# Species delimitation in xenarthran mammals at the genomic scale

Mathilde Barthe

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

**En Génétique et Génomique**

**École doctorale doctorale GAIA - Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau**

**Unité de recherche – Institut des Sciences de l'Evolution de Montpellier (ISEM) – UMR 5554**

## Species delimitation in xenarthran mammals at the genomic scale


**Présentée par Mathilde BARTHE**


**le 21 Juin 2024**

**Sous la direction de Frédéric Delsuc et Benoit Nabholz**

**Devant le jury composé de**

| | |
|---|---|
| Carole Smadja, Directrice de recherche, CNRS, Montpellier | Présidente |
| Violaine Nicolas-Colin, Professeure, MNHN, Paris | Rapportrice |
| Antoine Fouquet, Chargé de Recherche, CNRS, Toulouse | Rapporteur |
| Christelle Fraïsse, Chargée de Recherche, CNRS, Lille | Examinatrice |
| Nicolas Galtier, Directeur de recherche, CNRS, Montpellier | Examinateur |
| Benoit Nabholz, Professeur, Université de Montpellier, Montpellier | Co-directeur de thèse |
| Frédéric Delsuc, Directeur de recherche, CNRS, Montpellier | Co-directeur de thèse |

UNIVERSITÉ DE MONTPELLIER

*"Science never pursues the illusory aim of making its answers final, or even probable. Its advance is, rather, towards an infinite yet attainable aim: that of ever discovering new, deeper, and more general problems, and of subjecting our ever tentative answers to ever renewed and ever more rigorous tests."*

*"La science ne poursuit jamais l'objectif illusoire de rendre ses réponses définitives ou même probables. Elle s'achemine plutôt vers le but infini encore qu'accessible de toujours découvrir des problèmes nouveaux, plus profonds et plus généraux, et de soumettre ses réponses, toujours provisoires, à des tests toujours renouvelés et toujours affinés."*

*Karl Popper "Logique de la découverte scientifique"*
*Edition: Payot, pages: 286,287*

# CONTENTS

**General Introduction** **2**

1. Early Taxonomy: seeking order in nature to understand God's project . . . . . 3

2. Species concepts: Since "The Origin of species", or rather, since "the origin of species problem" . . . . . 4

3. Speciation: understanding the process to define taxonomic boundaries . . . . . 7

3.1 Speciation: understanding the process to define taxonomic boundaries . . 7

3.2 The reality of multiple lineages . . . . . 15

4. Species delimitation: the ongoing quest of finding boundaries within a continuous process . . . . . 16

4.1 Emergence of molecular delimitation . . . . . 17

4.2 Automatized delimitation methods . . . . . 18

4.3 The comparative approach . . . . . 20

5. Xenarthrans: a complex story or a story of species complexes . . . . . 21

5.1 Evolutionary history . . . . . 21

5.2 Taxonomy & ecology . . . . . 22

5.3 Phylogeny . . . . . 28

6. Objectives of this thesis . . . . . 30

References . . . . . 32

**Chapter I - Exon capture museomics deciphers the nine-banded armadillo species complex and identifies a new species endemic to the Guiana Shield** **51**

Abstract . . . . . 52

1. Introduction . . . 52
2 Materials & Methods . . . 57
2.1 Biological sampling . . . 57
2.2 DNA extractions and sequencing . . . 57
2.3 Dataset assembly . . . 58
2.4 Phylogenetic inferences . . . 60
2.5 Species partition and population genetic analyses . . . 61
2.6 Species validation . . . 62
2.7 Comparative approach . . . 63
3. Results . . . 63
3.1 Data filtering and quality . . . 63
3.2 Phylogenetic reconstructions . . . 64
3.3 Genetic structure and gene flow between lineages . . . 67
3.4 Species delimitation . . . 68
4. Discussion . . . 71
4.1 Disentangling cross-contamination and genotyping errors . . . 71
4.2 Species delimitation . . . 73
4.3 Integrative taxonomic support for four distinct species in the *Dasypus novemcinctus* complex . . . 73
5. Conclusion . . . 76
Acknowledgments . . . 76
References . . . 77
Data accessibility . . . 92
Supplementary Results . . . 92
Further comparison with literature . . . 92
Mito-nuclear phylogenetic discordance and introgression in nine-banded armadillos . . . 93

**Chapter II - Xenarthran evolutionary history unraveled by whole-genome sequencing at the species level 117**
Abstract . . . 117
1. Introduction . . . 118
2. Results . . . 121
2.1 Mitochondrial phylogeny . . . 121

2.2 Whole genome sequencing and genotyping . . . 125
2.3 Genetic structure and genomic divergence . . . 125
2.4 Genetic diversity, inbreeding, and demography . . . 130
2.5 Phylogenomic analyses . . . 135
2.6 Divergence time estimation . . . 137
3. Discussion . . . 141
3.1 Genomic evidence for new xenarthran species . . . 141
3.2 Conservation implications for xenarthran species . . . 145
3.3 Phylogenomic conflicts within xenarthrans . . . 148
3.4 Biogeography . . . 149
4. Conclusion . . . 150
5. Materials and Methods . . . 151
5.1 Biological sampling . . . 151
5.2 DNA extractions and sequencing . . . 151
5.3 Mitochondrial dataset . . . 151
5.4 Mitochondrial phylogenetic reconstruction . . . 152
5.5 Mitochondrial phylogenetic delimitation . . . 152
5.6 Whole genomes sequencing dataset . . . 153
5.7 Phylogenetic reconstruction . . . 154
5.8 Divergence time estimation . . . 154
5.9 Species delimitations . . . 155
5.10 Genetic diversity and inbreeding . . . 156
5.11 Demographic analysis . . . 156
References . . . 156
Supplementary . . . 168

**Discussion and perspectives** **202**

1. Sampling and sequencing strategies for species delimitation: dealing with non-invasive samples . . . 202
2. Systematic revision of xenarthran species . . . 207
2.1 Taxonomic revision . . . 207
2.2 Phylogenetic reconstruction . . . 209
3. Environment, demography, adaptation: what factors have driven speciation in xenarthran species? . . . 210

4. Is genetic differentiation between species homogeneous across mammals? . . . . 212
5. Personal reflection on taxonomy . . . . . 214
6. General conclusion . . . . . 214
References . . . . . 214

**Appendix 1: Exploring the influence of Andean elevation on xenarthran diversification 222**
1. Introduction . . . . . 222
2. Materials & Methods . . . . . 223
2.1 Estimating xenarthran diversification rates . . . . . 223
2.2 Environment-dependent models . . . . . 224
3 Results & Discussion . . . . . 224
3.1 Diversification rates and paleodiversity of xenarthrans . . . . . 224
3.2 Environmental factors driving diversification . . . . . 229
References . . . . . 234

**Appendix 2: Is island species divergence driven by local adaptation or demography? 238**
1. Introduction . . . . . 238
2. Materials & Methods . . . . . 241
2.1 Dataset assembly . . . . . 241
2.2 Extracting dwarfism and random genes . . . . . 242
2.3 Pn/Ps estimation and statistical analysis . . . . . 242
4. Results & Discussion . . . . . 243
References . . . . . 245

**Appendix 3: Is genetic differentiation between species homogeneous across mammals? 248**
References . . . . . 253

**Appendix 4: Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome 254**

**French summary 278**
Contexte général de la thèse . . . . . 278
Chapitre 1 : La muséomique et la capture d'exons permettent de déchiffrer le complexe d'espèces du tatou à neuf bandes (*Dasypus novemcinctus*) . . . . . 281

Chapitre 2 : Le séquençage du génome complet pour résoudre les incertitudes taxonomiques et phylogénétiques au sein des xénarthres . . . . . 284
Synthèse générale et perspectives . . . . . 289
References . . . . . 289

# Abstract

Xenarthrans (armadillos, anteaters, and sloths) are the least diverse of the four major clades of placental mammals, with only 39 currently recognized species. However, this emblematic Neotropical clade includes species complexes with genetically distinct lineages of uncertain taxonomic status. To address this issue, we conducted a comprehensive genomic analysis, encompassing 261 individual mitogenomes, an exon capture dataset for 71 individuals, and 94 whole genomes, representing 34 distinct xenarthran species. As we included numerous museum samples, we carefully cleaned up potential genotyping errors and cross contaminations that could blur species boundaries by mimicking gene flow. This nearly exhaustive genome-wide dataset allowed us to revise Xenarthra taxonomy employing multiple lines of evidence (species delimitation methods, estimation of gene flow, genomic differentiation, morphology, demography) raising the number of xenarthran species recognized to 43. This revised taxonomic framework enabled the reconstruction of the most comprehensive time-calibrated phylogeny of Xenarthra based on whole genomes, shedding new light on their evolutionary history. By disentangling incomplete lineage sorting and gene flow in discordant parts of the xenarthran phylogeny, we identified multiple events of gene flow suggesting the maintenance of contact zones during speciation events. Further exploring factors promoting speciation within xenarthrans pointed to the potential influence of Colombian and Venezuelan Andes in their diversification. Overall, this work contributes to a deeper understanding of xenarthran evolution but also provides insights into future directions for taxonomic investigations and conservation status assessment within this fascinating mammalian group.

*Keywords:* Species delimitation, molecular systematics, taxonomy, whole-genome, museomics, speciation, conservation genomics, phylogenomics, gene-tree discordance, neotropics, biogeography, mammals.

# Résumé

Les xénarthres (tatous, fourmiliers et paresseux) sont les moins diversifiés des quatre principaux clades de mammifères placentaires, avec seulement 39 espèces actuellement reconnues. Cependant, ce clade néotropical emblématique inclut des complexes d'espèces avec des lignées génétiquement distinctes dont le statut taxonomique reste incertain. Pour les évaluer, nous avons analysé des données génomiques, regroupant 261 mitogénomes, des données issues de la capture d'exons pour 71 individus, et 94 génomes complets, représentant 34 espèces de xénarthres. Comme nous avons inclus de nombreux échantillons de musées, nous avons soigneusement nettoyé les erreurs potentielles de génotypage et les contaminations croisées qui pouvaient brouiller les délimitations d'espèces en imitant le flux de gènes. Ainsi, grâce à un jeu de données quasi exhaustif de données génomiques pour les xénarthres, nous avons révisé leur taxonomie en utilisant de multiples éléments de soutient (méthodes de délimitation d'espèces, estimation du flux de gènes, différenciation génomique, morphologie, démographie), ce qui a permis d'augmenter le nombre d'espèces de xénarthres reconnues à 43. Ce cadre taxonomique révisé a permis de reconstruire la phylogénie calibrée dans le temps la plus complète des xénarthres basée sur des génomes complets, révélant plus précisément leur histoire évolutive. En démêlant le tri de lignées incomplet et le flux de gènes des arbres de gènes discordants au sein des xénarthres, nous avons identifié de multiples événements de flux de gènes, ce qui suggère le maintien de zones de contact au cours de certains événements de spéciation au sein de ce clade. Enfin, en étudiant les facteurs favorisant la spéciation des xénarthres, nous suggérons l'influence des Andes colombiennes et vénézuéliennes sur leur diversification. En conclusion, cette thèse a contribué à une meilleure compréhension de l'évolution des xénarthres, mais elle donne également des pistes potentielles pour orienter les futures études taxonomiques et l'évaluation du statut de conservation des espèces de ce groupe fascinant de mammifères.

*Mots-clés:* Délimitation d'espèces, systématique moléculaire, taxonomie, génomes complets, muséomique, spéciation, génomique de la conservation, phylogénomique, discordance des arbres de gènes, néotropiques, biogéographie, mammifères.

# General Introduction

Disclaimer: This PhD thesis does not constitute an official nomenclatural act.

Earth is home to a teeming diversity of organisms. Some studies are estimating 8.7 millions eukaryotic species on Earth and 86% await to be described (Mora et al., 2011). To better understand living organisms it is necessary to describe, name, and organize them. This task is the one of taxonomy that proposes a framework to simplify the reality by organizing organisms into comparable units (i.e. taxonomic ranks) whose basic unit is the species rank. It is based on two aspects: i) describing and naming taxa (i.e. alpha-taxonomy) and ii) organizing them into higher ranks (i.e. beta-taxonomy). Alpha-taxonomy is a crucial step for many fields of study. In particular, once a species has been defined, it is possible to extrapolate conclusions obtained from sampled individuals to the whole species. It also enables us to agree on a universal terminology that facilitates communication within the scientific community and more broadly in society. Finally, describing biodiversity is useful in many scientific areas, for example it has allowed macroevolutionary studies to highlight the uneven distribution of biodiversity across the planet and taxonomic groups (Hawkins, 2001). The necessity of the species as a reference unit is unequivocal for scientists but the taxonomic framework is inherited from a vision of the living world far removed from the one we have today.

# Early Taxonomy: seeking order in nature to understand God's project

As early as 3000 B.C. Shen Nung, Emperor of China, described 365 medicinal preparations from plants, animals and minerals, and organized them according to their medicinal properties (Lewis, 2010; Manktelow, 2010). In Egypt too, 1500 B.C. old papyrus and wall paintings described, illustrated, and named medicinal plants (Lewis, 2010; Manktelow, 2010). A major advance in taxonomy was allowed by Aristotle, a Greek scientist and philosopher from 384 to 322 B.C., who proposed to classify organisms according to their traits, physiology, behavior rather than their properties. For example, the presence of lungs in cetaceans influenced their grouping with other "viviparous quadrupeds" rather than with fish, which is still used today. Aristotle also grouped organisms with similar characteristics into genera, that he further divided into species. Until the 16th century, no major innovations from Aristotle's viewpoint were proposed, the main goal being to find the best criteria to classify organisms. Consequently, improvements to the field appeared with new optic technologies that allowed descriptions at a finer scale. Thus, Caesalpino (1519–1603) proposed classifying plants according to the structure of their seeds and fruits. Bauhin (1560-1624) refined Caesalpino's classifications by grouping the 6000 botanical species he registered in Pinax Theatri Botanici (Bauhin, 1623) according to their genus and their "natural affinities". He named some of them using a typography based on their genus followed by a single-word name, doing so, he introduced the binomial nomenclature that broadly influenced contemporary taxonomists and later Linnaeus (Cain, 1994). However, he did not conceive the species as a taxonomic rank as we do today (Cain, 1994). Few decades later, John Ray (1627-1705) proposed a definition for species where conspecificity comes from common parentage of individuals (Stevenson, 1947). Followed by Joseph Pitton de Tournefort (1656-1708), who defined genera as an essential aspect of taxonomy reflecting natural similarity between species. Although these terms have been used since Aristotle, these two early taxonomists provided clear concepts and thus limited discrepancies when used by different taxonomists. Both contributed greatly to botanical taxonomy and improved the classification system.

These 16th-century taxonomists laid foundations for naming and organizing living things that influenced Linnaeus and allowed him to propose a formal system of nomenclature (Manktelow, 2010; Wilson, 2010). In fact, the increasing number of species descriptions with-

out uniform taxonomic rules induced numerous synonyms (multiple names for the same species) and no proper description. To avoid confusion and inconsistency, Carl Von Linneaus, in the mid-18th century, proposed a formal framework for the nomenclature system, which he detailed in his book Systema Naturae (Linnaeus, 1758). Inspired by Bauhin, he first simplified the process of naming species by using binomial nomenclature based on their genus and a unique epithet to represent the species. The use of Latin for the binomial name facilitated the spread of his system, as this language was widely understood. Associated with this name, a diagnosis should describe characteristics of the species, and the use of a type specimen (a reference individual representing the species) was generalized. Finally, he provided a clearer vision of natural classification defining species as the basic taxonomic unit and then further aggregating taxa into higher taxonomic ranks such as genus, order, class and kingdom to reflect their biological similarity (Wilson, 2010). This nomenclature system revolutionized the field of taxonomy by providing formal rules for naming, describing, and classifying species that allowed the diversity of organisms to be recorded.

It is noteworthy that the main goal of taxonomy, which prevailed for many taxonomists at that time, was to describe and understand the organization of all forms created by God in order to understand his project. This was notably highlighted by Linneaus's own motto "Deus creavit, Linnaeus disposuit" meaning "Gods created, Linnaeus organized". From the creationist point of view, the origin of all species can be attributed to God (Barberousse and Samadi, 2010; Wood, 2008). Another dominant school of thought was the fixist theory, which suggests that organisms do not change over time (Barberousse and Samadi, 2010; Wood, 2008). Thus, at that time, species were seen as unchangeable since their creation:

> *"There are as many species as the infinite being created diverse forms in the beginning, which, following the laws of generation, produced as many others but always similar to them: Therefore there are as many species as we have different structures before us today".*
> Linnaeus, 1758

## Species concepts: Since "The Origin of species", or rather, since "the origin of species problem"

During the 19th century, suggestions that species might have changed since their creation emerged and widely impacted the species concept. One of the major transformism theories, which rely on inheritance of acquired characters, was proposed by Jean-Baptiste Lamarck

in 1809. His theory assumes spontaneous generation of organisms. Then, during their life, the environment induces modifications that are transmitted to their descendents. Fifty years later, Charles Darwin proposed his theory of natural selection in the famous On the Origin of Species (Darwin, 1859). Concomitantly with Alfred Russel Walace, they came to the conclusion that heritable variation could provide organisms with advantages to reproduce or survive in the environment, leading to the progressive adaptation of species and driving diversity. They revolutionized the field of biology by challenging the static/"fixist" view of species. According to his new theory, Darwin thought of species as subjective/arbitrary categories without biological reality:

> *"In short, we shall have to treat species in the same manner as those naturalists treat genera, who admit that genera are merely artificial combinations made for convenience. This may not be a cheering prospect, but we shall at least be freed from the vain search for the undiscovered and undiscoverable essence of the term species."*
>
> Darwin 1859, p.485

Thus, despite his title On the Origin of Species, Darwin was not particularly interested in defining species:

> *"Nor shall I here discuss the various definitions which have been given of the term species. No one definition has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of species. Generally the term includes the unknown element of a distinct act of creation"*
>
> Darwin 1859, p. 44

Even if Darwin's theory considerably impacted the vision of species by providing a totally new theoretical framework, it did not provide a clear species concept, thus, he did not have much impact on the field of taxonomy. His work only clarified the notion of "affinity" between taxa, giving it a biological meaning (De Queiroz, 1988).

> *"Systematists will be able to pursue their labors as at present but they will not be incessantly haunted by the shadowy doubt whether this or that form be in essence a species"*
>
> Darwin 1859, p. 484

In the early 20th century, theoretical geneticists and the rediscovery of Mendel's work on heredity confirmed Darwin's theory of natural selection (Ayala and Fitch, 1997). Even though they recognized the continuous evolution of species, Mayr and Dobzhansky disagreed with Darwin's view of species, believing that he misunderstood the nature of species and denied their distinctness to make them appear more continuous in order to support his theory of evolution (Mallet, 2010; Mayr, 1963).

> *"Some recent authors have dealt with the concept of species as if it were merely an arbitrary, man-made concept,[...]. The term 'species' refers to a concrete phenomenon of nature and this fact severely constrains the number and kinds of possible definitions. The word 'species' is, like the words 'planet' or 'moon', a technical term for a concrete phenomenon."*
>
> Mayr, 1963, p. 263

Mayr and Dobzhansky thought that reproductive isolation mechanisms would lead to discontinuities, creating natural units that would not be arbitrarily defined by humans (Mallet, 2010; Mayr, 1949). They proposed species concepts based on interbreeding and reproductive isolation that they define as the "Biological Species Concept" (T. Dobzhansky, 1935, 1937; Mayr, 1940, 1942, 1963).

> *"Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups"*
>
> Mayr, 1942, p. 120; modified according to the definition given in T. Dobzhansky, 1935

Although this definition was widely accepted by the scientific community at the time, numerous alternatives emerged. Simpson put evolution at the center of his concept introducing the Evolutionary Species Concept. He replaced the criterion of 'reproductively isolated' by the notion of "evolving separately from others" and thus added a time dimension (Simpson, 1961, p. 153). In 1976, Van Valen proposed the Ecological Species concept referring to the distinct niches occupied by species (Van Valen, 1976). Another one, the Monophyletic Species Concept, is based on shared derived characters (Rosen, 1979, p. 277, Mishler, 2000, p. 213). At the end of the 20th century, no fewer than 22 species definitions were recognized (Mayden, 1997). This inability to agree on a species definition is commonly known as the "species problem".

In 2007, De Queiroz identified that all species concepts rely on "separately evolving metapopulation lineages" (Queiroz, 2007). In addition to this primary property, species concepts are also based on secondary biological properties used as operational criteria for species delimitation (e.g. interbreeding, monophyly, etc.). These secondary biological properties are acquired as lineages diverge. However, they can arise at different times of the speciation process (Figure 1), leading to conflicting delimitations when different species concepts are used, which corresponds to a gray zone.

Even though De Queiroz's unified species concept allowed a major theoretical improvement of the species problem, the General Lineage Species Concept has only partially solved the problem because this fundamental property of species (i.e. separately evolving lineages)

is not operational as it cannot be used to delimit species (Queiroz, 2007; Zachos, 2018). Today, no consensus has been reached, as researchers still disagree on which species concept to use (Stankowski and Ravinet, 2021b).

