'alignement multiple de séquences nucléotidiques codantes qui prend en compte la traduction en acides aminés de ces séquences, tout en autorisant l'apparition de changements de cadre de lecture et de codons stop (**Ranwez et al. 2011**). Dans le cas où les séquences ne contiennent pas ce type d'évènements rares, notre programme MACSE (Multiple Alignment of Coding SEquences) produit des alignements d'une qualité proche de celle des meilleures solutions actuelles. Cependant, tout l'intérêt de MACSE réside dans sa capacité à aligner des séquences contenant des changements de cadre de lecture réels (ex. pseudogènes) ou apparents (erreurs de séquençage). MACSE permet ainsi de détecter des erreurs potentielles dans les séquences des bases de données

publiques, d'aligner des séquences de pseudogènes (Figure 18) ou encore des séquences particulièrement sujettes aux erreurs par insertion-délétion de nucléotides telles que celles produites par le séquençage 454. Cet outil est ainsi régulièrement utilisé dans des études de génomiques évolutives estimant le dN/dS sur un grand nombre d'alignements (ex. Assis et al. 2012) et a été également intégré dans des pipelines de métabarcoding pour permettant comme une étape de nettoyage des reads (ex. Yu et al. 2012). Nous sommes d'ailleurs en train d'étendre MACSE en une boîte à outils permettant notamment d'obtenir efficacement des alignements fiables de dizaines de milliers de séquences codantes pour les projets de métabarcoding (Leray et al. 2013). Le code source de MACSE peut-être téléchargé depuis le serveur web dédié à l'exécution en ligne du programme (<http://mbb.univ-montp2.fr/macse>).

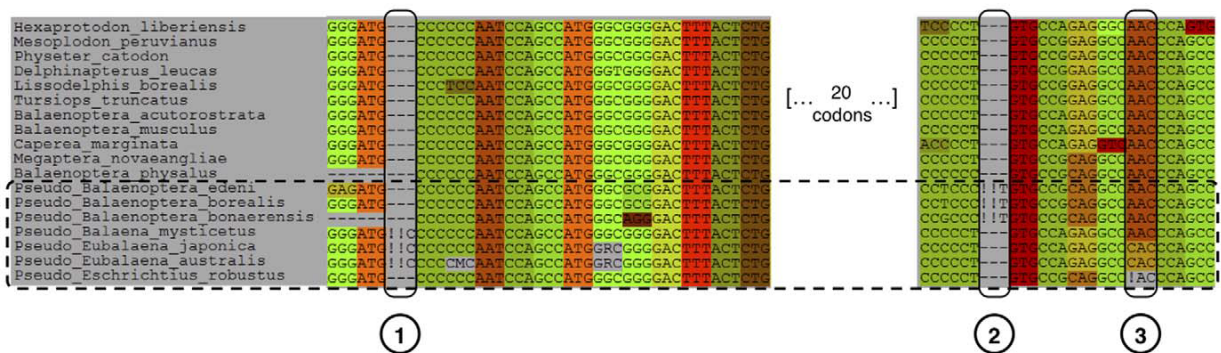


Figure 18 | Alignement de pseudogènes avec MACSE. Cet exemple montre l'alignement de séquences fonctionnelles et non fonctionnelles (pseudogènes encadrés sur la figure) du gène de l'améloblastine (AMBN) chez les Cétartiodactyles. Ce gène code pour une protéine de l'émail dentaire qui a perdu sa fonction chez les baleines mysticètes dont les dents ont été remplacées par des fanons kératineux. MACSE est capable d'aligner ces séquences sur le cadre de lecture en inférant la position de frameshifts codés par des points d'exclamation (!). La visualisation de l'alignement colorié par codons est réalisée par Seaview 4 (Gouy et al. 2010).

3. PROJET DE RECHERCHE

Mes objectifs de recherche à moyen terme s'inscrivent dans les thématiques de recherche du département Génome de l'ISEM. Ainsi, si je compte bien continuer à mener des projets autour des axes précédemment évoqués, je m'attacherai surtout à développer une thématique propre sur les mécanismes génomiques qui sous-tendent les phénomènes de convergence évolutive. Pour ce faire, j'utiliserai le modèle des mammifères myrmécophages sur lequel je viens d'obtenir un financement Consolidator de l'European Research Council (ERC) pour les cinq prochaines années. Ce projet s'imposera donc naturellement comme l'axe principal de mon programme de recherche dans les années à venir avec la constitution d'une équipe autour de cette thématique. Ceci me permettra ainsi d'affirmer mon autonomie au sein de l'ISEM.

3.1. CONVERGEANT : UNE APPROCHE INTEGRATIVE DE L'ÉVOLUTION CONVERGENTE

En dépit de large répartition au sein de l'arbre de vie, de nombreuses questions demeurent quant au fascinant phénomène d'évolution convergente. Les mammifères myrmécophages qui se nourrissent de fourmis et/ou termites constituent un exemple classique de convergence évolutive avec au moins cinq origines indépendantes au sein des placentaires (tatous, fourmiliers, oryctéropes, pangolins, et protèles). Les nombreuses convergences morphologiques, l'importance de la convergence moléculaire, et le rôle du microbiome dans l'adaptation au régime alimentaire sont de plus en plus reconnus. Cependant, des études comparatives à large échelle combinant l'évolution morphologique, la génomique, et la métagénomique du microbiome font encore défaut. Dans ce projet, nous proposons de tirer parti de l'ensemble unique de caractères convergents associés au régime myrmécophage pour étudier les mécanismes moléculaires sous-tendant l'adaptation phénotypique. En utilisant des méthodes de phénotypage quantitatives basées sur la microtomographie aux rayons X et les nouvelles technologies de séquençage, nous proposons de combiner les approches morphométrique, génomique, et métagénomique pour évaluer les modalités de l'évolution convergente du crâne des mammifères myrmécophages, de leurs génomes, et de leurs microbiomes oraux et intestinaux. Avec ce projet, nous visons à apporter des réponses à des questions fondamentales sur les mécanismes de l'évolution convergente. Ce projet sera ainsi le premier à appliquer une telle approche intégrative pour étudier l'interaction complexe entre la morphologie, le génome, et le microbiome des mammifères dans un cas classique de convergence associée à l'adoption d'un régime alimentaire spécialisé.

3.1.1 LA CONVERGENCE ÉVOLUTIVE

La convergence évolutive désigne le fait que des organismes phylogénétiquement éloignés évoluent indépendamment des adaptations similaires lorsqu'ils sont soumis aux mêmes pressions sélectives (McGhee 2011). Ce phénomène évolutif fascinant est très répandu dans l'arbre du vivant et se produit à diverses échelles phylogénétiques (voir <http://www.mapoflife.org> pour de nombreux exemples). Les phylogénies moléculaires récentes ont révélé l'ampleur de l'évolution convergente en mettant au grand jour de nombreux exemples insoupçonnés, y compris certains cas spectaculaires chez les mammifères placentaires (Madsen et al. 2001; Murphy et al. 2001; Delsuc et al. 2002). Ceci suggère que l'évolution morphologique est beaucoup plus contrainte qu'on ne le pensait, conduisant à un niveau élevé d'homoplasie des caractères morphologiques tant est si bien que leur analyse phylogénétique tend à soutenir des groupes écomorphologiques plutôt que des groupes phylogénétiques (Springer et al. 2007, 2013). D'importants progrès ont été réalisés récemment dans la compréhension des mécanismes moléculaires sous-tendant les phénomènes de convergence évolutive (Arendt & Reznick 2008; Christin et al. 2010). Cependant, il reste encore beaucoup à faire pour comprendre les diverses circonstances moléculaires de l'évolution convergente et les rôles respectifs que l'adaptation et les contraintes jouent dans ce phénomène évolutif (Losos 2011).

Avec l'avènement des nouvelles technologies de séquençage, il est désormais possible d'étudier les bases moléculaires de la convergence évolutive à l'échelle du génome (Elmer et Meyer 2011). Ces approches génomiques viennent compléter les études basées sur des gènes candidats qui ont très tôt révélé des cas spectaculaires d'évolution moléculaire convergente tel que l'exemple

désormais classique de l'adaptation convergente des lysozymes à l'herbivorie entre les ruminants et les colobes folivores (Stewart et al. 1987). Un exemple plus récent est l'identification de substitutions adaptatives convergentes dans le gène codant pour la protéine prestine impliquée dans l'audition, entre les chauves-souris écholocatrices et les cétacés odontocètes (Li et al. 2010a; Liu et al. 2010a,b). L'étude d'autres gènes de l'audition entre chauves-souris écholocatrices et non-écholocatrices a pointé des candidats supplémentaires et a suggéré que la convergence moléculaire adaptative pourrait être plus répandue que généralement admis (Davies et al. 2012). Ces prédictions ont été testées dans une première analyse menée à l'échelle du génome de l'écholocation entre chauves-souris et cétacés odontocètes qui a détecté plus de 200 gènes potentiellement convergents entre ces lignées (Parker et al. 2013). Cependant, cette étude a souffert de l'absence d'un modèle nul approprié pour les substitutions adaptatives convergentes et des ré-analyses ultérieures ont montré que, alors que l'évolution moléculaire convergente semble relativement courante dans les génomes, la convergence moléculaire adaptative liée à la convergence phénotypique est relativement rare (Thomas & Hahn 2015 ; Zou & Zhang 2015), comme récemment constaté dans le cas des mammifères marins (Foote et al. 2015).

3.1.2 LE MODELE DES MAMMIFERES MYRMECOPHAGES

Les mammifères myrmécophages représentent un exemple classique de convergence évolutive (McGhee 2011) et fournissent donc un modèle idéal pour étudier les processus sous-jacents. En effet, alors que plus de 200 espèces de mammifères incluent une proportion de fourmis et/ou termites dans leur régime alimentaire, seulement 22 espèces peuvent être considérées comme des myrmécophages spécialisés qui consomment plus de 90% de fourmis et/ou termites (Redford 1987). Les lignées myrmécophages ont évolué plusieurs fois de façon indépendante dans tous les grands clades de mammifères. Chez les Monotrèmes, les échidnés du genre *Tachyglossus* sont myrmécophages. Parmi les Marsupiaux, seul le numbat (*Myrmecobius fasciatus*) se nourrit presque exclusivement de termites et fourmis. La myrmécophagie a également évolué de façon indépendante dans cinq ordres de placentaires: les Cingulata au sein desquels le tatou géant (*Priodontes maximus*) s'est spécialisé dans la consommation de fourmis et termites, les Pilosa avec les trois genres de fourmiliers sud-américains (*Cyclopes*, *Myrmecophaga* et *Tamandua*), les Tubulidentata avec l'oryctérope (*Orycteropus afer*), les Pholidota avec trois genres de pangolins, et les Carnivora avec le protèle (*Proteles cristata*). Ces cinq lignées présentent différents degrés de spécialisation morphologique vers la myrmécophagie, reflétant probablement l'influence de contraintes phylogénétiques et le temps écoulé depuis l'adoption de ce régime alimentaire si particulier par ces lignées indépendantes (Figure 19).

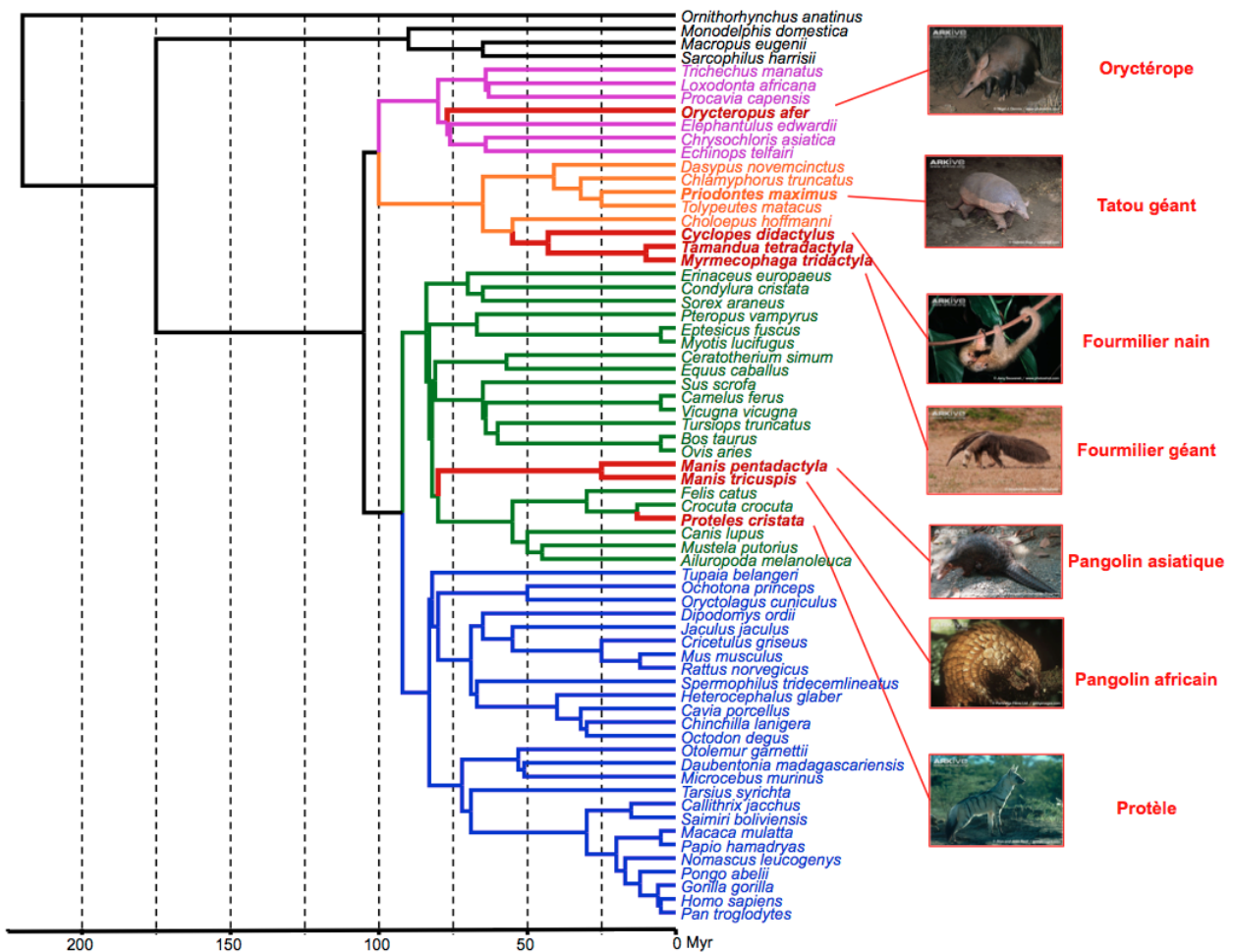


Figure 19 | Evolution convergente des placentaires myrmécophages. Les espèces myrmécophages (en rouge) sont replacées dans le contexte phylogénétique et temporel des génomes de mammifères disponibles. Photos ©ARKive.org.

Les tatous, les fourmiliers, l'oryctérope et les pangolins constituent les exemples les plus extrêmes de phénotypes myrmécophages et ils d'ailleurs longtemps été classés dans le même ordre des Edentés (Glass 1985). Ces animaux ont développé des adaptations morphologiques similaires, mais convergents telles que la réduction ou la perte des dents, un museau allongé avec une languette extensible, une salive visqueuse produite par des glandes salivaires hypertrophiées et des griffes puissantes utilisées pour creuser dans les nids de fourmis et de termites. En outre, du point de vue physiologique, les placentaires myrmécophages possèdent aussi un métabolisme basal faible probablement en rapport avec leur régime alimentaire relativement pauvre du point de vue nutritionnel et des structures olfactives bien développées (Redford & Dorea 1984). Le protèle est un myrmécophage strict se nourrissant exclusivement de termites du genre *Trinivertermes*. Malgré le fait qu'il soit le seul représentant d'une lignée de Hyanidae ayant divergé il y a 10 Myr au maximum (Eizirik et al. 2010), il présente des différences morphologiques marquées par rapport aux hyènes carnivores, avec une dentition réduite, un museau allongé et une langue élargie, et une morphologie globalement plus frêle.

3.1.3 LES MODALITES DE LA CONVERGENCE MORPHOLOGIQUE

Les placentaires myrmécophages fournissent donc un modèle particulièrement adapté à l'étude des multiples adaptations morphologiques et physiologiques

convergentes associées à la transition vers ce régime alimentaire hautement spécialisé. Springer et al. (2013) ont récemment montré que le signal écomorphologique dû à la convergence est si prononcé chez les placentaires que des analyses cladistiques basées sur 4541 caractères morphologiques (O'Leary et al. 2013) regroupent de manière artéfactuelle les lignées myrmécophages en un groupe monophylétique. Cette étude a ainsi identifié 98 caractères ayant convergés entre tatous, fourmiliers, oryctéropes et pangolins. Parmi ces caractères se trouvaient des caractères crâniens liés à l'allongement du museau et à la simplification dentaire. Si la modification de l'ontogenèse a été reconnue depuis longtemps comme une source importante de variation morphologique, l'hétérochronie et les processus hétérotypiques jouant un rôle central dans la génération de nouvelles morphologies, seules quelques tentatives ont été faites pour étudier le développement du crâne chez les myrmécophages. Les travaux antérieurs ont décrit des fœtus isolés ou se sont concentrés sur des aspects spécifiques de sa morphologie (ex. Martin 1916). Ces travaux restent anecdotiques et ne constituent pas une base comparative sur laquelle évaluer la façon dont les mammifères myrmécophages s'écartent des modèles de développement des autres mammifères. Le cadre phylogénétique aujourd'hui bien défini pour les placentaires (Meredith et al. 2011), permet désormais d'espérer pouvoir identifier des ensembles modulaires de caractères crâniens qui ont convergé au cours de l'adaptation à la myrmécophagie.

3.1.4 LES TRACES DE LA CONVERGENCE MOLECULAIRE

Les traces laissées par la convergence évolutive dans les génomes peuvent être de plusieurs types. Elles comprennent des événements de pseudogénération parallèles et de pertes de gènes orthologues en copie unique, des changements adaptatifs convergents dans les gènes protéiques, et l'évolution parallèle de familles de gènes par duplications/pertes. Nous avons déjà étudié le rôle potentiel du facteur de transcription Runt-like 2 (RUNX2) qui comprend une région répétée qui semble fonctionner comme un "bouton de réglage" pour déterminer la longueur du museau chez les carnivores (Sears et al. 2007). Cependant, nous n'avons trouvé aucune corrélation entre la répétition en tandem de glutamines et alanines de RUNX2 et la longueur du museau à l'échelle des placentaires y compris chez divers placentaires myrmécophages (**Pointer et al. 2012**). Cependant, il a été démontré que le relâchement des contraintes sélectives chez les placentaires montrant une simplification dentaire, y compris les édentés et les Cétacés à fanons, a conduit à des événements de pseudogénération parallèles dans les gènes de l'émail ENAM, AMBN et AMELX (Meredith et al. 2009, 2013 ; **Delsuc et al. 2015**). Notre travail en cours axé sur les myrmécophages montre que la réduction dentaire chez les xénarthres, les pangolins, et l'oryctérope a entraîné la perte convergente de la fonction de ces gènes de l'émail, ainsi que du gène DSPP qui est spécifique de la dentine (**Figure 20**). En outre, d'autres gènes potentiellement liés à des adaptations à la myrmécophagie tels que les enzymes digestives, les récepteurs du goût sucré et umami (gènes TAS1R; voir Jiang et al. 2012), et les gènes liés au métabolisme pourraient avoir été la cible de l'évolution convergente via des substitutions adaptatives parallèles. Enfin, l'évolution convergente de familles de gènes spécifiques peut également se produire par duplication adaptative comme cela a été montré pour les RNases digestives des colobes folivores (Zhang 2006) et/ou par perte de fonction comme les récepteurs olfactifs qui ont été façonnés par la spécialisation écologique chez les mammifères (Hayden et al. 2010). Les récepteurs olfactifs (OR gènes) et les récepteurs du goût amer (gènes TAS2R; Shi et Zhang 2006; Li et Zhang 2014) apparaissent ainsi comme de bons

candidats pour étudier l'évolution parallèle de familles des gènes dans les lignées myrmécophages qui partagent des stratégies de recherche de nourriture et des régimes alimentaires similaires.

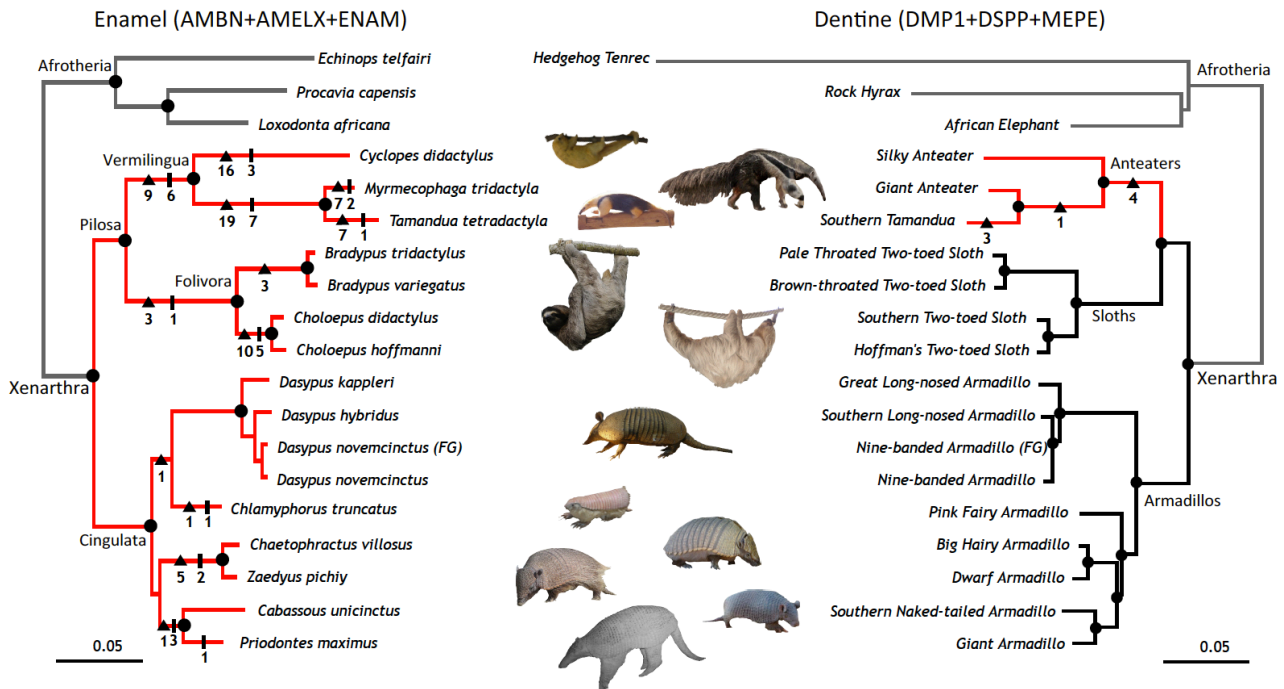


Figure 6 | Perte de fonction des gènes de l'émail et de la dentine chez les xénarthres. Les arbres sont basés sur l'analyse ML des principaux gènes de l'émail (à gauche) et de la dentine (à droite). Le nombre d'indels entraînant une rupture du cadre de lecture (triangles) et de codons stop (barres) ont été inférés par parcimonie avec l'optimisation ACCTRAN. Les nœuds avec bootstrap >99% sont indiqués par des cercles. Les trois gènes de l'émail ont perdu leur fonction (branches rouges) chez les xénarthres, alors que seul le gène spécifique de la dentine DSPP est non-fonctionnel chez les fourmiliers. Les autres gènes de la dentine (MEPE et DMP1) sont également impliqués dans la morphogénèse des os.

3.1.5 LE ROLE DU MICROBIOME

Cependant, en termes d'adaptation au régime alimentaire, le microbiote (i.e. l'ensemble des communautés de microorganismes commensaux) joue aussi potentiellement un rôle majeur (Karasov et al. 2011). Chez les vertébrés, le microbiome (i.e. l'ensemble des génomes du microbiote) représente un formidable potentiel d'adaptation pour l'hôte et son rôle essentiel dans l'adaptation des organismes est de plus en plus reconnu (McFall-Ngai et al. 2013). Peut-être l'un des exemples les plus remarquables est la présence dans le microbiome intestinal des populations japonaises de gènes spécifiques codant pour des enzymes spécialisées dans la dégradation des composants de la paroi des algues (Hehemann et al. 2010). Ces gènes ont été transférés horizontalement à partir de bactéries marines, vraisemblablement ingérées avec les fruits de mer, à la bactérie commensale de l'intestin humain *Bacteroides plebeius*. Dans ce cas, le microbiome intestinal a fourni une solution adaptative efficace en réponse à un changement récent du régime alimentaire impliquant une augmentation importante de la consommation d'algues. Le développement rapide des NGS a ouvert un champ nouveau d'études visant à caractériser la composition du microbiome humain et à comprendre son rôle fonctionnel au cours de l'évolution. Chez les mammifères, il a d'abord été démontré par

métabarcoding de l'ARNr 16S d'échantillons fécaux d'une diversité d'espèces que le régime alimentaire et la phylogénie ont influencé tous deux influencé l'évolution du microbiome intestinal (Ley et al. 2008). Le régime alimentaire apparaît néanmoins comme un facteur déterminant puisque la composition du microbiome intestinal a évolué de manière convergente chez les mammifères partageant les mêmes habitudes alimentaires (Muegge et al. 2011). En particulier, des différences importantes de communautés microbiennes intestinales ont été mis en évidence entre carnivores, omnivores et herbivores, indépendamment de leur phylogénie. De plus, des approches de métagénomique ont mis en évidence le fait que les microbiotes intestinaux des mammifères qui diffèrent dans leur composition partagent néanmoins un ensemble de gènes assurant des fonctions similaires (Muegge et al. 2011).

Les placentaires myrmécophages offre une opportunité supplémentaire pour étudier les mécanismes qui président à l'évolution des microbiomes à large échelle phylogénétique. Ils fournissent en effet un système modèle pour caractériser la composition taxonomique et le contenu fonctionnel en gènes des microbiomes intestinaux et oraux chez des espèces qui ont évolué de façon convergente depuis des millions d'années et qui partagent le même régime alimentaire hautement spécialisé. Les espèces myrmécophages pourraient en effet utiliser des bactéries symbiotiques leur permettant de digérer efficacement la chitine contenue dans les exosquelettes de fourmis et de termites, mais cette hypothèse n'a jamais été formellement testée. Le séjour Fulbright de trois mois que j'ai effectué en 2012 dans le laboratoire de Rob Knight (<https://knightlab.colorado.edu/>) à l'Université du Colorado à Boulder (USA), nous a permis d'analyser 93 nouveaux échantillons fécaux par barcoding 16S Illumina représentant la diversité des espèces myrmécophages et des espèces étroitement apparentées (**Delsuc et al. 2014**). L'analyse en réseau de ces données combinées avec celles de Muegge et al. (2011) a confirmé le regroupement significatif précédemment rapporté des espèces dicté à la fois de la phylogénie et du régime alimentaire (Figure 21). Ainsi, on peut distinguer les deux grands groupes d'herbivores correspondant aux espèces monogastriques (capibara, rhinocéros, éléphant et zèbre) et polygastriques (girafe, okapi, gazelle et mouton). Nos résultats ont également montré que les espèces myrmécophages semblent avoir convergé en termes de composition de leurs communautés bactériennes intestinales, les fourmiliers, les oryctéropes et les protèles se groupant ensemble sur le graphe. Fait intéressant, le protèle qui se nourrit uniquement de termites apparaît plus proche des myrmécophages, comme les fourmiliers et oryctéropes, que des autres espèces Carnivores, y compris de la hyène tachetée dont il y est phylogénétiquement très proche. Cette observation démontre que la composition du microbiome peut évoluer assez rapidement de façon convergente en réponse à la myrmécophilie. Ces résultats encourageants demandent néanmoins être confirmés par l'analyse d'échantillons fécaux supplémentaires représentant un large éventail d'espèces, avec un accent particulier sur les taxons insectivores afin de les comparer avec les myrmécophages stricts. Ces analyses devraient aussi logiquement être étendues au microbiome oral, qui joue également un rôle important dans l'adaptation au régime alimentaire, et pour lesquels pratiquement aucune donnée comparative n'est disponible pour l'instant chez les mammifères, à l'exception de l'homme, de la souris, et du chien (Dewhurst et al. 2012).

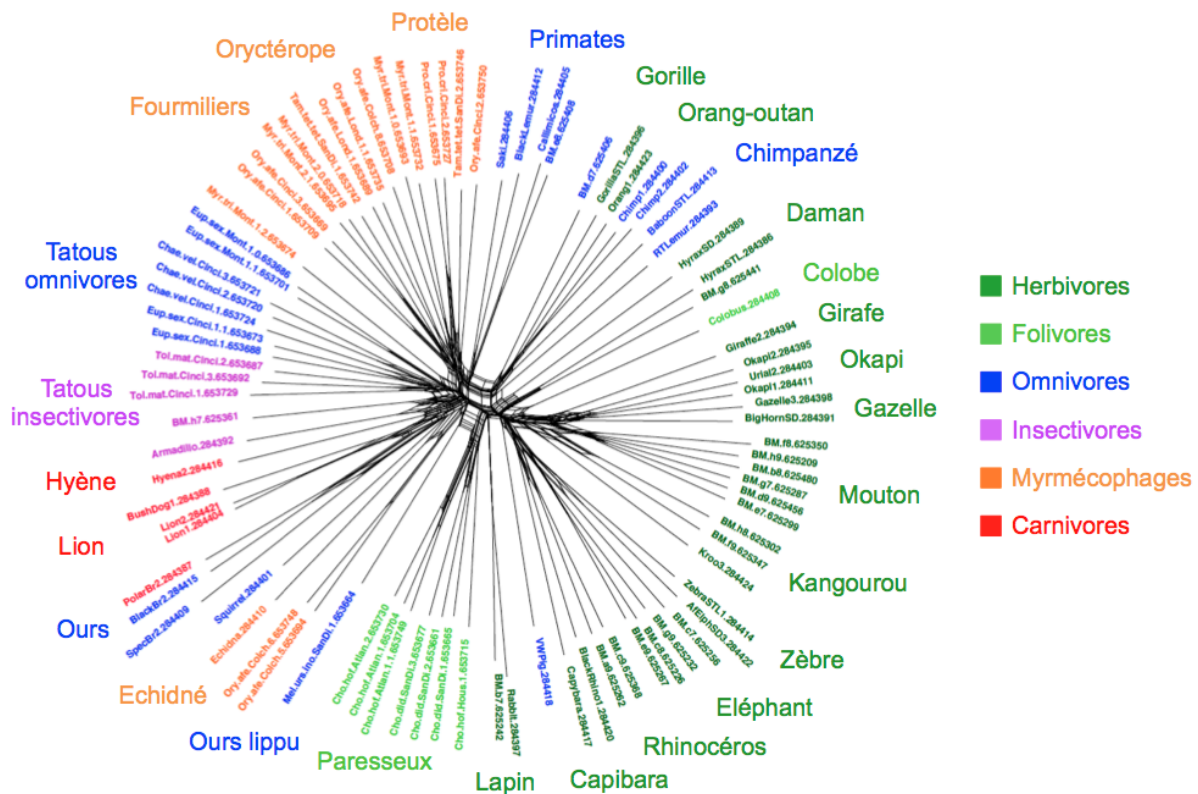


Figure 7 | Evolution des microbiomes intestinaux de mammifères. Le réseau phylogénétique représente les relations entre les microbiomes intestinaux de différents mammifères sur la base de leur composition taxonomique en bactéries et archées (Illustration tirée des Cahiers de Prospective du CNRS 2013 en Génomique Environnementale).

Au-delà de la composition taxonomique du microbiome des myrmécophages, leur contenu fonctionnel en gènes devra également être étudié par des approches de métagénomique (Muegge et al. 2011). Le cas de l'adaptation du panda géant (*Ailuropoda melanoleuca*) à son régime alimentaire très spécifique constitué uniquement de bambou offre un exemple particulièrement illustratif de la complémentarité des approches de génomique de l'hôte et de métagénomique du microbiome associé. Le séquençage du génome du panda géant a en effet révélé qu'il contient les enzymes digestives typiques des carnivores, mais qu'il est dépourvu des gènes codant pour les enzymes nécessaires à la digestion de la cellulose (Li et al. 2010b). Toutefois, le métabarcoding de l'ARNr 16S sur des échantillons fécaux a révélé la présence dans le microbiome intestinal du panda, de taxons bactériens pouvant dégrader la cellulose, et une approche métagénomique a retrouver des gènes codant pour les enzymes permettant de dégrader la cellulose et l'hémicellulose (Zhu et al. 2011). Ceci suggère que le panda géant utilise des bactéries symbiotiques pour dégrader la cellulose contenue dans les feuilles de bambou qu'il ingère en quantité au quotidien. Ce type d'études de métagénomique fonctionnelle du microbiome des mammifères est actuellement en expansion comme illustré par l'étude à large échelle de Muegge et al. (2011), et des travaux similaires menés chez le porc (Lamendella et al. 2011), le lynx ibérique (Alcaide et al. 2012), ou encore les cétacés (Sanders et al. 2015), fournissant ainsi des données comparatives pour notre étude sur les mammifères myrmécophages.

3.2. MISE EN ŒUVRE DU PROJET CONVERGEANT

Afin de revisiter le cas classique de l'évolution convergente chez les mammifères placentaires myrmécophages, ce projet mettra en œuvre les premières études comparatives de génomique et de métagénomique chez ces animaux. Les méthodes de séquençage de nouvelle génération permettent en effet d'obtenir les données nécessaires à un coût raisonnable. Le projet sera organisé en trois grandes tâches consacrées à l'étude de la convergence : (1) dans la morphologie des mammifères myrmécophages en se focalisant sur le développement du crâne et les origines de la simplification dentaire, (2) dans les génomes de placentaires myrmécophages, et (3) dans leurs microbiomes oraux et intestinaux.

3.2.1 TACHE 1 : CONVERGENCE MORPHOLOGIQUE ET SIMPLIFICATION DENTAIRE

Le crâne des myrmécophages a été l'objet d'une attention considérable notamment parce qu'ils montrent une réduction dentaire extrême atteignant son paroxysme chez les fourmiliers et les pangolins qui ont entièrement perdu les dents. Les principaux objectifs de cette tâche centrée sur la morphologie sont de révéler les modalités de la convergence menant à l'allongement du crâne et à la réduction dentaire observée dans les lignées indépendantes de myrmécophages. Les questions fondamentales auxquelles nous chercherons à répondre au travers de cette tâche sont les suivantes : (1) Peut-on caractériser l'origine développementale de la réduction dentaire ? ; (2) Peut-on identifier un ensemble de caractères convergents définissant le phénotype myrmécophage et qui pourraient correspondre à des modules développementaux du crâne ? ; et (3) Comment la variation du crâne a été influencée par le relâchement des contraintes sélectives en lien avec la myrmécophagie ?

Acquisition de Données

Pour la réalisation de cette tâche, nous nous appuierons sur des crânes et des séries ontogénétiques de spécimens complets (foetus entiers, juvéniles, sub-adultes et adultes) qui seront numérisés en 3D par microtomographie à rayons X (Figure 22) à partir de spécimens de collections empruntés au près de musées en Europe (MNHN: Museum National d'Histoire Naturelle de Paris; BMNH: British Museum of Natural History de Londres; ZMUC: Zoological Museum of the University of Cambridge; MNB: Museum für Naturkunde à Berlin) et aux Etats-Unis (AMNH: American Museum of Natural History à New York; USNM: Smithsonian Museum à Washington DC). Nous recueillerons des données numériques 3D de crânes et de spécimens complets (foetus entiers, juvéniles, sub-adultes et adultes) par tomographie par rayons X micro-ordinateur (scan μ CT). Nous envisageons la numérisation d'un total d'environ 500 spécimens au cours du projet. Pour chaque taxon, entre 5 et 20 stades ontogénétiques seront analysés afin de maximiser la représentation des différents événements développementaux. Les scans de crânes en 3D fourniront une énorme base de données à partir desquelles les analyses de morphométrie géométriques et des descriptions détaillées des tissus squelettiques seront conduites.

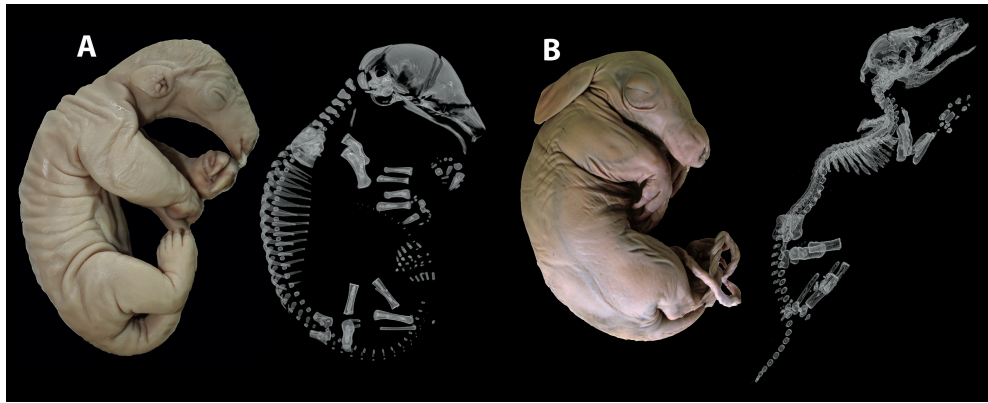


Figure 8 | Stades ontogéniques représentatifs de deux espèces myrmécophages et reconstructions tridimensionnelles par CT-scan de leur squelette (Hautier et al. 2011, 2013). A) Fourmilier (*Tamandua tetradactyla*). B) Oryctérope (*Orycteropus afer*).

Développement Prénatal et Réduction Dentaire

Bien que l'anatomie crânienne des myrmécophages ait été relativement bien étudiée, peu d'études ont porté sur le développement de leurs dents (Martin 1916; Ciancio et al. 2012). Les fourmiliers et les pangolins sont totalement édentés, au contraire des autres myrmécophages qui conservent une dentition simplifiée, généralement avec des racines peu profondes, et souvent réduite avec des dents composées uniquement de dentine. Martin (1916) a rapporté l'existence de dents rudimentaires dans la mâchoire inférieure de quelques tatous, qui sont soit trop petites pour être éruptées et donc réabsorbées, ou sont éruptées mais tombent peu après la naissance, étant que très superficiellement attachées à la mâchoire. Cependant, il est possible que d'autres groupes d'édentés puissent également présenter des bourgeons dentaires rudimentaires. Une telle observation a en effet été faite en utilisant la microtomographie à rayons X à haute résolution sur des séries ontogénétiques de paresseux, montrant qu'ils développent également des bourgeons dentaires supplémentaires qui n'éruptent pas chez les juvéniles (Hautier et al. 2016). La constitution d'une large base de données numérisées de fœtus d'espèces myrmécophages contribuera à une meilleure compréhension de l'ontogénèse dentaire prénatale et à l'établissement de critères développementaux permettant d'identifier des homologies avec les dents des autres mammifères. En recherchant la présence de bourgeons dentaires prénataux supplémentaires dans les mâchoires d'espèces myrmécophages, y compris chez les pangolins et les fourmiliers, nous espérons caractériser les processus évolutifs par lesquels la réduction dentaire se produit comme une première étape vers la découverte des changements génomiques sous-jacents.

Patrons de Covariation et Modularité du Crâne

L'évolution corrélée des traits est un facteur principal de l'évolution morphologique. Les résultats d'études précédentes ont suggéré que la variation des caractères crâniens est intimement associée à certains aspects de l'éruption dentaire chez les mammifères (Goswami 2006a). Au-delà des approches descriptives qualitatives, l'avènement des méthodes de morphométrie géométrique a marqué une étape importante dans l'analyse quantitative du phénotype. Nous envisageons donc d'utiliser ces méthodes basées sur des points homologues placés sur CT-scans 3D pour quantifier la variation du crâne chez les espèces myrmécophages. L'étude de séries ontogénétiques sera utilisée pour

caractériser les divergences dans les trajectoires ontogénétiques et allométriques entre les lignées convergentes de myrmécophages. Nous allons ensuite utiliser des analyses 2-blocks Partial Least Square (PLS) pour caractériser la covariation entre les différentes parties du crâne afin de révéler de potentiels patrons de modularité crânienne (ex. Hautier et al. 2012). Le concept de modularité renvoie à la croissance corrélée des traits liés à des groupes discrets ou modules, qui sont intégrés de façon interne, mais relativement indépendants les uns des autres (Goswami 2006a). La modularité fonctionnelle du crâne représente en effet des compromis entre les fonctions interagissant comme l'ouïe, la vue, l'odorat, la mastication, et la respiration. Chez les placentaires, le crâne est modulaire et peut être séparé en six modules (zygomatique-ptérygoïdien, voûte, orbite oro-nasale antérieure, molaire, et basicrâne), chacun ayant une intégration variable (Goswami 2006a). Ces études ont suggéré que l'intégration au sein de modules peut contraindre l'évolution morphologique et que la similitude alimentaire peut être corrélée avec une similitude d'intégration morphologique. Ces analyses ont également révélé des différences en terme d'intégration morphologique entre Primates et Carnivores. En utilisant le modèle myrmécophage, nous chercherons à tester si les niveaux d'intégration peuvent varier avec la réduction dentaire et si les patrons d'intégration ont convergé entre lignées indépendantes. Nous utiliserons des méthodes établies pour quantifier la modularité (Goswami 2006b) afin de comparer la disparité entre les six modules crâniens précédemment identifiés et d'explorer la relation entre l'intégration morphologique et le régime alimentaire. Si les modules varient de façon similaire entre les espèces myrmécophages, alors il doit être possible de déterminer comment l'intégration des traits morphologiques limite de l'évolution de leur crâne, probablement en limitant la variation disponible pour la sélection.

Variation Intraspécifique et Relâchement des Pressions de Sélection

Darwin a, le premier, émis l'hypothèse selon laquelle un organe lorsqu'il devient inutile devrait être plus variable puisque sa variation n'est plus soumise à la sélection naturelle (Darwin 1859). Les fourmiliers du genre *Tamandua* montrent justement une grande variation dans la forme du crâne relativement à d'autres mammifères (Hautier et al. 2014). Nous chercherons ainsi à caractériser en détail les niveaux de variation intraspécifiques de la forme du crâne chez différentes espèces de myrmécophages (notamment fourmiliers et pangolins) et si ces patrons de variation s'étendent à d'autres caractères de leur anatomie. Des études antérieures ont en effet montré que, de concert avec la réduction de leur dentition, les fourmiliers possèdent un nombre et une complexité réduite des muscles de la face par rapport à la plupart des placentaires (Naples 1985), ce qui leur confère une face relativement immobile. La réduction de la musculature faciale associée à la simplification dentaire peut potentiellement entraîner une variation accrue, comme cela a été proposé pour l'oreille interne des paresseux chez lesquels les contraintes fonctionnelles réduites ont abouti à un relâchement des pressions de sélection (Billet et al. 2012). Nous proposons donc d'examiner si les espèces myrmécophages présentent une plus grande variation de la forme du crâne par rapport à d'autres espèces de mammifères dans le but de tester l'hypothèse selon laquelle une demande fonctionnelle réduite sur un organe va diminuer les contraintes sélectives permettant à de multiples phénotypes de coexister dans la population.

3.2.2 TACHE 2 : CONVERGENCE MOLECULAIRE A L'ECHELLE GENOMIQUE

Les principaux objectifs de cette tâche génomique sont de révéler les adaptations génomiques qui sous-tendent l'évolution convergente du phénotype myrmécophage chez les tatous, les fourmiliers, l'oryctérope, les pangolins, et le protèle. Les questions évolutives abordées dans cette tâche sont: (1) Quelles adaptations génomiques ont évolué en réponse aux contraintes sélectives imposées par le mode de vie myrmécophage ? ; (2) Les mêmes adaptations génomiques sous-tendent-elles les phénotypes convergents observés dans ces lignées indépendantes ? ; et (3) Quelle sont l'étendue et les modalités de l'évolution moléculaire convergente dans les génomes de placentaires myrmécophages ?

Acquisition de Données

L'avènement des nouvelles technologies de séquençage rend désormais possible l'étude des bases moléculaires de la convergence à l'échelle du génome. Les modalités génomiques de l'évolution convergente sont diverses et comprennent notamment les événements de pseudogénéisation parallèles, l'évolution parallèle de familles de gènes par duplication/perte, et des changements adaptatifs convergents dans les gènes protéiques. Ces trois aspects seront abordés dans un projet reposant sur l'acquisition de données génomiques pour les taxons myrmécophages et leurs espèces apparentées. Sur la base de critères phylogénétiques et pratiques, le fourmilier nain (*Cyclopes didactylus*), le fourmilier géant ou tamanoir (*Myrmecophaga tridactyla*), le tatou géant (*Priodontes maximus*), le pangolin africain (*Smutsia temminckii*), la hyène tachetée (*Crocuta crocuta*) et le protèle (*Proteles cristata*) ont été sélectionnés pour le séquençage du génome. Comme nos analyses de la convergence moléculaire se concentreront sur les séquences codantes, nous opterons pour un séquençage Illumina de couverture moyenne 30X qui devrait suffire à obtenir l'assemblage de ces régions. Ces nouvelles données, combinées avec les génomes disponibles du tatou à neuf bandes (*Dasypus novemcinctus*), de l'oryctérope (*Orycteropus afer*) et du pangolin chinois (*Manis pentadactyla*) et les autres mammifères permettront d'effectuer des analyses comparatives de l'évolution moléculaire convergente.

Approche Gènes Candidats

Les gènes candidats les plus prometteurs pour détecter des traces de l'évolution moléculaire convergente chez les myrmécophages sont sans doute ceux liés à l'alimentation, la recherche de nourriture, et la digestion. Les gènes codant pour des enzymes digestives telles que les amylases, la gastrine, les pepsinogènes, les lysozymes, les RNAses, et d'autres protéases digestives semblent tous être de bons candidats. Les plus évidents sont peut-être les chitinases impliquées dans la dégradation de la chitine qui constitue la carapace des insectes. Chez les Vertébrés, les chitinases dégradent la chitine en N-acétyl-glucosamine qui entre ensuite dans la voie métabolique des sucres aminés. Chez les mammifères, deux gènes de chitinases sont impliqués dans l'hydrolyse de la chitine: CHIT1 (chitinase 1 ou chitotriosidase) et CHIA (chitinase acide ou AMCCase). Même si leur fonction et leurs origines évolutives sont relativement mal connues (Bussink et al. 2007), elles semblent avoir une activité digestive dans le suc gastrique humain (Cozzarini et al. 2009). Enfin, les gènes orthologues simple-copie des récepteurs du goût (TAS1R), qui sont responsables des goûts sucré et umami, pourraient également avoir été affectés par l'adoption du régime myrmécophage. En effet, il a été récemment montré que les gènes du récepteur du goût sucré

TAS1R2 ont subi des événements de pseudogénération parallèles dans les lignées hyper-carnivores telles que les phoques et les hyènes (Jiang et al. 2012). Pour tous ces gènes candidats, nous effectuerons des analyses détaillées de l'évolution moléculaire basées sur des estimations par site et par branche du rapport entre substitutions non-synonymes et substitutions synonymes (dN/dS) afin de révéler l'existence potentielle de pressions sélectives convergentes et/ou d'événements de pseudogénération parallèles entre les lignées myrmécophages comme nous l'avons déjà fait pour les gènes de l'émail et de la dentine.

Evolution des Familles Multigéniques

Nous étudierons également l'évolution convergente de familles de gènes spécifiques tels que les répertoires olfactifs (OR), et les récepteurs du goût amer de type 2 (TAS2R). L'évolution de chacune de ses deux familles a en effet été influencée par l'écologie des espèces (Hayden et al. 2010 ; Hayakawa et al. 2014) et notamment leur régime alimentaire (Hayden et al. 2014 ; Li & Zhang 2014). Celles-ci apparaissent donc comme de bonnes candidates, car les espèces myrmécophages partagent des stratégies de recherche de nourriture similaires et essentiellement basées sur l'odorat et le goût. Les séquences homologues pour chaque famille seront alignées en incluant séquences fonctionnelles et non-fonctionnelles grâce au programme MACSE (**Ranwez et al. 2011**). Ceci permettra de reconstruire les répertoires OR et TAS2R de chaque espèce en utilisant des méthodes phylogénétiques probabilistes, et ainsi d'estimer le pourcentage de pseudogènes dans chaque sous-répertoire afin d'identifier d'éventuelles expansions et/ou pertes convergentes. Des analyses de sélection basées sur le dN/dS par branche permettront aussi de détecter d'éventuelles sous-familles ayant subi des épisodes de sélection positive.

Convergence Moléculaire à l'Echelle du Génome

Nous évaluerons ensuite deux méthodes pour détecter l'évolution moléculaire convergente à l'échelle génomique. Tout d'abord, nous allons construire un pipeline inspiré de la méthode récemment préconisée par Thomas & Hahn (2015) sur la base de la comparaison des substitutions convergentes et divergentes entre toutes les paires d'espèces pour détecter les changements d'acides aminés convergents. Cette méthode offre une alternative à la méthode de détection de la convergence sur la base de la phylogénie basée sur les distributions de différences de vraisemblance par site entre la phylogénie des espèces et la topologie convergente alternative (Castoe et al. 2009). Cette dernière approche a été appliquée avec succès pour détecter la convergence moléculaire dans l'audition des gènes entre les lignées écholocatrices (Parker et al. 2013), mais il a été montré qu'elle souffrait d'un grand nombre de faux positifs en raison des difficultés inhérentes à la définition d'un modèle nul adéquat (Thomas & Hahn 2015; Zou & Zhang 2015). Cependant, au lieu d'utiliser des modèles homogènes de substitution d'acides aminés pour effectuer la reconstruction des séquences ancestrales comme récemment effectué pour les mammifères marins (Foote et al. 2015), nous nous appuyerons sur le modèle de mélange site-hétérogène Bayésien CAT (Lartillot et Philippe 2004) implémenté dans PhyloBayes (Lartillot et al. 2009). En étant plus performant pour détecter les changements multiples à un site donné, le modèle CAT devrait fournir une meilleure discrimination entre les acides aminés divergents et convergents. Dans notre cas, ce modèle sera particulièrement utile puisque les mammifères myrmécophages appartiennent à des lignées phylogénétiquement distinctes et anciennes.

En parallèle, nous développerons une méthode complémentaire inspirée de celle utilisée pour détecter les gènes potentiellement associés à la longévité dans le génome des mammifères (Jobson et al. 2010). Cette méthode sera adaptée pour identifier les gènes présentant un excès significatif (indiquant une sélection positive) ou un déficit (indiquant une sélection purificatrice) de substitutions non-synonymes par rapport aux substitutions synonymes dans tout ou partie des branches menant aux espèces myrmécophages par rapport aux autres mammifères. Par rapport à la méthode précédente, cette dernière présente l'avantage de ne pas s'appuyer sur quelques substitutions d'acides aminés parallèles, mais plutôt de détecter les gènes où une accélération ou un ralentissement du taux de substitution a eu lieu simultanément dans les lignées myrmécophages. Ce type d'analyse sera effectué pour les gènes individuels, mais aussi sur les gènes regroupés selon leur Gene Ontology (GO) afin de potentiellement identifier certaines catégories fonctionnelles qui auraient pu être affectées par l'évolution convergente vers la myrmécophilie. La confrontation des résultats des deux méthodes devrait permettre l'identification des gènes présentant un excès de substitutions convergentes adaptatives dans les lignées myrmécophages.

3.2.3 TACHE 3 : EVOLUTION ET DES MICROBIOMES ORAUX ET INTESTINAUX

L'objectif principal de cette tâche consacrée à l'évolution des microbiomes est de mieux comprendre le rôle joué par le microbiome dans l'adaptation convergente au régime myrmécophage chez les mammifères. Ceci impliquera tout d'abord une comparaison à large échelle des compositions taxonomiques des microbiomes oraux et intestinaux couvrant la diversité des mammifères, notamment en termes de régime alimentaire. Le contenu fonctionnel en gènes de ces mêmes microbiomes chez les espèces myrmécophages et leurs proches parents sera ensuite exploré par une approche métagénomique sur des échantillons buccaux et fécaux, dans le but de révéler les voies de dégradation particulières qui pourraient caractériser ces espèces. En particulier, nous testerons l'hypothèse selon laquelle les espèces myrmécophages pourraient utiliser des bactéries symbiotiques pour digérer la chitine contenue dans les exosquelettes des fourmis et termites. Les questions fondamentales que nous aborderons dans cette tâche sont les suivantes : (1) A quel point les microbiomes de lignées indépendantes de myrmécophages se ressemblent-ils en termes de composition taxonomique et de contenu fonctionnel en gènes ? ; (2) Les espèces myrmécophages ont-elles recruté indépendamment des communautés microbiennes semblables pour digérer les fourmis et termites ? ; et si oui, (3) ces communautés sont-elles composées des mêmes taxons bactériens ou bien de taxons bactériens différents assurant les mêmes fonctions ?

Acquisition de Données

Des échantillons fécaux et frottis buccaux seront collectés à partir d'animaux de zoos et d'individus sauvages lorsque cela sera possible. Deux missions de terrain seront organisées pour recueillir des échantillons de populations sauvages de pangolins (*S. temminckii*) et de protèle (*P. cristata*) dans la Réserve de Tswalu Kalahari en Afrique du Sud. Des échantillons d'oryctérope (*O. afer*) y ont déjà été récoltés avec succès lors d'une précédente mission grâce à la collaboration de Nico Avenant (National Museum Bloemfontein). Des échantillons de tatou géant (*P. maximus*) et de fourmilier géant (*M. tridactyla*) provenant du parc National

Emas au Brésil sont déjà disponibles au travers d'une collaboration avec Carly Vynne (National Fish US and Wildlife Foundation) sur l'analyse moléculaire de leur régime alimentaire. Trois autres missions d'échantillonnage seront organisées en Guyane pour recueillir des échantillons fécaux de fourmilier nain (*C. didactylus*) et de tamandua (*T. tetradactyla*) à partir de spécimens collectés sur le bord des routes dans le cadre d'un suivi de l'impact de la circulation sur la faune sauvage mené par l'Association Kwata dirigée par Benoit de Thoisy (Institut Pasteur de Cayenne). Enfin, une large collection d'échantillons oraux et fécaux de mammifères a déjà été assemblée dans le cadre d'un projet sur le microbiome des vertébrés mené en collaboration avec le laboratoire de Rob Knight (<https://knightlab.ucsd.edu/>).

Composition des Microbiomes Oraux et Intestinaux

Cette première partie consistera à étendre l'étude précédemment menée par métabarcoding d'ARNr 16S d'échantillons fécaux sur les mammifères myrmécophages (Delsuc et al. 2014) en incluant un grand nombre d'échantillons supplémentaires couvrant la diversité taxonomique et les différents régimes alimentaires des mammifères, avec un focus particulier sur les espèces insectivores. L'analyse préliminaire de données supplémentaires (environ 300 échantillons) en utilisant les distances UniFrac entre communautés microbiennes (Lozupone & Knight 2005) montre le regroupement des échantillons principalement par régime alimentaire (Figure 23). Cette analyse qui comprend à la fois des échantillons de zoo et de terrain montre qu'il ne semble pas exister de différence fondamentale entre les deux types d'environnement. En tout cas, si effet il y a, celui-ci ne semble pas être assez fort pour masquer le signal de composition majoritairement porté par le régime alimentaire à l'échelle globale des mammifères.

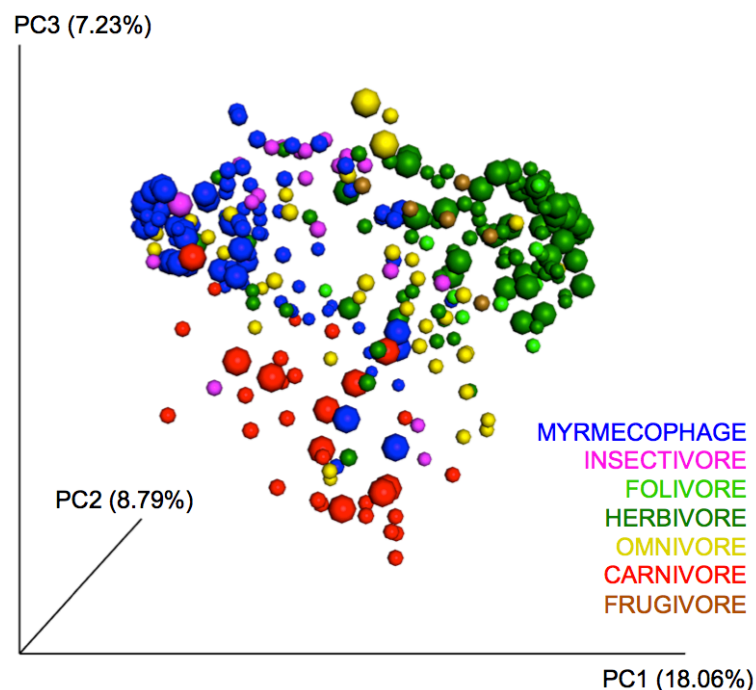


Figure 9 | La composition des microbiomes intestinaux de Mammifères est structurée principalement par le régime alimentaire. L'analyse en coordonnées principales (PCoA) est basée sur les distances UniFrac reflétant la proximité des communautés microbiennes en terme de diversité phylogénétique. Les points élargis représentent des échantillons de terrain par rapport aux échantillons de zoo.

Nous avons par ailleurs déjà recueilli des données de métabarcoding d'ARNr 18S afin de caractériser aussi la diversité eucaryote des échantillons de **Delsuc et al. (2014)**. Ces données offrent ainsi un premier aperçu moléculaire du régime alimentaire et de la charge parasitaire de populations sauvages de xénarthres. En guise d'exemple, la Figure 24 présente une comparaison de la diversité des Métazoaires détectés dans les fèces de tatou à neuf bandes (*D. novemcinctus*) qui est un insectivore généraliste, et le tatou géant (*P. maximus*) qui est lui myrmécophage. Le métabarcoding par ARNr 18S reflète fidèlement les différences de régime alimentaire entre les deux espèces avec une large diversité d'arthropodes, dont des fourmis et termites, observée chez le tatou à neuf bandes, alors que seuls des termites et des tiques sont détectés chez le tatou géant. Par ailleurs, ces données révèlent la présence de nématodes parasites du genre *Strongyloides* chez le tatou à neuf bandes et une large proportion de nématodes du genre *Aspidodera* chez le tatou géant. Ces nématodes sont bien connus pour infecter les mammifères néotropicaux, de nombreuses espèces du genre *Aspidodera* ayant d'ailleurs été récemment décrites chez différentes espèces de tatous (Jiménez et al. 2013).

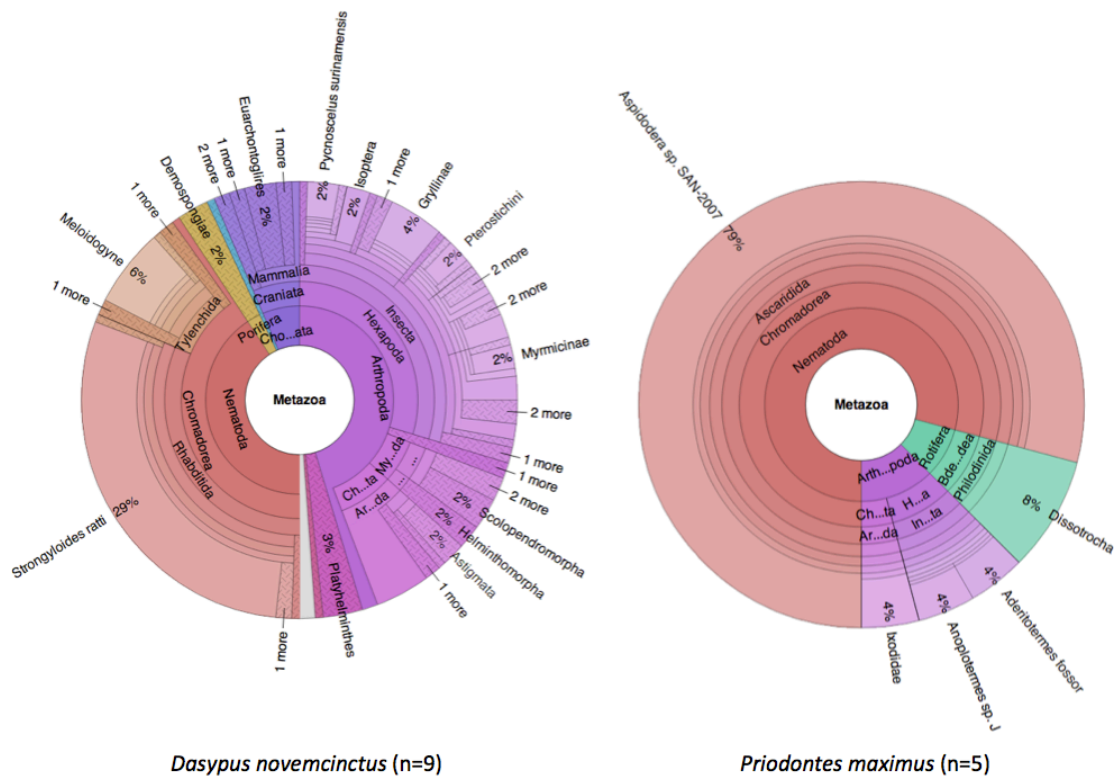


Figure 10 | Comparaison de la diversité des eucaryotes dans des échantillons fécaux de populations sauvages de tatous. Les graphes représentent les assignations taxonomiques obtenues pour les Métazoaires à partir de données de métabarcoding par ARNr 18S d'échantillons fécaux du tatou à neuf bandes (*D. novemcinctus*) qui est un insectivore généraliste et le tatou géant (*P. maximus*) qui est lui myrmécophage.

Le même type d'analyses combinant métabarcoding des ARNr 16S et 18S sera effectué sur des échantillons buccaux d'une diversité d'espèces de mammifères, afin d'étudier pour la première fois la structure et l'évolution du microbiome oral

dans le contexte de la phylogénie des mammifères afin de pouvoir le comparer au microbiome intestinal.

Exploration du Contenu Fonctionnel du Microbiome

Cette deuxième partie utilisera une approche métagénomique par séquençage shotgun d'échantillons fécaux et buccaux ciblés afin de caractériser et de comparer le contenu fonctionnel en gènes du microbiome du système digestif des mammifères. Nous mettrons notamment l'accent sur la recherche de gènes et de voies métaboliques impliqués dans la dégradation de la chitine, qui auraient pu être recrutés de façon convergente par les placentaires myrmécophages. Les gènes de chitinases sont très répandus chez les bactéries, les plantes et les champignons (Gooday 1990). La dégradation de la chitine est ainsi principalement assurée par des taxons bactériens tels que *Pseudomonas*, des bactéries entériques, des Actinomycètes, *Bacillus*, *Vibrio*. et *Clostridium*. Les chitinases sont également présentes chez les plantes, les invertébrés, et les vertébrés, y compris les mammifères. Il existe cependant trois sources possibles d'enzymes chitinolytiques dans le système digestif : l'animal lui-même, son microbiome intestinal, ou celui associé à la nourriture ingérée (Gooday 1990). Nos analyses métagénomiques des microbiomes oraux et intestinaux devraient permettre de distinguer entre ces différentes possibilités.

Cette approche métagénomique par séquençage shotgun Illumina sera appliquée à des échantillons oraux et fécaux représentatifs et sélectionnés à partir des résultats de métabarcoding précédemment obtenus pour les huit espèces suivantes: l'oryctérope (*O. afer*), le tatou géant (*P. maximus*), le fourmilier nain (*C. didactylus*), le fourmilier géant (*M. tridactyla*), le pangolin africain (*S. temminckii*), le pangolin chinois (*M. pentadactyla*), la hyène tachetée (*C. crocuta*) et le protèle (*P. cristata*). Ces données seront ensuite complétées par des données métagénomiques du même type obtenues pour une diversité de mammifères dans le cadre du projet sur le microbiome des vertébrés mené en collaboration avec le laboratoire de Rob Knight (<https://knightlab.ucsd.edu/>).

Le pipeline bioinformatique que nous allons utiliser pour traiter ces données est a été testé sur des données similaires déjà disponibles. Il s'appuie sur le projet MetAMOS (Treangen et al. 2013) qui fournit un pipeline flexible et permet de tester une variété d'algorithmes d'assemblage et d'annotation taxonomique et fonctionnelle des contigs produits. Nous extrairons également les contigs correspondants aux ARNr 16S et 18S par SortMeRNA (Kopylova et al. 2012) afin de déterminer le profil taxonomique des échantillons et de le comparer avec les données de métabarcoding obtenues pour les mêmes échantillons. Ce pipeline sera appliqué également aux données du même type déjà disponibles pour une quarantaine de mammifères (Muegge et al. 2011). Ces analyses nous permettront de caractériser respectivement les compositions taxonomique et fonctionnelle du microbiome intestinal à l'échelle des mammifères, et nous testerons plus particulièrement l'hypothèse d'un enrichissement en gènes impliqués dans la dégradation de la chitine chez les espèces insectivores et myrmécophages comme cela semble être le cas chez les cétacés zooplanctonophages (Sanders et al. 2015). Pour cela, nous nous focaliserons sur la détection d'enzymes impliquées dans des voies métaboliques particulières comme celle de la dégradation de la chitine par comparaison avec la base de données spécialisée CAZy qui référence les enzymes de la dégradation des hydrates de carbone auxquelles les chitinases appartiennent (Lombard et al. 2014). L'objectif est de construire des arbres phylogénétiques de référence pour les séquences des familles de Glycoside Hydrolases GH18 (Chitinases III, V) et

GH19 (Chitinases I, II, IV) en y incluant les séquences issues de nos données métagénomiques (Figure 25). Ceci nous permettra à la fois d'assigner les différentes séquences aux taxons bactériens pour lesquels des chitinases sont déjà connues et de tester statistiquement si les séquences issues des microbiomes de mammifères se distribuent en fonction de la taxonomie ou du régime alimentaire.

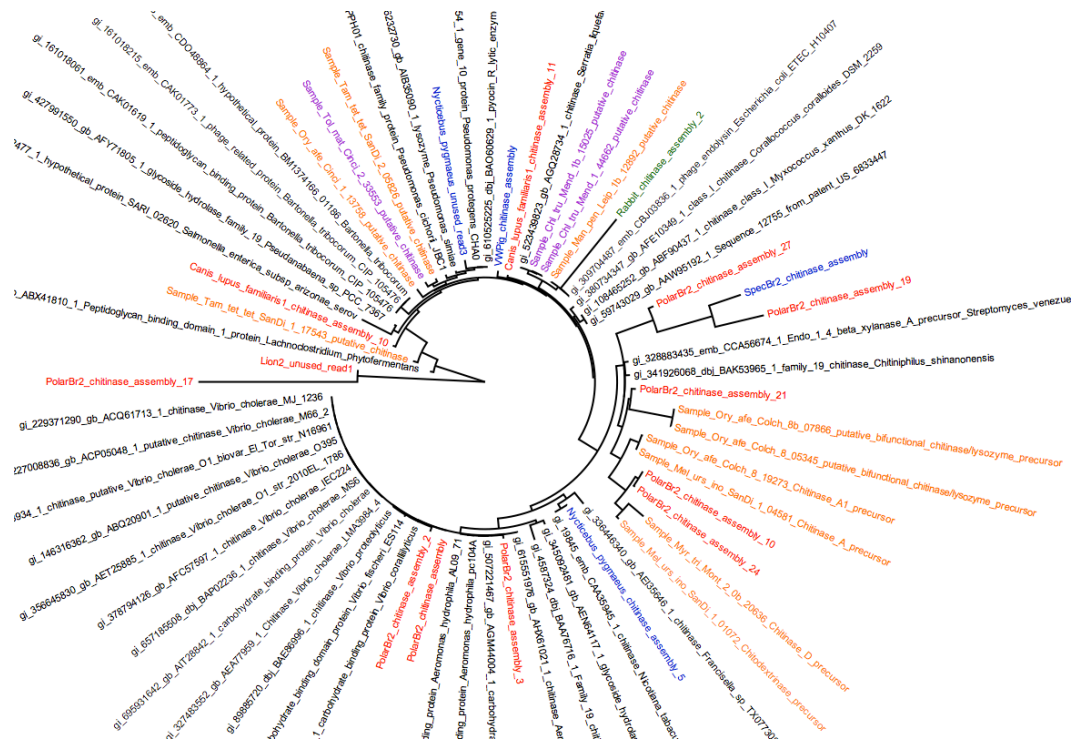


Figure 11 | Phylogénie des séquences de chitinases de la famille CAZY des Glycoside Hydrolases 19 (Chitinases I, II, IV). Les séquences issues du séquençage métagénomique d'échantillons fécaux de mammifères sont coloriées par régime alimentaire : Carnivore (Rouge), Herbivore (Vert), Omnivore (Bleu), Myrmécophage (Orange), et Insectivore (Violet).

4. CURRICULUM VITAE

4.1. INFORMATIONS PERSONNELLES

Frédéric Delsuc

Né le 11 octobre 1975 à Issoire (40 ans)

Marié à Nathalie Delsuc (née Escure) depuis le 7 Août 2004.

Deux enfants : Evan Delsuc né le 08 Décembre 2005 et Lémy Delsuc né le 26 Octobre 2009.

Adresse personnelle: 14 rue des Pétunias, 34070 Montpellier, France.

Page professionnelle: <http://www.isem.univ-montp2.fr/recherche/equipes/phylogenie-et-evolution-moleculaire/personnel/delsuc-frederic-2/>

Site web personnel: <http://fdelsuc.perso.neuf.fr/>

4.2. CURSUS UNIVERSITAIRE ET PROFESSIONNEL

Oct. 2009- : **Chargé de Recherche CNRS 1^{ère} Classe (section 22)**

Laboratoire : UMR5554 - Institut des Sciences de l'Évolution de Montpellier

Directeur : Agnès Mignot

Oct. 2005-Sept. 2009 : **Chargé de Recherche CNRS 2^{ème} Classe (section 22)**

Laboratoire : UMR5554 - Institut des Sciences de l'Évolution de Montpellier

Directeur : Jean-Christophe Auffray

Sept. 2004-Sept. 2005 : **Stage Post-Doctoral 2** (Université de Montréal, Québec, Canada)

Directeur : Pr. Hervé Philippe

Laboratoire : Centre de Recherche en Bioinformatique et Génomique Robert Cedergren

Sujet : La place des Tuniciers (Urochordés) au sein des animaux deutérostomiens : une approche phylogénomique et méthodologique.

Financement : Bourse Génome Québec (Salaire & Fonctionnement).

Fév. 2003-Fév. 2004 : **Stage Post-Doctoral 1** (Massey University, New Zealand)

Directeur : Pr. David Penny

Laboratoire : Allan Wilson Centre for Molecular Ecology and Evolution

Sujet : Réconciliation des arbres mitochondriaux et nucléaires en phylogénie des mammifères placentaires par le séquençage et l'analyse bioinformatique de génomes mitochondriaux complets de Xénarthres.

Financement : Bourse Lavoisier du Ministère Français des Affaires Étrangères (Salaire) et Allan Wilson Centre (Fonctionnement).

Nov. 1999-Déc. 2002 : **Thèse de Doctorat** (Université Montpellier 2)

Directeur : Pr. Emmanuel Douzery

Laboratoire : Institut des Sciences de l'Évolution de Montpellier (ISEM)

Titre : Phylogénie moléculaire des Xénarthres (tatous, fourmiliers et paresseux) : application des méthodes probabilistes à la reconstruction de leur histoire évolutive au sein des Mammifères placentaires

Mention : Très Honorable.

1998-1999 : **DEA** Biologie de l'Evolution et Ecologie (Université Montpellier 2)

Directeur : Pr. Emmanuel Douzery

Titre : Apports des phylogénies mitochondriales et nucléaires à la systématique et à la biogéographie des tatous, fourmiliers et paresseux (Mammalia, Xenarthra).

Mention : Bien (3^{ème})

1997-1998 : **Maîtrise** de Biologie des Populations et des Ecosystèmes (Université Montpellier 2)

Mention : Bien (11^{ème})

1996-1997 : **Licence** de Biologie des Organismes (Université Clermont-Ferrand 2)

Mention : Bien (1^{er} ex-aequo)

1993-1996 : **DEUG** Sciences Naturelles (Université Clermont-Ferrand 2)

Mention : Assez Bien (12^{ème})

1993 : **Baccalauréat** série D (Lycée Murat, Issoire, Puy-de-Dôme)

4.3. DISTINCTIONS

2015 : ERC Consolidator Grant 2015 décerné par l'European Research Council.

2012 : Bourse Fulbright 2012 décernée par la commission Franco-Américaine.

2010 : Médaille de bronze du CNRS décernée par la section 22.

2004 : Prix de Thèse décerné par la Société de Biologie de Montpellier Languedoc Roussillon.

4.4. PRODUCTION SCIENTIFIQUE

4.4.1 BIBLIOMETRIE

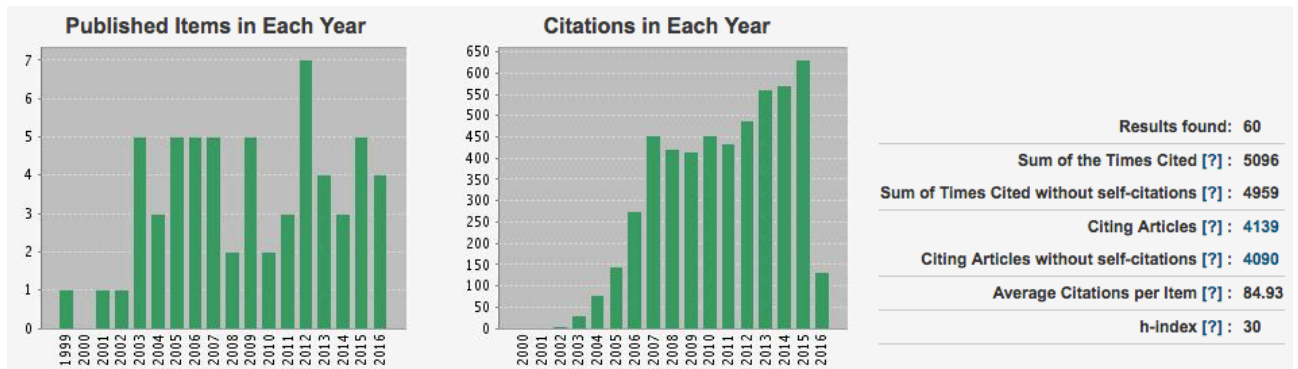
Source ISI Web of Science

60 publications ; Nombre total de citations = 5096 ; Nombre moyen de citations = 84 ;

H index = 30 ; i10 = 45.

14 publications avec plus de 100 citations dont 2 à plus de 500 citations.

18 en premier auteur ; 8 en senior auteur.



L'intégralité de mes publications est disponible au format pdf sur ma page web personnelle à l'adresse : http://fdelsuc.perso.neuf.fr/fd_pubs.html/

Profil Google Scholar : <http://tinyurl.com/ajfd65f>


4.4.2 PUBLICATIONS DANS REVUES A COMITE DE LECTURE



* : Contribution égale ; Etudiant encadré.



IF : Facteur d'Impact de l'année de publication (sauf 2015-2016)

ISI : Citations ISI Web of Knowledge ; **GS** : Citations Google Scholar.



- 65.** Philippe H., De Vienne D., Ranwez V., Roure B., Baurain D. & **Delsuc F.** Pitfalls of supermatrix phylogenomics. *European Journal of Taxonomy* soumis.
- 64.** Nardelli M., Ibáñez E.A., Dobler D., Justy F., **Delsuc F.**, Abba A.M., Cassini M.H. & Túnez J.I. Genetic structuring in a relictual population of screaming hairy armadillo (*Chaetophractus vellerosus*) in Argentina revealed by a set of novel microsatellite loci. *Genetica* soumis.
- 63.** Malicorne S., Vernochet C., Cornelis G., Mulot B., **Delsuc F.**, Heidmann O., Heidmann T. & Dupressoir A. Genomewide screening of endogenous retroviral envelope genes in the nine-banded armadillo (*Dasypus novemcinctus*; Xenarthra) reveals ongoing endogenization of a chimeric betaretrovirus using the ASCT2 receptor. *Journal of Virology* soumis.
- 62.** Botero-Castro F., Tilak M.-K., Justy F., Catzeflis F., **Delsuc F.** & Douzery E.J.P. In cold blood: compositional bias and positive selection drive the high evolutionary rate of vampire bats mitochondrial genomes. *Genome Biology and Evolution* en révision.
- 61.** Kadowaki K., Barbera C.G., Godsoe W., **Delsuc F.** & Mouquet N. Predicting biotic interactions and their variability in a changing environment. *Biology Letters* en révision.
- 60.** **Delsuc F.**, Gibb G.C., Kuch M., Billet G., Hautier L., Southon J., Rouillard J.-M., Fernicola J.C., Vizcaíno S.F., McPhee R.D.M & Poinar H.N. (2016). The phylogenetic affinities of the extinct glyptodonts. *Current Biology* 24: R155-R156. [**IF** : 9,6 ; **ISI** : 0 ; **GS** : 0]
- 59.** Gibb G.C., Condamine F., Kuch M., Enk J., de Moraes-Barros N., Superina M., Poinar H.N. & **Delsuc F.** (2016). Shotgun mitogenomics provides a

- reference phylogenetic framework and timescale for living xenarthrans. ***Molecular Biology and Evolution*** 33: 621-642. [IF : 9,1 ; ISI : 0 ; GS : 1]
58. Brozovic M.*, Martin C.*, Dantec C.*, Dauga D.*, Mendez M., Simion P., Percher M., Laporte B., Scornavacca C., Fujiwara S., Gineste M., Lowe E.K., Piette J., Sasakura Y., Takatori N., Brown T.C., **Delsuc F.**, Douzery E.J.P., Gissi C., McDougall A., Nishida H., Sawada H., Swalla B.J., Yasuo H. & Lemaire P. (2016). ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. ***Nuclear Acids Research*** 44: D808-D818. [IF : 9,1 ; ISI : 2 ; GS : 5]
57. Botero-Castro F., **Delsuc F.** & Douzery E.J.P. (2016). Thrice better than once: quality control guidelines to validate new mitogenomes. ***Mitochondrial DNA*** 27: 449-454. [IF : 1,7 ; ISI : 2 ; GS : 8]
56. Abba A.M., Cassini G.H., Valverde G., Tilak M.-K., Vizcaíno S.F., Superina M. & **Delsuc F.** (2015). Systematics of hairy armadillos and the taxonomic status of the Andean hairy armadillo (*Chaetophractus nationi*). ***Journal of Mammalogy*** 96: 673-689. [IF : 1,8 ; ISI : 3 ; GS : 4]
55. **Delsuc F.**, Gasse B. & Sire J.-Y. (2015). Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. ***BMC Evolutionary Biology*** 15: 148. [IF : 3,4 ; ISI : 2 ; GS : 2]
54. Fort P., Kajava A.V., **Delsuc F.** & Coux O. (2015). Evolution of proteasome regulators in Eukaryotes. ***Genome Biology and Evolution*** 7: 1363-1379. [IF : 4,2 ; ISI : 1 ; GS : 2]
53. **Delsuc F.** & Tilak M.-K. (2015). Naked but not *hairless*: the pitfalls of analyses of molecular adaptation based on few genome sequence comparisons. ***Genome Biology and Evolution*** 7: 768-774. [IF : 4,2 ; ISI : 2 ; GS : 3]
52. Tilak M.-K.*, Justy F.*, Debiais-Thibaud M., Botero-Castro F., **Delsuc F.** & Douzery E. J. P. (2015). A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. ***Conservation Genetics Resources*** 7: 37-40. [IF : 1,2 ; ISI : 1 ; GS : 3]
51. Malé P.-J.G., Bardon L., Besnard G., Coissac E., **Delsuc F.**, Engel J., Lhuillier E., Scotti-Saintagne C., Tinaut A. & Chave J. (2014). Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. ***Molecular Ecology Resources*** 14: 966-975. [IF : 5,6 ; ISI : 16 ; GS : 19]
50. Douzery E.J.P., Scornavacca C., Romiguier J., Belkhir K., Galtier N., **Delsuc F.** & Ranwez V. (2014). OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. ***Molecular Biology and Evolution*** 31: 1923-1928. [IF : 14,3 ; ISI : 10 ; GS : 13]
49. **Delsuc F.**, Metcalf J.L., Wegener Parfrey L., Song S.J., González A. & Knight R. (2014). Convergence of gut microbiomes in myrmecophagous mammals. ***Molecular Ecology*** 23: 1301-1317.  Highly Cited Paper [IF : 5,8 ; ISI : 22 ; GS : 35]
48. Botero-Castro F., Tilak M.-K., Justy F., Catzeflis F., **Delsuc F.** & Douzery E.J.P. (2013). Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed

- bats (Phyllostomidae). ***Molecular Phylogenetics and Evolution*** 69: 728-739. [IF : 4,1 ; ISI : 16 ; GS : 30]
47. Romiguier J., Ranwez V., **Delsuc F.**, Galtier N. & Douzery E.J.P. (2013). Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. ***Molecular Biology and Evolution*** 30: 2134-2144. [IF : 14,3 ; ISI : 33 ; GS : 48]
46. Nagy G.N., Marton L., Krámos B., Oláh J., Revész A, Vékey K., **Delsuc F.**, Hunyadi-Gulyás E., Medzihradzsky K.F., Lavigne M., Vial H., Cerdan R. & Vértessy B.G. (2013). Evolutionary and mechanistic insights into substrate and product accommodation of CTP:phosphocholine cytidyltransferase in *Plasmodium falciparum*. ***FEBS Journal*** 280: 3132-3148. [IF : 3,9 ; ISI : 7 ; GS : 9]
45. Rubinstein N.D., Feldstein T., Shenkar N., Botero-Castro F., Griggio F., Mastrototaro F., **Delsuc F.**, Douzery E.J.P., Gissi C. & Huchon D. (2013). Deep sequencing of mixed total DNA without barcodes allows efficient assembly of highly plastic ascidian mitochondrial genomes. ***Genome Biology and Evolution*** 5: 1185-1199. [IF : 4,5 ; ISI : 15 ; GS : 20]
44. Lartillot N. & **Delsuc F.** (2012). Joint reconstruction of divergence times and life history evolution in placental mammals using a phylogenetic covariance model. ***Evolution*** 66: 1773-1787. [IF : 4,9 ; ISI : 19 ; GS : 29]
43. Laguette N.*, Rahm N.*, Sobhian B., Chable-Bessia C., Munch J., Snoeck J., Sauter D., Switzer W.M., Heneine W., Kirchhoff F., **Delsuc F.***, Telenti A.* & Benkirane M.* (2012). Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. ***Cell Host & Microbe*** 11: 211-217.  Highly Cited Paper [IF : 12,6 ; ISI : 97 ; GS : 121]
42. **Delsuc F.**, Superina M., Tilak M.-K., Douzery E.J.P & Hassanin A. (2012). Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. ***Molecular Phylogenetics and Evolution*** 62: 673-680. [IF : 4,1 ; ISI : 29 ; GS : 31]
41. Chiari Y., Cahais V., Galtier N. & **Delsuc F.** (2012). Phylogenomic analyses support the position of turtles as the sister-group of birds and crocodiles (Archosauria). ***BMC Biology*** 10: 65.  Highly Cited Paper [IF : 6,5 ; ISI : 90 ; GS : 128]
40. Pointer M.A., Kamilar J.M., Warmuth V., Chester S.G.B., **Delsuc F.**, Mundy N.I., Asher R.J. & Bradley B.J. (2012). RUNX2 tandem repeats and the evolution of facial length in placental mammals. ***BMC Evolutionary Biology*** 12: 103. [IF : 3,3 ; ISI : 10 ; GS : 13]
39. Fort P.*, Albertini A.*, Van-Hua A., Berthomieu A., Roche S., **Delsuc F.**, Pasteur N., Capy P., Gaudin Y. & Weill M. (2012). Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution and potential functionality. ***Molecular Biology and Evolution*** 29: 367-379. [IF : 10,4 ; ISI : 28 ; GS : 39]
38. Hassanin A., **Delsuc F.**, Ropiquet A., Hammer C., van Vuuren BJ, Matthee C., Ruiz-Garcia M., Catzeflis F., Areskoug V., Nguyen T.T. & Couloux A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria) as revealed by a comprehensive analysis of mitochondrial

- genomes. **Comptes Rendus Biologies** 335: 32-50.  Highly Cited Paper [IF : 1,8 ; ISI : 108 ; GS : 133]
37. Atyame C.M., **Delsuc F.**, Pasteur N., Weill M. & Duron O. (2011). Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. **Molecular Biology and Evolution** 28: 2761-2772. [IF : 5,6 ; ISI : 32 ; GS : 40]
36. Dsouli N., **Delsuc F.**, Michaux J.R., De Stordeur E., Couloux A., Veuille M. & Duvallet G. (2011). Phylogenetic analyses of mitochondrial and nuclear data in haematophagous flies support the paraphyly of the genus *Stomoxys* (Diptera: Muscidae). **Infection, Genetics and Evolution** 11: 663-670. [IF : 3,1 ; ISI : 7 ; GS : 11]
35. Ranwez V., Harispe S., **Delsuc F.** & Douzery E.J.P. (2011). MACSE: Multiple Alignment of Coding Sequences accounting for frameshifts and stop codons. **PLoS ONE** 6: e22594. [IF : 4,9 ; ISI : 65 ; GS : 91]
34. Denoeud F., Henriot S., Mungpakdee S., Aury J.-M., Da Silva C., Brinkmann H., Mikhaleva J., Olsen L.C., Jubin C., Cañestro C., Bouquet J.-M., Danks G., Poulain J., Campsteijn C., Adamski M., Cross I., Yadetie F., Muffato M., Louis A., Butcher S., Tsagkogeorga G., Singh S., Jensen M.F., Huynh Cong E., Eikeseth-Otteraa H., Noel B., Anthouard V., Porcel-Setterblad B., Kachouri-Lafond R., Nishino A., Ugolini M., Chourrout P., Nishida H., Aasland R., Huzurbazar S., Westhof E., **Delsuc F.**, Lehrach H., Reinhardt R., Weissenbach J., Roy S.W., Artiguenave F., Postlethwait J.H., Manak J.R., Thompson E.M., Jaillon O., Du Pasquier L., Boudinot P., Liberles D.A., Volff J.-N., Philippe H., Lenhard B., Roest Crolius H., Wincker P. & Chourrout D. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. **Science** 330: 1381-1385.  [IF : 31,4 ; ISI : 99 ; GS : 128]
33. Tsagkogeorga G., Turon X., Galtier N., Douzery E.J.P. & **Delsuc F.** (2010). Accelerated evolutionary rate of housekeeping genes in tunicates. **Journal of Molecular Evolution** 71: 153-167. [IF : 2,3 ; ISI : 14 ; GS : 14]
32. Guindon S., **Delsuc F.**, Dufayard J.-F. & Gascuel O. (2009). Estimating maximum likelihood phylogenies with PhyML. **Methods in Molecular Biology** 537: 113-137. [IF : 1,3 ; ISI : 268 ; GS : 345]
31. Loughry W.J., Truman R.W., McDonough C.M., Tilak M.-K., Garnier S. & **Delsuc F.** (2009). Is leprosy spreading among nine-banded armadillos from the southeastern United States? **Journal of Wildlife Diseases** 45: 144-152. [IF : 1,4 ; ISI : 18 ; GS : 22]
30. Singh T. R.*, Tsagkogeorga G.*, **Delsuc F.**, Shenkar N., Loya Y., Douzery E.J.P. & Huchon D. (2009). Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. **BMC Genomics** 10: 534. [IF : 3,8 ; ISI : 28 ; GS : 41]
29. Tsagkogeorga G., Turon X., Hopcroft R.R., Tilak M.-K., Feldstein T., Shenkar N., Loya Y., Huchon D., Douzery E.J.P. & **Delsuc F.** (2009). An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. **BMC Evolutionary Biology** 9: 187. [IF : 4,3 ; ISI : 53 ; GS : 71]
28. Ranwez V., Clairon N., **Delsuc F.**, Pournali S., Auberval N., Diser S. & Berry V. (2009). PhyloExplorer: a web server to validate, explore and query

- phylogenetic trees. **BMC Evolutionary Biology** 9: 108. [IF : 4,3 ; ISI : 6 ; GS : 14]
27. Delsuc F., Tsagkogeorga G., Lartillot N. & Philippe H. (2008). Additional molecular support for the new chordate phylogeny. **Genesis** 46: 592-604. [IF : 2,2 ; ISI : 100 ; GS : 130]
26. Pantalacci S., Chaumot A., Benoît G., Sadier A., Delsuc F., Douzery E.J.P. & Laudet V. (2008). Conserved features and evolutionary shifts of the EDA signaling pathway involved in vertebrate skin appendage development. **Molecular Biology and Evolution** 24: 2573-2582. [IF : 7,3 ; ISI : 21 ; GS : 30]
25. Möller-Krull M.*, Delsuc F.*, Churakov G., Marker C., Superina M., Brosius J., Douzery E.J.P.* & Schmitz J.* (2007). Retroposed elements and their flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters and sloths). **Molecular Biology and Evolution** 24: 2573-2582. [IF : 6,7 ; ISI : 42 ; GS : 52]
24. Delsuc F., Superina M., Ferraris G., Tilak M.-K. & Douzery E.J.P. (2007). Molecular evidence for hybridisation between the two living species of South American ratites: potential conservation implications. **Conservation Genetics** 8: 503-507. [IF : 1,4 ; ISI : 5 ; GS : 8]
23. Slack K.E., Delsuc F., McLenachan P.A., Arnason U. & Penny D. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. **Molecular Phylogenetics and Evolution** 42: 1-13. [IF : 3,5 ; ISI : 70 ; GS : 99]
22. Ranwez V., Delsuc F., Ranwez S., Belkhir K., Tilak M.-K. & Douzery E.J.P. (2007). OrthoMaM: a database of orthologous genomic makers for placental mammal phylogenetics. **BMC Evolutionary Biology** 7: 241. [IF : 4,1 ; ISI : 66 ; GS : 84]
21. Chourrout D., Delsuc F., Chourrout P., Edvarsen R.B, Renfer E., Jensen M.F., Zhu B., de Jong P., Steele R.E. & Technau U. (2006). Minimal ProtoHox cluster inferred from comparison of bilaterian and cnidarian Hox complements. **Nature** 442: 684-687. [IF : 26,7 ; ISI : 113 ; GS : 138]
20. Delsuc F., Baurain D. & Philippe H. (2006). Vertebrate origins: does the tunic make the man? **Médecine/Sciences** 22: 688-690. [IF : 0,5 ; ISI : 10 ; GS : 12]
19. Jeffroy O., Brinkmann H., Delsuc F. & Philippe H. (2006). Phylogenomics: the beginning of incongruence? **Trends in Genetics** 22: 225-231.  [IF : 10,0 ; ISI : 264 ; GS : 341]
18. Douzery E.J.P., Delsuc F. & Philippe H. (2006). Molecular dating in the genomic era. **Médecine/Sciences** 22: 374-380. [IF : 0,5 ; ISI : 6 ; GS : 6]
17. Delsuc F., Brinkmann H., Chourrout D. & Philippe H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. **Nature** 439: 965-968.   [IF : 26,7 ; ISI : 831 ; GS : 1013]
16. Delsuc F., Brinkmann H. & Philippe H. (2005). Phylogenomics and the reconstruction of the tree of life. **Nature Reviews Genetics** 6: 361-375. [IF : 19,2 ; ISI : 564 ; GS : 787]
15. Holland B.R., Delsuc F. & Moulton V. (2005). Visualising conflicting evolutionary hypotheses in large collections of trees: using consensus

- networks to study the origins of placentals and hexapods. ***Systematic Biology*** 54: 66-76. [IF : 10,3 ; ISI : 45 ; GS : 60]
14. Knapp M., Stöckler K., Havell D., Delsuc F., Sebastiani F. & Lockhart P.J. (2005). Relaxed molecular clock provides evidence for long-distance dispersal of *Nothofagus* (southern beech). ***PLoS Biology*** 3: e14. [IF : 14,7 ; ISI : 145 ; GS : 188]
 13. Philippe H., Zhou Y., Brinkmann H., Rodrigue N. & Delsuc F. (2005). Heterotachy and long-branch attraction in phylogenetics. ***BMC Evolutionary Biology*** 5: 50. [IF : 4,5 ; ISI : 140 ; GS : 207]
 12. Philippe H., Delsuc F., Brinkmann H. & Lartillot N. (2005). Phylogenomics. ***Annual Review of Ecology, Evolution and Systematics*** 36: 541-562. [IF : 10,1 ; ISI : 157 ; GS : 218]
 11. Douzery E.J.P., Snell E.A., Bapteste E., Delsuc F. & Philippe H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? ***Proceedings of the National Academy of Sciences of the USA*** 101: 15386-15391.  [IF : 10,5 ; ISI : 307 ; GS : 421]
 10. Phillips M.J., Delsuc F. & Penny D. (2004). Genome-scale phylogeny and the detection of systematic biases. ***Molecular Biology and Evolution*** 21: 1455-1458. [IF : 6,1 ; ISI : 244 ; GS : 307]
 9. Delsuc F., Vizcaíno S.F. & Douzery E.J.P. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. ***BMC Evolutionary Biology*** 4: 11. [IF : 4,4 ; ISI : 73 ; GS : 148]
 8. Delsuc F. (2003). Army ants trapped by their evolutionary history. ***PLoS Biology***, 1: e37. [IF : 10,4 ; ISI : 1 ; GS : 8]
 7. Delsuc F., Phillips M.J. & Penny D. (2003). Comment on "Hexapod origins: monophyletic or paraphyletic?" ***Science*** 301: 1482d. [IF : 29,2 ; ISI : 99 ; GS : 139]
 6. Delsuc F., Stanhope M.J. & Douzery E.J.P. (2003). Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes. ***Molecular Phylogenetics and Evolution*** 28: 261-275. [IF : 2,8 ; ISI : 49 ; GS : 59]
 5. Douady C.J., Delsuc F., Boucher Y., Doolittle W.F. & Douzery E.J.P. (2003). Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. ***Molecular Biology and Evolution*** 20: 248-254.  [IF : 6,1 ; ISI : 315 ; GS : 401]
 4. Douzery E.J.P., Delsuc F., Stanhope M.J. & Huchon D. (2003). Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals, and incompatibility among fossil calibrations. ***Journal of Molecular Evolution*** 57: S201-S213. [IF : 3,1 ; ISI : 54 ; GS : 80]
 3. Delsuc F., Scally M., Madsen O., Stanhope M.J., de Jong W.W., Catzeflis F.M., Springer M.S. & Douzery E.J.P. (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. ***Molecular Biology and Evolution*** 19: 1656-1671. [IF : 5,3 ; ISI : 165 ; GS : 204]
 2. Delsuc F., Catzeflis F.M., Stanhope M.J. & Douzery E.J.P. (2001). The evolution of armadillos, anteaters, and sloths depicted by nuclear and

mitochondrial phylogenies: implications for the status of the enigmatic fossil *Eurotamandua*. **Proceedings of the Royal Society of London, Biological Sciences** 268: 1605-1615. [IF : 3,2 ; ISI : 72 ; GS : 98]

1. Huchon D.*, **Delsuc F.***, Catzeflis F.M. & Douzery E.J.P. (1999). Armadillos exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasyus novemcinctus* (Xenarthra). **Molecular Ecology** 8: 1743-1748. [IF : 3,4 ; ISI : 27 ; GS : 31]

4.4.3 CHAPITRES DE LIVRES

4. Raymond M., B. Godelle., Alizon S., Bourguet D., Caillaud D., Couvet D., David J., David P., Debussche M., **Delsuc F.**, Douzery E.J.P., Machon N., Pascal R., Robert A., Ronfort J., Sarrazin F., Schoenauer M., Tenaillon O., Huchard H., Theodorou K., Thompson J. (2010). Chap. 19 : Applications de la biologie évolutive. Dans « **Biologie évolutive** » édité par F. Thomas, T. Lefèvre & M. Raymond, pp. 749-788, De Boek Université, Bruxelles, Belgique. [GS : 0]
3. Douzery E.J.P., Blanquart S., Criscuolo A., **Delsuc F.**, Douady C., Lartillot N., Philippe H., Ranwez V. (2010). Chap. 6 : Phylogénie moléculaire. Dans « **Biologie évolutive** » édité par F. Thomas, T. Lefèvre & M. Raymond, pp. 183-244, De Boek Université, Bruxelles, Belgique. [GS : 2]
2. **Delsuc F.** & Douzery E.J.P. (2009). Armadillos, anteaters and sloths (Xenarthra). In **The Timetree of Life** edited by S. B. Hedges & S. Kumar, pp. 475-478, Oxford University Press, London, UK. [GS : 8]
1. **Delsuc F.** & Douzery E.J.P. (2008). Recent advances and future prospects in xenarthran molecular phylogenetics. Dans « **Biology of the Xenarthra** » édité par W. J. Loughry & S. F. Vizcaíno, pp. 11-23, University Press of Florida, Gainesville, FL. [GS : 42]

4.4.4 ACTES DE COLLOQUES A COMITE DE LECTURE

2. Pantalacci S., Chaumot A., **Delsuc F.**, Sadier A., Douzery E.J.P. & Laudet V. (2007). Evolution of the EDA pathway in vertebrates: a step towards tooth "Evo-Devo studies". **European Cells and Materials** 14 suppl. 2: 118. [Meeting abstract].
1. **Delsuc F.**, Superina M., Tilak M.-K., Douzery E.J.P. & Hassanin A. (2007). Advances in molecular and phylogenetic studies of Xenarthra. **Journal of Morphology** 268: 1066 [ICVM8-Abstracts].

4.4.5 AUTRES PUBLICATIONS

6. **Delsuc F.** (2015). Flawed evidence for convergent evolution of the circadian CLOCK gene in mole-rats. **bioRxiv** doi:10.1101/022004. [GS : 0]

5. Theulier Saint-Germain S., Lallement J.-C., Beaumes V. & **Delsuc F.** (2011). Quand l'évolution se prête au jeu... *La Lettre de l'OCIM* 136: 14-20. [GS : 0]
4. Douzery E.J.P. & **Delsuc F.** (2010). Les horloges de l'évolution. *Pour La Science* 397: 102-107. [GS : 0]
3. **Delsuc F.** & Douzery E.J.P. (2004). Les méthodes probabilistes en phylogénie moléculaire. (2) L'approche Bayésienne. *Biosystema : « Avenir et pertinence des méthodes d'analyses en phylogénie moléculaire »* 22: 75-86. [GS : 10]
2. **Delsuc F.** & Douzery E.J.P. (2004). Les méthodes probabilistes en phylogénie moléculaire. (1) Les modèles d'évolution des séquences et le maximum de vraisemblance. *Biosystema : « Avenir et pertinence des méthodes d'analyses en phylogénie moléculaire »* 22: 59-74. [GS : 5]
1. **Delsuc F.**, Mauffrey J.-F. & Douzery E.J.P. (2003). Une nouvelle classification des Mammifères. *Pour La Science* Janvier: 62-66. [GS : 5]

4.4.6 COMMUNICATIONS INVITEES

7. **Delsuc F.** (2015). A decade of advances in xenarthran molecular phylogenies. *3rd Montpellier Omics Days*, Montpellier (France).
6. **Delsuc F.** (2014). A decade of advances in xenarthran molecular phylogenies. *94th Annual Meeting of the American Society of Mammalogists*, Oklahoma City (USA).
5. **Delsuc F.** (2013). Building a species-level phylogenetic framework and timescale for Xenarthran evolution using next-generation sequencing data. *59th Brazilian Congress of Genetics*, Aguas de Lindoia (Brazil).
4. **Delsuc F.** (2012). Phylogenomics of tunicates based on next-generation sequencing data. *EFOR Network of functional Studies in Model Organisms*, Paris (France).
3. **Delsuc F.** (2010). Phylogenomic reconstruction in the next-generation sequencing era. *International Symposium "Frontiers in Biodiversity: a Phylogenetic Perspective"*, Barcelone (Espagne).
2. **Delsuc F.** (2010). Phylogenomics and next-generation sequencing data: a case study with 454 in tunicates. *German Zoological Society (DZG) Graduate Meeting "Systematics in the Age of Genomics"*, Hambourg (Allemagne).
1. **Delsuc F.** (2008). La phylogénomique : promesses et défis de la reconstruction phylogénétique à l'échelle des génomes. *Rencontres ALPHY : Génomique Evolutive, Bioinformatique, Alignement et Phylogénie*, Lyon (France).

4.4.7 COMMUNICATIONS DANS DES CONGRES

CO : Communication orale ; Orateur. **P** : Poster ; Concepteur.

- 45. Delsuc F.**, Belkhir K. & Scornavacca C. (2014). Exploring topological incongruence for detecting contaminations in phylogenomic datasets. *Mathematical and Computational Evolutionary Biology Meeting*, Le Hameau de l'Etoile (France). **CO**.
- 44. Botero-Castro F.**, Tilak M.-K, Justy F., Catzeflis F., **Delsuc F.** & Douzery E.J.P. (2014). Mutation bias and positive selection as driving forces of mitochondrial genome evolution in vampire bats (Chiroptera: Phyllostomidae). *Annual Meeting of the Society for Molecular Biology and Evolution*, Puerto Rico (USA). **P**.
- 43. Botero-Castro F.**, Tilak M.-K, Justy F., Catzeflis F., **Delsuc F.** & Douzery E.J.P. (2014). Evolution of the mitochondrial genome in vampire bats. *ALPHY/PhyloSIB 2014: Swiss French Meeting on Bioinformatics and Evolutionary Genomics*, Lausanne, (Switzerland). **CO**.
- 42. Botero-Castro F.**, Tilak M.-K, Justy F., Catzeflis F., **Delsuc F.** & Douzery E.J.P. (2014). NGS and mitogenomics of phyllostomid bats are useful for biogeography. *3rd Meeting of the Network for Neotropical Biogeography*, Bogota, (Colombia). **CO**.
- 41. Delsuc F.**, Metcalf J., Wegener Parfrey L., Song S.J., Gonzalez A. & Knight R. (2013). Convergence of gut microbiomes in myrmecophagous mammals. *Conférence de Génomique Environnementale*, Rennes (France). **CO**.
- 40. Villegas-Ramirez B.**, Tilak M.-K. & **Delsuc F.** (2013). How many species of armadillos, anteaters and sloths? A first attempt at developing DNA barcoding markers for xenarthran species delimitation and conservation. *59th Brazilian Congress of Genetics*, Aguas de Lindoia, Sao Paulo (Brazil). **CO**.
- 39. Piette J.**, Tiozzo S., McDougall, Philippe H., **Delsuc F.**, Douzery E.J.P. & Lemaire P. (July 2013). The pelagic tunicate *Salpa fusiformis* as a model organism for the study of Thaliaceans, remote Ascidians relatives. *7th Tunicate Meeting*, Naples (Italy). **CO**.
- 38. Ranwez V.**, Lechevallier F. & **Delsuc F.** (2013). Accurate alignment of thousands of barcoding sequences using the Multiple Alignment of Coding Sequences toolkit (MACSE-tk). *12th European Conference on Computational Biology*, Berlin (Germany). **P**.
- 37. Delsuc F.**, Rahm N., Benkirane M. & Lartillot N. (2013). Molecular evolution of HIV host restriction factors in Primates using a new bayesian method for estimating branch-specific dN/dS ratios. *Mathematical and Computational Evolutionary Biology Meeting*, Le Hameau de l'Etoile (France). **P**.
- 36. Chiari Y.**, Cahais V., Galtier N. & **Delsuc F.** (2012). Phylogenomics finally settles the position of turtles in the tree of amniotes. *Rencontres ALPHY : Génomique Evolutive, Bioinformatique, Alignement et Phylogénie*, Banyuls-sur-Mer (France). **CO**.
- 35. Gibb G.**, KÜch M., Poinar H.N. & **Delsuc F.** (2012). Next-generation mitogenomics: new results from modern and fossil xenarthrans. *6th Annual New-Zealand Phylogenetics Meeting*, Kaikoura (New Zealand). **CO**.

- 34.** Gibb G., Küch M., Poinar H.N. & **Delsuc F.** (2011). New insights into xenarthran phylogeny: a study of armadillos, anteaters and sloths using complete mitochondrial DNA. *15th Evolutionary Biology Meeting*, Marseille (France). **CO.**
- 33.** Gibb G., Sérol D., Tibi A., Tilak M.-K., Sire J.-Y. & **Delsuc F.** (2011). Retracing the genetic origins of enamel loss during the evolutionary history of xenarthran mammals. *13th Congress of the European Society for Evolutionary Biology*, Tuebingen (Germany). **P.**
- 32.** **Delsuc F.** (2010). An empirical comparison of species tree estimation methods in the context of mammalian phylogenomics. *Evolution Conference 2010: Joint Meeting of the Society for the Study of Evolution, Society of American Naturalists, and Society of Systematic Biologists*, Portland, Oregon (USA). **CO.**
- 31.** Tsagkogeorga G., Singh T.R., **Delsuc F.**, Blanquart S., Shenkar N., Loya Y., Douzery E.J.P. & Huchon D. (2010). Tunicate mitogenomics and phylogenetics: the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *Annual Meeting of the Society for Molecular Biology and Evolution*, Lyon (France). **P.**
- 30.** Douzery E.J.P., Belkhir K., Berry V., **Delsuc F.** & Ranwez V. (2010). OrthoMaM et PhyloExplorer : Deux outils pour l'exploration et l'exploitation des marqueurs et arbres phylogénétiques. *Réunion Nationale de Génomique Evolutive, Bioinformatique, Alignement et Phylogénie (ALPHY 2010)*, Marseille (France). **CO.**
- 29.** **Delsuc F.** (2009). The role of molecular data in xenarthran conservation. *10th International Mammalian Congress*, Mendoza (Argentina). **CO.**
- 28.** Philippe H., Baurain D., Brinkmann H., Roure B., **Delsuc F.**, Lartillot N. & Douzery E.J.P. (2009). Molecular dating of animal diversification. *Annual Meeting of the Society for Molecular Biology and Evolution*, Iowa City, Iowa (USA). **CO.**
- 27.** **Delsuc F.**, Tsagkogeorga G., Lartillot N. & Philippe H. (2008). More data and better models provide further evidence that tunicates are the closest living relatives of vertebrates. *Annual Meeting of the Society for Molecular Biology and Evolution*, Barcelone (Espagne). **P.**
- 26.** Pantalacci S., Chaumot A., Benoit G., Sadier A., **Delsuc F.**, Douzery E.J.P. & Laudet V. (2008). Conserved features and evolutionary shifts of the EDA signaling pathway involved in vertebrate skin appendage development. *Annual Meeting of the Society for Molecular Biology and Evolution*, Barcelone (Espagne). **CO.**
- 25.** Douzery E.J.P., **Delsuc F.**, Belkhir K., Galtier N., Ranwez S., Tilak M.-K. & Ranwez V. (2008). Evolutionary dynamics and phylogenetic content of coding sequences in mammalian genomes. *Annual Meeting of the Society for Molecular Biology and Evolution*, Barcelone (Espagne). **P.**
- 24.** **Delsuc F.**, Superina M., Tilak M.-K., Douzery E.J.P. & Hassanin A. (2007). Advances in molecular and phylogenetic studies of Xenarthra. *8th International Congress of Vertebrate Morphology*, Paris (France). **CO.**
- 23.** Tsagkogeorga G., Tilak M.-K., Turon X., Douzery E.J.P. & **Delsuc F.** (2007). Inferring Tunicate phylogeny using 18S rRNA secondary structure models. *4th International Tunicate Meeting*, Villefranche sur Mer (France). **CO.**

- 22.** Ranwez V., **Delsuc F.**, Tilak M.-K., Ranwez S. & Douzery E.J.P. (2007). OrthoMaM: a database of candidate coding markers for mammalian phylogenomics. *Evolution Conference 2007: Joint Meeting of the Society for the Study of Evolution, Society of American Naturalists, and Society of Systematic Biologists*, Christchurch (Nouvelle-Zélande). **CO.**
- 21.** **Delsuc F.***, Krull M.*, Churakov G., Superina M., Brosius J., Schmitz J. & Douzery E.J.P. (2007). Phylogenetic analyses of retroposed elements and sequence data further resolve the evolutionary history of xenarthran mammals. *Evolution Conference 2007: Joint meeting of the Society for the Study of Evolution, Society of American Naturalists, and Society of Systematic Biologists*, Christchurch (Nouvelle-Zélande). **[Modérateur de la session]. CO.**
- 20.** Tsagkogeorga G., Tilak M.-K., Turon X., Douzery E.J.P. & **Delsuc F.** (2006). Molecular phylogeny of tunicates inferred from 18S rRNA sequences using secondary structure models. *10th Evolutionary Biology Meeting*, Marseille (France). **P.**
- 19.** Fabre P.-H., **Delsuc F.** & Douzery E.J.P. (2006). Compatibility among vertebrate paleontological calibration constraints: an evaluation from mitochondrial and nuclear DNA relaxed molecular clocks. *54th Symposium of Vertebrate Palaeontology and Comparative Anatomy*, Paris (France). **P.**
- 18.** **Delsuc F.**, Brinkmann H., Chourrout D. & Philippe H. (2006). Tunicates are the closest living relatives of Vertebrates. *1st International Phylogenomics Conference*, Sainte-Adèle, Québec (Canada). **P.**
- 17.** Jeffroy O., Brinkmann H., **Delsuc F.** & Philippe H. (2006). Phylogenomics: the beginning of incongruence. *1st International Phylogenomics Conference*, Sainte-Adèle Québec (Canada). **CO.**
- 16.** **Delsuc F.**, Douzery E.J.P. & Philippe H. (2006). Digging to the root of the placental tree. *1st International Phylogenomics Conference*, Sainte-Adèle Québec (Canada). **CO.**
- 15.** Philippe H., Brinkmann H., Rodriguez-Ezpeleta N., Gray M.W., Derelle E., **Delsuc F.** & Douzery E.J.P. (2005). A new timescale for the evolution of eukaryotes inferred from a taxon-rich genome-scale dataset. *18th Annual Meeting of the CIAR Evolutionary Biology Program*, Parksville, Colombie Britannique (Canada). **P.**
- 14.** Douzery E.J.P., Snell E.A., Bapteste E., **Delsuc F.** & Philippe H. (2004). A revised timescale for the evolutionary history of Eukaryotes based on relaxed molecular clock analyses of a phylogenomic dataset. *17th Annual Meeting of the CIAR Evolutionary Biology Program*, Lac Carling, Québec (Canada). **P.**
- 13.** **Delsuc F.**, Douzery E.J.P., Snell E.A., Bapteste E. & Philippe H. (2004). Dating the evolutionary history of eukaryotes: Does a relaxed molecular clock reconcile molecules and fossils? *Colloque de Bioinformatique Robert-Cedergren*, Montréal, Québec (Canada). **[Prix de la meilleure présentation orale postdoctorale]. [Modérateur de la Session]. CO.**
- 12.** Slack K.E., **Delsuc F.**, Gibb G.C., Morgan-Richards M., Trewick S., Harrison G.L., Phillips M.J., McLenachan P.A., Cooper A., Arnason U. & Penny D. (2004). Avian evolution using complete mitochondrial genome sequences. *51th Annual Meeting of Genetics Society of Australia*, Melbourne (Australie). **P.**

11. Knapp M., Stöckler K., Havell D., **Delsuc F.**, Sebastiani F. & Lockhart P.J. (2004). Fangorn forest is not a Gondwanan relic. *Evolution Conference 2004: Joint Meeting of the Society for the Study of Evolution, Society of Systematic Biologists and Society of American Naturalists*, Fort Collins, Colorado (Etats-Unis). **P.**
10. Knapp M., Stöckler K., Havell D., **Delsuc F.**, Sebastiani F. & Lockhart P.J. (2004). Fangorn forest is not a Gondwanan relic. *Systematic Association of New-Zealand (SYSTANZ) General Meeting*, Whakapapa Village (Nouvelle-Zélande). **CO.**
9. Douzery E.J.P., Snell E.A., Baptiste E., **Delsuc F.** & Philippe H. (2003). The timing of eukaryotic evolution by a relaxed clock reconciles molecules and fossils. *Gordon Research Conference on the Origin of Life*, Bates College, Maine (Etats-Unis). **CO.**
8. **Delsuc F.**, Penny D., Lockhart P.J. & Douzery E.J.P. (2003). Secondary structure alignment and model sensitivity analysis: A case study using the 12S rRNA of xenarthrans. *Evolution Conference 2003: Joint Meeting of the Society for the Study of Evolution, Society of Systematic Biologists and Society of American Naturalists*, Chico, Californie (Etats-Unis). **P.**
7. Penny D., **Delsuc F.** & Phillips M.J. (2003). RY coding overcomes the bias in mitochondrial data. *Evolution Conference 2003: Joint Meeting of the Society for the Study of Evolution, Society of Systematic Biologists and Society of American Naturalists*, Chico, Californie (Etats-Unis). **CO.**
6. Douady C.J., **Delsuc F.**, Boucher Y., Doolittle W.F. & Douzery E.J.P. (2003). Bayesian posterior probabilities and maximum likelihood bootstrap proportions: oranges versus apples? *Annual New-Zealand Phylogenetics Meeting*, Kaikoura (Nouvelle-Zélande). **CO.**
5. Douady C.J., **Delsuc F.**, Boucher Y., Doolittle W.F. & Douzery E.J.P. (2002). Comparison of bootstrap support in phylogenetic trees robustness as inferred by Bayesian and Maximum Likelihood analyses. *16th Annual Meeting of the CIAR Evolutionary Biology Program*, Harrison Hot Spring, Colombie Britannique (Canada). **P.**
4. Douady C.J., **Delsuc F.**, Boucher Y., Doolittle W.F. & Douzery E.J.P. (2002). Comparison of bootstrap support in phylogenetic trees as inferred by Bayesian and Maximum Likelihood analyses. *Joint Meeting of the International Society of Molecular Evolution and the Society for Molecular Biology and Evolution*, Sorrento (Italie). **P.**
3. **Delsuc F.**, Scally M., Madsen O., Catzeflis F.M., Stanhope M.J., de Jong W.W., Springer M.S. & Douzery E.J.P. (2002). A molecular perspective on the evolutionary history of living xenarthrans. *Joint Meeting of the International Society of Molecular Evolution and the Society for Molecular Biology and Evolution*, Sorrento (Italie). **CO.**
2. **Delsuc F.**, Catzeflis F.M., Stanhope M.J. & Douzery E.J.P. (2000). The evolution of armadillos, anteaters, and sloths depicted by nuclear and mitochondrial phylogenies. *XV Jordanas Argentinas de Mastozoologia*, La Plata (Argentine). **CO.**
1. **Delsuc F.**, Huchon D., Catzeflis F. & Douzery E.J.P. (1998). Polymorphisme enzymatique et mitochondrial chez le tatou à neuf bandes (*Dasypus novemcinctus* : Mammalia, Xenarthra) : une perte de variabilité génétique lors

de la colonisation de l'Amérique du Nord. *Le Petit Pois Dérivé : XX^{ème} Réunion Annuelle du Groupe de Biologie et Génétique des Populations*, Lille (France). **CO**.

4.4.8 SEMINAIRES ET WORKSHOPS

CO : Communication orale ; Orateur.

- 28. Delsuc F.** (2015). Phylogeny and phylogeography of South-American mammals. *Labex CEBA annual meeting* (Cayenne, French Guiana). **CO**.
- 27. Delsuc F.** (2015). MicrobiAnt: Convergence of Gut Microbiomes in Myrmecophagous Mammals. *Colloque de restitution des projets BDL et APEGE du CNRS* (Paris, France). **CO**.
- 26. Delsuc F.** (2014). Next-generation mitogenomics for phylogenetic reconstruction. *2nd joint CNRS-CSIC workshop "Frontiers in Evolutionary Genomics"* (Gif-sur-Yvette, France). **CO**.
- 25. Delsuc F.** (2013). Phylogenetic reconstruction in the next-gen era. *CEBA Thematic School on Advanced Methods in Neotropical Biogeography* (Kourou, French Guiana). **CO**.
- 24. Delsuc F.** (2013). Mammalian Phylogenomics: the edge of the knowable. *Biological Sciences Institute* (Belo Horizonte, Brazil). **CO**.
- 23. Delsuc F.** (2013). Mammalian Phylogenomics: the edge of the knowable. *Center for Biodiversity Research Seminar* (Louvain, Belgium). **CO**.
- 22. Delsuc F.** (2012). Mammalian Phylogenomics: the edge of the knowable. *Seminars of the Center for Genomic Regulation* (Barcelona, Spain). **CO**.
- 21. Delsuc F.** (2012). Phylogénomique des mammifères : Où en est-on ? *Séminaire Invité du Laboratoire Evolution et Diversité Biologique de l'Université Paul Sabatier* (Toulouse, France). **CO**.
- 20. Delsuc F.** (2011). Mammalian phylogenomics and evolution: Where are we and what is next? *Seminars of the Department of Zoology of the University of Cambridge* (Cambridge, UK). **CO**.
- 19. Delsuc F.** (2010). L'histoire évolutive des mammifères racontée par leur ADN. Fête de la Science au Muséum d'Histoire Naturelle de Nîmes (Nîmes, France). **CO**.
- 18. Delsuc F.** (2009). La Phylogénomique : Promesses et défis de la reconstruction phylogénétique à l'échelle des génomes. *Séminaire Invité dans le Cadre du Cours de Master Adaptation et Phylogénie de l'Université Paris VI* (Banyuls-sur-Mer, France). **CO**.
- 17. Delsuc F.** (2008). Les promesses et défis de la phylogénomique illustrés par l'analyse des génomes de levures. *Séminaire Invité dans le Cadre du Cours de Master Analyse des Génomes de l'Institut Pasteur* (Paris, France). **CO**.
- 16. Tsagkogeorga G., Douzery E.J.P. & Delsuc F.** (2008). Updating tunicate phylogeny: Bayesian analyses of 18S rRNA sequences using mixture and

secondary structure models. *Bayesian Phylogeny Workshop of the Renyi Institute* (Budapest, Hungary). **CO**.

15. **Delsuc F.** (2007). Phylogenomics of Urochordates and its application for deciphering rate shifts in vertebrate proteins. *French-Israeli Cooperation Status Seminar on Bioinformatics* (Jerusalem, Israel). **CO**.
14. **Delsuc F.** (2007). La phylogénomique : Promesses et défis de la reconstruction phylogénétique à l'échelle des génomes. *Séminaire Invité du Laboratoire Evolution et Diversité Biologique de l'Université Paul Sabatier* (Toulouse, France). **CO**.
13. **Delsuc F.,** Douzery E.J.P. & Philippe H. (2007). Deciphering the phylogenetic signal at the root of the placental tree. *Dumont D'Urville Workshop on Applied Evolutionary Bioinformatics* (Kaikoura, New Zealand). **CO**.
12. **Delsuc F.** (2007). Les code-barres génétiques et leurs applications pour l'étude de la biodiversité. *Congrès Biotrace 2007 Biocertification des Origines par Voies Analytiques* (Montpellier, France). **CO**.
11. **Delsuc F.** (2006). L'histoire évolutive des mammifères racontée par leur ADN. *Fête de la Science* (Nîmes, France). **CO**.
10. **Delsuc F.** (2006). Quelques exemples des promesses et des dangers de la reconstruction phylogénétique à l'échelle des génomes. *Séminaire Interne du Laboratoire Génomes Populations Interactions* (Montpellier, France). **CO**.
9. **Delsuc F.,** Brinkmann H. & Philippe H. (2005). Phylogénomique des animaux : Qui sont les plus proches parents des vertébrés ? *Retraite Annuelle du Département de Biochimie de l'Université de Montréal* (Sainte-Adèle, Canada). **CO**.
8. **Delsuc F.,** Douzery E.J.P & Philippe H. (2005). Analyses phylogénomiques : Supermatrices versus superarbres. *Rencontres Annuelles Génome Québec* (Mont Orford, Canada). **CO**.
7. **Delsuc F.** (2004). Phylogénomique et inconsistance des méthodes de reconstruction d'arbres. *Réunion du Groupe de Travail en Phylogénie de l'Equipe Projet MultiLabos Bioinformatique pour la Biologie Moléculaire* (Montpellier, France). **CO**.
6. **Delsuc F.** (2003). Comparison of Maximum Likelihood and Bayesian indices of phylogenetic reliability. *Systematic Association of New Zealand (SYSTANZ) Workshop* (Palmerston North, New Zealand). **CO**.
5. Douady C.J., **Delsuc F.,** Boucher Y., Doolittle W.F. & Douzery E.J.P. (2002). Are Bayesian and maximum likelihood results fundamentally different? *Meeting of the European TMR Network Mammalian Phylogeny - Molecular Phylogeny of Mammalian Orders: A Model Study* (Sorrento, Italy). **CO**.
4. **Delsuc F.** & Douzery E.J.P. (2002). Testing the relaxed molecular clock to unveil the role of environmental changes on the evolutionary history of living xenarthrans. *Meeting of the European TMR Network Mammalian Phylogeny - Molecular Phylogeny of Mammalian Orders: A Model Study* (Sorrento, Italy). **CO**.
3. **Delsuc F.,** Scally M., Madsen O., Catzeflis F.M., Stanhope M.J., de Jong W.W., Springer M.S. & Douzery E.J.P. (2001). Phylogenetic relationships of living xenarthrans and the root of the placental tree. *Meeting of the European TMR*

Network Mammalian Phylogeny - Molecular Phylogeny of Mammalian Orders: A Model Study (Reykjavik, Iceland). **CO**.

- 2. Delsuc F.**, Catzeflis F.M. & Douzery E.J.P. (2000). Phylogenetic relationships among and between living xenarthrans. *Meeting of the European TMR Network Mammalian Phylogeny - Molecular Phylogeny of Mammalian Orders: A Model Study* (Montpellier, France). **CO**.
- 1. Delsuc F.** & Douzery E.J.P. (1999). Mitochondrial and nuclear phylogeny, systematics, and biogeography of armadillos, anteaters, and sloths (Mammalia, Xenarthra). *Meeting of the European TMR Network Mammalian Phylogeny - Molecular Phylogeny of Mammalian Orders: A Model Study* (Groningen, Netherlands). **CO**.

4.5. ENCADREMENT D'ÉTUDIANTS

4.5.1 POST-DOCTORANTS

- 2.** Paul Simion (Mars 2015 – Mars 2017). Phylogenomics and molecular evolution of tunicates. *Financement ANR TED*.
- 1.** Gillian C. Gibb (Janvier 2011 – Janvier 2012). Next-generation mitogenomics of modern and fossil xenarthrans. *Financement du Conseil Scientifique de l'Université Montpellier 2*.

4.5.2 DOCTORANTS

- 2.** Fidel Botero-Castro (Octobre 2011 – Décembre 2014). Systématique, phylogénie et évolution moléculaires de la famille Phyllostomidae (Mammalia : Chiroptera) : une approche mitogénomique comparative. *Ecole Doctorale SIBAGHE, Université Montpellier 2* (Allocation Ministérielle). Co-direction avec Emmanuel Douzery.
- 1.** Georgia Tsagakogeorga (Octobre 2006 – Décembre 2009). Phylogénie moléculaire, évolution, et développement des Tuniciers (Urochordés) : une approche phylogénomique. *Ecole Doctorale SIBAGHE, Université Montpellier II* (Allocation Ministérielle). Co-direction avec Emmanuel Douzery.

4.5.3 ETUDIANTS DE MASTER 2

- 5.** Clara Belfiore (1er Janvier – 31 Mai 2015). Délimitation d'espèces chez les tatous à long museau : une approche intégrative combinant morphologie géométrique et données moléculaires. *Master 2 Biologie Evolutive et Ecologie Darwin, Université de Montpellier*. Co-direction avec Lionel Hautier.
- 4.** Thomas Guillerme (15 Janvier – 15 Juin 2012). Horloges moléculaires et taux d'évolution chez les primates : une approche paléontologique et génomique. *Master 2 Biologie Géosciences Agroressources Environnement – Spécialité*

Paléontologie Phylogénie Paléobiologie – Université Montpellier 2 / Montpellier SupAgro. Co-direction avec Emmanuel Douzery.

- 3.** Fabien Trédez (15 Janvier – 15 Juin 2011). Evolution moléculaire du gène de l'énaméline (ENAM) en lien avec la microstructure de l'émail chez les rongeurs. *Master 2 Biologie Géosciences Agroressources Environnement – Spécialité Paléontologie Phylogénie Paléobiologie – Université Montpellier 2 / Montpellier SupAgro. Co-direction avec Laurent Marivaux.*
- 2.** Anaïs Tibi (15 Janvier – 15 Juin 2008). Phylogénie et évolution des gènes impliqués dans la formation de l'émail chez les mammifères. *Master 2 Recherche Biologie Géosciences Agroressources Environnement Spécialité Paléontologie Phylogénie Paléobiologie, Université Montpellier 2.*
- 1.** Georgia Tsagakogeorga (1^{er} Janvier – 30 Juin 2006). Phylogénie moléculaire et évolution des Tuniciers : Contribution de l'ARN ribosomique. *Master 2 Biologie Géosciences Agroressources Environnement Spécialité Biologie de l'Évolution et Écologie, Université Montpellier 2.*

4.5.4 ETUDIANTS DE MASTER 1

- 7.** Loïs Rancilhac (15 Mars 2016 – 15 Juin 2016). Délimitation d'espèces au sein des tatous à long museaux (genre *Dasypus*) par approches de coalescence multi-espèces. *Master 1 Ecologie-Biodiversité parcours Biologie Evolutive et Ecologie, Université de Montpellier.*
- 6.** Lucas Torres (15 Mars 2014 – 15 Juin 2014). Assemblage de génomes mitochondriaux de Tuniciers grâce aux techniques de séquençage de nouvelle génération et leur apport à la phylogénie. *Master 1 Ecologie-Biodiversité parcours Biologie Evolutive et Ecologie, Université Montpellier 2.*
- 5.** Maxime De Sario (15 Mars 2013 – 15 Juin 2013). Le séquençage à haute résolution pour obtenir de nouveaux mitogénomes à utiliser dans les phylogénies : focus sur un ordre au sein des Tuniciers, les Stolidobranches. *Master 1 Ecologie-Biodiversité parcours Biologie Evolutive et Ecologie, Université Montpellier 2.*
- 4.** Berenice Villegas-Ramirez (15 Janvier 2013 – 15 Juin 2013). A first assessment of the barcoding approach to species delimitation for the conservation of xenarthran mammals (Armadillos, Anteaters, and Sloths). *Erasmus Mundus Master Programme in Evolutionary Biology.*
- 3.** Karl-Stéphan Baczkowski (15 Janvier 2012 – 15 Juin 2012). Pseudogenization of taste receptor genes and specific feeding ecologies in mammals. *Master 1 Biosciences de l'Ecole Normale Supérieure de Lyon.*
- 2.** Delphine Sérol (15 Mars 2010 – 15 Juin 2010). Evolution moléculaire des gènes codant pour les protéines structurales des dents chez les Xénarthres. *Master 1 Biologie Géosciences Agroressources Environnement – Spécialité Biologie l'Évolution et Ecologie – Université Montpellier 2 / Montpellier SupAgro.*
- 1.** Benjamin Nigon (26 Mars – 26 Mai 2001). Evolution moléculaire et phylogéographie du tatou à neuf bandes (*Dasypus novemcinctus*). *Maîtrise du Magistère de Biologie Moléculaire & Cellulaire, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1.*

4.6. CONTRATS DE RECHERCHE

4.6.1 PORTEUR DE CONTRATS

8. Objet du contrat : "ConvergeAnt: An integrative approach to understanding convergent evolution in anteating mammals".

Porteur : Frédéric Delsuc

Date de signature : 11/04/2016

Source : ERC Consolidator Grant 2015.

Durée : 5 ans

Montant : 1 880 570€

Participation : 80 %

Partenaires associés au contrat :

UMR 5554, Institut des Sciences de l'Evolution, Université de Montpellier

7. Objet du contrat : "NeotroPhyl: Inferring the drivers of Neotropical diversification using an integrative macroevolutionary approach".

Porteur : Frédéric Delsuc

Date de signature : 01/04/2016

Source : Labex CEBA (Centre d'Etude de la Biodiversité Amazonienne).

Durée : 3 ans

Montant : 250 000 €

Participation : 20 %

Partenaires associés au contrat :

UMR 5554, Institut des Sciences de l'Evolution, Université de Montpellier

UMR 5174, Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier Toulouse III

UMR 1202, Biodiversité, Gènes et Communautés, Université de Bordeaux

USR 3456, Laboratoire Ecologie, Environnement, Interactions des Systèmes Amazoniens, Université de Guyane

6. Objet du contrat : "TATU2: Integrating state-of-the-art genomic and morphological tools to disentangle the biogeography of the long-nosed armadillo species complex (genus Dasypus)".

Porteur : Frédéric Delsuc

Date de signature : 01/04/2015

Source : Labex CEBA (Centre d'Etude de la Biodiversité Amazonienne).

Durée : 1 an

Montant : 19 600 €

Participation : 80 %

Partenaires associés au contrat :

UMR 5554, Institut des Sciences de l'Evolution, Université de Montpellier

Institut Pasteur de la Guyane

5. Objet du contrat : "TATU: Integrating state-of-the-art genomic and morphological tools to disentangle the biogeography of the long-nosed armadillo species complex (genus Dasypus)".

Porteur : Frédéric Delsuc

Date de signature : 01/04/2014

Source : Labex CEBA (Centre d'Etude de la Biodiversité Amazonienne).

Durée : 1 an
Montant : 19 390 €
Participation : 80 %
Partenaires associés au contrat :
UMR5554, Institut des Sciences de l'Evolution, Université de Montpellier
Institut Pasteur de la Guyane

4. Objet du contrat : "Evolution convergente du microbiome intestinal chez les mammifères myrmécophages".

Porteur : Frédéric Delsuc
Date de signature : 01/04/2013
Source : Appel à Projets en Génomique Environnementale (APEGE) CNRS INEE.
Durée : 1 an
Montant : 9 800 €
Participation : 100 %
Partenaires associés au contrat :
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2
Biofrontiers Institute, University of Colorado, Boulder.
McMaster Ancient DNA Centre, McMaster University, ON, Canada.

3. Objet du contrat : "Studying the convergent evolution of gut microbiomes in ant-eating mammals".

Porteur : Frédéric Delsuc
Date de signature : 01/07/2012
Source : Earth Microbiome Project
Durée : 3 mois
Montant : 3 000 US\$
Participation : 80 %
Partenaires associés au contrat :
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2
Biofrontiers Institute, University of Colorado, Boulder.

2. Objet du contrat : "Les nouvelles technologies de séquençage au service de l'évolution moléculaire des gènes codant pour les protéines structurales des dents chez les mammifères édentés".

Porteur : Frédéric Delsuc
Date de signature : 01/01/2010
Source : Conseil Scientifique de l'Université Montpellier 2
Durée : 1 an
Montant : 15 000 €
Participation : 80 %
Partenaires associés au contrat :
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2
McMaster Ancient DNA Centre, McMaster University, ON, Canada.

1. Objet du contrat : "Les nouvelles technologies de séquençage au service de l'évolution moléculaire des gènes codant pour les protéines structurales des dents de mammifères".

Porteur : Frédéric Delsuc
Date de signature : 01/01/2010
Source : PEPS CNRS INSB
Durée : 1 an
Montant : 10 000 €

Participation : 80 %
Partenaires associés au contrat :
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2
McMaster Ancient DNA Centre, McMaster University, ON, Canada.

4.6.2 PARTICIPATION A DES CONTRATS

6. Objet du contrat : ANR TED : "Rôle de la dérive neutre, des contraintes développementales et de la sélection naturelle dans la stabilité évolutive des morphologies embryonnaires des tuniciers".

Porteur : Patrick Lemaire
Date de signature : 01/09/2013
Source : Agence Nationale de la Recherche
Durée : 4 ans

Montant : 495 858 €
Participation : 15 %
Partenaires associés au contrat :
UMR 5237, Centre de Recherche de Biochimie Macromoléculaire
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2

5. Objet du contrat : ANR JAWS : "Vers les origines de la minéralisation chez les vertébrés: l'apport du séquençage de transcriptomes à grande échelle".

Porteur : Jean-Yves Sire
Date de signature : 01/09/2012
Source : Agence Nationale de la Recherche
Durée : 4 ans

Montant : 379 991 €
Participation : 10 %
Partenaires associés au contrat :
UMR 7138, Systématique, Adaptation, Evolution, Université Pierre et Marie Curie Paris 6
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2

4. Objet du contrat : " Modélisation de la biominéralisation de l'émail dentaire des mammifères : évolution et développement ".

Porteur : Fabrice Lihoreau
Date de signature : 01/01/2011
Source : Conseil Scientifique Université Montpellier 2
Durée : 1 ans

Montant : 12 000 €
Participation : 30 %
Partenaires associés au contrat :
UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2

3. Objet du contrat : ERC Advanced Grant: "Population phylogenomics : linking molecular evolution to species biology".

Porteur : Nicolas Galtier
Date de signature : 01/06/2009
Source : European Research Council
Durée : 4 ans
Montant : 2 000 000 €

Participation : 20%

Partenaires associés au contrat :

UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2

2. Objet du contrat : ANR Quenottes : "The EDA pathway and the evolution of teeth in rodents".

Porteur : Vincent Laudet

Date de signature : 06/11/2006

Source : Agence Nationale de la Recherche

Durée : 4 ans

Montant : 600 000 €

Participation : 25%

Partenaires associés au contrat :

UMR 5125, Paléoenvironnements et Paléobiosphère

UMR 5161, Laboratoire de biologie moléculaire de la cellule

UMR 5554, Institut des Sciences de l'Evolution

UMR 6046, Laboratoire de Géologie, Biochronologie et Paléontologie Humaine

1. Objet du contrat : High Council for Scientific and Technological Cooperation between France and Israel. Research networks in bioinformatics: "Phylogenomics of urochordates and its application for detecting rate shifts in vertebrate proteins".

Porteurs : Emmanuel Douzery et Dorothee Huchon

Date de signature : 01/04/2006

Source : Ministère des Affaires étrangères

Durée : 2 ans

Montant : 50 000 €

Participation : 50%

Partenaires associés au contrat :

George S. Wise Faculty of Life Sciences, Tel-Aviv University, Israel

UMR 5554, Institut des Sciences de l'Evolution, Université Montpellier 2

4.7. EVALUATION DE LA RECHERCHE

4.7.1 JURYS D'HABILITATION A DIRIGER DES RECHERCHES

1. Emmanuelle Jousselin (20 Octobre 2014). Le rôle des plantes-hôtes dans la diversification des insectes phytophages : approches phylogénétiques». *Université Montpellier 2* (Examinateur).

4.7.2 JURYS DE THESE

5. Florent Mazel (14 Décembre 2015). Comprendre et protéger la diversité des mammifères : une approche de biogéographie évolutive et fonctionnelle à l'échelle du globe. *Université Grenoble Alpes* (Examinateur). Direction : Wilfried Thuillier & Sébastien Lavergne.

- 4.** Paul Simion (27 Novembre 2014). Les Cténophores : de leur position dans l'arbre des métazoaires à leur diversité taxonomique. *Université Pierre et Marie Curie Paris 6* (Examineur). Direction : Michael Manuel.
- 3.** Kasie Raymann (19 Septembre 2014). Reconstructing the evolutionary relationships between Archaea and Eukaryotes: a phylogenomic approach. *Université Pierre et Marie Curie Paris 6* (Examineur). Direction : Simonetta Gribaldo.
- 2.** Salvador J. Capella-Gutiérrez (16 Novembre 2012). Analysis of multiple protein sequence alignments and phylogenetic trees in the context of phylogenomics studies. *Universidad Pompeu Fabra, Barcelone, Espagne* (Examineur). Direction : Toni Gabaldon.
- 1.** Juliette Arabi (6 Avril 2010). Phylogénie des Chélicérates et étude des taux de substitution dans leurs gènes mitochondriaux et nucléaires. *Museum National d'Histoire Naturelle* (Examineur). Direction : Alexandre Hassanin.

4.7.3 JURYS DE MASTER 2

- 3.** Membre du jury du Master 2 *Erasmus Mundus Master Programme in Evolutionary Biology (MEME)* de l'Université de Montpellier (2015).
- 2.** Membre du jury du Master 2 *Biologie Géosciences Agroressources Environnement - Spécialité Paléontologie Phylogénie Paléobiologie (PPP)* de l'Université Montpellier 2 (2008)
- 1.** Membre du jury du Master 2 *Biologie Géosciences Agroressources Environnement - Spécialité Biologie de l'Évolution et Écologie (BEE)* de l'Université Montpellier 2 (2009)

4.7.4 COMITES DE SELECTION

- 11.** Membre du comité de sélection de l'Université de Poitiers pour le poste MCF-1065 [2014].
- 10.** Membre du comité de sélection du Museum National d'Histoire Naturelle pour le poste MCM-4035 [2012].
- 9.** Membre du comité de sélection du Museum National d'Histoire Naturelle pour le poste MCM-4046 [2012].
- 8.** Membre du comité de sélection de l'Université Blaise Pascal (Clermont-Ferrand II) pour le poste MCF-0365 [2012].
- 7.** Membre du comité de sélection de l'Université Blaise Pascal (Clermont-Ferrand II) pour le poste MCF-0586 [2012].
- 6.** Membre du comité de sélection de l'Université Pierre et Marie Curie (Paris VI) pour le poste MCF-1589 [2010].
- 5.** Membre du comité de sélection de l'Université Blaise Pascal (Clermont-Ferrand II) pour le poste MCF-0351 [2010].

4. Membre du comité de sélection de l'Université Claude Bernard (Lyon I) pour le poste MCF-0851 [2010].
3. Membre du comité de sélection de l'Université Claude Bernard (Lyon I) pour le poste MCF-1299 [2010].
2. Membre du comité de sélection de l'Université Blaise Pascal (Clermont-Ferrand II) pour le poste MCF-0950 [2009].
1. Membre du comité de sélection de l'Université Claude Bernard (Lyon I) pour le poste MCF-0851 [2009].

4.7.5 COMITES D'ÉVALUATION

1. Membre du comité d'évaluation AERES de l'UMR 7205 OSEB [2013]

4.7.6 EXPERTISE

2010- : Expert pour le Fond National pour la Recherche Scientifique (FNRS, Belgique)

2009- : Membre nommé du groupe de spécialistes de l'IUCN pour la conservation des tatous, fourmiliers et paresseux (ASASG) (<http://www.xenarthrans.org/>).

4.7.7 ANIMATION DE LA RECHERCHE

2015- : Membre du comité de direction du Labex CEBA (Centre d'étude de la Biodiversité Amazonienne)

2015- : Membre élu du conseil d'unité de l'ISEM (Institut des Sciences de l'Évolution de Montpellier)

2012- : Membre du comité de pilotage de la plate-forme de biologie moléculaire du labex CeMEB (Centre Méditerranéen Environnement et Biodiversité)

2009-2012 : Membre du comité d'organisation des Séminaires hebdomadaires d'Écologie et Evolution de Montpellier (SEEM) de la SFR Montpellier-Environnement-Biodiversité MEB

4.7.8 ACTIVITES EDITORIALES

- Depuis 2006, je suis membre du comité éditorial de la revue *Systematic Biology*.
http://www.oxfordjournals.org/our_journals/sysbio/editorial_board.html
- J'exerce une activité de relecteur régulier pour les revues et organisations suivantes [nombre de manuscrits/projets] :
- *Systematic Biology* [11], *Molecular Biology and Evolution* [12], *Molecular Phylogenetics and Evolution* [6], *BMC Evolutionary Biology* [6], *Nature* [2],

Current Biology [2], Science Advances [1], Biology Letters [1], Nuclear Acid Research [1], PLoS Genetics [1], PLoS One [1], BMC Genetics [1], Journal of Molecular Evolution [1], Proceedings of the Royal Society B [1], Chromosome Research [1], Genome Biology [1], Genomics [1], Comparative and Functional Genomics [1], European Journal of Taxonomy [4], Journal of Biological Systems [1], Edentata [1], Senckenbergiana Biologica [1], Mammalia [1], Zootaxa [1].

- Fond National pour la Recherche Scientifique (FNRS, Belgique) [15], National Science Foundation (NSF, USA) [2], Agence Nationale de la Recherche (ANR, France) [1], German Research Foundation (DFG, Allemagne) [1], Swiss National Science Foundation (SNSF, Suisse) [1], Natural Sciences and Engineering Research Council of Canada (NSERC, Canada) [1], Fondation pour la Recherche en Biodiversité (FRB, France) [1].

4.8. ENSEIGNEMENT, FORMATION ET DIFFUSION

4.8.1 ENSEIGNEMENTS

2012-2013 [Total : 18,5 heures équivalent TD]

- Cours de Phylogénomique, Master 2 BEE. 3H CM.
- TD de Reconstruction phylogénétique ML et Bayésienne, Master 2 BEE. 5H TD.
- Cours de Reconstruction Phylogénétique Probabiliste, Master 2 SEP MNHN. 6H CM.

2011-2012 [Total : 40,5 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- TD de Reconstruction phylogénétique ML et Bayésienne, Master 2 IC. 10H TD.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- Cours de Reconstruction Phylogénétique Probabiliste, Master 2 SEP MNHN. 6H CM.

2010-2011 [Total : 40, 5 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- TD de Reconstruction phylogénétique ML et Bayésienne, Master 2 IC. 5H TD.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- Cours de Reconstruction Phylogénétique Probabiliste, Master 2 SEP MNHN. 6H CM.

2009-2010 [Total : 28,75 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- TD de Reconstruction phylogénétique ML et Bayésienne, Master 2 IC. 5H TD.
- Cours de Reconstruction Phylogénétique Probabiliste, Master 2 SEP MNHN. 6H CM.

- TD de Phylogénie et horloge moléculaire, Master 1 BGAE. 3H TD.

2008-2009 [Total : 38,5 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 IC. 5H TD.
- Cours Phylogénie et développement, Master 2 BioSanté. 2H CM.
- Cours Horloge moléculaire, Ecole doctorale du MNHN. 1H30 CM.
- Cours de Reconstruction Phylogénétique, Master 2 SEP MNHN. 6H CM.
- TD de Phylogénie et horloge moléculaire, Master 1 BGAE. 3H TD.

2007-2008 [Total : 26,5 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 IC. 5H TD.
- Cours Phylogénie et développement, Master 2 BioSanté. 2H CM.
- Cours Horloge moléculaire, Ecole doctorale du MNHN. 1H30 CM.

2006-2007 [Total : 34 heures équivalent TD]

- Conférence La Phylogénomique en pratique, Master 2 BGAE. 1H30 CM.
- Cours de Phylogénomique, Master 2 BGAE. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 BGAE. 5H TD.
- Cours Phylogénie et développement, Master 2 BioSanté. 2H CM.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- TD de Reconstruction phylogénétique Bayésienne, Master 2 IC. 5H TD.
- TD de Phylogénie et horloge moléculaire, Master 1 BGAE. 3H TD.
- Cours de Phylogénomique, Master 2 IC. 3H CM.
- Cours Horloge moléculaire, Ecole doctorale du MNHN. 1H30 CM.

4.8.2 FORMATIONS

- 7.** Formateur de la formation « One week training in phylogenetic reconstruction and phylogenomics » organisée par l'Université de Louvain-la-Neuve (Caroline Nieberding) [1 – 5 Juin 2015]. Total : 30H cours + TP (25 participants).
- 6.** Intervenant dans la formation CNRS organisée par le Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM) (Olivier Gascuel et Vincent Lefort) « Phylogénie Moléculaire ». [13 – 17 Oct 2014]. Total : 6,5H cours + TP (15 participants).
- 5.** Intervenant à la formation CNRS organisée par le Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM) (Olivier Gascuel et Vincent Lefort) « Phylogénie Moléculaire ». [12 – 15 Nov 2013]. Total : 6,5H cours + TP (17 participants).
- 4.** Intervenant à la formation CNRS organisée par le Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM) (Olivier Gascuel et Vincent Lefort) « Phylogénie Moléculaire ». [29 Mai – 1^{er} Juin 2012]. Total : 9,5H cours + TP (15 participants).

3. Intervenant à la formation CNRS organisée par le Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM) (Olivier Gascuel et Vincent Lefort) « Phylogénie Moléculaire ». [14-17 Juin 2011]. Total : 8H cours + TP (14 participants).
2. Co-organisateur avec Khalid Belkhir et formateur d'une Formation CNRS ouverte à la communauté Montpelliéraine. « Introduction à l'application des méthodes probabilistes de reconstruction phylogénétique (maximum de vraisemblance et approche Bayésienne) ». [15-18 Décembre 2009]. Total : 12H (25 participants).
1. Co-organisateur avec Khalid Belkhir et formateur d'une Formation CNRS ouverte à la communauté Montpelliéraine. « Introduction à l'application des méthodes probabilistes de reconstruction phylogénétique (maximum de vraisemblance et approche Bayésienne) ». [8-11 Décembre 2008]. Total : 12H (20 participants).

4.8.3 DIFFUSION SCIENTIFIQUE

Expositions

- Conseil scientifique pour le réaménagement de la galerie de zoologie du Muséum d'Histoire Naturelle de Nîmes <http://www.nimes.fr/index.php?id=465>
- Conseil scientifique pour l'exposition « Evolution : Des dinosaures aux OGM » présentée à la MJC de Castelnau-le-Lez (Hérault) dans le cadre de l'année Darwin , du Bicentenaire de l'UM2 et de la Fête de la science 2009.
- Conseil scientifique pour l'exposition permanente « L'arbre du vivant » du Musée Lecoq d'Histoire Naturelle de Clermont-Ferrand. <http://www.clermont-ferrand.fr/-L-arbre-du-vivant-.html>.

Radio

- Interview en français « Zähne aus Komposit » pour la radio allemande Deutschlandfunk (Joachim Budde). <http://www.dradio.de/dlf/sendungen/forschak/649343/>

Culture scientifique

- Participation à la Mission Peluches du service de Culture Scientifique de l'Université Montpellier 2 (<http://www.peluche.um2.fr/medias/?lang=fr>)

Web

- Conseil scientifique pour l'animation « Classer les espèces : le vivant réorganisé » de la Banque des savoirs (Conseil Général 91). <http://www.savoirs.essonne.fr/sections/ressources/images-animations/resource/classer-les-especes-le-vivant-reorganise/>.
- PRES TV Université Montpellier Sud de France. Peluchologie : Classification des espèces http://www.dailymotion.com/video/xgfc2o_pres-tv-peluchologie-la-classification-des-especes_school

Jeu éducatif

- Conseil scientifique pour le Jeu éducatif « Trivial Evolution » réalisé par l'École de l'ADN de Nîmes ayant pour objectif de sensibiliser le public à la théorie de l'évolution et la classification (systématique). http://www.ecole-adn.fr/?page_id=354.

Conférences publiques

- « L'histoire évolutive des mammifères racontée par leur ADN » dans le cadre de la Fête de la Science au Lycée Charles Gide d'Uzès (Octobre 2011).

- « L'histoire évolutive des mammifères » dans le cadre de la Fête de la Science au Musée d'Histoire Naturelle de Nîmes (Octobre 2006).

5. BIBLIOGRAPHIE

- Abascal F, Zardoya R, Telford MJ (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucl. Acids Res.* 38:W7-W13.
- Abubucker et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* 8:e1002358.
- Alcaide M et al. (2012). Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx. *PLoS One* 7:e51521.
- Assis R, Zhou Q, Bachtrog D (2012). Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol. Evol.* 4:1189-1200.
- Arendt J, Reznick D (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23:26-32.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013). GenBank. *Nucl. Acids Res.* 41:D36-D42.
- Billet G, Hautier L, Asher RJ, Schwarz C, Crumpton N, Martin T, Ruf I (2012). High morphological variation of vestibular system accompanies slow and infrequent locomotion in three-toed sloths. *Proc. Roy. Soc. B.* 279:3932-39.
- Bininda-Emonds OR, Gittleman JL, Steel MA (2002). The (super) tree of life: procedures, problems, and prospects. *Annu. Rev. Ecol. Syst.* 33:265-289.
- Bininda-Emonds OR (2005). transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* 6:156.
- Borrero LA, Martin FM (2012). Taphonomic observations on ground sloth bone and dung from Cueva del Milodón, Ultima Esperanza, Chile: 100 years of research history. *Quat. Int.* 278:3-11.
- Botero-Castro F, Tilak M-K, Justy F, Catzeflis F, Delsuc F, Douzery EJP (2013). Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol. Phylogenet. Evol.* 69:728-739.
- Botero-Castro F, Delsuc F, Douzery EJP (2016). Thrice better than once: quality control guidelines to validate new mitogenomes. *Mitoch. DNA* 27:449-454.
- Bradnam KR et al. (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2:1-31.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev VB, Golovanova LV, de la Rasilla M, Fortea J, Rosas A, Pääbo S (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318-321.

- Bussink AP, Speijer D, Aerts JM, Boot RG (2007). Evolution of mammalian chitinase (-like) members of family 18 glycosyl hydrolases. **Genetics** 177:959-970.
- Caporaso JG et al. (2010). QIIME allows analysis of high-throughput community sequencing data. **Nature Meth.** 7:335-6.
- Castoe TA, de Koning AJ, Kim HM, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. **PNAS** 106:8986-91.
- Chiari Y, Cahais V, Galtier N, Delsuc F (2012). Phylogenomic analyses support the position of turtles as the sister-group of birds and crocodiles (Archosauria). **BMC Biol.** 10:65.
- Chourrout D, Delsuc F, Chourrout P, Edvarsen RB, Renfer E, Jensen MF, Zhu B, de Jong P, Steele RE, Technau U (2006). Minimal ProtoHox cluster inferred from comparison of Cnidarian and Bilaterian Hox complements. **Nature** 442: 684-687.
- Christin PA, Weinreich DM, Besnard G (2010). Causes and evolutionary significance of genetic convergence. **Trends Genet.** 26:400-405.
- Churakov G, Kriegs JO, Baertsch R, Zemmann A, Brosius J, Schmitz J (2009). Mosaic retroposon insertion patterns in placental mammals. **Genome Res.** 19:868-875.
- Ciancio MR, Castro MC, Galliari FC, Carlini AA, Asher RJ (2011). Evolutionary implications of dental eruption in *Dasybus* (Xenarthra). **J. Mammal. Evol.** 19:1-8.
- Contet A, Pihan E, Lavigne M, Wengelnik K, Maheshwari S, Vial H, Douguet D, Cerdan R (2015). *Plasmodium falciparum* CTP: phosphocholine cytidyltransferase possesses two functional catalytic domains and is inhibited by a CDP-choline analog selected from a virtual screening. **FEBS Lett.** 589:992-1000.
- Cozzarini E, Bellin M, Norberto L, Polese L, Musumeci S, Lanfranchi G, Paoletti MG (2009). CHIT1 and AMCase expression in human gastric mucosa: correlation with inflammation and *Helicobacter pylori* infection. **Eur. J. Gastroenterol. Hepatol.** 21:1119-1126.
- Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. **Biol. Lett.** 8:783-786.
- Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. **PeerJ** 2:e243.
- Davies KT, Cotton JA, Kirwan JD, Teeling EC, Rossiter SJ (2012). Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. **Heredity** 108:480-9.
- Delsuc F, Scally M, Madsen O, Stanhope MJ, De Jong WW, Catzeflis FM, Springer MS, Douzery EJP (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. **Mol. Biol. Evol.** 19:1656-1671.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. **Nature** 439:965-968.
- Delsuc F, Douzery EJP (2008). Recent advances and future prospects in xenarthran molecular phylogenetics. In **Biology of the Xenarthra**, WJ Loughry, SF Vizcaíno, eds. University Press of Florida, Gainesville. Pp. 11-23.
- Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H (2008). Additional molecular support for the new chordate phylogeny. **Genesis** 46: 592-604.
- Delsuc F, Superina M, Tilak M-K, Douzery EJP, Hassanin A (2012). Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. **Mol. Phylogenet. Evol.** 62:673-680.
- Delsuc F, Metcalf JL, Wegener Parfrey L, Song SJ, Gonzalez A, Knight R (2014). Convergent evolution of gut microbiomes in myrmecophagous mammals. **Mol. Ecol.** 23:1301-1317.
- Delsuc F, Gasse B, Sire J-Y (2015). Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. **BMC Evol. Biol.** 15:148.
- Delsuc F, Philippe H, Tsagkogeorga G, Simion P, Turon X, Lopez-Legentil S, Piette J, Lemaire P, Douzery EJP (Soumis). Phylogeomics provides a reference phylogenetic framework and timescale for tunicate comparative genomics.

- Delsuc F, Emerling CE, Gibb GC, Tilak M-K, Sire JY (en prep). How edentates lost their teeth? Evidence for convergent loss of function of enamel and dentine genes in aardvarks, pangolins and xenarthrans.
- Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, Buckley CM, Davis IJ, Bennett ML, Marshall-Jones ZV (2012). The canine oral microbiome. **PLoS One** 7:e36067.
- Denoeud F et al. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. **Science** 330: 1381-1385.
- Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V (2014). OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. **Mol. Biol. Evol.** 31:1923-1928.
- DuBuc TQ, Ryan JF, Shinzato C, Satoh N, Martindale MQ (2012). Coral comparative genomics reveal expanded Hox cluster in the cnidarian–bilaterian ancestor. **Integr. Comp. Biol.** 52:835-841.
- Duggal NK, Emerman M (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. **Nat. Rev. Immunol.** 12:687-695.
- Eizirik E et al. (2010). Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. **Mol. Phylogenet. Evol.** 56:49-63.
- Elmer KR, Meyer A (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. **Trends Ecol. Evol.** 26:298-306.
- Engelmann GF (1985). The phylogeny of the Xenarthra. In **The evolution and ecology of armadillos, sloths, and vermilinguas**. GG Montgomery ed. Smithsonian Institution Press, Washington DC. Pp. 51-64.
- Finley D, Chen X, Walters KJ (2016). Gates, channels, and switches: elements of the proteasome machine. **Trends Biochem. Sci.** 41:77-93.
- Flicek P et al. (2014). Ensembl 2014. **Nucl. Acids Res.** 42:D749-D755.
- Foot AD et al. (2015). Convergent evolution of the genomes of marine mammals. **Nat. Genet.** 47:272-5.
- Fort P, Kajava AV, Delsuc F, Coux O (2015). Evolution of proteasome regulators in Eukaryotes. **Genome Biol. Evol.** 7:1363-1379.
- Galtier N, Duret L, Glémin S, Ranwez V (2009). GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. **Trends Genet.** 25:1-5.
- Garcia-Fernandez J (2005). The genesis and evolution of homeobox gene clusters. **Nature Rev. Genet.** 6: 881-892.
- Gaubert P, Wozencraft WC, Cordeiro-Estrela P, Veron G (2005). Mosaics of convergences and noise in morphological phylogenies: what's in a viverrid-like carnivoran? **Syst. Biol.** 54: 865-94.
- Gaudin TJ, Wible JR (2006). The phylogeny of living and extinct armadillos (Mammalia, Xenarthra, Cingulata): a craniodental analysis. In **Amniote paleobiology: perspectives on the evolution of mammals, birds and reptiles**. MT Carrano, TJ Gaudin, RW Blob, JR Wible eds. University of Chicago Press, Chicago. Pp. 153-98.
- Gee H (2006). Evolution: careful with that amphioxus. **Nature** 439:923-924.
- Gibb GC, Condamine F, Kuch M, Enk J, de Moraes-Barros N, Superina M, Poinar HN, Delsuc F (2016). Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. **Mol. Biol. Evol.** 3:621-642.
- Gnerre S et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. **PNAS** 108:1513-18.
- Gompel N, Prud'homme B (2009). The causes of repeated genetic evolution. **Dev. Biol.** 332:36-47.
- Gooday GW (1990). The ecology of chitin degradation. In **Advances in Microbial Ecology**, Vol 11, Marshall KC eds, Plenum Press, pp 387-430.
- Goswami A (2006a). Cranial modularity shifts during mammalian evolution. **Am. Nat.** 168:270-80.
- Goswami A (2006b). Morphological integration in the carnivoran skull. **Evolution** 60:169-83.

- Gouy M, Guindon S, Gascuel O (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Mol. Biol. Evol.** 27:221-224.
- Haas BJ et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols** 8:1494-512.
- Hallström BM, Kullberg M, Nilsson MA, Janke A (2007). Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. **Mol. Biol. Evol.** 24:2059-2068.
- Hallström BM, Janke A (2010). Mammalian evolution may not be strictly bifurcating. **Mol. Biol. Evol.** 27:2804-2816.
- Harris S, Pisani D, Gower DJ, Wilkinson M (2007). Investigating stagnation in morphological phylogenetics using consensus data. **Syst. Biol.** 56:125-129.
- Harris RS, Hultquist, JF, Evans DT (2012). The restriction factors of human immunodeficiency virus. **J. Biol. Chem.** 287:40875-40883.
- Hautier L et al. (2011). Skeletal ossification and sequence heterochrony in xenarthran evolution. **Evol. Dev.** 13:460-76.
- Hautier L, Lebrun R, Cox PG (2012). Patterns of covariation in the masticatory apparatus of hystricognathous rodents: implications for evolution and diversification: **J. Morph.** 273:1319-37.
- Hautier L et al. (2013). Patterns of ossification in southern vs. northern placental mammals. **Evolution** 67:1994-10.
- Hautier L, Billet G, Eastwood B, Lane J (2014). Patterns of morphological variation of extant sloth skulls and their implication for future conservation efforts. **Anat. Rec.** 297:979-08.
- Hautier L, Gomes Rodrigues H, Billet G, Asher RJ (2016). The hidden teeth of sloths: evolutionary vestiges and the development of a simplified dentition. **Sci. Rep.** in press.
- Hayakawa T, Suzuki-Hashido N, Matsui A, Go Y (2014). Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. **Mol. Biol. Evol.** 31:2018-2031.
- Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC (2010). Ecological adaptation determines functional mammalian olfactory subgenomes. **Genome Res.** 20:1-9.
- Hayden S, Bekaert M, Goodbla A, Murphy WJ, Dávalos LM, Teeling EC (2014). A cluster of olfactory receptor genes linked to frugivory in bats. **Mol. Biol. Evol.** 31:917-927.
- Hehemann JH et al. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. **Nature** 464:908-12.
- Hoffstetter R (1958). Xenarthra. In **Traité de paléontologie** Vol. 2, no. 6. P Piveteau ed. Masson et Cie, Paris. Pp. 535-636.
- Holland PWH (2013). Evolution of homeobox genes. **WIREs Dev. Biol.** 2013, 2:31-45.
- Jiang P et al. (2012). Major taste loss in carnivorous mammals. **PNAS** 109: 4956-61.
- Jiménez FA, Carreno RA, Gardner SL (2013). *Aspidodera kinsellai* n. sp. (Nematoda: Heterakoidea) from nine-banded armadillos in middle America with notes on phylogeny and host-parasite biogeography. **J. Parasitol.** 99:1056-1061.
- Jobson RW, Nabholz B, Galtier N (2010). An evolutionary genome scan for longevity-related natural selection in mammals. **Mol. Biol. Evol.** 27:840-47.
- Karasov WH, Martínez del Rio C, Caviedes-Vidal E (2011). Ecological physiology of diet and digestive systems. **Annu. Rev. Physiol.** 73: 69-93.
- Kopylova E, Noé L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. **Bioinformatics** 28:3211-3217.
- Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, Yatim A, Emiliani S, Schwartz O, Benkirane M (2011). SAMHD1 is the dendritic-and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. **Nature** 474:654-657.
- Laguette N, Rahm N, Sobhian B, Chable-Bessia C, Munch J, Snoeck J, Sauter D, Switzer WM, Heneine W, Kirchhoff F, Delsuc F, Telenti A, Benkirane M (2012). Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. **Cell Host Microbe** 11:211-217.

- Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. **BMC Microbiol.** 11:103.
- Lartillot N, Philippe H (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. **Mol. Biol. Evol.** 21:1095-109.
- Lartillot N, Brinkmann H, Philippe H (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. **BMC Evol. Biol.** 7:S4.
- Lartillot N, Lepage T, Blanquart S (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. **Bioinformatics** 25:2286-2288.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. **Front. Zool.** 10:34.
- Ley RE et al. (2008). Evolution of mammals and their gut microbes. **Science** 320:1647-51.
- Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJ (2013). Capturing protein-coding genes across highly divergent species. **Biotechniques** 54:321-326.
- Li D, Zhang J (2014). Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. **Mol. Biol. Evol.** 31:303-309.
- Li R et al. (2010b). The sequence and de novo assembly of the giant panda genome. **Nature** 463:311-7.
- Li Y, Liu Z, Shi P, Zhang J (2010a). The hearing gene Prestin unites echolocating bats and whales. **Curr. Biol.** 20:R55-6.
- Lim ES, Fregoso OI, McCoy CO, Matsen FA, Malik HS, Emerman M (2012). The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. **Cell Host Microbe** 11:194-204.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009). Coalescent methods for estimating phylogenetic trees. **Mol. Phylogenet. Evol.** 53:320-328.
- Liu Y, Cotton JA, Shen B, Han X, Rossiter SJ, Zhang S (2010a). Convergent sequence evolution between echolocating bats and dolphins. **Curr. Biol.** 20:R53-R54.
- Liu Y, Rossiter SJ, Han X, Cotton JA, Zhang S (2010b). Cetaceans on a molecular fast track to ultrasonic hearing. **Curr. Biol.** 20:1834-9.
- Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014). The carbohydrate-active enzymes database (CAZy) in 2013. **Nucl. Acids Res.** 42:D490-D495.
- Losos JB (2011). Convergence, adaptation, and constraint. **Evolution** 65:1827-40.
- Madsen O et al. (2001). Parallel adaptive radiations in two major clades of placental mammals. **Nature** 409:610-4.
- Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010). Convergence in pigmentation at multiple levels: mutations, genes and function. **Philos. Trans. R. Soc. Lond. B** 365:2439-50.
- Martin BE (1916). Tooth development in *Dasyus novemcinctus*. **J. Morphol.** 27:647-91.
- McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. **Genome Res.** 22:746-754.
- McDonald D et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. **ISME J.** 6:610-8.
- McFall-Ngai M et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. **PNAS** 110:3229-36.
- McGhee G (2011). Convergent evolution: limited forms most beautiful. *Vienna Series in Theoretical Biology*, Massachusetts Institute of Technology Press, Cambridge, USA. 322 pp.
- Meredith RW, Gatesy J, Murphy WJ, Ryder OA, Springer MS (2009). Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. **PLoS Genet.** 5:e1000634.
- Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C,

- Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. **Science** 334:521-4.
- Meredith RW, Gatesy J, Springer MS (2013). Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. **BMC Evol. Biol.** 13, 20.
- Möller-Krull M*, Delsuc F*, Churakov G, Marker C, Superina M, Brosius J, Douzery EJP, Schmitz J (2007). Retroposed elements and their flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters and sloths). **Mol. Biol. Evol.** 24:2573-2582.
- Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. **Science** 332:970-74.
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001). Molecular phylogenetics and the origins of placental mammals. **Nature** 409:614-618.
- Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. **Science** 294:2348-2351.
- Nagy GN, Marton L, Krámos B, Oláh J, Revész A, Vékey K, Delsuc F, Hunyadi-Gulyás E, Medzihradsky KF, Lavigne M, Vial H, Cerdan R, Vértessy BG (2013). Evolutionary and mechanistic insights into substrate and product accommodation of CTP:phosphocholine cytidyltransferase in *Plasmodium falciparum*. **FEBS J.** 280:3132-3148.
- Naples V (1985). The superficial facial musculature in sloths and vermilinguas (anteaters). In **The Evolution and Ecology of Armadillos, Sloths and Vermilinguas**, Montgomery GG ed, pp. 173-90.
- Nishihara H, Maruyama S, Okada N (2009). Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. **Proc. Natl. Acad. Sci. USA** 106:5235-5240.
- O'Leary MA et al. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. **Science** 339:662-7.
- Ordoñez GR, Hillier LW, Warren WC, Grützner F, López-Otín C, Puente XS (2008). Loss of genes implicated in gastric function during platypus evolution. **Genome Biol.** 9:R81.
- Orlando L, Gilbert MTP, Willerslev E (2015). Reconstructing ancient genomes and epigenomes. **Nat. Rev. Genet.** 16:395-408.
- Paijmans JL, Gilbert MTP, Hofreiter M (2013). Mitogenomic analyses from ancient DNA. **Mol. Phylogenet. Evol.** 69:404-416.
- Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ (2013). Genome-wide signatures of convergent evolution in echolocating mammals. **Nature** 502:228-31.
- Patterson B, Pascual R (1972) The fossil mammal fauna of South-America. In: **Evolution, mammals and southern continents** (eds. Keast A, Erk FC, Glass BP), pp. 247-309. State University of New-York Press, Albany.
- Pointer MA, Kamilar JM, Warmuth V, Chester SG, Delsuc F, Mundy NI, Asher RJ, Bradley BJ (2012). RUNX2 tandem repeats and the evolution of facial length in placental mammals. **BMC Evol. Biol.** 12:103.
- Putnam NH et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. **Nature** 453:1064-1071.
- Ranwez V, Delsuc F, Ranwez S, Belkhir K, Tilak M-K, Douzery EJP (2007). OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. **BMC Evol. Biol.** 7:241.
- Ranwez V, Clairon N, Delsuc F, Pourali S, Auberval N, Diser S, Berry V (2009). PhyloExplorer: a web server to validate, explore and query phylogenetic trees. **BMC Evol. Biol.** 9:108.
- Ranwez V, Harispe S, Delsuc F, Douzery EJP. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. **PLoS One** 6:e22594.
- Redford KH, Dorea JG (1984). The nutritional value of invertebrates with emphasis on ants and termites as food for mammals. **J. Zool.** 203:385-395.