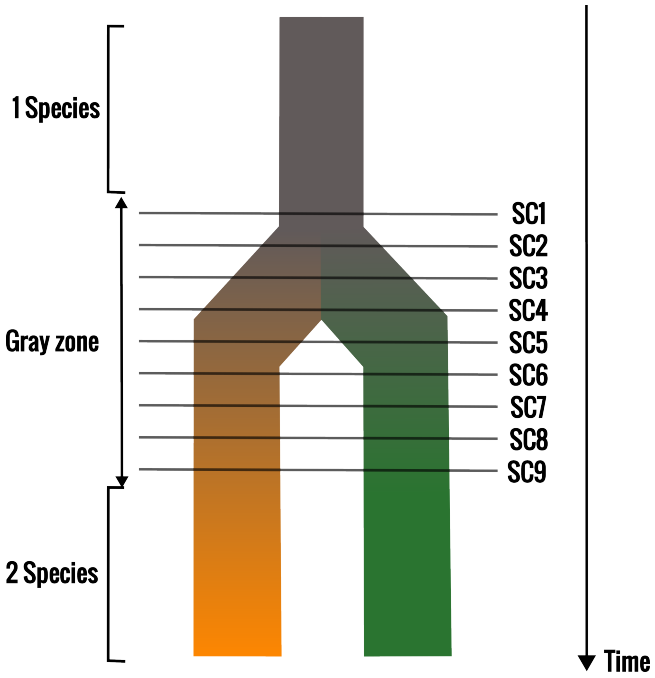


**Figure 1:** This dendrogram modified from Queiroz, 2007 represents an ancestral species (gray) splitting into two daughter lineages (orange and green). The color intensity represents their divergence. Different species concepts rely on different species criterion (SC). For example, the Ecological Species concept uses the property of distinct niches as a criterion to delimit species while the Monophyletic Species Concept uses the property of shared derived characters as a criterion. Horizontal lines symbolize when properties corresponding to species criteria (1 to 9) are acquired during the lineage divergence. The gray zone represents the period during which distinct species concepts may disagree because some properties are already acquired but not all. Outside this gray zone, all species concepts unanimously support the same delimitation. Finally, properties used as species criteria can appear in different orders in other speciation events.

# Speciation: understanding the process to define taxonomic boundaries

## The "simple" process of splitting a lineage into two

Although these decades of debate have not led to a consensual definition of species, our understanding of the formation of new species by the speciation process has improved considerably. We now recognize that this process involves a progressive interruption of gene flow between the two lineages until their complete reproductive isolation.

Two mechanisms contribute to the isolation of lineages by preventing hybridization: i) geographic isolation, and ii) reproductive isolation (Singh, 2022; Westram et al., 2022). The first mechanism is independent of genetics and corresponds to the environmental effect (geographic barrier or distances) to limit the probability of mating between lineages (Figure 2; T. Dobzhansky, 1937; Mayr, 1942). Geographic isolation is not considered to be a mechanism of reproductive isolation because two lineages with geographic isolation can still totally interbreed if they are brought into contact. The second mechanism involves genetic factors, barrier loci, that reduce interbreeding between lineages (Barton and Bengtsson, 1986; Lowry et al., 2008; Rieseberg et al., 1999). These barrier loci reduce the production or the fitness of hybrids by acting at different steps of reproduction: i) pre- or post-mating, and ii) pre- or post-zygotic (Figure 2, T. Dobzhansky, 1937; O. Seehausen et al., 2014). Generally, barrier loci contribute only partially to reproductive isolation, and various isolating barriers may accumulate during the speciation process until complete reproductive isolation of lineages (Coyne and Orr, 2004; Nosil et al., 2009b; Nosil and Schluter, 2011; Wu, 2001). This gradient in strength of reproductive isolation during the speciation process drives the concept of the speciation continuum (Mallet, 2008; O. Seehausen et al., 2014; Stankowski and Ravinet, 2021a).

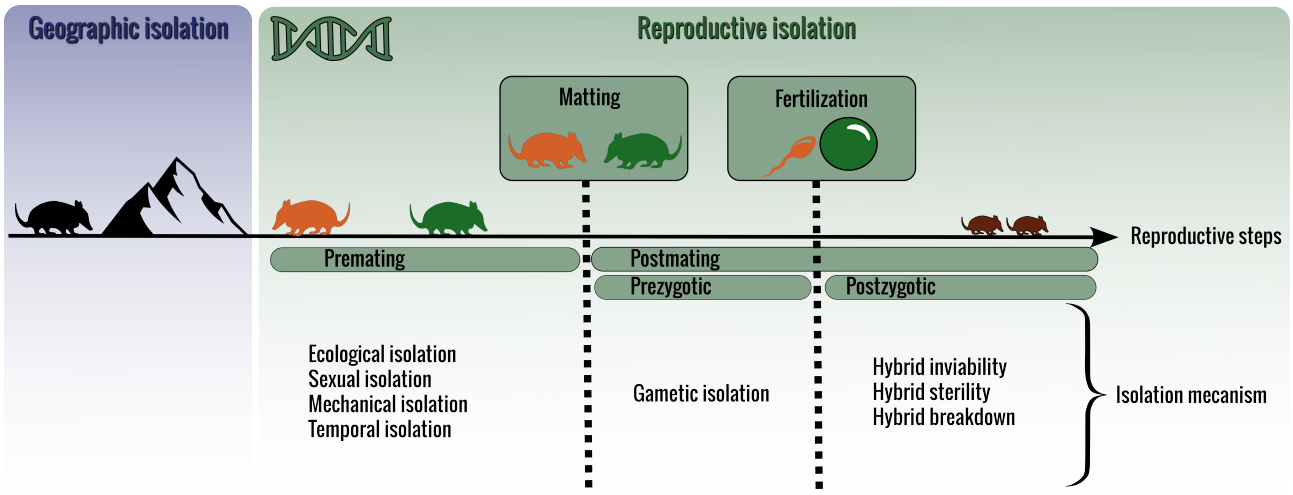


**Figure 2:** Isolation mechanisms caused by the environment (in blue; i.e. Geographic isolation) or by genetic factors (in green; i.e. Reproductive isolation). Genetic factors (i.e. barrier loci) induce lineage isolation at different stages of reproduction and through different isolation mechanisms.

Two scenarios have been proposed for the emergence of the first barrier(s) initiating the speciation process (O. Seehausen et al., 2014). The first requires geographic isolation to interrupt gene flow between lineages. During this isolated phase, the two lineages will accumu-

late mutations. In this context, T. H. Dobzhansky (1936) proposed a model starting from the ancestral genotype AABB, then the two lineages will evolve into AAbb and aaBB, respectively. By assuming that a-b are incompatible, any hybrid (AaBb) between the two lineages will have lower fitness. This model of two loci with negative epistatic interaction, known as Dobzhansky-Muller incompatibility, is a simple explanation, compatible with neutral evolution, to evolve incompatibilities without crossing a valley of low fitness (Figure 3; Bateson, 1909; T. H. Dobzhansky, 1936; Muller, 1942). The second scenario does not necessarily require geographic isolation, but involves disruptive selection that induces adaptation to local environments (Nosil, 2012). If hybrids formed between the two locally adapted lineages have a lower fitness, this locus of local adaptation can be considered a reproductive barrier. Remarkably, this ecological speciation scenario is compatible with gene flow and is also transposable to cases driven by sexual selection (Coyne and Orr, 2004; Doorn et al., 2009; Gavrilets, 2004).

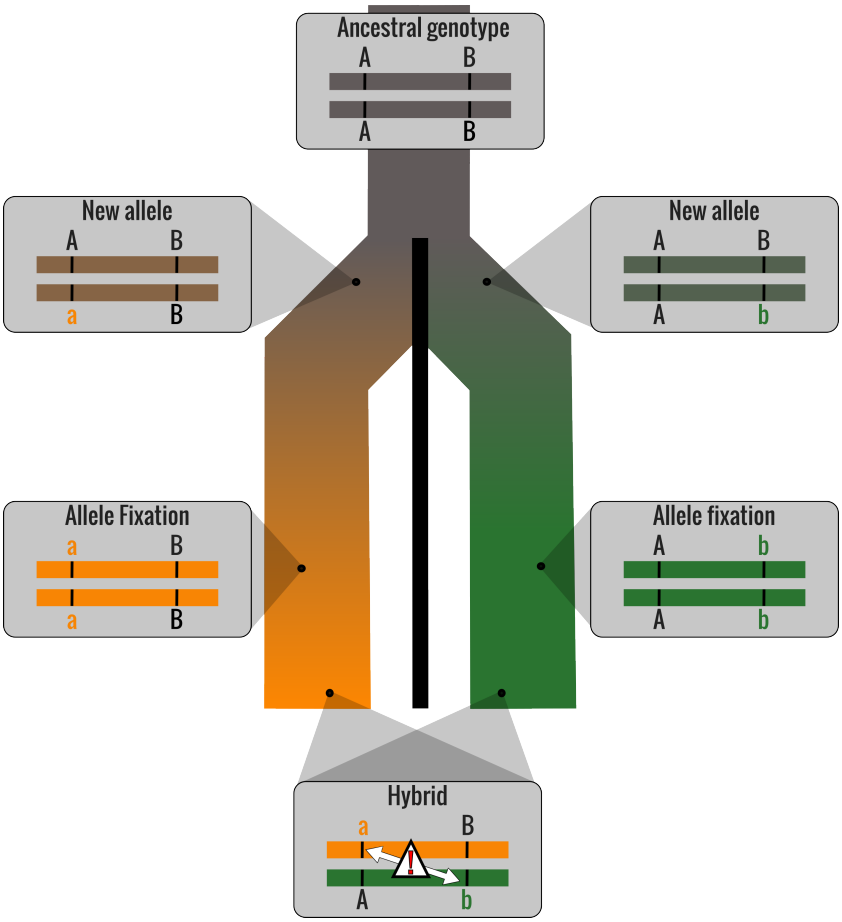


**Figure 3:** Emergence of Dobzhansky-Muller incompatibility between two lineages (orange and green, respectively) without gene flow (represented by the black line in the middle). The two lineages inherited the same ancestral genotype AABB. Independently, new alleles arise in the two lineages and become fixed (randomly or by natural selection) leading to aaBB and AAbb genotypes in the orange and green lineages, respectively. If a-b are incompatible, a hybrid (AaBb) between the two lineages will have a lower fitness.

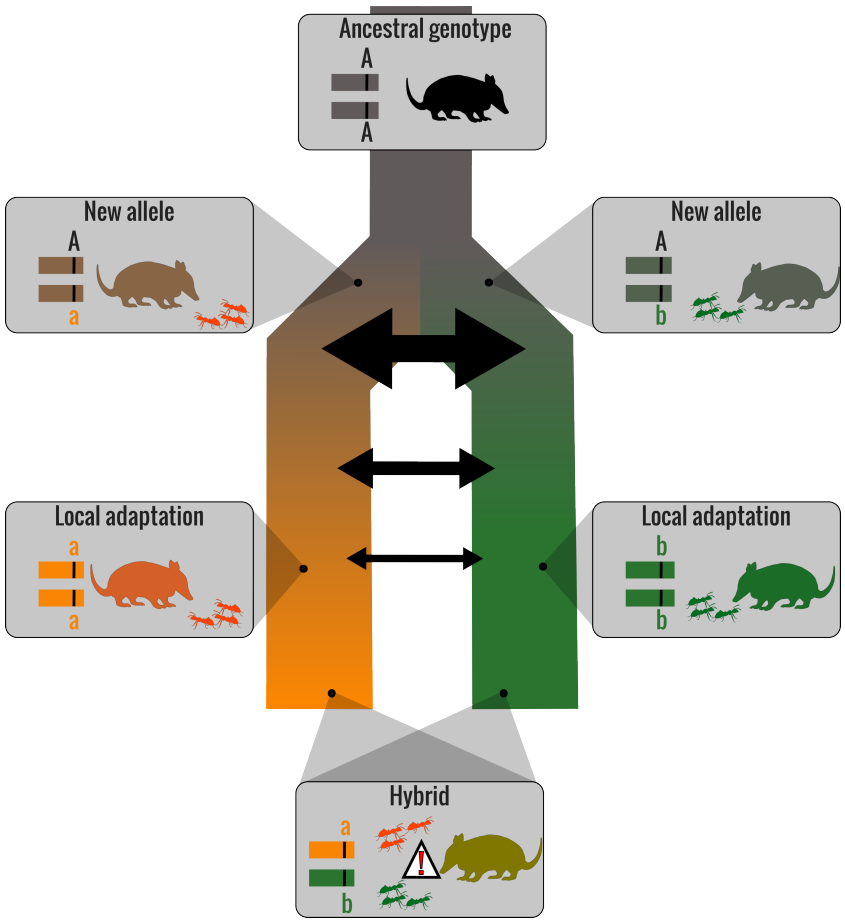


**Figure 4:** Distinct alleles (a and b) that induce adaptation to the local environments (represented by the orange and green ants) are fixed in the two lineages by natural selection. This scenario is compatible with gene flow (represented by arrows), but to be considered as barrier locus, hybrids formed must have a lower fitness.

After this initiation phase, in both scenarios, hybrids formed have a lower fitness. This is the substrate for the emergence of additional barriers, as any prezygotic barriers preventing the formation of these reduced-fitness hybrids would be selected. This is what is called reinforcement (Butlin and Smadja, 2018; Servedio and Noor, 2003). Intrinsic structures of the genome can also promote the accumulation of barrier loci (chromosomal inversions, gene duplications, changes in ploïdy, PRDM9 driven recombination; Campbell et al., 2018). Progressively, the genetic differentiation of barrier regions can spread to larger regions by divergence hitchhiking, because the barrier locus, by preventing gene flow, reduces recombination between lineages and increases divergence via hitchhiking of linked mutations. When new barriers arise in this region, it increases the size of the genetic region where hitchhiking occurs (Feder and Nosil, 2010; Smadja et al., 2008; Via, 2012; Via and West, 2008).

Some models have attempted to describe the dynamics of barrier accumulation during the speciation process (e.g. Orr's model supports a snowball accumulation of genetic incompatibilities (Orr, 1995)). However, it is still difficult to validate these theoretical models

(Gourbiere and Mallet, 2010; Kulmuni et al., 2020; Matute et al., 2010; Presgraves, 2010; Stankowski and Ravinet, 2021a). Notably because this process has no unique ending: i) in some cases lineages never reach total reproductive isolation (Nosil et al., 2009b), and ii) gene flow events (i.e. secondary contact) can reverse the process at any time, speciation is not unidirectional (Campagna et al., 2014; Kearns et al., 2018; O. L. E. Seehausen et al., 2008; Stankowski and Ravinet, 2021a; see Box 1 on page 11).

**Box 1: Molecular evolution of the speciation process**

During the speciation process, independent evolution of lineages and barrier loci influences the molecular evolution of genomes. At the beginning of the speciation process, the two daughter lineages inherited genetic diversity from the ancestral population. Each lineage sorts these variants: i) randomly for neutral mutations (Kimura, 1983), or ii) according to natural selection for advantageous or deleterious mutations (Figure 5; Wright, 1931). Thus, neutral mutations are sorted faster for lineages with low effective population size (Ne) or in genetic regions with low Ne (Crow and Kimura, 1970). Conversely selection is efficient in lineages with high Ne or in genetic regions with high Ne, sorting non-neutral mutations faster (Nei and Li, 1973). Simultaneously, each lineage also produces its own new variants, proportionally to its population size ($4N_e\mu$ (Kimura and Crow, 1964); Figure 5). This progressively induces the structuration of the polymorphism, and, as variants become fixed, leads to an increase in divergence. Noteworthy, this genetic differentiation can be amplified by barrier loci that increase fitness within their lineage (e.g. local adaptation, reinforcement barriers). Indeed, the fixation of these adapted loci accelerates lineage sorting of the linked mutations (i.e. those that recombination cannot separate) by provoking a selective sweep (Hermisson and Pennings, 2005; Nielsen, 2005).

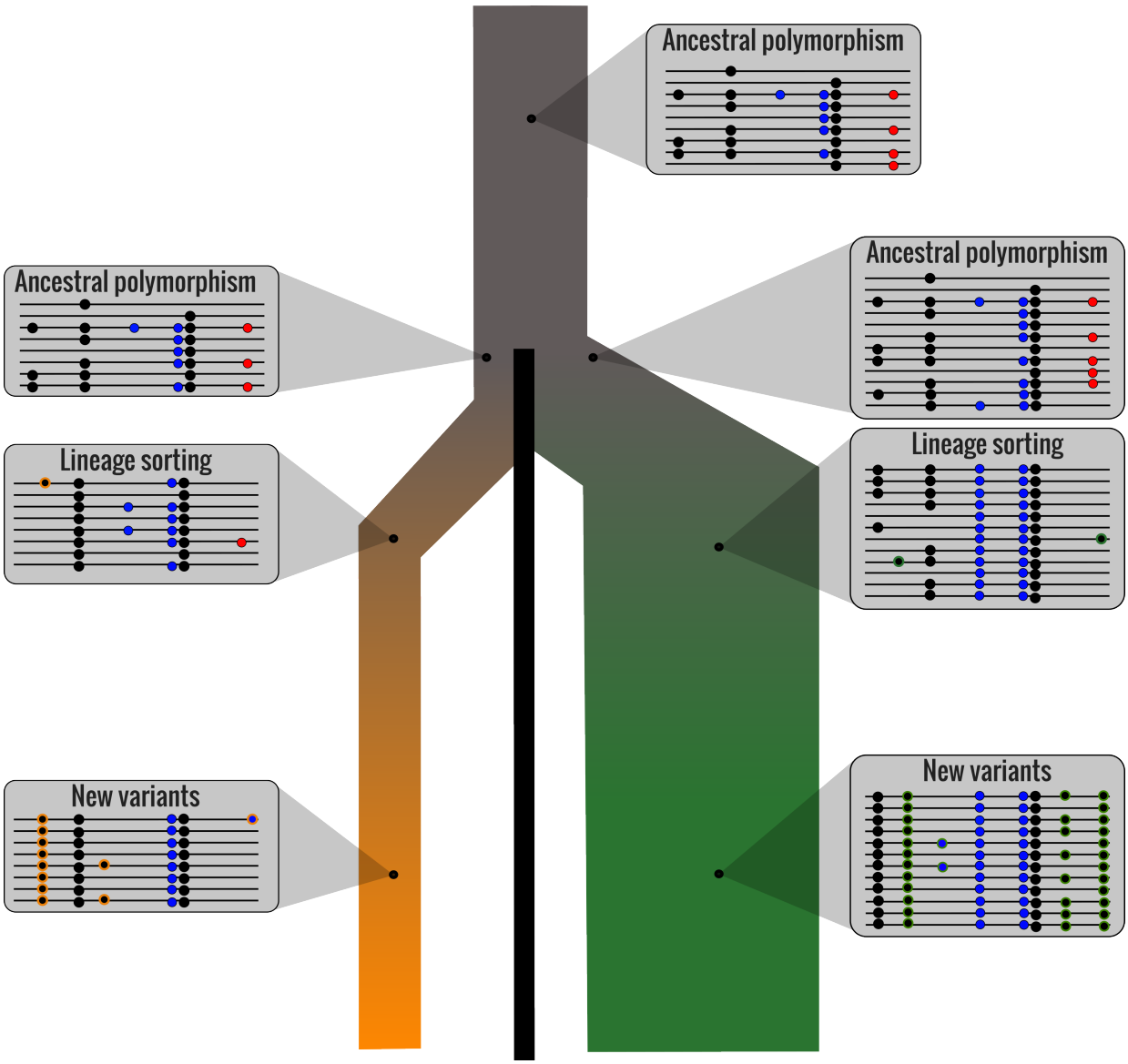


**Figure 5:** Polymorphism evolution during speciation without gene flow (represented by the vertical black line) of two lineages (orange and green) with different effective population sizes. The two daughter lineages inherit the same ancestral polymorphism composed of neutral mutations (black), advantageous mutations (blue), and deleterious mutations (red). The greater selection efficiency in large population size induces faster fixation or purge of non-neutral mutations. The greater drift in small population size induces a faster loss or fixation of neutral mutations. Simultaneously to this lineage sorting of ancestral polymorphism, new variants (circles outline of the lineage colors) arise.

In parallel, hybrids can be produced, and repeated backcrossing can result in introgression of genetic material from one lineage to another (Arnold and Arnold, 2006; Coyne and Orr, 2004). This gene flow can transfer neutral and/or adaptive mutations or genomic regions that will reduce genetic differentiation by homogenizing variants between lineages (Abbott et al., 2013; Feder et al., 2012). For nearly neutral mutations,

population size matters (see Box 2 on page 14; Ohta, 1992). Finally, barrier loci resist gene flow by being selected against due to their deleterious effect. Consequently, they maintain genetic differentiation at these loci and linked regions (Barton, 1979; Cruickshank and Hahn, 2014; Payseur, 2010).

This semi-permeability of the genome induces heterogeneity in the differentiation of regions, contributing to a typical mosaic pattern in which highly differentiated regions are designated as genomic islands (Feder et al., 2012; Harr, 2006; Malinsky et al., 2015; Nosil et al., 2009a; Ravinet et al., 2017; Turner et al., 2005; Wu, 2001). However, these genomic signatures of barrier loci are hard to distinguish from adaptive loci not involved in reproduction, evolutionary history, genetic architecture (gene density, mutation rate, recombination maps), or other linked loci (Cruickshank and Hahn, 2014; Noor and Bennett, 2009; Turner and Hahn, 2010).

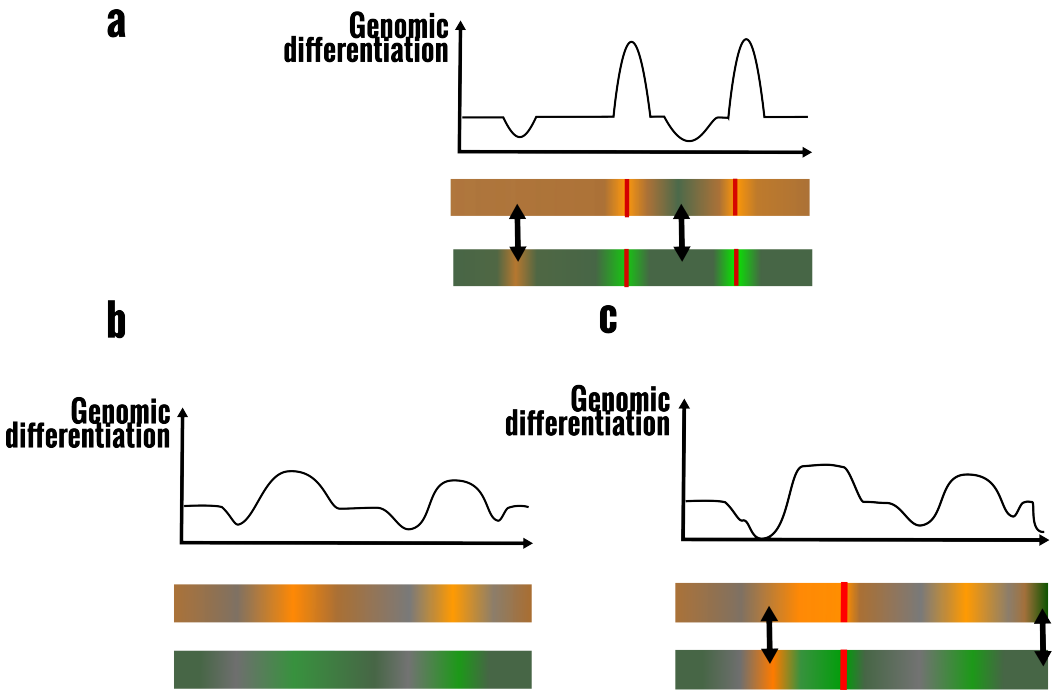


**Figure 6:** Genomic island expectation (a) and the complex reality (b and c). The genomic differentiation of two diverging lineages can be estimated along their genomes (represented under the graph, the color intensity reflects the mutation accumulation, and the color (i.e. orange or green) the lineage they come from). Loci introgressed from one lineage to another (represented by arrows) locally reduce this genetic differentiation contrary to barrier loci (red markers) which increase genetic differentiation. a) It is expected that high values of genetic differentiation (i.e. genomic islands) correspond to barrier loci. b) However, variation of genomic differentiation is inherent in the genome. c) Thus, identifying gene flow and barrier loci using genomic differentiation is not evident.

**Box 2: Weakly deleterious introgression fate**

The fitness effect of mutations depends on both effective population size (Ne) and selection coefficient (s) (Ohta, 1992). Indeed, for a same selection coefficient (-0.1), a mutation appears more deleterious in a large population size (Figure 7). This will induce asymmetrical introgression if lineages have different population sizes. Large population size will easily purge weakly deleterious mutations accumulated in the low population size haplotype, and low population size will benefit an already purged haplotype from large population size.

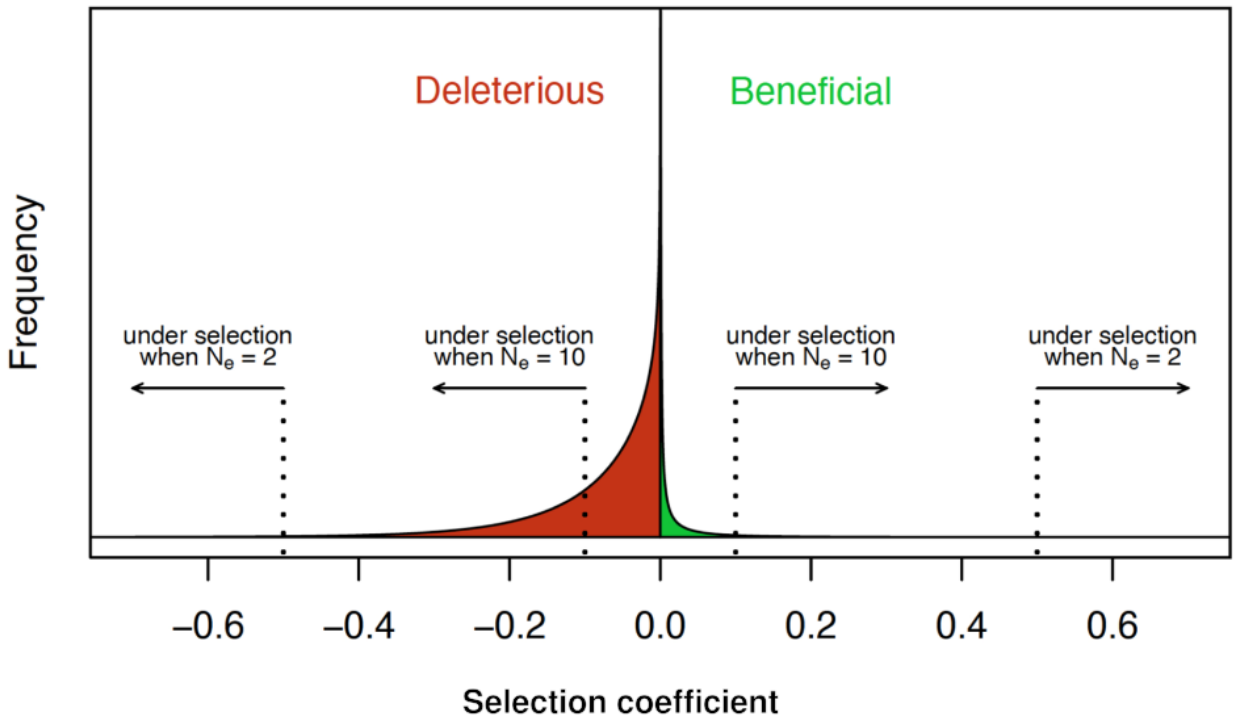


**Figure 7:** Conceptual representation of the fitness effects of mutations modified from Bao et al. (2022). Distribution of the selection coefficient of beneficial mutations (in green) and deleterious mutations (in red). The larger the population size, the greater the effect of the mutation on fitness ($Ne \times s$), and therefore the more effective selection is at fixing positive mutations or eliminating deleterious ones.

Even if we have described simple cases or mechanisms that lead to speciation, each speciation event results from a unique genetic pathway (among a multitude of other possible ones) resulting from complex historical scenarios that may have changed throughout the speciation process. Thus, while we have long searched for THE way to create species, the multitude of pathways leading to reproductive isolation now promote a multidimensional concept of the speciation continuum (hypercube; Dieckmann, 2004; Stankowski et al., 2023).

## The reality of multiple lineages

The reality is even more complicated, considering that populations can split into multiple lineages in a short period of time. If an additional split occurred before the complete sorting of the ancestral polymorphism, the multiple resulting lineages will share the polymorphism * (Box 1 on page 11; Hobolth et al., 2011; Hudson, 1983; J. F. Kingman, 1982; Neigel, 1985; Pamilo and Nei, 1988; Tajima, 1983; Takahata and Nei, 1985). Neutral variants will be sorted randomly and independently in each lineage. Thus, by chance, we expect an equivalent support for the three topologies (Figure 8). This phenomenon is called Incomplete Lineage Sorting (ILS) and contributes to discordant topologies between gene trees and the species tree (Figure 8; Maddison, 1997).

In addition, in such rapid speciation events, the multiple lineages can still interbreed. Thus, gene flow is not restricted to sister lineages, and if genetic material is transferred from a distant lineage, the phylogenetic reconstruction of this genetic region will not reflect speciation events. This mechanism also contributes to discordant topologies between gene trees (Figure 8; Maddison, 1997).

However, current known diversity does not necessarily reflect the biological diversity with which lineages may have exchanged during their evolution. It's essential to keep in mind potential gene flow with ghost lineages that also generate discordant topologies (Tricou et al., 2022a,b).

Genomes are therefore a patchwork of inherent variation of divergence, accentuated by i) evolution of regions involved in reproductive isolation, and ii) mixed up with various other taxa (ghost or extant). Considering this, reconstructing complex speciation scenarios using phylogenetic reconstruction based on strict bifurcation can sometimes seem inappropriate (Nosenko et al., 2013).

*Large population size can maintain more neutral polymorphism (Kimura and Crow, 1964); Box 1 on page 11

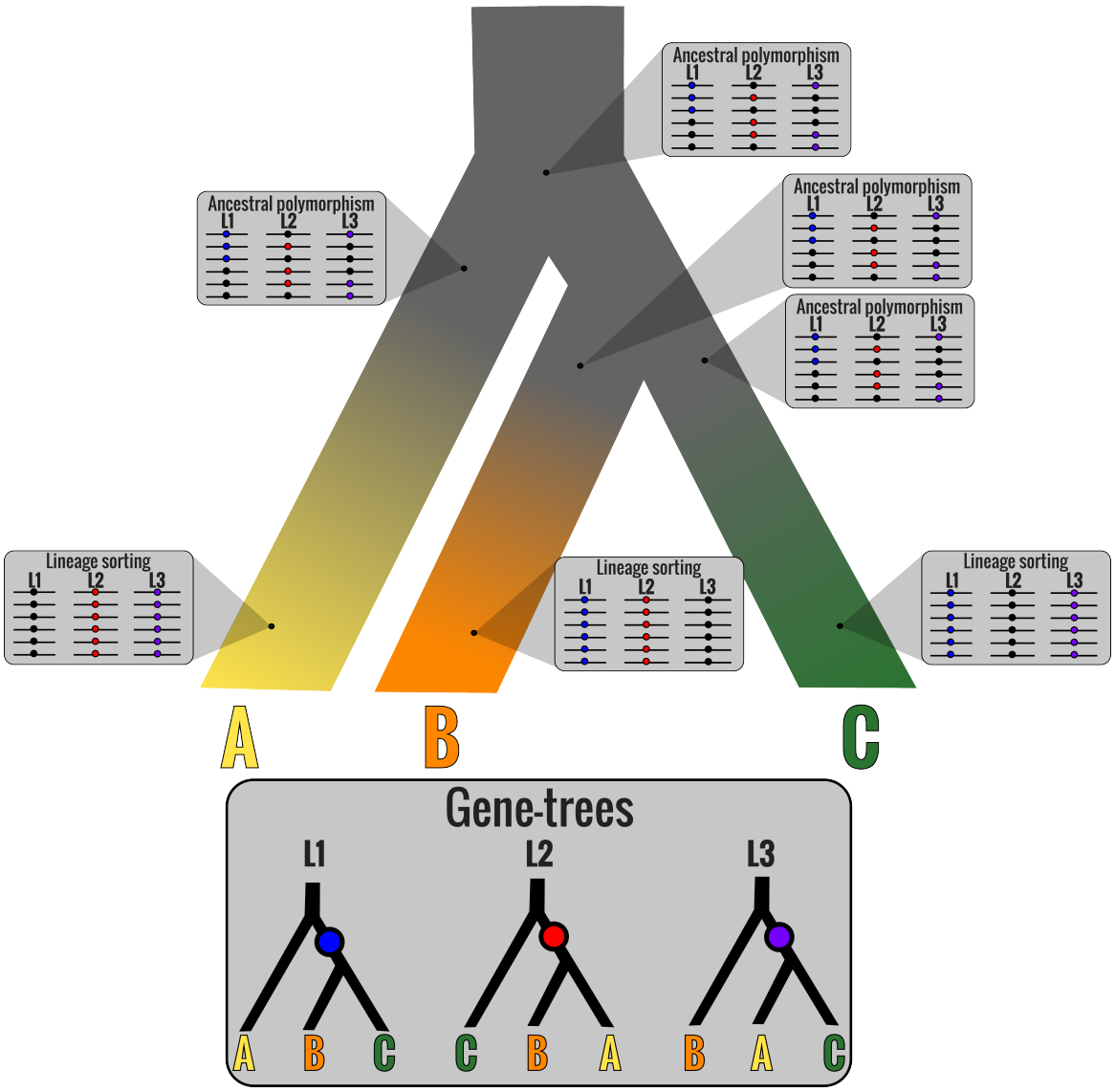


**Figure 8:** Gene tree discordances due to incomplete lineage sorting (ILS). Consider three independent loci (L1, L2, L3) with neutral variations (circles) in the ancestral population. If these polymorphic sites are not sorted before the next speciation events (e.g. large population size or fast speciation events), then the three daughter lineages (A, B and C) inherit this ancestral polymorphism. Within these three lineages, drift fixes one neutral allele at each locus. This random sorting supports by chance three distinct topologies. Thus neutral processes can induce discordant topologies between gene trees.

## Species delimitation: the ongoing quest of finding boundaries within a continuous process

The first step of taxonomy intends to find groups of individuals that belong to the same species: species delimitation. For some lineages that have clearly achieved the speciation process, assigning the species rank is unequivocal. The challenge arises for those with a more ambiguous "position" along the speciation continuum, falling into the so-called gray

zone. Indeed, taxonomy seeks discrete boundaries within the continuous speciation process and the only feasible approach to obtaining such discrete boundaries is to define an arbitrary threshold. However, despite several species concepts associated with operational criteria as thresholds, there is still no agreed threshold or consensual definition for species (Queiroz, 2007; Stankowski and Ravinet, 2021a). The problem has not been resolved with our understanding of the speciation process, which turned out to consist of a multitude of possible paths, making it difficult to identify a relevant threshold (Stankowski et al., 2023).

## Emergence of molecular delimitation

Historically, taxonomy had used comparative morphology to separate species, but the emergence of molecular data provided complementary support for delimitation with almost an unlimited number of characters. The latter also have the advantage of being strictly heritable unlike morphological characters that can be influenced by the environment (Subbotin and Moens, 2006). Since their first use in taxonomy during the 1960s (Fitch and Margoliash, 1967), new molecular-based methods arose. DNA-DNA hybridization, developed in the 1980s, allows estimation of genome similarity and has been used to delimit species that are genetically too distant (with less than 70% of their genome hybridizing; Wayne et al., 1987). This method considerably improved bacterial taxonomy, which has been laborious to resolve using morphological markers alone (Wayne et al., 1987). Later, the progress in DNA sequencing provided an unprecedented progress by making the nucleotide bases directly accessible (Hillis et al., 1996). Inference tree methods improved and initiated a delimitation method based on the reciprocal monophyly via the Genealogical Species Concept (Avise and Ball, 1990). The problem with this concept is the fractal characteristic of trees (i.e. monophyletic lineages can reflect species as well as populations; (Hey, 2001; Zachos, 2016)). Finally, the most promising delimitation approach of the 2000s was certainly the DNA-barcoding (Hebert et al., 2003). This method proposes to use a unique molecular marker with an appropriate evolution rate to distinguish every species, advocating for a gap between the distribution of interspecific variations and intraspecific variations that can serve as a species delimitation threshold. While powerful in numerous cases, this approach has been widely criticized notably because it rely on a single gene with limited phylogenetic resolution, of which divergence information is neither necessary nor sufficient to assess species boundaries (Moritz and Cicero, 2004; V. S. Smith, 2005; Will et al., 2005).

## Automatized delimitation methods

The development of bioinformatics allowed automated delimitation methods, providing more objective and reproducible methods and also allowed quantifying the support for species delimitation using statistics. Numerous molecular-based methods have been proposed but I will focus on the most popular ones (see Box 3). Carstens et al. (2013) and Smith and Carstens (2022) proposed to distinguish two types of delimitation methods: i) discovery methods, and ii) validation methods.

Discovery methods propose a species delimitation without requiring any a priori hypotheses about species delimitation. It includes distance-based methods such as the Automatic Barcode Gap Discovery (ABGD; Puillandre et al., 2012 or tree-based methods such as the Generalized Mixed Yule Coalescent (GMYC; Pons et al., 2006) or the Poisson Tree Process (PTP; Zhang et al., 2013). Initially developed for a single locus, they are increasingly applied to multilocus datasets relying on the assumption that the shared genealogical history would not bias branch length of the species tree. Luo et al., 2018 compared the effect of the number of loci on the performance of these methods to a multilocus delimitation method (BPP; Yang and Rannala, 2010) and found a restricted impact of the number of loci when gene flow is absent. However other methods such as ADMIXTURE (Alexander and Lange, 2011) or STRUCTURE (Pritchard et al., 2000) have specifically been designed for multilocus datasets and integrate gene flow, making these methods highly relevant and appropriate to various cases. All these discovery methods propose species delimitation hypotheses that can be further evaluated through validation methods.

Indeed, validation methods evaluate the support for an a priori species hypothesis. These include tree-based programs such as BPP or demographic-based models such as PHRAPL (Jackson et al., 2017). These methods rely on the Multi-Species Coalescent (MSC) models to evaluate evolutionary scenarios and estimate the number of supported species. Although these methods can take into account independent evolutionary histories along the genome by evaluating multilocus datasets, the demographic scenarios they explore can appear oversimplified (e.g. BPP does not take into account gene flow and PHRAPL allows only a limited number of individuals).

Although these methods have led to considerable advances in species delimitation, their sensitivity to over-splitting remains a weakness. In fact, the use of ever-increasing amounts of data makes it possible to identify genetic structuring at a finer resolution, which can be misinterpreted as species boundaries, while only representing population structure (Funk et al., 2012; Leaché et al., 2019; Sukumaran et al., 2021; Sukumaran and Knowles, 2017). In addition, statistical significance is more easily achieved when many loci are analyzed (Leaché et al., 2019).

**Box 3: Species delimitation methods**

**Automatic Barcode Gap Discovery (ABGD; Puillandre et al., 2012)**. This method estimates pairwise differences between sequences. The distribution of these genetic distances is expected to produce two distinct distributions (bimodal distribution) corresponding to intraspecific and interspecific distances. This method detects the gap between this bimodal distribution and attributes it to the species boundary.

**Generalized Mixed Yule Coalescent (GMYC; Pons et al., 2006)**. This method explores the transition in branching rate in an ultrametric tree, assuming that the between-species branching rate follows a Yule model (Yule, 1925) and the within-species rate follows a neutral coalescent process (J. F. C. Kingman, 1982). A maximum likelihood approach is used to find the time corresponding to this transition, which is attributed to the species boundary.

**Poisson Tree Process (PTP; Zhang et al., 2013).** This method distinguishes species from populations on a phylogenetic tree based on the transition in the substitution rate expected to follow two distinct Poisson distributions.

**ADMIXTURE (Alexander and Lange, 2011) and & STRUCTURE (Pritchard et al., 2000).** Through maximum likelihood for ADMIXTURE and a Bayesian approach for STRUCTURE, these two methods estimate the ancestry coefficient. In other words, they estimate the proportion of the genome of an individual that could be attributed to different ancestral gene pools. Individuals are assigned to genetic clusters with similar ancestries and the number of genetic clusters is estimated by minimizing the Hardy–Weinberg disequilibrium, and can serve as species delimitation.

**Bayesian Phylogenetics and Phylogeography (BPP; Yang and Rannala, 2010).** Us-

ing a multilocus dataset, BPP explores, through the MSC model, a wide range of evolutionary scenarios based on gene tree reconstruction. For each gene tree, it estimates the genetic diversity ($\theta$), the time of species divergence ($\tau$), and the number of species. This Bayesian approach requires setting priors on these model parameters and uses a reverse-jump Markov Chain Monte Carlo (rjMCMC) to estimate their posterior probability.

**PHRAPL (Jackson et al., 2017)** simulates gene trees under different demographic models (including migration, population size, time of coalescent event parameters). Then, the likelihood of the model given the gene tree dataset is estimated by the proportion of simulated gene tree topologies that match the observed one. This method tests an a priori delimitation hypothesis by comparing models supporting them as a unique or two distinct species. In addition, in cases where the best model supports two distinct species with gene flow, a genealogical divergence index (gdi) is estimated from the amount of gene flow and the time coalescent event to reflect the divergence between lineages. Values less than 0.2 are arbitrarily attributed to populations and more than 0.7 to species (Jackson et al., 2017).