- Redford KH (1987). Ants and termites as food: Patterns of mammalian myrmecophagy. In **Current Mammalogy**, Springer USA, pp. 349-399.
- Reiss KZ (2001). Using phylogenies to study convergence: the case of the ant-eating mammals. **Am. Zool.** 41: 507-25.
- Rodrigue N, Philippe H, Lartillot N (2010). Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. **PNAS** 107:4629-34.
- Rodrigue N, Lartillot N (2014). Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. **Bioinformatics** 30:1020-1021.
- Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP (2013). Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. **Mol. Biol. Evol.** 30:2134-2144.
- Rubinstein ND, Feldstein T, Shenkar N, Botero-Castro F, Griggio F, Mastrototaro F, Delsuc F, Douzery EJP, Gissi C, Huchon D (2013). Deep sequencing of mixed total DNA without barcodes allows efficient assembly of highly plastic ascidian mitochondrial genomes. **Genome Biol. Evol.** 5:1185-1199.
- Sanders JG, Beichman AC, Roman J, Scott JJ, Emerson D, McCarthy JJ, Girguis PR (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. **Nat. Commun.** 6:8285.
- Sanderson MJ, Donoghue MJ, Piel WH, Eriksson T (1994). TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. **Am. J. Bot.** 81:183.
- Sears KE, Goswami A, Flynn JJ, Niswander LA (2007). The correlated evolution of Runx2 tandem repeats, transcriptional activity, and facial length in Carnivora. **Evol. Dev.** 9:555-565.
- Shi P, Zhang J (2006). Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. **Mol. Biol. Evol.** 23:292-300.
- Singh TR*, Tsagkogeorga G*, Delsuc F, Shenkar N, Loya Y, Douzery EJP, Huchon D (2009). Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. **BMC Genomics** 10:534.
- Song S, Liu L, Edwards SV, Wu S (2012). Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. **Proc. Natl. Acad. Sci. USA** 109:14942-14947.
- Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, O'Brien SJ, Murphy WJ (2007). The adequacy of morphology for reconstructing the early history of placental mammals. **Syst. Biol.** 56:673-84.
- Springer MS, Meredith RW, Teeling EC, Murphy WJ (2013). Technical comment on "the placental mammal ancestor and the post-k-pg radiation of placentals". **Science** 341:613-14.
- Stern DL, Orgogozo V (2008). The loci of evolution: how predictable is genetic evolution? **Evolution** 62:2155-77.
- Stewart CB, Schilling JW, Wilson AC (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. **Nature** 330:401-4.
- Telford MJ, Moroz LL, Halanych KM (2016). Evolution: A sisterly dispute. **Nature** 529:286-287.
- Thomas GW, Hahn MW (2015). Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. **Mol. Biol. Evol.** 32:1232-1236.
- Thomas-Chollier M, Ledent V, Leyns L, Vervoort M (2010). A non-tree-based comprehensive study of metazoan Hox and ParaHox genes prompts new insights into their origin and evolution. **BMC Evol. Biol.** 10:73.
- Tilak M-K*, Justy F*, Debais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP (2015). A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. **Conserv. Genet. Res.** 7:37-40.
- Treangen et al. (2013). MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. **Genome Biol.** 14:R2.

- Tsagkogeorga G, Turon X, Hopcroft RR, Tilak M-K, Feldstein T, Shenkar N, Loya Y, Huchon D, Douzery EJP, Delsuc F (2009). An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. **BMC Evol. Biol.** 9:187.
- Tsagkogeorga G, Turon X, Galtier N, Douzery EJP, Delsuc F (2010). Accelerated evolutionary rate of housekeeping genes in tunicates. **J. Mol. Evol.** 71:153-167.
- Tsagkogeorga G, Cahais V, Galtier N (2012). The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. **Genome Biol. Evol.** 4:852-861.
- Voges D, Zwickl P, Baumeister W (1999). The 26S proteasome: a molecular machine designed for controlled proteolysis. **Annu. Rev. Biochem.** 68:1015-1068.
- Wang Z et al. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. **Nat. Genet.** 45:701-706.
- Whittington CM et al. (2010). Novel venom gene discovery in the platypus. **Genome Biol.** 11:R95.
- Wildman DE, Uddin M, Opazo JC, Liu G, Lefort V, Guindon S, Gascuel O, Grossman LI, Romero R, Goodman M (2007). Genomics, biogeography, and the diversification of placental mammals. **Proc. Natl. Acad. Sci. USA** 104:14395-14400.
- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. **Methods Ecol. Evol.** 3:613-623.
- Zardoya R, Meyer A (2001). The evolutionary position of turtles revised. **Naturwissenschaften** 88:193-200.
- Zhang J, Kumar S (1997). Detection of convergent and parallel evolution at the amino acid sequence level. **Mol. Biol. Evol.** 14:527-36.
- Zhang J (2006). Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. **Nat. Genet.** 38:819-23.
- Zhu L, Wu Q, Dai J, Zhang S, Wei F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome. **PNAS** 108:17714-9.
- Zou Z, Zhang J (2015). No genome-wide protein sequence convergence for echolocation. **Mol. Biol. Evol.** 32:1237-1241.

6. ANNEXES

6.1. ANNEXE 1 : PUBLICATIONS CHOISIES

Additional Molecular Support for the New Chordate Phylogeny

Frédéric Delsuc,^{1,2*} Georgia Tsagkogeorga,^{1,2} Nicolas Lartillot,³ and Hervé Philippe³

¹Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

²CNRS, Institut des Sciences de l'Évolution (UMR5554), CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

³Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, Québec, Canada

Received 6 May 2008; Revised 16 September 2008; Accepted 22 September 2008

Summary: Recent phylogenomic analyses have suggested tunicates instead of cephalochordates as the closest living relatives of vertebrates. In direct contradiction with the long accepted view of Euchordates, this new phylogenetic hypothesis for chordate evolution has been the object of some skepticism. We assembled an expanded phylogenomic dataset focused on deuterostomes. Maximum-likelihood using standard models and Bayesian phylogenetic analyses using the CAT site-heterogeneous mixture model of amino-acid replacement both provided unequivocal support for the sister-group relationship between tunicates and vertebrates (Olfactores). Chordates were recovered as monophyletic with cephalochordates as the most basal lineage. These results were robust to both gene sampling and missing data. New analyses of ribosomal rRNA also recovered Olfactores when compositional bias was alleviated. Despite the inclusion of 25 taxa representing all major lineages, the monophyly of deuterostomes remained poorly supported. The implications of these phylogenetic results for interpreting chordate evolution are discussed in light of recent advances from evolutionary developmental biology and genomics. *genesis* 46:592–604, 2008. © 2008 Wiley-Liss, Inc.

Key words: phylogenomics; deuterostomes; chordates; tunicates; cephalochordates; olfactores; ribosomal RNA; jackknife; evolution

INTRODUCTION

Besides its fundamental role in systematics, phylogenetic reconstruction is a prerequisite for understanding the evolution of organisms. The essential contribution of phylogenetics for understanding morphological diversity has perhaps been best exemplified in the case of animal evolution (Telford and Budd, 2003). The Cambrian explosion has produced a bewildering diversity of body plans whose origins and evolution can only be apprehended by undertaking an integrative approach through evolutionary developmental biology (Evo-Devo) (Conway-Morris, 2003). The knowledge of phylogenetic relationships, by allowing the polarization of character

transformations, sheds light on the extent of morphological convergence and reversal. A phylogenetic framework is therefore required for distinguishing ancestral characters from those representing morphological innovations. Comparative genomics is now providing the opportunity to track these morphological innovations back to the molecular level by revealing the patterns of gene acquisition/loss and giving clues to the molecular adaptations that underline the evolution of body plans (Cañestro *et al.*, 2007).

Animal taxonomy has deep roots. The study of morphological and embryological characters has allowed the definition of the major phyla but left their interrelationships almost unresolved (Nielsen, 2001). The advent of molecular data during the 1990s has revolutionized the traditional classification through a series of phylogenetic analyses of the 18S ribosomal RNA (rRNA) gene for an ever increasing number of key taxa (Aguinaldo *et al.*, 1997; Halanych *et al.*, 1995). This period culminated with the proposition of a new view of animal phylogeny at odds with the traditional paradigm of a steady increase toward morphological complexity, and revealing instead the major role played by secondary simplification from complex ancestors (Adoutte *et al.*, 2000; Lwoff, 1944). Despite these undeniable achievements, the resolving power provided by 18S rRNA and other single genes is nevertheless limited, and a number of open questions in animal phylogeny remained to be answered (Halanych, 2004).

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Frédéric Delsuc, CC064, Institut des Sciences de l'Évolution, UMR5554-CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. E-mail: frederic.delsuc@univ-montp2.fr

Contract grant sponsor: Contribution ISEM 2008-062 of the Institut des Sciences de l'Évolution (UMR5554-CNRS); Contract grant sponsor: Research Networks Program in Bioinformatics from the High Council for Scientific and Technological Cooperation between France and Israel

Published online 10 November 2008 in

Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/dvg.20450

The most recent advances in animal phylogeny have come from phylogenomics (Delsuc *et al.*, 2005), which considerably increases the resolving power by considering numerous concatenated genes from expressed sequence tags (ESTs) and complete genome projects (Philippe and Telford, 2006). Despite some troubled beginnings due to the shortcomings of using only a restricted set of taxa (Philippe *et al.*, 2005a), phylogenomics has provided strong corroborating support for the new animal phylogeny, essentially confirming the monophyly of Protostomia, Ecdysozoa, and Lophotrochozoa (Baurain *et al.*, 2007; Dunn *et al.*, 2008; Lartillot and Philippe, 2008; Philippe *et al.*, 2005b). Phylogenomics has also helped solving some longstanding mysteries such as the position of chaetognaths, which finally appear to belong to Protostomia (Marletaz *et al.*, 2006; Matus *et al.*, 2006) and also proposed unexpected phylogenetic affinities for enigmatic taxa such as *Buddenbrockia plumatellae* recently unmasked as a cnidarian worm (Jimenez-Guri *et al.*, 2007), or *Xenoturbella bocki*, representing a fourth deuterostome phylum on its own (Bourlat *et al.*, 2006).

Among the most groundbreaking results from recent phylogenomic studies was the identification of tunicates (or urochordates) as the closest living relatives of vertebrates, instead of cephalochordates as traditionally accepted (Delsuc *et al.*, 2006). Some hints of this unexpected result had been observed in previous large-scale phylogenetic studies including a single tunicate representative (Blair and Hedges, 2005; Philippe *et al.*, 2005b; Vienne and Pontarotti, 2006). However, a substantial increase in taxon sampling turned out to be required for recovering convincing support in favor of such an unorthodox relationship. In particular, the fact that the inclusion of the divergent appendicularian tunicate *Oikopleura dioica* did not disrupt the sister-group relationship between tunicate and vertebrates gave a good indication about the strength of the phylogenetic signal in its favor (Delsuc *et al.*, 2006). The grouping of tunicates and vertebrates had already been proposed on morphological grounds by Richard P.S. Jefferies who coined the name Olfactores after the presence a putatively homologous olfactory apparatus in fossils that were proposed to be precursors of tunicates and vertebrates (Jefferies, 1991). This phylogenetic result has nevertheless been the object of some skepticism. One reason for this maybe that it further invalidates the traditional textbook view of chordate evolution as a steady increase toward morphological complexity culminating with vertebrates, as betrayed by the use of the term Euchordates (literally “true chordates”) for denoting the grouping of cephalochordates and vertebrates (Gee, 2001). The lack of obvious morphological synapomorphies for Olfactores, apart from the presence of migratory neural crest-like cells (Jeffery, 2007; Jeffery *et al.*, 2004), and the apparent conflict with analyses of rRNA data which tend to favor Euchordates (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Winchell *et al.*, 2002) might also partly

explain the caution with which this result has been considered at first.

Phylogenomics, despite being a powerful approach, is however not immune to potential reconstruction artifacts. The possible pitfalls associated with phylogenomic studies include systematic errors that can be traced back to some kind of model misspecifications (Philippe *et al.*, 2005a) and caused mainly by heterogeneity of evolutionary rates among taxa (Lartillot *et al.*, 2007; Philippe *et al.*, 2005b) and compositional bias (Blanquart and Lartillot, 2008; Jeffroy *et al.*, 2006; Lartillot and Philippe, 2008; Phillips *et al.*, 2004). Empirical protocols have been designed to detect and reduce the impact of systematic error in genome-scale studies (Rodríguez-Ezpeleta *et al.*, 2007) but the ultimate solution lies in the development of improved models of sequence evolution (Felsenstein, 2004; Philippe *et al.*, 2005a; Steel, 2005). The reliance of current phylogenomic studies on a relatively limited number of highly expressed genes (Philippe and Telford, 2006) and the potential impact of missing data on phylogenomic inference (Hartmann and Vision, 2008; Philippe *et al.*, 2004) are also regularly cited as limitations of the phylogenomic approach.

The aim of this article is to evaluate the current evidence for the new chordate phylogeny by: (1) reanalyzing previous phylogenomic data using improved models of amino-acid replacement, (2) assembling and analyzing an updated phylogenomic dataset with more genes and more taxa, (3) assessing the impact of missing data and gene sampling on phylogenomic results, and (4) performing new analyses of rRNA data taking compositional bias into account.

MATERIALS AND METHODS

Phylogenomic Dataset Assembly

We built upon previous phylogenomic datasets assembled in the Philippe lab (Delsuc *et al.*, 2006; Jimenez-Guri *et al.*, 2007; Lartillot and Philippe, 2008; Philippe *et al.*, 2004, 2005b) to select a set of 179 orthologous markers showing sufficient conservation across metazoans to be useful for inferring the phylogeny of metazoans. Alignments were built and updated with available sequences downloaded from the Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/>) and the EST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>) of GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) using the program ED from the MUST package (Philippe, 1993). Unambiguously aligned regions were identified and excluded for each individual gene using the program GBLOCKS (Castresana, 2000) with a few manual refinements using NET from the MUST package. The complete list of genes with corresponding final numbers of amino-acid sites is available as Supporting Information.

The concatenation of the 179 genes was constructed with the program SCAFOS (Roure *et al.*, 2007) by defining 51 metazoan operational taxonomic units (OTUs)

including 25 deuterostomes representing all major lineages. When several sequences were available for a given OTU, the slowest evolving one was selected according to their degree of divergence using ML distances computed by TREE-PUZZLE (Schmidt *et al.*, 2002) under a WAG+F model (Whelan and Goldman, 2001) within SCAFOS. The percentage of missing data per taxon was reduced by creating some chimerical sequences for species belonging to the same OTU. The complete alignment consists of 179 genes and 51 taxa for 53,799 unambiguously aligned amino-acid sites with 32% missing data. To study the potential impact of missing data on phylogenetic inference (Hartmann and Vision, 2008; Philippe *et al.*, 2004; Wiens, 2006), a concatenation of the 106 genes with sequences available for at least 41 of the 51 OTUs was also constructed with SCAFOS. This reduced alignment consists of 106 genes and 51 taxa for 25,321 amino-acid sites and contains only 20% of missing data. The list of defined OTUs, chimerical sequences and percentages of missing data are available as Supporting Information. Individual gene alignments and their concatenations are available upon request.

Phylogenomic Analyses

Bayesian phylogenetic analyses of the two phylogenomic datasets were conducted using the program PHYLOBAYES 2.3c (<http://www.atgc-montpellier.fr/phylobayes/>) under the CAT+ Γ_4 site-heterogeneous mixture model (Lartillot and Philippe, 2004). For each dataset, four independent Monte Carlo Markov Chains (MCMCs) starting from a random topology were run in parallel for 20,000 cycles (1,500,000 generations), saving a point every cycle, and discarding the first 2,000 points as the burnin. Bayesian posterior probabilities (PP) were obtained from the 50% majority-rule consensus tree of the 18,000 MCMC sampled trees using the program READPB of PHYLOBAYES.

Maximum likelihood (ML) reconstruction on the new phylogenomic dataset was also performed using the program TREEFINDER version of March 2008 (Jobb *et al.*, 2004) under the empirical WAG+F+ Γ_8 model of amino-acid substitution. The α shape parameter of the Γ distribution was estimated along with the topology and the branch lengths. Reliability of nodes was estimated by bootstrap resampling with 100 pseudo-replicate datasets generated by the program SEQBOOT of the PHYLIP package (Felsenstein, 2001). The 100 corresponding ML heuristic searches were run in parallel on a computing cluster and the majority-rule consensus of the 100 resulting trees was computed using TREEFINDER.

Jackknife Procedure

The robustness of our phylogenomic inference with respect to gene sampling was assessed by applying a jackknife procedure. Fifty jackknife replicates of 50 genes drawn randomly from the full pool of 179 genes were generated. The only condition we imposed to this jackknife procedure was to require that each taxon is

represented by at least one gene in each replicate. The 50 jackknife supermatrices ranging from 11,163 to 17,181 amino-acid sites with 27 to 35% missing data were then analyzed using PHYLOBAYES under the CAT+ Γ_4 model. To ensure correct convergence of the MCMC on each replicate, an automated stopping rule was used. Specifically, for each jackknife replicate, two independent parallel (and synchronous) MCMC were run, until the posterior probability discrepancy between the two chains was less than 0.15 (maximum discrepancy over all bipartitions), and after removing the first 1,000 sampled trees of each chain as the burnin. A global majority-rule consensus tree was obtained from the 50 replicates as follows: for each jackknife replicate (D_r) taken in turn, we computed the frequency-table of all bipartitions (splits) observed in the sample collected from the posterior distribution $p(T|D_r)$. The frequencies associated to each bipartition were then averaged over the 50 replicates, and the resulting frequency table was used to build a consensus tree. The support values displayed by this Bayesian consensus tree are thus jackknife-resampled posterior probabilities (PP_{JK}). High PP_{JK} values indicate nodes that have high posterior probability support in most jackknife replicates.

Phylogenetic Analyses of Ribosomal RNA

The 46-taxa dataset of combined 18S+28S rRNA genes assembled by Mallatt and Winchell (2007) for studying deuterostome phylogeny was reanalyzed. This alignment contains a total of 3,925 unambiguously aligned nucleotide sites. A principal component analysis (PCA) of nucleotide composition was realized using the R statistical package (R Development Core Team, 2007). The best fitting model of nucleotide sequence evolution was evaluated using MODELTEST 3.7 (Posada and Crandall, 1998). The TIM+ Γ_4 +I transitional model (Posada and Crandall, 2001) was selected according to the Akaike information criterion. ML phylogenetic analysis of this nucleotide dataset was conducted with PAUP* 4.0b10 (Swofford, 2002) using a heuristic search with Tree Bisection Reconnection (TBR) branch swapping starting from a Neighbor-Joining (NJ) tree.

The nucleotide dataset was RY-coded by pooling puRines (AG = R) and pYrimidines (CT = Y) in an attempt to alleviate both compositional heterogeneity and substitutional saturation of transition events. This RY-coded dataset was then analyzed by conducting a ML heuristic search with TBR branch swapping on a NJ starting tree using PAUP* under the CF+ Γ_8 model for discrete characters (Cavender and Felsenstein, 1987). The α shape parameter of the Γ distribution was previously estimated during a ML heuristic search on the nucleotide dataset conducted with TREEFINDER under the GTR2+ Γ_8 two-state model.

Reliability of nodes was estimated for each dataset by nonparametric bootstrap resampling using 100 pseudo-replicates generated by SEQBOOT. The 100 corresponding ML heuristic searches using PAUP* with the previ-

ously estimated ML parameters, NJ starting trees, and TBR branch swapping were parallelized on a computing cluster. ML bootstrap percentages were obtained from the 50% majority-rule consensus tree of the 100 bootstrap ML trees using TREEFINDER.

RESULTS AND DISCUSSION

Effect of an Improved Model of Sequence Evolution

Our initial assessment of deuterostome phylogenetic relationships was based on a phylogenomic dataset encompassing 146 nuclear genes (33,800 amino-acids) from 38 taxa including 14 deuterostomes (Delsuc *et al.*, 2006). ML and Bayesian phylogenetic analyses conducted under the standard WAG+F+ Γ_4 model provided strong support for grouping tunicates with vertebrates (including cyclostomes), but also disrupted chordate monophyly because cephalochordates grouped with echinoderms, albeit with nonsignificant statistical support (Delsuc *et al.*, 2006). The limited taxon sampling available at the time for Ambulacraria (echinoderms and hemichordates), that is, a single echinoderm, prompted us to be cautious about this result and to call for the inclusion of xenoturbellidans, hemichordates, and more echinoderms before drawing definitive conclusions. In fact, a subsequent phylogenomic study did exactly what we pleaded for by adding a representative species for each of these three groups (Bourlat *et al.*, 2006). The inclusion of these taxa allowed retrieving the monophyly of chordates in Bayesian analyses using standard amino-acid models, although the alternative hypothesis of chordate paraphyly was still not statistically rejected by ML nonparametric tests (Bourlat *et al.*, 2006). Importantly, the strong statistical support for the monophyly of Olfactores was unaffected by taxon addition (Bourlat *et al.*, 2006).

Models accounting for site-specific modulations of the amino-acid replacement process, such as the CAT mixture model (Lartillot and Philippe, 2004), seem to offer a significantly better fit to real data than empirical substitution matrices currently used in standard models of amino-acid sequence evolution. Accounting for site-specific amino-acid propensities has also been shown to induce a significant improvement of phylogenetic reconstruction in difficult cases such as long-branch attraction (Baurain *et al.*, 2007; Lartillot *et al.*, 2007; Lartillot and Philippe, 2008). This improvement essentially lays in the ability of the CAT model to detect multiple conservative substitutions more efficiently than standard amino-acid models (Lartillot *et al.*, 2007).

To test for an eventual effect of model misspecification on previous phylogenomic analyses, we reanalyzed our previous dataset (Delsuc *et al.*, 2006) under the CAT+ Γ_4 model. This analysis provides strong corroborating support for the grouping of tunicates and vertebrates (see Fig. 1). However, in contrast with previous analyses using empirical amino-acid replacement matrices,

which favored a sister-group relationship between cephalochordates and echinoderms, the use of the CAT+ Γ_4 mixture model strongly supports the classical view of monophyletic chordates and deuterostomes (see Fig. 1). The fact that chordate polyphyly is disrupted both by a richer taxon sampling (Bourlat *et al.*, 2006), or upon the use of a more elaborate model, suggests that the previously observed grouping of cephalochordates and echinoderms (Delsuc *et al.*, 2006) was probably a phylogenetic reconstruction artifact. On the other hand, the fact that the grouping of tunicates and vertebrates is insensitive to the model used, adds further credence to the Olfactores hypothesis.

An Updated Phylogenomic Dataset

The continuously growing genomic databases allowed us to build an updated phylogenomic dataset that includes both more genes and more taxa than previously considered to address the question of deuterostome phylogeny. This new dataset of 179 genes for 51 taxa includes 25 deuterostomes representing all major lineages: Xenoturbellida (1 taxon), Hemichordata (1), Echinodermata (5), Cephalochordata (1), Tunicata (6), Cyclostomata (2) and Vertebrata (9), plus 26 selected slow evolving metazoan taxa including Cnidarians and Poriferans as the most distant outgroups. Chordates are particularly well sampled with the inclusion, for the first time, of six tunicate species covering the four major clades evidenced by 18S rRNA studies (Swalla *et al.*, 2000). This diverse taxon sampling is essential to further test the new chordate phylogeny recently revealed by phylogenomics (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006).

Bayesian (CAT+ Γ_4) and ML (WAG+F+ Γ_8) phylogenetic reconstructions conducted on this updated dataset (179 genes, 53,799 amino-acid sites, 51 taxa) resulted in a highly resolved tree (Fig. 2a). These analyses provided strong support for Ambulacraria (PP_{CAT+ Γ_4} = 1.0/BP_{WAG+F+ Γ_8} = 97), chordates (1.0/69) and olfactores (1.0/100). Xenambulacraria (*Xenoturbella* + Ambulacraria) and a basal position for the chaetognath *Spadella* among Protostomia were also moderately supported by our analyses (Fig. 2a). These results are compatible with a recent phylogenomic analysis, which also found strong support for Ambulacraria, chordates, and Olfactores when using the CAT mixture model (Dunn *et al.*, 2008). However, the monophyly of Deuterostomes is unresolved in both Bayesian and ML phylogenetic reconstructions (Fig. 2a).

The complete dataset obtained by concatenating all 179 genes contains 32% missing data. Previous studies of the impact of missing data on the accuracy of phylogenetic inference have concluded that probabilistic methods are relatively tolerant to missing data (Hartmann and Vision, 2008; Philippe *et al.*, 2004; Wiens, 2003, 2005), the most important factor being the absolute amount of available data for a given taxon. In phylogenomics, even incomplete taxa are usually represented by thousand of sites, and the impact of missing data on accuracy is

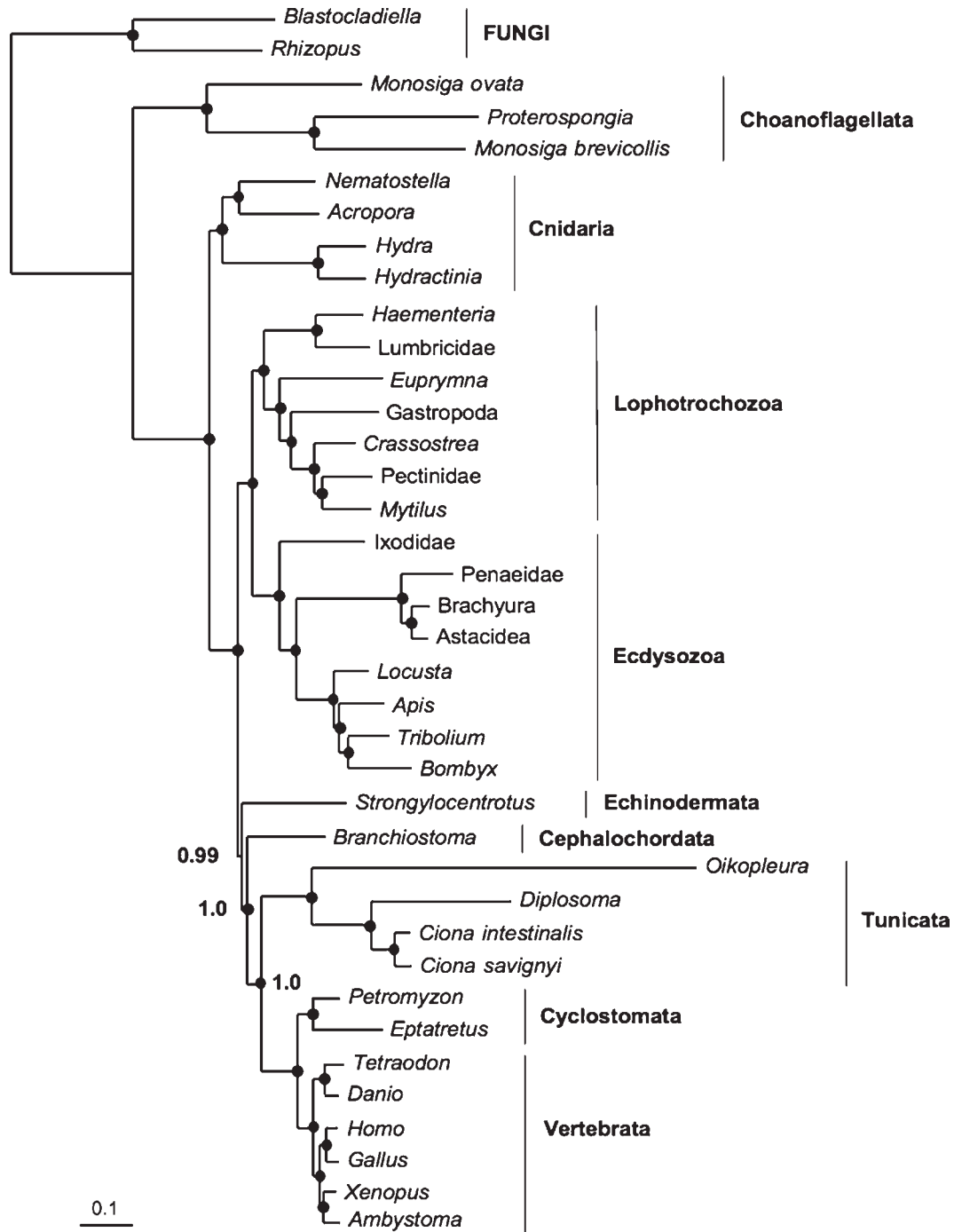


FIG. 1. Reanalysis of previous phylogenomic data using an improved model of sequence evolution. The Delsuc *et al.* (2006) phylogenomic dataset of 146 genes (38 taxa and 33,800 sites) was analyzed under the CAT+ Γ_4 site-heterogeneous mixture model of amino-acid replacement. Values at nodes represent Bayesian posterior probabilities (PP). Circles indicate nodes with maximal support PP = 1.0. The scale bar represents the estimated number of substitutions per site.

therefore relatively limited (Philippe *et al.*, 2004). Nonetheless, incomplete taxa might still be difficult to place with confidence especially when they represent isolated lineages such as *Xenoturbella* (65% missing data) and *Spadella* (75%) in our dataset. To control for a potential effect of missing data on our phylogenomic results, we

restricted our dataset to the 106 genes with sequences available for at least 41 of the 51 taxa. The concatenation of these 106 genes produces a matrix with 25,321 amino-acid sites that contains only 20% of missing data.

Bayesian and ML phylogenetic inference on this reduced dataset produced a tree fully congruent with

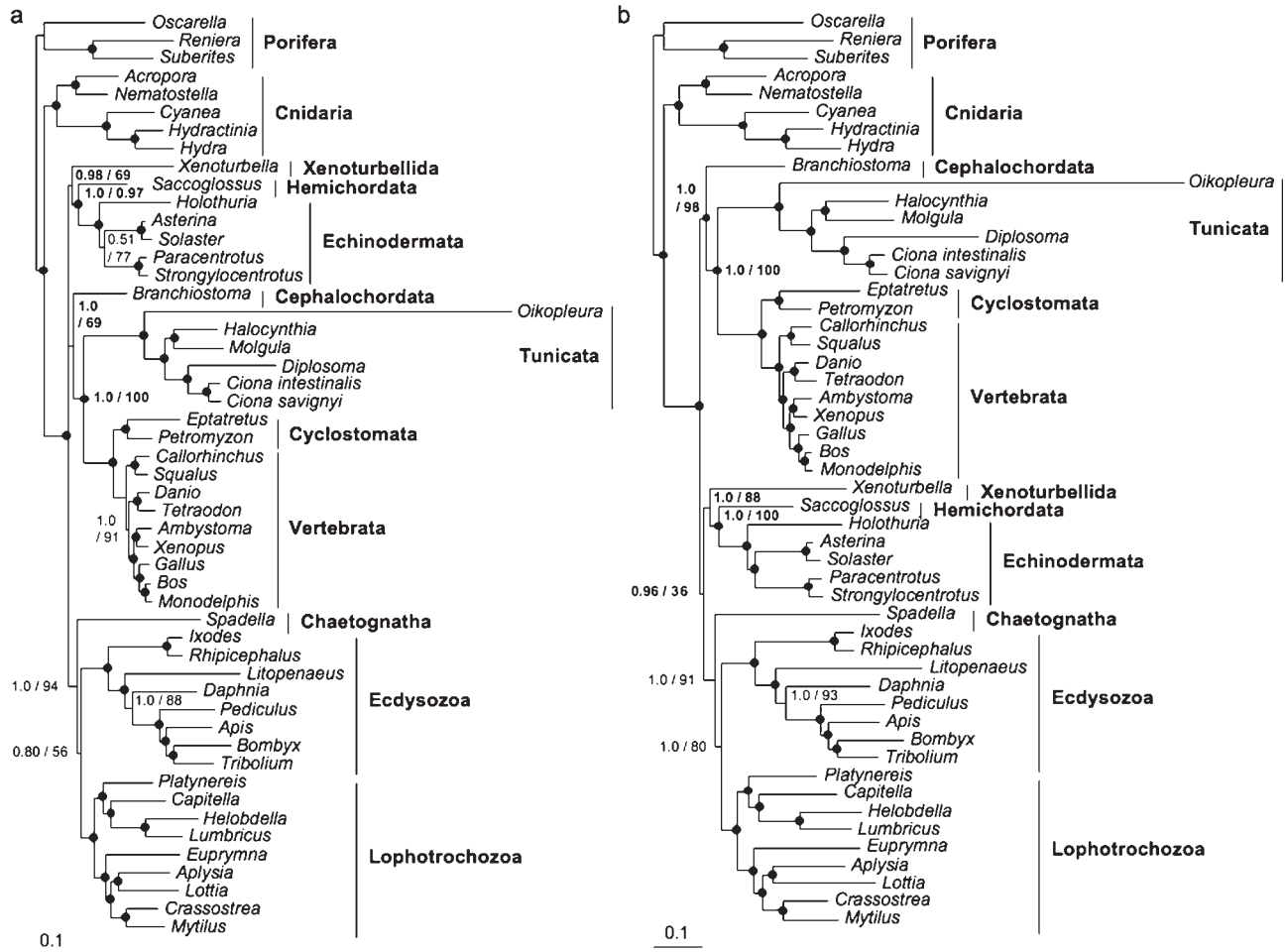


FIG. 2. Phylogenetic analyses of an updated phylogenomic dataset. (a) Bayesian consensus tree obtained using the CAT+ Γ_4 mixture model on the complete dataset based on the concatenation of 179 genes (51 taxa and 53,799 amino-acid sites) containing 32% missing data. (b) Bayesian inference using the CAT+ Γ_4 mixture model on the dataset reduced to the concatenation of the 106 genes for which sequences were available for at least 41 of the 51 taxa (25,321 amino-acid sites) containing only 20% of missing data. Values at nodes indicate Bayesian posterior probabilities (PP)/Maximum-likelihood bootstrap percentages (BP; 100 replicates) obtained under the WAG+ Γ_8 . Circles indicate strongly supported nodes with PP ≥ 0.95 and BP ≥ 95 . The scale bar represents the estimated number of substitutions per site.

the phylogenetic picture given by the complete dataset (Fig. 2b). In particular, the support for the monophyly of chordates was still maximal in terms of PP_{CAT+ Γ_4} , but BP_{WAG+F+ Γ_8} increased from 69 to 88%. The monophyly of Olfactores received maximal support in both cases and appeared not affected by missing data. Statistical support in terms of PP_{CAT+ Γ_4} and BP_{WAG+F+ Γ_8} was generally increased especially for locating incomplete taxa such as the *Xenoturbella* as the sister-group to Ambulacraria (from 0.98/69 to 1.0/88) and *Spadella* at the base of Protostomia (from 0.80/56 to 1.0/80). Altogether, reducing the amount of missing data, despite also reducing the total number of available sites, seems to result in a slight increase in bootstrap proportions. The only real noticeable difference between the two methods concerns the monophyly of deuterostomes. The Bayesian inference under the CAT mixture model suggests deuterostome paraphyly by supporting a basal position of chor-

dates within Bilateria (Fig. 2b) as previously reported (Lartillot and Philippe, 2008). However, ML retrieved the monophyly of deuterostomes, but with BP_{WAG+F+ Γ_8} of only 50%, leaving the monophyly of deuterostomes unresolved by our data.

Robustness of Phylogenomics to Gene Sampling

A legitimate question that can be directed to the phylogenomic approach is the degree to which the results are robust to the sample of genes used to infer phylogenetic trees. This potential concern was addressed by applying a jackknife statistical resampling protocol: fifty datasets were assembled by randomly drawing 50 genes from the total 179 genes, and subjected to Bayesian phylogenetic reconstruction using the CAT+ Γ_4 mixture model (see Methods). The resulting majority-rule consensus tree shows that the vast majority of inferred phy-

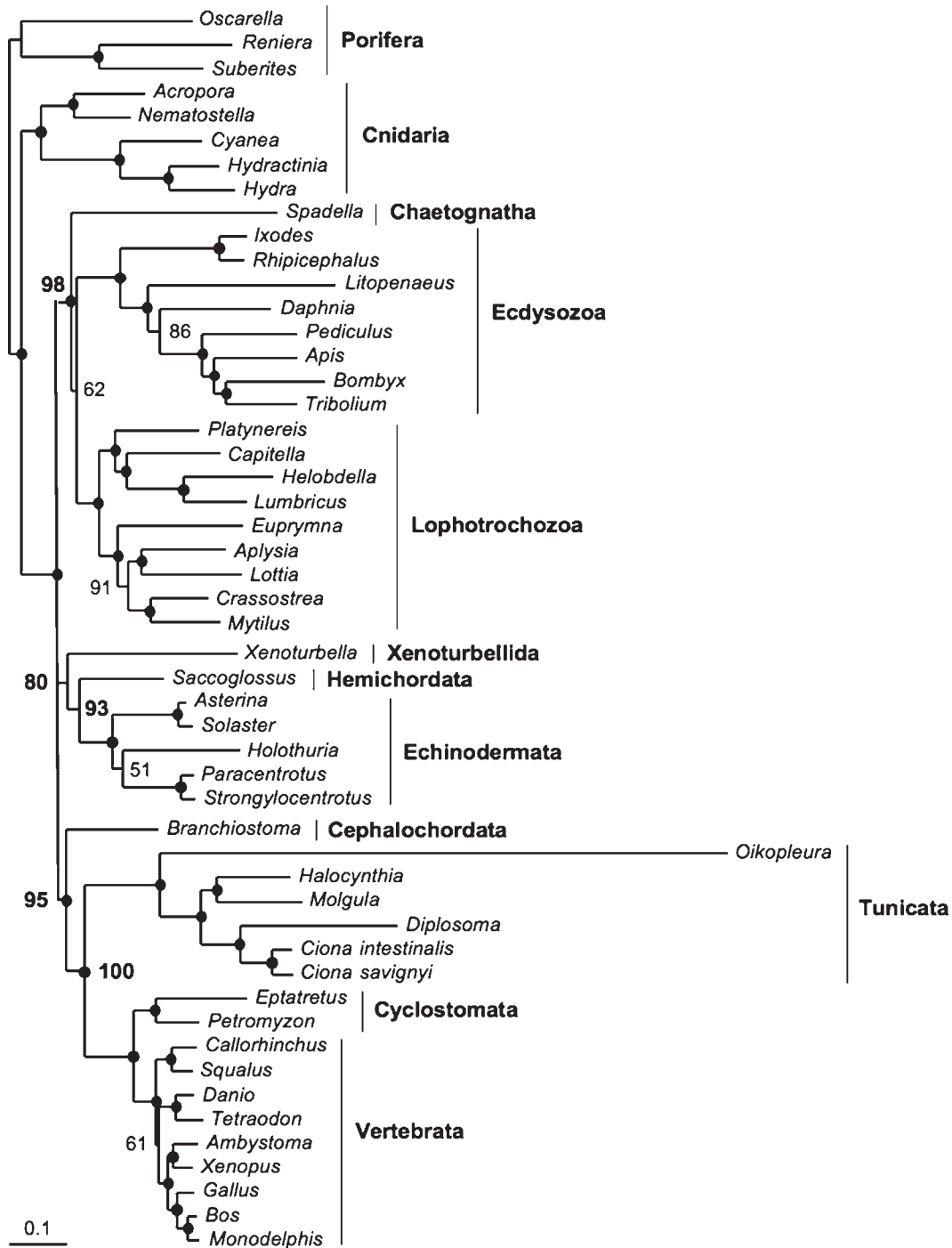


FIG. 3. Assessing the robustness of phylogenetic results to gene sampling using a jackknife procedure. The Bayesian phylogenetic inference was conducted under the CAT+ Γ_4 mixture model on 50 jackknife replicates of 50 genes over a total of 179. The tree presented is the weighted majority-rule consensus of all trees sampled every 10 cycles across the 50 replicates after removing the first 1000 trees in each MCMC as the burnin. Values at nodes represent corresponding jackknife-resampled posterior probabilities indices (PP_{JK}). Circles indicate highly repeatable nodes with PP_{JK} ≥ 95%. The scale bar represents the number of substitutions per site.

lognetic relationships are highly repeatable across the 50 jackknife replicates (see Fig. 3). Olfactores, Chordata, and Ambulacraria all received PP_{JK} of more than 90% indicating that phylogenetic support is not dependent upon a particular gene combination. Xenambulacraria

appears slightly more affected by gene sampling (PP_{JK} = 80%), but this relative instability might be explained by the poor gene representation available for *Xenoturbella* with only 98 genes over 179 (55%). The same kind of reasoning could apply to the relatively unstable posi-

tions of the chaetognath *Spadella* within protostomes ($PP_{JK} = 62\%$) and of *Holothuria* within echinoderms ($PP_{JK} = 51\%$) (see Fig. 3).

In fact, the only major clade whose monophyly appears to be influenced by gene sampling is deuterostomes for which PP_{JK} was less than 50% (see Fig. 3). In practice, this means that depending on the particular combination of 50 genes considered, deuterostomes might appear either monophyletic or paraphyletic, with the three possible topological alternatives retrieved in similar proportions: Deuterostomes (38%), basal chordates (28%), and basal Xenambulacraria (22%). Despite the inclusion of 25 taxa representing all major lineages in our dataset, these results confirm deuterostomes as one of the most difficult groups to resolve in the animal phylogeny despite its wide acceptance (see Lartillot and Philippe, 2008).

New Analyses of Ribosomal RNA Genes

The sister-group relationship between tunicates and vertebrates (Olfactores) observed in phylogenomics is in conflict with most (if not all) analyses of rRNA which favor cephalochordates as the closest relatives of vertebrates (Euchordates) (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Swalla *et al.*, 2000; Wada and Satoh, 1994; Winchell *et al.*, 2002). However, the statistical support for Euchordates in rRNA-based phylogenetic studies is moderate. Indeed, the first 18S rRNA study, based on a limited taxon sampling of deuterostomes, reported a bootstrap value of only 45% for Euchordates (Wada and Satoh, 1994). A subsequent 18S rRNA study considering only slowly evolving sequences for 16 deuterostomes found only a moderate bootstrap support of 71% for grouping cephalochordates and vertebrates (Cameron *et al.*, 2000). A study focused on tunicates also obtained moderate support for Euchordates (58 to 85% depending on the dataset and reconstruction method) but failed to support chordate monophyly likely because tunicate 18S rRNA sequences are rapidly evolving (Swalla *et al.*, 2000).

The next studies used the combination of 18S and 28S rRNAs. An investigation using 28 taxa for the two rRNA subunits found strong bootstrap support (89 to 97% depending on the method) for Euchordates (Winchell *et al.*, 2002). However, this study again failed to support chordate monophyly. Detailed analyses confirmed that tunicate genes have evolved rapidly and showed that they are compositionally biased toward AT, rendering tunicates virtually impossible to locate convincingly in the tree on the basis of rRNA data (Winchell *et al.*, 2002). Finally, increasing the sampling to 46 taxa for this 18S+28S rRNA data did not help in further resolving the relationships among the major groups of deuterostomes and even decreased the ML bootstrap support for Euchordates from 97% in the previous study to 50% (Mallatt and Winchell, 2007).

To gauge the extent to which the rRNA data conflicts with our phylogenomic results, we reanalyzed the 46-

taxa dataset of Mallatt and Winchell (2007). The heterogeneity of base composition in this dataset is well illustrated by the PCA presented in Figure 4a. At one extreme, tunicates (especially *Oikopleura*) are particularly AT-rich, and at the other extreme, Myxinidae (*Myxine* and *Eptatretus*) and the pterobranch hemichordate *Cephalodiscus* are highly GC-rich. We therefore compared phylogenetic reconstructions conducted on nucleotides and on RY-coded data, a coding scheme allowing reducing both substitutional saturation and nucleotide compositional bias (Fig. 4b). The two inferred ML trees appear mostly congruent except for two major topological shifts.

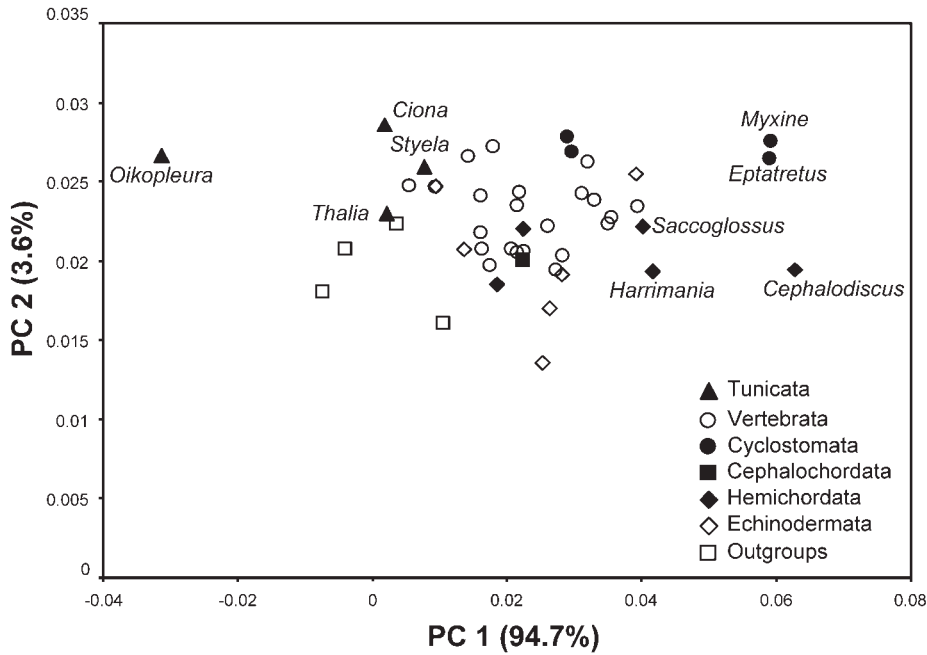
The strongest topological change occurred within hemichordates (Fig. 2b). Although the use of a standard DNA model strongly supports the paraphyly of enteropneusts by grouping the pterobranch *Cephalodiscus* with *Saccoglossus* and *Harrimania* (BP = 95), RY-coding allows recovering the monophyly of enteropneusts with high bootstrap support (BP = 90). This helps in understanding the conflict between 18S rRNA that supports enteropneust paraphyly (Cameron *et al.*, 2000; Halanych, 1995) and 28S rRNA that rather favors their monophyly (Mallatt and Winchell, 2007; Winchell *et al.*, 2002). This result is of particular importance because it potentially invalidates the controversial hypothesis that pterobranchs evolved from an enteropneust (Cameron *et al.*, 2000; Halanych, 1995) by suggesting that it is likely an artifact of 18S rRNA-based phylogenetic reconstructions due to shared nucleotide compositional bias between pterobranchs and Harrimaniidae (Fig. 4a).

Second, the support for the monophyly of Euchordates observed with nucleotides (BP = 84) disappeared in favor of the monophyly of Olfactores in the RY-coding dataset, albeit with no statistical support (BP = 44). This nevertheless strongly suggests that the high compositional bias of tunicate sequences has blurred the phylogenetic signal for Olfactores in previous analyses. Thus, according to our interpretation, reducing compositional bias and substitutional saturation by RY-recoding allows recovering a limited signal for Olfactores in agreement with our phylogenomic analysis of amino-acid data. It is worth noting however that rRNA does not statistically support chordate monophyly in both cases (Fig. 2b).

Molecular Phylogenetic Conclusions

Our aim was to reanalyze the phylogenetic relationships among chordates. The revision of the position of tunicates proposed by recent phylogenomic studies (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006; Dunn *et al.*, 2008) by concluding in favor of the monophyly of Olfactores, has not yet been considered as totally convincing, essentially because it is at odds with both the traditional view based on embryological and morphological characters (Rowe, 2004; Schaeffer, 1987), and with earlier molecular phylogenetic analyses based on rRNA (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Swalla *et al.*, 2000; Wada and Satoh, 1994; Winchell *et al.*, 2002). The

a



b

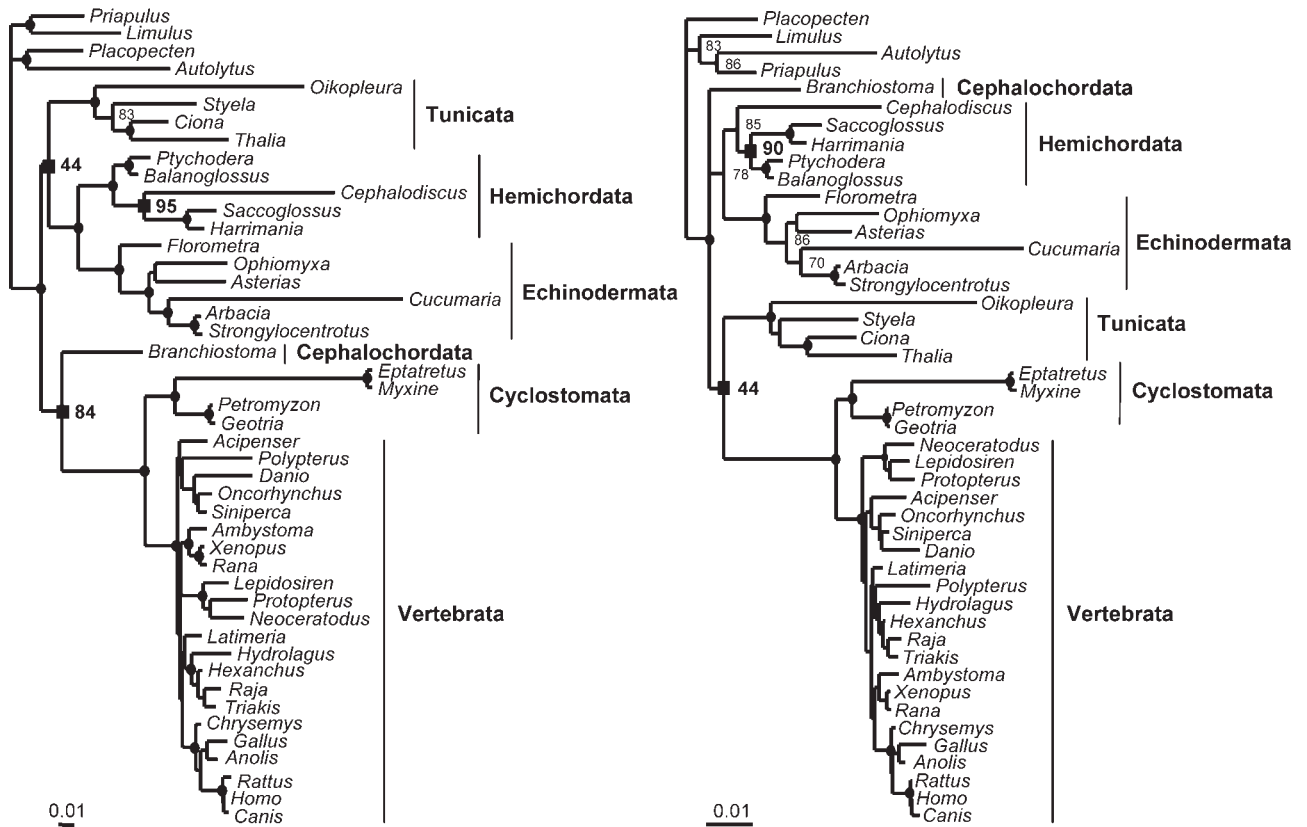


FIG. 4. New phylogenetic analyses of ribosomal RNA genes. (a) Principal component analysis of nucleotide composition of the combined 18S+28S rRNA dataset. The graph represents the projection of individuals on the first two axes, which explain more than 98% of the total variance. (b) Maximum-likelihood analyses of the 18S+28S dataset using the best-fitting standard DNA model (TIM+ Γ_4 +I) on nucleotides (left) and a two-state model (CF+ Γ_8) after RY-coding of nucleotides (right). ML bootstrap percentages are given at nodes when greater than 70 except within vertebrates. Circles indicate strongly supported nodes with BP ≥ 95 . Squares point to shifting nodes of interest between the two ML trees. Scale bars represent the number of substitutions per site.

unexpected sister-group relationship between echinoderms and cephalochordates observed in one of these studies (Delsuc *et al.*, 2006) may also have suggested the possibility that the monophyly of Olfactores was due to an artefactual attraction of cephalochordates with echinoderms (Bourlat *et al.*, 2006).

In this analysis, we have tried to address these points, essentially by reanalyzing both phylogenomic and rRNA data, under better taxonomic sampling and using more elaborate methods and probabilistic models. First, we demonstrate that, although the grouping of echinoderms and cephalochordates was indeed a probable artifact, disappearing upon the addition of several taxa or using an improved model of sequence evolution, the monophyly of Olfactores appears to be robust with respect to taxon sampling and model choice. Second, our reanalysis of rRNA data using RY-recoding also reveals a weak signal in favor of Olfactores, and suggests that the grouping of vertebrates and cephalochordates in former studies may have been an artifact driven by compositional biases. Altogether, our analyses allows a coherent interpretation of all empirical results observed thus far concerning chordate phylogeny, yielding further evidence in favor of the monophyly of both chordates and Olfactores.

At larger scale, however, we observe an overall lack of support for the monophyly of deuterostomes. Deuterostomes have nearly unanimously been considered as an unquestionable monophyletic group, a hypothesis backed up by traditional comparative analyses of embryological characters such as the fate of the blastopore (Nielsen, 2001), and morphological traits such as gill slits (Schaeffer, 1987). However, in our analyses, the status of deuterostomes seems to be sensitive to the model used, with CAT slightly favoring the paraphyletic configuration, and WAG the more traditional monophyly. In either case, the support measured by nonparametric resampling procedures (site-wise bootstrap or gene-wise jackknife) is weak.

Other phylogenomic studies (Dunn *et al.*, 2008; Lartillot and Philippe, 2008) also failed to obtain strong support for the relative phylogenetic positions of chordates and Ambulacraria. Moderate support for the monophyly of Deuterostomes was only obtained under empirical matrix models, the support disappearing when the CAT model was used instead (Dunn *et al.*, 2008; Lartillot and Philippe, 2008). Although profile mixture models such as CAT, whereas having a better fit than empirical matrices such as WAG, may have some inherent weaknesses as to their phylogenetic accuracy, the WAG empirical matrix fails in many cases, particularly when confronted to a high level of saturation (Lartillot *et al.*, 2007). This casts doubts on results that seem to receive support exclusively under this model, as is the case for deuterostome monophyly. More observations are needed to better gauge the relative merits of either type of model. Overall, although deuterostome monophyly still remains a reasonable working hypothesis to date, more work is needed before the question can be settled.

Corroborative Evidence for Olfactores Monophyly

The monophyly of Olfactores receives strong support from sequence-based phylogenomic inference. Rare genomic changes have also provided some evidence in its favor: the domain structure of cadherins (Oda *et al.*, 2002), a unique amino-acid insertion in fibrillar collagen (Wada *et al.*, 2006), and the distribution of micro RNAs (miRNAs) (Heimberg *et al.*, 2008). Finally, the *Branchiostoma floridae* genome helps confirming the sister-group relationship between tunicates and vertebrates in offering additional evidence from analyses of intron dynamics (Putnam *et al.*, 2008).

Cadherins are a superfamily of highly conserved adhesion molecules mediating cell communication and signaling that are pivotal for developmental processes of multicellular organisms. Their recent detection in the closest unicellular relatives of metazoans, the choanoflagellates, has highlighted their potential role in the origin of multicellularity (Abedin and King, 2008). Comparative studies on the classic cadherin subfamily has revealed that the structural element called Primitive Classic Cadherin Domain (PCCD) complex, otherwise termed non-chordate classic cadherin domain, is also present in cephalochordates, but has been lost in both tunicates and vertebrates (Oda *et al.*, 2002). The most parsimonious scenario is that this particular protein domain complex has been lost in the common ancestor of tunicates and vertebrates and constitutes a synapomorphy of Olfactores. However, cephalochordates possess two classic cadherin genes which originated by lineage-specific tandem duplication and that have a particular structure in lacking extracellular repeats found in all other investigated metazoans (Oda *et al.*, 2004). This derived state renders difficult to ascertain domain homology among chordate classic cadherin genes and casts doubt on its phylogenetic significance.

Further potential evidence for the clade Olfactores has been inferred from the evolution of fibrillar collagen genes within chordates. These genes represent important components of the notochord, the cartilage and mineralized bones in vertebrates. Phylogenetic analyses suggested that three ancestral fibrillar collagens gave rise to the gene diversity observed in living deuterostomes (Wada *et al.*, 2006). Comparative sequence analyses showed that tunicates and vertebrates share a molecular signature in the form of a six to seven amino-acid insertion in the C-terminus noncollagenous domain of one type of fibrillar collagens, that is absent in cephalochordates and echinoderms (Wada *et al.*, 2006). This insertion was interpreted as supporting the idea that vertebrates are more closely related to tunicates than to cephalochordates (Wada *et al.*, 2006). The homology of the insertion appears nevertheless difficult to assert with certainty given the high degree of sequence divergence observed in this region of the molecule. More tunicate fibrillar collagen sequences might help in better understanding the dynamics of this peculiar amino-acid insertion and the phylogenetic signal it conveys.

The comparison of miRNA repertoires in metazoans has also recently unearthed some potential signatures for the sister-group relationship of tunicates and vertebrates (Heimberg *et al.*, 2008). miRNAs are small non-coding RNAs involved in regulation of gene expression in eukaryotes and play an important role in the development of metazoans. Comparative genomic studies of miRNAs underlined that, during the evolution of metazoans, major body-plan innovations seemed to coincide with dramatic expansions of miRNA repertoires, suggesting a potential role in the increase of morphological complexity (Hertel *et al.*, 2006; Sempere *et al.*, 2006). The most recent study unveiled that three miRNA families (*mir-126*, *mir-135*, and *mir-155*) were likely acquired in the common ancestor of tunicates and vertebrates (Heimberg *et al.*, 2008). Taking into consideration that miRNAs might be only rarely secondarily lost once they have been recruited, this finding provides corroborative evidence for the clade Olfactores. It should be noted however that, of these three miRNA families, only *mir-126* constitutes an exclusive synapomorphy for Olfactores without subsequent secondary loss in descendant taxa confirmed by Northern analysis. Moreover, the profound reorganization of miRNA repertoire undergone by tunicates requires being cautious when interpreting acquisition of miRNAs as potential signatures for reconstructing their phylogenetic relationships (Fu *et al.*, 2008).

Additional sequence-based phylogenomic reconstructions and analyses of rare genomic changes have been issued along with the recently published draft sequence of a cephalochordate (*Branchiostoma floridae*) genome (Putnam *et al.*, 2008). The phylogenetic analysis of a concatenation of 1,090 orthologs from 12 complete genomes retrieved maximal Bayesian support for Olfactores and chordates, whereas the corresponding bootstrap support was maximal for Olfactores but of only 78% for chordate monophyly (Putnam *et al.*, 2008). Moreover, the analysis of individual gene phylogenies revealed twice more cases where Olfactores was favored over Euchordates than the reverse (Putnam *et al.*, 2008). Further evidence was obtained by analyzing the phylogenetic signal deduced from the dynamics of intron gain and loss among chordate genomes. Despite extensive intron losses along the tunicate lineage, a number of shared intron gain/loss events can be identified as a signature of tunicates and vertebrates common ancestry (Putnam *et al.*, 2008). Overall, the new evidence brought by the analysis of the *Branchiostoma floridae* genome essentially corroborates our phylogenetic results.

Implications for Chordate Evo-Devo

The additional evidence presented for the new chordate phylogeny provides a robust phylogenetic framework for (re)interpreting the evolution of morphological characters and developmental features. Inverting the phylogenetic position of tunicates and cephalochordates within monophyletic chordates highlights the preva-

lence of morphological simplification with characters that are likely ancestral for chordates, such as metameric segmentation, being lost secondarily in the tunicate lineage. On the other hand, the loss of preoral kidney and the presence of multiciliated epithelial cells might in fact constitute morphological synapomorphies for olfactores (Ruppert, 2005). The new chordate phylogeny further portrays tunicates as highly derived chordates with specialized lifestyles and developmental modes, whereas cephalochordates might have retained more ancestral chordate characteristics. We will use two examples to illustrate the importance of considering the new phylogenetic status of tunicates as the sister-group of vertebrates in the context of evolutionary developmental biology.

The first illustration concerns the evolutionary origin of such fundamental structures as the neural crest and olfactory placodes. Migratory neural crest cells and sensory placodes have long been considered as vertebrate innovations. Implicated respectively in the development of major tissues and sensory organs, their origin is generally correlated with the increase in morphological complexity of vertebrates. However, recent molecular developmental studies have revealed the presence in tunicates of migratory neural crest-like cells (Jeffery, 2006; Jeffery *et al.*, 2004) and olfactory placodes (Bassham and Postlethwait, 2005; Mazet *et al.*, 2005). When reinterpreted in light of the new chordate phylogeny, these results implied that both of these features did not evolve *de novo* in the vertebrate lineage, but rather evolved from specialized pre-existing structures in the common ancestor of vertebrates and tunicates.

The second example illustrates how the new phylogenetic context helps in understanding the genomic and developmental peculiarities of tunicates within chordates. The new phylogenetic picture implied that tunicate genomes have undergone significant genome reduction from the ancestral chordate genome (Holland, 2007). This genome compaction is also associated with a high rate of genomic evolution at the levels of both primary sequences (Delsuc *et al.*, 2006; Edvardsen *et al.*, 2004) and genome organization (Holland and Gibson-Brown, 2003). One of the most spectacular rearrangements of tunicate genomes is the loss of several Hox genes, the disintegration of the Hox cluster, and the loss of temporal colinearity in Hox gene expression during development (Ikuta *et al.*, 2004; Seo *et al.*, 2004). These observations raise the question of how tunicates, with their altered Hox clusters, are still able to develop a chordate body plan. In chordates, and deuterostomes more generally, temporal colinearity is regulated by the Retinoic-Acid (RA) signaling pathway which controls the antero-posterior patterning of the embryo (Cañestro *et al.*, 2006; Marlétaz *et al.*, 2006). However, axial patterning in tunicates seems to have become independent of RA-signaling, with the genes of the RA machinery even being lost in *Oikopleura* (Cañestro and Postlethwait, 2007). Functional studies have shown that if "*Oikopleura* can be considered as a classical RA-signal-

ing knock-down mutant naturally produced by evolution," it is still capable of developing a typical chordate body plan (Cañestro and Postlethwait, 2007). With cephalochordates, which possess the RA genomic toolkit, being basal among chordates, RA-signalling must have been present in the tunicate ancestor and secondarily lost in *Oikopleura* suggesting that appendicularians use alternative mechanisms for the development of chordate features (Cañestro *et al.*, 2007; Holland, 2007).

The new chordate phylogeny strengthens the view that tunicates and cephalochordates represent complementary models for studying vertebrate Evo-Devo (Schubert *et al.*, 2006). Tunicates are phylogenetically closer to vertebrates but appear both morphologically and molecularly highly derived. The diversity of their developmental modes offers the opportunity to study the evolution of alternative adaptive solutions to the typical chordate development. In having retained more ancestral features, cephalochordates provide an ideal outgroup for polarizing evolutionary changes that occurred in tunicates and vertebrates. With the cephalochordate *Branchiostoma floridae* genome (Putnam *et al.*, 2008) and the upcoming genome sequence of the appendicularian *Oikopleura dioica*, the newly established phylogenetic framework makes chordate comparative genomics appearing full of promises for the Evo-Devo community as exemplified in a recent work on the origin and evolution of the Pax gene family (Bassham *et al.*, 2008).

ACKNOWLEDGMENTS

The authors thank the associate editors Billie Swalla and José Xavier-Neto for giving the opportunity to write this article. The authors also thank John Mallatt for kindly providing his 18S-28S rRNA alignment, Julien Claude for help in using R statistical package, and two anonymous reviewers for their comments. The extensive phylogenomic calculations benefited from the ISEM computing cluster.

LITERATURE CITED

- Abedin M, King N. 2008. The premetazoan ancestry of cadherins. *Science* 319:946–948.
- Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R. 2000. The new animal phylogeny: Reliability and implications. *Proc Natl Acad Sci USA* 97:4453–4456.
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387:489–493.
- Bassham S, Cañestro C, Postlethwait JH. 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol* 6:35.
- Bassham S, Postlethwait JH. 2005. The evolutionary history of placodes: A molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. *Development* 132:4259–4272.
- Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? *Mol Biol Evol* 24:6–9.
- Blair JE, Hedges SB. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol* 22:2275–2284.
- Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 25:842–858.
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444:85–88.
- Cameron CB, Garey JR, Swalla BJ. 2000. Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. *Proc Natl Acad Sci USA* 97:4469–4474.
- Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* 305:522–538.
- Cañestro C, Postlethwait JH, Gonzalez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? *Evol Dev* 8:394–406.
- Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932–942.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
- Cavender JA, Felsenstein J. 1987. Invariants of phylogenies in a simple case with discrete states. *J Classif* 4:57–71.
- Conway-Morris S. 2003. The Cambrian "explosion" of metazoans and molecular biology: Would Darwin be satisfied? *Int J Dev Biol* 47:505–515.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968.
- Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 6:361–375.
- Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martinale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Edwardsen RB, Lerat E, Maeland AD, Flat M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D. 2004. Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. *J Mol Evol* 59:448–457.
- Felsenstein J. 2001. PHYLIP (Phylogenetic Inference Package) version 3.6. (distributed by the author). Seattle: Department of Genetics, University of Washington.
- Felsenstein J. 2004. *Inferring phylogenies*. Sunderland, MA, USA: Sinauer Associates, Inc. 645 p.
- Fu X, Adamski M, Thompson EM. 2008. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Mol Biol Evol* 25:1067–1080.
- Gee H. 2001. Deuterostome phylogeny: the context for the origin and evolution of the vertebrates. In: Ahlberg PE, editor. *Major events in early vertebrate evolution: Palaeontology, phylogeny, genetics, and development*. London: Taylor and Francis. pp 1–14.
- Halanych KM. 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Mol Phylogenet Evol* 4:72–76.
- Halanych KM. 2004. The new view of animal phylogeny. *Annu Rev Ecol Syst* 35:229–256.
- Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* 267:1641–1643.
- Hartmann S, Vision TJ. 2008. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol Biol* 8:95.
- Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* 105:2946–2950.
- Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7:25.
- Holland LZ. 2007. Developmental biology: A chordate with a difference. *Nature* 447:153–155.
- Holland LZ, Gibson-Brown JJ. 2003. The *Ciona intestinalis* genome: When the constraints are off. *Bioessays* 25:529–532.
- Ikuta T, Yoshida N, Satoh N, Saiga H. 2004. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci USA* 101:15118–15123.

- Jefferies RPS. 1991. Two types of bilateral symmetry in the Metazoa: Chordate and bilaterian. In: Bock GR, Marsh J, editors. *Biological asymmetry and handedness*. Chichester: Wiley. pp 94–127.
- Jeffery WR. 2006. Ascidian neural crest-like cells: Phylogenetic distribution, relationship to larval complexity, and pigment cell fate. *J Exp Zool B Mol Dev Evol* 306:470–480.
- Jeffery WR. 2007. Chordate ancestry of the neural crest: New insights from ascidians. *Semin Cell Dev Biol* 18:481–491.
- Jeffery WR, Strickler AG, Yamamoto Y. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* 431:696–699.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: The beginning of incongruence? *Trends Genet* 22:225–231.
- Jimenez-Guri E, Philippe H, Okamura B, Holland PWH. 2007. Buddenbrockia is a Cnidarian worm. *Science* 317:116–118.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
- Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7 Suppl 1:S4.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095–1109.
- Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci* 363:1463–1472.
- Lwoff A. 1944. *L'évolution physiologique: Étude des pertes de fonctions chez les microorganismes*. Paris: Hermann. 308 p.
- Mallatt J, Winchell CJ. 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol Phylogenet Evol* 43:1005–1022.
- Marlétaz F, Holland LZ, Laudet V, Schubert M. 2006a. Retinoic acid signaling and the evolution of chordates. *Int J Biol Sci* 2:38–47.
- Marlétaz F, Martin E, Perez Y, Papillon D, Caubit X, Lowe CJ, Freeman B, Fasano L, Dossat C, Wincker P, Weissenbach J, Le Parco Y. 2006b. Chaetognath phylogenomics: A protostome with deuterostome-like development. *Curr Biol* 16:R577–R578.
- Matus DQ, Copley RR, Dunn CW, Hejnol A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. *Curr Biol* 16:R575–R576.
- Mazet F, Shimeld SM. 2005. Molecular evidence from ascidians for the evolutionary origin of vertebrate cranial sensory placodes. *J Exp Zool B Mol Dev Evol* 304:340–346.
- Nielsen C. 2001. *Animal evolution, interrelationships of the living phyla*. Oxford, UK: Oxford University Press.
- Oda H, Akiyama-Oda Y, Zhang S. 2004. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: Distinct adhesive specificities and possible involvement in the development of multicell-layered structures. *J Cell Sci* 117:2757–2767.
- Oda H, Wada H, Tagawa K, Akiyama-Oda Y, Satoh N, Humphreys T, Zhang S, Tsukita S. 2002. A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. *Evol Dev* 4:426–434.
- Philippe H. 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res* 21:5264–5272.
- Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005a. Phylogenomics. *Annu Rev Ecol Syst* 36:541–562.
- Philippe H, Lartillot N, Brinkmann H. 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. *Mol Biol Evol* 22:1246–1253.
- Philippe H, Snell EA, Baptiste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Mol Biol Evol* 21:1740–1752.
- Philippe H, Telford MJ. 2006. Large-scale sequencing and the new animal phylogeny. *Trends Ecol Evol* 21:614–620.
- Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol* 21:1455–1458.
- Posada D, Crandall KA. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. *Syst Biol* 50:580–601.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
- R Development Core Team. 2007. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* 56:389–399.
- Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCAFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* 7 Suppl 1:S2.
- Rowe T. 2004. Chordate phylogeny and development. In: Cracraft J, Donoghue MJ, editors. *Assembling the tree of life*. Oxford: Oxford University Press. pp 384–409.
- Ruppert EE. 2005. Key characters uniting hemichordates and chordates: Homologies or homoplasies? *Can J Zool* 83:8–23.
- Schaeffer B. 1987. Deuterostome monophyly and phylogeny. *Evol Biol* 21:179–235.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
- Schubert M, Escriva H, Xavier-Neto J, Laudet V. 2006. Amphioxus and tunicates as evolutionary model systems. *Trends Ecol Evol* 21:269–277.
- Sempere LF, Cole CN, McPeck MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* 306:575–588.
- Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaot M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* 431:67–71.
- Steel M. 2005. Should phylogenetic models be trying to “fit an elephant”? *Trends Genet* 21:307–309.
- Swalla BJ, Cameron CB, Corley LS, Garey JR. 2000. Urochordates are monophyletic within the deuterostomes. *Syst Biol* 49:52–64.
- Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony and other methods version 4.0b10. Sunderland, MA: Sinauer.
- Telford MJ, Budd GE. 2003. The place of phylogeny and cladistics in Evo-Devo research. *Int J Dev Biol* 47:479–490.
- Vienne A, Pontarotti P. 2006. Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int J Biol Sci* 2:32–37.
- Wada H, Okuyama M, Satoh N, Zhang S. 2006. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol Dev* 8:370–377.
- Wada H, Satoh N. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. *Proc Natl Acad Sci USA* 91:1801–1804.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699.
- Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol* 52:528–538.
- Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst Biol* 54:731–742.
- Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *J Biomed Inform* 39:34–42.
- Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J. 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol Biol Evol* 19:762–776.

Research article

Open Access

An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models

Georgia Tsagkogeorga^{1,2}, Xavier Turon³, Russell R Hopcroft⁴, Marie-Ka Tilak^{1,2}, Tamar Feldstein⁵, Noa Shenkar^{5,6}, Yossi Loya⁵, Dorothee Huchon⁵, Emmanuel JP Douzery^{1,2} and Frédéric Delsuc*^{1,2}

Address: ¹Université Montpellier 2, Institut des Sciences de l'Evolution (UMR 5554), CC064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France, ²CNRS, Institut des Sciences de l'Evolution (UMR 5554), CC064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France, ³Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala S. Francesc 14, 17300 Blanes (Girona), Spain, ⁴Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska, USA, ⁵Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, 69978, Israel and ⁶Department of Biology, University of Washington, Seattle WA 98195, USA

Email: Georgia Tsagkogeorga - Georgia.Tsagkogeorga@univ-montp2.fr; Xavier Turon - xturon@ceab.csic.es; Russell R Hopcroft - hopcroft@ims.uaf.edu; Marie-Ka Tilak - Marie-Ka.Tilak@univ-montp2.fr; Tamar Feldstein - tfeldste@post.tau.ac.il; Noa Shenkar - nshenkar@u.washington.edu; Yossi Loya - yosiloya@post.tau.ac.il; Dorothee Huchon - huchond@post.tau.ac.il; Emmanuel JP Douzery - Emmanuel.Douzery@univ-montp2.fr; Frédéric Delsuc* - Frederic.Delsuc@univ-montp2.fr

* Corresponding author

Published: 5 August 2009

Received: 16 October 2008

BMC Evolutionary Biology 2009, 9:187 doi:10.1186/1471-2148-9-187

Accepted: 5 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2148/9/187>

© 2009 Tsagkogeorga et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tunicates have been recently revealed to be the closest living relatives of vertebrates. Yet, with more than 2500 described species, details of their evolutionary history are still obscure. From a molecular point of view, tunicate phylogenetic relationships have been mostly studied based on analyses of 18S rRNA sequences, which indicate several major clades at odds with the traditional class-level arrangements. Nonetheless, substantial uncertainty remains about the phylogenetic relationships and taxonomic status of key groups such as the Aplousobranchia, Appendicularia, and Thaliacea.

Results: Thirty new complete 18S rRNA sequences were acquired from previously unsampled tunicate species, with special focus on groups presenting high evolutionary rate. The updated 18S rRNA dataset has been aligned with respect to the constraint on homology imposed by the rRNA secondary structure. A probabilistic framework of phylogenetic reconstruction was adopted to accommodate the particular evolutionary dynamics of this ribosomal marker. Detailed Bayesian analyses were conducted under the non-parametric CAT mixture model accounting for site-specific heterogeneity of the evolutionary process, and under RNA-specific doublet models accommodating the occurrence of compensatory substitutions in stem regions. Our results support the division of tunicates into three major clades: 1) Phlebobranchia + Thaliacea + Aplousobranchia, 2) Appendicularia, and 3) Stolidobranchia, but the position of Appendicularia could not be firmly resolved. Our study additionally reveals that most Aplousobranchia evolve at extremely high rates involving changes in secondary structure of their 18S rRNA, with the exception of the family Clavelinidae, which appears to be slowly evolving. This extreme rate heterogeneity precluded resolving with certainty the exact phylogenetic placement of

Aplousobranchia. Finally, the best fitting secondary-structure and CAT-mixture models suggest a sister-group relationship between Salpida and Pyrosomatida within Thaliacea.

Conclusion: An updated phylogenetic framework for tunicates is provided based on phylogenetic analyses using the most realistic evolutionary models currently available for ribosomal molecules and an unprecedented taxonomic sampling. Detailed analyses of the 18S rRNA gene allowed a clear definition of the major tunicate groups and revealed contrasting evolutionary dynamics among major lineages. The resolving power of this gene nevertheless appears limited within the clades composed of Phlebobranchia + Thaliacea + Aplousobranchia and Pyuridae + Styelidae, which were delineated as spots of low resolution. These limitations underline the need to develop new nuclear markers in order to further resolve the phylogeny of this keystone group in chordate evolution.

Background

For more than a century, it has been known that tunicates (or urochordates) belong to chordates [1]. Traditionally occupying a basal position within chordates, they quickly became model organisms in evolutionary developmental studies aimed at understanding the origins of chordates. The complete genome sequence of the ascidian *Ciona intestinalis* revealed the essential toolkit of vertebrate genes in a small and compact genome that had not undergone the round of vertebrate-specific genome duplication events [2,3]. However, the *Ciona intestinalis* genome has also been evolving rapidly, with a number of lineage-specific innovations pointing to tunicates as a particularly interesting group for comparative genomic studies [4]. Meanwhile, recent phylogenomic analyses have refuted the classical view by demonstrating that tunicates, not cephalochordates, are the closest living relatives of vertebrates [5-9].

Despite their key position in the tree of life, our comprehension of the phylogenetic affinities within the tunicates is still limited. Traditionally, tunicates have been classified into three major classes – Ascidiacea, Appendicularia and Thaliacea – with distinct life-history traits and developmental modes. Among these classes, the Ascidiacea, commonly referred to as ascidians, is by far the most diversified group, with more than 2500 species classified on the basis of the structure of their branchial sac into three orders: Phlebobranchia, Aplousobranchia and Stolidobranchia [10-12]. Ascidians begin their lives as tadpole-like swimming larvae that later undergo a metamorphosis resulting in morphologically modified, sac-like sessile adults that can be either solitary or colonial. In contrast, the Thaliacea and Appendicularia classes consist of exclusively planktonic species. Appendicularia are peculiar in departing from the common developmental program of metamorphosis. They retain larval characteristics for their entire lifespan and are consequently of pivotal interest from an evolutionary developmental point of view [13].

Over the last decade, molecular phylogenetic analyses have been introduced to shed light on tunicate relationships. Following the pioneering study of Wada [14], the 18S rRNA gene had been the main marker used in reconstructing tunicate relationships at different taxonomic scales [15-19]. In all these studies, molecular phylogenies have contradicted the traditional classification, especially when higher taxonomic levels are considered. Molecular data refuted the division into the three classes (Ascidiacea, Thaliacea and Appendicularia), arguing instead for more complex evolutionary relationships among members of these morphologically distinct groups [20]. In overview, 18S rRNA data have clearly supported the paraphyletic nature of the class Ascidiacea by dividing the tunicates into the following three clades: Phlebobranchia + Thaliacea, Appendicularia, and Stolidobranchia, the last comprising Molgulidae and Pyuridae + Styelidae [15-19]. However, while this scheme is generally accepted, several major phylogenetic questions remain unanswered, among which at least three are prominent.

The first involves the long-standing question of the position of Appendicularia, which is crucial for understanding the evolution of body plans and developmental modes in tunicates [13]. One problem with positioning appendicularians in tunicate phylogeny is that the two 18S rRNA sequences from the genus *Oikopleura*, the only representatives currently sampled, evolved at an elevated rate [14]. This high evolutionary rate seems to be a genome-wide characteristic in *Oikopleura dioica* and leads to problematic long branches in inferred molecular phylogenies [6]. Consequently, in the majority of phylogenetic studies conducted so far, Appendicularia occupy a basal position in being the sister-group of all other tunicates [14,17,21]. This basal position, which might reflect a long-branch attraction artefact [17], has recently been challenged by analyses, including an improved taxon sampling, that instead favoured a sister-group relationship of Appendicularia with Stolidobranchia [19,20].

The second irresolution concerns the placement of the ascidian order Aplousobranchia within the tunicates. All

aplousobranch species are colonial and constituted by small zooids often embedded in a common tunic. In early 18S rRNA studies, it was proposed that Aplousobranchia form the sister-group of Appendicularia and that together they occupy the most basal position among tunicates [16]. Later on, phylogenies based on analyses of the mitochondrial cytochrome oxidase subunit I gene (*cox1*) suggested a close affinity of Aplousobranchia with Cionidae and Diazonidae (Phlebobranchia) [22], which is also supported by morphological characters [23]. Recently, it was shown that the spurious relationship of Aplousobranchia with Appendicularia [16] was an artefact due to the contamination of aplousobranch 18S rRNA sequences by protistan symbionts [18], and analyses of authentic 18S rRNA sequences clustered Aplousobranchia with Thaliacea [18]. However, the 18S rRNA sequences of Aplousobranchia seemed to evolve at extraordinarily high rates, yielding extremely long branches in the inferred phylogenetic trees [18]. This renders it particularly difficult to obtain a reliable placement of Aplousobranchia within the tunicates [19,20].

The third uncertainty pertains to the phylogenetic relationships of the thaliaceans within the ascidians. Despite the fact that rRNA studies have always placed Thaliacea and Phlebobranchia together [17-20], the exact phylogenetic position of Thaliacea remains unclear. This is mainly due to poor taxonomic sampling, with only three representatives of Thaliacea included to date in 18S rRNA phylogenies. The high evolutionary rate of 18S rRNA sequences in Thaliacea also precluded firmly resolving the phylogenetic relationships among its three constitutive orders.

Most of our knowledge of tunicate molecular phylogeny comes from analyses of 18S rRNA sequences using standard models of nucleotide sequence evolution [18,19]. Although among site-rate heterogeneity can be efficiently modelled by fitting a gamma-distribution [24] or by

assuming a proportion of invariable sites [25], these standard models make several assumptions, such as the independent evolution of nucleotide sites, and the spatial homogeneity of the substitution process across sites. In this respect, it has been demonstrated that neglecting the co-evolving paired sites in stems affects the estimation of bootstrap values [26,27] and also influences topological inference [28-30]. The spatial substitution pattern heterogeneity of sequence evolution may also produce misleading phylogenetic signals [31,32] which can severely affect phylogenetic reconstruction in cases of high substitutional saturation [33]. Our aim was thus to establish an updated phylogenetic framework for tunicate evolution based on analyses of the 18S rRNA gene, using the most realistic models that account for secondary structure and across-site heterogeneities in the evolutionary process. Phylogenetic analyses of a dataset incorporating 30 new complete 18S rRNA tunicate sequences allowed us to compare for the first time the use of RNA-specific paired-site substitution models [29,30,34] and the CAT mixture model [31], relaxing the hypothesis of a uniform substitution process by letting patterns of substitution be distinct at different sites.

Results and discussion

Improved phylogenetic models and 18S rRNA evolution

Table 1 summarises the selected models and methods used for phylogenetic inference of the two assembled datasets: a 110-taxon dataset including all tunicate sequences and a reduced 88-taxon dataset excluding the highly divergent aplousobranch species (see Material and Methods). The tree-reconstruction approaches we followed can be divided into three categories based on the type of substitution models used: (1) standard independent-site DNA models (e. g. GTR+ Γ +I); (2) doublet or paired-site substitution models (e. g. RNA6A+ Γ +I); and (3) the CAT-GTR non-parametric mixture model.

Table 1: Phylogenetic approaches and best-fitting models.

Datasets	Phylogenetic Reconstruction Approaches				
	Maximum Likelihood Standard DNA model	Standard DNA model	Bayesian Inference Doublet model		Mixture model
Complete dataset 110 taxa 1326 sites	GTR + Γ + I	GTR + Γ + I	Loops: GTR + Γ + I	Stems: RNA6C + Γ + I RNA7B + Γ RNA16A + Γ + I	CAT-GTR+ Γ
Reduced dataset 88 taxa 1650 sites	TN93 + Γ + I	GTR + Γ + I	Loops: TN93 + Γ + I	Stems: RNA6CA + Γ + I RNA7B + Γ RNA16A + Γ	CAT-GTR+ Γ

Contrary to standard models which assume that all sites evolve independently, doublet or paired-site models include a wide range of substitution models developed to overcome this assumption in rRNA molecules [35]. Ribosomal RNAs are a mosaic of two main structural motifs: single-stranded domains (loops) and double-stranded helices (stems) formed by Watson-Crick base pairing between nucleotides. During rRNA evolution, compensatory substitutions occur regularly in stems in order to maintain the counterpart-paired nucleotides. Doublet models, by considering pairs of nucleotides as states in the substitution matrices, account for this dependence.

There are three classes of doublet models – 6, 7 and 16 state models – differing in their treatment of non-Watson-Crick pairs (mismatches) [35]. For determining the most adequate model, we first selected the best-fitting standard DNA model for the unpaired loop partition, which we subsequently kept fixed in likelihood estimations, varying only the model for the stem partition [see Additional file 1]. For each of the three classes of doublet models, we used the AIC criterion to determine the best-fitting substitution model (i.e., the number of parameters of the transition matrix). However, we could not test whether the 6, 7 or 16 state models were the most appropriate because there is currently little agreement on whether it is possible to compare models with different matrix dimensions. Trees obtained under the best fitting model were almost identical among the three model classes (data not shown). We therefore only discuss those reconstructed using 6-state RNA models, which offer the best compromise between the increase in likelihood and the lowest number of free parameters incorporated.