## The comparative approach

To limit under- or over-splitting, it could be relevant to compare the genomic differentiation of candidate lineages with taxonomic references. Galtier, 2019 proposed a comparative approach for species delimitation in face of the lack of consensual threshold. Based on genetic differentiation, the threshold obtained from taxa with a consensual taxonomic status should be less arbitrary and maximize the consistency with current taxonomy. Earlier studies have already evaluated the relevance of different population genomic statistics to distinguish the taxonomic level. Hey and Pinho, 2012 examined two measures from isolation models with migration: one reflects gene flow (2NM) and the other separation time ($\tau$). More recently Rosel et al., 2017, studied different population genomic measures ($D_a$, $F_{ST}$, fixed differences, $\phi$ST). However, their results found overlaps in the distribution of these measures between populations and species pairs, so they concluded that these measures have little relevance for species delimitation. The only exception was for $D_a$ (i.e. net synonymous divergence), evaluated by Rosel et al., 2017, that allowed discerning the taxonomic rank of cetacean species, sub-species and populations: above 2% it corresponded to distinct species, while a value below 0.05% indicated populations. Independently, Roux et al., 2016 explored the interruption of gene flow in pairs of populations/species at the metazoan scale, and identi-

fied the same values of $D_a$ associated with evidence of gene flow below 0.05% and reduced gene flow above 2%.

This promising comparative framework provides species delimitation consistent with current taxonomy. However, it should be kept in mind that species with a consensual taxonomic status emerged from a unique speciation scenario that resulted in specific genetic differentiation values. Thus, their comparison with a distinct scenario can be inappropriate and misleading (i.e. the same reproductive isolation value associated with different values of genetic differentiation). Thus, as far as possible, taxonomic revision must include an integrative framework taking into account various criteria such as morphology, ecology, and behavior in order to provide the most robust support for delimitation (Dayrat, 2005; Padial et al., 2010).

These theoretical underpinnings of the speciation process, the species concept, and species delimitation methods illustrate the complexity of applying taxonomic revision. However, their practical applications are essential and have contributed to more accurate species descriptions in recent years (Costa-Araújo et al., 2021; McDonough et al., 2022; Miranda et al., 2018; Peres et al., 2021). For this reason, we investigated taxonomic boundaries within xenarthran species.

# Xenarthrans: a complex story or a story of species complexes

## Evolutionary history

Xenarthra (armadillos, anteaters, and sloths) is one of the four main clades of placental mammals, alongside Afrotheria, Laurasiatheria and Euarchontoglires (Murphy et al., 2001a,b). Crown xenarthrans emerged around 60.4-71.6 Mya in South America (Gibb et al., 2016; Vizcaíno and Bargo, 2014) and diversified during the so-called "splendid isolation", when South America was geographically isolated from other continents (Figure 9), reaching over 200 described genera (McKenna and Bell, 1997). During the Pliocene, South America was re-connected to North America by the emergence of the Isthmus of Panama (Figure 9). This allowed dispersal events from both continents known as the Great American Biotic Interchange (GABI), with notably, xenarthrans successfully colonizing Central and North America (MacDonald et al., 2007; McDonald, 2005; Patterson and Pascual, 1968). At the end of the

Pleistocene, around 11 000 years ago, a major extinction event affected many mammalian groups. Most xenarthrans became extinct, mainly the largest terrestrial forms (i.e. giant sloths and glyptodonts; Lyons et al., 2004). Today, only 14 genera and 39 extant species of armadillos, anteaters, and sloths remain from this past diversity (Abba et al., 2015; Feijó et al., 2018; Miranda et al., 2018; Wetzel et al., 2008).

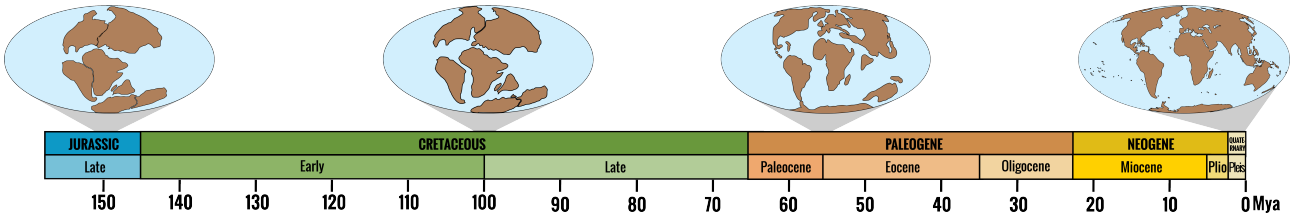


**Figure 9:** Paleogeographic maps of earth (modified from Encyclopaedia Britannica and Meseguer and Condamine, 2017) along the geological time scale back to 150 Mya.

## Taxonomy & ecology

Xenarthrans have evolved striking morphological adaptations with arboreal sloths (Folivora), myrmecophagous anteaters (Vermilingua), and armored armadillos (Cingulata), which together constitute a restricted diversity of 39 living species endemic to the Neotropics (Burgin et al., 2018; Feijó and Brandão, 2022).

Sloths comprise only seven species distributed in two genera: *Bradypus* and *Choloepus* (Gardner, 2005, 2007; Miranda et al., 2023). These species are fully adapted to the arboreal lifestyle and only come down to the ground occasionally. They are specialized on a folivorous diet even if some species also occasionally eat fruits. The two genera can be easily distinguished based on several anatomical characters (teeth, skull, postcranial Wetzel, 1985; Wetzel and Avila-Pires, 1980) but also based on their number of digits in their forelimbs, two-toed sloths (*Choloepus* spp., Choloepodidae) have two digits whereas three-toed sloths (*Bradypus* spp., Bradypodidae) have three. Within *Choloepus*, the Linnaeus's two-toed sloth (*Choloepus didactylus*) is distributed in Amazonia and the Guiana Shield, and the Hoffmann's two-toed sloth (*Choloepus hoffmanni*) presents a disjunct distribution from the West side of the Andes to Nicaragua in Central America, and in the Amazonian parts of Peru, Bolivia and Brazil in South America (Figure 10). Concerning the *Bradypus* genus, the widespread Brown-throated three-toed sloth (*Bradypus variegatus*) is distributed all the way from Honduras to the Atlantic forest of Brazil throughout Amazonia (Figure 10). The Pale-throated

three-toed sloth (*Bradypus tridactylus*) is mainly found in the Guiana Shield. The insular Pygmy three-toed sloth (*Bradypus pygmaeus*) is restricted to the island Escudo de Veraguas (off the coast of Panama), and is critically endangered due to its very small population size (estimated to about 500 individuals). Finally, two recently splitted species are found in the Atlantic forest of Brazil: the northern (*Bradypus torquatus*) and southern (*Bradypus crinitus*) maned three-toed sloths. Probably endangered, these species require a new assessment of their conservation status since their recent taxonomic revision (Miranda et al., 2023).

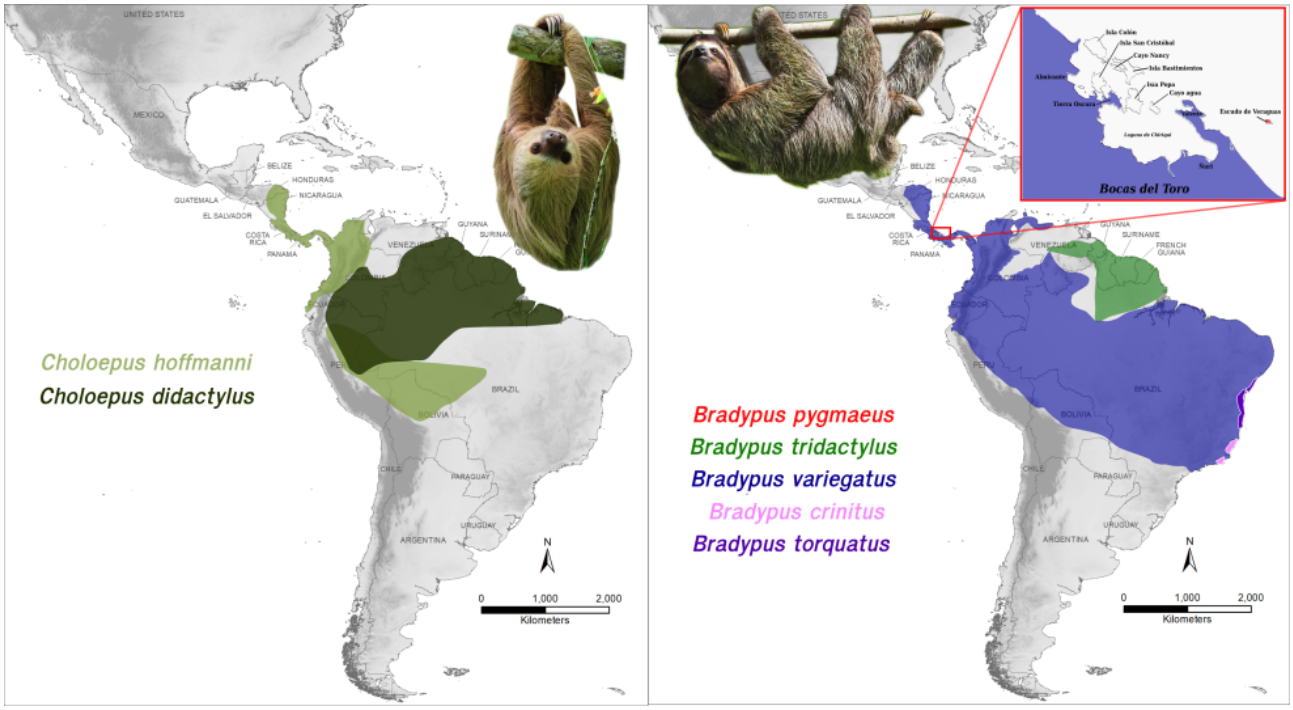


**Figure 10:** Distributions of the two species of two-toed sloths (*Choloepus*; left) and the five species of three-toed sloths (*Bradypus*; right). Adapted from IUCN (xenarthrans.org) and (Miranda et al., 2023). Photo credits: derekz65, benkelly (iNaturalist.org).

Anteaters are represented today by 10 species within the genera *Myrmecophaga*, *Tamandua* and *Cyclopes* (Gardner, 2005, 2007; Miranda et al., 2018). Due to their specialization to the myrmecophagous diet (i.e. mainly composed of ants and termites), striking morphological characteristics have evolved: powerful clawed forelegs, a long protractile tongue, prominent salivary glands, an elongated snout, and a complete loss of teeth (Endo et al., 2007, 2017; Naples, 1999; Reiss, 1997). The giant anteater (*Myrmecophaga tridactyla*) is completely terrestrial and is the most widespread anteater species, ranging from Honduras to northern Argentina (Figure 11). The two species of tamanduas are semi-arboreal with a prehensile tail that increases their mobility in the canopy. Their distributions are separated by the Andes, with the northern tamandua (*Tamandua mexicana*) on the western side of the Andes ranging up to Mexico while the southern tamandua (*Tamandua tetradactyla*) is found

on the eastern side of the Andes down to northern Argentina. Finally, silky anteaters (*Cyclopes*) are represented by strictly arboreal species. This genus has long been considered to include a single widely distributed species, the common silky anteater (*Cyclopes didactylus*). However, Coimbra et al., 2017 and Miranda et al., 2018 successively revised the taxonomy based on both morphological and molecular data, and elevated the Central American silky anteater (*Cyclopes dorsalis*), the Amboro silky anteater (*Cyclopes catellus*) and the Rio Negro silky anteater (*Cyclopes ida*) to the species level, and described three additional new species: Thomas's silky anteater (*Cyclopes thomasi*), the red silky anteater (*Cyclopes rufus*), and Xingu silky anteater (*Cyclopes xinguensis*). The distributions of these seven species seem to fit quite well with major river basins but remain partly uncertain.

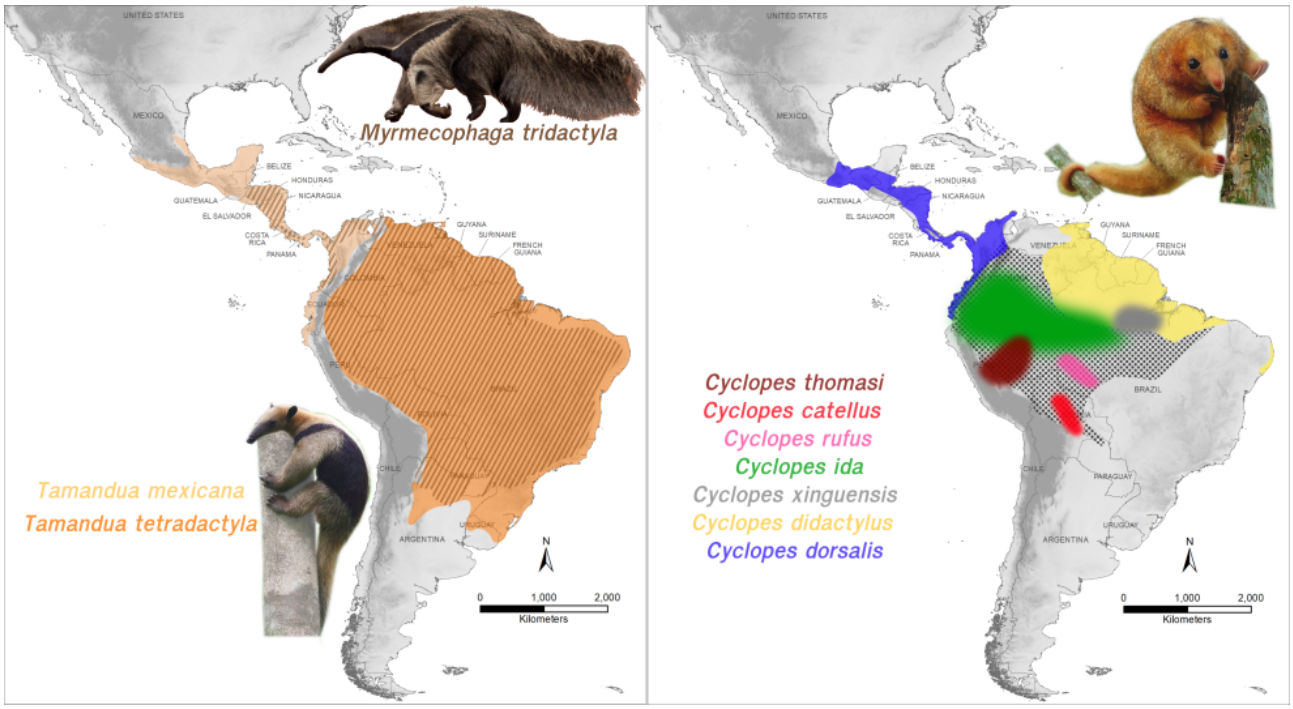


**Figure 11:** Distributions of the species of giant anteater (*Myrmecophaga tridactyla*) and the two species of tamanduas (*Tamandua* spp.; left) and the seven species of silky anteaters (*Cyclopes*; right). Adapted from IUCN (xenarthrans.org) and Miranda et al., 2018. The black dotted area represents an uncertain distribution. Photo credits: joao_andriola, lucaboscain, anthony2005 (iNaturalist.org).

Finally, armadillos are the most speciose group with 22 species classified within the two families of Dasypodidae and Chlamyphoridae (Feijo and Anacleto, 2021; Feijó et al., 2018, 2019; Gardner, 2005; Wetzel et al., 2008). These species are semi-fossorial and characterized by an armored carapace composed of dermal scutes covered by keratinous scales (Grassé, 1955). Dasypodidae comprise 8 species of long nosed armadillos that all belong to the *Dasypus* genus. These species are characterized by a peculiar reproductive system: an obligatory polyembryony. This unique reproductive system in vertebrates corresponds

to the split of the blastocyst that gives rise to two or more clonal embryos (Hamlett, 1933; Loughry et al., 1998). In addition, this genus include various distribution ranges, from the most widespread xenarthran species: the nine banded long nosed armadillo (*Dasypus novemcinctus*) distributed across the Americas to the most restricted distribution of the cingulate with the Yungas Lesser long nosed armadillo (*Dasypus mazzai*) limited to the northern subtropical dry forest of Argentina (Figure 12). The second family Chlamyphoridae include three subfamilies Euphractinae, Chlamyphorinae and Tolypeutinae. Euphractinae comprise four species distributed in eastern and southern South America mostly in dry and open regions with overlapping distributions in Argentina and Paraguay (Figure 12). Chlamyphorinae comprise only two species, the Pink Fairy Armadillo (*Chlamyphorus truncatus*) from Argentina and the Greater Fairy Armadillo (*Calyptophractus retusus*) from the Chaco region (in Argentina, Paraguay and Bolivia; Figure 12). These two species are very small compared to other cingulate species (only few centimeters for the former, and up to 20 cm for the later), and are well adapted to the fossorial life with a mole-like appearance (reduce eyes and ears with developed claws; Borghi et al., 2011; Slade, 1891; P. Smith, 2008; Smith and Owen, 2017). Finally, Tolypeutinae comprises 8 species: the Giant Armadillo (*Priodontes maximus*), unique representative of the genus Priodontes, two species of Three-banded Armadillo (*Tolypeutes* spp.), and five species of Naked-tailed Armadillo (*Cabassous* spp.; Figure 12).

**Table 1:** Xenarthra taxonomy detailed for the 39 current species adapted from urlwww.mammaldiversity.org

| Species name | Common Name | Order | Suborder | Family | Subfamily | Genus | Subgenus | Species Author | Species Year |
|---|---|---|---|---|---|---|---|---|---|
| Choloepus didactylus | Linnaeus's Two-toed Sloth | Pilosa | Folivora | Choloepodidae | – | Choloepus | – | Linnaeus | 1758 |
| Choloepus hoffmanni | Hoffmann's Two-toed Sloth | Pilosa | Folivora | Choloepodidae | – | Choloepus | – | W. Peters | 1858 |
| Bradypus torquatus | Northern Maned Three-toed Sloth | Pilosa | Folivora | Bradipodidae | – | Bradypus | – | Illiger | 1811 |
| Bradypus crinitus | Southern Maned Three-toed Sloth | Pilosa | Folivora | Bradipodidae | – | Bradypus | – | J. E. Gray | 1850 |
| Bradypus pygmaeus | Pygmy Three-toed Sloth | Pilosa | Folivora | Bradipodidae | – | Bradypus | – | Anderson & Handley | 2001 |
| Bradypus tridactylus | Pale-throated Three-toed Sloth | Pilosa | Folivora | Bradipodidae | – | Bradypus | – | Linnaeus | 1758 |
| Bradypus variegatus | Brown-throated Three-toed Sloth | Pilosa | Folivora | Bradipodidae | – | Bradypus | – | Schinz | 1825 |
| Cyclopes catellus | Amboro Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | O. Thomas | 1928 |
| Cyclopes thomasi | Thomas's Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | Miranda et al. | 2017 |
| Cyclopes ida | Rio Negro Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | O. Thomas | 1900 |
| Cyclopes rufus | Red Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | Miranda et al. | 2017 |
| Cyclopes xinguensis | Xingu Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | Miranda et al. | 2017 |
| Cyclopes dorsalis | Central American Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | J. E. Gray | 1865 |
| Cyclopes didactylus | Common Silky Anteater | Pilosa | Vermilingua | Cyclopedidae | – | Cyclopes | – | Linnaeus | 1758 |
| Myrmecophaga tridactyla | Giant Anteater | Pilosa | Vermilingua | Myrmecophagidae | – | Myrmecophaga | – | Linnaeus | 1758 |
| Tamandua mexicana | Northern Tamandua | Pilosa | Vermilingua | Myrmecophagidae | – | Tamandua | – | Saussure | 1860 |
| Tamandua tetradactyla | Southern Tamandua | Pilosa | Vermilingua | Myrmecophagidae | – | Tamandua | – | Linnaeus | 1758 |
| Dasypus kappleri | Greater Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Hyperoamdon | Krauss | 1862 |
| Dasypus beniensis | East Amazonian Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Hyperoamdon | Lönnberg | 1942 |
| Dasypus pastasae | West Amazonian Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Hyperoamdon | O. Thomas | 1901 |
| Dasypus septemcinctus | Brazilian Lesser Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Muletia | Linnaeus | 1758 |
| Dasypus pilosus | Hairy Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Dasypus | Fitzinger | 1856 |
| Dasypus sabanicola | Northern Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Dasypus | Mondolfi | 1968 |
| Dasypus mazzai | Yungas Lesser Long-nosed Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Dasypus | Yepes | 1933 |
| Dasypus novemcinctus | Nine-banded Armadillo | Cingulata | – | Dasipodidae | – | Dasypus | Dasypus | Linnaeus | 1758 |
| Euphractus sexcinctus | Six-banded Armadillo | Cingulata | – | Chlamyphoridae | Euphractinae | Euphractus | – | Linnaeus | 1758 |
| Chaetophractus villosus | Large Hairy Armadillo | Cingulata | – | Chlamyphoridae | Euphractinae | Chaetophractus | – | Desmarest | 1804 |
| Zaedyus pichiy | Pichi | Cingulata | – | Chlamyphoridae | Euphractinae | Zaedyus | – | Desmarest | 1804 |
| Chaetophractus vellerosus | Screaming Hairy Armadillo | Cingulata | – | Chlamyphoridae | Euphractinae | Chaetophractus | – | J. E. Gray | 1865 |
| Chlamyphorus truncatus | Pink Fairy Armadillo | Cingulata | – | Chlamyphoridae | Chlamyphorinae | Chlamyphorus | – | Harlan | 1825 |
| Calyptophractus retusus | Greater Fairy Armadillo | Cingulata | – | Chlamyphoridae | Chlamyphorinae | Calyptophractus | – | Burmeister | 1863 |
| Priodontes maximus | Giant Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Priodontes | – | Kerr | 1792 |
| Tolypeutes matacus | Southern Three-banded Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Tolypeutes | – | Desmarest | 1804 |
| Tolypeutes tricinctus | Brazilian Three-banded Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Tolypeutes | – | Linnaeus | 1758 |
| Cabassous tatouay | Greater Naked-tailed Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Cabassous | – | Desmarest | 1804 |
| Cabassous chacoensis | Chacoan Naked-tailed Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Cabassous | – | Wetzel | 1980 |
| Cabassous centralis | Northern Naked-tailed Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Cabassous | – | G. S. Miller | 1899 |
| Cabassous unicinctus | Amazon Naked-tailed Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Cabassous | – | Linnaeus | 1758 |
| Cabassous squamicaudis | Cerrado Naked-tailed Armadillo | Cingulata | – | Chlamyphoridae | Tolipeutinae | Cabassous | – | Lund | 1845 |


**Figure 12:** Distribution of the 22 species of armadillos. Adapted from IUCN (xenarthrans.org) and Feijo and Anacleto, 2021; Feijó et al., 2019. Photo credits: andresiade, silviolamothe, preli, Bradley Davis sclateria, Tomás Tamagno (iNaturalist.org), Mariella Superina, Quentin Martinez.

All xenarthran species are endemic to the Neotropics and represent the only living witnesses of the past diversity of this major placental order. Some of them have a wide distribution, encompassing geological formations or major topological structures such as the Andes, the Guiana Shield or Central America. Those formations have been identified as biogeographic barriers for numerous species (Cortés-Ortiz et al., 2003; Esquerré et al., 2019; Fouquet et al., 2012; Gutiérrez-García and Vázquez-Domínguez, 2013; Redondo et al., 2008; Weir and Price, 2011) and could also have been involved in the speciation process within xenarthrans. Some studies have notably identified diverging lineages with uncertain taxonomic status separated by such formations within *Dasypus novemcinctus* (Arteaga et al., 2020; Billet et al., 2017; Feijó et al., 2018, 2019; Gibb et al., 2016; Hautier et al., 2017), *Bradypus variegatus* (Ruiz-García et al., 2020), and *Cyclopes didactylus* (Coimbra et al., 2017; Miranda et al., 2018). This potentially hidden diversity is even more conceivable as 71.4% of xenarthrans genera have not been the subject of a taxonomic study in 50 years (Figure 5 of Feijó and Brandão, 2022). Feijó and Brandão, 2022 suggested that taxonomic stability in xenarthrans could rather be explained by a lack of taxonomic studies than a good understanding of their actual diversity. In fact, when considering the recently studied genera, important taxonomic changes have been implemented at the family, genus, subgenus, species, and subspecies levels (Feijó et al., 2018, 2019; Gibb et al., 2016; Miranda et al., 2018; Miranda et al., 2023). This recent taxonomic activity suggests that a lot of work remains to be done in this clade. The misperception of taxonomic boundaries within this clade may have induced an underestimation of the current diversity of this group, and consequently a potential misjudgement regarding their conservation status.

## Phylogeny

Figuring out the phylogenetic relationships of xenarthrans has long been challenging because of their peculiar morphology (Delsuc and Douzery, 2008). If their monophyly has always been recognized and well supported, notably by their particular vertebral articulation that led to their name (*xenos* = strange and *arthros* = articulation ; Engelmann, 1985; Gaudin, 1999), either their position within mammals and their intra-ordinal relationships have long been puzzling for morphologists (Engelmann, 1985; Novacek, 1992; Shoshani and McKenna, 1998).

The emergence of molecular studies provided an unprecedented advance in our under-

standing of xenarthran relationships. Notably, in their early molecular phylogenetic studies, Delsuc et al., 2003, 2001, 2002 resolved intra-ordinal relationships and confirmed previous morphological findings by grouping anteaters and sloths into Pilosa. Möller-Krull et al., 2007 and Delsuc et al., 2012 later included more xenarthran species encompassing all xenarthran genera in their newly defined molecular phylogenetic framework.

However, as highlighted by Delsuc et al., 2003, the internal relationships within the two armadillo subfamilies Tolypeutinae and Euphractinae appeared rather unresolved (short internal nodes and low statistical support values) even with different types of molecular data (respectively with non-coding retroposon flanking sequences and the concatenation of nuclear exons and two mitochondrial genes). This result was somewhat expected in Euphractinae (Abba et al., 2015), because the three genera (*Euphractus*, *Zaedyus*, *Chaetophractus*) appeared very similar and previous morphological studies found support for conflictual groupings (Abrantes and Bergqvist, 2006; Engelmann, 1985; Gaudin and Wible, 2006; Patterson et al., 1989). However, this was more surprising regarding Tolypeutinae in which morphological analyses consistently strongly supported the grouping of *Priodontes* and *Cabassous* into Priodontini, to the exclusion of *Tolypeutes* (Abrantes and Bergqvist, 2006; Cetica et al., 1998; Engelmann, 1985; Gaudin and Wible, 2006; McKenna and Bell, 1997). Gibb et al., 2016 provided the first molecular phylogenetic study including all xenarthran species described at the time. This mitogenomic study clearly supported the paraphyly of Priodontini by rather grouping *Cabassous* with *Tolypeutes* to the exclusion of *Priodontes*, but it did not allow deciphering the relationships within Euphractinae (Figure 13). The discordances with morphological studies and among molecular datasets have been suspected to be induced by ancient gene flow, incomplete lineage sorting (in relation to their fast diversification), or morphological convergence (Delsuc et al., 2003, 2002; Gibb et al., 2016). However, until now, phylogenetic studies (Abba et al., 2015; Delsuc and Douzery, 2008; Delsuc et al., 2012; Gibb et al., 2016; Möller-Krull et al., 2007) have been restricted to mitogenomes and a handful of nuclear markers that incompletely reflect xenarthran evolutionary history.

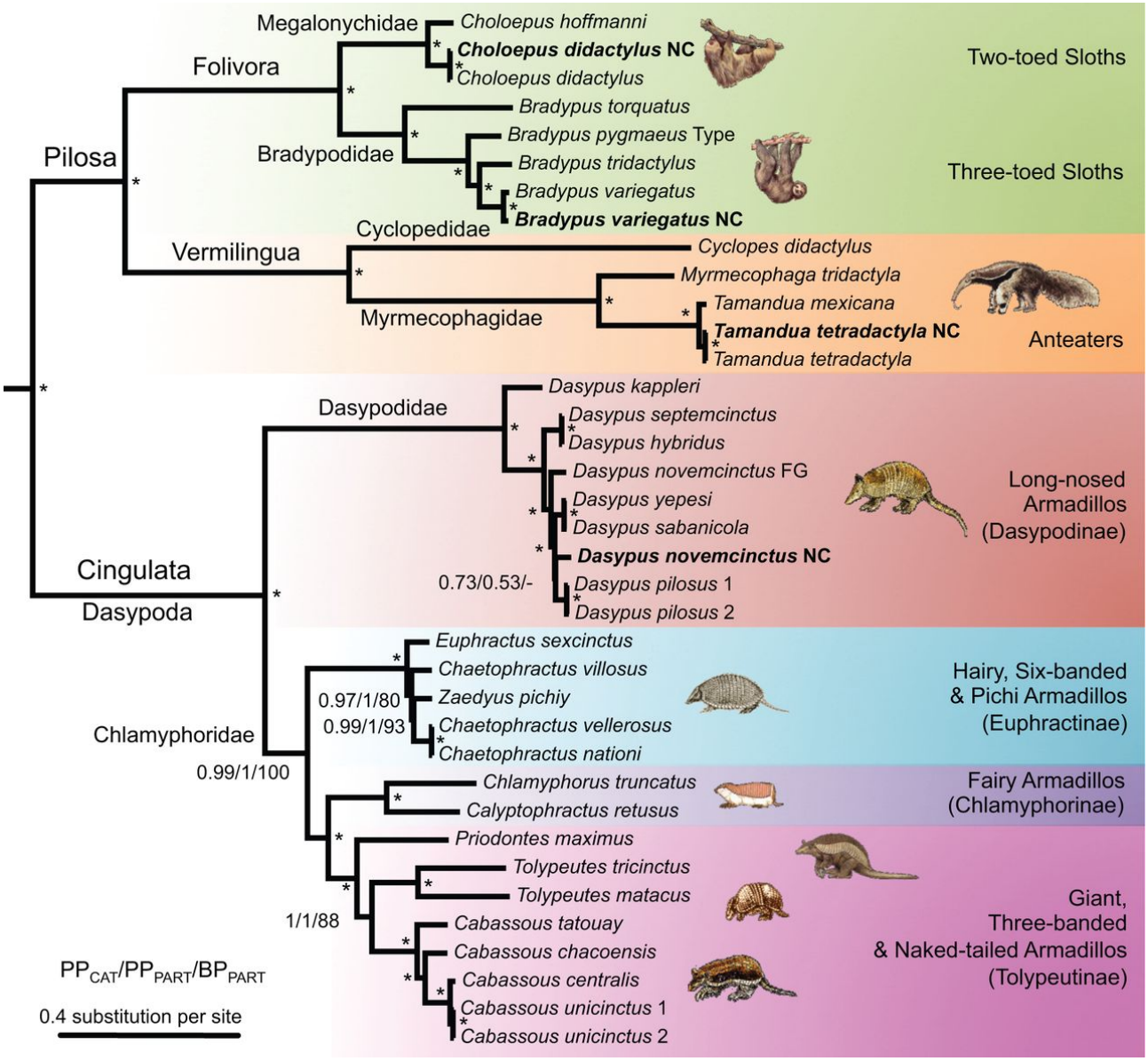


**Figure 13:** Phylogenetic reconstruction based on 33 mitogenomes representing the 31 xenarthran species described as of 2016 (Gibb et al., 2016).

## Objectives of this thesis

This thesis takes place in a context where our understanding of the speciation process is improving, but where the concept of species is still debated (Queiroz, 2007; Stankowski and Ravinet, 2021b). Even though the speciation process is continuous (Stankowski and Ravinet, 2021a), taxonomy has inherited a discrete vision, and the lack of a consensus on a threshold along this speciation continuum contributes to the difficulty of delimiting taxonomic species (Zachos, 2018). Nevertheless, species are the basic conservation unit, and taxonomic studies are urgently needed, especially for large mammals, which are more vulnerable to extinction (Feijó and Brandão, 2022).

With the recent advances and decreasing costs of DNA sequencing, it is now conceivable to produce genomic data to further examine xenarthran phylogeny and taxonomy. Addressing these issues requires acquiring genomic data from potentially threatened species. In our view, it was imperative to use the least invasive sampling methods possible by using museomics, roadkill, and freely available data, notwithstanding the challenges associated with these data. By gathering 95 available genomes (73 as mitogenomes and 22 as whole genome) and 183 newly generated resequencing genomic data (including 71 as exon captures and 72 as whole genomes), this PhD project tackles the following general challenges: i) What is the taxonomic status of recently evidenced xenarthran distinct lineages/species?; ii) Does considering gene flow and ILS allow a better understanding of the evolutionary history of xenarthran species?; iii) What factors (biogeography, local adaptation, demography) have contributed to speciation and diversification in this mammalian group?; and iv) Do the different taxonomic categories correspond to similar degree of genetic differenciatin among mammals?

We have addressed these questions in two chapters. The **first chapter** illustrates the power and challenges of museomics to delimit the nine-banded armadillo (*Dasypus novemcinctus*) species complex. Museum samples, more prone to cross-contamination and genotyping errors, can lead to inaccurate conclusions about genetic diversity and may even mimic gene flow, which is particularly problematic for species delimitation. After careful data cleaning, we reconstructed phylogenetic relationships, used several approaches to delimit species, and assess genetic exchange. This provided a better understanding of speciation events and to recognize four species (*D. novemcinctus*, *D. fenestratus*, *D. mexicanus*, and *D. guianensis*), including a new species endemic to the Guiana Shield, which is the first new armadillo species described in the last 30 years.

In the **second chapter**, we built up a nearly exhaustive whole genome dataset representing 88% of valid nominal xenarthran species. This allowed us to revise the taxonomy of this clade using the genetic differentiation between closely related species to provide a comparative taxonomic framework for delimiting two species complexes (*Bradypus* spp. and *Cyclopes* spp.). By evaluating genetic diversity, inbreeding and demography of xenarthran species, we propose future directions for conservation status assessment. Considering this revised taxonomic framework we reconstruct the most comprehensive time-calibrated phylogeny of

Xenarthra and further, disentangle the contribution of incomplete lineage sorting and gene flow in discordant topological signals within xenarthran phylogeny.

Finally, in a general discussion, I tackle non-invasive sampling strategies and their challenges. Subsequently, I summarize the systematic changes we have supported and discuss the implications and perspective of these findings for understanding xenarthran evolution. Finally, I raise broader questions about i) factors influencing speciation of xenarthran and ii) the genetic differentiation homogeneity of the species taxonomic status, supporting my points with new preliminary results in the Appendix.

# References

Abba, A. M. et al. (2015). "Systematics of hairy armadillos and the taxonomic status of the Andean hairy armadillo (Chaetophractus nationi)". In: *Journal of Mammalogy* 96.4. ISBN: 1545-1542 Publisher: Oxford University Press US, pp. 673–689.

Abbott, R. et al. (2013). "Hybridization and speciation". In: *Journal of evolutionary biology* 26.2. ISBN: 1010-061X Publisher: Wiley Online Library, pp. 229–246.

Abrantes, E. A. L. and L. P. Bergqvist (2006). "Proposta filogenética para os Dasypodidae (Mammalia: Cingulata)". In: *Paleontologia de vertebrados: grandes temas e contribuições científicas (V. Gallo, PM Brito, HMA Silva, and FJ Figueiredo, eds.). Interciência Ltda., Rio de Janeiro, Brazil*, pp. 261–274.

Alexander, D. H. and K. Lange (2011). "Enhancements to the ADMIXTURE algorithm for individual ancestry estimation". In: *BMC bioinformatics* 12. Publisher: Springer, pp. 1–6.

Arnold, M. L. and M. L. Arnold (2006). *Evolution through genetic exchange*. Oxford University Press. ISBN: 0-19-857006-6.

Arteaga, M. C. et al. (2020). "Conservation genetics, demographic history, and climatic distribution of the nine-banded armadillo (Dasypus novemcinctus): an analysis of its mitochondrial lineages". In: *Conservation genetics in mammals: integrative research using novel approaches*. ISBN: 3030333337 Publisher: Springer, pp. 141–163.

Avise, J. C. and R. Ball (1990). "Principles of genealogical concordance in species concepts and biological taxonomy". In: *Oxford surveys in evolutionary biology* 7. Publisher: Oxford University Press, pp. 45–67.

Ayala, F. J. and W. M. Fitch (1997). "Genetics and the origin of species: an introduction". In: *Proceedings of the National Academy of Sciences* 94.15. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 7691–7697.

Bao, K., R. H. Melde, and N. P. Sharp (2022). "Are mutations usually deleterious? A perspective on the fitness effects of mutation accumulation". In: *Evolutionary ecology* 36.5. ISBN: 0269-7653 Publisher: Springer, pp. 753–766.

Barberousse, A. and S. Samadi (2010). "Species from Darwin onward". In: *Integrative Zoology* 5.3. ISBN: 1749-4877 Publisher: Wiley Online Library, pp. 187–197.

Barton, N. (1979). "Gene flow past a cline". In: *Heredity* 43.3. Publisher: Nature Publishing Group, pp. 333–339. ISSN: 1365-2540.

Barton, N. and B. O. Bengtsson (1986). "The barrier to genetic exchange between hybridising populations". In: *Heredity* 57.3. ISBN: 1365-2540 Publisher: Nature Publishing Group, pp. 357–376.

Bateson, W. (1909). "Heredity and variation in modern lights". In: *Darwin and modern science*. Publisher: Cambridge University Press.

Bauhin, C. (1623). "Pinax theatri botanici". In: *Ludovici Regis, Basileae*.

Billet, G. et al. (2017). "The hidden anatomy of paranasal sinuses reveals biogeographically distinct morphotypes in the nine-banded armadillo (Dasypus novemcinctus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3593. ISSN: 2167-8359.

Borghi, C. E. et al. (2011). "Updated distribution of the pink fairy armadillo Chlamyphorus truncatus (Xenarthra, Dasypodidae), the world's smallest armadillo". In: *Edentata* 12.1. ISBN: 1413-4411 Publisher: BioOne, pp. 14–19.

Burgin, C. J. et al. (2018). "How many species of mammals are there?" In: *Journal of Mammalogy* 99.1. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 1–14.

Butlin, R. K. and C. M. Smadja (2018). "Coupling, reinforcement, and speciation". In: *The American Naturalist* 191.2. ISBN: 0003-0147 Publisher: University of Chicago Press Chicago, IL, pp. 155–172.

Cain, A. J. (1994). "Rank and sequence in Caspar Bauhin's Pinax". In: *Botanical journal of the Linnean Society* 114.4. ISBN: 1095-8339 Publisher: Oxford University Press, pp. 311–356.

Campagna, L. et al. (2014). "Secondary contact followed by gene flow between divergent mitochondrial lineages of a widespread Neotropical songbird (Zonotrichia capensis)". In: *Biological Journal of the Linnean Society* 111.4. ISBN: 0024-4066 Publisher: Oxford University Press, pp. 863–868.

Campbell, C. R., J. Poelstra, and A. D. Yoder (2018). "What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species". In: *Biological Journal of the Linnean Society* 124.4. Publisher: Oxford University Press UK, pp. 561–583. ISSN: 0024-4066.

Carstens, B. C. et al. (2013). "How to fail at species delimitation". In: *Molecular Ecology* 22.17. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 4369–4383.

Cetica, P. D. et al. (1998). "Evolutionary sperm morphology and morphometry in armadillos". In: *Journal of submicroscopic cytology and pathology* 30.2, pp. 309–314. ISSN: 1122-9497.

Coimbra, R. T. F. et al. (2017). "Phylogeographic history of South American populations of the silky anteater *Cyclopes didactylus* (Pilosa: Cyclopedidae)". In: *Genetics and Molecular Biology* 40. Publisher: Sociedade Brasileira de Genética, pp. 40–49. ISSN: 1415-4757, 1678-4685. DOI: 10.1590/1678-4685-GMB-2016-0040.

Cortés-Ortiz, L. et al. (2003). "Molecular systematics and biogeography of the Neotropical monkey genus, Alouatta". In: *Molecular Phylogenetics and Evolution* 26.1, pp. 64–81. ISSN: 1055-7903. DOI: 10.1016/S1055-7903(02)00308-1.

Costa-Araújo, R. et al. (2021). "An integrative analysis uncovers a new, pseudo-cryptic species of Amazonian marmoset (Primates: Callitrichidae: Mico) from the arc of deforestation". In: *Scientific Reports* 11.1. ISBN: 2045-2322 Publisher: Nature Publishing Group UK London, p. 15665.

Coyne, J. A. and H. A. Orr (2004). *Speciation*. Vol. 37. Sinauer Associates Sunderland, MA.

Crow, J. F. and M. Kimura (1970). "An introduction to population genetics theory." In: *An introduction to population genetics theory.* Publisher: New York, Evanston and London: Harper & Row, Publishers.

Cruickshank, T. E. and M. W. Hahn (2014). "Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow". In: *Molecular ecology* 23.13. Publisher: Wiley Online Library, pp. 3133–3157. ISSN: 0962-1083.

Darwin, C. (1859). "On the origin of species: facsimile of the first edition". In: Publisher: LONDON: JOHN MURRAY, ALBEMARLE STREET.

Dayrat, B. (2005). "Towards integrative taxonomy". In: *Biological journal of the Linnean society* 85.3. ISBN: 1095-8312 Publisher: Oxford University Press, pp. 407–417.

De Queiroz, K. (1988). "Systematics and the Darwinian revolution". In: *Philosophy of Science* 55.2. ISBN: 0031-8248 Publisher: Cambridge University Press, pp. 238–259.

Delsuc, F. and E. J. Douzery (2008). "Recent advances and future prospects in xenarthran molecular phylogenetics". In: *The biology of the Xenarthra* 11. Publisher: University Press of Florida Gainesville.

Delsuc, F., M. J. Stanhope, and E. J. P. Douzery (2003). "Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes". In: *Molecular Phylogenetics and Evolution* 28.2, pp. 261–275. ISSN: 1055-7903. DOI: 10.1016/S1055-7903(03)00111-8.

Delsuc, F. et al. (2001). "The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies: implications for the status of the enigmatic fossil Eurotamandua". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1476. Publisher: Royal Society, pp. 1605–1615. DOI: 10.1098/rspb.2001.1702.

Delsuc, F. et al. (2002). "Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting". In: *Molecular Biology and Evolution* 19.10, pp. 1656–1671. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a003989.

Delsuc, F. et al. (2012). "Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos". In: *Molecular Phylogenetics and Evolution* 62.2, pp. 673–680. ISSN: 1055-7903. DOI: https://doi.org/10.1016/j.ympev.2011.11.008.

Dieckmann, U. (2004). *Adaptive speciation*. Cambridge University Press. ISBN: 0-521-82842-2.

Dobzhansky, T. H. (1936). "Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids". In: *Genetics* 21.2. Publisher: Oxford University Press, p. 113.