In the third reconstruction approach, we used a mixture-model analysis to relax the assumption of a homogeneous evolutionary process across sites of the 18S rRNA molecule. The non-parametric mixture model CAT-GTR [31] was chosen for two reasons. First, previous studies using the CAT mixture model showed that it handles the substitutional saturation of the data better than classical homogeneous models [31], hence alleviating phylogenetic artefacts [33]; and second, because in contrast to other mixture models [32], it does not specify the number of mixture components *a priori*. More precisely, the CAT-GTR model assumes an infinite mixture (based on a Dirichlet process) of GTR matrices that differ only in their equilibrium base frequencies. Thus, the CAT model assigns to each site a frequency vector or so called "profile", with the number of different profiles (K) being a free parameter.

In statistical analyses, the selection of the model that best fits the data is a prerequisite. Currently, there is no generally accepted way to evaluate with confidence which of the afore-mentioned strategies (standard homogeneous

model, doublet models or CAT model) offers the best statistical fit for a given dataset. An adequate evaluation would require calculating the Bayes factor between each pair of models using thermodynamic integration [36], but such a strategy is not currently affordable computationally. Nevertheless, several studies have shown that doublet models outperform standard DNA models in phylogenetic reconstructions based on rRNA data [28,29,34,37].

With regards to the CAT-GTR mixture model versus standard site-homogeneous models, however, evaluations can be made directly from the posterior distribution of the number of profiles (K). If there is no heterogeneity across sites in the data, the CAT model will preferentially assign one single frequency vector for all sites [31]. In fact, the standard GTR model is nested within the CAT-GTR model, so that when the number of profiles is equal to one ($K = 1$), all sites evolve under a homogeneous GTR model. The distribution of the number of inferred profiles (K) for our two datasets shows that the frequency of the $K = 1$ class is zero in both cases (Figure 1). This implies that the standard GTR model has not been visited through the MCMC runs for either dataset. This means by extension that the site-heterogeneous CAT-GTR mixture model is statistically better at explaining the data than the standard homogeneous GTR model. Furthermore, the minimum number of predicted profiles is always higher than 20 (Figure 1), which suggests that 18S rRNA sequences are subject to more complex evolutionary pressures than those implied by the simple stem/loop partition.

This observation seems biologically relevant considering that beyond the pairing of nucleotides in stems, the rRNA spatial conformation is also affected by the base composition outside of these regions [38]. In this context, the maintenance of other structural motifs (T-loops, hook turns, GNRA tetraloops, bulged-G motifs, k-turns and A-minor motifs) is likely to be selectively constrained. Indeed, these motifs play an essential role in the enzymatic activity of rRNA and in its interaction with ribosomal proteins. Obviously, even if doublet models allow us to capture a major part of the constraints acting on rRNA, they are still far from modelling the overall evolutionary complexity of these molecules.

Evolutionary shifts in 18S rRNA sequences and phylogenetic relationships of Aplousobranchia

Most 18S rRNA sequences from species belonging to the divergent order Aplousobranchia are about 300 nucleotides longer than the other typical tunicate ones [18]. That is, sequences belonging to Didemnidae, Polyclinidae, and Polycitoridae, including the newly obtained *Cystodytes* sp., possess large insertions in multiple parts of the molecule when compared to the remaining tunicate

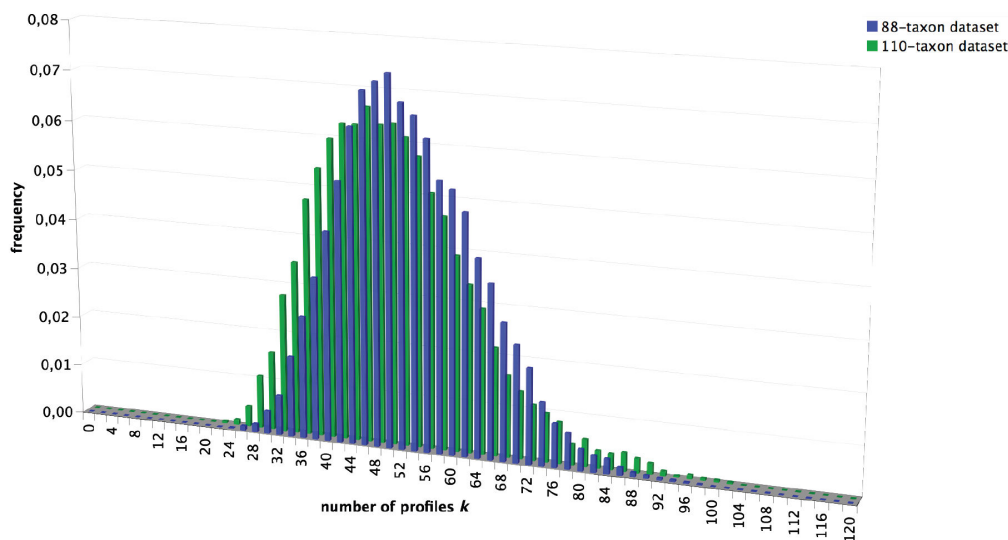


Figure 1

Estimated number of profiles K under the CAT-GTR+ Γ mixture model. The frequencies of the value (numbers of different profiles K), as estimated through the MCMC runs at the stationary stage for both the 110-taxa (green) and the 88-taxa (blue) datasets.

and outgroup species. These insertions take the form of elongations of already existing loops. More specifically, they occur in the unpaired helical regions between stems 4 and 5, and between stems 8 and 9, as well as within stems 18a-c, 29, 45a, E23_14b (Figure 2). Some of these insertions also induce changes to the secondary structure of 18S rRNA. Despite their high degree of divergence in terms of evolutionary rate and secondary structure, these unusual aplousobranch sequences are likely to be functional. Indeed, new helical structures are predicted in the elongated regions corresponding to helices 18, E23_14, 29 and 45 of the 18S rRNA molecule (Figure 2). Similar observations have also recently been reported in haplosclerid sponges [30] showing that a certain degree of freedom exists relative to the consensus 18S rRNA structure model. Knowing exactly which families belong to Aplousobranchia has been a matter of debate [23,39]. Here, we obtained sequences from two species belonging to the previously unsampled aplousobranch family Clavelinidae *sensu* Perez-Portela *et al.* [40] (*Clavelina meridionalis* and *Pycnoclavella aff. detorta*). Interestingly, these two sequences lack all the Aplousobranchia-specific structures described above, and thus show 18S rRNA sequence lengths more similar to other tunicate families.

Beside the greater gene length and the modified secondary structure, a clear compositional bias was detected for the Aplousobranchia by a Principal Component Analysis

(PCA) of base composition (Figure 3). The PCA revealed that the 18S rRNA sequences of the non-clavelinid Aplousobranchia have a different base composition than other tunicates and the outgroup species. That is, the sequences of Didemnidae, Polycitoridae and Polyclinidae are GC-rich. Conversely, Appendicularia and Molgulidae, are the most AT-rich according to the PCA results (Figure 3). The exact same compositional differences were evidenced when the loop and stem regions of the sequences were analysed separately (data not shown). Once again, the Clavelinidae sequences differ from those of other Aplousobranchia and have a base composition like that of typical tunicates and the outgroups (Figure 3).

The phylogenetic analysis of the complete tunicate dataset (110 taxa, including 95 tunicate species, and 1326 unambiguously aligned nucleotide sites: Table 1) revealed particularly long branches for the three Aplousobranchia families Polyclinidae, Polycitoridae and Didemnidae (Figure 4), as reported previously [18]. This dramatic rate acceleration was found regardless of the probabilistic method and the evolutionary model used. Furthermore, the newly obtained sequence from *Cystodytes sp.* (Polycitoridae) also had an elevated evolutionary rate and fell squarely within Aplousobranchia with high support. In striking contrast, the two sequences of the Clavelinidae evolved remarkably slowly compared to other Aplousobranchia species (Figure 4).

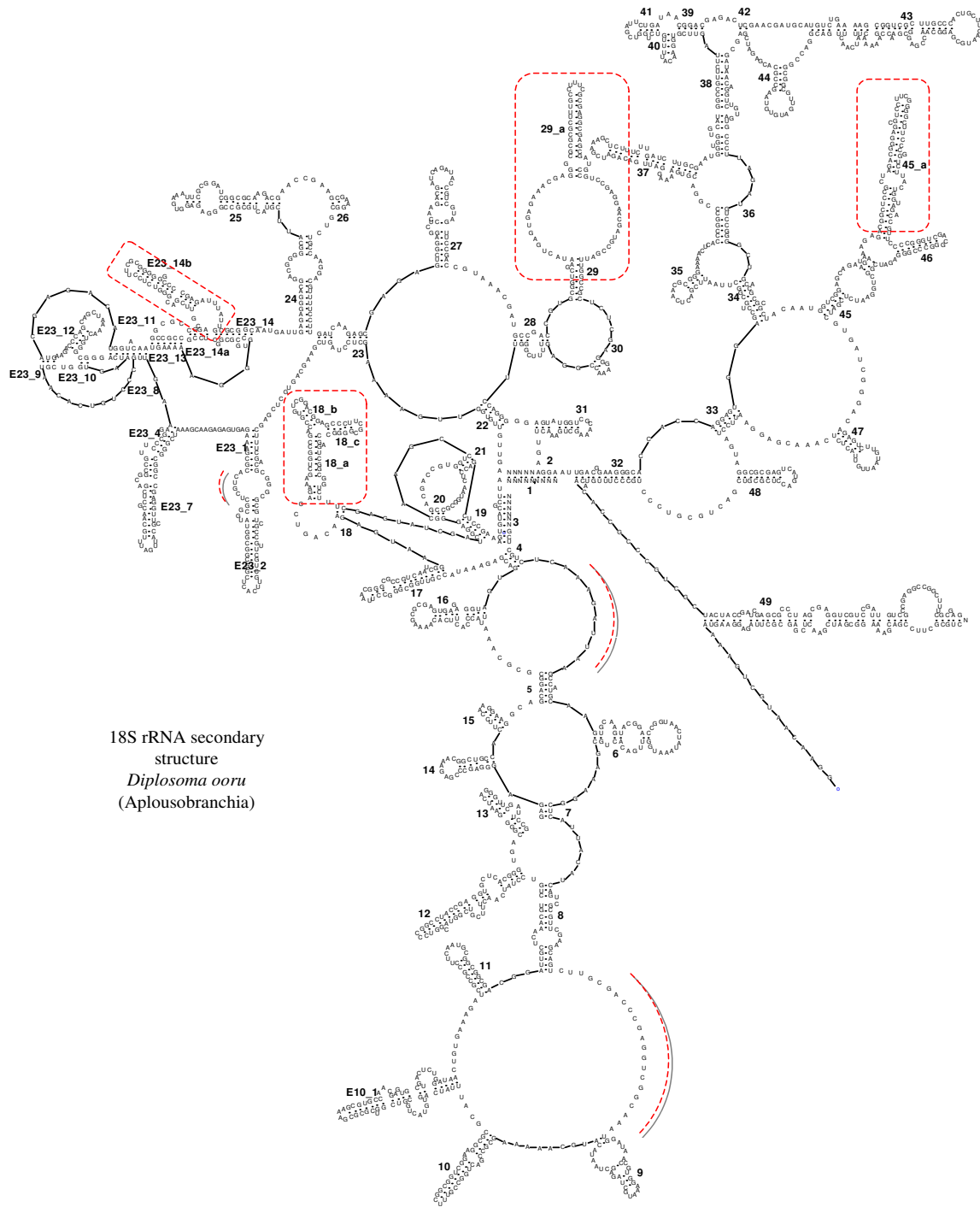


Figure 2
Predicted 18S rRNA secondary structure of a divergent aplousobranch sequence (*Diplosoma ooru*). New predicted structures unique to such divergent Aplousobranchia species (and absent in the conserved *Pycnoclavella aff. detorta* and *Clavelina meridionalis* sequences) are boxed in red. Red dotted lines indicate additional loop regions where major elongations occurred in other divergent aplousobranchs.

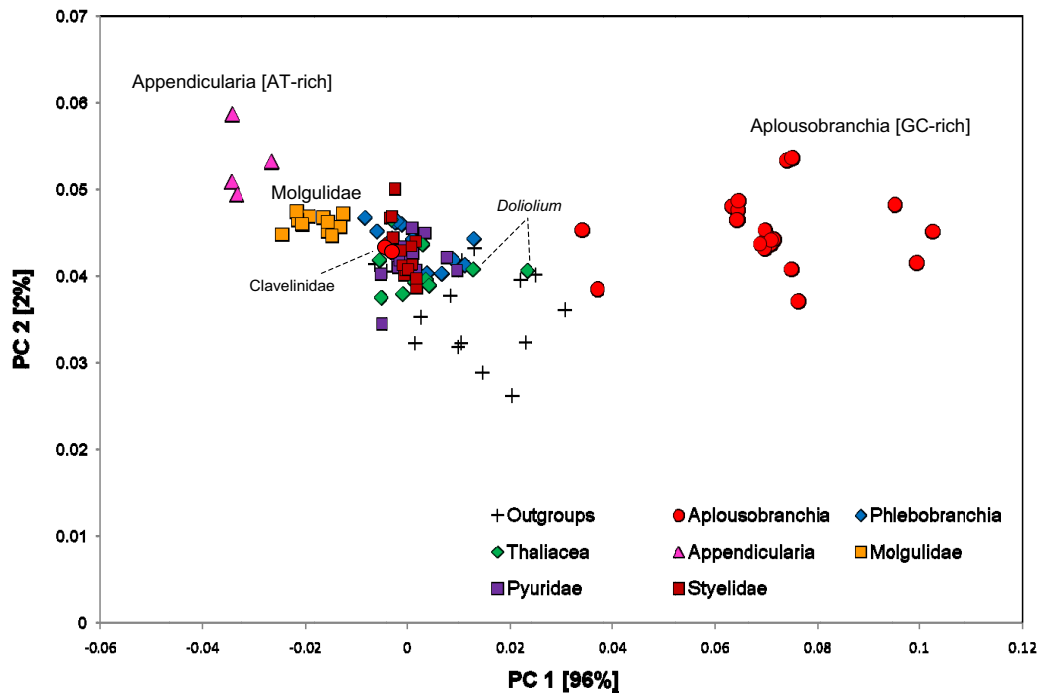


Figure 3
Analysis of base-composition heterogeneity. Principal component analysis (PCA) of the base composition of 18S rRNA from the 110-taxa dataset considering all nucleotide sites. The graph shows the first two principal components (PC), which contribute 96% and 2% of the total variance, respectively. The main component represents the variance along the AT versus GC axis, with the AT-rich Appendicularia, and the GC-rich Aplousobranchia at the two extremes.

From the phylogenetic point of view, the two slowly evolving Clavelinidae grouped together as a sister-group to the three divergent Aplousobranchia families in Bayesian analyses under the CAT mixture model and in Maximum Likelihood (ML) analyses, but with a lack of nodal support (Figure 4). This result is in agreement with what was recently proposed on the basis of the mitochondrial gene *cox1* [22]. However, monophyly of Aplousobranchia, including Clavelinidae, was not obtained in Bayesian analyses under secondary-structure models or site-homogeneous DNA models. These models instead yielded a tree topology where the divergent aplousobranch families (Polyclinidae, Polycitoridae and Didemnidae) clustered within Thaliacea to the exclusion of the Clavelinidae species. More precisely, a sister-group relationship to species of the genus *Doliolum* was obtained, but also with poor support values (data not shown).

Several arguments suggest that the previously proposed branching of Aplousobranchia within Thaliacea [18] might be a long-branch attraction (LBA) artefact [41] combined with a compositional effect. First, the topology is unstable and this grouping is not longer supported by the clavelinid sequences when the divergent aplousobranch species are removed (see Figure 5). Also, when the

Doliolum species were removed from the analyses, the Aplousobranchia affinity with the remaining Thaliacea was not recovered (data not shown). Second, *Doliolida* is the second fastest-evolving group in the Phlebobranchia + Thaliacea + Aplousobranchia clade and this could promote LBA to the fastest evolving Aplousobranchia. Third, the PCA of base composition revealed that the slightly GC-rich sequences from the *Doliolum* species are the closest in composition among tunicates to the deviant aplousobranch sequences (Figure 2), another potential cause of the observed phylogenetic artefact.

In conclusion, our results are consistent with the traditional view that the aplousobranch families Clavelinidae, Didemnidae, Polycitoridae and Polyclinidae form a monophyletic group which implies the occurrence within this order of drastic shifts in secondary structure, base composition, and evolutionary rate. Our phylogenetic analyses also provide firm support for their clustering with Phlebobranchia and Thaliacea species in a common clade. On the other hand, because of the peculiarities of 18S rRNA evolution in most species of Aplousobranchia, this gene does not allow establishing the phylogenetic relationships of this order with certitude. Also, the taxonomic sampling of 18S rRNA sequences is still insufficient

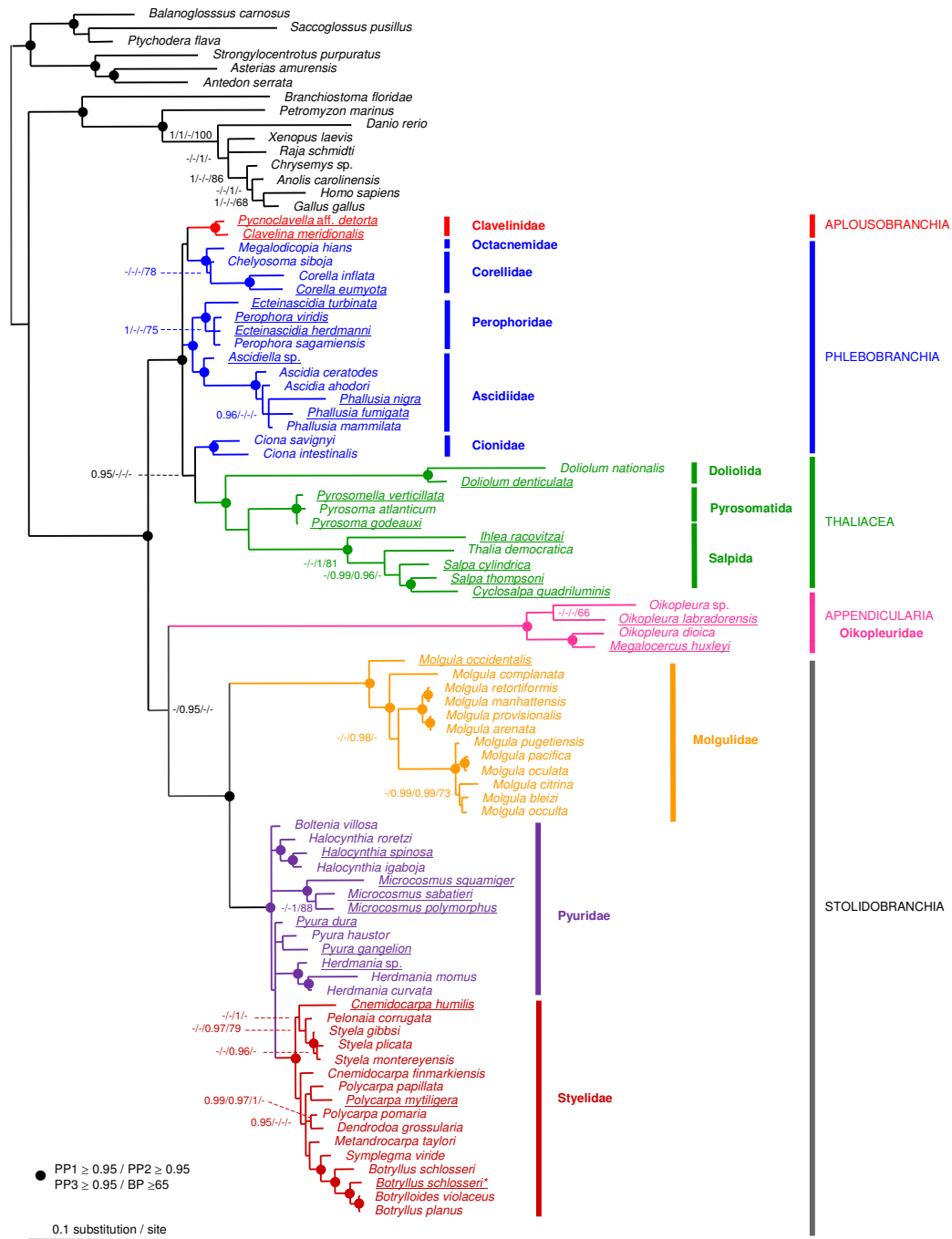


Figure 5
Phylogeny of tunicates inferred from a reduced 18S rRNA dataset (88 taxa and 1675 sites). Bayesian majority-rule consensus tree obtained under the CAT-GTR+ Γ mixture model implemented in PhyloBayes after exclusion of the fast-evolving Aplousobranchia species. Support values obtained using different reconstruction approaches are indicated at nodes in the following order: Bayesian posterior probabilities (PP) under: 1. PP1 = CAT-GTR+ Γ (PhyloBayes)/2. PP2 = RNA6A+ Γ +I and TN93+ Γ +I (Phase)/3. PP3 = GTR+ Γ +I (MrBayes)/and 4. BP = Maximum Likelihood bootstrap percentages (BP) under TN93+ Γ +I (PAUP*). Support values are displayed when PP \geq 0.95 and BP \geq 65. Dots indicate nodes for which all four reconstruction methods agree and provide PP \geq 0.95 and BP \geq 65. Newly obtained sequences are underlined, including an additional one from *Botryllus schlosseri* marked with an asterisk.

within Aplousobranchia, because the diversity of this group is huge and may comprise more than the five families here included [23,42]. Further work is needed for gaining an accurate picture of Aplousobranchia evolution within tunicates.

An updated 18S rRNA phylogenetic picture for tunicates

The growing interest in tunicates, along with the realisation that they are the sister-group of vertebrates [5-9], reinforces the importance of reconstructing a reliable phylogenetic framework for their evolutionary history. Previous molecular studies have revealed that tunicate phylogenetic relationships are difficult to resolve, mainly because of marked differences in evolutionary rates among lineages [16-19,43]. Among others, Appendicularia, Aplousobranchia, and Thaliacea have been the typical examples of rapidly evolving groups. In this respect, we aimed at revisiting tunicate 18S-rRNA phylogeny by using a wider taxonomic sampling targeted at these crucial groups, and by applying a range of phylogenetic-reconstruction approaches in a probabilistic framework. Under this general scheme, we re-evaluated previously proposed relationships and the resulting evolutionary scenarios.

I. Ascidiacea paraphyly and tunicate systematics

The different reconstruction methods applied to the complete dataset unambiguously supported the division of tunicates into the following groups: Phlebobranchia + Thaliacea + Aplousobranchia, Appendicularia and Stolidobranchia (Figure 4). Bayesian analyses provided maximal branch support (PP = 1.0) for the respective monophyly of each group, and good ML bootstrap support was obtained for the monophyly of the clade Phlebobranchia + Thaliacea + Aplousobranchia (BP = 86). A monophyletic Stolidobranchia is also retrieved with strong branch support values grouping monophyletic Molgulidae with Pyuridae + Styelidae (Figure 4). Interestingly, this dataset strongly favoured Appendicularia as the sister-group of Stolidobranchia (BP = 90) (Figure 4).

However, the inclusion of the divergent aplousobranch sequences in this 110-taxon analysis precludes a reliable examination of the phylogenetic relationships among the slower-evolving taxa. In order to obtain a reliable picture of tunicate relationships, the divergent aplousobranch sequences were removed from subsequent phylogenetic analyses, keeping only the slowly evolving Clavelinidae as representatives. This resulted in a dataset with 88 taxa, including 73 tunicates, but encompassing an expanded set of 1650 unambiguously aligned sites (Table 1). Probabilistic analyses of this dataset under standard DNA models using ML or a Bayesian approach yielded congruent results. Comparing these results to the ones obtained in Bayesian analyses using mixture and partitioned-doublet

models, some topological differences were observed (compare the different support values in Figure 5).

Concerning the relationships among the major tunicate lineages, all probabilistic analyses of this reduced dataset yielded a phylogenetic pattern similar to that for the complete data set (compare Figure 5 with Figure 4). The two ascidian orders Aplousobranchia and Phlebobranchia again grouped strongly with Thaliacea. Appendicularia was unambiguously monophyletic, as were also Molgulidae, and the grouping of Pyuridae with Styelidae. The monophyly of Stolidobranchia was also strongly recovered (Figure 5). However, in contrast with analyses of the complete dataset, the support for Appendicularia as the sister-group of Stolidobranchia was decreased (Figure 5). Indeed, only the partitioned Bayesian analysis using RNA secondary-structure models provided some support for this relationship (PP = 0.95).

Our results therefore confirm that Ascidiacea, as currently defined, is a paraphyletic group that includes Thaliacea and, most likely, Appendicularia, as previously suggested [14,16-20]. Traditionally, ascidians were subdivided morphologically in two different ways (or combinations thereof): 1) based on the structure of the branchial sac, dividing ascidians into Aplousobranchia, Phlebobranchia and Stolidobranchia, as established by Lahille [10-12]; and 2) based on the position of the gonads, which separates ascidians into Enterogona and Pleurogona, following Perrier [44] as modified by Garstang [45]. Enterogona includes the Aplousobranchia and Phlebobranchia, while Pleurogona consists of Stolidobranchia (Figures 4 and 5). The finding that 18S rRNA joins Phlebobranchia and Aplousobranchia into a common clade supports the classification scheme of Enterogona and Pleurogona. Yet, accepting this view, Thaliacea branched in our trees within Enterogona, and Appendicularia with Pleurogona. The gonad position in the thaliaceans indeed conforms to the enterogonid type (associated with the gut), while in the appendicularians the highly modified metamorphosis renders any inference about the gonad position difficult.

II. Phlebobranchia and the sister-group Cionidae

Within Phlebobranchia, the five represented families – Ascidiidae, Perophoridae, Octacnemidae, Corellidae and Cionidae – appeared reciprocally monophyletic with high support values, except for Corellidae (Figure 5). At the inter-familial level, strong support was obtained for grouping Octacnemidae with Corellidae, and Perophoridae with Ascidiidae within which *Ascidia* appears divergent from the closely related genera *Phallusia* and *Ascidia* (Figure 5). The inclusion of Cionidae within Aplousobranchia instead of in its traditional position within Phlebobranchia has been suggested on the basis of both morphological characters [23] and analyses of *cox1* [22].

However, this grouping was not recovered in any of our trees, which are consistent with the traditional phlebobranch position of Cionidae. Our results support the paraphyly of Phlebobranchia, with a close phylogenetic affinity between the *Ciona* species and Thaliacea in Bayesian analyses using the CAT-GTR mixture model (PP = 0.95). Although not supported by more conventional reconstruction methods (Figure 5), such a relationship provides a good working hypothesis for future phylogenomic studies aiming at determining the sister-group of model species from the *Ciona* genus.

III. Thaliacea relationships

The Thaliacea are a class of exclusively planktonic tunicates and consist of three main orders: Doliolida, Pyrosomatida and Salpida. They are distinct from the other tunicates, because all members of this class form specialized colonies and have complex life cycles, characterized by the alternation between sexual and asexual generations. There were two traditional views on the origin of the class: the first suggesting that Thaliacea derived from an Appendicularia-like ancestor (Herdman 1888; Seeliger 1893 – 1911; Newman 1933 cited in [46]), and the second deriving Thaliacea from an ascidian-like ancestor [47–49] or even placing the split more basally, with the ascidians and thaliaceans evolving from a common ancestor [45]. In parallel to this controversy, there has been considerable debate on whether the thaliaceans are monophyletic or not [46]. Within Thaliacea, however, the commonly accepted view places the pyrosomatids in a basal position, with the doliolids and salps being sister-groups, following a trend toward branchial sac simplification and muscle-band development [47].

Meanwhile, early molecular phylogenies supported a close relationship between Thaliacea and Phlebobranchia thereby suggesting Thaliacea arose from a sessile ascidian-like ancestor [14]. Then, Thaliacea was also reported to be monophyletic in more recent studies [16,19,20]. On the phylogenetic affinities among its three constitutive orders, the phylogenetic evidence appears more ambiguous. In fact, in half of the previous 18S-rRNA tunicate phylogenies, Doliolida were found to be the sister-group of Salpida [14,19], whereas in others a closer relationship was recovered between Pyrosomatida and Salpida to the exclusion of Doliolida [16,17].

All previous sampling of Thaliacea was limited to the species *Pyrosoma atlanticum*, *Doliolum nationalis*, and *Thalia democratica*, among which the latter two were characterized by relatively long branches. This evolutionary-rate heterogeneity is probably the reason for the reported conflicting relationships among the three orders. In our study, we subdivided these long branches by tripling the taxonomic sampling which rendered the strong statistical sup-

port for the clustering of Thaliacea with Phlebobranchia and Aplousobranchia even more reliable (Figure 5). Although we still lack some representative families, the monophyletic nature of Thaliacea and the respective monophyly of its three constitutive orders were also firmly supported in all reconstructed trees (Figure 5). This is in agreement with previous taxonomic views (e.g., [47,48]).

Regarding thaliacean interrelationships, our phylogeny suggests a sister-group relationship between Pyrosomatida and Salpida, with Doliolida species arising first (Figure 5). Such a relationship was obtained in Bayesian analyses using mixture (PP = 0.69) and secondary structure (PP = 0.56) models whereas analyses conducted under standard DNA models (PP = 0.69; BP = 57) tend to group the two fast-evolving groups Salpida and Doliolida as reported in the most recent study [19]. These contrasting results suggested a possible effect of model misspecification leading to a potential LBA artefact grouping the fast evolving Salpida and Doliolida within Thaliacea in analyses conducted under standard DNA models. If confirmed, this particular case probably represents an example where the use of the most complex and best fitting models of sequence evolution allows improving phylogenetic inference [29,33].

Our proposed phylogenetic hypothesis seems at first counterintuitive since both doliolids and salps have solitary generations, in contrast to pyrosomes, which are exclusively colonial and are generally considered as the least specialized thaliaceans. However, the notion of a basal Pyrosomatida from which Doliolida and Salpida derive is at odds with the fact that only Doliolida in this group retain the larval features of tail and notochord. In contrast, our phylogenetic picture implies a single event of tail loss within the thaliacean lineage. In agreement with this view of a basal Doliolida, comparative studies of sperm morphology have shown that doliolid sperm is more plesiomorphic than that of the pyrosomes or salps [46,50]. Likewise, the simple arrangement of gill slits and the particular morphogenesis and structure of stolons in pyrosomes and salps may represent synapomorphic characters of the two groups [51].

IV. The phylogenetic position of Appendicularia

Appendicularians or Larvaceans are small planktonic tunicates that have short generation times and characteristic tadpole body plans, consisting of a short trunk and a posterior tail with a notochord and a dorsal neural tube. From a developmental point of view they are unique among tunicates, since all species belonging to this class retain the tail and the notochord in adulthood [52]. For this reason, the relationships of Appendicularia to the other tunicate lineages have always been considered as

crucial for understanding the evolution of chordate body plans [14,16]. The traditional classification distinguishes three families within the class Appendicularia: Fritillariidae, Kowalevskiidae and Oikopleuridae [53]. However, all previous molecular phylogenies have been taxonomically restrained to species from the model genus *Oikopleura* (Oikopleuridae). While early studies of 18S rRNA indicated that Appendicularia emerged first among tunicates [14,17], more recent analyses suggested a sister-group relationship with Stolidobranchia [19]. Their phylogenetic position remains controversial mainly because *Oikopleura* 18S rRNA sequences are highly evolving [17,43].

Our results, including two additional species belonging to Oikopleuridae (*Oikopleura labradorensis* and *Megalocercus huxleyi*), did not seem to break up the long appendicularian ancestral branch enough to firmly establish the position of this order within tunicates. Indeed, whereas the complete dataset strongly favoured Appendicularia as the sister-group of Stolidobranchia (Figure 4), in the reduced but less heterogeneous dataset (Figure 5), only the partitioned Bayesian analysis using RNA-secondary structure models provided support for this relationship (PP = 0.95). These results do not support the basal emergence of Appendicularia [14,16,17], suggesting instead their placement as a sister-group to Stolidobranchia, as recently proposed [19]. The most parsimonious interpretation of this internal position would imply that the planktonic Appendicularia derived from a sessile ascidian-like ancestor.

Still, a reliable placement for Appendicularia based on 18S rRNA data is hampered by two potentially confounding factors: the persisting long branches characterizing the group, and the similarity of nucleotide composition in Appendicularia and Molgulidae (Figure 2). In both, 18S rRNA sequences are AT-rich and experienced accelerated evolutionary rates, leaving open the possibility of an artefactual attraction of the two groups due to systematic biases. Finally, phylogenomic studies, although based on just a few taxa, strongly support the basal position of Appendicularia within tunicates [6,7]. However, *Oikopleura dioica*, the sole member of the class with currently available genomic data, evolves at such an elevated rate that any conclusion concerning its exact phylogenetic position based on these data would be premature.

V. Stolidobranchia

Stolidobranchia appears as unambiguously monophyletic in all phylogenetic reconstructions performed (Figures 4 and 5). Within Stolidobranchia, Molgulidae formed a monophyletic group characterized by accelerated evolutionary rates compare to other stolidobranchs (Figure 5). The basal emergence of *Molgula occidentalis* was significantly supported (0.97/1.0/1.0/88), followed by *Molgula*

complanata as the sister-group of two reciprocally monophyletic groups that include other molgulids, in good agreement with a previous study of the group based on partial sequences [15].

Overall, phylogenetic affinities within Molgulidae were rather well resolved and stable, whereas Pyuridae, and especially, Styelidae showed greater topological instability (Figure 5). Despite the addition of 10 species belonging to both Pyuridae and Styelidae, phylogenetic relationships among and within the two families were crippled by low resolution, with two multifurcations occurring within Pyuridae and poor support values at several nodes within this clade (Figure 5). In our trees, the family Styelidae is monophyletic whereas Pyuridae appears paraphyletic (Figure 5). These results contrast with those from a recent study of the stolidobranchs based on partial 18S rRNA and *cox1* sequences where the reverse situation was observed [54]. This incongruence might stand in a rooting problem of the Pyuridae-Styelidae group whereby its fast-evolving, and distant molgulid sister-group, created LBA effects [54]. In fact, only the relationships within Botryllinae were well-supported by all reconstruction methods. Furthermore, the inclusion of Botryllinae within Styelidae was strongly supported, upholding the view that the latter is in fact a subfamily of Styelidae [47,55]. Yet, little evidence is provided for the exact position of the Botryllinae, because its proposed sister-group relation to the colonial *Metandrocarpa taylori* [19] was not significantly supported (Figure 5).

Conclusion

Our work has improved tunicate phylogeny based on 18S rRNA, in order to provide a reliable phylogenetic framework for the evolution of this key group of chordates. This was accomplished by a wider taxonomic sampling than in previous studies, yielding a more thorough representation of major tunicate lineages, and subdividing previously long branches. Our results showed that the non-parametric CAT-GTR mixture model has a better fit than standard DNA models on tunicate 18S rRNA data. The high number of site profiles inferred under the CAT-GTR model also suggested that the *a priori* partition of 18S rRNA data into stem and loop regions is likely an oversimplification of the complex evolutionary constraints acting on tunicate 18S rRNA sequences.

The 18S rRNA gene provided a clear view of the evolution of major tunicate lineages, but appeared less informative for relationships at lower taxonomic levels. This was reflected by the weak support values accorded to several nodes in the reconstructed trees, and by the unstable position of certain groups between phylogenies inferred using different reconstruction approaches. Among the spots of low resolution, the relationships at inter- and intra-famil-

ial levels within Phlebobranchia and within the Pyuridae – Styelidae group proved difficult to clarify. The development of new phylogenetic markers is thus necessary for the further comprehension of tunicate evolutionary history. An obvious candidate is the 28S rRNA gene, which is already available for some species and whose resolving power in combination with the 18S rRNA has been widely demonstrated [43,56]. Finally, the increasing availability of genomic information for tunicates also constitutes a promising source of future nuclear protein-coding markers.

Methods

Taxon sampling

The tissue samples used in this study were collected from the following species: class Ascidiacea: Phlebobranchia: *Perophora viridis*, *Ecteinascidia herdmanni*, *Ecteinascidia turbinata*, *Asciidiella* sp., *Corella eumyota*, *Phallusia nigra*, *Phallusia fumigata*; Aplousobranchia: *Pycnoclavella* aff. *detorta*, *Clavelina meridionalis*, *Cystodytes* sp.; Stolidobranchia: *Molgula occidentalis*, *Halocynthia spinosa*, *Pyura dura*, *Pyura gangelion*, *Herdmania* sp., *Microcosmus squamiger*, *Microcosmus sabatieri*, *Microcosmus polymorphus*, *Polycarpa mytiligera*, *Botryllus schlosseri*, *Cnemidocarpa humilis*, class Thaliacea: *Doliolum denticulata*, *Pyrosomella verticillata*, *Pyrosoma godeauxi*, *Ihlea racovitzai*, *Salpa cylindrica*, *Salpa thompsoni*, *Cyclosalpa* sp.; and class Appendicularia: *Oikopleura labradorensis*, *Megalocercus huxleyi*.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from 95% ethanol-preserved tissue samples using either the QIAamp DNA or the DNeasy Plant Mini kits following the manufacturer's protocols or following the procedure of Bernatzky and Tanksley [57]. The 18S rRNA gene was PCR-amplified either in one fragment (using the primer sets 18S1/18S2 or 18S1/18S_Herdm_R1) or in two overlapping fragments of approximately 1 kb, 18S1/18S4 and 18S3/18S2, using the following primers: 18S1 (Fwd) 5'-CCTGGTTGATCCTGCCAG-3', 18S2 (Rev) 5'-TAATGATCCATCTGCAGG-3', 18S3 (Fwd) 5'-TTAGAGTGTTCAAAGCAGGC-3', 18S4 (Rev) 5'-GATTAAGAAAACATTCTTGCC-3' and the newly designed 18S_Herdm_R1 (Rev) 5'-GATTRACCCGAGACCGCMATTYGCRIT-3'. PCR products were purified from 1.2% agarose gel using Gel Extraction Kit (Millipore) or using polyethylene glycol (PEG) in saline (NaCl). Most of the products were directly sequenced using Big Dye Terminator v1.1 (Applied Biosystems) on an ABI 310 sequencer. Some PCR products were ligated into pGEM-T Easy vector for cloning into One Shot TOP10 Competent Cells (Invitrogen), and five clones per species were sequenced. All chromatograms were manually corrected and assembled using the software Sequencher. The 30 new sequences have been deposited

in the EMBL database under Accession Numbers [FM244840](#) to [FM244869](#).

Data assembly and alignments

In addition to the 30 sequences obtained for this study, more tunicate 18S rRNA sequences with > 85% length coverage were recovered from GenBank [see Additional file 2]. Fifteen outgroups were chosen to evenly sample the diversity of the other phyla of Deuterostomes, with representative species belonging to Vertebrata and Cephalochordata (Chordata), and Hemichordata and Echinodermata (Ambulacraria). In all analyses, the outgroup was used for rooting trees *a posteriori*. Primary multiple alignments were performed using MAFFT [58] and were subsequently adjusted by eye. Following secondary structure models for the 18S rRNA molecule, available in the European Ribosomal Database [59], two partitions or character groups were assigned to the sequences: (1) paired *stem* – characters forming helices in the secondary structure and (2) unpaired *loop* – characters forming single strands. Further alignment adjustments were made manually and by using the script Xstem [29] in order to ensure that all sites in predicted helices form AU, GC Watson-Crick or GU wobble bonds with their partner nucleotide. Finally, ambiguously aligned sites as well as sites including gaps in more than 50% of the sequences were excluded using Gblocks [60] set to the following parameters: Minimum Number Of Sequences For A Conserved Position = $n/2 + 1$ (where n = number of taxa), Minimum Number Of Sequences For A Flanking Position $\approx 0.85 \times n$, Maximum Number Of Contiguous Nonconserved Positions = 8, Minimum Length Of A Block = 2 or 5, Allowed Gap Positions = With Half.

Two final datasets were assembled. The first one maximised taxonomic representation by including the Aplousobranchia sequences obtained by Yokobori *et al.* [18] which yielded a matrix of 110 taxa (95 Tunicates + 15 outgroups) and 1373 nucleotide sites. The inclusion of the divergent Aplousobranchia sequences had hindered the alignment of the 18S rRNA regions corresponding to helices E23_1-E23_1', E23_2-E23_2', E23_7-E23_7' and 49-49'. So, the corresponding sites were removed from subsequent phylogenetic analyses. In the second dataset, in order to reduce potential reconstruction artefacts associated with divergent sequences due to long-branch attraction effects [41], the fast-evolving Aplousobranchia sequences were removed. This second dataset had 88 taxa (73 Tunicates + 15 outgroups) and 1675 nucleotide sites.

Secondary structure prediction and visualisation

The 18S rRNA consensus secondary structure was obtained by following eukaryotic models available in the European Ribosomal RNA Database [59]. Potential structures for specific aplousobranch insertions were predicted

using the RNAfold Web Server [61] by choosing the minimum free energy algorithm and the option to avoid isolated base pairs. The predicted secondary structures were visualised using the program RNAviz [62].

Phylogenetic analyses

Phylogenetic analyses were conducted using Maximum Likelihood (ML) and Bayesian Inference (BI) approaches. In analyses using standard DNA models of sequence evolution, for the entire dataset as well as for the unpaired *loop* partition, the best fitting model was selected using MODELTEST for ML [63] and MrMODELTEST for BI [64] based on the Akaike Information Criterion (AIC) [65]. The best-fitting RNA-specific model for the paired *stem* partition was selected by calculating the AIC from log-likelihoods values previously estimated using the program Optimizer of the PHASE 2.0 package [66]. The selected models for each dataset and for the two different phylogenetic-reconstruction strategies are shown in Table 1.

All ML analyses were performed using PAUP* 4.0b10 [67] using a three round successive-approximation approach for estimating model parameters [68]. Starting from a neighbour-joining (NJ) tree, model parameters were estimated under the likelihood criterion and further kept fixed for heuristic searches with Tree Bisection Reconnection (TBR) branch swapping. The ML tree topology was then kept fixed and model parameters were re-estimated, with the whole process being repeated twice. Statistical support for the nodes was obtained by Bootstrap resampling with 100 pseudo-replicates generated by the program SeqBoot 3.5 of the PHYLIP package [69]. In all replicates, ML analyses were conducted in parallel on a computing cluster using PAUP* through the same heuristic search approach, with model parameters fixed at values estimated previously. Bootstrap Percentages (BP) were obtained from the 50% majority rule consensus of the 100 reconstructed trees using the program TREEFINDER [70]. In all ML analyses, the 6 Ambulacriaria species were declared as outgroup taxa.

Bayesian analyses were conducted using the programs MRBAYES 3.1.2 [71], mcmcphase from the PHASE package [66] and PHYLOBAYES 2.3 <http://www.atgc-montpellier.fr/phylobayes/> [31]. Although both MRBAYES and mcmcphase permit partitioned analysis using RNA secondary structure models, a more complete list of doublet models is implemented in PHASE. Thus in the present study, MRBAYES was principally used for conducting Bayesian phylogenetic reconstructions under a single DNA model of sequence evolution, while mcmcphase was used when both DNA and RNA secondary structure models were considered. Finally, PHYLOBAYES was used for

reconstructions under the mixture model CAT, allowing for a general substitution process (CAT-GTR) [31].

In BI analyses conducted with MRBAYES, two independent runs of four incrementally heated Metropolis-coupled Markov chains Monte Carlo (MCMCMC) were launched and run for 5,000,000 generations. In BI analyses conducted with mcmcphase, a single MCMC was launched for 10,000,000 generations. In both cases, parameters and trees were sampled every 100 generations. In PHYLOBAYES, two independent MCMCs were launched for 16,000 to 20,000 cycles with parameters and trees being sampled every cycle (20,000 cycles correspond to about 1,500,000 generations).

In all BI analyses, priors were set to default values and the convergence of chains was double checked. First, values of the marginal likelihood and independent run discrepancies were monitored through generations. Second, the posterior probabilities of all splits were plotted at 20 cycle increments over a run, using the AWTY system [72] [see Additional file 3]. The burnin value was determined when both indicators entered a stationary phase. This typically involved eliminating about 25% of the run samples. Bayesian Posterior Probabilities (PP) were obtained from the 50% majority-rule consensus of the trees sampled during the stationary phase.

Authors' contributions

FD, XT and EJPD conceived and initiated the study. XT, RRH, YL and NS collected samples and identified specimens. GT, MKT and TF carried out PCR amplifications and sequencing. EJPD, FD, DH and YL supervised the work in their respective labs. GT and FD performed the phylogenetic analyses and drafted the manuscript. All other authors assisted in revising the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Secondary structure model selection. The table presents the selection of the best-fitting model of rRNA sequence evolution for the 88-taxon and 110-taxon datasets based on the Akaike Information Criterion (AIC). The best-fitting models within each class (6, 7, and 16 states) of doublet models are bold-faced.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S1.pdf>]

Additional file 2

Species sampling, taxonomy and sequence accession numbers. The table indicate the taxonomy and the species sampling used in the present study with associated sequence Accession Numbers. The 30 new sequences obtained in this study are indicated with a star ().*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S2.pdf>]

Additional file 3

Monitoring the convergence of MCMC in Bayesian analyses. The figure illustrates the post-analysis of chain convergence in Bayesian analyses under the CAT-GTR+ Γ_4 model for the 88-taxon dataset, using the AWTY system (Nylander et al. 2007). A. Cumulative plot of clade posterior probabilities of the 20 more variable splits over a run of 20,000 cycles (i.e. 1,500,000 MCMC generations) sampled at every cycle. The vertical red line indicates the determined burn-in value of 5,000. B. Comparisons of clade posterior probabilities between the two independent MCMC runs.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S3.pdf>]

Acknowledgements

We would like to thank: Corinna Dubischar, Camille Martinand-Mari and Gabriel Gorsky for their help in obtaining precious tissue samples; Nicolas Galtier and Hervé Philippe for their helpful comments on this work; and Naomi Paz for English editing. We especially thank Nicolas Lartillot for his guidance on the use of the CAT model for this work. We also like to thank four anonymous reviewers for helpful comments and detailed suggestions that greatly improved the manuscript. This study is part of a Research Networks Program in Bioinformatics funded by the High Council for Scientific and Technological Cooperation between France and Israel (Grant # 3-3449 to DH, EJP and YL). The present work was carried out as part of the Galathea 3 expedition under the auspices of the Danish Expedition Foundation (Dansk Ekspeditionsfond). This is Galathea 3 contribution no. P42 and also contributes to the Census of Marine Zooplankton. This work benefited from the bioinformatics cluster of the Institut des Sciences de l'Evolution de Montpellier. This is contribution ISEM 2009-88 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 – CNRS).

References

- Kowalevski A: **Beiträge zur Entwicklungsgeschichte der Tunicaten.** *Nachrichten Gesellschaft Wissenschaften Göttingen* 1868, **19**:401-415.
- Cañestro C, Bassham S, Postlethwait JH: **Seeing chordate evolution through the *Ciona* genome sequence.** *Genome Biol* 2003, **4**:208.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, et al.: **The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins.** *Science* 2002, **298**:2157-2167.
- Holland LZ, Gibson-Brown JJ: **The *Ciona intestinalis* genome: when the constraints are off.** *Bioessays* 2003, **25**:529-532.
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, et al.: **Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida.** *Nature* 2006, **444**:85-88.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**:965-968.
- Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H: **Additional molecular support for the new chordate phylogeny.** *Genesis* 2008, **46**:592-604.
- Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al.: **Broad phylogenomic sampling improves resolution of the animal tree of life.** *Nature* 2008, **452**:745-749.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al.: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064-1071.
- Lahille MF: **Sur la classification des Tuniciers.** *C R Acad Sci Paris* 1886, **102**:1573-1575.
- Lahille MF: **Etude systématique des Tuniciers.** *C R Assoc Fr Adv Sci* 1887, **16**:667-677.
- Lahille MF: **Recherche sur les Tuniciers.** Paris: Faculté des Sciences de Paris; 1890.
- Nishino A, Satoh N: **The simple tail of chordates: phylogenetic significance of appendicularians.** *Genesis* 2001, **29**:36-45.
- Wada H: **Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny.** *Mol Biol Evol* 1998, **15**:1189-1194.
- Huber JL, da Silva KB, Bates WR, Swalla BJ: **The evolution of anural larvae in molgulid ascidians.** *Semin Cell Dev Biol* 2000, **11**:419-426.
- Stach T, Turbeville JM: **Phylogeny of Tunicata inferred from molecular and morphological characters.** *Mol Phylogenet Evol* 2002, **25**:408-428.
- Swalla BJ, Cameron CB, Corley LS, Garey JR: **Urochordates are monophyletic within the deuterostomes.** *Syst Biol* 2000, **49**:52-64.
- Yokobori S, Kurabayashi A, Neilan BA, Maruyama T, Hirose E: **Multiple origins of the ascidian-Prochloron symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences.** *Mol Phylogenet Evol* 2006, **40**:8-19.
- Zeng L, Jacobs M, Swalla B: **Coloniality and sociality has evolved once in Stolidobranch ascidians.** *Integr Comp Biol* 2006, **46**:255-268.
- Zeng L, Swalla B: **Molecular phylogeny of the protochordates: chordate evolution.** *Can J Zool* 2005, **83**:24-33.
- Kurabayashi A, Okuyama M, Ogawa M, Takeuchi A, Jing Z, Naganuma T, Saito Y: **Phylogenetic position of a deep-sea ascidian, *Megalodicopia hians*, inferred from the molecular data.** *Zool Sci* 2003, **20**:1243-1247.
- Turon X, Lopez-Legentil S: **Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata.** *Mol Phylogenet Evol* 2004, **33**:309-320.
- Kott P: **The Australian Ascidiacea. Part 2, Aplousobranchiata (1).** *Mem Qld Mus* 1990, **29**:1-299.
- Yang Z: **Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods.** *J Mol Evol* 1994, **39**:306-314.
- Lockhart PJ, Larkum AVW, Steel M, Waddell PJ, Penny D: **Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis.** *Proc Natl Acad Sci USA* 1996, **93**:1930-1934.
- Dixon MT, Hillis DM: **Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis.** *Mol Biol Evol* 1993, **10**:256-267.
- Galtier N: **Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites.** *Syst Biol* 2004, **53**:38-46.
- Jow H, Hudelot C, Rattray M, Higgs PG: **Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution.** *Mol Biol Evol* 2002, **19**:1591-1601.
- Telford MJ, Wise MJ, Gowri-Shankar V: **Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria.** *Mol Biol Evol* 2005, **22**:1129-1136.
- Voigt O, Erpenbeck D, Worheide G: **Molecular evolution of rDNA in early diverging Metazoa: first comparative analysis and phylogenetic application of complete SSU rRNA secondary structures in Porifera.** *BMC Evol Biol* 2008, **8**:69.
- Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.** *Mol Biol Evol* 2004, **21**:1095-1109.
- Pagel M, Meade A: **A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data.** *Syst Biol* 2004, **53**:571-581.

33. Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model.** *BMC Evol Biol* 2007, **7**(Suppl 1):S4.
34. Erpenbeck D, Nichols SA, Voigt O, Dohrmann M, Degnan BM, Hooper JN, Worheide G: **Phylogenetic analyses under secondary structure-specific substitution models outperform traditional approaches: case studies with diploblast LSU.** *J Mol Evol* 2007, **64**:543-557.
35. Savill NJ, Hoyle DC, Higgs PG: **RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods.** *Genetics* 2001, **157**:399-411.
36. Lartillot N, Philippe H: **Computing Bayes factors using thermodynamic integration.** *Syst Biol* 2006, **55**:195-207.
37. Dohrmann M, Janussen D, Reitner J, Collins AG, Worheide G: **Phylogeny and evolution of glass sponges (porifera, hexactinellida).** *Syst Biol* 2008, **57**:388-405.
38. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Rev Genet* 2007, **8**:921-931.
39. Monniot C, Monniot F, Laboute P: **Coral Reef Ascidiaceans of New Caledonia.** Paris: Orstom; 1991.
40. Perez-Portela R, Turon X: **Phylogenetic relationships of the Clavelinidae and Pycnoclavellidae (Ascidiacea) inferred from mtDNA data.** *Invertebr Biol* 2008, **127**:108-120.
41. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Syst Zool* 1978, **27**:401-410.
42. Moreno TR, Rocha RM: **Phylogeny of the Aplousobranchiata (Tunicata: Ascidiacea).** *Revista Brasileira de Zoologia* 2008, **25**:269-298.
43. Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J: **Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data.** *Mol Biol Evol* 2002, **19**:762-776.
44. Perrier E: **Note sur la Classification des Tuniciers.** *C R Acad Sci Paris* 1898, **124**:1758-1762.
45. Garstang W: **The morphology of the Tunicata, and its bearings on the phylogeny of the Chordata.** *Quart J Microsc Sci* 1928, **72**:51-187.
46. Holland LZ: **Fine structure of spermatids and sperm of *Doliolletta gegenbauri* and *Doliolum nationalis* (Tunicata: Thaliacea): implications for tunicate phylogeny.** *J Mar Biol* 1989, **110**:83-95.
47. Berrill NJ: **The Tunicata. With an Account of the British Species.** London: The Ray Society; 1950.
48. Brien P: **Embranchement des Tuniciers.** In *Traité de Zoologie, Tome XI Echinodermes, Stomocordés, Procordés* Edited by: Grassé PP. Paris: Masson; 1948.
49. Ujjanin BN: **Die Arten der Gattung *Doliolum* in Golfe von Neapel und den angrenzenden Meeresabschnitten.** Leipzig: Verlag von Wilhelm Engelmann; 1884.
50. Holland LZ: **Spermatogenesis in *Pyrosoma atlanticum* (Tunicata: Thaliacea: Pyrosomatida): Implications for tunicate phylogeny.** *J Mar Biol* 1990, **105**:451-470.
51. Lacalli TC: **Tunicate tails, stolons, and the origin of the vertebrate trunk.** *Biol Rev Camb Philos Soc* 1999, **74**:177-198.
52. Stach T, Winter J, Bouquet JM, Chourrout D, Schnabel R: **Embryology of a planktonic tunicate reveals traces of sessility.** *Proc Natl Acad Sci USA* 2008, **105**:7229-7234.
53. Fenaux R: **The classification of Appendicularia.** In *The biology of tunicates* Edited by: Bone Q. Oxford: Oxford University Press; 1998:295-306.
54. Perez-Portela R, Bishop JD, Davis AR, Turon X: **Phylogeny of the families Pyuridae and Styelidae (Stolidobranchiata, Ascidiacea) inferred from mitochondrial and nuclear DNA sequences.** *Mol Phylogenet Evol* 2009, **50**:560-570.
55. Kott P: **The Australian Ascidiacea. Part I, Phlebobranchia and Stolidobranchia.** *Mem Qld Mus* 1985, **23**:1-440.
56. Mallatt J, Winchell CJ: **Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star.** *Mol Phylogenet Evol* 2007, **43**:1005-1022.
57. Bernatzky R, Tanksley SD: **Genetics of actin-related sequences.** *Theor Appl Genet* 1986, **72**:314-321.
58. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment.** *Nucleic Acids Res* 2005, **33**:511-518.
59. Wuyts J, Perriere G, Peer Y Van De: **The European ribosomal RNA database.** *Nucleic Acids Res* 2004, **32**:D101-103.
60. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17**:540-552.
61. Gruber AR, Lorenz R, Bernhart SH, Neubeck R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**:W70-74.
62. De Rijk P, Wuyts J, De Wachter R: **RnaViz 2: an improved representation of RNA secondary structure.** *Bioinformatics* 2003, **19**:299-300.
63. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
64. Nylander JAA: **MrModeltest v2.3.** Evolutionary Biology Centre, Uppsala University.: Program distributed by the author; 2004.
65. Akaike H: **A new look at the statistical model identification.** *IEEE Trans Autom Contr AC* 1974, **19**:716-723.
66. Jow H, Gowri-Shankar V: **PHASE: a Software Package for Phylogenetics and Sequence Evolution.** University of Manchester, Manchester, UK: Distributed by the author; 2003.
67. Swofford D: **PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods).** In *Version 4.0b10 edn* Sunderland, Massachusetts: Sinauer Associates; 2002.
68. Sullivan J, Abdo Z, Joyce P, Swofford DL: **Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation.** *Mol Biol Evol* 2005, **22**:1386-1392.
69. Felsenstein J: **PHYLIP (PHYLogeny Inference Package).** In *Version 3.66 edn* Seattle: Department of Genome Sciences. University of Washington; 2001.
70. Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics.** *BMC Evol Biol* 2004, **4**:18.
71. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
72. Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL: **AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics.** *Bioinformatics* 2008, **24**:581-583.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



RESEARCH ARTICLE

Open Access

Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria)

Ylenia Chiari^{1,2*}, Vincent Cahais¹, Nicolas Galtier¹ and Frédéric Delsuc^{1*}

Abstract

Background: The morphological peculiarities of turtles have, for a long time, impeded their accurate placement in the phylogeny of amniotes. Molecular data used to address this major evolutionary question have so far been limited to a handful of markers and/or taxa. These studies have supported conflicting topologies, positioning turtles as either the sister group to all other reptiles, to lepidosaurs (tuatara, lizards and snakes), to archosaurs (birds and crocodiles), or to crocodylians. Genome-scale data have been shown to be useful in resolving other debated phylogenies, but no such adequate dataset is yet available for amniotes.

Results: In this study, we used next-generation sequencing to obtain seven new transcriptomes from the blood, liver, or jaws of four turtles, a caiman, a lizard, and a lungfish. We used a phylogenomic dataset based on 248 nuclear genes (187,026 nucleotide sites) for 16 vertebrate taxa to resolve the origins of turtles. Maximum likelihood and Bayesian concatenation analyses and species tree approaches performed under the most realistic models of the nucleotide and amino acid substitution processes unambiguously support turtles as a sister group to birds and crocodiles. The use of more simplistic models of nucleotide substitution for both concatenation and species tree reconstruction methods leads to the artefactual grouping of turtles and crocodiles, most likely because of substitution saturation at third codon positions. Relaxed molecular clock methods estimate the divergence between turtles and archosaurs around 255 million years ago. The most recent common ancestor of living turtles, corresponding to the split between Pleurodira and Cryptodira, is estimated to have occurred around 157 million years ago, in the Upper Jurassic period. This is a more recent estimate than previously reported, and questions the interpretation of controversial Lower Jurassic fossils as being part of the extant turtles radiation.

Conclusions: These results provide a phylogenetic framework and timescale with which to interpret the evolution of the peculiar morphological, developmental, and molecular features of turtles within the amniotes.

Background

Turtles (order Testudines) represent one of the most anatomically peculiar vertebrate groups. Their highly derived morphology relative to other vertebrates arose through profound structural changes associated with the origin of the shell [1]. Turtles have been described as having a conspicuously modified reptile body plan, and termed 'hopeful monsters', representing a successful phenotypic mutant with the potential to establish a new evolutionary lineage [2-7]. These morphological adaptations make it

difficult to compare turtles with other organisms and to establish the polarity of numerous anatomical characters as either being ancestral or derived. Because turtles represent one of the major groups of amniotes, resolving their phylogenetic position would fill an important gap in the evolutionary history of vertebrates, and contribute to the understanding of how such a key innovation as the turtle shell originated and which underlying genes are involved in its development [8].

The phylogenetic relationships of turtles within the amniotes have puzzled scientists for more than a century. Turtles have been classified both as basal to all other reptiles (including birds) and as nested within them, implying two radically different perspectives from

* Correspondence: yle@yleniachiar.i.it; frederic.delsuc@univ-montp2.fr
¹Institut des Sciences de l'Evolution, UMR5554-CNRS-IRD, Université Montpellier 2, Montpellier, France
Full list of author information is available at the end of the article

which to interpret the evolution of morphological, developmental, molecular, or ecological data. The classic view [9,10] places turtles as the sister group to all other reptiles, mostly on the basis of the lack of temporal fenestration in the skull, a character considered as being ancestral for Reptilia. This view reflects the traditional dichotomy of reptiles as Anapsida (lacking temporal fenestration) or Diapsida (with two temporal fenestrations). However, cladistic studies of morphological datasets have generated conflicting results, supporting both an anapsid [11-13] and a diapsid [14-16] affinity of turtles, depending on the fossil sampling considered and the morphological matrices used. Harris *et al.* [17] showed how the morphological characters used to assess the phylogenetic placement of the turtles within the tree of amniotes can lead to conflicting signals, and suggested, given these difficulties, that the answer to this long-standing controversy would most probably come from molecular data.

However, the use of molecular data has not yet settled the debate, as it has also provided somewhat conflicting results. In the large number of publications on the topic over the past decade, turtles have been grouped with Archosauria (birds and crocodiles) in most studies [18-22], but have also been grouped with crocodiles [23-26] or sometimes with Lepidosauria (tuatara, lizards, and snakes) [27]. The causes of the conflicting signals and/or lack of resolution obtained in most studies have been attributed to the limited number of genes considered, poor taxon sampling, substitution rate heterogeneity among genes and among taxa, and saturation or selection occurring at some of the markers [28]. Moreover, statistical tests performed to evaluate alternative topologies based on these early molecular sequence datasets usually failed to reach significance, probably because of the reduced number of genes included, but also possibly because of heterogeneity in gene trees.

With the advent of genomic data, the comparative datasets increased in size, but the issue of turtle phylogeny remained unresolved. The first investigation of genome structure and composition in reptiles identified a similarity in genomic signatures between turtles and crocodiles [29]. A recent multigene study offered the first convincing support for the grouping of turtles and archosaurs [30], but this result was contradicted by a newer study based on the distribution of microRNAs (miRNA), which strongly suggested an alternative turtles plus lizards clade [31]. Finally, a recent phylogenomic study based on reptile transcriptomic data did not find compelling support to distinguish between turtles plus crocodiles and turtles plus archosaurs, despite including a large number of genes [32]. In that study, the analysis of the largest dataset strongly supported a topology with the turtle as the sister group to the crocodile, whereas

analyses after removing potential paralogs favoured a turtle plus archosaurs clade, albeit with reduced statistical support [32].

In the present study, we used a phylogenomic approach [33] to resolve the position of the turtles within the amniotes, and estimated the time of their origin using a dataset comprising 248 nuclear protein-coding genes for 16 vertebrates. We applied phylogenetic reconstruction methods and models of sequence evolution, explicitly accounting for substitution rate heterogeneities among taxa and among genes, and maximum likelihood (ML) species tree analyses accounting for gene tree discordance. We also used various relaxed molecular clock Bayesian approaches to reconstruct a timescale for the evolutionary history of the amniotes.

Results

Phylogenetic results

Our phylogenomic dataset provides strong support for the phylogenetic position of turtles as a sister group to Archosauria within Amniota based on concatenation analyses (Figure 1). All of our Bayesian and ML analyses of the concatenated amino-acid dataset recovered this topology with maximal ML bootstrap support (BP) and Bayesian posterior probabilities (PP) irrespective of the model used ($BP_{ML} = 100$; $BP_{PARTG} = 100$; $PP_{BAY} = 1.0$; $PP_{CAT} = 1.0$) (Figure 1a; Table 1). The same result was obtained from analyses of the complete nucleotide dataset with ML and Bayesian analyses when a mixed model partitioned by codon was applied ($BP_{PARTC} = 100$; $PP_{PARTC} = 1.0$), and in Bayesian analyses conducted under the site-heterogeneous CAT-GTR + G4 mixture model ($PP_{CAT} = 1.0$) (Figure 1b; Table 1). Conversely, ML and Bayesian phylogenetic reconstructions from the complete nucleotide dataset using a single site-homogeneous GTR + G model for the whole concatenation ($BP_{ML} = 76$; $PP_{BAY} = 1.0$), and a mixed model partitioned by gene ($BP_{PARTG} = 54$) tended to support an alternative topology grouping turtles with crocodylians (Table 1).

Likelihood-based comparisons of partitioned models based on the Akaike information criterion (AIC) showed that partitioning by codon position using the GTR + G model was by far the best partition scheme ($AIC_{CONCAT} = 2,109,010$; $AIC_{ByGene} = 2,082,688$; $AIC_{ByCodon} = 2,008,142$). The fact that only the suboptimal and poorly fitting models supported a turtles + crocodylians relationship suggests that this topology is a phylogenetic reconstruction artefact, most likely the result of the inability of these models to account efficiently for site-specific heterogeneities in the substitution process. The better fit offered by the codon position partition scheme over the gene partition scheme indicates that the main source of heterogeneity lies in the codon positions, most probably because of multiple substitutions accumulating at third codon positions.

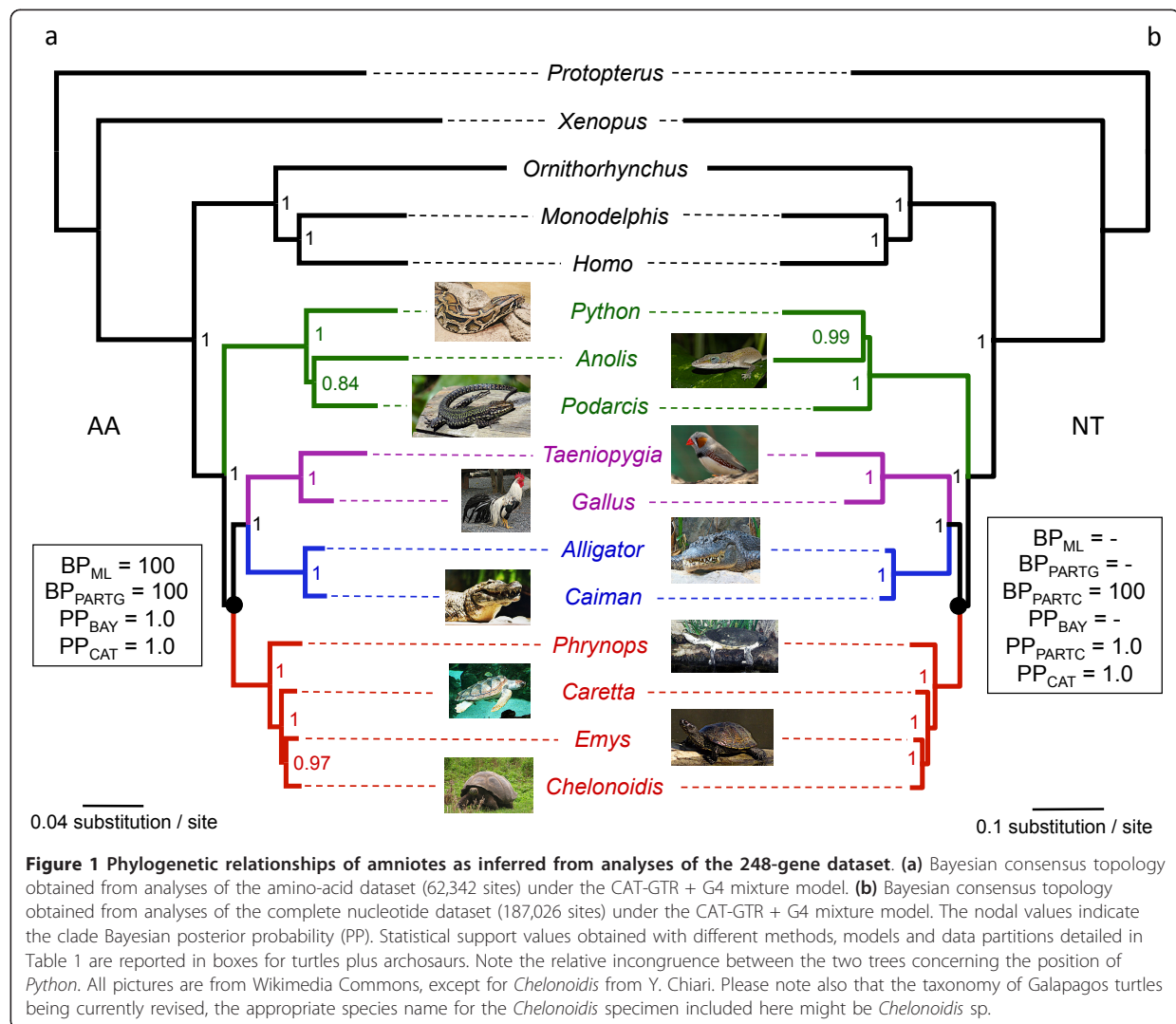


Table 1 Statistical support for the phylogenetic position of turtles based on the various reconstruction methods, substitution models, and data partitions.

	Amino acids		Nucleotides	
	All positions	All positions	Positions 1 + 2	Positions 3
Total sites	62,342	187,026	124,684	62,342
Constant sites	41,170 (66.0%)	99,638 (53.3%)	92,128 (73.9%)	7,510 (11.2%)
Informative sites	8,749 (14.0%)	54,880 (29.3%)	14,009 (11.2%)	40,871 (65.6%)
RaxML LG + G / GTR + G	Turtles + Archosaurs BP _{ML} = 100	Turtles + Crocodiles BP _{ML} = 76	Turtles + Archosaurs BP _{ML} = 100	Turtles + Crocodiles BP _{ML} = 100
RaxML GTR + G partitioned by gene	Turtles + Archosaurs BP _{PARTG} = 100	Turtles + Crocodiles BP _{PARTG} = 54	-	-
RaxML GTR + G partitioned by codon	-	Turtles + Archosaurs BP _{PARTC} = 100	-	-
MrBayes WAG + G / GTR + G	Turtles + Archosaurs PP _{BAY} = 1.0	Turtles + Crocodiles PP _{BAY} = 1.0	Turtles + Archosaurs PP _{BAY} = 1.0	Turtles + Crocodiles PP _{BAY} = 1.0
MrBayes GTR + G partitioned by codon	-	Turtles + Archosaurs PP _{PARTC} = 1.0	-	-
PhyloBayes CAT-GTR + G	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0	Turtles + Archosaurs PP _{CAT} = 1.0

Comparisons of ML-based saturation plots [34] between the amino-acid and the complete nucleotide datasets (Figure 2a) did not reveal clear evidence for global substitutional saturation of the complete nucleotide dataset relative to the amino-acid dataset, despite a slightly lower slope (0.36 versus 0.50, respectively). However, as expected in protein-coding genes conserved at this level of divergence, substitutional saturation was particularly pronounced at the third codon positions (Figure 2b). In cases in which the substitutional saturation of third codon positions was particularly high, excluding this third codon position partition from the dataset would be expected to result in less biased phylogenetic reconstructions. In agreement with this prediction, all ML and Bayesian reconstructions performed on the nucleotide dataset after exclusion of third codon positions provide unambiguous support ($BP_{ML} = 100$; $PP_{BAY} = 1.0$; $PP_{CAT} = 1.0$) for regrouping turtles and archosaurs (Table 1). Conversely, ML and Bayesian analyses of concatenated third codon positions using a single GTR + G model returned maximal support ($BP_{ML} = 100$; $PP_{BAY} = 1.0$) for the topology clustering turtles with crocodylians (Table 1). Only the CAT-GTR + G4 mixture model seemed to be able to deal efficiently with the saturated third codon positions dataset by strongly supporting the turtles + archosaurs clade ($PP_{CAT} = 1.0$).

These analyses indicate that substitutional saturation at third codon positions is so strong in this phylogenomic dataset that it is able to impede phylogenetic reconstruction when inappropriate models of sequence evolution are used.

Statistical tests between competing topologies confirmed the above results (Table 2). The approximately unbiased (AU) likelihood-based test showed that all proposed alternative hypotheses to the sister group relationship of turtles with archosaurs were rejected based on the amino-acid dataset, irrespective of the model used. In concordance with the results of saturation analyses, the complete nucleotide dataset did not distinguish statistically between the competing alternatives of turtles plus archosaurs and turtles plus crocodylians. These more equivocal and method-dependent results, when nucleotide sequences were used, are suggestive of conflicting phylogenetic signals between codon positions. However, the alternative topologies proposing turtles as the sister group to other reptiles (including birds), and grouping turtles with squamates (lizards and snakes) received no support from our data.

Finally, given the fact that the internal branch lengths connecting the main reptiles lineages seemed to be relatively short in trees obtained from concatenated analyses (Figure 1), we also explored the potential influence of the

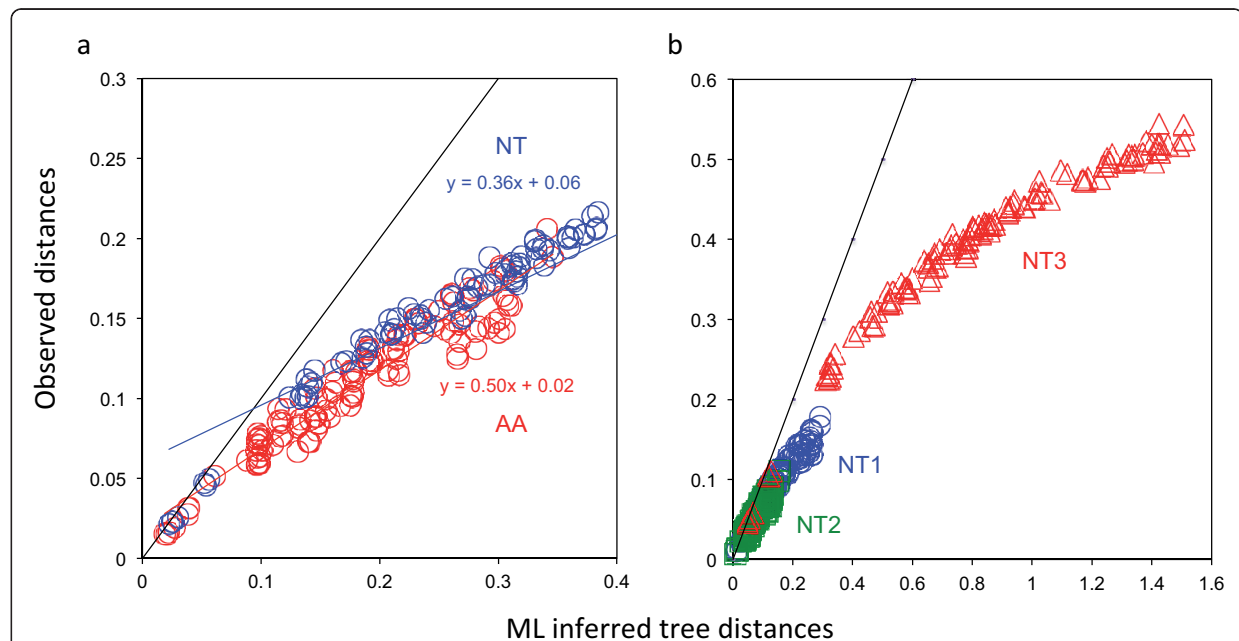


Figure 2 Analyses of substitution saturation at each codon position. Maximum likelihood saturation plots [34] were compared (a) between the complete amino-acid and nucleotide datasets, and (b) between the codon positions of the complete nucleotide dataset. The observed pairwise distances between the 16 taxa were directly computed from sequence alignments, and the corresponding inferred pairwise tree distances calculated from branch lengths of the ML topology. The $Y = X$ line marks the theoretical limit where the number of observed substitutions equals the number of inferred substitutions. The slope of the linear regression indicates the amount of substitution saturation; the smaller the slope, the greater the number of inferred multiple substitutions.

Table 2 Results of the approximately unbiased likelihood-based statistical test for comparing alternative topologies using different data types and partitions.

Topologies	Amino acids		Nucleotides		
	Concatenated	Partitioned by gene	Concatenated	Partitioned by gene	Partitioned by codon
Turtles + archosaurs	best	best	0.26	0.85	0.91
Turtles + crocodylians	< 0.001*	< 0.001*	0.74	0.15	0.09
Turtles + squamates	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
Turtles + other reptiles	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*

*Significant.

underlying gene-tree heterogeneity created by deep coalescence events, which might lead to statistical inconsistency of concatenation-based methods in the anomaly zone [35,36]. The results obtained using the maximum pseudo-likelihood for estimating species trees (MP-EST) approach showed high consistency with the results of our concatenation-based analyses (Figure 3). Indeed, the species tree reconstructed from the amino-acid ML gene trees unambiguously supported (BP = 100) the grouping of turtles and archosaurs (Figure 3a), whereas the species tree based on nucleotide ML gene trees supported (BP = 87) a conflicting turtles plus crocodylians clade (Figure 3b), as previously shown in concatenation-based analyses using suboptimal models of sequence evolution. In fact, only six amino-acid and three nucleotide ML gene trees were fully compatible with their corresponding species trees. These figures illustrate the large extent of gene-tree heterogeneity in this dataset, which probably reflects the large effect of stochastic error on individual gene-tree inference. We interpret these congruent results between concatenation and species tree inference as good evidence that the source of the statistical inconsistency resulting in the grouping of turtles with crocodiles does not come from potential discordances between gene trees and the species tree, but rather from the influence of substitutional saturation of third codon positions in individual gene-tree inference.

Molecular dating results

Detailed results from the molecular dating analyses performed under auto-correlated models of molecular clock relaxation are presented in Table 3. Divergence date estimates varied depending on the methods and datasets used, but were nevertheless consistent between the two programs we used (MCMCTree and PhyloBayes). We generally found more consistency with published estimates for the results obtained with PhyloBayes under the CAT-GTR + G site-heterogeneous mixture model (Table 3) than for the results obtained with the site-homogeneous LG + G / WAG + G and GTR + G models. Our analyses based on the CAT-GTR + G model placed the divergence between turtles and archosaurs around the Permian-Triassic boundary at a mean of 255 Mya (range 274 to 233 Mya), the separation of crocodylians

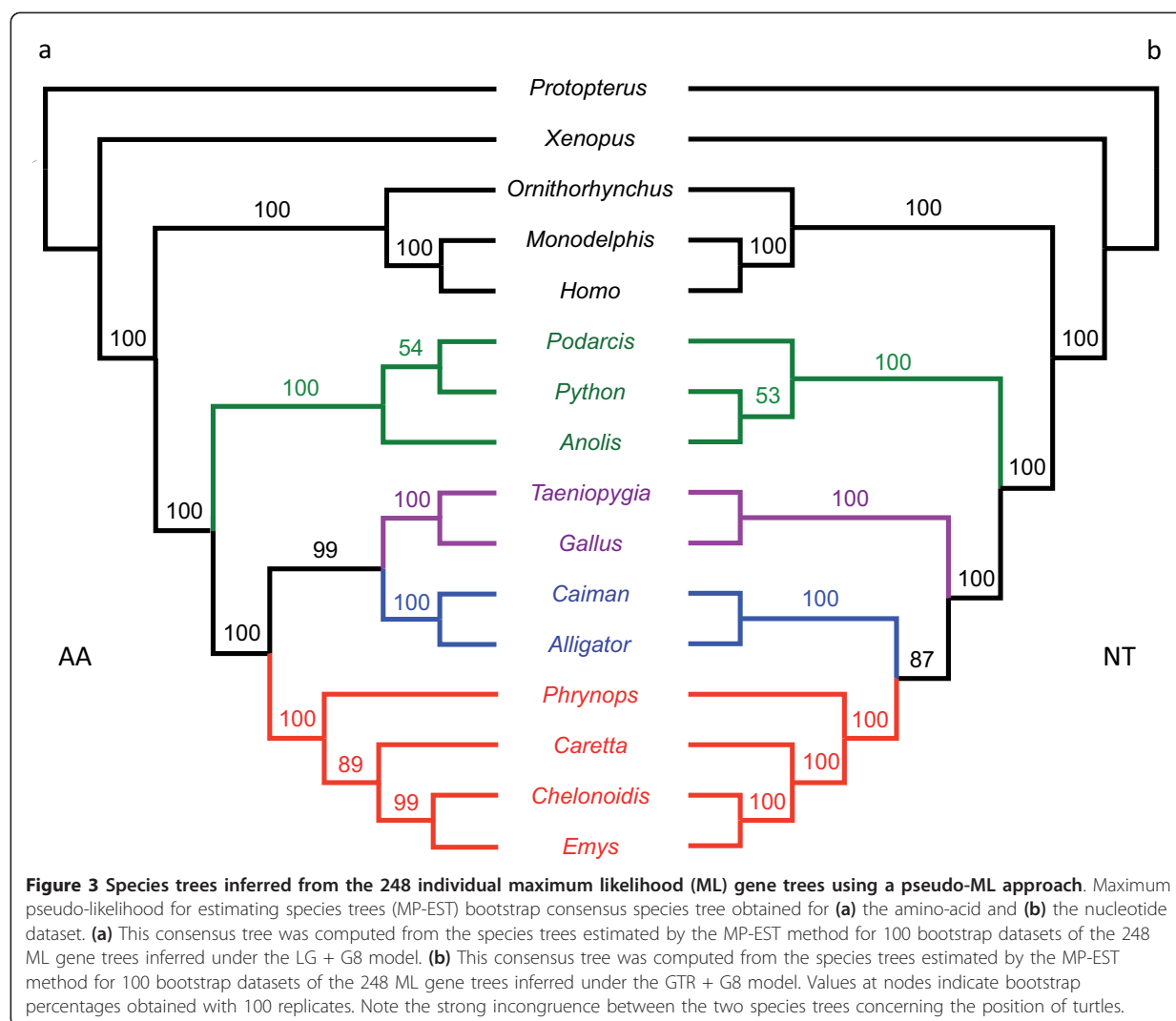
and birds in the Upper Triassic with a mean of 219 Mya (249 to 186 Mya), and the most recent common ancestor (MRCA) of living turtles (corresponding to the separation between Pleurodira and Cryptodira) in the Upper Jurassic with a mean of 157 Mya (207 to 104 Mya) depending on whether amino acids or nucleotides are considered (Table 3). The chronogram obtained from the analysis of the nucleotide dataset using the CAT-GTR + G model is shown in Figure 4.

Strikingly different results were obtained when using uncorrelated models of clock relaxation (see Additional file 1). Again, dating estimates were fairly consistent between the different program implementations (BEAST, MCMCTree, and PhyloBayes), but using uncorrelated rate models generally led to much smaller age estimates than the ones obtained under auto-correlated rate models. For example, using uncorrelated models, the MRCA of living turtles was estimated to be half the age of that found with auto-correlated models, with mean estimates ranging from 81 to 64 Mya versus 167 to 107 Mya, respectively. Other estimates, such as the caiman/alligator divergence, were reduced by two-thirds, resulting in unreasonably recent estimations relative to the TimeTree values (see Additional file 1).

Discussion

Phylogenomics and the position of turtles

Previous phylogenetic investigations of amniote phylogeny have failed to reach a clear consensus on the phylogenetic position of turtles, as the various studies have often produced ambiguous and sometimes conflicting results. The causes for this probably stem from the intrinsic difficulty of this phylogenetic problem, which involves ancient divergences. Most of the previous molecular studies addressing this question used either small datasets based on a few nuclear genes [19,24] or genetically linked mitochondrial genes [22,23,37]. In general, phylogenetic analyses based on using mitochondrial data tended to recover a sister group relationship between turtles and Archosauria [20-22,37,38], whereas some of the nuclear data favoured a clade of turtles with crocodylians [23,24,26]. The only exception to this pattern is the study by Iwabe *et al.* [19], who reported statistical



support for the turtles plus archosaurs clade, but this was based on only two nuclear genes and a minimal taxon sampling.