Dobzhansky, T. (1935). "A critique of the species concept in biology". In: *Philosophy of Science* 2.3. ISBN: 0031-8248 Publisher: Cambridge University Press, pp. 344–355.

— (1937). "Genetic nature of species differences". In: *The American Naturalist* 71.735. ISBN: 0003-0147 Publisher: Science Press, pp. 404–420.

Doorn, G. S. van, P. Edelaar, and F. J. Weissing (2009). "On the origin of species by natural and sexual selection". In: *Science* 326.5960. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 1704–1707.

Endo, H. et al. (2007). "Three-dimensional CT examination of the mastication system in the giant anteater". In: *Zoological Science* 24.10. ISBN: 0289-0003 Publisher: BioOne, pp. 1005–1011.

Endo, H. et al. (2017). "Macroscopic and CT examinations of the mastication mechanism in the southern tamandua". In: *Mammal study* 42.2. ISBN: 1343-4152 Publisher: BioOne, pp. 89–96.

Engelmann, G. (1985). "The phylogeny of Xenarthra". In: *The evolution and ecology of armadillos, sloths, and vermilinguas*. Publisher: Smithsonian Institution Press, pp. 51–64.

Esquerré, D. et al. (2019). "How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae)". In: *Evolution* 73.2. _eprint: https://onlinelibrary.wile pp. 214–230. ISSN: 1558-5646. DOI: 10.1111/evo.13657.

Feder, J. L., S. P. Egan, and P. Nosil (2012). "The genomics of speciation-with-gene-flow". In: *Trends in Genetics* 28.7. Publisher: Elsevier, pp. 342–350. ISSN: 0168-9525.

Feder, J. L. and P. Nosil (2010). "The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation". In: *Evolution* 64.6. ISBN: 1558-5646 Publisher: Blackwell Publishing Inc Malden, USA, pp. 1729–1747.

Feijo, A. and T. C. Anacleto (2021). "Taxonomic revision of the genus Cabassous McMurtrie, 1831 (Cingulata: Chlamyphoridae), with revalidation of Cabassous squamicaudis (Lund, 1845)". In: *Zootaxa* 4974.1. ISBN: 1175-5334, pp. 47–78–47–78.

Feijó, A. and M. V. Brandão (2022). *Taxonomy as the first step towards conservation: an appraisal on the taxonomy of medium-and large-sized Neotropical mammals in the 21st century*. Vol. 39. Publication Title: Zoologia (Curitiba). SciELO Brasil. ISBN: 1984-4670.

Feijó, A., B. D. Patterson, and P. Cordeiro-Estrela (2018). "Taxonomic revision of the long-nosed armadillos, Genus Dasypus Linnaeus, 1758 (Mammalia, Cingulata)". In: *PLoS One* 13.4. Publisher: Public Library of Science San Francisco, CA USA, e0195084. ISSN: 1932-6203.

Feijó, A. et al. (2019). "Phylogeny and molecular species delimitation of long-nosed armadillos (Dasypus: Cingulata) supports morphology-based taxonomy". In: *Zoological Journal of the Linnean Society* 186.3. Publisher: Oxford University Press UK, pp. 813–825. ISSN: 0024-4082.

Fitch, W. M. and E. Margoliash (1967). "Construction of phylogenetic trees: a method based on mutation distances as estimated from cytochrome c sequences is of general applicability." In: *Science* 155.3760. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 279–284.

Fouquet, A. et al. (2012). "Multiple quaternary refugia in the eastern Guiana Shield revealed by comparative phylogeography of 12 frog species". In: *Systematic Biology* 61.3. ISBN: 1076-836X Publisher: Oxford University Press, p. 461.

Funk, W. C. et al. (2012). "Harnessing genomics for delineating conservation units". In: *Trends in Ecology & Evolution* 27.9. ISBN: 0169-5347 Publisher: Elsevier, pp. 489–496.

Galtier, N. (2019). "Delineating species in the speciation continuum: A proposal". In: *Evolutionary Applications* 12.4. Publisher: Wiley Online Library, pp. 657–663. ISSN: 1752-4571.

Gardner, A. L. (2005). "Mammal Species of the World: a taxonomic and geographic reference. The Johns Hopkins University Press, Baltimore, Estados Unidos," in.

— (2007). *Mammals of South America, volume 1: marsupials, xenarthrans, shrews, and bats*. University of Chicago Press. ISBN: 0-226-28242-2.

Gaudin, T. J. and J. R. Wible (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*, Chicago (IL). University of Chicago Press, pp. 153–198.

Gaudin, T. (1999). "The morphology of xenarthrous vertebrae (Mammalia: Xenarthra)". In: *Fieldiana: Geology* 41, pp. 1–38.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species (MPB-41)*. Vol. 41. Princeton university press. ISBN: 0-691-11983-X.

Gibb, G. C. et al. (2016). "Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans". In: *Molecular Biology and Evolution* 33.3, pp. 621–642. ISSN: 0737-4038. DOI: 10.1093/molbev/msv250.

Gourbiere, S. and J. Mallet (2010). "Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball"". In: *Evolution* 64.1. ISBN: 1558-5646 Publisher: Blackwell Publishing Inc Malden, USA, pp. 1–24.

Grassé, P. P. (1955). "Ordre des édentés". In: *Traité de zoologie* 17.2. Publisher: Masson, pp. 1182–1246.

Gutiérrez-García, T. A. and E. Vázquez-Domínguez (2013). "Consensus between genes and stones in the biogeographic and evolutionary history of Central America". In: *Quaternary Research* 79.3. ISBN: 0033-5894 Publisher: Cambridge University Press, pp. 311–324.

Hamlett, G. W. D. (1933). "Polyembryony in the armadillo: genetic or physiological?" In: *The Quarterly Review of Biology* 8.3. ISBN: 0033-5770 Publisher: Williams and Wilkins, pp. 348–358.

Harr, B. (2006). "Genomic islands of differentiation between house mouse subspecies". In: *Genome research* 16.6. ISBN: 1088-9051 Publisher: Cold Spring Harbor Lab, pp. 730–737.

Hautier, L. et al. (2017). "Beyond the carapace: skull shape variation and morphological systematics of long-nosed armadillos (genus Dasypus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3650. ISSN: 2167-8359.

Hawkins, B. A. (2001). "Ecology's oldest pattern?" In: *Trends in Ecology & Evolution* 16.8. ISBN: 0169-5347 Publisher: Elsevier, p. 470.

Hebert, P. D. et al. (2003). "Biological identifications through DNA barcodes". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1512. ISBN: 0962-8452 Publisher: The Royal Society, pp. 313–321.

Hermisson, J. and P. S. Pennings (2005). "Soft sweeps: molecular population genetics of adaptation from standing genetic variation". In: *Genetics* 169.4. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 2335–2352.

Hey, J. (2001). *Genes, categories, and species: the evolutionary and cognitive cause of the species problem*. Oxford University Press. ISBN: 0-19-534931-8.

Hey, J. and C. Pinho (2012). "Population genetics and objectivity in species diagnosis". In: *Evolution* 66.5. Publisher: Wiley Online Library, pp. 1413–1429. ISSN: 0014-3820.

Hillis, D. M., C. Moritz, and B. K. Mable (1996). *Molecular systematics*. Vol. 23. Sinauer.

Hobolth, A. et al. (2011). "Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection". In: *Genome research* 21.3. ISBN: 1088-9051 Publisher: Cold Spring Harbor Lab, pp. 349–356.

Hudson, R. R. (1983). "Properties of a neutral allele model with intragenic recombination". In: *Theoretical population biology* 23.2. ISBN: 0040-5809 Publisher: Elsevier, pp. 183–201.

Jackson, N. D. et al. (2017). "PHRAPL: phylogeographic inference using approximate likelihoods". In: *Systematic Biology* 66.6. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 1045–1053.

Kearns, A. M. et al. (2018). "Genomic evidence of speciation reversal in ravens". In: *Nature Communications* 9.1. ISBN: 2041-1723 Publisher: Nature Publishing Group UK London, p. 906.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press. ISBN: 0-521-31793-2.

Kimura, M. and J. F. Crow (1964). "The number of alleles that can be maintained in a finite population". In: *Genetics* 49.4. Publisher: Oxford University Press, p. 725.

Kingman, J. F. (1982). "On the genealogy of large populations". In: *Journal of applied probability* 19 (A). ISBN: 0021-9002 Publisher: Cambridge University Press, pp. 27–43.

Kingman, J. F. C. (1982). "The coalescent". In: *Stochastic processes and their applications* 13.3. ISBN: 0304-4149 Publisher: Elsevier, pp. 235–248.

Kulmuni, J. et al. (2020). *Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers*. Vol. 375. Issue: 1806 Pages: 20190528 Publication Title: Philosophical Transactions of the Royal Society B. The Royal Society. ISBN: 0962-8436.

Leaché, A. D. et al. (2019). "The spectre of too many species". In: *Systematic Biology* 68.1. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 168–181.

Lewis, M. J. (2010). *Classification of Living Organisms*. The Rosen Publishing Group, Inc. ISBN: 1-4358-9535-5.

Linnaeus, C. (1758). *Systema naturae*. Vol. 1. Stockholm Laurentii Salvii.

Loughry, W. J. et al. (1998). "Polyembryony in armadillos: an unusual feature of the female nine-banded armadillo's reproductive tract may explain why her litters consist of four genetically identical offspring". In: *American Scientist* 86.3. ISBN: 0003-0996 Publisher: JSTOR, pp. 274–279.

Lowry, D. B. et al. (2008). "The strength and genetic basis of reproductive isolating barriers in flowering plants". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1506. ISBN: 0962-8436 Publisher: The Royal Society London, pp. 3009–3021.

Luo, A. et al. (2018). "Comparison of methods for molecular species delimitation across a range of speciation scenarios". In: *Systematic Biology* 67.5. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 830–846.

Lyons, S. K., F. A. Smith, and J. H. Brown (2004). "Of mice, mastodons and men: human-mediated extinctions on four continents". In: *Evolutionary Ecology Research* 6.3. ISBN: 1522-0613 Publisher: Evolutionary Ecology, Ltd., pp. 339–358.

MacDonald, G. H., S. F. Vizcaíno, and M. S. Bargo (2007). "Skeletal anatomy and the fossil history of the Vermilingua". In: *The Biology of the Xenarthra/Vizcaíno, Sergio Fabián; Loughry, WJ*.

Maddison, W. P. (1997). "Gene trees in species trees". In: *Systematic biology* 46.3. ISBN: 1076-836X Publisher: Society of Systematic Biologists, pp. 523–536.

Malinsky, M. et al. (2015). "Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake". In: *Science* 350.6267. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 1493–1498.

Mallet, J. (2008). "Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1506. ISBN: 0962-8436 Publisher: The Royal Society London, pp. 2971–2986.

— (2010). "Group selection and the development of the biological species concept". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1547. ISBN: 0962-8436 Publisher: The Royal Society, pp. 1853–1863.

Manktelow, M. (2010). "History of taxonomy". In: *Lecture from Dept. of Systematic Biology, Uppsala University* 29.

Matute, D. R. et al. (2010). "A test of the snowball theory for the rate of evolution of hybrid incompatibilities". In: *Science* 329.5998. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 1518–1521.

Mayden, R. L. (1997). "A hierarchy of species concepts: the denouement in the saga of the species problem". In.

Mayr, E. (1940). "Speciation phenomena in birds". In: *The American Naturalist* 74.752. ISBN: 0003-0147 Publisher: Science Press, pp. 249–278.

— (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press. ISBN: 0-674-86250-3.

— (1949). "The species concept: semantics versus semantics". In: *Evolution* 3.4. ISBN: 0014-3820 Publisher: Wiley Online Library, pp. 371–372.

— (1963). *Animal species and evolution*. Harvard University Press. ISBN: 0-674-86530-8.

McDonald, H. G. (2005). "Palecology of extinct xenarthrans and the Great American Biotic Interchange". In: *Bulletin of the Florida Museum of Natural History* 45.4, pp. 319–340.

McDonough, M. M. et al. (2022). "Phylogenomic systematics of the spotted skunks (Carnivora, Mephitidae, Spilogale): additional species diversity and Pleistocene climate change as a major driver of diversification". In: *Molecular Phylogenetics and Evolution* 167. ISBN: 1055-7903 Publisher: Elsevier, p. 107266.

McKenna, M. C. and S. K. Bell (1997). *Classification of mammals: above the species level*. Columbia University Press. ISBN: 0-231-52853-1.

Meseguer, A. S. and F. L. Condamine (2017). "Ancient tropical extinctions contributed to the latitudinal diversity gradient". In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 236646.

Miranda, F. R. et al. (2018). "Taxonomic review of the genus Cyclopes Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species". In: *Zoological Journal of the Linnean Society* 183.3, pp. 687–721. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlx079.

Miranda, F. R. et al. (2023). "Taxonomic revision of maned sloths, subgenus Bradypus (Scaeopus), Pilosa, Bradypodidae, with revalidation of Bradypus crinitus Gray, 1850". In: *Journal of Mammalogy* 104.1. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 86–103.

Mishler, B. D. (2000). "The phylogenetic species concept (sensu Mishler and Theriot): monophyly, apomorphy, and phylogenetic species concepts". In: *Species concepts and phylogenetic theory: a debate. Columbia University Press, New York*, pp. 44–54.

Möller-Krull, M. et al. (2007). "Retroposed Elements and Their Flanking Regions Resolve the Evolutionary History of Xenarthran Mammals (Armadillos, Anteaters, and Sloths)". In: *Molecular Biology and Evolution* 24.11, pp. 2573–2582. ISSN: 0737-4038. DOI: 10.1093/molbev/msm201.

Mora, C. et al. (2011). "How many species are there on Earth and in the ocean?" In: *PLoS biology* 9.8. ISBN: 1545-7885 Publisher: Public Library of Science, e1001127.

Moritz, C. and C. Cicero (2004). "DNA barcoding: promise and pitfalls". In: *PLoS biology* 2.10. ISBN: 1545-7885 Publisher: Public Library of Science San Francisco, USA, e354.

Muller, H. J. (1942). "Isolating mechanisms, evolution, and temperature." In: *Biol. Symp.* Vol. 6, p. 71.

Murphy, W. J. et al. (2001a). "Molecular phylogenetics and the origins of placental mammals". In: *Nature* 409.6820. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6820 Primary_atype: Research Publisher: Nature Publishing Group, pp. 614–618. ISSN: 1476-4687. DOI: 10.1038/35054550.

Murphy, W. J. et al. (2001b). "Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics". In: *Science* 294.5550. Publisher: American Association for the

Advancement of Science Section: Report, pp. 2348–2351. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1067179.

Naples, V. L. (1999). "Morphology, evolution and function of feeding in the giant anteater (Myrmecophaga tridactyla)". In: *Journal of Zoology* 249.1. ISBN: 1469-7998 Publisher: Cambridge University Press, pp. 19–41.

Nei, M. and W.-H. Li (1973). "Linkage disequilibrium in subdivided populations". In: *Genetics* 75.1. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 213–219.

Neigel, J. E. (1985). "Phylogenetic relationships of mitochondrial DNA under various demographic models of speciations." In: *Evolutionary processes and theory*. Publisher: Academic Press, pp. 515–534.

Nielsen, R. (2005). "Molecular signatures of natural selection". In: *Annu. Rev. Genet.* 39. ISBN: 0066-4197 Publisher: Annual Reviews, pp. 197–218.

Noor, M. A. and S. M. Bennett (2009). "Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species". In: *Heredity* 103.6. ISBN: 1365-2540 Publisher: Nature Publishing Group, pp. 439–444.

Nosenko, T. et al. (2013). "Deep metazoan phylogeny: when different genes tell different stories". In: *Molecular phylogenetics and evolution* 67.1. ISBN: 1055-7903 Publisher: Elsevier, pp. 223–233.

Nosil, P. (2012). *Ecological speciation*. Oxford Ecology and Evolution. ISBN: 0-19-958711-6.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos (2009a). "Divergent selection and heterogeneous genomic divergence". In: *Molecular ecology* 18.3. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 375–402.

Nosil, P., L. J. Harmon, and O. Seehausen (2009b). "Ecological explanations for (incomplete) speciation". In: *Trends in ecology & evolution* 24.3. ISBN: 0169-5347 Publisher: Elsevier, pp. 145–156.

Nosil, P. and D. Schluter (2011). "The genes underlying the process of speciation". In: *Trends in ecology & evolution* 26.4. Publisher: Elsevier, pp. 160–167. ISSN: 0169-5347.

Novacek, M. J. (1992). "Mammalian phytogeny: shaking the tree". In: *Nature* 356.6365. ISBN: 0028-0836 Publisher: Nature Publishing Group UK London, pp. 121–125.

Ohta, T. (1992). "The nearly neutral theory of molecular evolution". In: *Annual review of ecology and systematics* 23.1. ISBN: 0066-4162 Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 263–286.

Orr, H. A. (1995). "The population genetics of speciation: the evolution of hybrid incompatibilities." In: *Genetics* 139.4. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 1805–1813.

Padial, J. M. et al. (2010). "The integrative future of taxonomy". In: *Frontiers in Zoology* 7.1. ISBN: 1742-9994 Publisher: BioMed Central, pp. 1–14.

Pamilo, P. and M. Nei (1988). "Relationships between gene trees and species trees." In: *Molecular biology and evolution* 5.5. ISBN: 1537-1719, pp. 568–583.

Patterson, B. and R. Pascual (1968). "The fossil mammal fauna of South America". In: *The Quarterly Review of Biology* 43.4. ISBN: 0033-5770 Publisher: Stony Brook Foundation, Inc., pp. 409–451.

Patterson, B., W. Segall, and W. D. Turnbull (1989). "ear region in Xenarthrans (= Edentata: Mammalia)". In: Publisher: Field Museum of Natural History.

Payseur, B. A. (2010). "Using differential introgression in hybrid zones to identify genomic regions involved in speciation". In: *Molecular ecology resources* 10.5. Publisher: Wiley Online Library, pp. 806–820. ISSN: 1755-098X.

Peres, P. H. et al. (2021). "Revalidation of Mazama rufa (Illiger 1815)(Artiodactyla: Cervidae) as a distinct species out of the complex Mazama americana (Erxleben 1777)". In: *Frontiers in genetics* 12. ISBN: 1664-8021 Publisher: Frontiers Media SA, p. 742870.

Pons, J. et al. (2006). "Sequence-based species delimitation for the DNA taxonomy of undescribed insects". In: *Systematic Biology* 55.4. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 595–609.

Presgraves, D. C. (2010). "Speciation genetics: search for the missing snowball". In: *Current biology* 20.24. ISBN: 0960-9822 Publisher: Elsevier, R1073–R1074.

Pritchard, J. K., M. Stephens, and P. Donnelly (2000). "Inference of population structure using multilocus genotype data". In: *Genetics* 155.2. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 945–959.

Puillandre, N. et al. (2012). "ABGD, Automatic Barcode Gap Discovery for primary species delimitation". In: *Molecular Ecology* 21.8, pp. 1864–1877. ISSN: 1365-294X. DOI: 10.1111/j.1365-294X.2011.05239.x.

Queiroz, K. de (2007). "Species concepts and species delimitation". In: *Systematic Biology* 56.6. Publisher: Society of Systematic Zoology, pp. 879–886. ISSN: 1076-836X.

Ravinet, M. et al. (2017). "Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow". In: *Journal of evolutionary biology* 30.8. Publisher: Wiley Online Library, pp. 1450–1477. ISSN: 1010-061X.

Redondo, R. A. F. et al. (2008). "Molecular systematics of the genus Artibeus (Chiroptera: Phyllostomidae)". In: *Molecular Phylogenetics and Evolution* 49.1, pp. 44–58. ISSN: 1055-7903. DOI: 10.1016/j.ympev.2008.07.001.

Reiss, K. Z. (1997). "Myology of the feeding apparatus of myrmecophagid anteaters (Xenarthra: Myrmecophagidae)". In: *Journal of Mammalian Evolution* 4. ISBN: 1064-7554 Publisher: Springer, pp. 87–117.

Rieseberg, L. H., J. Whitton, and K. Gardner (1999). "Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species". In: *Genetics* 152.2. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 713–727.

Rosel, P. E. et al. (2017). "Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences". In: *Marine Mammal Science* 33 (S1). Publisher: Wiley Online Library, pp. 76–100. ISSN: 0824-0469.

Rosen, D. E. (1979). "Fishes from the uplands and intermontane basins of Guatemala: revisionary studies and comparative geography. Bulletin of the AMNH; v. 162, article 5". In: Publisher: New York: American Museum of Natural History.

Roux, C. et al. (2016). "Shedding light on the grey zone of speciation along a continuum of genomic divergence". In: *PLoS Biology* 14.12. Publisher: Public Library of Science San Francisco, CA USA, e2000234. ISSN: 1544-9173.

Ruiz-García, M. et al. (2020). "Molecular phylogenetics of Bradypus (three-toed sloth, Pilosa: Bradypodidae, Mammalia) and phylogeography of Bradypus variegatus (brown-

throated three-toed sloth) with mitochondrial gene sequences". In: *Journal of Mammalian Evolution* 27.3. Publisher: Springer, pp. 461–482. ISSN: 1573-7055.