Resolving the branching patterns of the major lineages of amniotes requires gathering a considerable amount of informative sequence data from independent markers with adequate taxon sampling. Shen *et al.* [30] recently investigated this question using 23 (mostly nuclear) markers for 28 vertebrates, and estimated that with their taxon sampling, a total sequence length of more than 13,000 nucleotides from independent nuclear markers is necessary to provide significant statistical support for resolving the controversial relationships between turtles, birds, and crocodylians. Our phylogenomic results, based on 248 nuclear markers, corroborate their predictions about the challenge represented by resolving this phylogenetic question, and add support to the sister group relationship between turtles and archosaurs (birds plus

crocodylians). Furthermore, our statistical analyses reject any alternative hypotheses to the sister group relationship of turtles to Archosauria (Table 2), thus advancing the resolution of this long-standing controversial issue of vertebrate evolutionary history.

However, as illustrated by the occurrence of conflicting signals in our phylogenetic analyses, the phylogenomic approach is not immune to statistical inconsistency [39], as highlighted here in the cases of ML and Bayesian analyses of nucleotide datasets under a single concatenated GTR + G model, and in species tree inference from nucleotide gene trees, which showed inconsistencies that are most likely due to saturation at third codon positions. The fact that turtles group with crocodylians in concatenation analyses is probably due to a long-branch attraction (LBA) artefact causing the faster evolving squamates to be attracted towards mammals and the outgroups (see Additional file 2). The same grouping of turtles + crocodyles

Turtles and archosaurs MRCA	229 (200 to 253)	228 (208 to 252)	249 (233 to 270)	228 (188 to 256)	225 (212 to 240)	258 (238 to 274)	244/265
Archosaurs MRCA	200 (166 to 231)	188 (162 to 217)	211 (186 to 236)	209 (168 to 241)	201 (180 to 221)	226 (199 to 249)	238/245
Turtles MRCA	121 (63 to 183)	126 (85 to 166)	147 (104 to 185)	107 (53 to 183)	133 (93 to 171)	167 (120 to 207)	207/211
<i>Caretta/Emys + Cheloniid</i>	77 (40 to 127)	83 (47 to 117)	99 (64 to 139)	69 (34 to 129)	90 (52 to 125)	115 (71 to 154)	97/99
<i>Emys/Chelonoidis</i>	65 (25 to 116)	71 (38 to 102)	87 (54 to 124)	52 (16 to 112)	74 (39 to 106)	95 (56 to 131)	70/70
<i>Caiman/Alligator</i>	72 (15 to 156)	54 (33 to 86)	61 (35 to 93)	60 (11 to 160)	39 (21 to 66)	46 (24 to 87)	70/72
<i>Podarcis/Python + Anolis</i>	171 (119 to 218)	150 (116 to 183)	134 (100 to 172)	167 (108 to 219)	150 (122 to 181)	119 (90 to 151)	190/178
<i>Python/Anolis</i>	147 (81 to 198)	136 (98 to 171)	127 (93 to 167)	141 (58 to 199)	135 (101 to 161)	105 (74 to 136)	171/163

Abbreviations: MRCA, most recent common ancestor

^aValues are mean (95% Credibility interval).

^bValues taken from the TimeTree database [92].

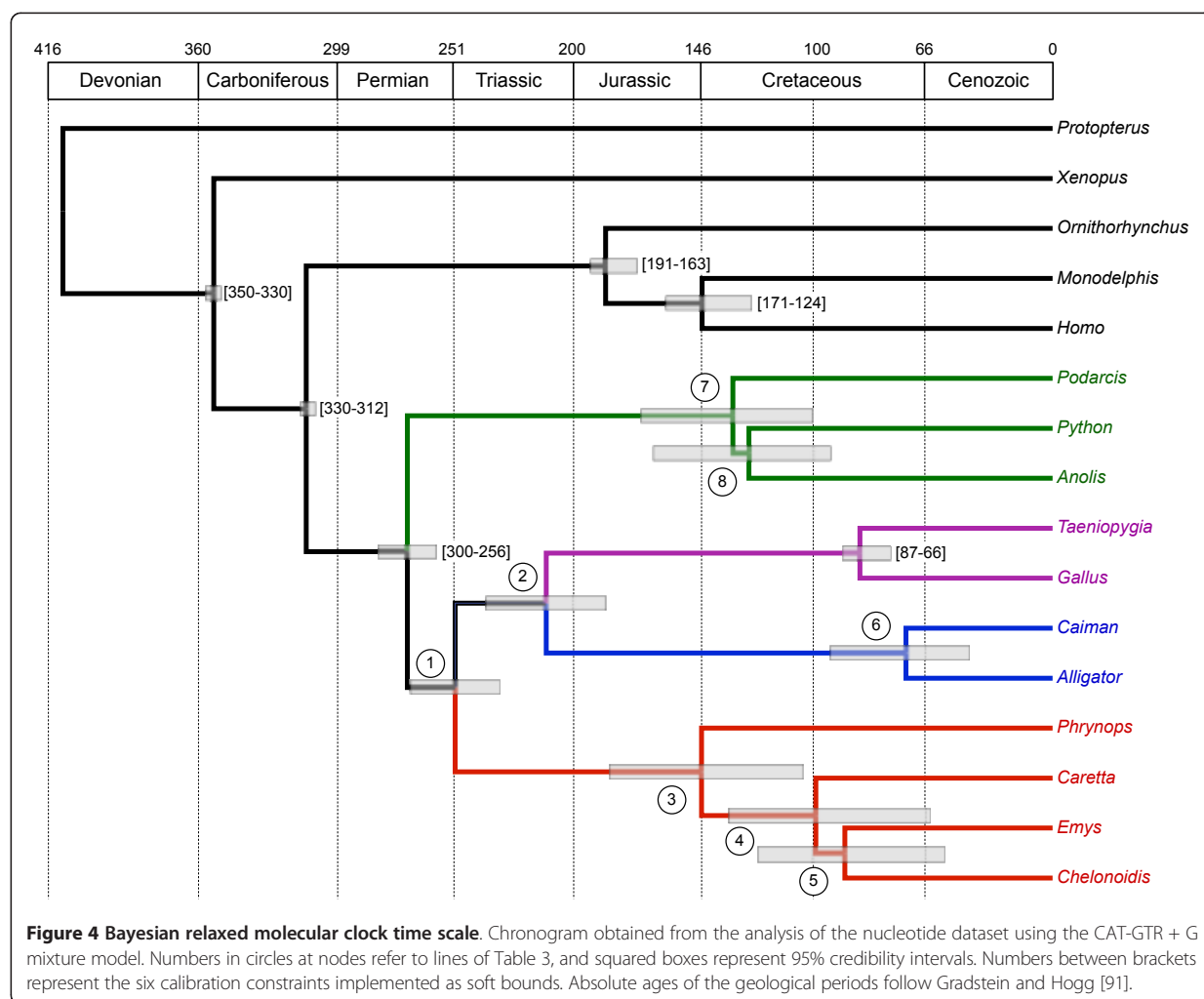
was also retrieved with strong support by Tzika *et al.* [32] based on ML and Bayesian analyses of amino-acid datasets using site-homogeneous empirical models and a reduced taxon set. Those authors evoked the same kind of LBA artefact to explain what they considered as an artefactual result, as support for it disappeared in analyses including fewer sites but with fewer missing data [32]. These observations confirm that phylogenomic reconstruction can lead to artefacts, especially when the taxon sampling is reduced and model assumptions are violated [40]. When analysing large concatenations, use of best-fit models is required to account specifically for heterogeneities among genes and codons in the substitution process, and to alleviate the deleterious effects of substitution saturation. Similarly, we found that when using mixed models for analysing protein-coding gene concatenations, partitioning by codon position outperformed the widely used gene-partitioning scheme. The CAT-GTR + G mixture model nevertheless offers the most efficient solution to account explicitly for site heterogeneities in the substitution process as typically observed in phylogenomic datasets [41].

As illustrated by our results, statistical inconsistency is not restricted only to concatenation-based phylogenetic reconstruction methods. Although specifically designed to accommodate potential gene-tree discordances, species tree inference methods also seem to be sensitive to misspecification of the substitution model used to infer gene trees. Indeed, the species tree obtained from the nucleotide dataset also strongly supported the artefactual grouping of crocodiles and turtles (Figure 3b). Therefore, in addition to their accounting for gene-tree heterogeneity,

the use of the best-fitting substitution models seems to be equally important for these methods [42]. These results also indicate a potential problem of stochastic error in reconstructing gene trees for which only a limited number of sites is available, and the consequent effect on species tree inference. Ultimately, species trees can only be as good as the gene trees from which they are inferred.

Finally, it is worth noting that a recent analysis of miRNA phylogenetic distribution [31] supported a branching order (turtles + squamates) that was strongly rejected by our data. This is not the first instance of a conflict between miRNA and sequence-based phylogenetic studies, as shown by the case of acoels for instance [43,44]. Thus, our study suggests caution is needed when using miRNA markers in phylogenetic reconstructions, as they might not be as free from homoplasy as sometimes considered [45]. For example, secondary loss of multiple families of miRNAs have already been reported in tunicates [46]. Our results imply that the four miRNAs families exclusively shared by turtles and lizards [31] either have been lost secondarily in archosaurs, or have been independently recruited in turtle and lizard genomes. Upcoming reptile genomic data [47,48] should allow verification of these predictions.

The well-supported sister group relationship of the turtles to the archosaurs has important implications for the evolution of morphological, developmental, and ecological characters of amniotes. It implies, as previously proposed



[15], that turtles experienced a secondary closure of the temporal fenestration, which therefore appears to be a derived character, rather than a reflection of the ancestral condition, as has long been assumed. In addition, because a basal position of turtles within reptiles is supported by the timing of events in organogenesis [49], our results suggest the occurrence of significant convergence in developmental timing characters, and advocate for the re-interpretation of sequence heterochrony data in the light of the phylogenetic position of turtles supported by our phylogenomic analyses. Finally, the assumption of an aquatic origin of turtles (the hypothesis that was brought forward due to the proposed sister group relationship between turtles and an extinct group of marine reptiles (Sauropterygia) and Lepidosauria [16]) also needs to be reconsidered. A recent study suggested, for example, that stem turtles could have occupied both terrestrial and aquatic habitats [50].

The proposed phylogeny of amniotes also provides a more solid background from which to investigate the evolution of the sex-determining systems and genomic characteristics of reptiles. Whereas mammals and birds have only genetic sex determination, non-avian reptiles have both genetic and temperature-dependent sex determination. Temperature-dependent sex determination also occurs in crocodylians, tuatara, and the majority of turtles, whereas it is less common in squamates [51,52]. Studies on this subject have relied on a traditional view of the vertebrate phylogeny, with turtles being basal to the other reptiles (including birds) (compare, for example Janzen & Krenz [51] with Janes *et al.* [53]). Although the phylogenetic scenario supported by our data would not change the main conclusion that temperature-dependent sex determination evolved multiple times within amniotes, it does provide a basis from which to further investigate the possible adaptive evolutionary value of the

temperature-dependent sex determination in amniotes and the evolution of sex chromosomes.

Recently, Matsuda *et al.* [54] reported a high degree of conservation between the chromosomes of a turtle (*Pelodiscus sinensis*) and the common chicken, in accordance with an archosaurian affinity of turtles. These authors also suggested that although no specific sex chromosomes could be identified in the studied turtle, which has genetically determined sex, the ancestral avian Z sex chromosome has been conserved in the turtle genome. However, other features of the genome, such as its average genome size, GC content, and distribution of transposable elements show a marked similarity between turtles and crocodiles, to the exclusion of birds [29,55]. By rejecting the turtles plus crocodylians grouping, our analysis could possibly be interpreted as evidence for a parallel evolution of these genomic features in the two lineages, or, perhaps more plausibly, recent evolution of bird-specific features in the avian lineage.

Models of molecular clock relaxation

In our molecular dating analyses, we found discrepancies between the results obtained using standard substitution models (LG + G and GTR + G) and the CAT-GTR + G mixture model with both amino acids and nucleotides. These differences probably stem from underlying differences in branch-length estimates between the two types of models, indicating the need to apply the most appropriate models of sequence evolution currently available for conducting molecular dating analyses [56]. Our results indicate that the CAT-GTR + G mixture model better accounts for the site-specific heterogeneities of our concatenated protein-coding gene datasets. Therefore, we consider that the divergence date estimates obtained under this model are the most reliable.

However, these small discrepancies between estimates obtained under different substitution models are almost negligible as compared with the large differences in estimates between the auto-correlated and uncorrelated models of rate change. In our case, the use of uncorrelated models generally led to unreasonably recent dating estimates for all nodes relative to the values reported in the literature [90]. These results seriously question the adequacy of the uncorrelated models of molecular clock relaxation parameters for estimating divergence times, at least with our dataset. Based on Bayes factor comparisons, Lepage *et al.* [56] showed that auto-correlated models provide a significantly better fit than the uncorrelated gamma model, especially for large phylogenomic datasets. These results were recently confirmed in an empirical phylogenomic study focusing on the estimation of arthropod divergence times, for which the assumption of rate autocorrelation seemed to be the most realistic way of modelling evolutionary rate variations across the tree

[56,57]. For these reasons, we consider the results from the auto-correlated relaxed clock analyses under the CAT-GTR + G substitution model as our most reliable dating estimates (Table 3; Figure 4).

Paleontological implications

Our auto-correlated relaxed clock analyses based on the CAT-GTR + G model support a divergence between turtles and Archosauria around 255 Mya (274-233 Mya), which is in agreement with the estimates recently reported by Shen *et al.* [30]. The dating obtained for other nodes also seems to be mostly compatible with current knowledge. For example, the Testudinoidea MRCA corresponding to the divergence between *Emys* and *Chelonoidis* is estimated at a mean of 91 Mya (range 131 to 54 Mya), which is comparable with that obtained by Lourenço *et al.* [58]. Exceptions concern squamates and crocodylians, for which our estimates indicated a more recent time than generally reported (Table 3). We note that the confidence intervals are relatively large, however, as would be expected for such indirect estimates, in which dates are estimated jointly with the process of substitution-change variations over time [59].

The single major difference between our estimates and the previously published divergence dates concern the MRCA of living turtles. This is a controversial issue in the paleontological literature, with proposed ages of divergence between the two main turtle lineages (Pleurodira and Cryptodira) varying from the Upper Triassic to the Upper Jurassic. This debate is mostly due to the rarity and the need for better characterization of turtle fossils older than 160 Mya [60]. Our analyses suggest that the MRCA of Chelonia (Pleurodira plus Cryptodira) is on average 157 Mya (range 207 to 104 Mya), taking the average mean and extremes of 95% credibility intervals for the CAT-GTR + G model amino-acid and nucleotide results. This means that an Upper Triassic origin (229 to 200 Mya) of extant turtle lineages is rejected, and that although a Lower Jurassic origin (200 to 176 Mya) is still possible, it seems unlikely. Remarkably, a similar conclusion was reached in a recent study using a fully independent molecular dataset, which only included turtle sequences and within-turtle fossil calibrations [58]. Our 95% credibility intervals for the turtle ancestral node (185 to 104 Mya with amino acids, and 207 to 120 Mya with nucleotides) are nevertheless wider and probably less realistic in including the Lower Cretaceous (146 to 97 Mya). However, taken together, these two analyses begin to suggest that the origin of Chelonia may be in the Middle or Upper Jurassic (176 to 146 Mya) or later. If so, two controversial fossils, *Proterochersis* and *Kayentachelys*, attributed respectively to the Cryptodira and Pleurodira clades, would be currently misplaced on the turtle lineage. These placements, proposed by Gaffney [61], have been a subject of intense

debate [62,63] (references therein). This would also have important implications for molecular clock analyses, because these fossils are usually used for calibrating both turtle [64] and amniote [30] trees. Considering this, it may be prudent to consider these fossils *Testudines incertae sedis* until additional data can be obtained to confirm their phylogenetic placement.

Conclusions

As already shown in cases of other difficult phylogenetic questions, we found analyses of phylogenomic data to be useful in resolving the uncertain placement of turtles within the phylogeny of amniotes. In fact, our analyses show that the hypothesis of a sister group relationship between turtles and crocodylians is most likely a phylogenetic reconstruction artefact related to substitution saturation. When this artefact is taken into account by using the best models of sequence evolution currently available, we found strong support in all cases for identifying turtles as a sister group to Archosauria, to the exclusion of any alternative phylogenetic hypothesis. Our results confirming turtles as derived diapsids have important implications for understanding the evolution of morphological characters and for interpreting developmental and genomic data in amniotes. Finally, our results shed light on another debated topic by contesting the ancient Lower Jurassic origin of the two main extant lineages of turtles. Indeed, our molecular dating places the MRCA of living turtles in the Upper Jurassic period, a more recent estimate than previously reported, and one that questions the interpretation of controversial Lower Jurassic fossils considered as 'crown turtles'.

Methods

Transcriptome data acquisition

Blood samples were obtained from four species of turtle for which genomic data were not already available: *Phrynosops hilarii*, *Caretta caretta*, *Chelonoidis nigra*, and *Emys orbicularis*, representing the two suborders Pleurodira and Cryptodira. We also took a jaw sample from a crocodylian (*Caiman crocodylus*) and a liver sample from a lacertid lizard (*Podarcis* sp.). A jaw sample from a lungfish (*Protopterus annectens*) was used to provide an outgroup for the phylogenetic analyses.

RNA was extracted and checked for quality and quantity in accordance with previously described protocols [65,66]. Transcriptome sequencing using the 454 GS-FLX Titanium platform (454 Life Sciences, Branford, Connecticut, USA) was performed by GATC Biotech (Konstanz, Germany). We also retrieved 454 sequencing reads from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) for two additional reptile species: a snake (*Python molurus bivittatus*; SRX072633, SRX072634, SRX057862, and SRX018167

[67,68]) and an alligator (*Alligator mississippiensis*; SRX012365 [69]). All raw sequencing reads were cleaned from sequencing adaptors and then assembled into *de novo* contigs for each of the nine species using either CAP3 [70] or PCAP [71] assembly software. The basic statistics on raw reads and contig assemblies are indicated in Table 4.

Data assembly

For constructing our phylogenomic dataset, we designed an analytical pipeline aiming at conservatively selecting a set of single-copy orthologous genes that would also minimize the amount of missing data in the assembled dataset (see Additional file 3). We relied on the EnsemblCompara phylogenetic assessment of orthology [72] by downloading the 7,943 (Ensembl, release 60) coding sequences (CDSs) of the 1:1 orthologous genes shared by *Homo sapiens*, *Monodelphis domestica*, *Gallus gallus*, and *Taeniopygia guttata*. We then added all CDSs that are predicted to be 1:1 orthologous genes between *H. sapiens* and *Ornithorhynchus anatinus*, *H. sapiens* and *Anolis carolinensis*, and *H. sapiens* and *Xenopus tropicalis*. This resulted in 4,305 1:1 orthologous CDSs that are shared by these seven core species. A best reciprocal hit (BRH) strategy was then used to identify 1:1 orthologs from the contigs of the nine assembled transcriptomes. We performed tBLASTx searches, using the 4,305 CDSs of *G. gallus* against each contig set (parameters: length > 100 nucleotides, score > 100, e-value < $1e^{-100}$, and identity > 50%). Another BLAST search was then performed for each matching contig against the full CDS set of *G. gallus* (17,934 genes), and only contigs with a significant BRH on exactly the same CDS were conserved. This BRH step led to 2,118 1:1 orthologous CDSs, for which at least one contig from the nine transcriptomes was added to the initial set of seven species. These datasets were then filtered taxonomically using the PhyloExplorer software [73], to keep only the 367 datasets that contained at least one turtle and one crocodile.

The resulting 367 CDS multiFASTA files were then aligned with MACSE [74] using the next-generation sequencing default settings option. This allowed us to align the newly assembled contigs against the seven reference CDSs from Ensembl, while respecting the coding frame by inserting frameshifts and stop codons in assembled contigs where needed. The 367 nucleotide alignments were then manually curated and trimmed, based on MASCE annotations of frameshifting events. Datasets in which turtle and crocodile sequences did not overlap were excluded. ML trees were then inferred from the remaining 331 alignments using PHYML (version 3.0) [75] with SPR moves on a BIONJ starting tree under the GTR + G8 model. We next excluded genes for which amniotes were not monophyletic as they are likely to

Table 4 Basic sequencing and assembly statistics for the seven newly sequenced and the two additional transcriptomes included in this study.

Species	Number of 454 reads	Mean read length after cleaning	Number of contigs	Mean contig length
<i>Phrynosops hilarii</i> *	454,711	303	45,554	548
<i>Caretta caretta</i> *	377,038	250	34,440	450
<i>Chelonoidis nigra</i> *	426,721	270	55,272	481
<i>Emys orbicularis</i> *	522,009	357	39,166	588
<i>Alligator mississippiensis</i>	436,439	266	29,309	400
<i>Caiman crocodilus</i> *	343,080	217	22,025	378
<i>Podarcis</i> sp.*	303,076	198	20,756	357
<i>Python molurus bivittatus</i>	950,283	232	36,334	446
<i>Protopterus annectens</i> *	666,034	231	37,810	469

*Newly sequenced in this study.

correspond to orthology assessment problems of the *Xenopus* and/or *Protopterus* sequences used as outgroups. Ambiguously aligned codons were filtered out from the resulting 260 alignments by using Gblocks [76] with default parameters. After excluding the datasets containing less than 300 nucleotide sites, the concatenation of the final 248 CDS datasets represented a total of 187,026 nucleotide sites (62,342 amino-acid sites) for 16 taxa, with only 35% missing data. These two final datasets have been deposited in the Dryad digital repository [77]. A table indicating the chicken Ensembl gene identification numbers (IDs) and official gene names of the 248 genes used is provided (see Additional file 4).

Phylogenetic analyses

Phylogenetic analyses were performed using both ML and Bayesian reconstruction methods on the nucleotide and amino-acid datasets. ML analyses of the different concatenations (amino acids, all nucleotide sites, codon positions 1 + 2, and codon positions 3) were first conducted using RAxML (version 7.2.8) [78] using a single LG + F + G model for amino acids, and a single GTR + G model for nucleotide datasets. We also performed ML searches under mixed models partitioned by gene (248 partitions) and codon position (3 partitions) using the same LG + F + G and GTR + G models for each amino-acid and nucleotide partitions, respectively. In mixed-model analyses, branch lengths were optimized individually per partition. ML bootstrap values were computed by repeating the original ML heuristic search on 100 bootstrap pseudoreplicates for each dataset. The AU statistical test for comparing alternative topologies was computed using CONSEL [79] from site-wise log-likelihood values estimated by RAxML for four *a priori* competing phylogenetic hypotheses for the position of turtles.

Bayesian phylogenetic inferences were conducted using MrBayes (version 3.1.2) [80] using a single WAG + G model for amino acids and a single GTR + G model for nucleotide datasets. We also applied the Bayesian approach using a mixed model partitioned by codon

position (three partitions) with a GTR + G model for each nucleotide partition. In this mixed-model analysis, all model parameters including branch lengths were unlinked between partitions. All Bayesian analyses were computed using four incrementally heated Metropolis-coupled Markov chain Monte Carlo (MCMCMC) run for 1,000,000 generations, with trees and associated model parameters sampled every 100 generations. The initial 1000 sampled trees (10%) were discarded as the burn-in, and the 50% majority-rule Bayesian consensus tree and associated clade PPs were computed from the remaining 9000 trees.

We also performed Bayesian phylogenetic analyses under the site-heterogeneous CAT-GTR + G4 mixture model [81] on both amino-acid and nucleotide datasets using PhyloBayes (version 3.3b) [82]. To avoid potential biases associated with a large proportion of invariable sites in estimating the site-heterogeneous CAT profiles, analyses of codon positions 1 + 2 and codon position 3 were conducted with constant sites excluded (-dc option). In each individual analysis, two independent MCMC chains starting from a random tree were run for 20,000 cycles, with trees and associated model parameters being sampled every 10 cycles until 2,000 trees were sampled. The initial 200 trees (10%) sampled in each MCMC run were discarded as the burn-in period. The 50% majority-rule Bayesian consensus tree and the associated PP_{CAT} were then computed from the remaining 3,600 trees combined from the two independent runs.

Finally, to account for potential discordances between gene trees and the species tree, we used the pseudo-ML approach implemented in the MP-EST program [83]. The species tree was inferred under the coalescent model from the 248 individual ML gene trees obtained by PHYML with SPR moves on a BIONJ starting tree under the GTR + G8 model for nucleotides and the LG + G8 model for amino acids. The reliability of the species tree inference was assessed using a nonparametric bootstrapping procedure resampling sites within individual genes

[84]. PHYML, using the same settings as described above, was first used to infer 100 ML bootstrap trees from each individual gene dataset, then the initial MP-EST species tree inference was repeated 100 times from each of the 248 individual bootstrap gene-tree replicates. Bootstrap percentages were finally obtained by computing the 50% majority-rule consensus tree from the resulting 100 bootstrap MP-EST species trees.

Molecular dating analyses

Dates of divergence between amniotes were estimated using the Bayesian relaxed molecular clock approaches implemented in PhyloBayes, MCMCTree from the PAML (version 4.5) package [85], and BEAST (version 1.7) [86], using the Bayesian consensus topology obtained from nucleotides (Figure 1b). With PhyloBayes, both amino-acid and nucleotide datasets were analysed under the CAT + GTR + G4 mixture model, and under the standard LG + G and GTR + G models, respectively. Besides PhyloBayes, two alternative molecular dating programs were used to replicate the analyses under distinct implementations of the Bayesian exploration of clock-relaxed models. The standard WAG + G and GTR + G models were used in MCMCTree and BEAST for analysing amino-acid and nucleotide datasets, respectively. In MCMCTree, we used the ML approximation by first calculating the ML estimates of the branch lengths, the gradient vector and Hessian matrix, using the BaseML and CodeML programs of PAML. Despite the fact that auto-correlated models of clock relaxation have been shown to provide a significantly better fit than uncorrelated models on phylogenomic datasets [56,57], all analyses were conducted under both models of molecular clock relaxation for comparison purposes. As BEAST implements only uncorrelated relaxed clock models, we used the uncorrelated \log_{normal} model.

Six fossil calibrations compatible with our tree were selected from Benton *et al.* [87]: (1) *Xenopus/Homo* (350 Myr to 330 Myr), (2) *Gallus/Homo* (330 to 312), (3) *Anolis/Gallus* (300 to 256), (4) *Ornithorhynchus/Homo* (191 to 163), (5) *Monodelphis/Homo* (171 to 124), and (6) *Gallus/Taeniopygia* (87 to 66). These calibration constraints were used with soft bounds [88] under a birth-death prior in PhyloBayes and MCMCTree, because this strategy has been shown to provide the best compromise for dating estimates [89]. The prior on the root age corresponding to the *Protopterus/Homo* split was set at 419 to 408 Myr [90]. In BEAST, we used normal distributions with 95% confidence intervals covering these constraints as calibration priors with a birth-death process on the tree.

In PhyloBayes, all calculations were conducted by running two independent MCMC chains for 20,000 cycles, sampling posterior rates and dates every 10 cycles until 2000 points were collected. Posterior estimates of

divergence dates were then computed from the last 1800 samples of each chain after accounting for the initial burn-in period (10%). In MCMCTree, two independent MCMC chains were run with the following parameters: burn in = 1,000,000; sampling frequency = 100; number of samples = 10,000,000. The first 1,000,000 iterations were thus discarded as burn-in, and then the MCMC was run for 100,000,000 iterations, sampling every 100 iterations. The 10,000,000 samples were then summarized to estimate mean divergence date and associated 95% credibility intervals. Finally, BEAST was set up using a single MCMC run for 5,000,000 and 10,000,000 generations for analysing the amino-acid and nucleotide datasets, respectively. Each chain was sampled every 10,000 generations to generate 5,000 and 10,000 samples, of which the first 10% were excluded as the burn-in before computing the mean divergence time estimates and associated 95% credibility intervals.

Additional material

Additional file 1: Table S1: Detailed results of Bayesian relaxed molecular clock analyses obtained under different uncorrelated models for the eight unconstrained nodes.

Additional file 2: Figure S1: Maximum likelihood analyses of the nucleotide dataset. ML phylograms with branch lengths obtained using RAxML with a single concatenated GTR + G model for analysing (a) the complete nucleotide dataset, (b) codon positions 1 + 2, and (c) third codon positions only.

Additional file 3: Figure S2: Analytical pipeline used for assembling the phylogenomic dataset.

Additional file 4: Table S2: Chicken Ensembl gene IDs and official gene names of the 248 markers used in this study.

Acknowledgements

We thank Marion Ballenghien, Pierre-André Crochet, Amélie Laencina, Cédric Libert, João Lourenço, Samuel Martin, Jean-Baptiste Sénégas, Jean-Yves Sire, and Marco Zuffi for their help with tissue samples. We are grateful to the Zoo du Lunaret (Ville de Montpellier), La Ferme aux Crocodiles (Pierrelatte), and to the Cestmed of the Seaquarium (Le Grau-du-Roi) for allowing the sampling of the various reptiles, and to the CITES office in Montpellier for help in obtaining sampling permits. We also thank Khalid Belkhir for helping with the initial bioinformatics analyses, and João Lourenço and Julien Claude for helpful discussion on the dating results. Phylogenetic analyses largely benefited from the ISEM computing cluster and the Montpellier Bioinformatics Biodiversity platform. We are grateful to the three anonymous reviewers for their helpful comments. This project was supported by a European Research Council (ERC) grant to Nicolas Galtier (ERC PopPhyl 232971). YC has been partially supported by a FCT postdoctoral fellowship SFRH/BPD/73515/2010. This work is contribution ISEM 2012-083 of the Institut des Sciences de l'Evolution de Montpellier. The cover image was provided by Ylenia Chiari.

Author details

¹Institut des Sciences de l'Evolution, UMR5554-CNRS-IRD, Université Montpellier 2, Montpellier, France. ²CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.

Authors' contributions

NG, FD and YC conceived and designed the study. YC and FD collected biological material. YC carried out RNA extractions and coordinated

transcriptomes sequencing. VC performed sequence database construction, contigs assembly, and BRH calculations. FD and YC constructed the phylogenomic dataset. FD conducted the phylogenetic and dating analyses. YC, NG and FD wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 26 November 2011 Accepted: 27 July 2012

Published: 27 July 2012

References

- Rieppel O: **The relationships of turtles within amniotes.** In *Biology of turtles*. Edited by: Wyneken J, Godfrey MH, Bels V. Boca Raton, FL: CRC Press, Taylor and 2007:345-353.
- Burke AC: **Development of the turtle carapace: implications for the evolution of a novel bauplan.** *J Morphol* 1989, **199**:363-378.
- Burke AC: **Turtles again.** *Evol Dev* 2009, **11**:622-624.
- Rieppel O: **Turtles as hopeful monsters.** *Bioessays* 2001, **23**:987-991.
- Rieppel O: **How did the turtle get its shell?** *Science* 2009, **325**:154-155.
- Theißen G: **The proper place of hopeful monsters in evolutionary biology.** *Theory Biosci* 2006, **124**:349-369.
- Theißen G: **Saltational evolution: hopeful monsters are here to stay.** *Theory Biosci* 2009, **128**:43-51.
- Kuratani S, Kuraku S, Nagashima H: **Evolutionary developmental perspective for the origin of turtles: the folding theory for the shell based on the developmental nature of the carapacial ridge.** *Evol Dev* 2011, **13**:1-14.
- Gauthier J, Kluge AG, Rowe T: **Amniote phylogeny and the importance of fossils.** *Cladistics* 1988, **4**:105-209.
- Williston SW: **The phylogeny and classification of reptiles.** *J Geol* 1917, **25**:411-421.
- Lee MSY: **Reptile relationships turn turtle.** *Nature* 1997, **389**:245-246.
- Lee MSY: **Molecules, morphology, and the monophyly of diapsid reptiles.** *Contrib Zool* 2001, **70**:1-22.
- Wilkinson M, Thorley J, Benton MJ: **Uncertain turtle relationships.** *Nature* 1997, **387**:466.
- Hill RV: **Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling.** *Syst Biol* 2005, **54**:530-547.
- Rieppel O, deBraga M: **Turtles as diapsid reptiles.** *Nature* 1996, **384**:453-455.
- Rieppel O, Reisz RR: **The origin and early evolution of turtles.** *Annu Rev Ecol Syst* 1999, **20**:1-22.
- Harris S, Pisani D, Gower DJ, Wilkinson M: **Investigating stagnation in morphological phylogenetics using consensus data.** *Syst Biol* 2007, **56**:125-129.
- Hugall AF, Foster R, Lee MS: **Calibration choice, rate smoothing, and the pattern of tetrapod diversification the long nuclear gene RAG-1.** *Syst Biol* 2007, **56**:543-563.
- Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K: **Sister group relationship of turtles to the bird-crocodylian clade revealed by nuclear DNA-coded proteins.** *Mol Biol Evol* 2005, **22**:810-813.
- Janke A, Erpenbeck D, Nilsson M, Arnason U: **The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny.** *Proc Biol Sci* 2001, **268**:623-631.
- Kumazawa Y, Nishida M: **Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles.** *Mol Biol Evol* 1999, **16**:784-792.
- Zardoya R, Meyer A: **Complete mitochondrial genome suggests diapsid affinities of turtles.** *Proc Natl Acad Sci USA* 1998, **95**:14226-14231.
- Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M: **Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes.** *Gene* 2000, **259**:139-148.
- Hedges SB, Poling LL: **A molecular phylogeny of reptiles.** *Science* 1999, **283**:998-1001.
- Kirsch JAW, Mayer GC: **The platypus is not a rodent: DNA hybridization, amniote phylogeny and the palimpsest theory.** *Phil Trans R Soc Lond B* 1998, **353**:1221-1237.
- Mannen H, Li SS: **Molecular evidence for a clade of turtles.** *Mol Phylogenet Evol* 1999, **13**:144-148.
- Becker RE, Valverde RA, Crother BI: **Proopiomelanocortin (POMC) and testing the phylogenetic position of turtles (Testudines).** *J Zool Syst Evol Res* 2011, **49**:148-159.
- Zardoya R, Meyer A: **The evolutionary position of turtles revised.** *Naturwissenschaften* 2001, **88**:193-200.
- Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV: **Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome.** *Proc Natl Acad Sci USA* 2007, **104**:2767-2772.
- Shen XX, Liang D, Wen JZ, Zhang P: **Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles.** *Mol Biol Evol* 2011, **28**:3237-3252.
- Lyson TR, Sperling EA, Heimberg AM, Gauthier JA, King BL, Peterson KJ: **MicroRNAs support a turtle + lizard clade.** *Biol Lett* 2012, **8**:104-107.
- Tzika AC, Helaers R, Schramm G, Milinkovitch MC: **Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Saurospida lineages and the phylogenetic position of turtles.** *Evodevo* 2011, **2**:19.
- Delsuc F, Brinkmann H, Philippe H: **Phylogenomics and the reconstruction of the tree of life.** *Nat Rev Genet* 2005, **6**:361-375.
- Philippe H, Forterre P: **The rooting of the universal tree of life is not reliable.** *J Mol Evol* 1999, **49**:509-523.
- Degnan JH, Rosenberg NA: **Discordance of species trees with their most likely gene trees.** *PLoS Genet* 2006, **2**:e68.
- Kubatko LS, Degnan JH: **Inconsistency of phylogenetic estimates from concatenated data under coalescence.** *Syst Biol* 2007, **56**:17-24.
- Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP: **Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome.** *Mol Phylogenet Evol* 2003, **29**:289-297.
- Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T: **Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes.** *Syst Biol* 1999, **48**:138-152.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H: **Phylogenomics: the beginning of incongruence?** *Trends Genet* 2006, **22**:225-231.
- Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, Baurain D: **Resolving difficult phylogenetic questions: why more sequences are not enough.** *PLoS Biol* 2011, **9**:e1000602.
- Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model.** *BMC Evol Biol* 2007, **7**(Suppl 1):S4.
- Edwards SV: **Is a new and general theory of molecular systematics emerging?** *Evolution* 2009, **63**:1-19.
- Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ: **Acoelomorph flatworms are deuterostomes related to Xenoturbella.** *Nature* 2011, **470**:255-258.
- Sempere LF, Martinez P, Cole C, Baguna J, Peterson KJ: **Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes.** *Evol Dev* 2007, **9**:409-415.
- Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ: **The deep evolution of metazoan microRNAs.** *Evol Dev* 2009, **11**:50-68.
- Fu X, Adamski M, Thompson EM: **Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*.** *Mol Biol Evol* 2008, **25**:1067-1080.
- Castoe TA, de Koning JA, Hall KT, Yokoyama KD, Gu W, Smith EN, Feschotte C, Uetz P, Ray DA, Dobry J, et al: **Sequencing the genome of the Burmese python (*Python molurus bivittatus*) as a model for studying extreme adaptations in snakes.** *Genome Biol* 2011, **12**:406.
- St John JA, Braun EL, Isberg SR, Miles LG, Chong AY, Gongora J, Dalzell P, Moran C, Bed'hom B, Abzhanov A, et al: **Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes.** *Genome Biol* 2012, **13**:415.
- Werneburg I, Sanchez-Villagra MR: **Timing of organogenesis support basal position of turtles in the amniote tree of life.** *BMC Evol Biol* 2009, **9**:82.
- Benson RBJ, Domokos G, Várkonyi PL, Reisz RR: **Shell geometry and habitat determination in extinct and extant turtles (Reptilia: Testudinata).** *Paleobiology* 2011, **37**:547-562.
- Janzen FJ, Krenz JG: **Phylogenetics: Which was first, TSD or GSD? In Temperature-dependent sex determination in Vertebrates.** Edited by: Valenzuela N, Lance VA. Washington, DC: Smithsonian Books; 2004:121-130.

52. Janzen FJ, Phillips PC: Exploring the evolution of environmental sex determination, especially in reptiles. *J Evol Biol* 2006, **19**:1775-1784.
53. Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV: Genome evolution in Reptilia, the sister group of mammals. *Annu Rev Genomics Hum Genet* 2010, **11**:239-264.
54. Matsuda Y, Nishida-Umehara C, Tarui H, Kuroiwa A, Yamada K, Isobe T, Ando J, Fujiwara A, Hirao Y, Nishimura O, et al: Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. *Chromosome Res* 2005, **13**:601-615.
55. Olmo E: Trends in the evolution of reptilian chromosomes. *Integr Comp Biol* 2008, **48**:486-493.
56. Lepage T, Bryant D, Philippe H, Lartillot N: A general comparison of relaxed molecular clock models. *Mol Biol Evol* 2007, **24**:2669-2680.
57. Rehm P, Borner J, Meusemann K, von Reumont BM, Simon S, Hadrys H, Misof B, Burmester T: Dating the arthropod tree based on large-scale transcriptome data. *Mol Phylogenet Evol* 2011, **61**:880-887.
58. Lourenco JM, Claude J, Galtier N, Chiari Y: Dating cryptodiran nodes: Origin and diversification of the turtle superfamily Testudinoidea. *Mol Phylogenet Evol* 2012, **62**:496-507.
59. Lanfear R, Welch JJ, Bromham L: Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol Evol* 2010, **25**:495-503.
60. Danilov IG, Parham JF: A reassessment of some poorly known turtles from the middle Jurassic of China, with comments on the antiquity of extant turtles. *J Vert Paleontol* 2008, **28**:306-318.
61. Gaffney ES: Triassic and early Jurassic turtles. In *The beginnings of the age of dinosaurs*. Edited by: Padian K. Cambridge: Cambridge University Press; 1986:183-187.
62. Gaffney ES, Jenkins FA: The cranial morphology of *Kayentachelys*, an Early Jurassic cryptodire, and the early history of turtles. *Acta Zoologica* 2010, **91**:335-368.
63. Joyce WG, Sterli J: Congruence, non-homology, and the phylogeny of basal turtles. *Acta Zoologica* 2012, **93**:149-159.
64. Near TJ, Meylan PA, Shaffer HB: Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am Nat* 2005, **165**:137-146.
65. Chiari Y, Galtier N: RNA extraction from sauropsids blood: evaluation and improvement of methods. *Amphibia-Reptilia* 2011, **32**:136-139.
66. Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N: Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol Ecol Resour* 2011, **11**:650-661.
67. Castoe TA, Fox SE, Jason de Koning A, Poole AW, Daza JM, Smith EN, Mockler TC, Secor SM, Pollock DD: A multi-organ transcriptome resource for the Burmese Python (*Python molurus bivittatus*). *BMC Res Notes* 2011, **4**:310.
68. Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE, Poole AW, Vemulapalli V, Daza JM, Mockler T, et al: Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol Evol* 2011, **3**:641-653.
69. Kunstner A, Wolf JB, Backstrom N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK, et al: Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol* 2010, **19**(Suppl 1):266-276.
70. Huang X, Madan A: CAP3: A DNA sequence assembly program. *Genome Res* 1999, **9**:868-877.
71. Huang X, Wang J, Aluru S, Yang SP, Hillier L: PCAP: a whole-genome assembly program. *Genome Res* 2003, **13**:2164-2170.
72. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E: EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 2009, **19**:327-335.
73. Ranwez V, Clairo N, Delsuc F, Pourali S, Auberval N, Diser S, Berry V: PhyloExplorer: a web server to validate, explore and query phylogenetic trees. *BMC Evol Biol* 2009, **9**:108.
74. Ranwez V, Harispe S, Delsuc F, Douzery EJP: MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 2011, **6**:e22594.
75. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010, **59**:307-321.
76. Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000, **17**:540-552.
77. Dryad Repository. [http://dx.doi.org/10.5061/dryad.87b01fq0].
78. Stamatakis A: RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006, **22**:2688-2690.
79. Shimodaira H, Hasegawa M: CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 2001, **17**:1246-1247.
80. Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, **19**:1572-1574.
81. Lartillot N, Philippe H: A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 2004, **21**:1095-1109.
82. Lartillot N, Lepage T, Blanquart S: PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 2009, **25**:2286-2288.
83. Liu L, Yu L, Edwards SV: A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol* 2010, **10**:302.
84. Efron B: Nonparametric estimates of standard error - the Jackknife, the Bootstrap and other methods. *Biometrika* 1981, **68**:589-599.
85. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, **24**:1586-1591.
86. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012.
87. Benton MJ, Donoghue PCJ, Asher RJ: Calibrating and constraining molecular clocks. In *The timetree of life*. Edited by: Hedges SB, Kumar S. New York Oxford University Press; 2009:35-86.
88. Yang Z, Rannala B: Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 2006, **23**:212-226.
89. Inoue J, Donoghue PC, Yang Z: The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst Biol* 2010, **59**:74-89.
90. Muller J, Reisz RR: Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* 2005, **27**:1069-1075.
91. Gradstein FM, Ogg JG: The geological time scale. In *The timetree of life*. Edited by: Hedges SB, Kumar S. New York Oxford University Press; 2009:223-230.
92. Kumar S, Hedges SB: TimeTree2: species divergence times on the iPhone. *Bioinformatics* 2011, **27**:2023-2024.

doi:10.1186/1741-7007-10-65

Cite this article as: Chiari et al.: Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology* 2012 **10**:65.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit





Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae)

Fidel Botero-Castro, Marie-ka Tilak, Fabienne Justy, François Catzeflis, Frédéric Delsuc, Emmanuel J.P. Douzery*

Institut des Sciences de l'Evolution, UMR 5554-CNRS-IRD, Université Montpellier 2, Montpellier, France

ARTICLE INFO

Article history:

Received 22 February 2013

Revised 19 June 2013

Accepted 3 July 2013

Available online 10 July 2013

Keywords:

Chiroptera
Illumina
Mitogenomics
Nuclear DNA
Evolution

ABSTRACT

Leaf-nosed bats (Phyllostomidae) are one of the most studied groups within the order Chiroptera mainly because of their outstanding species richness and diversity in morphological and ecological traits. Rapid diversification and multiple homoplasies have made the phylogeny of the family difficult to solve using morphological characters. Molecular data have contributed to shed light on the evolutionary history of phyllostomid bats, yet several relationships remain unresolved at the intra-familial level. Complete mitochondrial genomes have proven useful to deal with this kind of situation in other groups of mammals by providing access to a large number of molecular characters. At present, there are only two mitogenomes available for phyllostomid bats hinting at the need for further exploration of the mitogenomic approach in this group. We used both standard Sanger sequencing of PCR products and next-generation sequencing (NGS) of shotgun genomic DNA to obtain new complete mitochondrial genomes from 10 species of phyllostomid bats, including representatives of major subfamilies, plus one outgroup belonging to the closely-related mormoopids. We then evaluated the contribution of mitogenomics to the resolution of the phylogeny of leaf-nosed bats and compared the results to those based on mitochondrial genes and the RAG2 and VWF nuclear makers. Our results demonstrate the advantages of the Illumina NGS approach to efficiently obtain mitogenomes of phyllostomid bats. The phylogenetic signal provided by entire mitogenomes is highly comparable to the one of a concatenation of individual mitochondrial and nuclear markers, and allows increasing both resolution and statistical support for several clades. This enhanced phylogenetic signal is the result of combining markers with heterogeneous evolutionary rates representing a large number of nucleotide sites. Our results illustrate the potential of the NGS mitogenomic approach for resolving the evolutionary history of phyllostomid bats based on a denser species sampling.

© 2013 Published by Elsevier Inc.

1. Introduction

Among chiropteran families, the Neotropical leaf-nosed bats (family Phyllostomidae, suborder Yangochiroptera; Teeling *et al.*, 2002) are one of the groups that have attracted the most attention from ecological and evolutionary standpoints. This focus of interest stems from their high diversity that encompasses more than 10% of all extant bat species. Moreover, among ca. 160 recognized species of phyllostomids, at least a dozen new species have only recently been described (Dávalos and Corthals, 2008; Gregorin and Ditchfield, 2005; Larsen *et al.*, 2011; Mantilla-Meluk and Baker, 2010; McCarthy *et al.*, 2006; Muchhala *et al.*, 2005; Solari and Baker, 2006; Taddei and Lim, 2010; Velazco, 2005; Velazco *et al.*, 2010;

Woodman, 2007). Yet, cladistic studies based on comprehensive morphological datasets (Wetterer *et al.*, 2000) have encountered difficulties resolving phylogenetic relationships within the family. This might be explained by the astonishing ecological diversification, mirroring the considerable species diversity of phyllostomids, and leading to a number of evolutionary convergences that makes their phylogeny difficult to resolve using morpho-anatomical characters. For instance, it has recently been shown that nectarivory evolved twice independently in distantly related subfamilies (Datzmann *et al.*, 2010; Rojas *et al.*, 2011).

The use of molecular data has provided a clearer picture of their evolutionary history, validating new species discoveries, and refining taxonomic assessments (Larsen *et al.*, 2010a; Redondo *et al.*, 2008; Velazco and Simmons, 2011). Leaf-nosed bats are currently the group of mammals with by far the most comprehensive cytochrome *c* oxidase subunit 1 (COI) gene sampling (Clare *et al.*, 2007), with more than 8000 COI barcode sequences publicly

* Corresponding author. Fax: +33 4 67 14 36 10.

E-mail address: emmanuel.douzery@univ-montp2.fr (E.J.P. Douzery).

available for 117 species (BOLD database as of February 2013). Moreover, molecular phylogenetic studies based on a few standard mitochondrial markers such as 12S and 16S ribosomal RNAs (rRNAs), COI, and cytochrome b (CYTB) (Baker et al., 2003; Hooper et al., 2008a) and some nuclear exons like those of the RAG2, VWF, and BRCA1 genes, as well as the 3'-UTR of the PLCB4 gene, and the short intron of the PEPCK gene (Baker et al., 2003, 2000; Datzmann et al., 2010) have greatly improved our understanding of the phyllostomid phylogeny.

Two striking examples best illustrate the contribution of molecular data in providing a more resolved picture of the evolutionary history of leaf-nosed bats. First, for a long time the genera *Carollia* and *Rhinophylla* were considered close relatives, mainly on the basis of tooth shape similarity (Wright et al., 1999). Molecular data have shown that they actually belong to two different subfamilies (Baker and Bleier, 1971; Baker et al., 2003, 2000; Wright et al., 1999), which might have diverged more than 20 million years ago (Mya) (Datzmann et al., 2010). Second, the genus *Tonatia* has been traditionally circumscribed on the basis of ears shape in the "round-eared bats" group (Porter et al., 2003), whereas molecular data led to the distinction of *Lophostoma* from *Tonatia*. These two genera may even not be close relatives, with *Lophostoma* being phylogenetically closer to the clade including *Phyllostomus* and *Phylloderma* (Hoffmann et al., 2008; Porter et al., 2003). Similarly, divergence times may involve ca. 20 million years (Myr) of separated history (Datzmann et al., 2010).

At present, 56 genera classified into 11 subfamilies have been recognized within phyllostomid bats according to recent molecular phylogenetic studies and taxonomic revisions. Starting from the deepest node of the Phyllostomidae phylogeny as proposed by several authors (Baker et al., 2003; Datzmann et al., 2010; Simmons, 2005; Wetterer et al., 2000), and moving towards more recent branching points, these subfamilies are (i) Macrotoninae (e.g., big-eared bats), (ii) Micronycterinae (e.g., little big-eared bats), (iii) Desmodontinae (vampire bats), (iv) Lonchorhininae (sword-nosed bats), (v) Phyllostominae (spear-nosed bats), (vi) Glossophaginae (long-tongued, long-nosed, long-tailed, single-leaf, and banana bats), (vii) Lonchophyllinae (nectar bats), (viii) Carollinae (e.g., short-tailed leaf-nosed bats), (ix) Glyphonycterinae (e.g., tricolored bats), (x) Rhinophyllinae (e.g., little fruit bats) and (xi) Stenodermatinae (e.g., neotropical fruit-eating bats). These subfamilies have followed contrasting evolutionary paths after they diverged from each other, leading to a great morphological and ecological heterogeneity among their members (Baker et al., 2012; Datzmann et al., 2010; Dávalos and Jansa, 2004; Hoffmann et al., 2008; Hooper et al., 2008a,b; Larsen et al., 2010b; Rojas et al., 2011; Solari et al., 2009). Some phylogenetic relationships among these subfamilies and their members are strongly supported, as for example the deep branching positions of Micronycterinae and Desmodontinae, or else the paraphyly of nectarivorous phyllostomids (Datzmann et al., 2010), for which a convergence in adaptive patterns has also been recently suggested for the CYTB gene (Dávalos et al., 2012). Some other relationships are still contentious because of conflictive results obtained with different markers, as illustrated by the case of the vampyressine bats (Hooper and Baker, 2006; Hooper et al., 2008a; Lim, 2003; Porter and Baker, 2004). Such conflict has been shown to arise from both biological and methodological sources (Dávalos et al., 2012).

One way to improve the phylogenetic resolution within families of vertebrates is to combine the information from several mitochondrial and nuclear markers. Recently, complete mitochondrial genomes have been shown to provide compelling phylogenetic signal for resolving evolutionary relationships within mammalian families as, for example, Ursidae (Krause et al., 2008), Mustelidae (Yu et al., 2011a), Delphinidae (Vilstrup et al., 2011), Elephantidae (Rohland et al., 2007), and Bovidae and Cervidae (Hassanin et al.,

2012; Wada et al., 2010). A recent study within the bird family Icteridae (New World blackbirds) has also demonstrated the enhanced power of complete mitochondrial genomes to resolve phylogenetic relationships, in comparison to the more traditional CYTB and ND2 genes (Powell et al., 2013). Indeed, mitochondrial genomes provide a set of unambiguously orthologous markers evolving under contrasted selective pressures and, thus, at evolutionary rates which are variable among genes (Reyes et al., 1998). Consequently, individual mitochondrial genes can provide phylogenetic signal at different taxonomic levels. Also, mitochondrial genes exhibit faster rates of molecular evolution than nuclear genes, and could thus potentially be informative enough to resolve recent divergence events, a feature reinforced by the fact that their coalescence time is shorter than for nuclear genes (Moore, 1995). However, potential shortcomings of mitochondrial DNA (mtDNA) as a phylogenetic marker are also widely acknowledged: (i) the substitutional saturation potentially reinforced by the location of the mitogenome in a metabolically active and highly oxidative environment, (ii) the detection of selection-like patterns in some taxa and genes (Castoe et al., 2008; Dávalos et al., 2012; Foote et al., 2011; Jiang et al., 2007; Tomasco and Lessa, 2011; Yu et al., 2011b), (iii) the linkage of all mitochondrial markers into a single locus (Galtier et al., 2009), and (iv) the discordance with the actual species tree in cases of introgression and/or hybridization as recently revealed in African fruit bats (Nesi et al., 2011). These advantages and shortcomings signify that mitogenomes should be supplemented by nuclear markers in a combination that has proven successful on the basis of simulations taking in account the phylogenetic resolving power of each kind of marker (Sánchez-Gracia and Castresana, 2012).

Despite the attractiveness of the mitogenome as a marker of the evolutionary history of mammals, its sequencing is not always straightforward. Extensive variability among taxa generates hyper-variable regions, which hinder PCR primer design and standard amplification and sequencing procedures. Building a comparative mitogenomics framework can also become a harder task in largely diversified clades, for which having a wide taxonomic sampling is mandatory to infer reliable macroevolution and diversification patterns. To date, only 18 mitochondrial genomes of bats (as of February 2013) are available in public databases, with only two species of Phyllostomidae belonging to the single genus *Artibeus*: the Jamaican fruit bat (*Artibeus jamaicensis*) (Pumo et al., 1998) and the great fruit-eating bat (*Artibeus lituratus*) (Meganathan et al., 2012). Moreover, recent works have evidenced the need of improving data collection for leaf-nosed bats (Dávalos et al., 2012).

With the improvement of next-generation sequencing (NGS) technologies, acquiring complete mitogenomes is getting easier. The high-coverage capacities of NGS approaches based on short DNA reads appears especially suited to make use of samples from ethanol and/or DMSO preserved tissue collections and museum specimens (Mason et al., 2011; Miller et al., 2009; Rowe et al., 2011). Recent examples in mammals include ancient DNA studies using both 454 sequencing on rhinos (Willerslev et al., 2009) and Illumina technology on mammoths (Enk et al., 2011). To evaluate the contribution of a mitogenomic approach to reconstruct chiropteran phylogeny, and with a special focus on leaf-nosed bats, we sampled representatives of several major phyllostomid clades, and we sequenced the whole mitochondrial genomes of 11 species using both standard Sanger and NGS strategies. We show that the Illumina approach can be easily applied to quickly obtain complete mitochondrial genomes from fresh or frozen tissue samples using a shotgun genome sequencing approach. Using the mitogenomic dataset, we assessed the phylogenetic signal contained in individual mitochondrial genes, we compared it to the signal of two widely used nuclear exons, and we evaluated the effect of different gene combination schemes on tree inference in the context of the phyllostomid phylogeny.

2. Material and methods

2.1. Taxonomic sampling

To evaluate the phylogenetic signal provided by complete mitochondrial genomes in phyllostomid bats, we chose a set of ingroup and outgroup species satisfying the following conditions: (i) inclusion of members of several major subfamilies following the taxonomy proposed by Baker et al. (2003), which ensured covering the major clades of phyllostomid bats; (ii) availability of nuclear gene sequences (here, RAG2 and VWF) for comparative purposes. Thus, we included at least one representative of the seven ingroup families Micronycterinae, Desmodontinae, Phyllostominae, Glossophaginae, Carollinae, Rhinophyllinae, and Stenodermatinae. The four subfamilies Macrotinae, Lonchorhininae, Lonchophyllinae, and Glyphonycterinae were not sampled. Outgroup taxa for Phyllostomidae belonged to the closely-related families Mormoopidae and Mystacinidae, and to the more distantly related Vespertilionidae, Rhinolophidae, and Pteropodidae. Two species (*Bos taurus* and *Canis lupus*) were also used as non-chiropteran laurasiatherian outgroups. Species sampling and accession numbers of all sequences are given in Table 1.

2.2. Sequencing of complete mitochondrial genomes

We used tissue samples preserved in 95% ethanol and stored in the mammalian tissue collection of the Institut des Sciences de l'Evolution de Montpellier (Catzefflis, 1991). Tissue and voucher numbers are provided in Supplementary Table 1. Whole mitochondrial genomes for 10 phyllostomid and one mormoopid bats were sequenced following two strategies.

First, mitogenomes of three phyllostomid species – *Brachyphylla cavernarum*, *Carollia perspicillata* and *Sturnira tildae* – were PCR amplified from genomic DNA extracted with the standard phenol–chloroform protocol. To increase the likelihood of finding efficient

PCR primers, we chose these three species because they were close to the only two phyllostomid bats (*Artibeus* sp.) for which mitochondrial genomes were already available. Primers were designed so that they regularly covered the whole mitogenome in overlapping fragments of at least 100 nucleotides (nt). The list of primers and the protocol of amplification are provided as Supplementary material (Supplementary Material 1 and Supplementary Table 2). The ND3–ND4 region was amplified using the primers proposed by Hofer and Baker (2006). PCR products were then purified with magnetic beads (Agencourt AMPure XP), or, when unspecific bands were present, fragments were extracted from 1% agarose gels and then purified using the GFX™ PCR, DNA and gel band purification kit (GE Healthcare). Purified PCR amplicons were then sequenced in both directions with the BigDye® Terminator v3.1 kit (Applied Biosystems) on an Applied ABI Prism 3130 XL sequencer. Resulting sequences were manually edited and automatically assembled with Sequencher v 4.3 (Gene Code Corporation) and then aligned using the mitochondrial genome of *Artibeus jamaicensis* as a reference. Sequences of protein-coding genes were validated after translation to check for the absence of frameshifts and / or stop codons.

Secondly, to alleviate the PCR amplification difficulties associated with higher sequence divergence of more distantly related taxa, we developed an original protocol to sequence the mitogenomes of eight additional species (seven phyllostomids and one mormoopid) using a high-throughput NGS Illumina approach. For these samples, we used the QIAGEN® Blood Tissue extraction kit to extract total genomic DNA. This method was chosen because it generally results in better DNA yields and purity required for adequate NGS library construction. Indeed, when extracting DNA with the phenol–chloroform protocol, phenol traces can hamper subsequent steps in library construction by altering, for example, the measures of absorbance on which rely DNA concentration measurements. For each sample, an aliquot containing 5 µg of RNA-free total genomic DNA at a concentration of ~200 ng/µl was provided to the GATC-Biotech company (Konstanz, Germany)

Table 1
Species and accession numbers of sequences used in this work. Taxa names in bold correspond to sequences obtained in this study. In the “sequence source” column, numbers correspond to: 1 = Sanger sequencing; 2 = NGS-Illumina sequencing; 3 = Public databases. Corresponding subfamilies are indicated for phyllostomid bat species.

Family	Subfamily	Species	Sequence source	Accession number		
				Mitogenome	RAG2	VWF
Bovidae		<i>Bos taurus</i>	3	NC_006853.1	ENSBTAG 00000031309	ENSBTAG 00000012265
Canidae		<i>Canis lupus familiaris</i>	3	NC_002008.4	ENSCAFG 00000015228	ENSCAFG 00000015228
Pteropodidae		<i>Pteropus vampyrus</i>	3	Ensembl~v57 Scaffold_17814	JN398310.1	JN398278.1
Rhinolophidae		<i>Rousettus aegyptiacus</i>	3	NC_007393.1	EU617927.1	DQ445694.1
		<i>Rhinolophus monoceros</i>	3	NC_005433.1	AF447528.1*	AF447546.1*
Vespertilionidae		<i>Myotis lucifugus</i>	3	Ensembl~v57 Scaffold_144518	AM265673.1	JN415062.1
		<i>Plecotus auritus</i>	3	NC_015484.1	GU328100.1	AB079840.1
Mystacinidae		<i>Mystacina tuberculata</i>	3	NC_006925.1	AY141021.1	AY245421.1
Mormoopidae		<i>Pteronotus rubiginosus</i>	2	HG003312	HG380334	HG380337
Phyllostomidae	Micronycterinae	<i>Micronycteris megalotis</i>	2	HF947304	HG380333	HG380344
	Desmodontinae	<i>Desmodus rotundus</i>	2	HG003310	HG380331	HG380342
	Phyllostominae	<i>Lophostoma silvicolum</i>	2	HG003311	HG380330	HG380340
		<i>Tonatia saurophila</i>	2	HG003315	HG380332	HG380343
		<i>Vampyrum spectrum</i>	2	HG003316	AF316495.1	HG380338
	Glossophaginae	<i>Anoura caudifer</i>	2	HG003307	HG380327	HG380339
		<i>Brachyphylla cavernarum</i>	1	HG003308	HG380328	HG380336
	Carollinae	<i>Carollia perspicillata</i>	1	HG003309	HG380329	HG380335
	Rhinophyllinae	<i>Rhinophylla pumilio</i>	2	HG003313	HG380326	HG380345
	Stenodermatinae	<i>Artibeus jamaicensis</i>	3	NC_002009.1	FN641674.1	FN645666.1
		<i>Sturnira tildae</i>	1	HG003314	HG380325	HG380341

* Given that for *Rhinolophus monoceros* there are not any sequences available for RAG2 and VWF genes, accession numbers correspond to those available for *Rhinolophus creaghi*.

for library construction and sequencing. A library of tagged genomic DNA fragments was built for each species. Library preparation included DNA fragmentation, sizing, ligation, indexation, and pooling of samples into an equimolar mix for subsequent run on a single Illumina HiSeq2000 lane. Both library construction and sequencing were conducted under the conditions and protocols practiced by GATC-Biotech.

2.2.1. Mitochondrial genomes assembly

Mitochondrial genomes were *de novo* assembled from the Illumina reads as follows. First, raw single-ended 96-nt reads were assembled into contigs using ABySS (Simpson et al., 2009), version 1.2.0. Different assemblies were generated by varying k-mer values from 48 (half length of a read) to 64 (default maximal k-mer size allowed by ABySS). Similarity searches using BLASTN (Altschul et al., 1990) were then performed on non-redundant ABySS contigs to recover mitochondrial DNA-like matches, using the *Artibeus jamaicensis* mitochondrial genome as a query. To account for the potential high divergence of the target sequences, we set the e-value for a positive match to 1e-05. Moreover, since rapidly evolving protein-coding genes can be difficult to recover based on nucleotide similarity, we also performed TBLASTN searches using the 13 *Artibeus jamaicensis* mitochondrial proteins as a query. This strategy, taking advantage of the amino acid translation of subject contigs for all possible reading frames, also increases the probability to find the targeted mitochondrial sequences.

The final assembly of matching contigs into supercontigs was then realized with CAP3 (Huang and Madan, 1999) under default parameters. This method has proven useful for reference-free transcriptome assembly in non-model animals (Cahais et al., 2012). Visualization and minor manual editing of the resulting sequences was done with MUST (Philippe, 1993) and Seaview v4 (Gouy et al., 2010) to obtain the final mitogenomic assemblies. Site coverage was evaluated for each species from the collection of Illumina reads using custom BASH and R scripts available upon request. Read mapping and annotation of the mitogenomes were performed using GENEIOUS® Pro (Drummond et al., 2011). Reads were mapped only if 24 nucleotides consecutively matched the reference sequence, with a maximum 10% of single mismatches over the read length, a minimum of 95% similarity in overlapping regions, and a maximum 10% of indels not exceeding a gap size of 3.

2.3. Sequencing of nuclear genes

We sequenced the RAG2 and VWF nuclear markers in species for which they were not available in public databases (Table 1). The RAG2 gene was amplified using the primers RAG2F12 (fwd): 5'-TAACCATCTAAAAGTGAAGC-3' and RAG2R901 (rev) 5'-GTTTTCTGTTCTTCATTCAC-3' with an annealing temperature of 52 °C. The VWF exon was amplified using two overlapping pairs of primers: (i) VWF50 (fwd): 5'-CCCGTATGTGGAAGACACC-3' and VWF649 (rev): 5'-AGCTGATAATCTCGTCCCTTCG-3' with an annealing temperature of 50 °C; (ii) VWF489 (fwd): 5'-GGGCTGAAGAAGAA-GAAAGTC-3' and VWF1050 (rev): 5'-GCTGTGCTCGGACACGTAC T-3' with an annealing temperature of 51 °C. The amplification protocol was the same as for mitochondrial fragments using the corresponding temperatures of annealing (see Supplementary material). Sequences were aligned using MUSCLE (Edgar, 2004) and verified by eye in Seaview v4 which resulted in final alignments of 1366 sites for RAG2 and 1239 sites for VWF.

2.4. Phylogenetic analyses

Mitochondrial genomes were aligned using the GENEIOUS aligner with default parameters and then corrected by eye to ensure that alignment of protein coding genes was in agreement with

reading frame and to minimize the number of uninformative gaps in the rRNAs, tRNAs and control region (CR) sequences. The total mitogenome alignment comprising 17,680 sites was filtered with GBLOCKS (Castresana, 2000) under the following parameters: a minimum of 11 sequences for a conserved position, a minimum of 17 sequences for a flanking position, a maximum of 5 contiguous non-conserved positions, a minimum of 10 positions for a block, and a maximum of 50% of gaps per position. This step conserved 89% of the original alignment (15,737 unambiguously aligned sites), keeping the entire sequences of protein coding genes, but removing hypervariable regions from rRNAs, tRNAs and CR.

The mitogenome phylogeny of phyllostomid bats was reconstructed using maximum likelihood (ML) and Bayesian inference (BI). Maximum likelihood analyses were performed under the general time reversible (GTR) model of nucleotide exchangeabilities, with a Gamma (Γ) distribution and a fraction of invariable (I) sites to account for the among-site heterogeneity in substitution rates. Statistical support of nodes was measured by bootstrap percentages (BS) after 100 replicates. All these ML analyses were conducted using PAUP* v4.0b10 (Swofford, 2002), with a neighbor-joining starting tree, and a tree bisection–reconnection branch swapping. Bayesian inference was performed with Phylobayes v3.3 (Lartillot et al., 2009). To account for the potential heterogeneity of the substitution pattern among the different regions of the mitogenome, the CAT site-heterogeneous mixture model was used (Lartillot and Philippe, 2004). The model of DNA sequence evolution also incorporated the GTR + Γ options (CAT-GTR + Γ). Statistical support for nodes was measured by the corresponding posterior probabilities (PP). We used a Dirichlet prior for nucleotide frequency profiles, and exponential priors for the nucleotide exchangeability, among-site rate heterogeneity, and branch length parameters. Trees were sampled every 10 cycles until reaching 10,000 trees. The convergence of three independent Markov chains Monte Carlo (MCMC) was evaluated with the *bpcomp* procedure: chains were stopped when the maximum PP difference for a given node among the three chains became less than 0.1 with a burnin of 1000 cycles.

2.5. Single-gene versus combined-gene approaches

In order to compare the phylogenetic signal carried out by each individual mitochondrial and nuclear gene to the one provided after gene concatenation, we performed a series of ML analyses with comparable taxon sampling. Using PAUP* under the GTR + Γ + I model, we inferred the phylogeny of phyllostomid bats for each of the two nuclear genes (RAG2 and VWF) as well as for a total of 17 independent mitochondrial partitions: the 13 protein-coding genes, each of the two rRNAs (12S and 16S), the concatenation of all 22 tRNAs, and the CR. These partitions were extracted from the alignment previously curated with GBLOCKS.

Then, with the aim of comparing the phylogenetic signal of the different mitochondrial and nuclear partitions within the family, we evaluated the bootstrap statistical support for several clades: Phyllostomidae (node A), Phyllostominae (B), Glossophaginae (C) and Stenodermatinae (D). Two nodes corresponding to rapid diversifications were also evaluated to assess the performance of mitochondrial genomes at resolving this kind of event: the Phyllostominae + [Glossophaginae + Carollinae + Rhinophyllinae + Stenodermatinae] clade (E) and the Glossophaginae + [Carollinae + Rhinophyllinae + Stenodermatinae] clade (F).

3. Results

3.1. Sequencing of mitochondrial genomes

Eleven new complete bat mitochondrial genomes were obtained for 10 species of phyllostomids and one species of mormoopids by

using both classical (three species) and NGS Illumina (eight species) sequencing. For the NGS approach, the total number of 96-nt reads ranged from 3,909,139 (*Anoura caudifer*) to 7,725,965 (*Desmodus rotundus*) (Table 2). From these, 0.02% to 0.35% corresponded to mitochondrial reads, and no direct correlation was identified between the total and mitochondrial number of reads. Length of assembled mitochondrial genomes varied between 16,616 (*Pteronotus rubiginosus*) and 16,671 (*Rhinophylla pumilio*) nucleotides. The total length of the three mitogenomes obtained by standard sequencing was 16,647 nt (*S. tildae*), 16,711 nt (*C. perspicillata*) and 16,785 nt (*Brachyphylla cavernarum*). Median coverage – defined as the median of the number of reads covering any given site – varied greatly from one species to another with a minimum of 6× for *Lophostoma silvicolum* and a maximum of 87× for *Micronycteris megalotis*. A drop in coverage is observed for all species – and this is more pronounced for well-covered mitogenomes – in the region corresponding to the WANCY tRNAs cluster (including the replication origin of the light strand, O_L), and at the beginning of the COI and ND4 genes (Fig. 1: arrows). A decrease in coverage is also observed in the tPhe-peripheral domain of the control region due to uncertainty in the number of satellite repeats.

Different sequencing technologies may exhibit biases in base composition (Aird et al., 2011). If not detected, artifacts in base composition can lead to misleading assemblies and potentially erroneous phylogenetic inferences. To check for these potential biases, and as a way to compare the standard and Illumina sequencing outcomes, we evaluated the base composition of

the sequenced mitochondrial genomes (Table 3). A chi-square test with respect to expected base frequencies did not reject the null hypothesis of a homogeneous distribution across taxa, and this regardless of the sequencing method used (p -value <0.001).

3.2. Phylogenetic reconstructions

We compared the phylogenetic signal of mitochondrial genomes versus nuclear exons, and single-gene versus concatenation for both kinds of genes. Phylogenetic signal and statistical support of the nodes of single-gene mitochondrial and nuclear phylogenies varied greatly from one gene to another, and also as a function of the clade under focus (Table 4). A common trend for individual mitochondrial and nuclear genes is the fact that nodes defining relationships among subfamilies were either not recovered or weakly supported (Table 4: nodes E, F). The concatenation of the five mitochondrial markers used in previous works (12S and 16S rRNA, and the COI, CYTB and ND1 protein-coding genes) improved the statistical support as compared to the single-gene approach. However, clades E and F remained weakly supported, and clade B (Phyllostominae) was not recovered.

Expectedly, phylogenies inferred from concatenated mitogenomic sequences appeared better resolved and well supported with BS > 70 and PP > 0.99 for most nodes. This result is observed, with few exceptions, for both ML and BI approaches (see Table 4 and Fig. 2). Complete mitochondrial genomes provided support

Table 2
Statistics associated to the sequencing of mitogenomes using NGS-Illumina technology in 7 phyllostomid and one mormoopid bats.

Species	Total number of reads	Mitochondrial reads		Mitogenome length (nt)	Median coverage
		Number	%		
<i>Anoura caudifer</i>	6,499,175	5400	0.08	16,546	31×
<i>Desmodus rotundus</i>	7,725,965	1495	0.02	16,665	8×
<i>Lophostoma silvicolum</i>	5,560,092	1158	0.02	16,666	6×
<i>Micronycteris megalotis</i>	4,388,964	15,240	0.35	16,589	87×
<i>Pteronotus rubiginosus</i>	3,830,451	9779	0.26	16,616	55×
<i>Rhinophylla pumilio</i>	7,302,121	4,359	0.06	16,671	25×
<i>Tonatia saurophila</i>	4,664,854	13,867	0.30	16,628	79×
<i>Vampyrum spectrum</i>	3,909,179	4242	0.11	16,637	24×

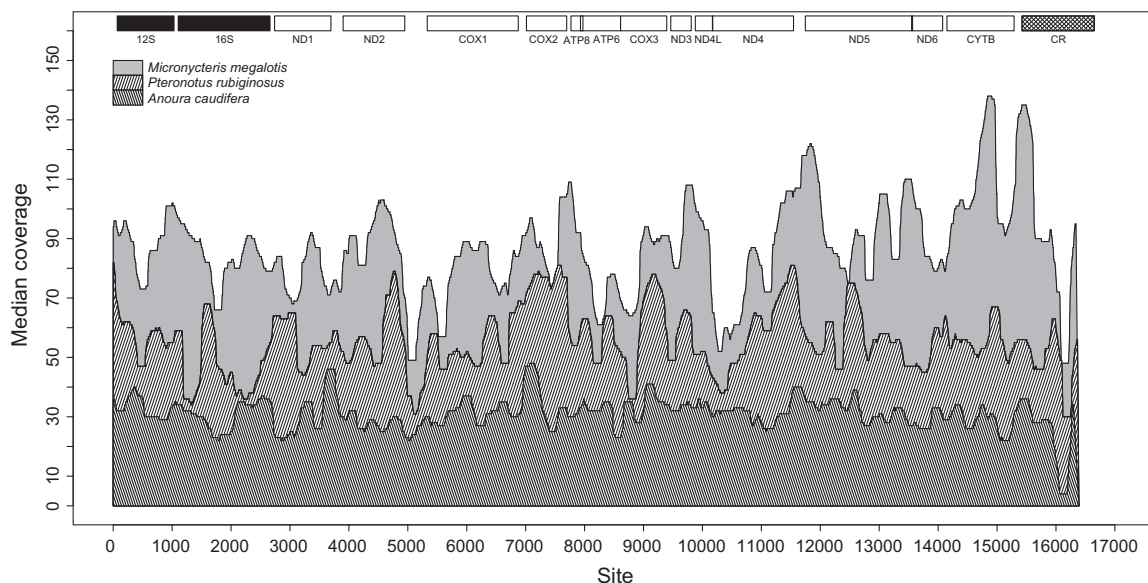


Fig. 1. Coverage for the assembled mitochondrial genomes of *Micronycteris megalotis*, *Pteronotus rubiginosus*, and *Anoura caudifer* using Illumina next-generation sequencing (NGS). The mitochondrial map is indicated on the top: rRNAs in black, protein-coding genes in white, CR in grey, with spaces corresponding to tRNAs.

Table 3

Base composition for the 11 mitochondrial genomes of phyllostomid and mormoopid bats sequenced with either Sanger or Illumina approaches.

Species	%A	%C	%G	%T	Sequencing
<i>Pteronotus rubiginosus</i>	33.6	27.4	13.4	25.6	Illumina
<i>Micronycteris megalotis</i>	32.4	29.3	13.5	24.8	Illumina
<i>Desmodus rotundus</i>	29.7	32.0	15.6	22.7	Illumina
<i>Vampyrum spectrum</i>	30.2	31.1	14.7	24.0	Illumina
<i>Tonatia saurophila</i>	33.5	27.5	12.6	26.4	Illumina
<i>Lophostoma silvicolom</i>	32.2	28.9	13.0	25.9	Illumina
<i>Anoura caudifer</i>	32.2	25.9	13.6	28.3	Illumina
<i>Brachyphylla cavernarum</i>	32.0	27.2	13.8	27.0	Sanger
<i>Carollia perspicillata</i>	31.5	27.7	13.7	27.1	Sanger
<i>Rhinophylla pumilio</i>	31.9	24.9	13.8	29.4	Illumina
<i>Sturnira tildae</i>	32.0	30.0	13.3	24.7	Sanger

for the close relationship between the New-Zealand family Mystacinidae and the New World bats (BS = 97; PP = 1). Within phyllostomids, nodes defining the family and some intra-familial relationships were also strongly supported (BS > 95). This was notably the case for deep divergences of the subfamilies Micronycterinae and Desmodontinae, and for the node grouping the two species of Glossophaginae (Fig. 2).

Nodes with lower support involved the monophyly of Phyllostominae (BS = 67; PP = 1), the monophyly of the two Stenodermatinae species (*A. jamaicensis* and *S. tildae*; BS = 65; PP = 0.99), the grouping of Glossophaginae with the clade Carollinae + Rhinophyllinae + Stenodermatinae (BS = 78; PP = 0.82), and the grouping of Phyllostominae with the clade Glossophaginae + Carollinae + Rhinophyllinae + Stenodermatinae (BS = 68; PP = 0.99). The major split within Chiroptera between Yinpterochiroptera and Yangochiroptera is supported by the ML tree (BS = 100), whereas *Rhinolophus* appears closer to the other echolocating bats than to Pteropodidae under the BI analysis. However, in all cases, clade supports are much higher for mitogenomic analyses than those recovered using the single-gene approach (Table 4).

Individual nuclear genes were able to solve some of the relationships in the phylogeny, but statistical support remained

globally weak, especially within Phyllostomidae (BS < 70). Moreover, nuclear markers were less informative for recent divergences and for rapid diversification events. For instance, the VWF gene resolved deep divergences like those of *Micronycteris megalotis* and *Desmodus rotundus*, and more internal subfamilies like Stenodermatinae, but it was unable to define the branching pattern within Phyllostominae and Glossophaginae. The RAG2 gene, which is less variable than VWF, was able to resolve deep divergences in the chiropteran phylogeny, but failed to decipher the relationships among phyllostomid species. When comparing both topologies, some topological discrepancies can be observed, but the conflicting nodes were weakly supported and often corresponded to short internal branches.

With respect to the single-gene inference, both phylogenetic resolution and node support were improved by the concatenation of the two nuclear markers. However, mitogenomes were still more informative and provided a globally stronger statistical support than the nuclear concatenation (Fig. 3 and Table 4). Finally, when both mitogenomes and nuclear markers were concatenated, there was an increase in statistical support for the nodes that were still weakly supported by the mitogenomic inference alone (Fig. 4).

4. Discussion

4.1. The promises of next-generation mitogenomics for phylogenetics

In the last two decades, the acquisition of a large number of molecular markers has significantly contributed to mammalian phylogenetics and systematics. The development of PCR and sequencing techniques and their subsequent improvements allowed, first, the accumulation of mitochondrial sequences for a few standard markers (e.g., CYTB and CO1 genes) and second, the development and use of single-copy nuclear exons such as RBP3 (Stanhope et al., 1992). However, the efficiency of PCR-based DNA sequencing is impacted by factors such as primer specificity, which is altered by the amount of genetic divergence between taxa, and the limited size of the amplicons, which is directly correlated with

Table 4

Bootstrap support provided by individual and concatenated mitochondrial and nuclear genes for different clades within the family Phyllostomidae. Dashes (–) indicate that the clade was not recovered when using the gene under focus. Values in bold indicate significantly supported clades (i.e., BS > 70). The 12S + 16S + ND1 + COI + CYTB concatenate involve 5 mitochondrial genes classically used for chiropteran phylogenetics. Clade letters refer to Fig. 4: Phyllostomidae (A), Phyllostominae (B), Glossophaginae (C), Stenodermatinae (D), Phyllostominae + (Glossophaginae + Carollinae + Rhinophyllinae + Stenodermatinae) (E), and Glossophaginae + (Carollinae + Rhinophyllinae + Stenodermatinae) (F).

Genes/partitions	Length of the alignment (sites)	Variable sites (%)	Clades					
			A	B	C	D	E	F
12S rRNA	908	42.8	81	–	36	6	–	–
16S rRNA	1440	43.0	86	–	78	88	–	23
ND1	957	51.8	66	6	12	42	6	–
ND2	1044	63.6	51	66	72	16	40	26
COI	1545	41.8	26	7	49	47	–	–
CO2	684	47.8	74	–	77	–	–	26
ATP8	224	60.7	22	–	22	–	8	–
ATP6	681	52.3	–	–	16	–	–	–
CO3	785	45.9	6	14	–	8	5	11
ND3	348	58.0	37	–	–	5	–	–
ND4L	297	57.6	–	–	–	–	7	–
ND4	1378	56.5	68	15	86	–	12	9
ND5	1820	58.5	77	5	70	45	–	10
ND6	528	64.8	60	–	52	79	–	–
CYTB	1140	51.2	46	5	22	–	–	–
CR	554	63.4	94	–	66	41	25	–
22 tRNAs	1492	41.4	6	–	66	87	52	–
12S + 16S + ND1 + COI + CYTB	5990	45.6	100	–	96	98	34	39
Whole mitogenome	15,737	51.3	100	67	100	65	68	78
Nuclear VWF	1239	39.0	95	–	–	–	–	–
Nuclear RAG2	1363	27.2	100	–	95	86	–	–
Nuclear VWF + RAG2	2605	32.8	100	–	48	95	52	19
Mitochondrial + nuclear concatenation	18,342	48.6	100	99	100	100	100	76

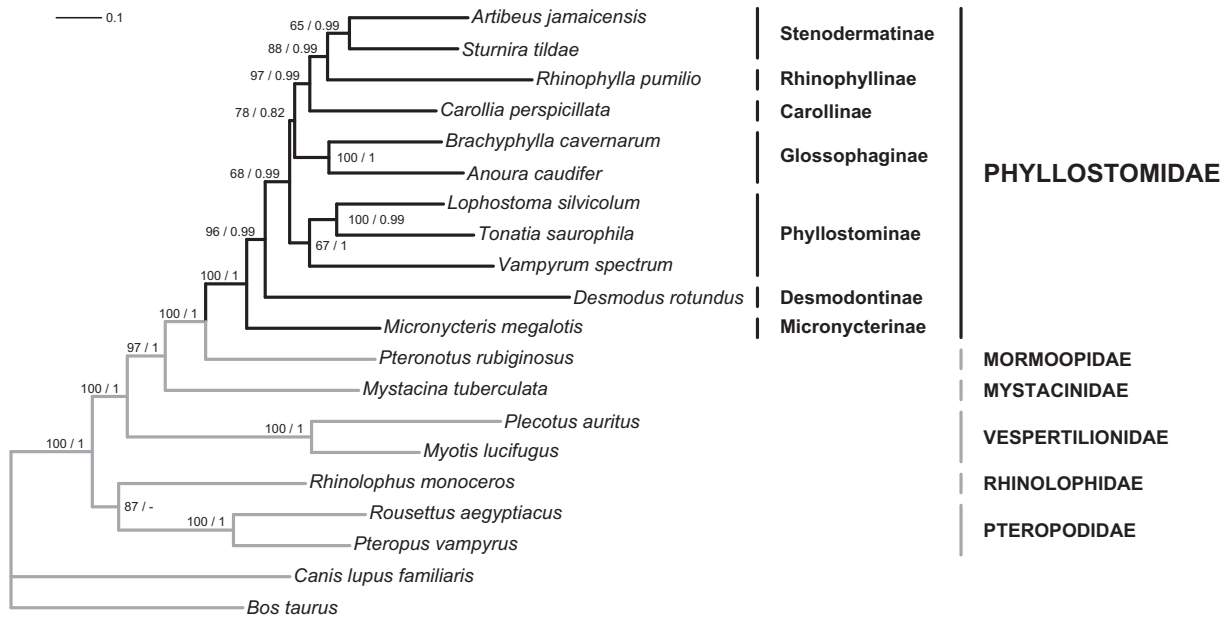


Fig. 2. Phylogeny of phyllostomid bats inferred from complete mitogenomes using maximum likelihood (ML) (a) and Bayesian inference (BI) (b) approaches. Values on nodes correspond to bootstrap support (BS) and posterior probability (PP) values, respectively. Black branches connect the ingroup taxa (Phyllostomidae) and gray branches connect outgroup taxa. For the Phyllostomidae family, subfamilies are indicated. Branch lengths are expressed as a number of nucleotide substitutions per site.

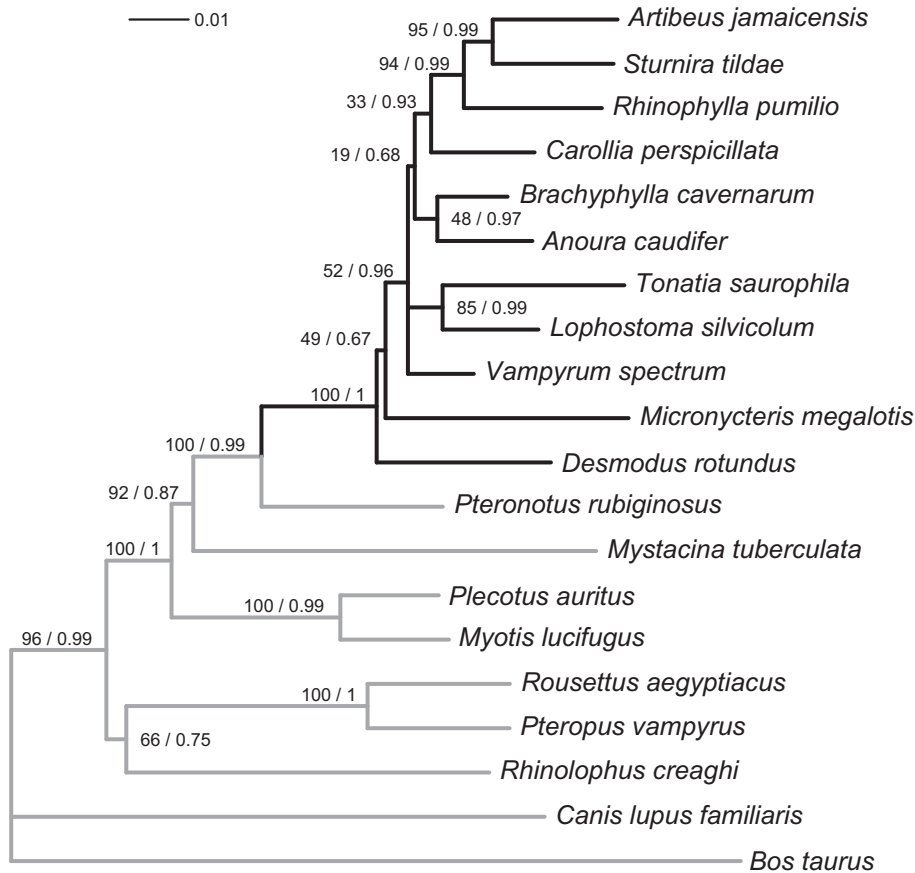


Fig. 3. Maximum likelihood molecular phylogeny of phyllostomid bats inferred from a concatenation of RAG2 and VWF gene sequences. Values on nodes correspond to bootstrap support (BS) and posterior probability (PP) values, respectively.

the quality of the DNA template especially for museum specimens. Besides, several complications can be encountered for the different kinds of targeted markers. The need of specific primers and custom

amplification conditions for each of the taxa under focus, especially when the evolutionary rate of targeted genes and / or their flanking regions is high, renders PCR amplification tedious at a large scale.

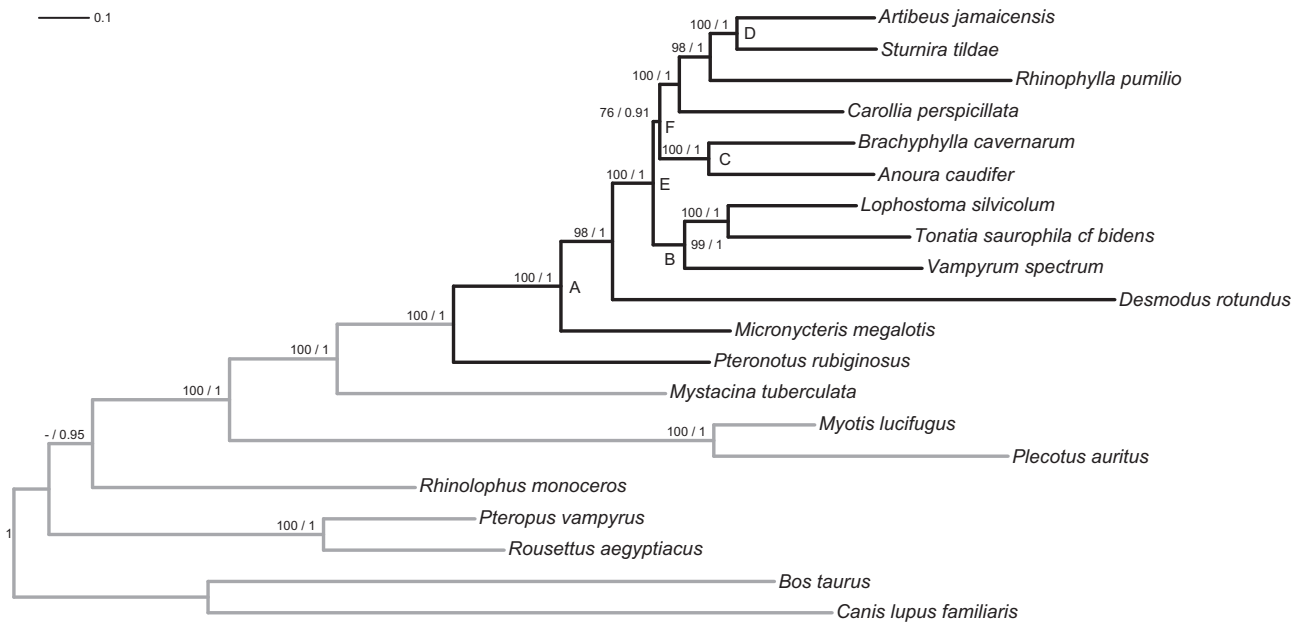


Fig. 4. Maximum likelihood molecular phylogeny of phyllostomid bats inferred from a concatenation of the sequences of complete mitochondrial genomes and the nuclear markers VWF and RAG2. Values on nodes correspond to BS and PP, respectively. Capital letters label the clades used for comparison of the different partitions (see Table 4).

Also, the development of nuclear markers is complicated by the mosaic structure of mammalian genes, which require targeting exons of reasonable length while avoiding the occurrence of large introns (Ranwez et al., 2007). Finally, in the case of mitochondrial markers, a recurrent problem is the co-amplification of nuclear copies of mitochondrial fragments (Hassanin et al., 2010; Richly and Leister, 2004).

Another critical aspect of standard PCR-based sequencing is the amount of DNA required to complete, for instance, a whole mitochondrial genome sequencing, which is highly dependent upon the source of genetic material (tissues, hairs, bones, skin, blood, or saliva). The DNA yielded by samples from museum specimens is often degraded and allowed only short amplicons to be produced (Cooper et al., 1992; Haddrath and Baker, 2001; Rowe et al., 2011; Thomas et al., 1990). To circumvent this problem, procedures of sequence capture and enrichment in mitochondrial fragments have been developed (Enk et al., 2011; Mason et al., 2011; Willerslev et al., 2009). Our results show that a shotgun approach with high-coverage sequencing can also be useful to quickly obtain complete mitogenomes. Based on the sequencing of short DNA fragments, Illumina NGS is especially suited for the analysis of small amount of tissues or samples from museum specimen in which DNA is naturally scarce and / or fragmented. Moreover, the possibility of using tags to label individual sample libraries allows sequencing several taxa simultaneously, offering thus a powerful approach for comparative mitogenomics of highly diversified groups such as phyllostomid bats. Note, however, that alternative approaches without tagging have been explored to assemble highly divergent sequences (Rubinstein et al., 2013; Timmermans et al., 2010).

4.2. The ratio of mitochondrial over nuclear DNA

Given the overrepresentation of mitochondrial DNA compared to nuclear DNA in most tissues (Robin and Wong, 1988; Veltri et al., 1990), the shotgun Illumina NGS approach allows assembling whole mitochondrial genomes with adequate coverage from less than 10 millions reads in most cases. This means that, in theory, 200 millions of reads (i.e., the yield of a single HiSeq lane) can lead

to at least 20 complete mammalian mitochondrial genomes. Taking *Micronycteris megalotis* as an example, we assume that each cell of the initial tissue contained two copies of 2 Gb of nuclear DNA (a typical nuclear genome size in bats: see e.g., *Myotis lucifugus* in Ensembl v.70) versus a number of mitochondria each containing on average two copies of the mitogenome (Robin and Wong, 1988). Under the assumption that the DNA amplification steps during the construction of the Illumina libraries have not biased the initial ratio of mitochondrial to nuclear sequences, and with a mitogenome length of 16,589 bp and 0.35% of mitochondrial reads (Table 2), this yields a putative total of $0.35 \times 2 \times 2.10^9 / (2 \times 16,589 \times 100)$, i.e., ca. 420 mitochondria per cell. This value falls in the range of the estimates available for mammalian cells from the literature (Robin and Wong, 1988). However, the variation of the mitogenome coverage among species (cf. Table 2) can be explained by a variable number of mtDNA circles per mitochondrion and of mitochondria per cell and tissue (Fuke et al., 2011; Robin and Wong, 1988; Veltri et al., 1990).

Given this excess of mitochondria versus nuclei number in each cell, and as a consequence of their relative per-cell number of copies, nuclear elements of mitochondrial origin (numts) are expected to be in a smaller proportion than those corresponding to functional mitochondrial genes (Maricic et al., 2010). In other words, the probability of getting one numt read among millions of other nuclear ones will still be smaller than that of sequencing a true mitochondrial read, except in the case of a massive amplification of numts after their transfer to the nucleus. If numt reads are less abundant and if they display some level of sequence divergence as compared to their mitochondrial counterpart, they are less likely to be assembled in the final mitogenome.

4.3. The assembly and coverage of the mitogenomes

A valuable advantage of NGS comes from the resulting high coverage, which significantly eases mitogenome assembly. With more reads overlapping on a given genomic region, *de novo* assembly strategies are facilitated, even for divergent taxa. Moreover, higher coverage ensures higher sequence reliability as each position is supported by several independent reads with associated quality

score. This allows the identification of reads exhibiting minority polymorphisms likely corresponding to sequencing errors, mitochondrial heteroplasmy and / or numts. In case of limited NGS read coverage, we however identified two factors that can play an important role in the performance and efficiency of the assembly and mapping steps: the phylogenetic position of the closest reference mitogenome, and the evolutionary rate of the targeted taxa.

A striking example of these situations is illustrated by the case of the common vampire bat (*Desmodus rotundus*), whose mitogenome was weakly covered despite the high number reads produced. Indeed, this species not only diverged early within phyllostomid bats (Baker et al., 2003; Datzmann et al., 2010), but its mitogenome has accumulated a higher number of nucleotide substitutions as compared to other phyllostomid bats, as attested by the longer branch observed in inferred phylogenies (Fig. 2). Our initial reference mitogenome was the Jamaican fruit-eating bat (*A. jamaicensis*) since only two *Artibeus* species were available among phyllostomid bats. These species belong to the more recently diverged subfamily of frugivorous bats, the Stenodermatinae. Because of the high sequence divergence between *Desmodus* and *Artibeus*, some regions of the mitogenome – e.g., variable parts of the 12S rRNA, 16S rRNA, and ND5 protein-coding gene – were much more difficult to recover by nucleotide-based similarity searches. The missing regions corresponding to protein coding genes were only identified through the TBLASTN strategy. Otherwise, faster-evolving regions corresponding to rRNAs, tRNAs, and control region can be more tricky to recover using nucleotide BLAST and mapping tools. When coverage is adequate, *de novo* assembly should provide contigs containing these regions. When the sequencing coverage is low, denser taxonomic sampling may provide closer relatives of the group under focus, thus increasing the efficiency of the mapping approach.

About targeted sequences of mitochondrial genes, we expected that on average each fragment would be sequenced in a similar number of copies and that the coverage would be uniform. However, our results show that the Illumina reads coverage is not uniform along the mitochondrial genome (Fig. 1). Some drops in median coverage are observed in the region corresponding to the WANCY tRNA cluster preceding the COI gene, and in the ND4 gene. These observations have also been reported in previous mitogenomic studies of mammals using NGS sequencing (Mason et al., 2011; Rowe et al., 2011). These decreases in read coverage are likely due to the presence of secondary structures of the DNA impairing their efficient sequencing by NGS techniques. For instance, GC content (especially GGC motifs) and inverted repetitions can inhibit single-base elongation (Nakamura et al., 2011). In the nuclear genome, it has also been found that CpG islands and promoters are often less covered than the rest of the genome (Wang et al., 2011).

4.4. The phylogenetic signal of mitogenomes in bats

Single mitochondrial and nuclear genes contribute to resolve relationships at high taxonomic level in bats such as the branching pattern among subfamilies. However, their phylogenetic signal is not sufficient when it comes to resolve more recent nodes like intergeneric relationships within phyllostomid subfamilies (Table 3). Although there are some differences in branching order among the topologies resulting from single-gene analyses, none of them is strongly supported statistically. The conflicting nodes often involved short internal branches therefore reflecting the lack of information rather than true conflicting signals. In the case of the two nuclear markers, RAG2 exhibits a lower proportion of variable sites as compared to VWF. Also, this marker seems to be sensitive to taxon sampling as previous studies considering a larger number

of species obtained better topological resolution and statistical support (Baker et al., 2000).

It has been shown that the concatenation of individual genes into a single supermatrix provides more phylogenetic signal by increasing the amount of informative sites (de Queiroz and Gatesy, 2007; Gadagkar et al., 2005). Recently, it has also been suggested that a combination of nuclear and mitochondrial markers should provide strong phylogenetic signal thanks to the combination of sites with heterogeneous evolutionary rates and, thus, complementary phylogenetic signal (Sánchez-Gracia and Castresana, 2012). Mitochondrial genomes intrinsically offer a concatenation of markers including rRNAs, tRNAs, protein-coding genes and a non-coding control region, which results in a sum of alignments characterized by heterogeneous gene-specific substitution patterns (Reyes et al., 1998). In the context of the phyllostomid phylogeny, Datzmann et al. (2010) built a mitochondrial dataset that combined the 12S and 16S rRNAs, and the protein-coding COI, CYTB, and ND1 genes. We inferred the phylogeny of phyllostomid bats using the same five genes to evaluate the signal of such a concatenation with respect to our taxon sampling. We obtained a fairly well resolved phylogeny with high bootstrap values for several nodes. However, the phylogenetic signal provided by this set of genes was not strong enough to recover several clades, including Phyllostominae (node B, BS = 62), and nodes E and F were only weakly supported (BS = 34 and 39, respectively) (see Table 3). The same nodes remained either unresolved or weakly supported even after correcting for saturation and with increased taxonomic sampling (Baker et al., 2012; Dávalos et al., 2012).

Complete mitogenomes contribute to resolve difficult nodes. Indeed, bootstrap values for nodes such as E and F, which were not well supported by single genes, increased significantly when complete mitogenomic sequences were used. Besides, most nodes of the phylogeny were strongly supported (Fig. 2), even those only supported by nuclear markers. Similarly, a concatenation of the two nuclear genes resulted in an improvement of bootstrap support compared to single-gene trees. Moreover, phylogenetic signal of whole mitochondrial genomes is comparable to that of a concatenation of nuclear and mitochondrial markers, and this holds true not only for topological resolution, but also for associated statistical support values. With respect to chiropterans, the ML analysis, but not the Bayesian one, recovered the major split between Yinpterochiroptera (Rhinolophidae + Pteropodidae) and Yangochiroptera (echolocating bats) (Teeling et al., 2002). Because a single taxon was included as a representative of each family for non-phyllostomid bats, the parameter-rich Bayesian CAT mixture model may lack information to retrieve this result. We also recovered the close relationship of the New Zealand short-tailed bat (*Mystacina tuberculata*) and South American bats (Mormoopidae + Phyllostomidae) (Hoofer et al., 2003; Teeling et al., 2003, 2005). Within Phyllostomidae, we retrieved the deep divergences of *Micronycteris megalotis* and *Desmodus rotundus* (members of the *Micronycterinae* and *Desmodontinae* subfamilies, respectively) (Baker et al., 2003; Datzmann et al., 2010). Similarly, the monophyly of the subfamilies Phyllostominae and Stenodermatinae is strongly supported, but this would need to be confirmed by expanding the taxon sampling. It is however noticeable that the relationships among phyllostomid bats inferred from comparative mitogenomics are in agreement with those based on individual mitochondrial and nuclear genes coupled with denser taxon sampling (Baker et al., 2012, 2003, 2000; Datzmann et al., 2010).

Conflicting branching and the weak support observed for some nodes are problems frequently encountered when working with highly diversified groups, which often include divergence events followed by rapid diversifications. This diversification pattern generates short branches in the phylogeny and the few accumulated

substitutions are not informative enough to clarify the evolutionary history at these nodes. This effect can be clearly observed in topologies resulting from single-gene analyses. Within the phylogeny of phyllostomid bats, this seems to have occurred twice: within the subfamily Phyllostominae and after the divergence of the clade formed by Glossophaginae, Carollinae, Rhinophyllinae and Stenodermatinae (see nodes E and F in Figs. 3 and 4).

Although all mitochondrial genes are linked by the same organismal history, they can efficiently recover the species tree. One advantage of mtDNA lies in its reduced effective population size compared to nuclear genes which makes it more likely to retrace the species evolutionary history (Moore, 1995). Moreover, in our case, a concatenation of mitochondrial genes provided much more variable sites than the two studied nuclear genes. Conversely, a problem of mitogenomics markers is their saturation with respect to multiple substitutions. Their usefulness is therefore restricted to rather low taxonomic scales such as intra-familial relationships in mammals. For instance, some saturated sites have been identified in protein-coding COI and CYTB, and in the loops of the 16S rRNA genes of phyllostomid bats (Dávalos et al., 2012). This underlines the importance of improving taxon and site sampling as well as the models of sequence evolution to better take into account the heterogeneities in the substitution process.

5. Conclusions

We have shown the advantages of coupling a high-throughput sequencing technology like Illumina with a comparative mitogenomics approach as a first step towards resolving the phylogeny of a highly diversified family of bats. This combination allowed assembling a large and informative dataset, which can easily be extended at moderate cost for including a representative sample of phyllostomid species. Moreover, we demonstrated that mitogenomes assembled from NGS data can provide adequate phylogenetic signal for resolving intra-familial relationships in mammals. Finally, the use of mitogenomic sequences is not restricted to barcoding and phylogenetic purposes. Complete mitogenomes can also be used to investigate other challenging questions about their molecular evolution such as the determinants of evolutionary rate variations among taxa (Nabholz et al., 2008) and among genes (Nabholz et al., 2013).

Acknowledgments

This work has been supported by the Agence Nationale de la Recherche “Bio-Informatique” (ANR-10-BINF-01 “Ancestrôme”). We thank all the people who generously contributed to the collection of tissue samples used in this study: Burton Lim, Cynthia Steiner, Omar Linares, Anya Cockle, Marguerite Delaval, Jacques Cuisin and Gilles Peroz. We would also like to thank Hendrik Poinar for constructive discussions about NGS strategies. We are thankful to two anonymous reviewers for their helpful comments. This publication is contribution N° 2013-097 of the Institut des Sciences de l'Évolution de Montpellier (UMR 5554 – CNRS - IRD).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2013.07.003>.

References

- Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Baker, R.J., Bleier, W.J., 1971. Karyotypes of bats of the subfamily Carollinae (Mammalia: Phyllostomatidae) and their evolutionary implications. *Experientia* 7, 220–222.
- Baker, R.J., Porter, C.A., Patton, J.C., Van Den Bussche, R.A., 2000. Systematics of bats of the family Phyllostomidae based on RAG2 DNA sequences. *Occas. Papers Mus. Texas Tech Univ.* 202, i+1–16.
- Baker, R.J., Hofer, S.R., Porter, C.A., Van Den Bussche, R.A., 2003. Diversification among New World leaf-nosed bats: An evolutionary hypothesis and classification inferred from digenomic congruence of DNA sequence. *Occas. Papers Mus. Texas Tech Univ.* 230, i+1–32.
- Baker, R.J., Bininda-Emonds, O.R.P., Mantilla-Meluk, H., Porter, C.A., Van Den Bussche, R.A., 2012. Molecular timescale of diversification of feeding strategy and morphology in New World leaf-nosed bats (Phyllostomidae): a phylogenetic perspective. In: Gunnell, G.F., Simmons, N.B. (Eds.), *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, Cambridge, pp. 385–409.
- Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., Galtier, N., 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Res.* 12, 834–845.
- Castoe, T.A., Jiang, Z.J., Gu, W., Wang, Z.O., Pollock, D.D., 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE* 3, e2201.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Catzeffis, F.M., 1991. Animal tissue collections for molecular genetics and systematics. *Trends Ecol. Evol.* 6, 168.
- Clare, E.L., Lim, B.K., Engstrom, M.D., Eger, J.L., Hebert, P.D.N., 2007. DNA barcoding of neotropical bats: species identification and discovery within Guyana. *Mol. Ecol. Notes* 7, 184–190.
- Cooper, A., Mourer-Chauviré, C., Chambers, G.K., von Haeseler, A., Wilson, A.C., Pääbo, S., 1992. Independent origins of New Zealand moas and kiwis. *Proc. Natl. Acad. Sci.* 89, 8741–8744.
- Datzmann, T., von Helversen, O., Mayer, F., 2010. Evolution of nectarivory in phyllostomid bats (Phyllostomidae Gray, 1825, Chiroptera: Mammalia). *BMC Evol. Biol.* 10, 165.
- Dávalos, L.M., Cortals, A., 2008. A new species of Lonchophylla (Chiroptera: Phyllostomidae) from the eastern Andes of northwestern South America. *Am. Mus. Novit.* 3635.
- Dávalos, L.M., Jansa, S.A., 2004. Phylogeny of the Lonchophyllini (Chiroptera: Phyllostomidae). *J. Mammal.* 85, 404–413.
- Dávalos, L.M., Cirranello, A.L., Geisler, J.H., Simmons, N.B., 2012. Understanding phylogenetic incongruence: lessons from phyllostomid bats. *Biol. Rev.* 87, 991–1024.
- de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. *Trends Ecol. Evol.* 22, 34–41.
- Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2011. Geneious v5.5. <<http://www.geneious.com/>>.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Enk, J., Devault, A., Debruyne, R., King, C.E., Treangen, T., O'Rourke, D., Salzberg, S.L., Fisher, D., MacPhee, R., Poinar, H., 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol.* 12, R51.
- Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P., de Fonseca, R.R., 2011. Positive selection on the killer whale mitogenome. *Biol. Lett.* 7, 116–118.
- Fuke, S., Kubota-Sakashita, M., Kasahara, T., Shigeyoshi, Y., Kato, T., 2011. Regional variation in mitochondrial DNA copy number in mouse brain. *Biochim. Biophys. Acta* 1807, 270–274.
- Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Exp. Zool. (Mol. Dev. Evol.)* 304B, 64–74.
- Galtier, N., Nabholz, B., Glémin, S., Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541–4550.
- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.
- Gregorin, R., Ditchfield, A.D., 2005. New genus and species of nectar-feeding bat in the tribe Lonchophyllini (Phyllostomidae: Glossophaginae) from Northeastern Brazil. *J. Mammal.* 86, 403–414.
- Haddrath, O., Baker, A.J., 2001. Complete mitochondrial DNA genome sequence of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. Roy. Soc. Lond. B* 268, 939–945.
- Hassanin, A., Bonillo, C., Nguyen, B.X., Cruaud, C., 2010. Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA* 21, 68–76.
- Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-García, M., Catzeffis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *CR Biol.* 335, 32–50.

- Hoffmann, F.G., Hooper, S.R., Baker, R.J., 2008. Molecular dating of the diversification of Phyllostominae bats based on nuclear and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 49, 653–658.
- Hooper, S.R., Baker, R.J., 2006. Molecular systematics of Vampyressine bats (Phyllostomidae: Stenodermatinae) with comparison of direct and indirect surveys of mitochondrial DNA variation. *Mol. Phylogenet. Evol.* 39, 424–438.
- Hooper, S.R., Reeder, S.A., Hansen, E.W., Van Den Bussche, R.A., 2003. Molecular phylogenetics and taxonomic review of noctilionoid and vespertilionoid bats (Chiroptera: Yangochiroptera). *J. Mammal.* 84, 809–821.
- Hooper, S.R., Flanary, W.E., Bull, R.J., Baker, R.J., 2008a. Phylogenetic relationships of vampyressine bats and allies (Phyllostomidae: Stenodermatinae) based on DNA sequences of a nuclear intron (TSHB-I2). *Mol. Phylogenet. Evol.* 47, 870–876.
- Hooper, S.R., Solari, S., Larsen, P.A., Bradley, R.D., Baker, A.J., 2008b. Phylogenetics of the fruit-eating bats (Phyllostomidae: Artibeina) inferred from mitochondrial DNA sequences. *Occas. Papers. Mus. Tex. Tech. Univ.* 277.
- Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Jiang, Z.J., Castoe, T.A., Austin, C.C., Burbrink, F.T., Herron, M.D., McGuire, J.A., Parkinson, C.L., Pollock, D.D., 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol. Biol.* 7, 123.
- Krause, J., Unger, T., Noçon, A., Malaspinas, A.-S., Kolokotronis, S.-O., Stiller, M., Soibelzon, L., Spriggs, H., Dear, P.H., Briggs, A.W., Bray, S.C., O'Brien, S.J., Rabeder, G., Matheus, P., Cooper, A., Slatkin, M., Pääbo, S., Hofreiter, M., 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.* 8, 220.
- Larsen, P.A., Marchán-Rivadeneira, M.R., Baker, R.J., 2010a. Taxonomic status of Andersen's fruit-eating bat (*Artibeus jamaicensis aequatorialis*) and revised classification of *Artibeus* (Chiroptera: Phyllostomidae). *Zootaxa* 2648, 45–60.
- Larsen, P.A., Marchán-Rivadeneira, M.R., Baker, R.J., 2010b. Natural hybridization generates mammalian lineage with species characteristics. *Proc. Natl. Acad. Sci.* 107, 11447–11452.
- Larsen, P.A., Siles, L., Pedersen, S.C., Kwiecinski, G.G., 2011. A new species of *Miconycteris* (Chiroptera: Phyllostomidae) from Saint Vincent, Lesser Antilles. *Mammal Biol.* 76, 687–700.
- Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109.
- Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288.
- Lim, B.K., 2003. Differentiation and species status of the Neotropical yellow-eared bats *Vampyressa pusilla* and *V. thyone* (Phyllostomidae) with a molecular phylogeny and review of the genus. *Acta Chiropterologica* 5, 15–29.
- Mantillo-Meluk, H., Baker, R.J., 2010. New species of *Anoura* (Chiroptera: Phyllostomidae) from Colombia, with systematic remarks and notes on the distribution of the *A. geoffroyi* complex. *Occas. Papers. Mus. Tex. Tech. Univ.* 292.
- Maricic, T., Whitten, M., Pääbo, S., 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5, e14004.
- Mason, V.C., Li, G., Helgen, K.M., Murphy, W.J., 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Res.* 21, 1695–1704.
- McCarthy, T.J., Albuja, V.L., Alberico, M.S., 2006. A new species of choacoan *Sturmira* (Chiroptera: Phyllostomidae: Stenodermatinae) from western Ecuador and Colombia. *Annls. Carnegie Mus.* 75, 97–110.
- Meganathan, P.R., Pagan, H.J.T., McCulloch, E.S., Stevens, R.D., Ray, D.A., 2012. Complete mitochondrial genome sequences of three bats species and whole genome mitochondrial analyses reveal patterns of codon bias and lend support to a basal split in Chiroptera. *Gene* 492, 121–129.
- Miller, W., Drautz, D.L., Janecka, J.E., Lesk, A.M., Ratan, A., Tomsho, L.P., Packard, M., Zhang, Y., McClellan, L.R., Qi, J., Zhao, F., Gilbert, T.P., Dalén, L., Arsuaga, J.L., Ericson, P.G.P., Huson, D.H., Helgen, K.M., Murphy, W.J., Götherström, A., Schuster, S.C., 2009. The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Res.* 19, 213–220.
- Moore, W.S., 1995. Inferring Phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49, 718–726.
- Muchhala, N., Mena, V.P., Albuja, V.L., 2005. A new species of *Anoura* (Chiroptera: Phyllostomidae) from the Ecuadorian Andes. *J. Mammal.* 86, 457–461.
- Nabholz, B., Glemis, S., Galtier, N., 2008. Strong variations of mitochondrial mutation rate across mammals – the longevity hypothesis. *Mol. Biol. Evol.* 25, 120–130.
- Nabholz, B., Ellegren, H., Wolf, J.B.W., 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Mol. Biol. Evol.* 30, 272–284.
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., Kanaya, S., 2011. Sequence-specific error profile of Illumina sequencers. *Nucl. Acids Res.* 39, e90.
- Nesi, N., Nakouné, E., Cruaud, C., Hassanin, A., 2011. DNA barcoding of African fruit bats (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*. *C R Biol.* 334, 544–554.
- Philippe, H., 1993. MUST: a computer package of management utilities for sequences and trees. *Nucl. Acids Res.* 21, 5264–5272.
- Porter, C.A., Baker, R.J., 2004. Systematics of *Vampyressa* and related genera of phyllostomid bats as determined by cytochrome-b sequences. *J. Mammal.* 85, 126–132.
- Porter, C.A., Hooper, S.R., Van Den Bussche, R.A., Lee, T.E., Baker, R.J., 2003. Systematics of round-eared bats (*Tonatia* and *Lophostoma*) based on nuclear and mitochondrial DNA sequences. *J. Mammal.* 84, 791–808.
- Powell, A.F.L., Barker, F.K., Lanyon, S.M., 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data. *Mol. Phylogenet. Evol.* 66, 69–79.
- Pumo, D.E., Finamore, P.S., Franek, W.R., Phillips, C.J., Tarzami, S., Balzarano, D., 1998. Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other Eutherian mammals. *J. Mol. Evol.* 47, 709–717.
- Ranwez, V., Delsuc, F., Ranwez, S., Belkhir, K., Tilak, M., Douzery, E.J.P., 2007. OrthoMAM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* 7, 241.
- Redondo, R.A.F., Brina, L.P.S., Silva, R.F., Ditchfield, A.D., Santos, F.R., 2008. Molecular systematics of the genus *Artibeus* (Chiroptera: Phyllostomidae). *Mol. Phylogenet. Evol.* 49, 44–58.
- Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* 15, 957–966.
- Richly, E., Leister, D., 2004. NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* 21, 1081–1084.
- Robin, E.D., Wong, R., 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Phys.* 136, 507–513.
- Rohland, N., Malaspinas, A.S., Pollack, J.L., Slatkin, M., Matheus, P., Hofreiter, M., 2007. Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* 5, e207.
- Rojas, D., Vale, A., Ferrero, V., Navarro, L., 2011. When did plants become important to leaf-nosed bats? Diversification of feeding habits in the family Phyllostomidae. *Mol. Ecol.* 20, 2217–2228.
- Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K., Moritz, C.C., 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Mol. Ecol. Resour.* 11, 1082–1092.
- Rubinstein, N.D., Feldstein, T., Shenkar, N., Botero-Castro, F., Griggio, F., Mastrotoato, F., Delsuc, F., Douzery, E.J.P., Gissi, C., Huchon, D., 2013. Deep sequencing of mixed total DNA without barcodes allows efficient assembly of highly plastic Ascidian mitochondrial genomes. *Genome Biol. Evol.* <http://dx.doi.org/10.1093/gbe/evt1081>.
- Sánchez-Gracia, A., Castresana, J., 2012. Impact of deep coalescence on the reliability of species tree inference from different types of DNA markers in Mammals. *PLoS ONE* 7, e30239.
- Simmons, N.B., 2005. Order Chiroptera. In: Wilson, D.E., Reeder, D.M. (Eds.), *Mammal species of the world: a taxonomic and geographic reference*, third ed. Johns Hopkins University Press, Baltimore, pp. 312–529.
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123.
- Solari, S., Baker, R.J., 2006. Mitochondrial DNA sequence, karyotypic, and morphological variation in the *Carollia castanea* species complex (Chiroptera: Phyllostomidae) with a description of a new species. *Occas. Papers. Mus. Tex. Tech. Univ.* 254.
- Solari, S., Hooper, S.R., Larsen, P.A., Brown, A.D., Bull, R.J., Guerrero, J.A., Ortega, J., Carrera, J.P., Bradley, R.D., Baker, R.J., 2009. Operational criteria for genetically defined species: analysis of the diversification of the small fruit-eating bats, *Dermanura* (Phyllostomidae: Stenodermatinae). *Acta Chiropterol.* 11, 279–288.
- Stanhope, M.J., Czelusniak, J., Si, J.-S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. *Mol. Phylogenet. Evol.* 1, 148–160.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Taddei, V.A., Lim, B.K., 2010. A new species of *Chiroderma* (Chiroptera: Phyllostomidae) from Northeastern Brazil. *Braz. J. Biol.* 70, 381–386.
- Teeling, E.C., Madsen, O., Van Den Bussche, R.A., de Jong, W.W., Stanhope, M.J., Springer, M.S., 2002. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proc. Natl. Acad. Sci. USA* 99, 1431–1436.
- Teeling, E.C., Madsen, O., Murphy, W.J., Springer, M.S., O'Brien, S.J., 2003. Nuclear gene sequences confirm an ancient link between New Zealand's short-tailed bat and South American noctilionoid bats. *Mol. Phylogenet. Evol.* 28, 308–319.
- Teeling, E.C., Springer, M.S., Madsen, O., Bates, P., O'Brien, S.J., Murphy, W.J., 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307, 580–584.
- Thomas, W.K., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *J. Mol. Evol.* 31, 101–112.
- Timmermans, M.J., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T., Pons, J., Vogler, A.P., 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucl. Acids Res.* 38, e197.
- Tomasco, I.H., Lessa, E.P., 2011. The evolution of mitochondrial genomes in subterranean caviomorph rodents: Adaptation against a background of purifying selection. *Mol. Phylogenet. Evol.* 61, 64–70.

- Velazco, P.M., 2005. Morphological phylogeny of the bat genus *Platyrrhinus* Saussure, 1860 (Chiroptera: Phyllostomidae) with the description of four new species. *Fieldiana: Zoology*, N.S. 105, 1–54.
- Velazco, P.M., Simmons, N.B., 2011. Systematics and taxonomy of great striped faced-bats of the genus *Vampyroides* Thomas, 1900 (Chiroptera: Phyllostomidae). *Am. Mus. Novit.*, 3710.
- Velazco, P.M., Gardner, A.L., Patterson, B.D., 2010. Systematics of the *Platyrrhinus helleri* species complex (Chiroptera: Phyllostomidae), with descriptions of two new species. *Zool. J. Linn. Soc.* 159, 785–812.
- Veltri, K.L., Espiritu, M., Singh, G., 1990. Distinct genomic copy number in mitochondria of different mammalian organs. *J. Cell. Phys.* 143, 160–164.
- Vilstrup, J.T., Ho, S.Y., Foote, A.D., Morin, P.A., Kreb, D., Krützen, M., Parra, G.J., Robertson, K.M., de Stephanis, R., Verborgh, P., Willerslev, E., Orlando, L., Gilbert, M.T.P., 2011. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol. Biol.* 11, 65.
- Wada, K., Okumura, K., Nishibori, M., Kikkawa, Y., Yokohama, M., 2010. The complete mitochondrial genome of the domestic red deer (*Cervus elaphus*) of New Zealand and its phylogenetic position within the family Cervidae. *Anim. Sci. J.* 81, 551–557.
- Wang, W., Wei, Z., Lam, T.-W., Wang, J., 2011. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Sci. Rep.* 1, srep00055.
- Wetterer, A.L., Rockman, M.V., Simmons, N.B., 2000. Phylogeny of phyllostomid bats (Mammalia: Chiroptera): data from diverse morphological systems, sex chromosomes, and restriction sites. *Bull. Am. Mus. Nat. Hist.* 248, 1–200.
- Willerslev, E., Gilbert, M.T.P., Binladen, J., Ho, S.Y., Campos, P.F., Ratan, A., Tomsho, L.P., de Fonseca, R.R., Sher, A., Kuznetsova, T.V., Nowak-Kemp, M., Roth, T.L., Miller, W., Schuster, S.C., 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evol. Biol.* 9, 95.
- Woodman, N., 2007. A new species of nectar-feeding bat, genus *Lonchophylla*, from western Colombia and western Ecuador (Mammalia: Chiroptera: Phyllostomidae). *Proc. Biol. Soc. Wash.* 120, 340–358.
- Wright, A.J., van Den Bussche, R.A., Lim, B.K., Engstrom, M.D., Baker, R.J., 1999. Systematics of the genera *Carollia* and *Rhinophylla* based on the cytochrome-B gene. *J. Mammal.* 80, 1202–1213.
- Yu, L., Peng, D., Liu, J., Luan, P., Liang, L., Lee, H., Lee, M., Ryder, O.A., Zhang, Y., 2011a. On the phylogeny of Mustelidae subfamilies: analysis of seventeen nuclear non-coding loci and mitochondrial complete genomes. *BMC Evol. Biol.* 11, 92.
- Yu, L., Wang, X., Ting, N., Zhang, Y., 2011b. Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion* 11, 497–503.

SPECIAL ISSUE: NATURE'S MICROBIOME

Convergence of gut microbiomes in myrmecophagous mammals

FRÉDÉRIC DELSUC,^{*,†‡} JESSICA L. METCALF,[‡] LAURA WEGENER PARFREY,[‡] SE JIN SONG,^{‡§} ANTONIO GONZÁLEZ[‡] and ROB KNIGHT^{†‡¶**}

**Institut des Sciences de l'Évolution, UMR 5554-CNRS-IRD, Université Montpellier 2, Montpellier, France, †Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA, ‡Biofrontiers Institute, University of Colorado, Boulder, CO 80309, USA, §Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA, ¶Department of Computer Science, University of Colorado, Boulder, CO 80309, USA, **Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA*

Abstract

Mammals have diversified into many dietary niches. Specialized myrmecophagous (ant- and termite-eating) placental mammals represent a textbook example of evolutionary convergence driven by extreme diet specialization. Armadillos, anteaters, aardvarks, pangolins and aardwolves thus provide a model system for understanding the potential role of gut microbiota in the convergent adaptation to myrmecophagy. Here, we expand upon previous mammalian gut microbiome studies by using high-throughput barcoded Illumina sequencing of the 16S rRNA gene to characterize the composition of gut microbiota in 15 species representing all placental myrmecophagous lineages and their close relatives from zoo- and field-collected samples. We confirm that both diet and phylogeny drive the evolution of mammalian gut microbiota, with cases of convergence in global composition, but also examples of phylogenetic inertia. Our results reveal specialized placental myrmecophages as a spectacular case of large-scale convergence in gut microbiome composition. Indeed, neighbour-net networks and beta-diversity plots based on UniFrac distances show significant clustering of myrmecophagous species (anteaters, aardvarks and aardwolves), even though they belong to phylogenetically distant lineages representing different orders. The aardwolf, which diverged from carnivorous hyenas only in the last 10 million years, experienced a convergent shift in the composition of its gut microbiome to become more similar to other myrmecophages. These results confirm diet adaptation to be a major driving factor of convergence in gut microbiome composition over evolutionary timescales. This study sets the scene for future metagenomic studies aiming at evaluating potential convergence in functional gene content in the microbiomes of specialized mammalian myrmecophages.

Keywords: 16S rRNA sequencing, convergence, gut microbiome, mammals, microbial diversity

Received 20 April 2013; revision received 14 August 2013; accepted 15 August 2013

Introduction

The radiation of extant mammals has resulted in more than 5000 living species, which diversified into a wide variety of diet niches ranging from broadly generalized

to highly specialized (Feldhamer 2007). Host-associated microbiota play a major role in diet specialization across vertebrates (Ley *et al.* 2008a; Karasov *et al.* 2011) to the point that they might be considered an integral part of the phenotype (McFall-Ngai *et al.* 2013). In mammals, high-throughput 16S rRNA gene sequencing of faecal samples from a diversity of species has shown that both diet and phylogeny have driven the evolution

Correspondence: Rob Knight, Fax: +1-303-492 7744; E-mail: rob.knight@colorado.edu

of the gut microbiome (Ley *et al.* 2008b). At large taxonomic scales, diet appears to be a major driving factor, as gut microbiomes have evolved convergently in mammals sharing the same feeding habits (Muegge *et al.* 2011). In particular, large differences in gut microbial communities have been demonstrated among carnivores, omnivores and herbivores. Within herbivores, a clear divide can be seen between foregut fermenters, which consist essentially of artiodactyls, and the phylogenetically diverse hindgut-fermenting taxa, which include horses, capybaras, rabbits and elephants (Ley *et al.* 2008b; Muegge *et al.* 2011). Cases of convergence driven by the herbivorous diet are also found among vertebrates with the folivorous stinkbird or hoatzin (Godoy-Vitorino *et al.* 2011) and herbivorous fish (Sullam *et al.* 2012) showing similarities in gut microbiome composition to ruminant mammals and in invertebrates with ants in which herbivory has also driven the convergence of gut symbionts (Russell *et al.* 2009; Anderson *et al.* 2012). At shallower taxonomic scales, such as within bears (Ley *et al.* 2008b), great apes (Ochman *et al.* 2010) and between bat families (Phillips *et al.* 2012), host phylogeny and microbiome composition seem to have codiverged. For instance, despite its exclusive bamboo diet, the giant panda still hosts a gut microbiome similar to other bears (Ley *et al.* 2008b).

Mammalian myrmecophages (anteater and termite eaters) provide an opportunity for further understanding the mechanisms that drive the evolution of the gut microbiome by disentangling the effects of diet and phylogeny. Whereas more than 200 mammalian species include a significant portion of ants and/or termites in their diet, only 22 of them can be considered specialized myrmecophages eating more than 90% of ants and/or termites (Redford 1987). Most of these species feed opportunistically on both types of social insects, with only a few species specialized on either ants or termites. In fact, the aardwolf is the only true termite-eating specialist. Ant-eating specialists are restricted to arboreal species such as the pygmy anteater (*Cyclopes didactylus*), which is the only anteater species fully specialized on ants (Miranda *et al.* 2009), and the long-tailed pangolin (*Manis tetradactyla*), which probably feeds specifically on arboreal ant species (Redford 1987). Mammalian myrmecophages can thus be considered as a highly specialized group of insectivores (i.e. a term generally used to cover the eating of all terrestrial invertebrates) feeding exclusively on social insects of low nutritional value (Redford & Dorea 1984).

Myrmecophagous mammals represent a textbook example of phenotypic evolutionary convergence (Feldhamer 2007; McGhee 2011) and therefore provide a model system in which to characterize the taxonomic composition of the gut microbiome in convergently

evolved species sharing the same highly specialized diet. Lineages specialized to eat exclusively ants and/or termites have independently evolved multiple times in mammalian evolutionary history: in monotremes, the short-beaked echidna (*Tachyglossus aculeatus*); in marsupials, the numbat (*Myrmecobius fasciatus*); in placental mammals across five different orders, tolpeutine armadillos (*Priodontes*, *Cabassous* and *Tolypeutes*), the three anteater genera (*Cyclopes*, *Myrmecophaga* and *Tamandua*), pangolins (*Manis*), the aardvark (*Orycteropus afer*) and the aardwolf (*Proteles cristata*) (Redford 1987). These five placental lineages represent different degrees of morphological specialization towards myrmecophagy, probably reflecting both phylogenetic constraints and the time since this peculiar feeding habit evolved. Giant armadillos, anteaters aardvarks, and pangolins constitute the most extreme examples of specialized myrmecophagous phenotypes (Reiss 2001). These animals have developed similar but convergent morphological adaptations such as the reduction in or loss of teeth, an elongated muzzle with an extensible tongue, viscous saliva produced by hypertrophied salivary glands and powerful claws used to dig into ant and termite nests. Placental myrmecophages also share a relatively low metabolic rate due to their nutritionally poor diet (McNab 1984). Indeed, exclusively feeding on social insects imposes strong energetic constraints because most of the protein value is locked in their recalcitrant chitin exoskeleton (Redford & Dorea 1984). Chitinase genes are found in mammalian genomes, but their exact role in digestion is still unclear (Bussink *et al.* 2007). However, chitinolytic bacteria are ecologically widespread (Gooday 1990), and some species have been identified in the mammalian digestive tract (Simůnek *et al.* 2001). This raises the hypothesis that specialized myrmecophagous mammals might rely on symbiotic bacteria for degrading chitin exoskeletons to optimize their protein nutritional intake.

Recent molecular phylogenetic advances have unambiguously demonstrated that myrmecophagous placentals are anciently diverged, independent lineages with more than 80 million years separating myrmecophagous xenarthrans (tolpeutine armadillos and anteaters), aardvarks, pangolins and aardwolves (Springer *et al.* 2003; Delsuc *et al.* 2004; Meredith *et al.* 2011). Specialized myrmecophages therefore provide an especially good model for studying convergence of the gut microbiome over large evolutionary timescales. In this work, we expand upon previous studies of placental gut microbiome evolution using high-throughput Illumina barcoded 16S rRNA amplicons of faecal samples from representatives of all placental myrmecophagous lineages and their close relatives. We characterize the taxonomic composition of their gut microbiota with the objective of

answering several fundamental ecological and evolutionary questions: 'How distinct are the microbiomes of ant-eating mammals in terms of taxonomic composition compared with other mammals?' 'Have entire gut bacterial communities been affected by convergence towards myrmecophagy or do they still show signs of phylogenetic constraints?' 'Did myrmecophagous host species independently recruit similar gut microbes?' 'Can we identify bacterial taxa specific to myrmecophages that may be potential chitin degraders?'

Materials and methods

Data acquisition

Faecal samples from myrmecophagous placentals and phylogenetically related species were provided by European (Colchester UK, London UK, Leipzig DE, and Montpellier FR) and US (Atlanta GA, Cincinnati OH, Houston TX, and San Diego CA) zoos. We also used faecal samples of 30 giant anteaters (*Myrmecophaga tridactyla*) and five giant armadillos (*Priodontes maximus*) collected in 2006 in Emas National Park (Brazil) under IBAMA licence no. 02001.00215/07-21 provided by the Brazilian Institute on Environment and Natural Resources for a study using scat detection dogs (Vynne *et al.* 2011). Faeces were also collected in 2012 from 10 nine-banded armadillos (*Dasyurus novemcinctus*) previously captured by staff of the Merritt Island National Wildlife Refuge in Florida (USA). Finally, two wild pichis (*Zaedyus pichiy*) and a pink fairy armadillo (*Chlamyphorus truncatus*) were sampled in 2009 with Research Permit 339/08 issued by the Dirección de Recursos Naturales Renovables of Mendoza Province (Argentina). These last three samples were lyophilized and shipped immediately after collection. For zoo samples, faeces were usually swabbed within 48 h of defaecation. For field samples, swabs were made from previously frozen or lyophilized faecal material. A total of 93 swab samples from 15 species were collected (Table S1, Supporting information).

DNA extraction, PCR amplification and amplicon preparation for sequencing followed the protocols described in Caporaso *et al.* (2012) and can be found on the Earth Microbiome Project (EMP; Gilbert *et al.* 2010) web page (<http://www.earthmicrobiome.org/emp-standard-protocols/>). Briefly, faecal swabs were extracted using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Total genomic DNA was subjected to PCR amplification targeting a ~300-bp fragment of the 16S rRNA variable region 4 (V4) using the universal bacterial primer set 515F/806R, which amplifies bacterial and archaeal 16S genes near universally (e.g. Walters *et al.* 2011; Caporaso *et al.* 2012). Three

replicate PCRs were performed for each DNA sample, and amplicons generated from each set of three reactions were subsequently pooled and quantified using PicoGreen. Negative controls included no template controls for DNA extraction and PCR amplification. Finally, all barcoded amplicons were pooled in equal concentrations for sequencing. The amplicon pool was purified using the MoBio UltraClean PCR Clean-up kit and sequenced on the Illumina MiSeq sequencing platform (MiSeq Control Software 2.0.5 and Real-Time Analysis software 1.16.18) at the BioFrontiers Institute Next-Generation Genomics Facility at University of Colorado, Boulder, USA. We analysed the single-end sequencing read from the 515f primer (GTGCCAGCMGCCGCG GTAA).

16S rRNA data processing and taxonomic assignment

Raw 16S rRNA amplicon sequences were processed using the QIIME suite of software tools (version 1.6.0-dev) (Caporaso *et al.* 2010a). Sequences were demultiplexed and quality-filtered according to default parameters within QIIME. These sequences were then clustered into operational taxonomic units (OTUs) with a sequence similarity threshold of 97% with UCLUST (Edgar 2010) within QIIME. We assigned sequences to OTUs in two ways, first with an open-reference protocol that captures the full diversity within our data set and second with a closed-reference protocol that enables comparison with previously published studies.

For the open-reference approach, we followed the subsampling open-reference protocol with default parameters in QIIME (http://qiime.org/tutorials/open-reference_illumina_processing.html). Briefly, sequence reads were initially clustered against the October 2012 release of the Greengenes (DeSantis *et al.* 2006; McDonald *et al.* 2012) 97% reference data set (<http://greengenes.secondgenome.com>). The majority of sequences, 85%, matched the reference database, and these OTUs received the taxonomic classification standardized in Greengenes. Sequences that did not match the Greengenes data set at 97% were subsequently clustered into de novo OTUs at 97% similarity with UCLUST. The representative sequences of all OTUs were then aligned to the Greengenes reference alignment using PyNAST (Caporaso *et al.* 2010b), and this alignment was used to construct a phylogenetic tree using FastTree (Price *et al.* 2010) within QIIME. This tree was used subsequently for phylogenetically informed diversity analyses. Sequences that did not align to Greengenes with a 70% similarity threshold were assumed to be non-16S and thus artefactual and removed from further analysis. These de novo OTUs were then assigned their

taxonomies to the finest level possible with the RDP classifier (Wang *et al.* 2007) retrained on the Greengenes October 2012 reference data set using an 80% confidence threshold.

The second OTU picking approach utilized closed-reference OTU picking against the Greengenes 97% reference database from 4 February 2011 (Files available at http://qiime.org/home_static/dataFiles.html). Subsequent analyses used the Greengenes reference tree and taxonomy assignments (McDonald *et al.* 2012). These new data were then combined with previously published data sets. Because all closed-reference data sets are processed using the same method and the same references for taxonomic assignments, using a closed-reference approach allows for the combination of different data sets and comparisons across different studies. Our data as well as previously published data are available in the EMP/QIIME database (<http://www.microbio.me/qiime/> and <http://www.microbio.me/emp>). For our 93 samples, the closed-reference and open-reference data sets were compared with Procrustes analysis to assess whether we were recovering the same beta-diversity patterns with each data set. This analysis rotates or transforms the points in one PCoA plot to try to match the corresponding points in the second plot while still preserving the relative distances between points within each plot. The goodness of fit (M^2) and statistical significance of the goodness of fit (P) were then measured to determine the level of correspondence between the two sets. Finally, we also compared proportions of phyla that were not assigned to Greengenes by comparing taxa summary charts. For both OTU picking pipelines, low-abundance OTUs (OTUs representing <0.00005% of the total reads in the data set) were filtered out as recommended for Illumina-generated sequence data (Bokulich *et al.* 2013).

Comparative analyses with other mammals

Source tracking. Before combining our closed-reference data set with additional mammalian data, we controlled for potential contamination by soil and other confounding environmental factors. We suspected that some samples, in particular field-collected samples, may have been contaminated postdefaecation with soil or other environmental bacteria. Also, it has been reported that faecal samples of armadillo species can contain high percentages of soil particles (Anacleto 2007; Vaz *et al.* 2012), and some myrmecophagous animals such as pangolins are fed with an insectivore diet mixed with soil in captivity (Yang *et al.* 2007). Because we are not able to confidently distinguish between these two scenarios, we filtered out samples that either displayed evidence of contamination or assigned to a low number

of taxa previously described from mammal gut communities. To detect potential contamination, we used a Bayesian approach to estimate the proportion of each sample derived from a priori-defined source communities (Knights *et al.* 2011). As source data sets, we included 16S rRNA data from a set of 42 soil samples (Ramirez *et al.* 2010; Eilers *et al.* 2012) sequenced using the same 16S rRNA V4 region primers (Table S2, Supporting information) and a representative and diverse set of mammal gut communities (Muegge *et al.* 2011). We excluded samples with more than 10% assignment to our representative soil community. Because our sample set focused on myrmecophagous mammals, which includes host species with potentially highly unique gut microbiomes, we used a fairly liberal filtering threshold (i.e. very low percentage) and removed samples for which <0.01% of the community assigned to our representative source of mammal gut communities. Our goal with these filtering thresholds was to remove samples that were probably not representative of the host's gut microbiome.

Statistical analyses. We performed comparative analyses of myrmecophagous, insectivorous, omnivorous, folivorous, herbivorous and carnivorous mammals by combining our filtered data set with 39 available samples that were originally sequenced as part of the Muegge *et al.*'s. (2011) study. We resequenced the V4 16S region for 16 of these samples on the Illumina platform using the 515f/806r primer set following EMP protocols as described above (Table S3, Supporting information). Previous work has shown that sequence reads of this length amplified over the V2 or V4 region are sufficient for accurate and repeatable taxonomic identification to at least the family level and for global community characterization (Liu *et al.* 2008; Caporaso *et al.* 2011). Because we had both V2 and V4 16S region sequence data for 16 samples, we compared these data sets to confirm their similarity using a Procrustes analysis of PCoA scores based on unweighted UniFrac distances.

Because we included 23 samples sequenced on the 454 platform (Muegge *et al.* 2011), we rarefied our closed-reference comparative data set at a level of 1100 sequences per sample to avoid biases caused by differences in sequencing depth of samples. The composition of each sample was summarized at various taxonomic levels using QIIME. The Greengenes reference tree was used to perform beta-diversity comparisons by computing phylogeny-based UniFrac distances (Lozupone & Knight 2005) between samples within QIIME. For exploring relationships between host diet and host phylogeny, we then used both principal coordinate analysis (PCoA) on UniFrac distance matrices within

QIIME and phylogenetic network analysis of community similarity (Parks & Beiko 2012) by reconstructing networks from UniFrac distance matrices with Splits-Tree4 (Huson & Bryant 2006) using the neighbour-net agglomerative method (Bryant & Moulton 2004). To assess the relative importance of diet vs. host phylogeny in shaping the gut microbiota, we performed two-factor crossed analysis of similarity (ANOSIM) with the factors diet and host order using the software package PRIMER v6 (Clarke & Gorley 2006). To address the uneven sampling across diet and host order categories, we randomly subsampled even numbers of samples for each category and ran ANOSIM on diet and host order separately within QIIME. This subsampling procedure was repeated 100 times.

Finally, we identified bacterial genera significantly associated with an ant-eating diet using the script `otu_category_significance.py` in QIIME. We performed this analysis using a related-taxon approach where myrmecophagous species were compared with their nonmyrmecophagous relatives, allowing us to better control for the potential effect of phylogenetic distance

between certain myrmecophagous and nonmyrmecophagous species. Four separate comparisons were possible given the taxonomic breadth of our data set: (i) aardvark vs. rock hyrax, (ii) aardwolf vs. nonmyrmecophagous species in the order Carnivora, (iii) giant anteater vs. sloths and (iv) southern tamandua vs. sloths. For each comparison, we used an analysis of variance (ANOVA) to identify bacterial genera that are found at significantly higher relative abundance in myrmecophagous species compared with their nonmyrmecophagous relatives. After correction for multiple comparisons, this provided a list of the taxa potentially associated with myrmecophagous mammals (Table 1), for which we calculated the mean relative abundance found across the different diet classes (Table S5, Supporting information). To ensure that microbial taxa potentially unique to this study were not overlooked, these analyses were performed on both the closed-reference (i.e. only sequences matching those in the Greengenes database are retained) and open-reference (i.e. nonmatching sequence clusters are also retained) data sets, respectively, rarefied to 1100 and 12 000 sequences per sample.

Table 1 Bacterial genera significantly higher in mean relative abundance (RA) in the guts of myrmecophagous (Myr) species than closely related nonmyrmecophagous (NM) species. Taxa in bold are those detected in more than one comparison

Myr species	NM related species	Bacterial genus (Family)	Myr RA	NM RA	P^\dagger
Aardvark	Rock hyrax	None	NA	NA	NA
Aardwolf	Hyena, lion, bush dog, polar bear	<i>Prevotella</i> (Prevotellaceae)	0.1041	0.0025	0.0001
		<i>Streptococcus</i> (Streptococcaceae)	0.1327	0.0005	0.0341
		<i>Dialister</i> (Veillonellaceae)	0.0055	0	0.0021
		<i>Klebsiella</i> (Enterobacteriaceae)	0.01	0	0.0147
		Unclassified (Erysipelotrichaceae)	0.1123	0.0007	0.0111
		Unclassified (Clostridiaceae)	0.1173	0.0216	0.0182
		<i>Faecalibacterium</i> (Ruminococcaceae)	0.025	0.0027	0.0478
		<i>Eubacterium</i> (Lachnospiraceae)	0.0186	0	0.0489
Giant anteater	Linnaeus two-toed sloth, Hoffmann's two-toed sloth	Unclassified (Clostridiaceae)	0.0286	0.0063	0.007
		<i>Streptococcus</i> (Streptococcaceae)	0.3031	0.0023	0.001
		<i>Eubacterium</i> (Erysipelotrichaceae)*	0.0566	0.0036	0.0019
		<i>Bifidobacterium</i> (Bifidobacteriaceae)	0.0251	0.0025	0.039
		<i>Blautia</i> (Lachnospiraceae)*	0.0241	0.0009	0.0086
		Unclassified (Streptococcaceae)*	0.0128	0.0001	0.0009
		<i>Lactococcus</i> (Streptococcaceae)*	0.0124	0.0001	0
		<i>Lactobacillus</i> (Lactobacillaceae)*	0.0052	0.0004	0.0083
		<i>Peptococcus</i> (Peptococcaceae)	0.0048	0	0.0243
		<i>Turcibacter</i> (Erysipelotrichaceae)	0.0075	0.0014	0.0331
		<i>Sarcina</i> (Clostridiaceae)*	0.0013	0	0
Southern tamandua	Linnaeus two-toed sloth, Hoffmann's two-toed sloth	<i>Catenibacterium</i> (Erysipelotrichaceae)	0.0283	0	0.0081
		<i>Bulleidia</i> (Erysipelotrichaceae)	0.0201	0.0002	0.0095
		<i>Blautia</i> (Lachnospiraceae)	0.0173	0.0005	0
		<i>Roseburia</i> (Lachnospiraceae)*	0.0105	0.0003	0.0451
		<i>Peptococcus</i> (Peptococcaceae)	0.0017	0	0.0029
		<i>Collinsella</i> (Coriobacteriaceae)*	0.0015	0	0

*Additional taxa detected using the open-reference data set.

†FDR corrected.

Results

We generated 16S rRNA gene sequences from a total of 93 new faecal samples from 15 mammalian species representing all major ant-eating lineages, including ant-eaters, aardvarks, pangolins and aardwolves, and closely related species such as armadillos and sloths (Table S1). The MiSeq run resulted in 2 030 814 sequence reads of 115–145 bp after quality filtering with default QIIME parameters. One sample failed to amplify (wild giant armadillo sample Pri.max.Seat.05007 with five sequences in total), resulting in a data set of 92 successfully sequenced new mammal gut samples with an average of 22 074 sequences per sample.

Comparison of OTU picking methods

We compared the results from the two different OTU picking protocols to determine what impact they had on the biological conclusions reached from our data set. These methods were the open-reference method (using the Greengenes reference alignment, but allowing for new OTU clusters) and the closed-reference OTU picking (discarding sequences that do not cluster with 97% similarity with the Greengenes reference set). Open-reference OTU picking utilizes more of the sequence data as reads that do not match the reference data set with 97% similarity are clustered into new OTUs, classified to the finest taxonomic level possible and retained in the data set. However, this approach precludes the comparison of data sets sequenced with different primer sets such as the data set from Muegge *et al.* (2011). Closed-reference OTU picking, on the other hand, allows for comparison of data sets sequenced using nonoverlapping primer regions because only the reads that cluster with the reference data set are retained in the analysis. As the reference data set is composed of full-length 16S rRNA sequences, reads generated from different primer sets can be assigned to the same OTU.

Open-reference OTU picking resulted in a total data set of 1 814 980 sequences being assigned to OTUs. The number of assigned sequences per sample ranged from 11 466 to 35 420 with a median of 19 516. Closed-reference OTU picking resulted in ~82% of the open-reference sequences (i.e. 1 481 672 sequences), aligning with Greengenes OTU's. The number of sequences per sample ranged from 8196 to 31 701 with a median of 15 512. Of the 188 336 sequence reads that did not assign to the reference data set, most sequences were assigned to either Firmicutes (42%) or Bacteroidetes (32%) (Fig. S1, Supporting information). The two OTU picking methods nevertheless yielded very similar bacterial communities (Mantel test comparing UniFrac distance matrices from each OTU picking method

$R^2 = 0.95$, $P = 0.001$). A plot of the Procrustes analysis comparing the PCoA based on unweighted UniFrac distances obtained from open-reference and closed-reference OTU assignment methods for the 92 new samples again shows that these are similar (Procrustes $M^2 = 0.041$, $P = 0.001$) and illustrates that the global sample clustering patterns projected on the first two axes of the PCoA match almost exactly (Fig. 1). These results suggest that the closed-reference data set can be used for downstream comparative analyses incorporating critical samples from previously published data sets that also used Greengenes as a reference set (Muegge *et al.* 2011). Thus, we utilized the closed-reference data set for the majority of our analyses.

Filtering out samples that were not probably representative of the host's gut microbiome

Of the 92 new samples successfully sequenced as part of this study, 56 samples resulted in an assignment to a mammalian gut source community of more than 0.01% (Table S4 and Fig. S2, Supporting information). Among these, one wild giant armadillo sample (Pri.max.Seat.04J09) also had a 10% assignment to a soil source community, leaving 55 samples of the initial 92 for

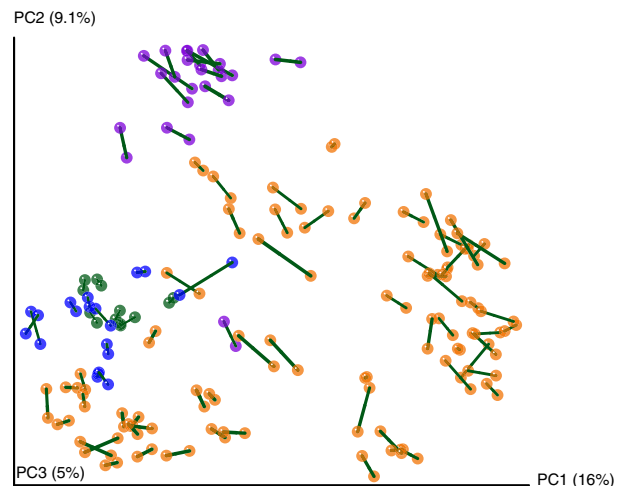


Fig. 1 Procrustes analysis comparing unweighted UniFrac distances for open-reference and closed-reference OTU assignment methods of the 92 mammalian gut samples. This kind of analysis attempts to stretch and rotate the points in one matrix, here obtained by PCoA, to be as close as possible to the points in the other matrix, thus preserving the relative distances between points within each matrix. Circles represent bacterial communities of samples characterized by either open- or closed-reference OTU assignments coloured by diet (blue: omnivore; green: folivore; purple: insectivore; orange: myrmecophage). Each samples' two circles are connected to each other by lines representing the distortion between the two matrices.

subsequent analyses. In total, 5 of the 92 samples resulted in a substantial proportion of their bacterial community assigning to soil. These were field-collected samples from the giant anteater (*Myrmecophaga tridactyla*), giant armadillo (*Priodontes maximus*) and aardvark (*Orycteropus afer*) samples from the Colchester Zoo, for which we suspected soil contamination during sampling. Notably, one of the giant armadillo samples (Pri.max.Seat.3.04L11) assigned 56% of its bacterial community to the source soil data set. Most samples with a low assignment to a mammalian gut community also assigned a high proportion of their community to an 'unknown' source, including 27 samples with 100% assignment to an 'unknown' source. Of these, 19 were giant anteater and giant armadillo faeces located by scat detection dogs in Emas National Park (Brazil). These faecal samples potentially were exposed to a number of contamination sources, including rain and soil. Although we included soil as a source community in our source-tracking analysis, our soil sample set was representative of major soil types and not specific to the field sites or zoo soils associated directly with faecal samples included in this study. Also included in the samples assigning to an 'unknown' source were five aardvark samples from the Colchester Zoo, for which we suspected problems during shipment (i.e. the samples had thawed and stayed at ambient temperature for almost a week), and a wild pichi (*Zaedyus pichiy*) sample. Unfortunately, the only two Chinese pangolin (*Manis pentadactyla*) samples that we managed to collect from the Leipzig Zoo also belonged to this category and were therefore also excluded. Because this source-tracking analysis depended on a closed-reference data set that included a wide taxonomic range of mammals, it is unlikely that full assignment to 'unknown' was due to samples having an entirely unique mammalian gut microbiome, but rather likely due to contamination from sources not included in our source-tracking analysis.

Combining closed-reference mammalian gut data sets

We combined our 55 samples with those from 39 mammals in Muegge *et al.*'s study. (2011), including 23 originally sequenced for the V2 region and 16 samples resequenced for the V4 region. Through a Procrustes analysis, these 16 resequenced samples were confirmed to have similar patterns despite the use of different 16S regions (V2 and V4) and different sequencing technologies (454 and Illumina) (Fig. S3, Supporting information). Therefore, we assembled a comparative data set containing mammalian gut samples from 94 individuals and used only V2 data in cases where no V4 data were available.

We show the taxa relative abundance of the corresponding microbial communities in the phylogenetic context of the 47 mammalian species considered (Fig. 2). Taxa summary charts at the phylum level confirm that mammalian gut microbial communities are dominated by members of a reduced number of bacterial phyla, with Firmicutes, Bacteroidetes, Proteobacteria and Tenericutes being the most abundant in our data set. Large variations in the proportion of these bacterial phyla are seen across the mammalian phylogeny. For example, large proportions of Firmicutes and Bacteroidetes characterize gut microbiomes of Artiodactyla, whereas Firmicutes are mostly predominant in Carnivora. Xenarthran gut microbiomes seem to have relatively high proportions of Proteobacteria, which are especially abundant in anteaters. Within Xenarthra, sloths are characterized by a high abundance of Bacteroidetes, as it is also the case in elephants and hyraxes within Afrotheria.

The comparison of taxa summary plots computed at the family level and ordered by either host taxonomy or host diet (Fig. S4, Supporting information) highlights the confounding effects of phylogeny and diet on mammalian gut microbiome evolution that have been shown in previous studies (Ley *et al.* 2008b; Muegge *et al.* 2011). The two taxa charts display evident cases of clustering of similar gut microbial community composition in species belonging to the same mammalian order or sharing the same diet. For example, herbivorous and folivorous species belonging to different orders (Diprotodontia, Hyracoidea, Proboscidea, Lagomorpha, Perisodactyla and Artiodactyla for herbivores, and Pilosa (sloths) and primates (colobus monkey) for folivores) appear to have similar gut microbiota compositions characterized by high proportions of Bacteroidales and Ruminococcaceae. These phylogenetically diverse species sharing similar diets seem to have converged on gut microbiomes of similar taxonomic compositions confirming herbivory as a major driver of gut microbiome evolution in mammals (Ley *et al.* 2008b).

To further explore the roles of phylogeny and diet in shaping mammal gut microbial communities, we compared beta-diversity based on UniFrac distances computed between the 94 samples. The neighbour-net network reconstructed from the unweighted UniFrac distance matrix revealed interesting patterns of clustering by both host taxonomy and diet (Fig. S5, Supporting information). However, we also noticed in this network a potential effect of zoo vs. wild sampling, with some field-collected samples of giant anteaters, giant armadillos and pichi clustering with the wild nine-banded armadillo samples. This effect, potentially driven by the distinctiveness of the numerous nine-banded armadillo samples, is best visualized in a PCoA of unweighted

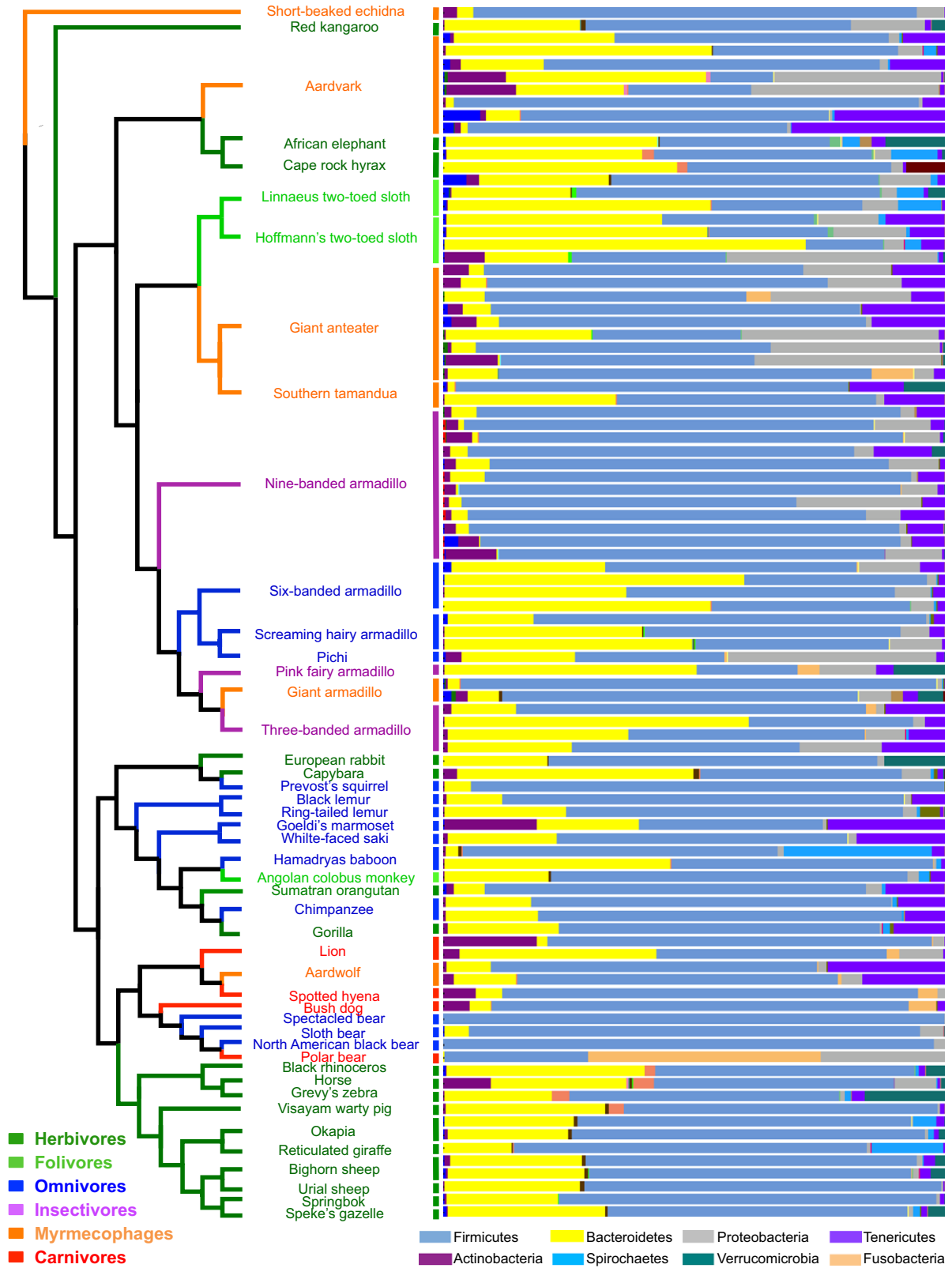


Fig. 2 Phylogeny of the 47 mammalian species represented by their corresponding gut microbiome compositions estimated at the phylum level from a total of 94 faecal samples. The mammalian phylogenetic tree coloured by diet (see inset legend) follows Delsuc *et al.* (2012) for xenarthrans and Meredith *et al.* (2011) for other mammals. The eight most represented bacterial phyla are indicated.

UniFrac distances (Fig. S6, Supporting information). Although this effect appears less important than clustering by taxonomy (ANOSIM host order: $R = 0.628$, $P = 0.001$) or by diet (ANOSIM host diet: $R = 0.394$, $P = 0.001$; Table S6, Supporting information), it is nevertheless statistically significant (ANOSIM captive vs. wild $R = 0.178$, $P = 0.002$). As most samples in the mammalian comparative data set were of captive origins, we excluded field-collected samples to prevent the influence of this potentially confounding variable in downstream analyses. This resulted in a final data set of 69 captive mammals, with all diet categories still represented including myrmecophages, insectivores, carnivores, omnivores, folivores and herbivores.

The neighbour-net network reconstructed from the unweighted UniFrac distance matrix on this reduced data set still shows a clear pattern of clustering by both host diet and taxonomy (Fig. 3). In particular, all

herbivores except the gorilla form a well-defined cluster within which we retrieved the classic divide into two distinct groups corresponding to foregut (artiodactyls) and hindgut (horses, rhinos, capybaras, hyraxes and elephants) fermenters. This analysis also revealed the distinctiveness of the gut microbiomes of the folivorous two-toed sloths (*Choloepus didactylus* and *Choloepus hoffmanni*), which form a well-defined group clearly separated from other mammals, but nevertheless close to herbivores. Primates (lemurs, sakis, baboons, gorillas and chimps) that are primarily omnivorous also group together. Similarly, most members of Carnivora (hyena, bush dog, lions and bears) also belong to a distinct cluster with the notable exceptions of the sloth bear and aardwolves that are well separated. Aardwolves are in fact part of a well-defined cluster of myrmecophagous species together with the large majority of aardvark and anteater samples. Among myrmecophages, only two

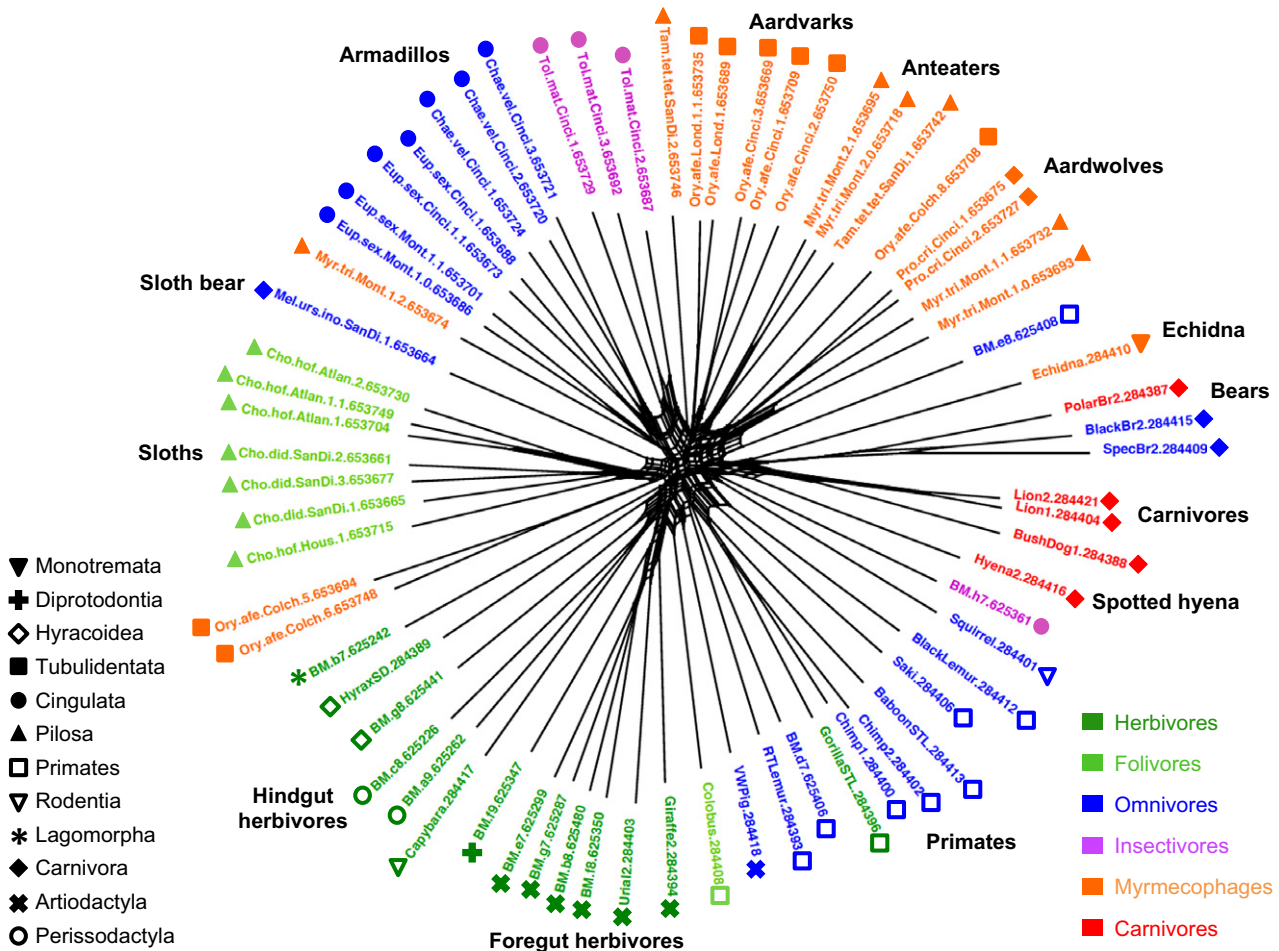


Fig. 3 Neighbour-net network reconstructed from unweighted UniFrac distances between 69 mammalian gut samples collected from captive animals. Samples are coloured by diet, and symbols indicate mammalian orders (see inset legend). Neighbour-net is an agglomerative method based on the neighbour-joining algorithm and is useful for representing the overall structure of the data as a network while accounting for uncertainty.

aardvarks, a giant anteater sample, and the distantly related echidna (*Tachyglossus aculeatus*) appear separated from the other species. Finally, armadillos form another obvious group divided into two distinct clusters corresponding to omnivorous (hairy and six-banded armadillos) and insectivorous (three-banded armadillos) species and it appears close to myrmecophages in the network.

Beta-diversity analysis of this 69-sample captive data set using PCoA of unweighted UniFrac distances confirmed a strong signal for sample clustering by both diet and host order (Fig. 4). The major clusters previously identified in the neighbour-net network are visually evident on the PCoA plot, with herbivores and myrmecophages being distributed at the two extremes of the first axis. Also, sloths appear to form a distinct group, which is nevertheless close to herbivores. This clustering pattern by both diet and host order is significant as assessed by a crossed ANOSIM (diet $R = 0.559$, $P = 0.001$; Table S6, Supporting information; host order $R = 0.62$, $P = 0.001$). However, because our data set contains uneven numbers of diet and host order categories, we randomly subsampled the data so that each category had the same number of samples and reran the ANOSIM 100 times. Our results are robust to uneven sampling as both factors were significant at $P = 0.001$ in all 100 subsample tests. Examining the pairwise ANOSIM results across diet categories shows that myrmecophages are significantly different from most other diets ($P = 0.001$), although a comparison of variances within groups using a Bartlett test (Bartlett 1937) shows that the amount of dispersion among these groups is not equal (Table S6, Supporting information), which can

also lead to statistically significant differences. Insectivores, however, are not significantly different from myrmecophages ($R = 0.196$, $P = 0.137$). Accordingly, in the PCoA (Fig. 4), samples from the insectivorous three-banded armadillo (Order Cingulata, pink circle) appear superimposed with the myrmecophagous aardvark (Order Tubulidentata, orange square) and anteater samples (Order Pilosa, orange triangle). Two other notable results from this beta-diversity analysis are the fact that both the termite-eating aardwolf (*Proteles cristata*) and the sloth bear (*Melursus ursinus*), which includes a significant portion of termites and ants in its diet (Joshi *et al.* 1997), are outliers within Carnivora. The aardwolf firmly clusters with other specialized myrmecophages (anteaters and aardvarks), rather than with other members of Carnivora, including its closest relative the spotted hyena (*Crocuta crocuta*). The sloth bear also appears clearly separated from the other members of Carnivora including its close bear relatives.

Finally, we explored the distinctiveness of myrmecophage gut microbiota composition by performing comparisons with their closest relatives using ANOVA (Table 1). The comparisons of the relative abundance of all bacterial taxa present indicate that the aardwolf gut microbiota is significantly enriched in *Prevotella* (Prevotellaceae), *Streptococcus* (Streptococcaceae), *Dialister* (Veillonellaceae), *Klebsiella* (Enterobacteriaceae), *Faecalibacterium* (Ruminococcaceae), *Eubacterium* (Lachnospiraceae) as well as unclassified genera of Erysipelotrichaceae and Clostridiaceae, as compared to other members of Carnivora. Similar comparisons conducted between anteaters and their sloth relatives also revealed a significant enrichment in *Streptococcus* and other

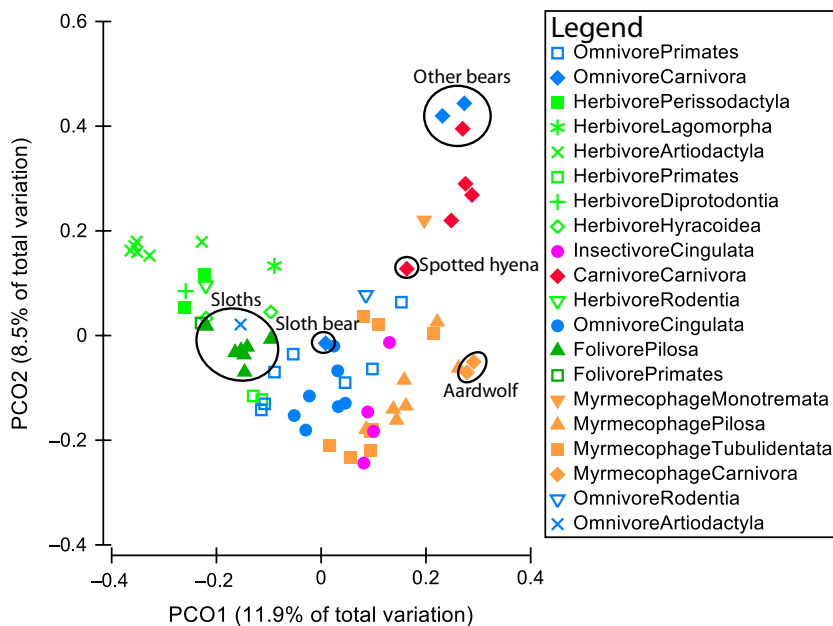


Fig. 4 Mammalian faecal bacterial communities clustered using principal coordinates analysis of the unweighted UniFrac distance matrix. Each point corresponds to a faecal sample coloured according to diet with different symbols corresponding to host order (see legend inset). The respective positions of the sloths, spotted hyena, aardwolf, sloth bear and other bears are indicated. In several cases, host taxa gut microbiomes are clustered by diet. A couple of striking examples include the sloth bear, which has a gut microbiome that is more similar to that of sloths than other bears; and the aardwolf, which has a gut microbiome that clusters more closely with the microbiomes of myrmecophagous host taxa rather than with that of its closest relative, the spotted hyena.

bacterial taxa in anteaters' gut microbiotae. However, when all related-taxon comparisons are considered, very few genera appear to be consistently more abundant in the myrmecophagous than in nonmyrmecophagous species. Furthermore, when mean relative abundances of these bacterial taxa are calculated across all diet groups represented in the data set (Table S5), it becomes less clear as to whether these same taxa may be considered to be uniquely associated with myrmecophagous mammals, with insectivores in general or even more broadly with omnivorous mammals that include insects in their diet. For example, although *Streptococcus* is found to be significantly higher in abundance in aardwolves and giant anteaters compared with their nonmyrmecophagous relatives, it appears to be enriched at similar levels in omnivores as in myrmecophages when compared with the other diet categories (Table S5).