Seehausen, O. L. E. et al. (2008). "Speciation reversal and biodiversity dynamics with hybridization in changing environments". In: *Molecular ecology* 17.1. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 30–44.

Seehausen, O. et al. (2014). "Genomics and the origin of species". In: *Nature Reviews Genetics* 15.3. ISBN: 1471-0056 Publisher: Nature Publishing Group UK London, pp. 176–192.

Servedio, M. R. and M. A. Noor (2003). "The role of reinforcement in speciation: theory and data". In: *Annual Review of Ecology, Evolution, and Systematics* 34.1. ISBN: 1543-592X Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 339–364.

Shoshani, J. and M. C. McKenna (1998). "Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data". In: *Molecular phylogenetics and evolution* 9.3. ISBN: 1055-7903 Publisher: Elsevier, pp. 572–584.

Simpson, G. G. (1961). *Principles of animal taxonomy*. Columbia University Press. ISBN: 0-231-88859-7.

Singh, B. N. (2022). "Reproductive isolating mechanisms and speciation". In: *Journal of Scientific Research* 66.1.

Slade, D. D. (1891). "On the genus Chlamydophorus". In: *The American Naturalist* 25.294. ISBN: 0003-0147 Publisher: Ferris Bros., pp. 540–548.

Smadja, C., J. Galindo, and R. Butlin (2008). "Hitching a lift on the road to speciation". In: *Molecular Ecology* 17.19. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 4177–4180.

Smith, M. L. and B. C. Carstens (2022). "Species delimitation using molecular data". In: *Species problems and beyond*. CRC Press, pp. 145–160.

Smith, P. (2008). "Chaco Fairy Armadillo Calyptophractus retusus". In: *FAUNA Paraguay Handbook of the Mammals of Paraguay* 20, pp. 1–5.

Smith, P. and R. D. Owen (2017). "Calyptophractus retusus (Cingulata: Dasypodidae)". In: *Mammalian Species* 49.947. ISBN: 0076-3519 Publisher: Oxford University Press US, pp. 57–62.

Smith, V. S. (2005). "DNA barcoding: Perspectives from a "Partnerships for Enhancing Expertise in Taxonomy"(PEET) debate". In: *Systematic Biology* 54.5. ISBN: 1076-836X Publisher: Taylor & Francis, pp. 841–844.

Stankowski, S. and M. Ravinet (2021a). "Defining the speciation continuum". In: *Evolution* 75.6. ISBN: 0014-3820 Publisher: Wiley Online Library, pp. 1256–1273.

— (2021b). "Quantifying the use of species concepts". In: *Current Biology* 31.9. ISBN: 0960-9822 Publisher: Elsevier, R428–R429.

Stankowski, S. et al. (2023). "Selection on many loci drove the origin and spread of a key innovation". In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 2023.02. 13.528213.

Stevenson, I. P. (1947). "John Ray and his contributions to plant and animal classification". In: *Journal of the History of Medicine and Allied Sciences* 2.2. ISBN: 1468-4373 Publisher: Oxford University Press, pp. 250–261.

Subbotin, S. A. and M. Moens (2006). "Molecular taxonomy and phylogeny." In: *Plant Nematology*. CAB International Wallingford UK, pp. 399–418.

Sukumaran, J., M. T. Holder, and L. L. Knowles (2021). "Incorporating the speciation process into species delimitation". In: *PLoS Computational Biology* 17.5. ISBN: 1553-734X Publisher: Public Library of Science San Francisco, CA USA, e1008924.

Sukumaran, J. and L. L. Knowles (2017). "Multispecies coalescent delimits structure, not species". In: *Proceedings of the National Academy of Sciences* 114.7. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 1607–1612.

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations". In: *Genetics* 105.2. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 437–460.

Takahata, N. and M. Nei (1985). "Gene genealogy and variance of interpopulational nucleotide differences". In: *Genetics* 110.2. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 325–344.

Tricou, T., E. Tannier, and D. M. de Vienne (2022a). "Ghost lineages can invalidate or even reverse findings regarding gene flow". In: *PLoS Biology* 20.9. ISBN: 1544-9173 Publisher: Public Library of Science San Francisco, CA USA, e3001776.

Tricou, T., E. Tannier, and D. M. de Vienne (2022b). "Ghost lineages highly influence the interpretation of introgression tests". In: *Systematic Biology* 71.5. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 1147–1158.

Turner, T. L. and M. W. Hahn (2010). "Genomic islands of speciation or genomic islands and speciation?" In: *Molecular ecology* 19.5. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 848–850.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin (2005). "Genomic islands of speciation in Anopheles gambiae". In: *PLoS biology* 3.9. ISBN: 1544-9173 Publisher: Public Library of Science San Francisco, USA, e285.

Van Valen, L. (1976). "Ecological species, multispecies, and oaks". In: *Taxon*. ISBN: 0040-0262 Publisher: JSTOR, pp. 233–239.

Via, S. (2012). "Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1587. ISBN: 0962-8436 Publisher: The Royal Society, pp. 451–460.

Via, S. and J. West (2008). "The genetic mosaic suggests a new role for hitchhiking in ecological speciation". In: *Molecular ecology* 17.19. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 4334–4345.

Vizcaíno, S. F. and M. S. Bargo (2014). "Loss of ancient diversity of xenarthrans and the value of protecting extant armadillos, sloths and anteaters". In: *Edentata* 15.2014. ISBN: 1413-4411 Publisher: BioOne, pp. 27–38.

Wayne, L. G. et al. (1987). "Report of the ad hoc committee on reconciliation of approaches to bacterial systematics". In: *International Journal of Systematic and Evolutionary Microbiology* 37.4. ISBN: 1466-5026 Publisher: Microbiology Society, pp. 463–464.

Weir, J. T. and M. Price (2011). "Andean uplift promotes lowland speciation through vicariance and dispersal in Dendrocincla woodcreepers". In: *Molecular Ecology* 20.21. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-294X.2011.05294.x, pp. 4550–4563. ISSN: 1365-294X. DOI: 10.1111/j.1365-294X.2011.05294.x.

Westram, A. M. et al. (2022). "What is reproductive isolation?" In: *Journal of evolutionary biology* 35.9. ISBN: 1010-061X Publisher: Wiley Online Library, pp. 1143–1164.

Wetzel, R. M. et al. (2008). "Order Cingulata". In: *Mammals of South America* 1. Publisher: Chicago University of Chicago Press, pp. 128–156.

Wetzel, R. M. (1985). "The identification and distribution of recent Xenarthra (= Edentata)". In: *The evolution and ecology of armadillos, sloths, and vermilinguas*. Publisher: Smithsonian Institution Press Washington, DC, pp. 5–21.

Wetzel, R. M. and F. d. Avila-Pires (1980). "Identification and distribution of the recent sloths of Brazil (Edentata)". In: *Revista Brasileira de Biologia* 40.4, pp. 831–836.

Will, K. W., B. D. Mishler, and Q. D. Wheeler (2005). "The perils of DNA barcoding and the need for integrative taxonomy". In: *Systematic biology* 54.5. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 844–851.

Wilson, E. O. (2010). "The major historical trends of biodiversity studies". In: *Systema Naturae 250—The Linnaean Ark*.

Wood, T. C. (2008). *SPECIES VARIABILITY AND CREATIONISM*. Center for Origins Research Bryan College. Dayton, Tennessee.

Wright, S. (1931). "Evolution in Mendelian populations". In: *Genetics* 16.2. Publisher: Oxford University Press, p. 97.

Wu, C.-I. (2001). "The genic view of the process of speciation". In: *Journal of evolutionary biology* 14.6. Publisher: Wiley Online Library, pp. 851–865. ISSN: 1010-061X.

Yang, Z. and B. Rannala (2010). "Bayesian species delimitation using multilocus sequence data". In: *Proceedings of the National Academy of Sciences* 107.20. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 9264–9269.

Yule, G. U. (1925). "II.—A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S". In: *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character* 213.402. ISBN: 0264-3960 Publisher: The Royal Society London, pp. 21–87.

Zachos, F. E. (2018). "Mammals and meaningful taxonomic units: the debate about species concepts and conservation". In: *Mammal Review* 48.3. Publisher: Wiley Online Library, pp. 153–159. ISSN: 0305-1838.

— (2016). *Species concepts in biology*. Vol. 801. Springer.

Zhang, J. et al. (2013). "A general species delimitation method with applications to phylogenetic placements". In: *Bioinformatics* 29.22. ISBN: 1460-2059 Publisher: Oxford University Press, pp. 2869–2876.

# CHAPTER I - EXON CAPTURE MUSEOMICS DECIPHERS THE NINE-BANDED ARMADILLO SPECIES COMPLEX AND IDENTIFIES A NEW SPECIES ENDEMIC TO THE GUIANA SHIELD

Mathilde Barthe[1,*], Loïs Rancilhac[1,2,3], Maria C. Arteaga[4], Anderson Feijó[5,6], Marie-Ka Tilak[1], Fabienne Justy[1], W. J. Loughry[7], Colleen M. McDonough[7], Benoit de Thoisy[8,9], François Catzeflis[1,†], Guillaume Billet[10], Lionel Hautier[1,11], Benoit Nabholz[1,12], and Frédéric Delsuc[1,*]

[1]1Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ. Montpellier, CNRS, IRD, Montpellier, France

[2]Animal ecology, Department of Ecology and Genetics, Uppsala University, Sweden

[3]Department of biology, University of Cyprus, Nicosia, Cyprus

[4]Department of Conservation Biology, CICESE, Ensenada, Baja California, México

[5]Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, IL, USA

[6]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[7]Department of Biology, Valdosta State University, Valdosta, GA, USA

[8]Institut Pasteur de la Guyane, Cayenne, French Guiana

[9]Kwata NGO, Cayenne, French Guiana

[10]Centre de Recherche en Paléontologie – Paris (CR2P), CNRS/MNHN/Sorbonne Université, Muséum national d'Histoire naturelle, Paris, France

[11]Mammal Section, Life Sciences, Vertebrate Division, The Natural History Museum, London, UK

[12]Institut universitaire de France, Paris, France

[†]Deceased

Nomenclatural statement: A Life Science Identifier (LSID) number was obtained for this publication: urn:lsid:zoobank.org:pub:F079EE52-4134-409F-90D1-DC77FFB7F54B

[*]Corresponding authors: mathilde.barthe.pro@gmail.com; frederic.delsuc@umontpellier.fr

## Abstract

The nine-banded armadillo (*Dasypus novemcinctus*) is the most widespread xenarthran species across the Americas. Recent studies have suggested it is composed of four morphologically and genetically distinct lineages of uncertain taxonomic status. To address this issue, we used a museomic approach to sequence 80 complete mitogenomes and capture 997 nuclear loci for 71 *Dasypus* individuals sampled across the entire distribution. We carefully cleaned up potential genotyping errors and cross contaminations that could blur species boundaries by mimicking gene flow. Our results unambiguously support four distinct lineages within the *D. novemcinctus* complex. We found cases of mito-nuclear phylogenetic discordance but only limited contemporary gene flow confined to the margins of the lineage distributions. All available evidence including the restricted gene flow, phylogenetic reconstructions based on both mitogenomes and nuclear loci, and phylogenetic delimitation methods consistently supported the four lineages within *D. novemcinctus* as four distinct species. Comparable genetic differentiation values to other recognized *Dasypus* species further reinforced their status as valid species. Considering congruent morphological results from previous studies, we provide an integrative taxonomic view to recognise four species within the *D. novemcinctus* complex: *D. novemcinctus*, *D. fenestratus*, *D. mexicanus*, and *D. guianensis* sp. nov., a new species endemic of the Guiana Shield that we describe here. The two available individuals of *D. mazzai* and *D. sabanicola* were consistently nested within *D. novemcinctus* lineage and their status remains to be assessed. The present work offers a case study illustrating the power of museomics to reveal cryptic species diversity within a widely distributed and emblematic species of mammals.


**Keywords**: Species delimitation, Multilocus phylogeny, Integrative taxonomy, Museomics, Phylogeography, Mito-nuclear discordance.

## Introduction

Species represent the basic units of taxonomy and are the quintessential objects of evolutionary biology, but their delineation is still problematic. Such difficulties mainly stem from the fact that taxonomy defines discrete boundaries, while speciation is continuous (Dres and Mallet 2002; Mallet et al. 2007; Stankowski and Ravinet 2021). This

process corresponds to the transition from free gene exchange within populations to the cessation of gene flow (Mayr 1942; Dobzhansky 1982; complete reproductive isolation; Coyne and Orr 2004; Galtier 2019). When the speciation process is complete, the taxonomic status as distinct species becomes unambiguous. However, species delimitation becomes less clear cut during this differentiation process. The use of different delimitation criteria associated with different species concepts can lead to inconsistencies between studies, as incipient species will not always evolve the same characteristics at the same stages of speciation (de Queiroz 2007). In order to overcome such conceptual flaws, de Queiroz (2007) proposed a unified species concept, the generalized lineage concept (GLC), in which species represent metapopulation lineages that evolve independently from one another. While such a definition provided a significant conceptual advance, it is not operational for species delimitation (Zachos and Asher 2018; Kollár et al. 2022). It has, however, highlighted the methodological difficulties of delineating species within a speciation continuum (Zachos and Asher 2018). The analysis of molecular data in statistical frameworks has improved the reproducibility of analyses and the transparency of species delimitation protocols, allowing quantification of the degree of support for different taxonomic hypotheses. These methods have been widely used for species delimitation and have provided valuable insights into species boundaries, particularly when used in combination with morphological and ecological evidence in an integrative approach (Padial et al. 2010; Nascimento et al. 2021).

Nevertheless, molecular methods should be carefully considered because the use of ever-increasing amounts of data may lead to over-splitting. By identifying genetic structuring at a finer resolution, these methods can lead to the delineation of the smallest identifiable unit (Funk et al. 2012; Sukumaran and Knowles 2017; Leaché et al. 2019; Sukumaran et al. 2021), while statistical significance can be more easily achieved when many loci are analyzed (Leaché et al. 2019). For these reasons, Carstens et al. (2013) recommended relying on concordance across various species delimitation algorithms depending on different assumptions. Using a comparative approach could also constitute an alternative way to confirm genetically-defined candidate species in a homogeneous way, and thus limit under- or over-splitting. It has thus been proposed that genomic differentiation of taxa with a consensus taxonomic rank can be used as a reference to delineate lineages with comparable genomic differentiation values (Hey and Pinho 2012; Roux et al. 2016; Riesch et al. 2017; Rosel et al. 2017; Galtier 2019). This genome-wide comparative approach has proven useful for

delimiting species of aardwolves for instance (Allio et al. 2021). Finally, genomic data provide an unprecedented opportunity to understand the speciation process itself by investigating the phylogeographic history of the focal lineages, and by doing so, define species boundaries more precisely. Comprehensive analyses of many unlinked molecular markers can yield information on the levels of past and contemporary gene flow, hereby providing insight on the levels of reproductive isolation, especially in the context of secondary contact zones (Dufresnes et al. 2015).

The growing use of integrative and comparative approaches has reshuffled the classification of several mammalian groups once considered taxonomically well-known. In the Neotropics alone about 100 species of medium and large mammals have been recognized in the last 15 years (Burgin et al. 2018; Feijó and Brandão 2022), including new species of anteaters (Miranda et al. 2018), olingos (Helgen et al. 2013), porcupines (Pontes et al. 2013), rabbits (Ruedas 2017), and primates (Boubli et al. 2019; Costa-Araújo et al. 2019, 2021; Gusmão et al. 2019). Among those, xenarthrans represent an emblematic group whose diversity has just begun to be recognized in the past decade. They are one of the major lineages of placental mammals and are currently represented by 39 species of anteaters, sloths, and armadillos (Gibb et al. 2016; Feijó in press). Long-nosed armadillos of the genus *Dasypus* are the most speciose group of xenarthrans, occurring throughout most of the Americas, and have a complex taxonomic history with the recent recognition of three distinct species within the greater long-nosed armadillo (*Dasypus kappleri*) for instance (Feijó et al. 2018). Presently, eight extant *Dasypus* species are recognized but available evidence suggests cryptic diversity, particularly within the nine-banded armadillo (*Dasypus novemcinctus*), a species with a Panamerican distribution spanning from northern Argentina to the United States (Feijó et al. 2018).

Indeed, an early molecular study revealed a high genetic divergence between the invasive US populations and populations from French Guiana (Huchon et al. 1999). Subsequent studies have identified four distinct lineages within this species including a potentially undescribed species restricted to the Guiana Shield (hereafter Guianan lineage; Hautier et al. 2017), as originally proposed by Gibb et al. (2016). Morphological studies based on the structure of paranasal sinuses (Billet et al. 2017), geometric morphometrics of skull shape (Hautier et al. 2017), and whole specimen anatomy (Feijó et al. 2018), as well as molecular studies based on a few markers (Feijó et al. 2019; Arteaga et al. 2020) have confirmed the distinctiveness of this Guianan lineage. Another

lineage with both characteristic skull shape (Hautier et al. 2017) and paranasal sinuses (Billet et al. 2017) includes armadillos distributed throughout South America on the eastern side of the Andes (hereafter Southern lineage; Hautier et al. 2017). Subsequent molecular data have shown that this lineage might also include *D. mazzai* and *D. sabanicola* (Feijó et al. 2019), although it was previously suggested that the latter two should be maintained as distinct species given their morphological differences (Feijó et al. 2018). Two additional lineages distributed (i) on the western side of the Andes in South America and southern Central America (hereafter Central lineage; Hautier et al. 2017) and (ii) in northern Central America and the US (hereafter Northern lineage; Hautier et al. 2017) have been further highlighted based on mitochondrial control region and microsatellite analyses (Arteaga et al. 2011, 2012). These two lineages were again confirmed by geometric morphometrics of skull shape (Hautier et al. 2017) and, to a lesser extent, the structure of paranasal sinuses (Billet et al. 2017). Up to now, however, the taxonomic status of these four lineages, as well as their phylogenetic relationships with *D. pilosus*, *D. mazzai*, and *D. sabanicola*, remain uncertain. In fact, *D. pilosus* was found to be nested within the *D. novemcinctus* complex despite its uniquely hairy carapace and particularly elongated skull (Gibb et al. 2016; Feijó et al. 2019). A better understanding of the evolutionary history of the *D. novemcinctus* complex has broader implications as splitting a widespread species into potentially more geographically restricted species would necessitate reevaluating the conservation status of each (Abba & Superina, 2010; Feijó & Brandão 2022; Superina et al., 2014).

In addition to the taxonomic conundrum involving *D. novemcinctus* and related species, its evolutionary history remains controversial. Based on a short fragment of the mitochondrial control region, Arteaga et al. (2020) identified individuals belonging to the Northern lineage overlapping with the Southern lineage in most of South America. This result contrasts with morphological studies supporting a clearly defined and homogenous Southern morpho-group (Billet et al. 2017; Hautier et al. 2017) and the molecular study of Feijó et al. (2019) that found Mexican nine-banded armadillos clustering with populations from the United States and clearly diverging from South American ones. In addition, a population in western Mexico (thereafter western Mexico population) was identified as part of the Central lineage by Arteaga et al. (2011, 2012, 2020), while morphological studies identified a unique morpho-group within the Northern range (Billet et al. 2017; Hautier et al. 2017). Both ancient introgression and contemporary gene flow between these parapatric lineages could result in mito-nuclear discor-

dance (Toews and Brelsford 2012) and heterogeneity across nuclear gene trees, as previously suggested by microsatellite analyses (Arteaga et al. 2011, 2012). These heterogeneities might lead to incongruences across phylogenetic analyses, depending on the history of the markers analyzed, and in turn cause species boundaries defined based on molecular data to disagree with those inferred from other lines of evidence (i.e. morphology) (Degnan & Rosenberg, 2009; de Queiroz, 2007; Toews & Brelsford, 2012; Wiens et al., 2010). Furthermore, identifying contemporary gene flow is important to define the taxonomic status of evolutionary lineages, as it reflects their levels of reproductive isolation. Better understanding genealogical discrepancies, and separating historical and contemporary introgression typically requires the analysis of multiple loci.

To disentangle the *D. novemcinctus* species complex, we relied on a museomic approach (Raxworthy and Smith 2021) to ensure a geographically and taxonomically representative sample. We used shotgun sequencing and exon capture to obtain the complete mitogenome as well as 997 nuclear loci for respectively 80 and 71 individuals including 48 museum specimens. In a context where species concepts and the resulting delimitation methods are still controversial, we referred to the GLC of species (de Queiroz 2007) and considered taxonomic status as a hypothesis to be explored through various analyses (Carstens et al. 2013; Zachos and Asher 2018). We thus considered the four morphotypes of *D. novemcinctus* identified by Hautier et al. (2017) as our main species delimitation hypothesis. Because our sampling includes numerous museum specimens with expected low contents of endogenous DNA, we also carefully assessed the effect of sequencing errors, genotyping errors, and cross-contaminations that could both blur phylogenetic signal and mimic gene flow (Pompanon et al., 2005; Raxworthy & Smith, 2021; Robledo-Arnuncio & Gaggiotti, 2017). Indeed, these different types of errors are common, and despite increasing awareness, relatively few studies have focused on this problem and its consequences for subsequent analyses (Bonin et al. 2004; Pompanon et al. 2005; Mastretta-Yanes et al. 2015; Robledo-Arnuncio and Gaggiotti 2017). Incorrect genotype calls are expected to have a significant impact on population genetic analyses, in particular by leading to inaccurate conclusions about genetic diversity and population structure (Bonin et al. 2004; Herrmann et al. 2010; Zhang and Hare 2012; Petrou et al. 2019). Some of these errors, such as cross-contamination, could even mimic gene flow (Robledo-Arnuncio & Gaggiotti, 2017; Petrou et al., 2019), which is particularly problematic for species delimitation. Appropriate data cleaning should improve species delineation but disentangling

errors from biological signals of introgression remains a challenge. For this reason, we performed thorough cross-contamination exploration and carefully cleaned reads from sequencing and contamination errors.