Discussion

Diversity of mammalian gut microbiomes and the problem of wild vs. zoo sampling

Our study is the first large-scale comparative study of the gut microbiome in mammals that includes a significant number of samples collected in the field. Although previous studies reported no effect of captivity (Ley *et al.* 2008b; Muegge *et al.* 2011), we observed systematic differences between wild and zoo specimens. These differences are not altogether unexpected given the large differences in diet that may exist between wild and captive settings. The effect might be especially noticeable in animals for which diet in captivity is markedly different than in nature, which is the case for xenarthrans in general and especially myrmecophagous mammals (Superina 2011). It is also worth noting that of the samples that were detected as containing soil microbes, a large proportion were from wild animals. For example, field-collected samples from giant anteaters and giant armadillos showed signs of soil contamination with notably one of the giant armadillo samples assigning 56% to soil. Although our application of source tracking allowed us to identify faecal samples that may have truly been contaminated with soil, we recognize that a signature of soil microbes may not be an indication of contamination, but rather a natural feature of myrmecophagous mammals feeding on ant and termite nests, as soil is ingested along with their prey (Redford 1987; Anacleto 2007; Vaz *et al.* 2012). For example, soil particles are generally found in most armadillo species faeces (Anacleto 2007) with an estimate of 17% in the nine-banded armadillo (Beyer *et al.* 1994). Also, some ant-eating species kept in captivity, such as the Chinese pangolins we sampled, are main-

tained on a specific diet incorporating additional chitin and soil to facilitate ingestion (Yang *et al.* 2007). Also, we could not exclude the possibility that for field-collected samples of myrmecophages, a certain proportion of the 16S rRNA sequences obtained might actually come from ants and especially termites, which contain a diversity of associated bacteria (Köhler *et al.* 2012).

Diet differences between wild and captive herbivores and omnivores, on the other hand, are probably smaller than for species with more specialized diets such as carnivores and insectivores for which special diets have to be designed in captivity (Yang *et al.* 2007; Superina 2011). Accordingly, only minor differences were reported in studies comparing the gut microbiomes of wild and captive pandas (Zhu *et al.* 2011) and of domestic vs. feral goats (De Jesus-Laboy *et al.* 2011). The few wild samples included in Ley *et al.* (2008b) and Muegge *et al.* (2011) were from omnivorous and herbivorous species, which may explain why the differences were not significant.

Studies of host-associated microbiota in wild vertebrate populations have recently begun to flourish, with host taxa covered to date including primates (Ochman *et al.* 2010; Yildirim *et al.* 2010), North American moose (Ishaq & Wright 2012), bats (Phillips *et al.* 2012), capybaras (Garcia-Amado *et al.* 2012), hoatzins (Godoy-Vitorino *et al.* 2012), iguanas (Lankau *et al.* 2012) and house mice (Linnenbrink *et al.* 2013). These studies have generally revealed diverse gut microbiota in wild populations and the influence of environmental factors such as geography that can affect both interspecific (Phillips *et al.* 2012) and intraspecific variations (Godoy-Vitorino *et al.* 2012; Linnenbrink *et al.* 2013). This suggests that broad sampling across the range of each species might be necessary to gain a full understanding of microbial diversity. For example, in wild populations of house mice, biogeography was identified as the main driving factor of microbiome structure (Linnenbrink *et al.* 2013), and similar patterns have been observed in humans (Yatsunenkov *et al.* 2012), although geographical variation is often also confounded by variation in diet and host genetics.

In our analyses, field-sampled nine-banded armadillos (*Dasypus novemcinctus*) appear to have distinct gut microbiota from other armadillos and from mammals in general. This species is the most widespread xenarthran species ranging from northern Argentina to the southeastern United States and is thus adapted to a wide range of environmental conditions. It is also worth noting that nine-banded armadillos entered Texas only relatively recently (around 1850 A.D.). They quickly become invasive and rapidly spread eastwards to meet founding populations accidentally introduced in Florida later (Taulman & Robbins 1996; Loughry *et al.* 2009). This recent invasion was accompanied by severe genetic

bottlenecks, likely due to successive founder effects (Huchon *et al.* 1999). The individuals we sampled came from the removal programme of an invasive population conducted at Merritt Island National Wildlife Refuge in Florida in 2012. The gut microbiota of these individuals are likely not to be representative of the full range of gut microbiome diversity of this species, as invasive populations have shifted from a mostly myrmecophagous diet in their native range to a more omnivorous diet in the newly colonized northern area (Redford 1986; Sikes *et al.* 1990). There is the possibility that the invasive population may have also shifted its microbiota structure, or, intriguingly, the microbiome may have contributed to its success as an invasive species. Alternatively, the observed distinctiveness of nine-banded armadillo gut microbiota might also reflect a wild vs. captive pattern as suggested by its clustering with other field-collected samples in the neighbour-net analysis (Fig. S5). Our results highlight the need for more comparative sampling in wild specimens to capture the essence of intraspecific variation in the gut microbiome across species range.

Convergence of gut microbiota in myrmecophagous mammals

Our expanded analyses including wild and captive animals and a wider range of diets confirm both diet and phylogeny as major drivers of the microbiome composition as shown in previous studies (Ley *et al.* 2008b; Muegge *et al.* 2011). The previously unstudied myrmecophagous species appear to span a large area of the beta-diversity plots, which highlight marked differences in the composition of their gut bacterial communities. Although the microbiota of specialized myrmecophagous mammals show substantial variation, they cluster significantly with respect to the rest of the mammals. However, we note that in the reduced 69 captive animal data set, statistical significance for comparisons against certain diet types (i.e. myrmecophage vs herbivore and myrmecophage vs carnivore) could be driven by differences in dispersion between groups rather than true clustering (see Table S6, Supporting information). Yet supporting statistical evidence, microbial clustering by diet is also evident in neighbour-net networks where the gut microbiota of myrmecophages appear significantly distinct from that of animals with other diets, except insectivores. Indeed, the gut microbiota of insectivorous three-banded armadillos (*Tolypeutes matacus*) were clustered with that of aardvark and anteaters in the PCoA and appeared close to myrmecophage gut communities in network analyses. The fact that three-banded armadillos are opportunistic insectivores eating substantial quantities of ants and termites (Redford 1985; Bolković *et al.* 1995)

might explain this similarity in gut microbiota composition. However, phylogenetic inertia might also play a role as the gut communities of omnivorous armadillos (*Chaetophractus* and *Euphractus*) cluster with those of the three-banded armadillos in the neighbour-net network. The future inclusion of additional insectivorous taxa from different orders such as bats, shrews and tenrecs will allow disentangling the two effects. These analyses illustrate the utility of network representations for comparing bacterial communities based on measures such as UniFrac as a complement to classical PCoA. They provide a useful alternative to clustering methods such as UPGMA, which do not take into consideration the occurrence of conflicting signals in the data when visualizing the similarity between bacterial communities (Parks & Beiko 2012).

Our results revealed that aardvarks, anteaters and aardwolves possess similar gut microbiota despite representing highly distinct phylogenetic lineages that diverged some 100 million years ago (Mya; Meredith *et al.* 2011). This strongly suggests a major role for diet in driving the convergent composition of the gut microbiome in these specialized myrmecophages, paralleling the case of hindgut fermenter herbivores composed of phylogenetically diverse lineages such as horses, rhinos, capybaras, hyraxes and elephants (Muegge *et al.* 2011). In contrast, groups such as primates and carnivores show signs of phylogenetic inertia, with members of these groups forming well-defined clusters consisting of species with different diets. This clearly is not the case within myrmecophages where anteaters have diverged considerably from their closest relatives in xenarthrans (i.e. sloths), aardvarks appear very distinct from the other afrotherians (i.e. elephants and hyraxes), and the aardwolf has diverged substantially from the other members of Carnivora (hyenas, lions, bush dog and bears). This latter case represents an interesting contrast to the case of the giant panda, which clusters with other Ursidae despite its distinct bamboo diet (Ley *et al.* 2008b).

Both of these cases represent recent transitions to a highly specialized diet in the two distinct groups of Carnivora (Eizirik *et al.* 2010). The aardwolf (*Proteles cristata*) is a strict myrmecophage, almost exclusively feeding on the termite genus *Triniventermes* with some seasonal variations (Kruuk & Sands 1972; Redford 1987). The aardwolf represents the sister lineage to all other hyenas (Koepfli *et al.* 2006) from which it diverged <10 Mya (Eizirik *et al.* 2010). Despite this relatively recent divergence, this species shows marked morphological and physiological differences from carnivorous hyenas, such as a reduced dentition, elongated muzzle and tongue, and digestive tract modifications such as a muscular stomach and relatively short small intestine, to facilitate the rapid processing of termites (Anderson *et al.* 1992).

The giant panda (*Ailuropoda melanoleuca*) diverged from all other ursids some 19 Mya (Krause *et al.* 2008; Eizirik *et al.* 2010), but the fossil record estimate for its dietary shift to a bamboo-rich diet is at least 7 Mya (Jin *et al.* 2007). Despite the fact that the aardwolf and the giant panda present some parallel aspects in their diet adaptation, the trajectories followed by their gut microbiota nevertheless differ greatly. The panda has retained a carnivore-like gut microbiome (Ley *et al.* 2008b) and uses specific genes from its bacterial taxa to facilitate cellulose and lignin degradation, even though their significance for the host physiology has yet to be demonstrated (Zhu *et al.* 2011; Fang *et al.* 2012). In the case of the aardwolf, its gut microbiome appears to have shifted in overall composition to converge with other myrmecophages. In contrast to the giant panda, the sloth bear (*Melursus ursinus*) is an omnivore specialized on ants and termites when available, but also exhibits seasonal variations in its diet (Joshi *et al.* 1997). The sloth bear diverged from other Ursinae around 7 Mya (Krause *et al.* 2008), and morphological adaptations to its myrmecophagous diet include an elongated snout with bare lips and the lack of upper incisors. It is interesting to note that this species is the only member of Carnivora (excluding the aardwolf) that does not cluster with other bears in neighbour-net networks and beta-diversity plots. This suggests that despite its more recent divergence from other bears and its less specialized diet than the panda, the sloth bear experienced an overall shift in the composition of its gut microbiota.

Finally, our results also provide new insights into the evolution of sloth gut microbiota. Sloths are poorly studied arboreal folivores with multichambered stomachs allowing the fermentation of plant material in a manner analogous to foregut fermenters such as artiodactyls (Foley *et al.* 1995). In our analyses, two-toed sloths (genus *Choloepus*) present distinct gut microbial communities, but nevertheless appear close to herbivores in neighbour-net analyses and in intermediate position between omnivores and herbivores in beta-diversity plots. They do not cluster firmly with the colobus monkey (*Colobus guereza*), the only other folivorous species, nor with artiodactyls or hindgut fermenters. There is thus some degree of convergence in gut microbiota composition between folivorous sloths and mammalian herbivores, but the sloth microbiome seems to be rather distinct suggesting substantial differences in digestion modes in these xenarthrans relative to other mammals.

Potential chitin degraders and prospects for future metagenomic surveys

The myrmecophagous diet imposes strong nutritional challenges to mammals because a large proportion of

the protein value of termites and ants consists of the chitin constituting their exoskeletons. In terms of nutritional values, ants and termites are not especially different from other terrestrial invertebrates; however, their larval and alate forms contain much more fat and are thus the frequent prey of many species of mammals (Redford & Dorea 1984). Apart from a muscular stomach thought to assist in the mechanical processing of large amounts of social insects, myrmecophagous mammalian digestive tracts do not show obvious anatomical adaptations that might reflect an increased potential for chitin degradation (Stevens & Hume 2004). Nevertheless, the macroscopic inspection of faecal samples in diet characterization studies (Kruuk & Sands 1972; Taylor *et al.* 2002; Miranda *et al.* 2009) and the results of some rare functional assays (Cooper & Withers 2004) suggest that substantial variations exist among mammalian myrmecophages with regard to chitin degradation and assimilation.

The degradation of chitin in nature is primarily carried out by bacterial taxa such as pseudomonads, enteric bacteria, gliding bacteria, actinomycetes and members of the genera *Bacillus*, *Vibrio* and *Clostridium* (Gooday 1990). Endogenous chitinases have also been reported in plants, invertebrates, fungi and vertebrates including mammals. Thus, there are three possible sources of chitinolytic enzymes in the digestive system: from the animal itself, from its associated gut microbiome or from the ingested food (Gooday 1990). Chitinase genes are found in vertebrate genomes, but their exact function and role in digestion are not fully understood in mammals (Bussink *et al.* 2007; Funkhouser & Aronson 2007). These enzymes are nevertheless believed to be important for chitin degradation in marine fishes feeding on crustaceans (Gutowska *et al.* 2004), and chitinase activity has been detected in the stomach of insectivorous nine-banded armadillos (Smith *et al.* 1998). It has also been recently demonstrated that these enzymes might also have a digestive activity in the human gastric juice (Paoletti *et al.* 2007). However, there is also evidence that chitinolytic bacteria are an integral part of mammalian gut microbiomes (Simunek *et al.* 2001). Some studies have suggested that the forestomach of minke whales (*Balaenoptera acutorostrata*) includes bacterial taxa that can digest the chitinous exoskeleton of krill (Mårtensson *et al.* 1994) and that insectivorous bats use symbiotic gut bacteria to assimilate the chitin of their insect prey (Whitaker *et al.* 2004).

A key question is thus whether myrmecophagous mammals possibly use symbiotic bacteria to help them digest the chitin of ant and termite exoskeletons that would account at least partially for the observed convergence in the composition of their gut microbiota. If this process occurs, are the same bacterial taxa

recruited independently in different myrmecophagous lineages or do different bacterial taxa provide the same functions? Although we tentatively identified some bacterial genera whose abundance is significantly increased in myrmecophagous species, among which *Lactococcus* is a promising candidate with species known to degrade chitin such as *L. lactis* (Vaaje-Kolstad *et al.* 2009), we do not have an understanding of the functional gene content associated with these taxa. For example, processes such as gene loss and lateral gene transfer occurring among gut and environmental bacteria (Hehemann *et al.* 2010) could lead to a weak correspondence between the taxonomic classification and the functional role of host-associated microbes. An investigation of the functional gene content of myrmecophagous gut microbial communities through shotgun metagenomic approaches could help elucidate whether these and other taxa provide such functional roles.

A previous shotgun metagenomic survey illustrated the point that mammalian gut microbial communities can differ in their composition, but nevertheless share a functional core of genes ensuring similar functions (Muegge *et al.* 2011). Such functional metagenomic studies of the mammalian gut microbiome are currently expanding, as illustrated by recent surveys conducted in pig (Lamendella *et al.* 2011), Iberian lynx (Alcaide *et al.* 2012) and pygmy loris (Xu *et al.* 2013), providing comparative data for future studies of myrmecophagous mammals. The study on the endangered Iberian lynx (*Lynx pardalis*) suggests the intriguing possibility that the lynx microbiome evolved to digest not only the meat of its exclusive prey (the European rabbit, *Oryctolagus cuniculus*) but also the plants constitutive of its diet (Alcaide *et al.* 2012). For myrmecophagous animals, such studies would focus on genes and pathways involved in chitin degradation, which might have been convergently recruited in ant-eating placentals. Future genomic, metagenomic and functional studies of myrmecophagous placental species and their associated oral and gut microbiomes should help reveal the complex interplay between the host genome and its associated microbiome in the adaptation to the myrmecophagous diet. Testing whether adaptation to myrmecophagy consists primarily of differences in which taxa are present, in which genes supply essential functions, or in gene expression thus remains a key challenge for the field, although the present study provides a substantial advance in our understanding of the overall patterns of microbial communities associated with this extreme dietary adaptation.

Acknowledgements

We would like to thank Gail Ackermann, Catherine Nicholas, Chris Lauber, Donna Berg-Lyons, Matthew J.

Gebert, Gregory Humphrey and Yoshiki Vázquez Baeza for their help in data generation and processing. This study would not have been possible without the contribution of the following people and institutions in facilitating access to biological samples: Carly Vynne, Jim Loughry, Mariella Superina, David Gomis, Cédric Libert and Yann Raulet (Zoo de Lunaret Montpellier), Cynthia Steiner and Josephine Braun (San Diego Zoo), Andreas Bernhard (Leipzig Zoo), Angela Ryan and Amanda Ferguson (Regent's Park, Zoological Society of London), Sarah Forsyth (Colchester Zoo), Terri Roth (Cincinnati Zoo), Joseph Mendelson (Atlanta Zoo), Joseph Flanagan and Lauren Howard (Houston Zoo). We wish to thank the National Ecological Observatory Network (a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc.) for donation of the soil samples that we used in source-tracking analyses. Finally, we thank the subject editor and three anonymous referees for numerous constructive and thoughtful comments. This work has been financially supported by grants from the France-US Fulbright Commission to FD and from the Howard Hughes Medical Institute to RK. This is publication ISEM 2013-108 of the Institut des Sciences de l'Evolution de Montpellier.

References

- Alcaide M, Messina E, Richter M *et al.* (2012) Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx. *PLoS One*, **7**, e51521.
- Anacleto TCS (2007) Food habits of four armadillo species in the Cerrado area, Mato Grosso, Brazil. *Zoological Studies*, **46**, 529–537.
- Anderson MD, Richardson PRK, Woodall PF (1992) Functional analysis of the feeding apparatus and digestive tract anatomy of the aardwolf *Proteles cristatus*. *Journal of Zoology*, **228**, 423–434.
- Anderson KE, Russell JA, Moreau CS *et al.* (2012) Highly similar microbial communities are shared among related and trophically similar ant species. *Molecular Ecology*, **21**, 2282–2296.
- Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A*, **160**, 268–282.
- Beyer WN, Connor EE, Gerould S (1994) Estimates of soil ingestion by wildlife. *The Journal of Wildlife Management*, **58**, 375–382.
- Bokulich NA, Subramanian S, Faith JJ *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 57–59.
- Bolković ML, Caziani SM, Protomastro JJ (1995) Food habits of the three-banded armadillo (Xenarthra: Dasypodidae) in the dry Chaco, Argentina. *Journal of Mammalogy*, **76**, 1199–1204.
- Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.
- Bussink AP, Speijer D, Aerts JM, Boot RG (2007) Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. *Genetics*, **177**, 959–970.

- Caporaso JG, Bittinger K, Bushman FD *et al.* (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.
- Caporaso JG, Lauber CL, Walters WA *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA*, **108**(Suppl. 1), 4516–4522.
- Caporaso JG, Lauber CL, Walters WA *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, **6**, 1621–1624.
- Cooper CE, Withers PC (2004) Termite digestibility and water and energy contents determine the water economy index of numbats (*Myrmecobius fasciatus*) and other myrmecophages. *Physiological and Biochemical Zoology*, **77**, 641–650.
- Clarke KR, Gorley RN (2006) PRIMER v6. *User manual/tutorial. Plymouth routine in multivariate ecological research. Plymouth Marine Laboratory.*
- De Jesus-Laboy KM, Godoy-Vitorino F, Piceno YM *et al.* (2011) Comparison of the fecal microbiota in feral and domestic goats. *Genes*, **3**, 1–18.
- Delsuc F, Vizcaino SF, Douzery EJP (2004) Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology*, **4**, 11.
- Delsuc F, Superina M, Tilak MK, Douzery EJP, Hassanin A (2012) Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Molecular Phylogenetics and Evolution*, **62**, 673–680.
- DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied Environmental Microbiology*, **72**, 5069–5072.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
- Eilers KG, Debenport S, Anderson S, Fierer N (2012) Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biology and Biochemistry*, **50**, 58–65.
- Eizirik E, Murphy WJ, Koepfli KP *et al.* (2010) Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **56**, 49–63.
- Fang W, Fang Z, Zhou P *et al.* (2012) Evidence for lignin oxidation by the giant panda fecal microbiome. *PLoS One*, **7**, e50312.
- Feldhamer GA (2007) *Mammalogy: Adaptation, Diversity, Ecology*. John Hopkins University Press, Baltimore, Maryland.
- Foley WJ, Engelhardt WV, Charles-Dominique P (1995) The passage of digesta, particle size, and in vitro fermentation rate in the three-toed sloth *Bradypus tridactylus* (Edentata: Bradypodidae). *Journal of Zoology*, **236**, 681–696.
- Funkhouser JD, Aronson NN (2007) Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evolutionary Biology*, **7**, 96.
- Garcia-Amado MA, Godoy-Vitorino F, Piceno YM *et al.* (2012) Bacterial diversity in the cecum of the world's largest living rodent (*Hydrochoerus hydrochaeris*). *Microbial Ecology*, **63**, 719–725.
- Gilbert JA, Meyer F, Jansson J *et al.* (2010) The Earth Microbiome Project: meeting report of the “1 EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6 2010. *Standards in Genomic Sciences*, **3**, 249–253.
- Godoy-Vitorino F, Goldfarb KC, Karaoz U *et al.* (2011) Comparative analyses of foregut and hindgut bacterial communities in hoatzins and cows. *ISME Journal*, **6**, 531–541.
- Godoy-Vitorino F, Leal SJ, Diaz WA *et al.* (2012) Differences in crop bacterial community structure between hoatzins from different geographical locations. *Research in Microbiology*, **163**, 211–220.
- Gooday GW (1990) The ecology of chitin degradation. *Advances in Microbial Ecology*, **11**, 387–430.
- Gutowska MA, Drazen JC, Robison BH (2004) Digestive chitinolytic activity in marine fishes of Monterey Bay, California. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology*, **139**, 351–358.
- Hehemann JH, Correc G, Barbeyron T *et al.* (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, **464**, 908–912.
- Huchon D, Delsuc F, Catzeflis FM, Douzery EJP (1999) Armadillos exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasypus novemcinctus* (Xenarthra). *Molecular Ecology*, **8**, 1743–1748.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- Ishaq SL, Wright AD (2012) Insight into the bacterial gut microbiome of the North American moose (*Alces alces*). *BMC Microbiology*, **12**, 212.
- Jin C, Ciochon RL, Dong W *et al.* (2007) The first skull of the earliest giant panda. *Proceedings of the National Academy of Sciences, USA*, **104**, 10932–10937.
- Joshi AR, Garshelis DL, Smith JL (1997) Seasonal and habitat-related diets of sloth bears in Nepal. *Journal of Mammalogy*, **78**, 584–597.
- Karasov WH, Martinez del Rio C, Caviedes-Vidal E (2011) Ecological physiology of diet and digestive systems. *Annual Review of Physiology*, **73**, 69–93.
- Knights D, Kuczynski J, Charlson ES *et al.* (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, **8**, 761–763.
- Koepfli KP, Jenks SM, Eizirik E *et al.* (2006) Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Molecular Phylogenetics and Evolution*, **38**, 603–620.
- Köhler T, Dietrich C, Scheffrahn RH, Brune A (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental Microbiology*, **78**, 4691–4701.
- Krause J, Unger T, Nocon A *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, **8**, 220.
- Kruuk H, Sands WA (1972) The aardwolf (*Proteles cristata*, Sparrman 1783) as predator of termites. *African Journal of Ecology*, **10**, 211–227.

- Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiology*, **11**, 103.
- Lankau EW, Hong PY, Mackie RI (2012) Ecological drift and local exposures drive enteric bacterial community differences within species of Galapagos iguanas. *Molecular Ecology*, **21**, 1779–1788.
- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI (2008a) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, **6**, 776–788.
- Ley RE, Hamady M, Lozupone C *et al.* (2008b) Evolution of mammals and their gut microbes. *Science*, **320**, 1647–1651.
- Linnenbrink M, Wang J, Hardouin EA *et al.* (2013) The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, **22**, 1904–1916.
- Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, **36**, e120.
- Loughry WJ, Truman RW, McDonough CM, Tilak MK, Garnier S, Delsuc F (2009) Is leprosy spreading among nine-banded armadillos in the southeastern United States? *Journal of Wildlife Diseases*, **45**, 144–152.
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.
- Mårtensson PE, Nordøy ES, Blix AS (1994) Digestibility of krill (*Euphausia superba* and *Thysanoessa* sp.) in minke whales (*Balaenoptera acutorostrata*) and crabeater seals (*Lobodon carcinophagus*). *The British Journal of Nutrition*, **72**, 713–716.
- McDonald D, Price MN, Goodrich J *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal*, **6**, 610–618.
- McFall-Ngai M, Hadfield MG, Bosch TC *et al.* (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences, USA*, **110**, 3229–3236.
- McGhee GR (2011) *Convergent Evolution: Limited Forms Most Beautiful*. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.
- McNab BK (1984) Physiological convergence amongst ant-eating and termite-eating mammals. *Journal of Zoology*, **203**, 485–510.
- Meredith RW, Janecka JE, Gatesy J *et al.* (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, **334**, 521–524.
- Miranda F, Veloso R, Superina M, Zara FJ (2009) Food habits of wild silky anteaters (*Cyclopes didactylus*) of São Luis do Maranhão, Brazil. *Edentata*, **8**, 1–5.
- Muegge BD, Kuczynski J, Knights D *et al.* (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, **332**, 970–974.
- Ochman H, Worobey M, Kuo CH *et al.* (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, **8**, e1000546.
- Paoletti MG, Norberto L, Damini R, Musumeci S (2007) Human gastric juice contains chitinase that can degrade chitin. *Annals of Nutrition and Metabolism*, **51**, 244–251.
- Parks DH, Beiko RG (2012) Measuring community similarity with phylogenetic networks. *Molecular Biology and Evolution*, **29**, 3947–3958.
- Phillips CD, Phelan G, Dowd SE *et al.* (2012) Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Molecular Ecology*, **21**, 2617–2627.
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
- Ramirez KS, Lauber CL, Knight R, Bradford MA, Fierer N (2010) Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology*, **91**, 3463–3470.
- Redford KH (1985) Foods habits of armadillos (Xenarthra: Dasypodidae). In: *The Evolution and Ecology of Sloths, Armadillos, and Vermilinguas*, pp. 429–437. Smithsonian Institution Press, Washington, D.C.
- Redford KH (1986) Dietary specialization and variation in two mammalian myrmecophages (variation in mammalian myrmecophagy). *Revista Chilena de Historia Natural*, **59**, 201–208.
- Redford KH (1987) Ants and termites as food: patterns of mammalian myrmecophagy. In: *Current Mammalogy*, vol. 1, pp. 349–399. Plenum Press, New York.
- Redford KH, Dorea JG (1984) The nutritional value of invertebrates with emphasis on ants and termites as food for mammals. *Journal of Zoology*, **203**, 385–395.
- Reiss KZ (2001) Using phylogenies to study convergence: the case of the ant-eating mammals. *American Zoologist*, **41**, 507–525.
- Russell JA, Moreau CS, Goldman-Huertas B *et al.* (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 21236–21241.
- Sikes RS, Heidt GA, Elrod DA (1990) Seasonal diets of the nine-banded armadillo (*Dasypus novemcinctus*) in a northern part of its range. *American Midland Naturalist*, **123**, 383–389.
- Simůnek J, Hodrová B, Bartonová H, Kopecný J (2001) Chitino-lytic bacteria of the mammal digestive tract. *Folia Microbiologica*, **46**, 76–78.
- Smith SA, Robbins LW, Steiert JG (1998) Isolation and characterization of a chitinase from the nine-banded armadillo, *Dasypus novemcinctus*. *Journal of Mammalogy*, **79**, 486–491.
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proceedings of the National Academy of Sciences, USA*, **100**, 1056–1061.
- Stevens CE, Hume ID (2004) *Comparative Physiology of the Vertebrate Digestive System*. Cambridge University Press, Cambridge, UK.
- Sullam KE, Essinger SD, Lozupone CA *et al.* (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology*, **21**, 3363–3378.
- Superina M (2011) Husbandry of a pink fairy armadillo (*Chlamyphorus truncatus*): case study of a cryptic and little known species in captivity. *Zoo Biology*, **30**, 225–231.
- Taulman JF, Robbins LW (1996) Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. *Journal of Biogeography*, **23**, 635–648.

- Taylor WA, Lindsey PA, Skinner JD (2002) The feeding ecology of the aardvark *Orycteropus afer*. *Journal of Arid Environments*, **50**, 135–152.
- Vaae-Kolstad G, Bunaes AC, Mathiesen G, Eijsink VG (2009) The chitinolytic system of *Lactococcus lactis* ssp. *lactis* comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of alpha- and beta-chitin. *FEBS Journal*, **276**, 2402–2415.
- Vaz VC, Santori RT, Jansen AM, Delciellos AC, D'Andrea PS (2012) Notes on food habits of armadillos (Cingulata, Dasypodidae) and anteaters (Pilosa, Myrmecophagidae) at Serra Da Capivara National Park (Piauí State, Brazil). *Edentata*, **13**, 84–89.
- Vynne C, Skalski JR, Machado RB *et al.* (2011) Effectiveness of scat-detection dogs in determining species presence in a tropical savanna landscape. *Conservation Biology*, **25**, 154–162.
- Walters WA, Caporaso JG, Lauber CL *et al.* (2011) Primer-Prospector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, **27**, 1159–1161.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.
- Whitaker JO Jr, Dannelly HK, Prentice DA (2004) Chitinase in insectivorous bats. *Journal of Mammalogy*, **85**, 15–18.
- Xu B, Xu W, Yang F *et al.* (2013) Metagenomic analysis of the pygmy loris fecal microbiome reveals unique functional capacity related to metabolism of aromatic compounds. *PLoS One*, **8**, e56565.
- Yang CW, Chen S, Chang CY *et al.* (2007) History and dietary husbandry of pangolins in captivity. *Zoo Biology*, **26**, 223–230.
- Yatsunencko T, Rey FE, Manary MJ *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–227.
- Yildirim S, Yeoman CJ, Sipos M *et al.* (2010) Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS One*, **5**, e13963.
- Zhu L, Wu Q, Dai J, Zhang S, Wei F (2011) Evidence of cellulose metabolism by the giant panda gut microbiome. *Proceedings of the National Academy of Sciences, USA*, **108**, 17714–17719.

F.D., J.L.M and R.K. designed research. F.D. collected samples. R.K. and A.G. contributed new reagents and analytical tools. F.D., J.L.M., L.W.P and S.J.S. performed research and analysed data. F.D. and J.L.M. wrote the first draft of the manuscript. F.D., J.L.M., L.W.P, S.J.S., A.G. and R.K. participated in revising the final manuscript.

Data accessibility

The new 16S rRNA sequence data produced in this study are part of the Earth Microbiome Project and can

be freely accessed from a dedicated public database at <http://www.microbio.me/emp/> under study ID 1056 “Delsuc_microbi_ant”. The data are also available at EBI under Accession no. ERP003782 <http://www.ebi.ac.uk/ena/data/view/ERP003782>. Finally, mapping files and OTU tables used for this study may be downloaded from DRYAD at doi:10.5061/dryad.390 ng.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information and basic statistics on the 93 samples newly sequenced in this study.

Table S2 Information on the soil samples included in the source tracking analysis.

Table S3 Information and basic statistics on the 39 additional mammalian samples included in our comparative study.

Table S4 Results from the source-tracking analysis using a previously published set of mammal gut samples, as well as soil samples, as potential sources to determine which of the 92 new samples were potentially contaminated with soil.

Table S5 Mean relative abundance (RA) of the taxa from Table 1 in the six diet groups: myrmecophages, insectivores, omnivores, carnivores, herbivores, and folivores.

Table S6 Results of ANOSIM and the Bartlett test of homogeneity of variances between diet groups for the full (All) and reduced (Captive) data sets, including pairwise comparisons of diet categories.

Fig. S1 Proportion of Phyla represented in sequences that did not hit Greengenes.

Fig. S2 Results from the source-tracking analysis using a previously published set of mammal gut samples, as well as soil samples, as potential sources to determine which of the 92 samples were potentially contaminated with soil.

Fig. S3 Procrustes analysis comparing unweighted UniFrac distances for 16 mammalian samples sequenced for V2 using 454 technology (Muegge *et al.* 2011) and resequenced as part of this study for the V4 region using an Illumina platform.

Fig. S4 Taxa summary plots at the order level for 94 fecal samples representing 47 mammalian species ordered by host diet (upper panel) and by host order (lower panel).

Fig. S5 Neighbor-Net network based on unweighted UniFrac distances between 94 mammalian gut samples colored by diet (see inset legend).

Fig. S6 Principal coordinate analysis (PCoA) of 94 mammalian samples (with potentially contaminated samples removed) based on unweighted UniFrac distances and colored by wild (red) vs. captive (blue) status.

Shotgun Mitogenomics Provides a Reference Phylogenetic Framework and Timescale for Living Xenarthrans

Gillian C. Gibb,^{1,2} Fabien L. Condamine,^{1,3,4} Melanie Kuch,⁵ Jacob Enk,⁵ Nadia Moraes-Barros,^{6,7} Mariella Superina,⁸ Hendrik N. Poinar,^{*,5} and Frédéric Delsuc^{*,1}

¹Institut des Sciences de l'Évolution, UMR 5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France

²Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand

³Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden

⁴Department of Biological Sciences, University of Alberta, Edmonton, AL, Canada

⁵McMaster Ancient DNA Centre, Department of Anthropology and Biology, McMaster University, Hamilton, ON, Canada

⁶Cíbio/Inbio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal

⁷Laboratório de Biologia Evolutiva e Conservação de Vertebrados (Labec), Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

⁸Laboratorio de Endocrinología de la Fauna Silvestre, IMBECU, CCT CONICET Mendoza, Mendoza, Argentina

*Corresponding author: E-mail: poinarh@mcmaster.ca; frederic.delsuc@umontpellier.fr.

Associate editor: Nicolas Vidal

Abstract

Xenarthra (armadillos, sloths, and anteaters) constitutes one of the four major clades of placental mammals. Despite their phylogenetic distinctiveness in mammals, a reference phylogeny is still lacking for the 31 described species. Here we used Illumina shotgun sequencing to assemble 33 new complete mitochondrial genomes, establishing Xenarthra as the first major placental clade to be fully sequenced at the species level for mitogenomes. The resulting data set allowed the reconstruction of a robust phylogenetic framework and timescale that are consistent with previous studies conducted at the genus level using nuclear genes. Incorporating the full species diversity of extant xenarthrans points to a number of inconsistencies in xenarthran systematics and species definition. We propose to split armadillos into two distinct families Dasypodidae (dasypodines) and Chlamyphoridae (euphractines, chlamyphorines, and tolpeutines) to better reflect their ancient divergence, estimated around 42 Ma. Species delimitation within long-nosed armadillos (genus *Dasypus*) appeared more complex than anticipated, with the discovery of a divergent lineage in French Guiana. Diversification analyses showed Xenarthra to be an ancient clade with a constant diversification rate through time with a species turnover driven by high but constant extinction. We also detected a significant negative correlation between speciation rate and past temperature fluctuations with an increase in speciation rate corresponding to the general cooling observed during the last 15 My. Biogeographic reconstructions identified the tropical rainforest biome of Amazonia and the Guiana Shield as the cradle of xenarthran evolutionary history with subsequent dispersions into more open and dry habitats.

Key words: mammals, Xenarthra, shotgun Illumina sequencing, molecular phylogenetics, mitochondrial genomes, molecular dating.

Introduction

Xenarthra (armadillos, anteaters, and sloths), with only 31 living species currently recognized (Wetzel 1985; Gardner 2008), is the least diversified of the 4 major groups of placental mammals (Meredith et al. 2011). As old, South American endemics, their evolutionary history has been shaped by their geographical isolation for the greater part of the Cenozoic until the Great American Biotic Interchange (GABI) triggered by the formation of the Isthmus of Panama (Marshall et al. 1982; Bacon et al. 2015). Xenarthrans have been major components during this interchange, with many taxa successfully dispersing into Central and North America (Patterson and Pascual 1968; McDonald 2005). The reduced number of extant species is likely the result of extinctions, especially during the terminal

Pleistocene extinctions that occurred around 11,000 years ago. This major extinction event seems to have preferentially affected the largest terrestrial forms, such as giant ground sloths and glyptodonts (Lyons et al. 2004). Surviving xenarthran species can be regarded as unique witnesses of the oldest South American endemic radiation of placental mammals (Delsuc et al. 2004). They thus represent an ideal model for understanding the biogeographical patterns and diversification processes at work during South America's "splendid isolation" (Simpson 1980; Moraes-Barros and Arteaga 2015).

The last decade has seen much progress in elucidating xenarthran phylogeny, thanks to new molecular data. Most of these studies have focused on their position within placental mammals because morphological studies placed them as a sister group of other placentals, referred to as the Epitheria

© The Author 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

hypothesis (Novacek 1992). The seminal molecular phylogenetic studies of placentals have shown convincingly that armadillos, anteaters, and sloths (*Xenarthra*) constitute one of the four major placental clades, establishing them as an essential component of the early placental radiation alongside Afrotheria, Laurasiatheria, and Euarchontoglires (Madsen et al. 2001; Murphy, Eizirik, Johnson, et al. 2001; Murphy, Eizirik, O'Brien, et al. 2001). Yet, despite studies using multi-gene data (Delsuc et al. 2002; Amrine-Madsen et al. 2003; Meredith et al. 2011), retroposon insertions (Churakov et al. 2009; Nishihara et al. 2009) and genome-wide data (McCormack et al. 2012; Song et al. 2012; Romiguier et al. 2013), their exact position within placentals remains contentious.

Within *Xenarthra*, molecular studies have converged upon a robust phylogeny of the 14 recognized genera (Delsuc et al. 2002, 2003, 2012; Möller-Krull et al. 2007). This phylogenetic framework has served for specifying the timing of their diversification in South America during the Cenozoic (Delsuc et al. 2004). These studies have also helped refine xenarthran taxonomy with, for instance, the recognition of two distinct families within anteaters (Delsuc et al. 2001; Barros et al. 2008) to reflect the deep divergence (about 40 Ma) estimated between the pygmy anteater (*Cyclopes didactylus*; Cyclopedidae) and the giant anteater and tamanduas (Myrmecophagidae). Molecular data have also recently revealed an ancient divergence of fairy armadillos, supporting the classification of the two living species in two distinct genera (*Chlamyphorus* and *Calyptophractus*) and grouped into the distinct subfamily Chlamyphorinae (Delsuc et al. 2012).

Despite these significant advances, a fully resolved species-level phylogeny is still lacking for *Xenarthra*. The reason for this includes the rarity of many of its constitutive species and a dated taxonomy with persistent uncertainty on species delimitations and distributions due to a lack of basic field data (Superina et al. 2014). Establishing a good reference framework would be critical for the conservation of this peculiar placental group, which includes a number of endangered species. According to the last IUCN Red list Assessment (IUCN Red List of Threatened Species 2015), 6 of the 31 living species (19%) were considered threatened, while the conservation status of 5 species could not be assessed due to lack of data. One particularly striking example is the iconic Brazilian three-banded armadillo for which no molecular data currently exist and for which only scarce biological information can be found in the literature (Superina et al. 2014; Feijó et al. 2015). More generally, 9 species among the 31 currently recognized xenarthrans still have not been investigated via molecular means are thus not represented in GenBank.

The scarcity of molecular data for this group is perhaps best reflected by the fact that only four xenarthran complete mitochondrial genomes are available, only two of which have been published (Arnason et al. 1997, 2002). With current advances in DNA sequencing, mitogenomes have become a standard for estimating well-sampled species-level phylogenies in numerous mammalian groups (Hassanin et al. 2012;

Finstermeier et al. 2013; Guschanski et al. 2013; Mitchell et al. 2014). In facilitating access to museum specimens and recently extinct species, next-generation sequencing techniques promise the development of museomics (Mason et al. 2011; Rowe et al. 2011; Fabre et al. 2014) and a rebirth of ancient DNA studies based on complete mitogenomes (Enk et al. 2011; Paijmans et al. 2013). Here we have used an Illumina shotgun sequencing approach to obtain and assemble 33 mitogenomes representing all living xenarthran species using both modern tissues and historic museum samples, including a type specimen. This allowed us to establish a robust phylogenetic framework and timescale that we have used as a reference to evaluate the systematics and species delineation within xenarthrans, and to study their diversification history and biogeography with respect to the paleoenvironmental changes that took place in South America throughout the Cenozoic.

Results and Discussion

Shotgun Mitogenomics and Phylogenetics

Using shotgun Illumina sequencing of genomic DNA and a combination of de novo assembly and mapping methods, we were able to successfully assemble 33 new complete mitochondrial genomes from individuals representing all 31 living xenarthran species (table 1). Using ancient DNA laboratory conditions to avoid potential contaminations, and cross-contamination, we extracted DNA from a variety of modern (ear biopsies, internal organs, and blood) and historical (dried skins and bones) tissue samples, whose collection dates ranged from 1896 to 2011 and have been stored in different collections since that time (table 1). Not surprisingly, the quality and quantity of DNA varied dramatically among samples. Nevertheless, shotgun Illumina sequencing from each total genomic DNA extract allowed the successful assembly of the mitogenome, with mean depth of coverage varying from 2,319X for the pygmy sloth (*Bradypus pygmaeus*) to 7X for the six-banded armadillo (*Euphractus sexcinctus*). The total number of reads required to obtain reasonable coverage was highly variable, as was also the proportion of mitochondrial reads recovered from each sample. At the two extremes of the range stand the pink fairy armadillo (*Chlamyphorus truncatus*) with 11.76% of mitochondrial reads from which we obtained a 244X coverage depth with only 570,194 total reads, and one of the two hairy long-nosed armadillo (*Dasybus pilosus*) samples that contained only 0.02% of mitochondrial reads and required the use of 44,847,476 total reads to reach 23X depth of coverage (table 1). These results confirmed that there is no a priori predictor of the final mitogenome coverage that may be obtained for a given sample, because it appears to be mainly dependent on the mitochondrial/nuclear cell ratio in the sampled tissue (Tilak et al. 2015). Also, predicting the amount of final mitogenome coverage from shotgun data is further complicated for museum specimens by the fraction of endogenous versus exogenous DNA in the sample. Our results nevertheless illustrate the utility of a shotgun approach for mammalian mitogenomics (Enk et al. 2011; Botero-Castro et al. 2013) enabling the efficient use of

Table 1. Sample Details and Assembly Statistics of the 33 New Xenarthran Mitochondrial Genomes.

Species	Common Name	Collection Number	Origin	Collection Date	Sample Type	Total Reads	Mito Reads	MtDNA (%)	Mean Coverage Depth	Accession Number
<i>Bradypus pygmaeus</i>	Pygmy sloth	USNM 579179	Panama	1991	Internal organ	31,262,238	383,944	1.23	2,319X	KT818523
<i>Bradypus torquatus</i>	Maned sloth	BA449/11	Brazil	2011	Skin biopsy	39,749,058	61,728	0.16	375X	KT818524
<i>Bradypus tridactylus</i>	Pale-throated three-toed sloth	ISEM T-5013	French Guiana	2006	Ear biopsy	2,016,278	11,916	0.59	42X	KT818525
<i>Bradypus variegatus</i>	Brown-throated three-toed sloth	MVZ 155186	Peru	1978	Internal organ	726,173	30,043	4.14	121X	KT818526
<i>B. variegatus</i>	Brown-throated three-toed sloth									NC_006923
<i>Choloepus didactylus</i>	Southern two-toed sloth	MNHN 1998-1819	French Guiana	1997	Internal organ	670,499	7,289	1.09	26X	KT818537
<i>Cho. didactylus</i>	Southern two-toed sloth									NC_006924
<i>Choloepus hoffmanni</i>	Hoffmann's two-toed sloth	ISEM T-6052	Panama	2001	Kidney	1,258,442	3,174	0.25	11X	KT818538
<i>Cyclopes didactylus</i>	Pygmy anteater	MNHN 1998-234	French Guiana	1995	Internal organ	1,832,592	5,630	0.31	21X	KT818539
<i>Myrmecophaga tridactyla</i>	Giant anteater	ISEM T-2862	French Guiana	2001	Internal organ	1,102,261	7,215	0.65	26X	KT818549
<i>Tamandua mexicana</i>	Northern tamandua	MVZ 192699	Mexico	1977	Internal organ	1,552,407	4,167	0.27	16X	KT818551
<i>Tamandua tetradactyla</i>	Southern tamandua	ISEM T-6054	French Guiana	2001	Ear biopsy	14,370,168	6,373	0.04	30X	KT818552
<i>Ta. tetradactyla</i>	Southern tamandua									NC_004032
<i>Dasybus kappleri</i>	Greater long-nosed armadillo	ISEM T-3365	French Guiana	2001	Internal organ	1,059,009	46,167	4.36	166X	KT818541
<i>Dasybus hybridus</i>	Southern long-nosed armadillo	ZVC M2010	Uruguay	1976	Tail bone	30,663,498	38,967	0.13	123X	KT818540
<i>Dasybus septemcinctus</i>	Seven-banded armadillo	ISEM T-3002	Argentina	2000	Ear biopsy	1,878,745	2,174	0.12	8X	KT818546
<i>Dasybus novemcinctus</i>	Nine-banded armadillo	ISEM T-1863	French Guiana	1995	Ear biopsy	4,525,766	2,104	0.05	8X	KT818542
<i>Da. novemcinctus</i>	Nine-banded armadillo		USA							NC_001821
<i>Dasybus pilosus</i>	Hairy long-nosed armadillo	MSB 49990	Peru	1980	Dried skin	21,384,760	11,901	0.06	54X	KT818544
<i>Da. pilosus</i>	Hairy long-nosed armadillo	LSUMZ 21888	Peru	1978	Dried skin	44,847,476	8,525	0.02	23X	KT818543
<i>Dasybus sabanicola</i>	Northern long-nosed armadillo	USNM 372834	Venezuela	1966	Dried skin	32,354,212	10,323	0.03	46X	KT818545
<i>Dasybus yepesi</i>	Yunga's long-nosed armadillo	MLP 30.III.90.2	Argentina	1988	Rib bone	1,518,470	4,203	0.28	13X	KT818547
<i>Chaetophractus nationi</i>	Andean hairy armadillo	ISEM T-LP1	Bolivia	2008	Blood	790,237	2,453	0.31	9X	KT818534
<i>Chaetophractus vellerosus</i>	Screaming hairy armadillo	ISEM T-CV1	Argentina	2005	Internal organ	1,212,063	9,552	0.79	34X	KT818533
<i>Chaetophractus villosus</i>	Large hairy armadillo	ISEM T-NP390	Argentina	2001	Ear biopsy	978,540	4,889	0.50	18X	KT818535
<i>Euphractus sexcinctus</i>	Six-banded armadillo	ISEM T-1246	NA	NA	Dried skin	2,885,506	2,029	0.07	7X	KT818548
<i>Zaedyus pichiy</i>	Pichi	ISEM T-6060	Argentina	2005	Internal organ	1,939,442	3,202	0.17	12X	KT818555
<i>Calyptophractus retusus</i>	Greater fairy armadillo	ZSM T-Bret	Bolivia	1974	Internal organ	1,766,903	3,063	0.17	10X	KT818532
<i>Chlamyphorus truncatus</i>	Pink fairy armadillo	ISEM T-CT1	Argentina	2005	Internal organ	570,194	67,049	11.76	244X	KT818536
<i>Priodontes maximus</i>	Giant armadillo	ISEM T-2353	Argentina	2000	Skin biopsy	13,662,361	5,123	0.04	19X	KT818550
<i>Tolypeutes matacus</i>	Southern three-banded armadillo	ISEM T-2348	Argentina	2000	Ear biopsy	1,356,101	6,051	0.45	20X	KT818553
<i>Tolypeutes tricinctus</i>	Brazilian three-banded armadillo	JB21	Brazil	2007	Ear biopsy	13,307,971	20,076	0.15	97X	KT818554

(continued)

Table 1. Continued

Species	Common Name	Collection Number	Origin	Collection Date	Sample Type	Total Reads	Mito Reads	MtDNA (%)	Mean Coverage Depth	Accession Number
<i>Cabassous centralis</i>	Northern naked-tailed armadillo	AMNH MO-10752	Costa Rica	1896	Skull bone	35,461,781	14,828	0.04	51X	KT818527
<i>Cabassous chacoensis</i>	Chacoan naked-tailed armadillo	ISEM T-2350	Argentina	2000	Tail biopsy	910,714	4,848	0.53	18X	KT818528
<i>Cabassous tatouay</i>	Greater naked-tailed armadillo	ZVC M365	Uruguay	1966	Dried skin	40,034,772	14,614	0.04	39X	KT818529
<i>Cabassous unicinctus</i>	Southern naked-tailed armadillo	MNHN 1999-1068	French Guiana	1995	Internal organ	2,099,343	19,302	0.92	66X	KT818531
<i>Cab. unicinctus</i>	Southern naked-tailed armadillo	ISEM T-2291	French Guiana	2000	Internal organ	1,021,049	4,205	0.41	16X	KT818530
<i>Dugong dugon</i>	Dugong									NC_003314
<i>Loxodonta Africana</i>	African elephant									NC_000934
<i>Orycteropus afer</i>	Aardvark									NC_002078

NOTE.—NA: not available; USNM: National Museum of Natural History, Washington, USA; ISEM: Institut des Sciences de l'Evolution, Montpellier, France; MVZ: Museum of Vertebrate Zoology, Berkeley, USA; MNHN: Museum National d'Histoire Naturelle, Paris, France; ZVC: Colección de Vertebrados de la Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay; MSB: Museum of Southwestern Biology, Albuquerque, USA; LSUMZ: Louisiana State University Museum of Natural Science, Baton Rouge, USA; MLP: Museo de La Plata, La Plata, Argentina; AMNH: American Museum of Natural History, New York, USA; ZSM: Zoologische Staatssammlung München, Munich, Germany.

museum specimens for full mitogenome assembly (Rowe et al. 2011).

The complete mitogenomes allowed us to construct a highly informative phylogenetic data set containing 15,006 sites for 40 taxa (37 xenarthrans and 3 outgroups). The best way to analyze mitogenomic data continues to be debated (Leavitt et al. 2013; Powell et al. 2013). Despite representing a single linked locus, selection pressures and evolutionary rates are highly heterogeneous across the mitochondrial DNA molecule (Galtier et al. 2006; Nabholz et al. 2012) and particular substitution patterns and base composition biases exist among sites and strands (Reyes et al. 1998). One way to accommodate this rate heterogeneity is to use partitioned models implemented in both maximum likelihood (ML) and Bayesian approaches, which have been shown to improve phylogenetic inference (Chiari et al. 2012; Kainer and Lanfear 2015). However, determining the best partition scheme currently requires the a priori definition of biologically relevant partitions (Lanfear et al. 2012). In our case, the optimal partition scheme selected by PartitionFinder (Lanfear et al. 2012) attributed a GTR+G+I model to four partitions, basically capturing the substitution rate heterogeneity among codon positions of protein-coding genes, RNAs, and ND6 that is the only gene to be encoded on the heavy strand (supplementary table S1, Supplementary Material online). A valuable alternative to partitioned models is provided by the Bayesian site-heterogeneous CAT-GTR (general time reversible) mixture model (Lartillot and Philippe 2004), in which the optimal number of site-specific substitution pattern categories is directly estimated from the data. The application of this model for analyzing mammalian mitogenomic data is only starting, but it has already been rather promising (Hassanin et al. 2012; Botero-Castro et al. 2013; Fabre et al. 2013). In the case of our xenarthran data set, ML and Bayesian implementations of the optimal partitioned model and Bayesian inference under the CAT-GTR mixture model gave exactly the same, fully resolved topology apart from one unsupported node within the genus *Dasyus* (fig. 1).

The xenarthran mitogenomic tree shows a fair amount of branch length heterogeneity among groups, with fast evolving clades including anteaters and Dasypodinae, and slow evolving clades such as sloths and hairy armadillos (fig. 1). Lineage-specific variation in evolutionary rates in mammalian mitochondrial genomes has been previously characterized (Gissi et al. 2000). Such variation has been linked to differences in mutation rates that correlate well with longevity in mammals (Nabholz et al. 2008). As a result, the mammalian mitochondrial clock is particularly erratic (Nabholz et al. 2009) and substitution rate variations among lineages should be accounted for in dating analyses by using relaxed clock models (Thorne et al. 1998). The selection of the clock model is, nevertheless, often arbitrary and appears mostly dependent upon the software choice, with an overwhelming majority of studies relying on BEAST (Drummond et al. 2012), thus generally using an uncorrelated gamma (UGAM) also

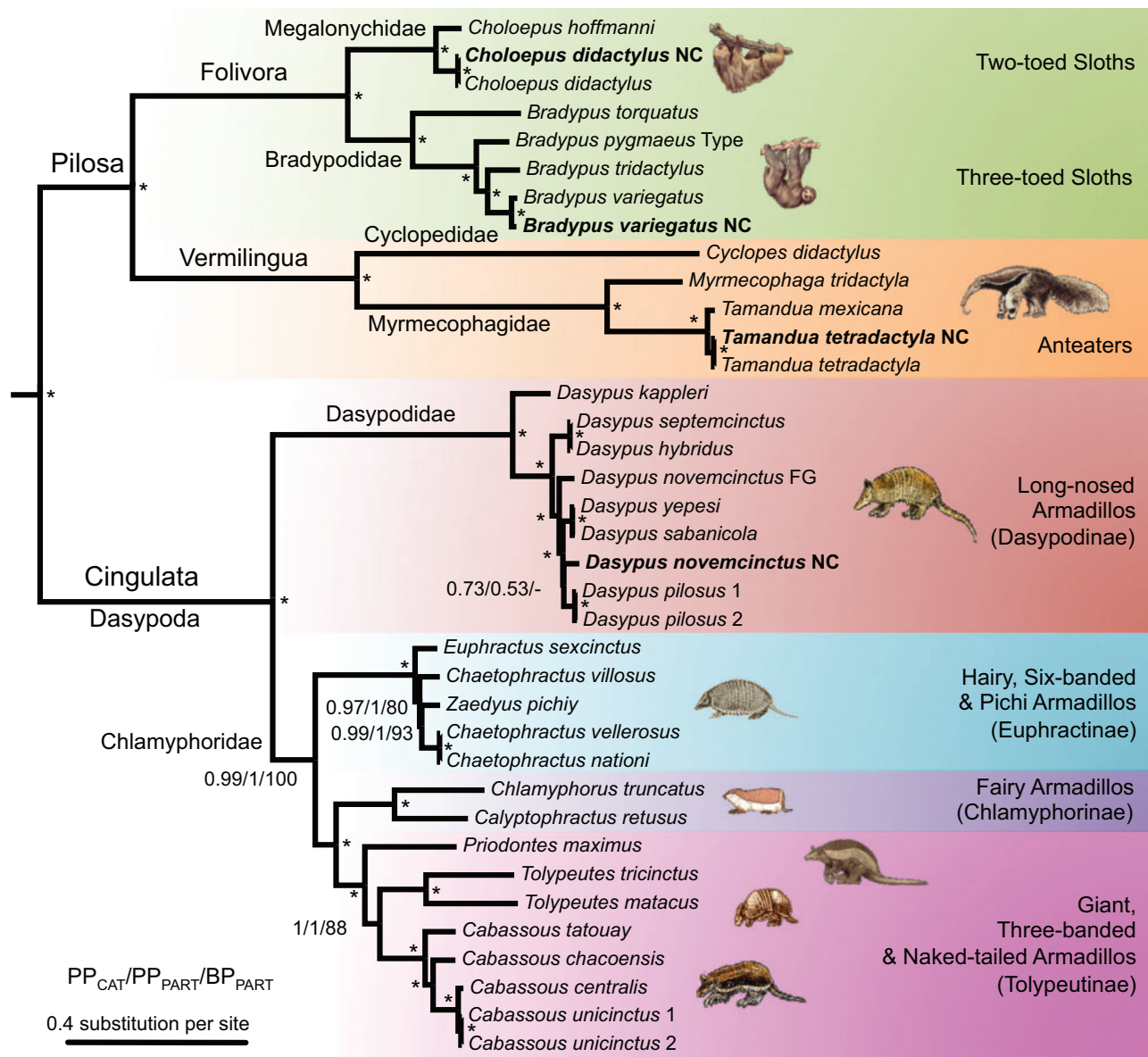


Fig. 1. Phylogenetic relationships of all extant xenarthran species. Bayesian consensus phylogram obtained using PhyloBayes under the CAT-GTR-G mixture model. Values at nodes indicate Bayesian posterior probabilities (PP) obtained under the CAT-GTR-G4 mixture model and maximum-likelihood bootstrap percentages (BP) obtained under the optimal partitioned model, respectively. Asterisks (*) mark branches that received maximal support from both methods. The afrotherian outgroup is not shown (full tree provided as [supplementary fig. S1, Supplementary Material](#) online). NC: GenBank reference mitogenomes (in bold); FG: French Guiana. Type: Museum type specimen.

known as UCLN) relaxed clock model (Drummond et al. 2006). However, it has been shown that autocorrelated rate models, such as the autocorrelated log-normal model (LN; Thorne et al. 1998), generally offer a better fit, especially with large data sets above the species level (Lepage et al. 2007; Rehm et al. 2011). We thus compared the fit of the UGAM and LN models with a strict molecular clock (CL) model using cross-validation tests. The latter showed that the relaxed clock models both largely outperform the strict clock model (UGAM vs. CL: 14.42 ± 9.12 ; LN vs. CL: 18.47 ± 4.86), and among relaxed clock models, LN fits our data better than UGAM (LN vs. UGAM: 4.05 ± 7.87). Accordingly, we present and discuss the divergence times obtained with the autocorrelated LN relaxed clock model (fig. 2 and table 2).

Phylogenetic Framework and Timescale for Living Xenarthrans

Our analyses provide the first comprehensive phylogeny including all living species of Xenarthra. We obtained a fully resolved tree with high ML bootstrap and Bayesian support values, except for one node within the genus *Dasyopus* (fig. 1). This mitogenomic topology is entirely congruent with previous studies conducted at the genus level using nuclear exons (Delsuc et al. 2002), combinations of mitochondrial and nuclear genes (Delsuc et al. 2001, 2003, 2012), and retroposons and their flanking noncoding sequences (Möller-Krull et al. 2007). Furthermore, the newly estimated timescale (fig. 2 and table 2) is compatible with previous time estimates provided by the analyses of nuclear exons alone (Delsuc et al. 2004) or

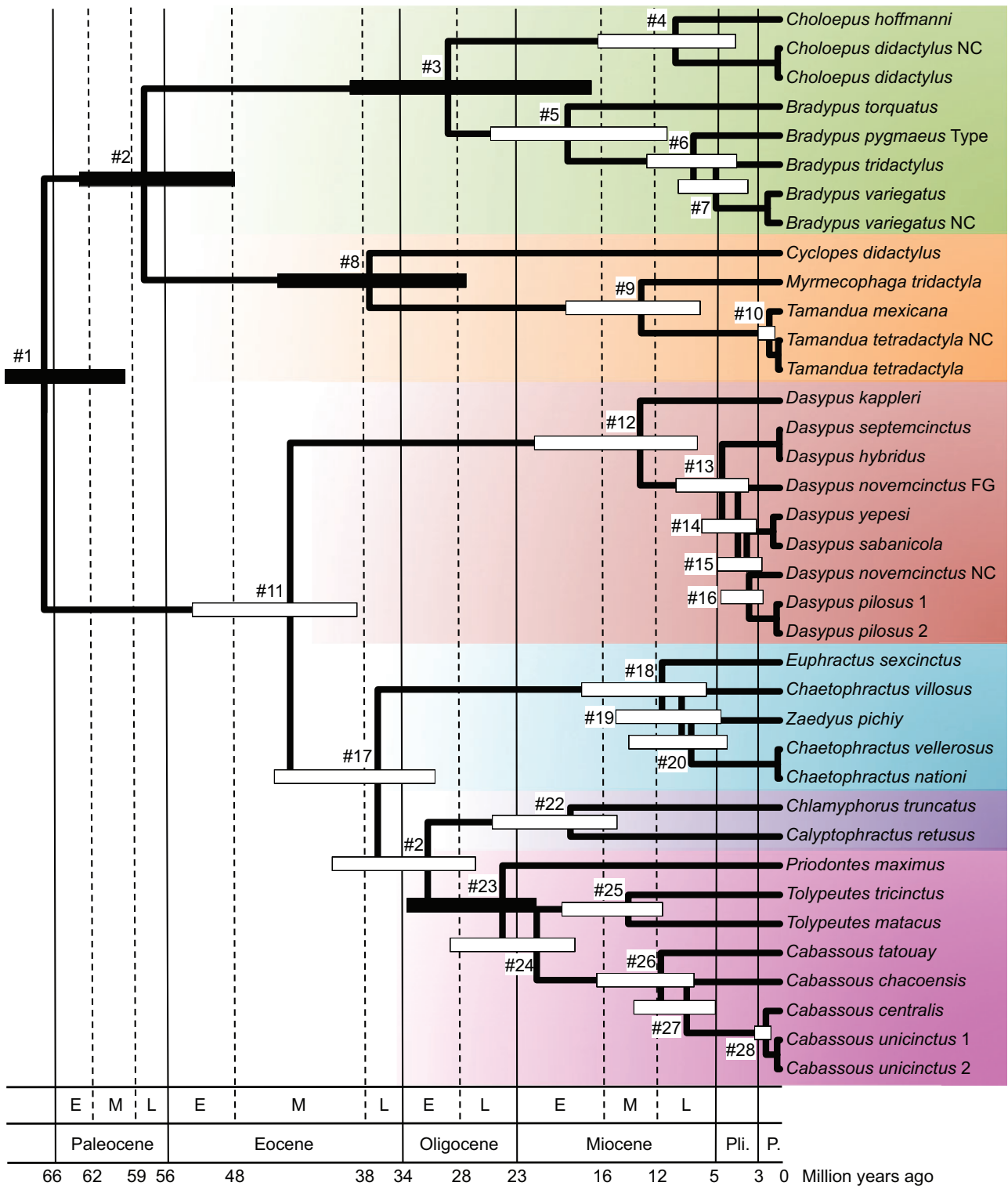


FIG. 2. Molecular timescale for all extant xenarthran species. The Bayesian chronogram was obtained using a rate-autocorrelated LN relaxed molecular clock model using PhyloBayes under the CAT-GTR-G mixture model with a birth–death prior on the diversification process, and six soft calibration constraints. Node bars indicate the uncertainty around mean age estimates based on 95% credibility intervals. Divergence dates less than 0.5 Ma are not represented. Plain black node bars indicate nodes used as a priori calibration constraints. Numbers at nodes refer to table 2. The afrotherian outgroup is not shown (full tree provided as supplementary fig. S2, Supplementary Material online). Vertical lines delimitate the main geological periods of the Cenozoic following the 2012 Geological Time Scale of the Geological Society of America (Gradstein et al. 2012). E = early; M = middle; L = late; Pli. = Pliocene; P. = Pleistocene.

Table 2. Divergence Time Estimates for All Xenarthran Nodes Inferred Using the Site-Heterogeneous CAT-GTR+G4 Substitution Model and an Autocorrelated LN Relaxed Molecular Clock Model.

Nodes	This study	Delsuc et al. (2012)	Delsuc et al. (2004)
1. Xenarthra ^a	67.7 ± 3.0 [60.4–71.6]	67.8 ± 3.4 [61.3–74.7]	64.7 ± 4.9 [55.3–74.6]
2. Pilosa ^a (anteaters + sloths)	58.4 ± 4.1 [48.6–64.7]	60.1 ± 3.6 [53.1–67.2]	55.2 ± 4.9 [45.8–65.2]
3. Folivora ^a (sloths)	29.9 ± 6.5 [16.5–39.6]	28.3 ± 3.4 [22.0–35.2]	20.8 ± 3.3 [15.0–27.8]
4. Megalonychidae (two-toed sloths)	9.2 ± 3.5 [3.5–16.7]	N.A.	N.A.
5. Bradypodidae (three-toed sloths)	19.0 ± 4.7 [9.6–27.0]	N.A.	N.A.
6. <i>Bradypus pygmaeus</i> /others	7.7 ± 2.4 [3.6–12.6]	N.A.	N.A.
7. <i>Bradypus tridactylus</i> / <i>Bradypus variegatus</i>	5.7 ± 1.8 [2.6–9.5]	N.A.	N.A.
8. <i>Vermilingua</i> ^a (anteaters)	37.8 ± 4.9 [26.9–46.2]	45.5 ± 3.7 [38.4–52.8]	40.0 ± 4.4 [31.8–49.0]
9. Myrmecophaga/Tamandua	12.7 ± 3.3 [7.0–19.8]	13.6 ± 2.1 [9.9–18.2]	10.1 ± 1.8 [6.9–14.1]
10. <i>Tamandua mexicana</i> / <i>Tamandua tetradactyla</i>	1.0 ± 0.4 [0.4–2.0]	N.A.	N.A.
11. Cingulata (armadillos)	44.9 ± 3.5 [38.3–52.1]	42.3 ± 3.8 [35.1–50.0]	39.7 ± 4.5 [31.3–49.1]
12. Dasypodinae (long-nosed armadillos)	12.4 ± 3.4 [7.2–20.4]	11.2 ± 2.0 [7.8–15.6]	7.3 ± 1.6 [4.6–10.9]
13. <i>Dasyopus septemcinctus</i> + <i>Dasyopus hybridus</i> /others	5.1 ± 1.7 [2.7–9.2]	N.A.	N.A.
14. <i>Dasyopus novemcinctus</i> FG/others	3.7 ± 1.2 [2.0–6.8]	N.A.	N.A.
15. <i>Dasyopus sabanicola</i> + <i>Dasyopus yepesi</i> /others	2.9 ± 1.0 [1.5–5.4]	N.A.	N.A.
16. <i>Dasyopus novemcinctus</i> NC/ <i>Dasyopus pilosus</i>	2.8 ± 0.9 [1.5–5.1]	N.A.	N.A.
17. Chlamyphoridae	37.2 ± 3.4 [31.5–44.7]	34.5 ± 3.6 [27.8–41.9]	32.9 ± 4.1 [25.2–41.5]
18. Euphractinae (hairy armadillos)	11.0 ± 2.8 [6.8–17.8]	8.3 ± 1.6 [5.5–11.8]	6.2 ± 1.4 [3.8–9.3]
19. <i>Chaetophractus villosus</i> /others	9.1 ± 2.4 [5.5–15.1]	7.1 ± 1.4 [4.7–10.3]	N.A.
20. <i>Zaedyus pichiy</i> /others	8.2 ± 2.3 [4.9–13.7]	N.A.	N.A.
21. Chlamyphorinae/Tolypeutinae	32.6 ± 3.1 [27.9–40.0]	32.9 ± 3.6 [26.3–40.2]	N.A.
22. Chlamyphorinae (fairy armadillos)	19.4 ± 2.7 [15.2–25.9]	17.3 ± 2.7 [12.4–23.0]	N.A.
23. Tolypeutinae ^a	25.7 ± 2.7 [22.4–32.7]	26.1 ± 3.2 [20.2–32.9]	21.8 ± 3.3 [15.8–28.9]
24. <i>Tolypeutes</i> / <i>Cabassous</i>	22.5 ± 2.6 [19.0–29.0]	24.2 ± 3.1 [18.5–30.7]	20.5 ± 3.2 [14.7–27.3]
25. <i>Tolypeutes</i>	14.1 ± 2.0 [11.0–19.1]	N.A.	N.A.
26. <i>Cabassous</i>	10.9 ± 1.9 [8.0–15.5]	N.A.	N.A.
27. <i>Cabassous chacoensis</i> /others	8.6 ± 1.6 [6.0–12.4]	N.A.	N.A.
28. <i>Cabassous centralis</i> / <i>Cabassous unicinctus</i>	1.3 ± 0.3 [0.8–2.1]	N.A.	N.A.

NOTE.—Mean posterior estimates, associated standard errors, and 95% credibility intervals are expressed in Ma (mean date ± SD [95% Cred]).

SD: standard deviation; 95% Cred: 95% credibility interval; FG: French Guiana; NC: GenBank reference mitogenome (specimen from the USA); N.A.: not applicable

^aUsed as a priori calibration constraints.

in combination with mitochondrial genes (Delsuc et al. 2012). The few discrepancies concern nodes for which the species sampling has been substantially increased such as Folivora, Dasypodinae, Euphractinae, and Tolypeutinae (table 2). For these nodes, the newly inferred dates appear older than previous estimates performed at the genus level as expected with a denser species sampling. Such global congruence with previous nuclear-based phylogenetic and dating analyses, allows being confident that ancient introgression and/or hybridization events have not significantly affected the mitogenomic tree of xenarthrans. A number of new surprising and important inferences are to be drawn from our mitogenomic framework with respect to phylogenetic relationships and species delimitation within the different xenarthran groups.

Sloths (*Pilosa*; *Folivora*)

The six living species of sloths belong to two genera, with two-toed sloths (genus *Choloepus*) and three-toed sloths (genus *Bradypus*) having been placed in two distinct families (Megalonychidae and Bradypodidae, respectively) to reflect their numerous morphological differences and a probably diphyletic origin from two different fossil groups (Webb 1985). Their independent adaptation to the arboreal lifestyle

also led to a number of anatomical convergences related to their peculiar suspensory locomotion (Nyakatura 2012). Our results confirm this deep dichotomy with a divergence date between the two genera around 30 Ma (fig. 2 and table 2), which appears more ancient than previously estimated with nuclear data (Delsuc et al. 2004). This difference might stem from our increased taxon sampling, because only a single representative species of each genus was previously considered. Their considerable molecular divergence nevertheless supports the classification of the two modern sloth genera into distinct families.

Within two-toed sloths, the new mitochondrial genome sequence obtained for the Southern two-toed sloth (*Choloepus didactylus*) appears almost identical to the reference mitogenome (NC_006924) deposited in GenBank (99.8% pairwise identity). As expected, the Hoffmann's two-toed sloth (*Choloepus hoffmanni*) is more divergent (pairwise distance of 7.2% with *Cho. didactylus*). The divergence time between the two toed-sloth species is estimated at about 9 Ma (fig. 2 and table 2). *Choloepus hoffmanni* presents two disjunct northern and southern populations. A recent study estimated the divergence between northern and southern

Cho. hoffmanni mitochondrial lineages at about 7 Ma (Moraes-Barros and Arteaga 2015), but it did not include *Cho. didactylus*. Obtained from a captive individual most likely coming from Panama, our *Cho. hoffmanni* mitogenome belongs to the northern population. Therefore, we could not discard the possibility that a southern *Cho. hoffmanni* sequence will not belong to the *Cho. didactylus* lineage. The occurrence of hybrids in captivity raises the question of whether hybridization also occurs in natural populations, especially between the southern populations of *Cho. hoffmanni* and individuals of *Cho. didactylus* inhabiting north-central Peru and south-western areas of Brazil (Steiner et al. 2010). This evidence coupled with the significant variation in chromosome number observed in South American *Choloepus* ($2n = 53\text{--}67$; Hayssen 2010) indicates the need for a taxonomic review of both *Cho. hoffmanni* and *Cho. didactylus*. This will require the analysis of additional mitochondrial and nuclear data for an extensive sampling, especially along the southern distribution of *Cho. hoffmanni*, where it is sympatric with *Cho. didactylus*.

Concerning three-toed sloths, the endangered maned sloth (*Bradypus torquatus*) is the sister group of the three other described species (fig. 1). This phylogenetic position is in agreement with previous studies based on a few mitochondrial genes (Barros et al. 2003; Moraes-Barros et al. 2011). Our dating estimates confirm the maned sloth as an ancient Atlantic forest endemic, which may have diverged from other sloths as early as 19 Ma (fig. 2 and table 2). Such an old divergence date associated with its distinctive morphological characters would argue for a classification of *B. torquatus* in its own genus (*Scaeopus*), as advocated by Barros et al. (2008). As the maned sloth is one of the most threatened xenarthran species (Superina, Plese, et al. 2010), its phylogenetic distinctiveness should be considered in future conservation assessments.

The critically endangered pygmy sloth (*B. pygmaeus*) is restricted to Isla Escudo de Veraguas, in the islands of Bocas del Toro (Panama). Anderson and Handley (2001) described this insular population as a distinct species on the basis of morphometric analyses showing a reduced body size compared with the mainland and other island populations of the brown-throated three-toed sloth (*Bradypus variegatus*). Our analyses, based on the sequencing of the type specimen (USNM 579179), show that *B. pygmaeus* constitutes a distinct lineage within three-toed sloths that is clearly separated from *Bradypus tridactylus* and *B. variegatus* (fig. 1) from which it diverged some 8 Ma (fig. 2 and table 2). By sequencing a palethroated three-toed sloth (*B. tridactylus*) specimen from French Guiana, where only this species occurs, we were able to confirm that the GenBank reference mitogenome (NC_006923) was originally misidentified as *B. tridactylus* and in fact belongs to the brown-throated three-toed sloth (*B. variegatus*), as previously shown by Moraes-Barros et al. (2011). Indeed, our newly sequenced *B. variegatus* specimen from Peru (MVZ 155186) is 97.3% identical to NC_006923, whereas the pairwise distance with *B. tridactylus* reaches 9.5%. The divergence date between *B. variegatus* and *B. tridactylus* is estimated here around 6 Ma (fig. 2 and table 2). However,

similar to the Hoffmann's two-toed sloth, the brown-throated three-toed sloth includes divergent mitochondrial lineages (Moraes-Barros et al. 2011; Moraes-Barros and Arteaga 2015) and occurs as far north as Honduras. Therefore, we cannot exclude the possibility that individuals from mainland Central America (Panama, Costa Rica, Nicaragua, and Honduras) not included in this study might belong to the distinct pygmy sloth (*B. pygmaeus*) lineage rather than to *B. variegatus*.

Anteaters (*Pilosa*; *Vermilingua*)

Anteaters are the least diverse xenarthran group with only four described species classified in three genera. The mitogenomic tree (fig. 1) confirms the phylogenetic distinctiveness of the monotypic pygmy anteater (*Cyc. didactylus*) from the two closely related genera *Tamandua* and *Myrmecophaga*. The early divergence of the pygmy anteater is estimated around 38 Ma, whereas *Tamandua* and *Myrmecophaga* diverged much more recently around 13 Ma (fig. 2 and table 2). This profound dichotomy appears fully compatible with previous estimates (Delsuc et al. 2004, 2012) and confirms the rationality of dividing anteaters into two distinct families Cyclopedidae (*Cyclopes*) and Myrmecophagidae (*Tamandua* and *Myrmecophaga*), as proposed by Barros et al. (2008). This taxonomic distinction also reflects the numerous morphological differences observed between the two main anteater lineages (Gaudin and Branham 1998). Regarding lesser anteaters, our new sequence of the southern tamandua (*Tamandua tetradactyla*) from French Guiana confirmed the identification of the GenBank reference mitogenome (NC_004032), the two sequences being 99.4% identical (fig. 1). However, our mitogenomic data revealed a very limited genetic differentiation between northern (*Tamandua mexicana*) and southern (*Ta. tetradactyla*) tamanduas, with a pairwise distance of only 2.8% (2.1% on Cytochrome c oxidase subunit 1; COX1). The two tamanduas are considered two distinct species with parapatric distributions separating on each side of the Northern Andes in Venezuela, Colombia, Ecuador, and Peru (Superina, Miranda, et al. 2010). However, the species diagnoses are based on differences in coat coloration, body size, skull characters, and number of caudal vertebrae that can be quite variable within populations (Wetzel 1985). Our whole mitochondrial genome data question the species status of these anteaters and encourage future nuclear studies aiming at delimitating species within the genus *Tamandua*.

Armadillos (*Cingulata*; *Dasyopoda*)

Armadillos are the most speciose xenarthran group with 21 extant species and 9 genera. Our phylogenetic results unambiguously support the monophyly of each of the four armadillo subfamilies Dasyopodinae, Euphractinae, Chlamyphorinae, and Tolypeutinae (fig. 1), as previously recognized with a concatenation of nuclear and mitochondrial genes (Delsuc et al. 2012). The first dichotomy separates long-nosed armadillos (Dasyopodinae) from the other three subfamilies, which form a monophyletic group ($PP_{CAT} = 0.99/PP_{PART} = 1.0/BP_{PART} = 100$). This deep divergence occurred early in the Cenozoic at an estimated time of about 45 Ma (fig. 2 and table 2). Given this remarkably ancient divergence

date relative to other placental families (Meredith et al. 2011), we propose splitting armadillos into two distinct families: Dasypodidae and Chlamyphoridae. The proposed use of Chlamyphoridae is based on the taxonomic rank elevation of the oldest constitutive subfamily that is Chlamyphorinae Bonaparte 1850. Within Chlamyphoridae, the mitogenomic tree (fig. 1) supports the grouping of Chlamyphorinae (fairly armadillos) with Tolypeutinae (giant, three-banded, and naked-tailed armadillos) to the exclusion of Euphractinae (hairy armadillos), in line with previous studies including nuclear noncoding (Möller-Krull et al. 2007) and protein-coding (Delsuc et al. 2012) sequences. The early branching of Euphractinae is estimated around 37 Ma, relatively quickly followed by the separation between Chlamyphorinae and Tolypeutinae, circa 33 Ma (fig. 2 and table 2).

Dasypodinae. This subfamily includes seven species of long-nosed armadillos belonging to the single genus *Dasypus*. Species in this genus are characterized by an elongated nose that can be functionally related to the use of the tongue to gather ants, termites, and a diversity of soil invertebrates (Loughry and McDonough 2013). An unusual particularity, thought to be shared by all species belonging to this genus, is their reproduction by obligate polyembryony in which the female systematically gives birth to genetically identical litters (Galbreath 1985). The most common and best studied is the nine-banded armadillo (*Dasypus novemcinctus*), which has the largest distribution from Argentina to North America as a consequence of its ongoing invasion of the southern United States (Taulman and Robbins 2014; Feng and Papes 2015). The greater long-nosed armadillo (*Dasypus kappleri*) is the largest of the group, and it has been proposed on morphological grounds to classify this species in its own subgenus *Hyperoambon* (Wetzel and Mondolfi 1979). The hairy long-nosed armadillo (*Da. pilosus*), which is an endemic of Peru, is also morphologically distinctive in being the only armadillo possessing a carapace entirely covered with dense fur. This peculiarity has led some authors to propose its taxonomic distinction in the subgenus *Cryptophractus* (Wetzel and Mondolfi 1979). A recent morphological study, which was mainly based on the analysis of the structure of its osteoderms, even proposed to raise it to the genus level (Castro et al. 2015). The remaining species constitute a complex of morphologically similar taxa with historical taxonomic uncertainty. The southern long-nosed armadillo (*Dasypus hybridus*) and the seven-banded armadillo (*Dasypus septemcinctus*) are particularly hard to distinguish, with globally parapatric distributions that might overlap in southern Brazil, northern Argentina, and Paraguay (Abba and Superina 2010). The species status of the northern long-nosed armadillo (*Dasypus sabanicola*) has also been historically hard to establish (Wetzel 1985), and the Yunga's lesser long-nosed armadillo (*Dasypus yepesi*) has only been recently recognized as a distinct species (Vizcaino 1995).

The phylogenetic tree obtained from the mitogenomes (fig. 1) clearly identifies *Da. kappleri* as the sister group to all other long-nosed armadillo species. Molecular dating, estimates its early divergence at more than 12 Ma (fig. 2

and table 2). This fairly ancient date, coupled with well-characterized morphological differences such as the presence of unique scutes on the knee, would argue for its placement in the distinct genus *Hyperoambon*, originally proposed as a subgenus by Wetzel and Mondolfi (1979). Second to diverge within Dasypodinae is a clade composed of *Da. hybridus* and *Da. septemcinctus* whose mitogenome sequences appear almost identical (99.3% identity). This observation is consistent with noted morphological similarity and historical synonymy between these two taxa (Abba and Superina 2010). However, given the fact that our *Da. septemcinctus* species is from northern Argentina and our *Da. hybridus* is from Uruguay, both in potential areas of sympatry between the two taxa, we cannot exclude the possibility of misidentification, hybridization, and/or introgression being responsible for the observed mitogenomic similarity. Further clarification of the taxonomic status of these two species would require collecting additional mitochondrial and nuclear data for specimens coming from the two extremes of their ranges in central Argentina for *Da. hybridus* and northern Brazil for *Da. septemcinctus*.

The next diverging lineage is represented by an individual identified as *Da. novemcinctus* from French Guiana, which unambiguously represents a distinct branch in the mitogenomic tree (fig. 1). In an early phylogeographic study of nine-banded armadillos based on the mitochondrial D-loop, we observed that individuals from French Guiana were indeed very distant from the ones of the invasive US populations to which the reference mitochondrial genome for *Da. novemcinctus* belongs (Huchon et al. 1999). The two mitogenomes are indeed fairly divergent with a pairwise distance of 5.6% (5.7% on COX1). The divergence date between the French Guianan lineage and other long-nosed armadillos is estimated around 3.7 Ma (fig. 2 and table 2), which strongly suggests that the French Guianan lineage might represent a previously unrecognized species. This potentially new species is the sister group of a clade regrouping *Da. yepesi* and *Da. sabanicola* on one side, and *Da. pilosus* and *Da. novemcinctus* on the other side. Only the position of the hairy long-nosed armadillo (*Da. pilosus*) appears unstable with low statistical support (fig. 1). Although *Da. sabanicola* and *Da. yepesi* are restricted to very distinct localities of South America respectively in Venezuela/Colombia and north-eastern Argentina, they possess very similar mitogenomes (98.7% identity). Our results confirm that the taxonomic status of both species is questionable and needs further review (Abba and Superina 2010). Further morphological and molecular species delimitation studies will be needed to fully understand the species boundaries within long-nosed armadillos.

Finally, the hairy long-nosed armadillo (*Da. pilosus*) appears to constitute a distinct lineage of long-nosed armadillos (fig. 1), but its molecular divergence does not seem to match its morphological distinctiveness. Indeed, its pairwise distance with both *Da. novemcinctus* and with the *Da. sabanicola/ Da. yepesi* clade is about 5%. The divergence date between *Da. pilosus* and *Da. novemcinctus* is estimated around 2.8 Ma (fig. 2 and table 2). Our phylogenetic reconstruction also strongly contradicts the

results of Castro et al. (2015) who found *Da. pilosus* to be the sister group to all other long-nosed armadillos based on cladistic analysis of morphological characters. Based on these results and the peculiar structure of its osteoderms, they proposed resurrecting *Cryptophractus* as the genus name for the hairy long-nosed armadillo. Our results do not support such a taxonomic reassessment and argue in favor of conserving the hairy long-nosed armadillo in the genus *Dasybus*. More generally, our mitogenomic topology for long-nosed armadillos reveals many conflicts with the cladistic analysis of Castro et al. (2015). This suggests that osteodermal characters are of limited taxonomic value in being highly homoplastic, and should therefore be used with caution in phylogenetic analyses.

Euphractinae. Euphractine armadillos constitute an ecologically homogeneous and morphologically similar group with five traditionally recognized species classified in the three genera *Chaetophractus*, *Euphractus*, and *Zaedyus* (Abba and Superina 2010). The genera *Euphractus* and *Zaedyus* are monospecific and include the six-banded armadillo (*E. sexcinctus*) and the pichi (*Zaedyus pichiy*), respectively. The genus *Chaetophractus* classically encloses three species, the large hairy armadillo (*Chaetophractus villosus*), the screaming hairy armadillo (*Chaetophractus vellerosus*), and the Andean hairy armadillo (*Chaetophractus nationi*). The interrelationships between the three genera have been difficult to decipher with both morphological (Engelmann 1985; Gaudin and Wible 2006) and molecular data (Delsuc et al. 2002, 2003; Möller-Krull et al. 2007) likely due to their rapid diversification (Delsuc et al. 2004). A recent study investigated the phylogenetic relationships among all five species using an integrative approach based on skull geometric morphometrics and molecular phylogenetics (Abba et al. 2015). It was proposed that *Cha. nationi* should be considered a synonym of *Cha. vellerosus* based on shared mitochondrial haplotypes and a close proximity at the nuclear level coupled with a very similar morphology because the two species only differ in size. Moreover, phylogenetic analyses of a combination of six noncoding nuclear markers and two nuclear exons suggested the paraphyly of the genus *Chaetophractus*, with *Cha. vellerosus* being more closely related to *Z. pichiy* than to *Cha. villosus*. The relative positions of the large hairy armadillo (*Cha. villosus*) and the six-banded armadillo (*E. sexcinctus*) nevertheless remained uncertain, as conflicting positions were obtained with the noncoding and protein-coding partitions (Abba et al. 2015).

The complete mitochondrial genome sequences confirm that the threatened and geographically restricted Andean hairy armadillo (*Cha. nationi*) could not be genetically distinguished from the widespread screaming hairy armadillo (*Cha. vellerosus*), with 99.8% mitogenomic identity between the two taxa. This result reinforces the proposition of taxonomically synonymizing these two species by retaining only *Cha. vellerosus* (Abba et al. 2015). Moreover, the mitogenomic tree (fig. 1) offers some additional support for the paraphyly of the genus *Chaetophractus* caused by the strongly supported sister group relationship of *Z. pichiy* with *Cha. vellerosus/Cha.*

nationi ($PP_{CAT} = 0.99/PP_{PART} = 1/BP_{PART} = 93$). As with nuclear data, the early branching of *E. sexcinctus* is slightly less supported by ML but nevertheless appears quite robust ($PP_{CAT} = 0.97/PP_{PART} = 1/BP_{PART} = 80$). Molecular dating results (fig. 2 and table 2) also confirm a rapid diversification of Euphractinae, with speciation events occurring within a few million years between 8 and 11 Ma, a period marked by the appearance of more arid areas in the Southern Cone of South America where most of these species are distributed. Definitive resolution of the relationships among hairy armadillos might require large-scale genomic data to account for possible discordances between gene trees and the species tree resulting from incomplete lineage sorting.

Chlamyphorinae. This recently recognized subfamily (Delsuc et al. 2012) consists of only two species of fairy armadillos or pichiciegos that count among the most elusive mammals due to their nocturnal and subterranean habits. The pink fairy armadillo (*Chl. truncatus*) is restricted to sandy plains of central Argentina, while the greater fairy armadillo (*Calyptophractus retusus*) is found in the Gran Chaco of northern Argentina, Paraguay, and eastern Bolivia (Abba et al. 2010). The mitogenomic tree (fig. 1) corroborates earlier results by strongly supporting the monophyly of fairy armadillos (Delsuc et al. 2012) and their sister group relationship with tolpeutines (Möller-Krull et al. 2007; Delsuc et al. 2012). Molecular dating (fig. 2 and table 2) also confirms the considerably old divergence of the two species (ca. 19 Ma) and their ancient split from tolpeutine armadillos (ca. 33 Ma). These results underline again the phylogenetic distinctiveness of the two fairy armadillo species and argue in favor of their classification in two distinct genera within their own subfamily. The phylogenetic uniqueness of fairy armadillos, combined with their scarcity in the wild, make pleas for increased conservation attention of these atypical mammals.

Tolypeutinae. This subfamily includes seven species classified in three genera. Two tribes are classically recognized (Wetzel 1985; McKenna and Bell 1997): Priodontini grouping the giant armadillo (*Priodontes maximus*) with naked-tailed armadillos of the genus *Cabassous*, and Tolypeutini consisting solely of two species of three-banded armadillos (genus *Tolypeutes*). Giant and naked-tailed armadillos are typically fossorial and are equipped with large anterior claws used for digging. Three-banded armadillos are ground dwelling and morphologically distinctive in being the only armadillos capable of entirely rolling into a ball by locking their carapace as a defensive strategy. In contrast to morphological data that always favored the monophyly of the *Priodontes* and *Cabassous* genera on the basis of numerous anatomical similarities (Engelman 1985; Gaudin and Wible 2006) and characteristically spoon-shaped spermatozoa (Cetica et al. 1998), the phylogenetic relationships within the family Tolypeutinae have been notoriously difficult to resolve with molecular data (Delsuc and Douzery 2008). The concatenation of nuclear exons and two mitochondrial genes has basically left the issue unresolved (Delsuc et al. 2002, 2003, 2012), whereas analyses of noncoding retroposon flanking sequences offered some support for a sister group relationship between

Cabassous and *Tolypeutes* (Möller-Krull et al. 2007). The mitogenomic picture (fig. 1) is congruent with noncoding nuclear data in supporting the paraphyly of the tribe Priodontini by grouping *Tolypeutes* with *Cabassous* to the exclusion of *Priodontes* ($PP_{CAT} = 1/PP_{PART} = 1/BP_{PART} = 88$). This suggests that the morphological characters related to fossoriality used to define this tribe might either have been acquired convergently by giant and naked-tailed armadillos or, more probably, represent symplesiomorphies inherited from a fossorial ancestor.

Concerning Tolypeutini, we collected the first molecular data for the flagship Brazilian three-banded armadillo (*Tolypeutes tricinctus*). This endangered endemic of the North-Eastern Brazilian Caatinga biome was chosen as a mascot to increase awareness about biodiversity and ecosystem conservation. Our mitochondrial genome data revealed an unexpectedly high sequence divergence with its sister species, the southern three-banded armadillo (*Tolypeutes matacus*). The pairwise distance between the two mitogenomes of these morphologically and ecologically similar species reaches 12% (11.9% on COX1). Accordingly, molecular dating estimated a deep divergence of circa 14 Ma between the two allopatrically distributed species (fig. 2 and table 2). The considerable phylogenetic distinctiveness revealed for the Brazilian three-banded armadillo reinforces the conservation concerns expressed for a species considered to be one of the most threatened Brazilian mammals (Feijó et al. 2015).

Our mitogenomic study is the first to include all four recognized species of the conspicuous and fossorial naked-tailed armadillos. The greater naked-tailed armadillo (*Cabassous tatouay*) is the first to diverge, followed by the Chacoan naked-tailed armadillo (*Cabassous chacoensis*) and the two closely related northern (*Cabassous centralis*) and southern (*Cabassous unicinctus*) naked-tailed armadillos (fig. 1). The emergence of *Cab. tatouay* appears quite ancient (ca. 11 Ma), as is the separation of *Cab. chacoensis* from *Cab. centralis* and *Cab. unicinctus* (ca. 9 Ma), which are estimated to have diverged much more recently, less than 2 Ma (fig. 2 and table 2). The mitogenomes of the northern and southern naked-tailed armadillos appear very similar in sequence (98.0% identity), with a pairwise distance of only 1.9% based on COX1. This situation is reminiscent of the case revealed between the northern and southern tamanduas, with closely related species presenting parapatric distributions in Central and South America only interrupted by the Northern Andes. Additional nuclear studies would be warranted for further defining the taxonomic status of *Cab. centralis* and *Cab. unicinctus* that appear only weakly differentiated based on their mitogenomes.

Diversification and Historical Biogeography of Xenarthra

The strongly resolved tree obtained for all living xenarthran species allowed us to derive a reference timescale that can be used to study the patterns and processes underlying their diversification. Seeking the causes of species diversification and extinction by teasing apart the role of abiotic (e.g.,

physical environmental changes) and/or biotic (e.g., species interactions) factors (Benton 2009), is now made possible by the use of different birth–death models (Morlon 2014). We applied a suite of diversification models to sequentially consider the effects of past environmental changes (Condamine et al. 2013), rate variation through time (Stadler 2011; Rabosky 2014), and diversity-dependent processes (Etienne et al. 2012) on the macroevolutionary history of Xenarthra.

It was previously proposed that xenarthran diversification has been influenced by paleoenvironmental changes triggered by Andean uplift and sea level fluctuations in South America during the Cenozoic (Delsuc et al. 2004). At that time, no models of diversification integrating the effect of environmental variables were available to formally test the synchronicity of some cladogeneses with periods of cooling and Andean uplift. Such explicit models are now available (Condamine et al. 2013), and when applied to our data, the best temperature-dependent model showed that the speciation rate over the entire xenarthran timetree correlates negatively with temperature (fig. 3, table 3, and supplementary table S2, Supplementary Material online). This pattern is the opposite of the one found, for instance, in Cetacea (Condamine et al. 2013), and may be explained by the fact that a number of rapid speciation events in the xenarthran tree and especially within armadillos occurred in the last 10–15 Ma, during a period of intense cooling (Zachos et al. 2001). This continuous drop in temperature since the middle Miocene, followed by the setup of the *circum* Antarctic current and the last Andean uplift phase (Garzione et al. 2008), caused the aridification of South America and the formation of dry biomes such as Caatinga and Cerrado in the North, and Chaco and pampas in the Southern Cone (Simon et al. 2009; Hoorn et al. 2010).

The Bayesian analysis of macroevolutionary mixtures (BAMM, Rabosky 2014) show that a single macroevolutionary rate explains the diversification of the group over time (table 3 and supplementary fig. S3, Supplementary Material online), and that the net diversification rate tends to increase through time driven by a higher speciation rate in the last 15 Ma (fig. 4). This corroborates the results obtained with the paleoenvironmental model. Furthermore, time-dependent diversification analyses (table 3 and supplementary table S2, Supplementary Material online) also portray Xenarthra as an old and species-poor, but nevertheless successful clade with a low diversification rate throughout the Cenozoic characterized by a high species turnover driven by an intermediate, but constant extinction rate. The BAMM and TreePar models are thus congruent on inferring no detectable diversification rate shift and constant extinction through time, but they disagree on the speciation rate inference for which only BAMM estimates an increase over time. This discrepancy in estimates might be explained by the relatively small size of the xenarthran timetree, which includes only 32 tips, but may increase with the addition of extinct xenarthran genomes. The inference of extinction on the xenarthran phylogeny is nevertheless in agreement with the fossil record, which documents a relatively high rate of extinction in this clade (Patterson and Pascual 1968; Simpson 1980). Notably here, extinction is

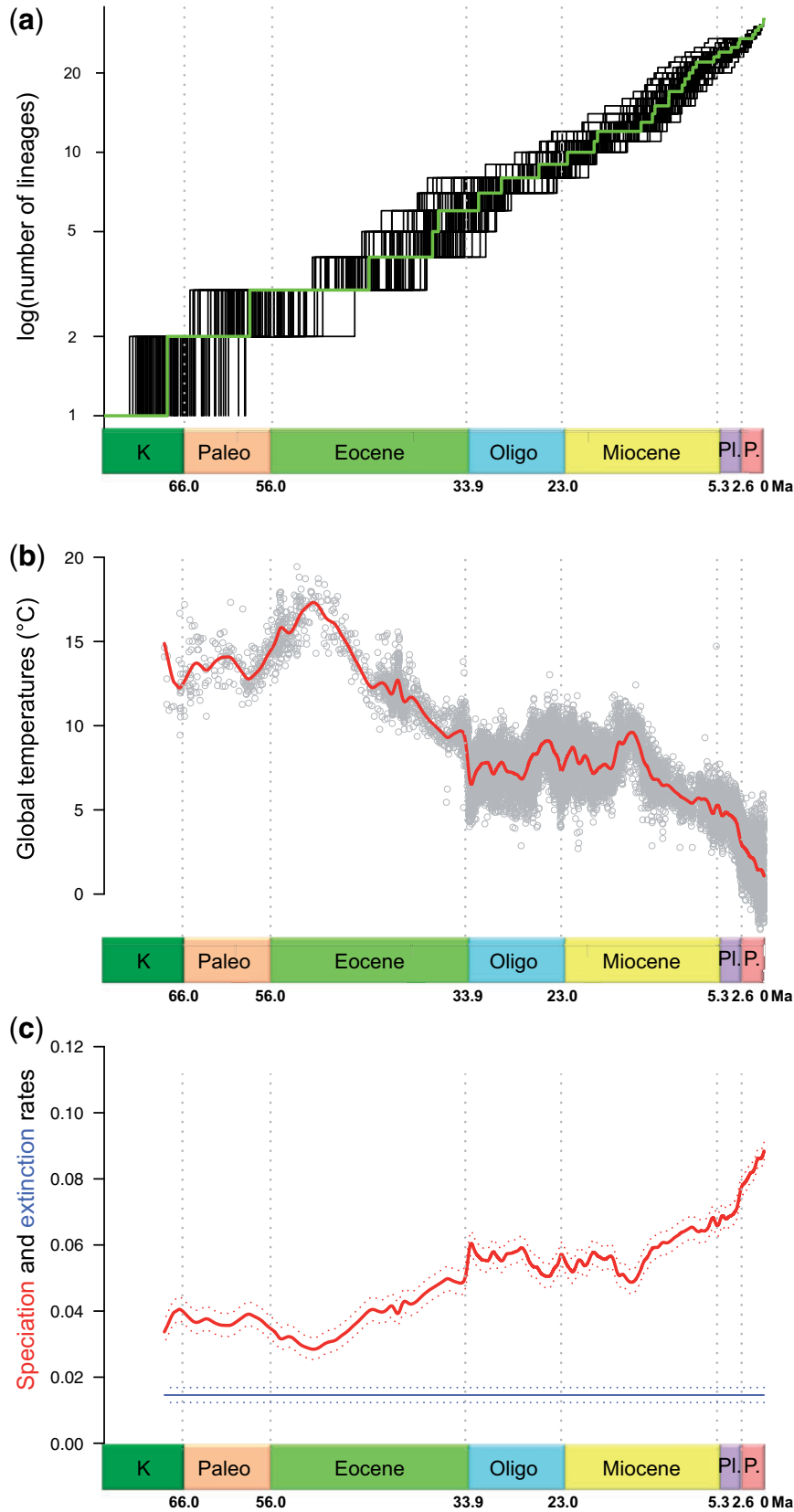


Fig. 3. Diversification pattern of Xenarthra. (a) Lineages-through-time plot constructed from 100 Bayesian posterior trees showing a steady accumulation of species through time. (b) Past fluctuations of temperatures over the Cenozoic (data plotted from Zachos et al. 2001, 2008). (c) Speciation (bold curve) and extinction rates through time for xenarthrans obtained from the relationship between diversification and paleotemperatures estimated using the approach of Condamine et al. (2013). The best model indicates a negative correlation between speciation and past temperatures and no dependence on extinction. K = Cretaceous; Paleo. = Paleocene; Oligo. = Oligocene; Pl. = Pliocene; P. = Pleistocene.

Table 3. Summary of Diversification Analyses Results.

Type of Birth-Death	Method Used	Reference	Data Used	Settings	Result
Paleoenvironmental dependence (rates vary “continuously” as a function of time)	RPANDA (<i>fit_env</i>)	Condamine et al. (2013)	100 posterior chronograms	7 ML models testing whether rates vary or not (exponential and linear variation)	Speciation is “negatively” linked to past temperatures, and “constant extinction”
Among clade and time variation of rates	BAMM	Rabosky (2014)	Bayesian chronogram	Bayesian model testing rate shift(s) among clade and through time (Poisson prior = 1.0)	No significant rate shift detected: Speciation increased through time, “constant extinction”
Time dependence (rates vary “discretely” as a function of time)	TreePar (<i>bd.shifts.optim</i>)	Stadler (2011)	100 posterior chronograms	4 ML models testing from no rate shift to 3 rate shifts	No global rate shift detected, a constant birth-death is supported
Diversity dependence (rates vary as a function of the number of species)	DDD (<i>dd_ML</i>)	Etienne et al. (2012)	Bayesian chronogram	5 ML models testing whether speciation declines with diversity and/or extinction increases with diversity	The clade has reached its carrying capacity, with extinction increasing as diversity increases

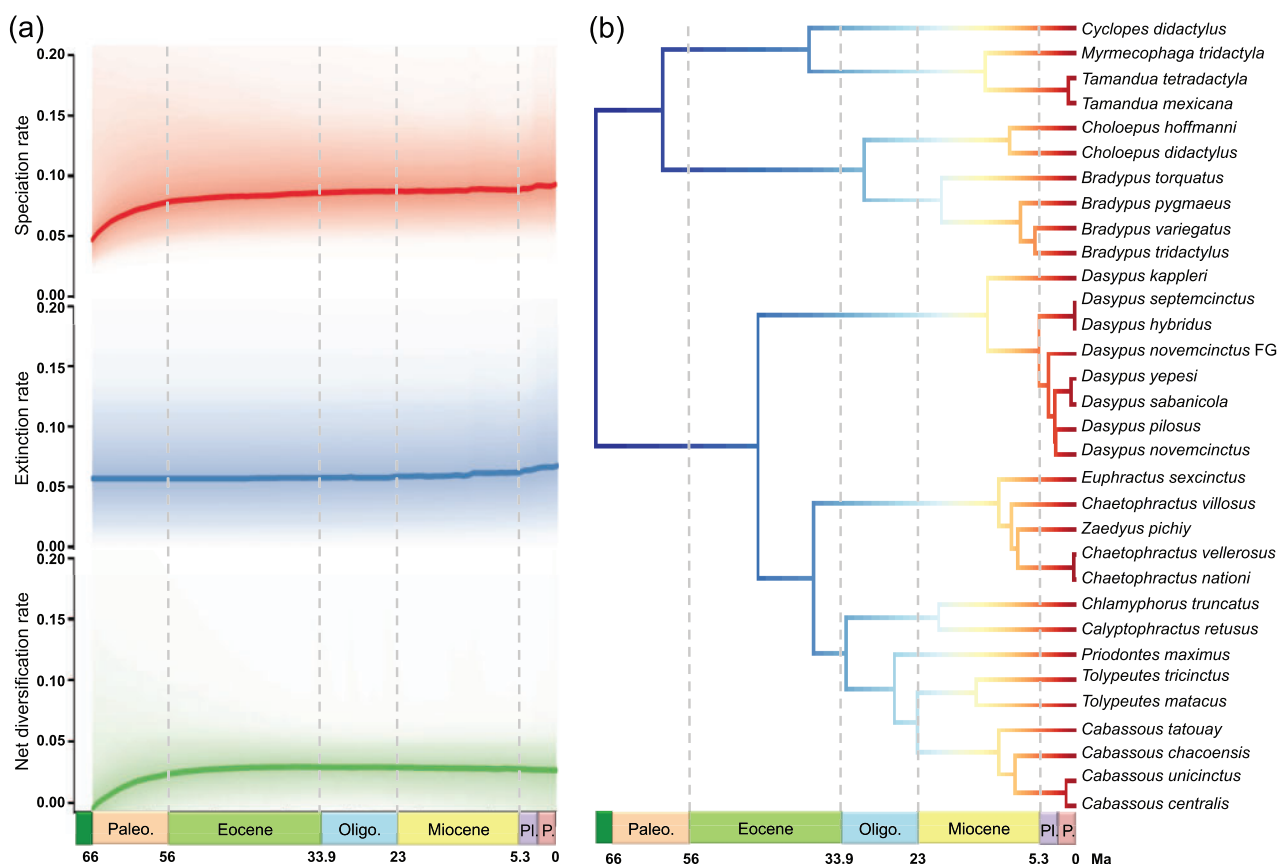


FIG. 4. Bayesian analysis of macroevolutionary mixtures in Xenarthra. (a) Bayesian reconstruction of rate variations in speciation, extinction, and net diversification through time. (b) Maximum a posteriori probability shift configuration represented as a phylorate plot showing variations in speciation rates (cool colors = slow, warm = fast) along each branch of the xenarthran phylogeny. Each unique color section of a branch represents the mean of the marginal posterior density of speciation rates on a localized segment of a phylogenetic tree. The rate variation pattern for lineages involves a uniform, although slight, temporal acceleration in speciation rates. “*Dasyopus novemcinctus* FG” denotes the French Guiana lineage. Paleo. = Paleocene; Oligo. = Oligocene; Pl = Pliocene; P = Pleistocene.

estimated to be high but nevertheless constant, confirming that these models are able to detect extinction signal even in species-poor clades (Morlon et al. 2011; Jansa et al. 2014; Beaulieu and O’Meara 2015).

Interestingly, xenarthrans do not seem to have been particularly affected by the GABI, with, for instance, giant sloths and glyptodonts successfully colonizing Central and North America. Xenarthra was finally depauperated by the

extinction of the largest forms at the terminus of the Pleistocene (Lyons et al. 2004), a very recent event that is hardly detectable with current methods of macroevolutionary analyses. This effect is especially resonant for Pilosa in which the successful northern emigrants such as giant ground sloths were drawn to extinction (McDonald 2005). Ultimately, fossil data would have to be integrated into diversification analyses to adequately model the macroevolutionary history of this species-poor clade (Fritz et al. 2013; Silvestro et al. 2014).

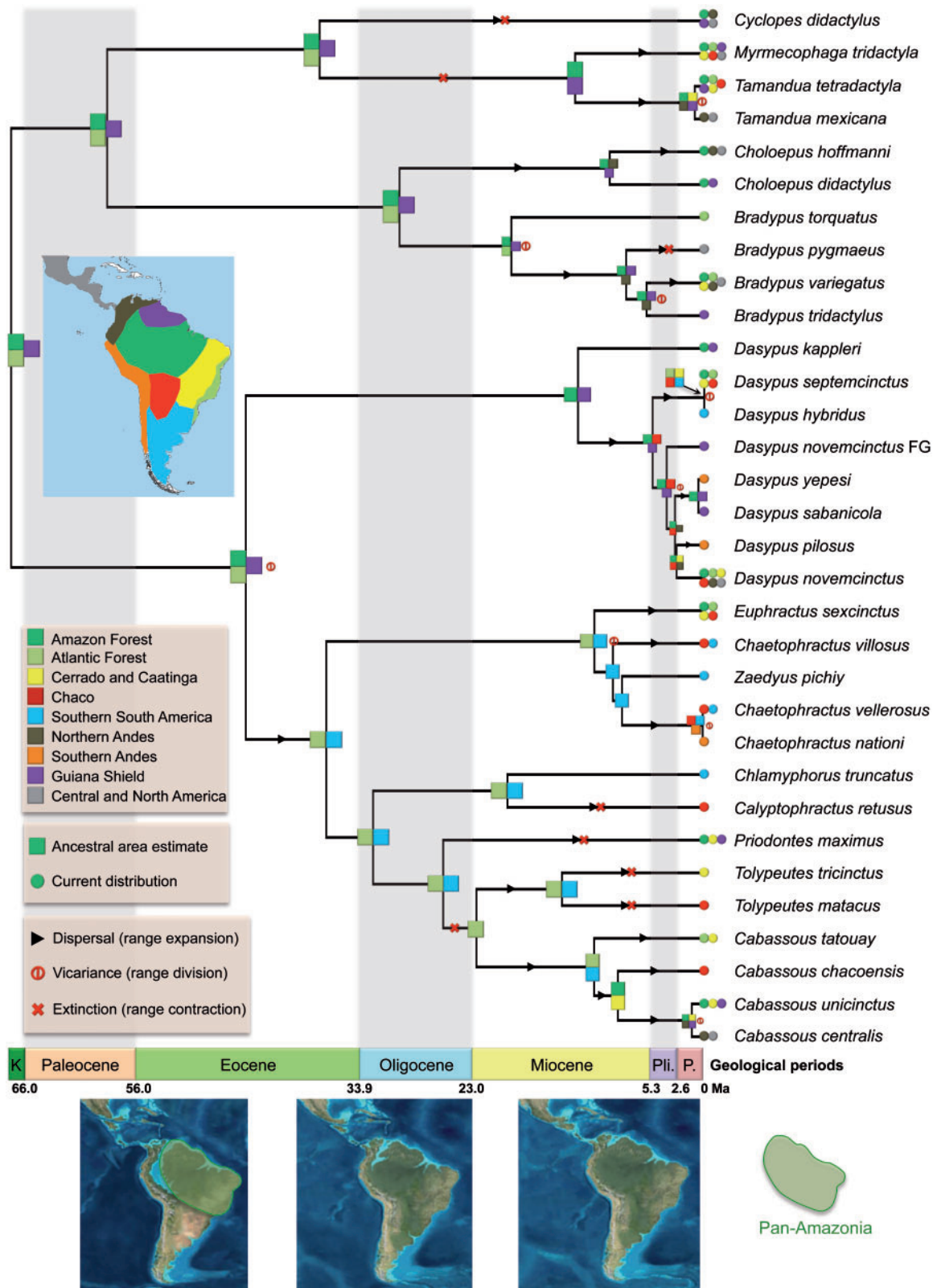
The role of biotic interactions was finally assessed using the diversity-dependent diversification (DDD) model of Etienne et al. (2012). The DDD model depicts Xenarthra as a clade that has reached its carrying capacity, with extinction increasing as diversity increases (table 3 and supplementary table S2, Supplementary Material online). Thus, niche partitioning in the Neotropics may be a dominant factor in shaping the pattern of species richness in Xenarthra. Evidence from current geographical distributions shows that in clades such as Tolyptinae and Dasypodinae, only little overlap in distribution ranges is observed (Abba and Superina 2010; Superina, Miranda, et al. 2010; Superina, Plese, et al. 2010). In line with these observations, a detailed study has recently identified the tropical rainforest in the Amazon Basin as an area of high ecological diversity for xenarthrans, indicating a high disparity between pairs of coexisting species (Fergnani and Ruggiero 2015).

Biogeographic analyses could provide further insight into this question, and help understand the role of niche partitioning in time and the consequences on the resulting current biodiversity. We thus used the Dispersal-Extinction-Cladogenesis (DEC, Ree and Smith 2008) model on the xenarthran timetree to estimate ancestral biogeographic ranges by taking into account the connectivity among areas through time, as well as the dispersal abilities between areas according to the regional biome evolution (fig. 5). The biogeographic analysis identifies Pan-Amazonia (tropical lowland rainforest of Amazonia and Guiana Shield) and Atlantic forest as the cradle of Xenarthra evolution for most of the Paleogene (left map on fig. 5). This fits well with the high phylogenetic and ecological diversities observed for xenarthrans in the Amazonian region (Fergnani and Ruggiero 2015). Nonetheless, the common ancestor of Chlamyphoridae (Euphractinae, Chlamyphorinae, and Tolyptinae) subsequently dispersed toward the Southern Cone in the late Eocene (central map on fig. 5). We also estimated that different species of armadillos colonized Central America after the closing of the Isthmus of Panama (Pliocene), but not in the middle or late Miocene when land connections were first made (Montes et al. 2015). Interestingly, only 9 of the 31 phylogenetic events are explained by allopatric speciation, 4 of which involved the Andes and 5 other biome divergences, such as tropical forest and savannah. Moreover, most of the allopatric speciation events are recent (last 8 My), indicating an important role of vicariance due to the building of the northern Andes, especially in northern Colombia for species with parapatric distributions in South and Central America,

such as *Cab. unicinctus*/*Cab. centralis* or *Ta. tetradactyla*/*Ta. mexicana* (Moraes-Barros and Arteaga 2015).

The pattern inferred within Pilosa and Cingulata shows notable differences in the biogeographic processes. Regarding allopatric speciation (vicariance)/range expansion (dispersal)/range contraction (local extinction), we found 3/8/3 events for Pilosa versus 5/17/5 for Cingulata. These results are suggestive of a relatively stable biogeographic history in the clade Pilosa, and a more dynamic history in Cingulata. Indeed anteaters and sloths appeared more stable in time and centered on the Pan-Amazonian region (Amazonia and Guiana Shield) and the Atlantic forest. The formation of the Cerrado and Caatinga isolated Amazonia from the Atlantic Forest about 9 Ma (Simon et al. 2009) and might be associated with the diversification of *B. torquatus* restricted to the Atlantic forest and its sister group constituted of the other three-toed sloths (*B. pygmaeus*, *B. tridactylus*, and *B. variegatus*) within Pan-Amazonia further expanding into Central America. The final rise of the Northern Andes probably explains the vicariance between the two *Tamandua* species. These findings corroborate patterns of Xenarthra's diversification discussed by Moraes-Barros and Arteaga (2015) who proposed a Western South America origin for *Bradypus*. These authors also pointed to a West to East dispersal through Amazonia for *B. variegatus* with a later colonization of the Atlantic forest during the Pleistocene.

On the contrary, armadillos display a more dynamic pattern with many inferred events of range expansions within all groups, also compensated by several local extinctions. For instance, we evidenced six independent colonizations of the Chaco region that likely occurred during the middle to late Miocene cooling and the aridification from the southern Amazon region (right map on fig. 5). The Chaco is indeed an area where current taxonomic richness is high in Xenarthra (Fergnani and Ruggiero 2015). The creation of new open habitats (biomes) in response to this general cooling (Simon et al. 2009) probably also promoted the diversification of Chlamyphoridae, especially within Euphractinae, with many species being now restricted to the different biomes: *Z. pichiy* in the Southern Cone (semiarid steppe grasslands); *Chl. truncatus* in the Central Desert; *Cab. chacoensis*, *Cal. retusus*, and *To. matacus* in the Chaco; *To. tricinctus* in the Caatinga; and *Cab. tatouay* in the Cerrado (Anacleto et al. 2006; Abba et al. 2012). Few species have a large repartition covering different biomes, such as *P. maximus*, *Cab. unicinctus*, and *E. sexcinctus*. The concurrent colonizations of arid areas and positive correlation of xenarthran diversification with cooler temperatures (cooler climate favors drier conditions) indicate that our diversification and biogeographical results are concordant and corroborate the invasion of the Southern cone within the last 10–15 My from a tropical origin. This is consistent with the reported increase in ecological diversity observed for xenarthrans in this region (Fergnani and Ruggiero 2015). Overall, the historical biogeography of Xenarthra is best explained by a progressive biome specialization of species due to the Cenozoic differentiation of biomes toward the present that probably led to more opportunities to disperse and diversify. The diversity-dependence pattern is likely attributed



Downloaded from <http://mbe.oxfordjournals.org/> at BUPHARM on February 19, 2016

FIG. 5. Historical biogeography of living xenarthrans. The biogeographical range estimation was inferred under the Dispersal-Extinction-Cladogenesis model taking into account the change of connectivity and dispersal ability between areas defined as the main biomes of the American continent. Paleogeographic maps depict the tectonic evolution of South America adapted from Blakey (2008). “*Dasyopus novemcinctus* FG” denotes the French Guiana lineage. K = Cretaceous; Pli. = Pliocene; P. = Pleistocene.

to allopatry within clades and secondary sympatry established between clades.

Conclusions

Our exhaustive data set establishes Xenarthra as the first major clade of placental mammals to be completely sequenced at the species level for mitogenomes from both modern and archival tissues. These data provide a reference phylogenetic framework and timescale, setting the scene for studying the diversification and biogeography of xenarthrans. The phylogenetic scaffold defined here will also be particularly useful for phylogenetic analyses based on ancient DNA of the numerous recently extinct taxa contained in this group. Finally, the wealth of molecular data generated here will be important for forthcoming studies on the phylogeography, species delimitation, barcoding, and conservation efforts of this understudied group of placentals.

Materials and Methods

Biological Samples

The complete taxon sampling we used in this study is detailed in [table 1](#). The samples notably include several tissues linked to specimen vouchers deposited in International Natural History Museums including a type specimen (*B. pygmaeus*; USNM 579179). Most other tissue samples are conserved in the Mammalian Tissue Collection of the Institut des Sciences de l'Evolution de Montpellier (Catzeffis 1991). The maned sloth (*B. torquatus*) sample was collected under Brazilian permit Ibama/MMA 19267-3/14597869, and the Brazilian three-banded armadillo (*To. tricinctus*) with permit Ibama/MMA 42354-1.

DNA Extraction, Illumina Library Preparation, and Sequencing

All DNA extractions and library preparations took place in a separate, dedicated clean room (Biobubble) of McMaster Ancient DNA Centre, used strictly for low template samples.

DNA Extractions

For bone samples, up to 50 mg of bone materials were split into small pieces (1–5 mm) that were demineralized with 0.75 ml of 0.5 M EDTA pH 8 at room temperature overnight with agitation, and the supernatants removed following centrifugation. For soft tissue samples, up to 50 mg were cut into small pieces (1–5 mm). The one tissue preserved in Dimethylsulfoxid (DMSO) was washed multiple times with 0.1 × Tris-EDTA buffer (TE) pH 8 and blotted dry to remove this reducing agent before further processing. Pelleted bone and tissue samples were then digested with 0.5 ml of a Tris HCl–based proteinase K solution with 20 mM Tris–Cl pH 8, 0.5% sodium lauryl sarcosine (Fisher Scientific), 5 mM calcium chloride, 1% polyvinylpyrrolidone (Fisher Scientific), 50 mM dithiothreitol, 2.5 mM N-phenacyl thiazolium bromide (Prime Organics), and 250 µg/ml proteinase K. Proteinase digestion was performed at 55 °C for a couple of hours, with agitation. Following centrifugation the digestion supernatants were removed and extracted of organics using phenol:chloroform:isoamyl alcohol (25:24:1), and the resulting postcentrifugation aqueous solution was extracted with

chloroform. We then concentrated the final aqueous phase with 30 kDA Amicon Ultra 0.5 ml centrifugal filters (Millipore) at 14,000 × g, with three washes using 0.1 × TE buffer pH 8 plus 0.05% Tween-20 to provide a desalted concentrate of 50 µl.

Genomic DNA Fragmentation

For 26 extracts, DNA fragmentation was done by digestion of 1 µg of DNA extract with NEBNext dsDNA Fragmentase (New England Biolabs). The reactions were purified using the MinElute Polymerase Chain Reaction (PCR) Purification kit (Qiagen) and eluted in 20 µl buffer EB. The remaining seven extracts were sonicated using a Covaris S220 according to manufacturer's protocol for a median fragment length of 200 bp with a reduced input volume of 50 µl and a maximum of 10 µg of DNA (5 or 10 µl of extract).

Library Preparation and Indexing

We used between 0.1 and 1 µg of sheared DNA in Illumina library preparations as described in Meyer and Kircher (2010) with the modification of a blunt-end repair reaction volume of 40 µl or 70 µl and replacing all SPRI bead clean-ups with MinElute purifications to 20 µl buffer EB. We did not heat deactivate the Bst polymerase following the fill-in step and instead purified the reaction with MinElute to 20 µl buffer EB. A first set of libraries was index amplified using the common P5 and a set of unique P7 indexing primers (Meyer and Kircher 2010) in 50 µl reactions consisting of 1 × PCR buffer II, 2.5 mM MgCl₂, 250 µM Deoxynucleotide (dNTP) mix, 200 nM each forward (P5) and reverse (P7) primer, 2.5 U Taq Gold, and 2 µl (100 ng) of template library. Thermal cycling conditions were as follows: initial denaturation at 95 °C for 4 min, 12 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 10 min. Amplifications were performed using a MJ thermocycler (BioRad). A second set of libraries was dual-index-amplified using unique P5 and P7 indexing primers (Kircher et al. 2012) in 50 µl reactions consisting of 1 × AccuPrime Pfx Reaction mix, 0.5 × EvaGreen, 500 nM each forward (P5) and reverse (P7) primer, 1.25 U AccuPrime Pfx DNA Polymerase, and 5 µl (250–2,500 ng) of template library. Thermal cycling conditions were as follows: initial denaturation at 95 °C for 2 min, 13 cycles of 95 °C for 15 s, 60 °C for 30 s, 68 °C for 30 s, and a final extension at 68 °C for 10 min. Amplifications were performed in real time with a CFX96 Real-time PCR platform (BioRad). All indexed libraries were finally purified with MinElute to 50 µl EB.

Illumina Sequencing

The 33 libraries were sequenced on 3 different lanes. The first run including 19 libraries was processed on an Illumina Genome Analyzer IIx using 72 bp single-end reads. The second run of four libraries was also processed on an Illumina Genome Analyzer IIx but using 72 bp paired-end reads. These first two runs were subcontracted to Ambry Genetics (Aliso Viejo, CA, USA). The third run of 10 libraries was run on an Illumina HiSeq 2500 instrument with 100 bp paired-end reads at the Donnelly Sequencing Centre of the University of Toronto (Canada). Initial data processing and base calling, including extraction of cluster intensities, was

done using RTA1.8 (SCS version 2.8). Sequence quality filtering script was executed in the Illumina CASAVA software (ver 1.7.0, Illumina, Hayward, CA).

Mitochondrial Genomes Assembly

Raw sequence reads were first trimmed to remove any adapter and index tag sequences using CutAdapt (Martin 2011). De novo assembly of the trimmed reads was performed using the ABySS assembler (Simpson et al. 2009) with default parameters. We used a range of kmers to optimize contig lengths and identical contigs resulting from the use of different kmers were identified and collapsed using CD-HIT (Fu et al. 2012). Mitochondrial contigs were then extracted using BLASTN similarity searches against the closest reference mitogenome. If not already full length, contigs were assembled into complete mitochondrial genomes with Geneious R7 (Kearse et al. 2012). Using Geneious, mitogenomes were checked by mapping the trimmed reads to the mitogenome assemblies with the low sensitivity option, and scanned by eye to confirm appropriate mapping, particularly in regions with repeats. Any gaps in the contig assemblies were progressively filled by extending the contigs using iterations of the mapping procedure on the consensus sequence as implemented in Geneious. Final mitochondrial genomes were annotated by alignment with published xenarthran mitogenomes, and protein-coding regions were checked to confirm no indels or stop codons were present. The 33 new xenarthran mitogenomes have been deposited in GenBank under accession numbers KT818523–KT818555.

Phylogenetic Reconstructions

The 33 new mitochondrial genomes were added to the 4 available xenarthran mitogenomes and to 3 afrotherian species that were used as outgroups: the dugong (*Dugong dugon*; NC_003314), the African savannah elephant (*Loxodonta africana*; NC_000934), and the aardvark (*Orycteropus afer*; NC_002078). Individual mitochondrial genes were aligned with the MAFFT v7.017 plugin (Katoh and Standley 2013) in Geneious, using the amino acid translation for protein-coding genes. Unambiguously aligned sites were then selected by Gblocks (Castresana 2000) with default relaxed settings and codon options for protein-coding genes. The final data set contained 14,917 sites for 40 taxa and is available as [supplementary material](#). PartitionFinder v1.1 (Lanfear et al. 2012) was used to find the optimal partition schemes and models of sequence evolution for RAxML, using the greedy algorithm starting from 42 a priori defined partitions corresponding to the 3 codon positions of each of the 13 protein-coding genes, 12S ribosomal RNA (rRNA), 16S rRNA, and the combined transfer RNAs (tRNAs). Branch lengths have been unlinked among partitions and the Bayesian information criterion was used for selecting the best-fitting partition scheme.

ML inference was implemented with RAxML v7.8 (Stamatakis et al. 2008) using separate general time-reversible models with gamma distribution for each of the four best-fit partitions selected by PartitionFinder. Statistical reliability of the ML tree was evaluated with nonparametric bootstrapping

(100 replications) through the Thorough Bootstrap option of RAxML under the optimal partitioned model to obtain ML bootstrap percentages (BP_{PART}).

Bayesian phylogenetic inference under a mixed model was conducted using the MPI version of MrBayes 3.2.3 (Ronquist et al. 2012) using separate GTR+G8+I models for each of the four selected partitions, as determined by PartitionFinder, with parameters being unlinked across partitions. Two independent runs of four incrementally heated MCMCMC starting from a random tree were performed. MCMCMC was run for 10,000,000 generations, with trees and associated model parameters being sampled every 1,000 generations. The initial 2,500 trees in each run were discarded as burn-in samples after convergence checking. The 50% majority-rule Bayesian consensus tree and the associated posterior probabilities (PP_{PART}) were then computed from the 15,000 combined trees sampled in the 2 independent runs. Bayesian phylogenetic reconstruction under the CAT-GTR-G4 mixture model (Lartillot and Philippe 2004) was conducted using PhyloBayes-MPI 1.5a (Lartillot et al. 2013). Two independent Markov Chain Monte Carlo (MCMC) starting from a random tree was run for 35,000 cycles (2,400,000 tree generations), with trees and associated model parameters being sampled every 10 cycles. The initial 350 trees (10%) sampled in each MCMC run were discarded as burn-in after checking for convergence in both likelihood and model parameters (tracecomp subprogram), and clade posterior probability (bpcomp subprogram). The 50% majority-rule Bayesian consensus tree and the associated posterior probabilities (PP_{CAT}) were then computed from the remaining combined 6,300 (2 × 3,150) trees using bpcomp.

Molecular Datings

Molecular dating analyses were performed under a Bayesian relaxed molecular framework using PhyloBayes 3.3f (Lartillot et al. 2009). In all dating calculations, the tree topology was fixed to the majority-rule consensus tree previously inferred in Bayesian analyses. Dating analyses were conducted using the site-heterogeneous CAT-GTR+G4 mixture model and a relaxed clock model, as recommended by Lartillot et al. (2009) with a birth–death prior on divergence times (Gernhard 2008) combined with soft fossil calibrations (Yang and Rannala 2006). We used the following five, well-justified afrotherian and xenarthran calibration intervals defined by Meredith et al. (2011): 1) Paenungulata (maximum age 71.2 Ma, minimum age 55.6 Ma); 2) Xenarthra (maximum age 71.2 Ma, minimum age 58.5 Ma); 3) Pilosa (maximum age 65.5 Ma, minimum age 31.5 Ma); 4) Folivora (maximum age 40.6 Ma, minimum age 15.97 Ma); and 5) Vermilingua (maximum age 61.1 Ma, minimum age 15.97 Ma). We also added a recently proposed calibration point within armadillos based on the oldest armadillo fossil skull identified as a stem Tolypeutinae and found in the late Oligocene of Desadean in Bolivia at 26 Ma (Billet et al. 2011). This finding allows us to set the minimum age for Tolypeutinae at a conservative 23.0 Ma, corresponding to the upper boundary of late Oligocene. The maximum age for the origin of Tolypeutinae was set at 37.8

Ma using the lower boundary of the late Eocene, because the oldest fossils of the closest outgroup (Euphractinae) trace back to the Casamayoran, at least 36 Ma (Kay et al. 1999). The prior on the root of the tree (Placentalia) was set at 100 Ma according to Meredith et al. (2011).

Although Lepage et al. (2007) and Rehm et al. (2011) showed that the autocorrelated LN relaxed clock model generally offers the best fit, we compared it with the UGAM model (Drummond et al. 2006) and a strict molecular clock (CL) model. These three clock models were compared against each other using the same prior settings (see above) in a cross-validation procedure as implemented in PhyloBayes. The cross-validation tests were performed by dividing the original alignment in 2 subsets of 13,426 sites (learning set) and 1,491 sites (test set). The overall procedure was repeated over 10 random splits for which a MCMC chain was run on the learning set for a total 1,100 cycles sampling posterior rates and dates every cycle. The first 100 samples of each MCMC were excluded as the burn-in period for calculating the cross-validation scores averaged across the 10 replicates. The final dating calculations were conducted under the best fitting model by running two independent MCMC chains for a total 50,000 cycles, sampling posterior rates and dates every 10 cycles. The first 500 samples (10%) of each MCMC were excluded as the burn-in after checking for convergence in both likelihood and model parameters. Posterior estimates of divergence dates were then computed from the remaining 4,500 samples of each MCMC using the readdiv subprogram. The values reported in table 2 are averages over the 2 independent chains.

Diversification Analyses

We used either the Bayesian chronogram or 100 randomly sampled chronograms obtained from the post burn-in posterior distribution of the PhyloBayes dating analyses to estimate diversification rates with different methods. These trees have been restricted to 32 taxa to better reflect the current xenarthran species diversity by excluding redundant taxa. To visualize the tempo and mode of diversification of the group, we first reconstructed lineages-through-time plots and then used a suite of ML models and a Bayesian model of diversification (Morlon 2014). For each type of ML diversification model, we computed the corrected Akaike information criterion (AICc). We then checked the support for the selected model against all models nested within it using likelihood ratio test (LRT). The scenario supported by LRT with the lowest AICc was considered the best fit. Bayes factors were used to assess model fit in the Bayesian framework.

Paleoenvironment-Dependent Diversification Model

To test the effect that past environmental change might have had on the diversification of Xenarthra, we used a model derived from the one of Morlon et al. (2011) that allows speciation and extinction rates to vary according to an environmental variable, which itself varies through time (Condamine et al. 2013), such as past variations in temperature (Zachos et al. 2001, 2008). Following the approach of Morlon et al. (2011), we designed four models to be tested:

- 1) BCSTDCST, a time constant birth–death model (null model);
- 2) BVARDCST, speciation rate is exponentially varying with temperature and extinction rate is constant;
- 3) BCSTDVAR, speciation rate is constant and extinction rate is exponentially varying with temperature;
- and 4) BVARDVAR, speciation and extinction rates are both exponentially varying with temperature.

We also repeated these three rate-variable models with a linear dependence with temperature (because we have no prior expectation about how speciation or extinction might vary with temperature).

These models rely on a past environmental variable describing how the environment varied through time. For temperature, we relied on the well-known Cenozoic temperature curves published by Zachos et al. (2001, 2008). We used the R package *pspline* to reconstruct smooth lines of the paleo-data for the environmental variable. In other words, a smooth line is introduced in the birth–death model to represent the paleoenvironment through time, and at each point in time the model refers to this smooth line to obtain the value of the temperature. Given the dated phylogeny, the model then estimates speciation and extinction rates as a function of this value (Condamine et al. 2013).

Time-Dependent Diversification Models

We assessed whether diversification rates remained constant during the evolutionary history of Xenarthra. We first used BAMM to estimate speciation and extinction rates through time along the xenarthran phylogeny (Rabosky 2014). BAMM allows studying complex evolutionary processes on phylogenetic trees, potentially shaped by a heterogeneous mixture of distinct evolutionary dynamics of speciation and extinction across clades. The method is designed to automatically detect rate shifts and sample distinct evolutionary dynamics that best explain the whole diversification dynamics of the clade. In BAMM, the speciation rate is allowed to vary exponentially through time while extinction is maintained constant (Rabosky, Donnellan, et al. 2014). Subclades in the tree might diversify faster (or slower) than others, and BAMM allows detecting these diversification rate shifts and comparing how many and where these shifts might occur. BAMM is implemented in a C++ command line program and the BAMMtools R package (Rabosky, Grundler, et al. 2014). We set 4 MCMC running for 10 million generations and sampled every 10,000 generations. Other parameters were set to default values except the Poisson process prior that we set to 1.0 following the authors' recommendation (Rabosky, Grundler, et al. 2014). We performed four independent runs (with a burn-in of 15%) using different seeds, and we used effective sample size to assess the convergence of the runs, considering values above 200 as indicating convergence. The posterior distribution was used to compute the best global rates of diversification through time, to estimate the configuration of the diversification rate shifts by evaluating alternative diversification models as compared by Bayes factors.

The TreePar package (Stadler 2011) was used to assess speciation and extinction rates through time, and to specifically detect potential rapid and global changes in diversification rates that might be due to environmental factors, such as

climatic shifts following major geological events. We employed the “bd.shifts.optim” function that allows estimating discrete changes in speciation, extinction rates, and mass extinction events in undersampled phylogenies (Stadler 2011). At each time t , the rates are allowed to change and the species may undergo a shift in diversification. TreePar analyses were run with the following settings: start = 0, end = crown age estimated by dating analyses, grid = 0.1 Myr, and posdiv = FALSE to allow the diversification rate to be negative (i.e., allows for periods of declining diversity).

Diversity-Dependent Diversification Model

We investigated whether lineages diversified rapidly in their early stages and have reached equilibrium (or are being bounded), suggesting that diversity is saturated toward the present, as niches became occupied and diversification rates slowed down. We used the method of Etienne et al. (2012) to explore the effect of diversity on speciation and extinction rates. The function “dd_ML” was used to fit five models: 1) speciation declines linearly with diversity and no extinction (DDL), 2) speciation declines linearly with diversity and extinction (DDL + E), 3) speciation declines exponentially with diversity and extinction (DDX + E), 4) extinction increases linearly with diversity (DD + EL), and 5) extinction increases exponentially with diversity (DD + EX). The initial carrying capacity was set to the current species diversity, and the final carrying capacity was estimated according to the models and parameters.

Biogeographic Reconstructions

The ancestral range estimation was performed using BioGeoBEARS (Matzke 2014). The analyses were carried out using the same time-calibrated phylogeny as used for the diversification analyses. We used the DEC model (Ree and Smith 2008) to conduct local optimizations and estimate the ancestral character state of each node according to the current distribution and biogeographical model the user introduces in the analyses. The root was left unconstrained but optimized by the method. We did not assess whether the inclusion of the founder-event speciation (parameter J) significantly improved the likelihood because DEC-J is appropriate for island-dwelling clades (Matzke 2014).

A geographic model was incorporated to include operational areas that are defined as geographic ranges shared by two or more species and delimited by geological or oceanic features, which may have acted as barriers to dispersal. The distribution of anteaters, armadillos, and sloths ranges from southern North America to southern South America (Abba and Superina 2010; Superina, Plese, et al. 2010; Superina, Miranda, et al. 2010). We further divided these two regions into smaller biogeographic identities to obtain higher resolution in the inference of the ancestral area of origin. Using tectonic reconstructions, notably the evolution of past Amazonian landscapes (e.g., Hoorn et al. 2010), the model comprised nine component areas: 1) Amazonia (Amazon Forest plus Amazon Basin); 2) Mata Atlántica (Atlantic Forest along the Brazilian coast); 3) Cerrado (tropical savannah) and Caatinga (desert and xeric shrublands); 4) Chaco

region and Pantanal (semiarid lowland and adjacent wetlands); 5) southern South America (grasslands); 6) Southern and Central Andes (Chile, western Bolivia, and Peru); 7) Northern Andes (Ecuador, Colombia, and western Venezuela); 8) Guiana Shield (eastern Venezuela, Guyana, Suriname, French Guiana), and 9) Central and North America (from Panama to southern USA). The adjacency matrix was designed while taking into account the geological history and the biological plausibility of combined ranges (Hoorn et al. 2010). We discarded ranges larger than six areas in size that were not subsets of observed species ranges (*Da. novemcinctus* had the largest range with six areas). Distributional data were compiled from monographs and IUCN data (Abba and Superina 2010; Superina, Plese, et al. 2010; Superina, Miranda, et al. 2010).

We did not split our sample into smaller geographic areas for several reasons. The complexity of the geological history of this region makes it difficult to accurately reconstruct past distributions of land and sea at any given point in time. For instance, many uncertainties remain about the appearance and disappearance of the Pebas System, or the exact timing of the rising of the Andes (Hoorn et al. 2010). Moreover, our goal was to investigate the dominant biogeographical processes that shaped the xenarthran distribution pattern. For the latter, we were particularly interested in assessing the relative contributions of vicariance and dispersal.

Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work largely benefited from the help of the following individuals and institutions who provided tissue samples: François Catzeflis (Institut des Sciences de l'Evolution, Montpellier, France), Jean-François Mauffrey, Philippe Gaucher, Eric Hansen, François Ouhoud-Renoux, Jean-Christophe Vié, Philippe Cerdan, Michel Blanc, and Rodolphe Paowé (French Guiana), Sergio Vizcaíno (Museo de La Plata, La Plata, Argentina), Jorge Omar García and Rodolfo Rearte (Complejo Ecológico Municipal, Presidencia Roque Sáenz Peña, Argentina), Daniel Hernández (Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay), John Trupkiewicz (Philadelphia Zoo, USA), Darrin Lunde (National Museum of Natural History, Washington, USA), Jim Patton and Yuri Leite (Museum of Vertebrate Zoology, Berkeley, USA), Ross MacPhee (American Museum of Natural History, New York, USA), Jonathan Dunnun and Joseph Cook (Museum of Southwestern Biology, Albuquerque, USA), Donna Dittman and Mark Hafner (Louisiana State University Museum of Natural Science, Baton Rouge, USA), Gerhard Haszprunar and Michael Hiermeier (Zoologische Staatssammlung München, Munich, Germany), Géraldine Véron (Museum National d'Histoire Naturelle, Paris, France), Agustín Jiménez-Ruiz, Guido Valverde, and Guillermo Pérez-Jimeno. Our thanks also

extend to Trish McLenachan and David Penny who sequenced and made freely available the first sloth mitogenomes, Lionel Hautier and Philippe Gaubert for taxonomical advice, and two anonymous referees for helpful comments. This work was supported by grants from the EU's Seventh Framework Programme (No 286431) to N.M.-B., the Centre National de la Recherche Scientifique (CNRS), the Scientific Council of Université Montpellier 2 (UM2), and Investissement d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01) to F.D., and the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs program to H.N.P. This is contribution ISEM 2015-227-S of the Institut des Sciences de l'Evolution de Montpellier.

References

- Abba AM, Cassini GH, Valverde G, Tilak MK, Vizcaíno SF, Superina M, Delsuc F. 2015. Systematics of hairy armadillos and the taxonomic status of the Andean hairy armadillo (*ChaetophRACTUS nationi*). *J Mammal*. 96:673–689.
- Abba AM, Superina M. 2010. The 2009/2010 armadillo Red List assessment. *Edentata* 11:135–184.
- Abba AM, Tognelli MF, Seitz VP, Bender JB, Vizcaíno SF. 2012. Distribution of extant xenarthrans (Mammalia: Xenarthra) in Argentina using species distribution models. *Mammalia* 76:123–136.
- Amrine-Madsen H, Koepfli KP, Wayne RK, Springer MS. 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol Phylogenet Evol*. 28:225–240.
- Anacleto TCS, Diniz JAF, Vital MVC. 2006. Estimating potential geographic ranges of armadillos (Xenarthra, Dasypodidae) in Brazil under niche-based models. *Mammalia* 70:202–213.
- Anderson RP, Handley CO. 2001. A new species of three-toed sloth (Mammalia: Xenarthra) from Panama, with a review of the genus *Bradypus*. *Proc Biol Soc Wash*. 114:1–33.
- Arnason U, Adegok JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci U S A*. 99:8151–8156.
- Arnason U, Gullberg A, Janke A. 1997. Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. *Mol Biol Evol*. 14:762–768.
- Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proc Natl Acad Sci U S A*. 112:6110–6115.
- Barros MC, Sampaio I, Schneider H. 2003. Phylogenetic analysis of 16S mitochondrial DNA data in sloths and anteaters. *Genet Mol Biol*. 26:5–12.
- Barros MC, Sampaio I, Schneider H. 2008. Novel 12S mtDNA findings in sloths (Pilosa, Folivora) and anteaters (Pilosa, Vermilingua) suggest a true case of long branch attraction. *Genet Mol Biol*. 31:793–799.
- Beaulieu JM, O'Meara BC. 2015. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* 69:1036–1043.
- Benton MJ. 2009. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science* 323:728–732.
- Billet G, Hautier L, de Muizon C, Valentin X. 2011. Oldest cingulate skulls provide congruence between morphological and molecular scenarios of armadillo evolution. *Proc R Soc Lond B*. 278:2791–2797.
- Blakey RC. 2008. Gondwana paleogeography from assembly to breakup—a 500 million year odyssey. In: Fielding CR, Frank TD, Isbell JL, editors. *Resolving the late Paleozoic ice age in time and space*. Boulder (CO): Geological Society of America Special Paper 441. p. 1–28.
- Botero-Castro F, Tilak MK, Justy F, Catzeflis F, Delsuc F, Douzery EJP. 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol*. 69:728–739.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.
- Castro MC, Ciancio MR, Pacheco V, Salas-Gismondi RM, Bostelmann JE, Carlini AA. 2015. Reassessment of the hairy long-nosed armadillo "*Dasybus pilosus*" (Xenarthra, Dasypodidae) and revalidation of the genus *CryptophRACTUS* Fitzinger, 1856. *Zootaxa* 3947:30–48.
- Catzefflis FM. 1991. Animal tissue collections for molecular genetics and systematics. *Trends Ecol Evol*. 419:565–571.
- Cetica PD, Solari AJ, Merani MS, de Rosas JC, Burgos MH. 1998. Evolutionary sperm morphology and morphometry in armadillos. *J Submicrosc Cytol Pathol*. 30:309–314.
- Chiari Y, Cahais V, Galtier N, Delsuc F. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol*. 10:65.
- Churakov G, Kriegs JO, Baertsch R, Zemann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res*. 19:868–875.
- Condamine FL, Rolland J, Morlon H. 2013. Macroevolutionary perspectives to environmental change. *Ecol Lett*. 16:72–75.
- Delsuc F, Catzeflis FM, Stanhope MJ, Douzery EJP. 2001. The evolution of armadillos, anteaters, and sloths depicted by nuclear and mitochondrial phylogenies: implications for the status of the enigmatic fossil *Eurotamandua*. *Proc R Soc Lond B*. 268:1605–1615.
- Delsuc F, Douzery EJP. 2008. Recent advances and future prospects in xenarthran molecular phylogenetics. In: Vizcaíno SF, Loughry WJ, editors. *The Biology of the Xenarthra*. Gainesville: University Press of Florida. p. 11–23.
- Delsuc F, Scally M, Madsen O, Stanhope MJ, De Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol Biol Evol*. 19:1656–1671.
- Delsuc F, Stanhope MJ, Douzery EJP. 2003. Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes. *Mol Phylogenet Evol*. 28:261–275.
- Delsuc F, Superina M, Tilak MK, Douzery EJP, Hassanin A. 2012. Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Mol Phylogenet Evol*. 62:673–680.
- Delsuc F, Vizcaíno SF, Douzery EJP. 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evol Biol*. 4:11.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:699–710.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 29:1969–1973.
- Engelmann GF. 1985. The phylogeny of the Xenarthra. In: Montgomery GG, editor. *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas*. Washington and London: Smithsonian Institution Press. p. 51–64.
- Enk J, Devault A, Debruyne R, King CE, Treangen T, O'Rourke D, Salzberg SL, Fisher D, MacPhee R, Poinar H. 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol*. 12:R51.
- Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, Purvis A, Phillimore AB. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc R Soc Lond B*. 279:1300–1309.
- Fabre PH, Jönsson KA, Douzery EJP. 2013. Jumping and gliding rodents: mitogenomic affinities of Pedetidae and Anomaluridae deduced from an RNA-Seq approach. *Gene* 531:388–397.
- Fabre PH, Vilstrup JT, Raghavan M, Der Sarkissian C, Willerslev E, Douzery EJP, Orlando L. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museumics. *Biol Lett*. 10.

- Feijó A, Garbino GST, Campos B ATP, Rocha PA, Ferrari SF, Langguth A. 2015. Distribution of *Tolypeutes* Illiger, 1811 (Xenarthra: Cingulata) with comments on its biogeography and conservation. *Zool Sci.* 32:77–87.
- Feng X, Papes M. 2015. Ecological niche modelling confirms potential north-east range expansion of the nine-banded armadillo (*Dasypus novemcinctus*) in the USA. *J Biogeogr.* 42:803–807.
- Fernani PN, Ruggiero A. 2015. Ecological diversity in South American mammals: their geographical distribution shows variable associations with phylogenetic diversity and does not follow the latitudinal richness gradient. *PLoS One* 10:e0128264.
- Finstermeier K, Zinner D, Brameier M, Meyer M, Kreuz E, Hofreiter M, Roos C. 2013. A mitogenomic phylogeny of living primates. *PLoS One* 8:e69504.
- Fritz SA, Schnitzler J, Eronen JT, Hof C, Böhning-Gaese K, Graham CH. 2013. Diversity in time and space: wanted dead and alive. *Trends Ecol Evol.* 28:509–516.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Galbreath GJ. 1985. The evolution of monozygotic polyembryony in *Dasypus*. In: Montgomery GG, editor. *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas*. Washington and London: Smithsonian Institution Press. p. 243–245.
- Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. 2006. Mutation hot spots in mammalian mitochondrial DNA. *Genome Res.* 16:215–222.
- Gardner AL. 2008. Magnorder Xenarthra. In: Gardner AL, editor. *Mammals of South America: Volume 1 Marsupials, Xenarthrans, Shrews, and Bats*. Chicago (IL): University of Chicago Press. p. 127–128.
- Garzone CN, Hoke GD, Libarkin JC, Withers S, MacFadden B, Eiler J, Ghosh P, Mulch A. 2008. Rise of the Andes. *Science* 320:1304–1307.
- Gaudin TJ, Branham DG. 1998. The phylogeny of the Myrmecophagidae (Mammalia, Xenarthra, Vermilingua) and the relationship of *Eurotamandua* to the Vermilingua. *J Mammal Evol.* 5:237–265.
- Gaudin TJ, Wible JR. 2006. The phylogeny of living and extinct armadillos (Mammalia, Xenarthra, Cingulata): A craniodental analysis. In: Carrano MT, Gaudin TJ, Blob RW, Wible JR, editors. *Amniote Paleobiology*. Chicago: The University of Chicago Press. p. 153–198.
- Gissi C, Reyes A, Pesole G, Saccone C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Mol Biol Evol.* 17:1022–1031.
- Gradstein FM, Ogg JG, Schmitz MD, Ogg GM. 2012. *The geologic time scale 2012*, volume set. Boston: Elsevier.
- Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gillissen E, Sonet G, Nagy ZT, et al. 2013. Next-generation museum genomics disentangles one of the largest primate radiations. *Syst Biol.* 62:539–554.
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskoug V, Nguyen TT, et al. 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol.* 335:32–50.
- Haysen V. 2011. *Choloepus hoffmanni* (Pilosa: Megalonychidae). *Mammalian Species* 43:37–55.
- Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution and biodiversity. *Science* 330:927–931.
- Huchon D, Delsuc F, Catzeflis FM, Douzery EJP. 1999. Armadillos exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasypus novemcinctus* (Xenarthra). *Mol Ecol.* 8:1743–1748.
- IUCN Red List of Threatened Species. 2015. Version 2014.3. Available from: www.iucnredlist.org. Accessed on 15 April 2015.
- Jansa SA, Barker FK, Voss RS. 2014. The early diversification history of didelphid marsupials: a window into South America's "Splendid Isolation". *Evolution* 68:684–695.
- Kainer D, Lanfear R. 2015. The effects of partitioning on phylogenetic inference. *Mol Biol Evol.* 2:1611–1627.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kay RF, Madden RH, Vucetich GM, Carlini AA, Mazzoni MM, Re GH, Heizler M, Sandeman H. 1999. Revised geochronology of the Casamayoran South American land mammal age: climatic and biotic implications. *Proc Natl Acad Sci U S A.* 96:13235–13240.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40:e3.
- Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29:695–1701.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095–1109.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 62:611–615.
- Leavitt JR, Hiatt KD, Whiting MF, Song H. 2013. Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: a phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. *Mol Phylogenet Evol.* 67:494–508.
- Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. *Mol Biol Evol.* 24:2669–2680.
- Loughry WJ, McDonough CM. 2013. The nine-banded armadillo: a natural history. Norman (OK): University of Oklahoma Press.
- Lyons SK, Smith FA, Brown JH. 2004. Of mice, mastodons and men: human-mediated extinctions on four continents. *Evol Ecol Res.* 6:339–358.
- Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong WW, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610–614.
- Marshall LG, Webb SD, Sepkoski JJ, Raup DM. 1982. Mammalian evolution and the great American interchange. *Science* 215:1351–1357.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17:10–12.
- Mason VC, Li G, Helgen KM, Murphy WJ. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Res.* 21:1695–1704.
- Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst Biol.* 63:951–970.
- McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species tree analysis. *Genome Res.* 22:746–754.
- McDonald HG. 2005. Paleogeography of extinct xenarthrans and the Great American Biotic Interchange. *Bull Fla Mus Nat Hist.* 45:313–333.
- McKenna MC, Bell SK. 1997. *Classification of mammals above the species level*. New York: Columbia University Press.
- Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334:521–524.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010: pdb.prot5448.
- Mitchell KJ, Pratt RC, Watson LN, Gibb GC, Llamas B, Kasper M, Edson J, Hopwood B, Male D, Armstrong KN, et al. 2014. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Mol Biol Evol.* 31:2322–2330.
- Möller-Krull M, Delsuc F, Churakov G, Marker C, Superina M, Brosius J, Douzery EJP, Schmitz J. 2007. Retroposed elements and their

- flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths). *Mol Biol Evol.* 24:2573–2582.
- Montes C, Cardona A, Jaramillo C, Pardo A, Silva JC, Valencia V, Ayala VC, Pérez-Angel LC, Rodríguez-Parra LA, Ramirez V, et al. 2015. Middle Miocene closure of the Central American seaway. *Science* 348:226–229.
- Moraes-Barros N, Arteaga MC. 2015. Genetic diversity in Xenarthra and its relevance to patterns of Neotropical biodiversity. *J Mammal.* 96:690–702.
- Moraes-Barros N, Silva JA, Morgante JS. 2011. Morphology, molecular phylogeny, and taxonomic inconsistencies in the study of *Bradypus* sloths (Pilosa: Bradypodidae). *J Mammal.* 92:86–100.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecol Lett.* 17:508–525.
- Morlon H, Parsons TL, Plotkin J. 2011. Reconciling molecular phylogenies with the fossil record. *Proc Natl Acad Sci U S A.* 108:16327–16332.
- Murphy WJ, Eizirik E, Johnson WJ, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
- Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.
- Nabholz B, Ellegren H, Wolf JB. 2012. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Mol Biol Evol.* 30:272–284.
- Nabholz B, Glemin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.
- Nabholz B, Glémin S, Galtier N. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol Biol.* 9:54.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A.* 106:5235–5240.
- Novacek MJ. 1992. Mammalian phylogeny: shaking the tree. *Nature* 356:121–125.
- Nyakatura JA. 2012. The convergent evolution of suspensory posture and locomotion in tree sloths. *J Mammal Evol.* 19:225–234.
- Pajjmans JL, Gilbert MTP, Hofreiter M. 2013. Mitogenomic analyses from ancient DNA. *Mol Phylogenet Evol.* 69:404–416.
- Patterson B, Pascual R. 1968. The fossil mammal fauna of South America. *Q Rev Biol.* 43:409–451.
- Powell AF, Barker FK, Lanyon SM. 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. *Mol Phylogenet Evol.* 66:69–79.
- Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9:e89543.
- Rabosky DL, Donnellan SC, Grundler M, Lovette IJ. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Syst Biol.* 63:610–627.
- Rabosky DL, Grundler M, Anderson C, Shi JJ, Brown JW, Huang H, Larson JG. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol Evol.* 5:701–707.
- Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst Biol.* 57:4–14.
- Rehm P, Borner J, Meusemann K, von Reumont BM, Simon S, Hadrys H, Misof B, Burmester T. 2011. Dating the arthropod tree based on large-scale transcriptome data. *Mol Phylogenet Evol.* 61:880–887.
- Reyes A, Gissi C, Pesole G, Saccone C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol Biol Evol.* 15:957–966.
- Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol Biol Evol.* 30:2134–2144.
- Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.
- Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi K, Moritz CC. 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Mol Ecol Res.* 11:1082–1092.
- Silvestro D, Schnitzler J, Liow LH, Antonelli A, Salamin N. 2014. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst Biol.* 63:349–367.
- Simon MF, Grether R, de Queiroz LP, Skema C, Pennington RT, Hughes CE. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc Natl Acad Sci U S A.* 108:20359–20364.
- Simpson GG. 1980. Splendid isolation: the curious history of mammals in South America. New Haven (CT): Yale University Press.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117–1123.
- Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci U S A.* 109:14942–14947.
- Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proc Natl Acad Sci U S A.* 108:6187–6192.
- Stamatakis A, Hoover PM, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol.* 57:758–771.
- Steiner CC, Houck ML, Ryder OA. 2010. Species identification and chromosome variation of captive two-toed sloths. *Zoo Biol.* 29:1–13.
- Superina M, Miranda FR, Abba AM. 2010. The 2010 anteater Red List assessment. *Edentata* 11:96–114.
- Superina M, Pagnutti N, Abba AM. 2014. What do we know about armadillos? An analysis of four centuries of knowledge about a group of South American mammals, with emphasis on their conservation. *Mammal Rev.* 44:69–80.
- Superina MT, Plese N, de Moraes-Barros N, Abba AM. 2010. The 2010 sloth red list assessment. *Edentata* 11:115–134.
- Taulman JF, Robbins LW. 2014. Range expansion and distributional limits of the nine-banded armadillo in the United States: an update of Taulman & Robbins (1996). *J Biogeogr.* 41:1626–1630.
- Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 15:1647–1657.
- Tilak MK, Justy F, Debais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP. 2015. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. *Conserv Genet Res.* 7:37–40.
- Vizcaino SF. 1995. Identificación específica de las “mulitas”, género *Dasyopus* L. (*Mammalia, Dasypodidae*), del noroeste argentino. Descripción de una nueva especie. *Mastozool Neotrop.* 2:5–13.
- Webb SD. 1985. The interrelationships of tree sloths and ground sloths. In: Montgomery GG, editor. *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas*. Washington: Smithsonian Institution. p. 105–112.
- Wetzel RM. 1985. The identification and distribution of recent Xenarthra (= Edentata). In: Montgomery GG, editor. *The evolution and ecology of armadillos, sloths, and vermilinguas*. Washington and London: Smithsonian Institution Press. p. 5–21.
- Wetzel RM, Mondolfi E. 1979. The subgenera and species of long-nosed armadillos, Genus *Dasyopus* L. In: Eisenberg JF, editor. *Vertebrate Ecology in the Northern Neotropics*. Washington: The National Zoological Park, Smithsonian Institution. p. 39–63.
- Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using fossil calibrations with soft bounds. *Mol Biol Evol.* 23:212–226.
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–693.
- Zachos JC, Dickens GR, Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451:279–283.

Correspondence

The phylogenetic affinities of the extinct glyptodonts

Frédéric Delsuc^{1,*}, Gillian C. Gibb^{1,2},
Melanie Kuch³, Guillaume Billet⁴,
Lionel Hautier¹, John Southon⁵,
Jean-Marie Rouillard⁶,
Juan Carlos Fernicola⁷,
Sergio F. Vizcaíno⁸, Ross D.E. MacPhee⁹,
and Hendrik N. Poinar^{3,*}

Among the fossils of hitherto unknown mammals that Darwin collected in South America between 1832 and 1833 during the Beagle expedition [1] were examples of the large, heavily armored herbivores later known as glyptodonts. Ever since, glyptodonts have fascinated evolutionary biologists because of their remarkable skeletal adaptations and seemingly isolated phylogenetic position even within their natural group, the cingulate xenarthrans (armadillos and their allies [2]). In possessing a carapace comprised of fused osteoderms, the glyptodonts were clearly related to other cingulates, but their precise phylogenetic position as suggested by morphology remains unresolved [3,4]. To provide a molecular perspective on this issue, we designed sequence-capture baits using *in silico* reconstructed ancestral sequences and successfully assembled the complete mitochondrial genome of *Doedicurus* sp., one of the largest glyptodonts. Our phylogenetic reconstructions establish that glyptodonts are in fact deeply nested within the armadillo crown-group, representing a distinct subfamily (Glyptodontinae) within family Chlamyphoridae [5]. Molecular dating suggests that glyptodonts diverged no earlier than around 35 million years ago, in good agreement with their fossil record. Our results highlight the derived nature of the glyptodont morphotype, one aspect of which is a spectacular increase in body size until their extinction at the end of the last ice age.

Although the phylogenetic unity of order Cingulata has never been seriously questioned, how its three constituent groups (armadillos, glyptodonts, and pampatheres) are related to one another has been difficult to resolve in fine detail. Of special interest in this regard is the

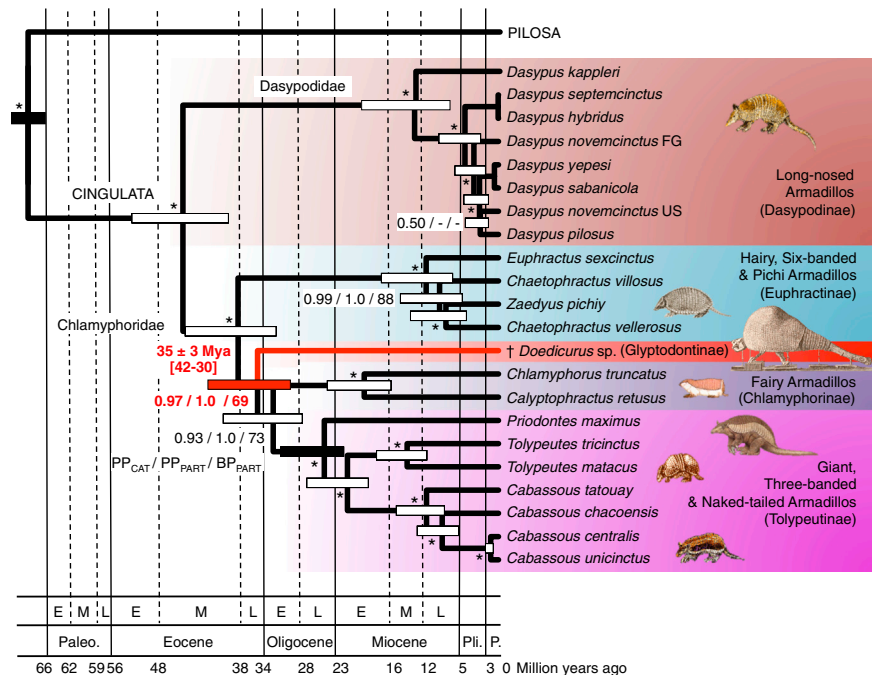


Figure 1. Phylogenetic position of glyptodonts.

Phylogeny and molecular timescale of extant armadillos including the extinct glyptodont *Doedicurus* sp. (in red). Bayesian chronogram was obtained using a rate-autocorrelated log-normal relaxed molecular clock model using PhyloBayes under the CAT-GTR-G mixture model with a birth death prior on the diversification process, and six soft calibration constraints. Mean divergence dates and associated 95% credibility intervals are represented as node bars. Plain black node bars indicated calibration constraints. The main geological periods follow Geological Time Scale of the Geological Society of America (E = Early, M = Middle, L = Late; Paleo. = Paleocene, Pli. = Pliocene, P. = Pleistocene). Statistical support values obtained from three different phylogenetic reconstruction methods (PP_{CAT}: Bayesian Posterior Probability under the CAT-GTR+G mixture model; PP_{PART}: Bayesian PP under the best partition model; BP_{PART}: Maximum likelihood Bootstrap Percentage under the best partition model) are indicated with stars corresponding to nodes with PP > 0.95 and BP > 90. The full chronogram and phylogram are provided in Figure S2.

recent proposal that, despite numerous differences in body size and carapace structure, glyptodonts do not constitute a sister-group to armadillos, as traditionally assumed [2], but are instead nested within them [4,6]. This hypothesis is, however, based on a restricted set of cranio-dental characters. Here, we put this proposition to the test by analyzing the mitochondrial genome of a specimen of the late surviving glyptodont *Doedicurus*. One of the largest members of its clade, with an estimated body mass of ~1.5 tons [7], *Doedicurus* exhibited numerous distinctive characters, famously including a club-shaped, armored tail adorned with spikes, presumably used in intraspecific combat.

Using ancient DNA (aDNA) extraction techniques, we recovered endogenous DNA from a carapace fragment (MACN Pv 6744) dated to 12,015 ± 50 ¹⁴C radiocarbon years before present (Supplemental information). Utilizing

a recently assembled dataset encompassing all modern xenarthran species [5], we reconstructed, *in silico*, a set of ancestral mitogenomic sequences, which permitted the synthesis of a set of target capture RNA baits. Baits constructed in this way may allow for a more specific sequence capture of phylogenetically distant ancient specimens than baits based solely on available modern sequences. This permitted the reconstruction of a nearly complete mitochondrial genome of *Doedicurus* at 76x coverage. Illumina reads mapping to the newly assembled *Doedicurus* mitogenome were 45 base pairs on average and displayed C-to-T damage patterns at both 3' and 5' ends, characteristic of authentic aDNA. We have ruled out the possibility of the inadvertent enrichment of nuclear copies of mitochondrial origin (NUMTs) by performing additional phylogenetic controls (Supplemental information).

By comparing our ancient mitogenome to those of living xenarthrans (Figure 1), we were able to confidently place *Doedicurus* within armadillos as the sister-group of a clade composed of Chlamyphorinae (fairy armadillos) and Tolypeutinae (three-banded, naked-tailed and giant armadillos; Supplemental information). This clearly contradicts the old view that glyptodonts must have diverged from other cingulates at a very early point in their phylogenetic history, on the grounds that, for example, they possessed such features as a completely fused carapace lacking movable bands [3]. Our results are more compatible but still incongruent with recent morphological cladistic analyses [4,6] that position glyptodonts within a more inclusive but nevertheless paraphyletic Euphractinae.

To examine the consequences of this novel phylogenetic placement, we incorporated *Doedicurus* into the morphological character matrix of Billet *et al.* [6], but were unable to identify any exclusive synapomorphies justifying grouping of the former with Chlamyphorinae + Tolypeutinae. In our study, only two characters, pertaining to the shape and position of the mandibular coronoid process, might qualify as potential synapomorphies, but only under the assumption that both have reverted to ancestral states in three-banded armadillos. This analysis nevertheless revealed other morphological similarities between glyptodonts and fairy armadillos (Supplemental information).

We estimate that glyptodonts diverged from Chlamyphorinae + Tolypeutinae 35 ± 3 million years ago, close to the Eocene–Oligocene transition (Figure 1). This molecular estimate is compatible with the age of the oldest and widely accepted glyptodont remains (Mustersan *Glyptatelus* osteoderms [8], ca. 36–38 million years old [9]). Tarsal bones from the Early Eocene locality of Itaboraí (Brazil), currently dated to more than 50 Myr [9]), have been interpreted as glyptodont, but the elements in question are better interpreted as belonging to indeterminate dasypodoids [10]. According to our results, they might belong to stem cingulates that evolved before basal divergences occurred within the armadillo crown group, an event we date to ca. 45 million years ago (Figure 1).

While our results are based strictly on the comparison of mitogenomes,

the global congruence observed with previous nuclear-based phylogenies as well as molecular dating analyses provides convincing evidence for the proposed xenarthran evolutionary history [5]. On this evidence, glyptodonts (Glyptodontinae) comprised a distinct, Late Paleogene lineage of chlamyphorid armadillos [5]. Such a radical repositioning of glyptodonts within the armadillo crown group has major consequences for interpreting aspects of cingulate evolution. For example, the dome-shaped, tightly-fused carapace of glyptodonts has long been thought to be fundamentally different from that of armadillos and pampatheres, in which the carapace consists of articulated sections. Our results imply that the unarticulated carapace is in fact a derived feature, which in turn provides an explanation for the apparent presence of movable bands in some Miocene glyptodonts [3].

Glyptodonts were a group of ambulatory specialized herbivores that reached giant size bracketed between two extant clades of armadillos that do not share either of these characteristics. Based on our new phylogenetic framework, we performed a statistical reconstruction of ancestral body masses. According to our analysis, the mean ancestral body mass estimate of the last common ancestor of Glyptodontinae + Chlamyphorinae + Tolypeutinae was a mere 6 kg (95% credibility interval: 1–19 kg), implying a spectacular increase in glyptodont body mass during the Neogene (Supplemental Information). This inference is in line with the fossil record, which indicates that glyptodonts evolved from medium-sized forms in the Miocene (e.g., *Propalaeohoplophorus*, ~80 kg) to become true megafauna in the Pleistocene (e.g., *Glyptodon clavipes*, ~2,000 kg) before disappearing with most other South American large mammals some 10,000 years ago [7].

SUPPLEMENTAL INFORMATION

Supplemental Information including results, acknowledgements, experimental procedures and two figures can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2016.01.039>.

REFERENCES

1. Fernicola, J.C., Vizcaino, S.F., and De Iuliis, G. (2009). The fossil mammals collected by Charles Darwin in South America during his travels on board the HMS Beagle. *Revist. Asoc. Geol. Argent.* 64, 147–159.

2. Hoffstetter, R. (1958). Xenarthra. In *Traité de paléontologie*, Vol. 2, no. 6., P. Piveteau, ed. (Masson et Cie: Paris), pp. 535–636.
3. Engelmann, G.F. (1985). The phylogeny of the Xenarthra. In *The evolution and ecology of armadillos, sloths, and vermillings*. G.G. Montgomery, ed. (Smithsonian Institution Press: Washington DC), pp. 51–64.
4. Gaudin, T.J., and Wible, J.R. (2006). The phylogeny of living and extinct armadillos (Mammalia, Xenarthra, Cingulata): a craniodental analysis. In *Amniote paleobiology: perspectives on the evolution of mammals, birds and reptiles*, M.T. Carrano, T.J. Gaudin, R.W. Blob, and J.R. Wible, eds. (University of Chicago Press: Chicago), pp. 153–198.
5. Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H.N., and Delsuc, F. (in press). Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. *Mol. Biol. Evol.* <http://dx.doi.org/10.1093/molbev/msv250>.
6. Billet, G., Hautier, L., De Muizon, C., and Valentin, X. (2011). Oldest cingulate skulls provide congruence between morphological and molecular scenarios of armadillo evolution. *Proc. Biol. Sci.* 278, 2791–2797.
7. Vizcaino, S.F., Cassini, G.H., Toledo, N., and Bargo, M.S. (2012). On the evolution of large size in mammalian herbivores of Cenozoic faunas of southern South America. In *Bones, clones and biomes: an 80-million year history of recent Neotropical mammals*, B. Patterson and L. Costa, eds. (University of Chicago Press: Chicago), pp. 76–101.
8. Simpson, G.G. (1948). The beginning of the age of mammals in South America. Part 1. Introduction. *Systematics: Marsupialia, Edentata, Condylarthra, Litopterna, and Notoprogonia*. *Bull. Am. Mus. Nat. Hist.* 97, 1–227.
9. Woodburne, M.O., Goin, F.J., Bond, M., Carlini, A.A., Gelfo, J.N., López, G.M., Iglesias, A., and Zimicz, A.N. (2014). Paleogene land mammal faunas of South America; a response to global climatic changes and indigenous floral diversity. *J. Mammal. Evol.* 21, 1–73.
10. Bergqvist, L.P., Abrantes, E.A., and Dos Santos Avilla, L. (2004). The Xenarthra (Mammalia) of São José de Itaboraí basin (Upper Paleocene, Itaboraian), Rio de Janeiro, Brazil. *Geodiversitas* 26, 323–337.

¹Institut des Sciences de l'Évolution, UMR 5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France. ²Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand.

³McMaster Ancient DNA Centre, Department of Anthropology, Biology and Biochemistry, McMaster University, Hamilton, Canada.

⁴Sorbonne Universités, CR2P, UMR 7207, CNRS, Muséum National d'Histoire naturelle, Université Paris 06, Paris, France. ⁵Keck-Carbon Cycle AMS facility, Department of Earth System Science, University of California, Irvine, CA, USA. ⁶MYcroarray, Ann Arbor, MI 48105, United States / Chemical Engineering Department, University of Michigan, Ann Arbor, MI, USA. ⁷Consejo Nacional de Investigaciones Científicas y Técnicas, Sección Paleontología de Vertebrados, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires, Argentina. ⁸División Paleontología Vertebrados, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, CONICET, La Plata, Argentina. ⁹Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY, USA.

*E-mail: Frederic.Delsuc@umontpellier.fr (F.D.), poinarh@mcmaster.ca (H.N.P.)