# Materials & Methods

## Biological sampling

A total of 80 armadillo tissue samples were obtained over the years for this study through multiple sources (Table S1). Our sampling notably includes 46 individuals sampled in the form of dried skin pieces, of which 38 were obtained from museum specimens collected between 1894 and 2000, and stored in 12 different collections. The remaining samples were fresh tissue biopsies. In accordance with the policy of sharing benefits and advantages (APA TREL1916196S/224), biological material from French Guiana has been deposited in the JAGUARS collection supported by the Collectivité Territoriale de Guyane and the Direction Générale des Territoires et de la Mer Guyane.

## DNA extractions and sequencing

Total genomic DNA extractions from fresh tissue biopsies preserved in 95% ethanol were performed using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions. Museum dried skin samples were processed sequentially under a dedicated UV hood, cleaned between samples to minimize cross-contamination. DNA extractions were then performed in batches of 12 samples including an extraction blank, using the same DNeasy Blood & Tissue Kit under a UV hood in a dedicated clean room. The only minor modification was to reduce the final elution volume to 70 $\mu$l instead of 100 $\mu$l. Illumina libraries were constructed from DNA extracts according to the cost-effective version of the Meyer and Kircher (2010) protocol proposed by Tilak et al. (2015). Mitogenomes were obtained through shotgun Illumina sequencing, using single-end 100 bp reads for the fresh samples (sequenced at GATC Biotech, Konstanz, Germany) and paired-end 150 bp reads for the museum specimens (sequenced by Daicel Arbor Biosciences, Ann Arbor, MI, USA). To obtain complementary nuclear data, single-copy orthologous exons shared by the nine-banded armadillo (*D. novemcinctus*) and the Hoffmann's two-toed sloth (*Choloepus hoffmanni*) were selected from the OrthoMam v8 database (Douzery et al. 2014) based on their size, that was set to be around 200 bp, until a total of 1000 exons

was reached. The sequences of these 1000 exons plus 400 bp of flanking introns (200 bp on each side) were then extracted from the Dasnov3.1 nine-banded armadillo genome assembly (GCF_000208655.2) and used to design RNA capture probes resulting in a final set of 16,146 probes targeting 997 nuclear loci of about 600 bp length each. We verified that these 997 loci provide a representative sampling of the nine-banded armadillo genome by locating them on the latest chromosome-scale reference assembly (GCF_030445035.1) (Fig. S1). Probe design and synthesis, library preparation on DNA extracts from museum specimens, capture reactions on previously constructed libraries, and Illumina sequencing of the 997 nuclear loci were outsourced to Daicel Arbor Biosciences.

## Dataset assembly

*Mitochondrial dataset.* The raw reads obtained by shotgun sequencing were cleaned with FastP v0.21.0 (Chen et al. 2018). The reference mitogenome of *Dasypus novemcinctus* (NC_001821.1; Arnason et al. 1997) was used to map the reads of each individual using bwa mem v.0.7.17 (Li 2013) with default parameters. Samtools v1.9 (Li et al. 2009) and Picard v2.25.5 (Picard Toolkit 2019) were respectively used to convert the mapping files and to order and index reads according to their position on the reference genome. MarkDuplicates v2.25.5 (Picard Toolkit 2019) was used to mark duplicate reads. Samtools depth v1.9 (Li et al. 2009) was used to estimate depth coverage (Fig. S2). Variant calling for haploid data (–ploidy 1) was then performed with Freebayes v1.3.1 (Garrison and Marth 2012). Finally, the vcf file was converted with bcftools consensus v1.14 (Danecek et al. 2021) using haploid parameters and filtering out positions with less than 5x depth coverage to obtain sequences in fasta format. Furthermore, we assembled two additional mitochondrial datasets of shorter fragments including samples from previous studies on 16S rRNA (Abba et al. 2018) and D-loop (Arteaga et al. 2020) to evaluate the consistency between these sequences and our new mitogenomes (see Supplementary Results for more information).

*Contamination exploration.* We implemented a method inspired by Green et al. (2008, 2010) for detecting human contamination in Neanderthal genomes. To identify cross-contamination between lineages, we identified the mitochondrial diagnostic positions (DPs) for each of the four lineages of the species complex and *D. pilosus* using the apolist command of PAUP v4.0a (Swofford 1998). A total of 122 DPs were identified for the Guianan lineage, 31 for the Southern lineage, 26 for the Central lineage, 28 for the Northern lineage, and 143 for *D. pilosus*. Then, for each individual, we recorded

the proportion of the lineage DPs supported by at least three reads (Fig. S3). The frequency of reads supporting these positions (i.e., the number of reads with the diagnostic allele divided by the total number of reads mapping at the position) was also taken into account (Fig. S3). Proportions of DPs and read frequencies were estimated using bam-readcount (https://github.com/genome/bam-readcount) and the custom script Estimate_lineage_support.py (https://github.com/Mathilde-Barthe/Mitochondrial_ support). The support for a specific lineage was then estimated as the product between the proportion of DPs supporting this lineage and their read frequency (Lineage support p̄roportion of DPs x read frequency; Fig. S3) and plotted using the custom script Plot_lineage_support.R (https://github.com/ Mathilde-Barthe/ Mitochondrial_support) and the ggplot2, dplyr, colorspace, cowplot, grid, gridExtra, ggpubr R packages (Auguie and Antonov 2017; R Core Team 2018; Villanueva and Chen 2019; Wilke 2019; Zeileis et al. 2019; Kassambara 2020; Wickham et al. 2023). DPs of *D. pilosus* were used as controls as laboratory experiments (DNA extractions and library preparations) for these two individuals were performed separately from those of the *D. novemcinctus* complex. Thus, their proportion within *D. novemcinctus* samples represents shared genotyping errors expected between individuals that did not cross-contaminated each other.

*Nuclear dataset.* Reads from exon capture sequencing were cleaned with FastP v0.21.0 (Chen et al. 2018). The 997 targeted sequences used to define the capture probes were used as references to map the reads of each individual using bwa mem v.0.7.17 (Li 2013) with default parameters for paired-end data. As for the mitogenomes, we used Samtools v1.9 (Li et al. 2009), Picard v2.25.5 (Picard Toolkit 2019) and MarkDuplicates v2.25.5 (Picard Toolkit 2019) to respectively convert the mapping files, order and index reads according to their position on the reference genome, and mark duplicate reads. Variant calling for diploid data was then performed with Freebayes v1.3.1 (Garrison and Marth 2012). Positions with a coverage under 10x and a quality score below 20 were considered ambiguous and called as 'N'. The vcf file was modified with a custom script to ensure that heterozygous positions were supported by reads at frequency 0.3 to 0.7, otherwise, the most frequent allele was called as a homozygous position. Depth of coverage was estimated with samtools depth v1.9 (Li et al. 2009). We excluded from the analysis 3,541 sequences (57 loci per individual on average) with more than 40% missing data (Table S2) and nine individuals with more than 55% average missing data across loci (see Table S1 for details). We also excluded one locus containing less than three

individuals as this is the minimal number required to reconstruct a phylogenetic tree. Note however that only seven loci contained less than 10 individuals. Next, we used the "–hardy" option from vcftools to filter out 159 loci in which at least 75% of the individuals in a lineage were heterozygous at a single position. Inbreeding coefficient (F) and heterozygosity (He) were estimated with the "–het" option from vcftools and a custom script, respectively. Using these cleaned vcf files, 25,742 variant positions out of a total of 506,355 nucleotide positions across the 837 retained loci were reconstructed in fasta format using the VCF2FastaFreebayes custom script (https://github.com/ benoitnabholz/VCF2Fasta). Finally, bcftools consensus v1.14 was used to obtain the fasta sequence, with heterozygous sites encoded as IUPAC characters.

## Phylogenetic inferences

The reconstructed mitogenomes were aligned with mafft v7.310 (Katoh and Standley 2013) with default parameters. Mitochondrial CDSs and rRNAs were concatenated with AMAS (Borowiec 2016) and sites with more than 50% missing data were removed with trimal v1.4 (Capella-Gutiérrez et al. 2009), resulting in 13,924 nucleotide positions. The mitogenomic tree was reconstructed by maximum likelihood (ML) using IQ-TREE v2.1.4 (Minh et al. 2020b) with a partitioned model ("-spp" option) and using ModelFinder ("-m TESTNEW" option; Kalyaanamoorthy et al. 2017) to select the best-fitting model. The same methodology was applied to reconstruct ML phylogenies for the two short mitochondrial fragment datasets of 16S rRNA (212 bp; Abba et al. 2018) and D-loop (387 bp; Arteaga et al. 2020). Individual nuclear loci were aligned with mafft v7.310 with default parameters and concatenated with the "concat" option from AMAS. A ML tree was reconstructed with IQ-TREE using the best-fitting loci-partitions and corresponding substitution models selected by Bayesian Information Criteria (BIC) in ModelFinder ("-m TESTNEW" option; Kalyaanamoorthy et al. 2017). Node support was assessed using 1000 ultrafast bootstrap replicates and associated gene- and site-concordance factors (Minh et al. 2020a). Individual gene trees were reconstructed by ML with IQ-TREE, with the best-fitting model selected using ModelFinder. The gene trees were not reconstructed for two loci without any SNP and three loci represented by only three individuals. The remaining 832 ML gene trees were then used to infer a phylogeny while taking incomplete lineage sorting into account using Astral v5.7.7 (Mirarab & Warnow, 2015), a quartet-based method consistent with the Multispecies Coalescent (MSC) model. In all phylogenetic reconstructions, *D. kappleri* was

used as an outgroup.

## Species partition and population genetic analyses

In order to corroborate the species hypothesis provided by the four morphotypes identified by Hautier et al. (2017), the species partition based on molecular data was investigated through population genetic analyses and species discovery methods.

*Admixture analysis.* Using variant calling files obtained from the captured nuclear loci, we excluded sites with more than two alleles with bcftools v1.8, sites with more than 25% missing data with the max-missing option of vcftools v0.1.17 (Danecek et al. 2011), and linked positions with the indep-pairwise option of the plink software (Purcell et al. 2007) with a threshold of 0.1. We excluded individuals of *D. kappleri*, *D. septemcinctus* and *D. pilosus* because they were represented by only two individuals each and uneven sampling is known to bias this type of analysis (Puechmaille 2016). The filtered vcf file of 19,872 SNPs was then converted to bed format with plink v1.9 (Chang et al. 2015) in order to estimate the number of genetic clusters in our dataset and individual ancestries with Admixture v1.3.0 (Alexander et al. 2009). Finally, graphical rendering was performed in Rstudio v4.2.0 using the plotADMIXTURE.r script (https://github.com/speciationgenomics/scripts/blob/master/plotADMIXTURE.r).

*Genetic variation.* A principal component analysis (PCA) of genetic variation was performed on the diploid sequences of the nuclear loci using the program popPhyl_PCA (https://github.com/popgenomics/ popPhyl_PCA). We focused on the 56 individuals of the four lineages of the *D. novemcinctus* complex and the two individuals of *D. pilosus*. Individuals of *D. kappleri* and *D. septemcinctus* were excluded because they were considered out of the complex. Two individuals (DNO-MC21, DPI-L29) with low completion were excluded from this analysis to increase the number of SNPs represented by all individuals. Thus, 3,350 SNPs shared by the remaining 56 individuals were retained. Rstudio v4.2.0 (R Core Team 2018; Posit Team 2022) using the script plotPCA.R (https://github.com/popgenomics/ popPhyl_PCA) and the packages shiny, plotly, tidyverse, and shinycssloaders were used to plot the results (Wickham et al. 2019; Sali and Attali 2020; Sievert 2020).

*Tree-based discovery methods.* We used the ML tree reconstructed from the concatenated nuclear dataset to conduct species delimitation using the Bayesian version of PTP (bPTP-ML) v0.51 (Zhang et al. 2013). This method distinguishes species from popula-

tions based on the transition in the substitution rate expected to follow two distinct Poisson distributions. In addition, we reconstructed an ultrametric tree by calibrating the ML tree using ancestral dates with IQ-TREE with the same options plus the "–date-tip 0" option by setting the root node ("–date-root" option) to 12.4 MYA according to Gibb et al. (2016). Using this ultrametric tree, species delimitation was performed under the Generalized Mixed Yule Coalescent method (GMYC; Pons et al. 2006) from the R package splits v1 (Ezard et al. 2009) using default parameters. Contrary to PTP, this method uses time to explore the transition in branching rate, expecting between-species branching rate to follow a Yule model and within-species rate a neutral coalescent process. Both PTP and GMYC were designed for single-locus datasets, but they are increasingly applied to concatenated multilocus datasets relying on the assumption that the shared genealogical history between loci would not bias branch lengths of the species tree (Luo et al. 2018). In the absence of gene flow, Luo et al. (2018) found a restricted impact of concatenating multiple loci on the performance of these methods. However, in the case of ongoing gene flow, these methods were less performant (Luo et al. 2018).

## Species validation

Species validation methods were performed considering the species partition best supported by discovery methods and the comparative approach. First, species delimitation under the MSC model (Rannala and Yang 2003) was run in BPP v4.4.1 (Yang 2015) using the consensus sequences of the nuclear loci. The four *D. novemcinctus* lineages defined in phylogenetic analyses (Fig. 1, see results for more details) and the three other species of our dataset (*D. kappleri*, *D. septemcinctus* and *D. pilosus*) were defined as candidate species. We used the species delimitation algorithm 1 with the finetune parameter e = 2 while also allowing inference of the species tree. We defined parameters $\tau$ and $\Theta$, respectively as divergence time and population size, by a gamma law of parameter a = 2 and b = 1000. After a burn-in phase of 50,000 generations, Markov Chain Monte Carlo sampling was done with a step of 5 during 500,000 steps. Secondly, PHRAPL, a species delimitation method that considers gene flow, was also applied. We evaluated the support as species of: i) Central (A) and Northern (B) lineages, and ii) Guianan (A) and Southern (B) lineages separately, using respectively *D. septemcinctus* or *D. pilosus* as a third lineage (C) and *D. kappleri* as outgroup. For these four configurations, three individuals within lineages A and B, and two within lineages C were subsampled from the

832 ML gene trees. This was replicated 10 times with replacement. Gene trees that did not contain enough lineage representatives were excluded. These observed gene trees were compared to 10000 simulated trees covering 48 demographic scenarios and three delimitation hypotheses (i.e., A, B, C represent one, two, or three distinct species). The best model was selected based on AIC.

## Comparative approach

Confirming that the species delimitation is consistent with current taxonomy can limit under- or over-splitting. For this reason, genetic differentiation between taxa with a consensual taxonomic rank can provide reference values to evaluate the candidate species.

*Population genomics statistics.* The script ABCstat_global.txt from the DILSmcsnp program (last accessed 1st October 2022; Fraïsse et al. 2021) was used to estimate population genomics statistics including mean pairwise divergence between lineages (Dxy) and net divergence (computed as $Da = Dxy - \frac{(Pi1+Pi2)}{2}$, where $Pi1$ and $Pi2$ are the pairwise nucleotide diversity of populations 1 and 2, respectively). Similarly, the script seq_stat_2pop_2N was used to calculate population genetic statistics using only two diploid genomes (Allio et al. 2021). Heterozygosity was computed for each individual ($Pi1$, $Pi2$), and total nucleotide diversity was computed as the mean pairwise divergence between all chromosomes, within and between individuals ($Pi_Tot$) (https://github.com/benoit nabholz/seq_stat_2pop). Then, the Genetic Differentiation Index, which is conceptually similar to Fst (Allio et al. 2021), was estimated as $1 - \frac{(Pi1,Pi2)/2}{Pi_tot}$.

# Results

## Data filtering and quality

In order to detect contamination, we estimated mitochondrial support for the different lineages using diagnostic positions. Except for one outlier (DNO-MC1000), all individuals had at least 70% of their reads supporting a single lineage (Fig. S4). Only two individuals had reads supporting two different mitochondrial lineages in large proportions, suggesting a significant amount of cross-contamination. Sample MC1000 had 57% of its mitochondrial reads supporting the Southern lineage while 32% supported the Guianan lineage, moreover, these reads recovered at least 97.5% of the DPs of these two lineages. Sample MC40 had 100% of Northern lineage DPs supported by 92% of

the mitochondrial reads but also had 52% of Central lineage DPs supported by 8% of the mitochondrial reads (Fig. S4). For these two individuals, the mitogenomes reconstructed from the majority of the reads were used in subsequent analyses as we considered this amount of mitochondrial contamination would not affect the majority-rule consensus sequence. The final mitochondrial dataset included 81 *Dasypus* individuals (2 *D. kappleri*, 2 *D. s. septemcinctus*, 1 *D. s. hybridus*, 2 *D. pilosus*, 1 *D. sabanicola*, 1 *D. mazzai*, 12 Guianan lineage, 33 Southern lineage, 13 Central lineage, and 13 Northern lineage), as well as a total of 13,924 unambiguously aligned nucleotide sites.

Reads originating from contaminations and those containing errors can be misinterpreted as heterozygous positions during variant calling. To alleviate their potential effect in the nuclear dataset, we found that the most effective filter was to keep only heterozygous positions with a read frequency between 0.3 and 0.7 per individual (Fig. S5); otherwise they were considered homozygous for the most frequent allele. We also excluded 159 loci, for which 598 sites had more than 75% heterozygous individuals per lineage. These loci showed high deviation from Hardy-Weiberg equilibrium expectations (Pval $< 0.1$) and could be the result of hidden paralogy. Combining these two filters turned unexpected negative inbreeding coefficient values to positive and led to a more even distribution of heterozygosity values among individuals (Fig. S5). This suggests that large genetic diversity values of a few outlier individuals observed before filtering were artificially inflated by sample cross-contamination or erroneous alleles. This approach allowed us to keep individuals detected as contaminated by filtering out the contaminated allele. Finally, after additional cleaning based on the sequencing depth of coverage (Fig. S6), we obtained a dataset of 837 capture loci alignments, representing a total of 506,355 sites, for 62 individuals (2 *D. kappleri*, 2 *D. s. septemcinctus*, 2 *D. pilosus*, 1 *D. sabanicola*, 1 *D. mazzai*, 10 Guianan lineage, 21 Southern lineage, 9 Central lineage, and 14 Northern lineage), with an average of 780 loci represented per sample (out of the 997 originally targeted) and 118x average depth of coverage (Table S2). Of the 837 loci, 94% had less than 20% missing data and 98% were represented by at least 20 individuals (Fig. S7).

## Phylogenetic reconstructions

The phylogenetic relationships of the genus *Dasypus* reconstructed from the mitogenomic and nuclear datasets are presented in Figure 1. Maximum likelihood reconstruction using the mitogenome (Fig. 1a), the partitioned concatenation of the 837 nuclear loci (Fig.

1b), and the species tree inference from the 832 individual gene trees (Fig. S8), all supported the four lineages previously recognized within *D. novemcinctus*, while the two individuals of *D. sabanicola* and *D. mazzai* fell within the Southern lineage (Fig. 1).

However, although the monophyly of these lineages, as well as internal nodes, received strong Bootstrap Support (BS = 100), their relationships differed between the nuclear and mitochondrial datasets (Fig. 1a,b). In both reconstructions, the lineages occurring north and west of the Andes (*D. pilosus*, Central and Northern lineages) formed a clade, within which the latter two were sister lineages. In the mitochondrial phylogeny (Fig. 1a), the Southern lineage was sister to the aforementioned clade, while the Guianan lineage diverged first within the *D. novemcinctus* complex. Conversely, in the nuclear analyses (Fig. 1b), the Guianan and Southern lineages were sister, splitting the *D. novemcinctus* complex into two clades separated by the Andes. Another important case of mito-nuclear discordance concerns populations from Western Mexico, which carry a Central lineage mitogenome while belonging to the Northern lineage based on their nuclear genomes (Fig. 1). Two further individuals, MC304 and AM78, carrying a Southern lineage mitochondrial haplotype, clustered with the Central and Guiana lineages, respectively, in nuclear analyses. Both were samples close to the geographic limits between these respective lineages.

Both topological conflicts and low gene- and site- concordance factors (respectively gCF and sCF, Fig. 1b) indicate high levels of discordance in our dataset. When accounting for ILS, the topologies recovered were similar to concatenation analyses (Fig. S8). However, we also found that the two main mito-nuclear discordances (changing position of the Guianan lineage and presence of Southern lineage haplotypes in Mexico) were concomitant with signals of introgression in the nuclear data (Supplementary Results, Fig. S9).

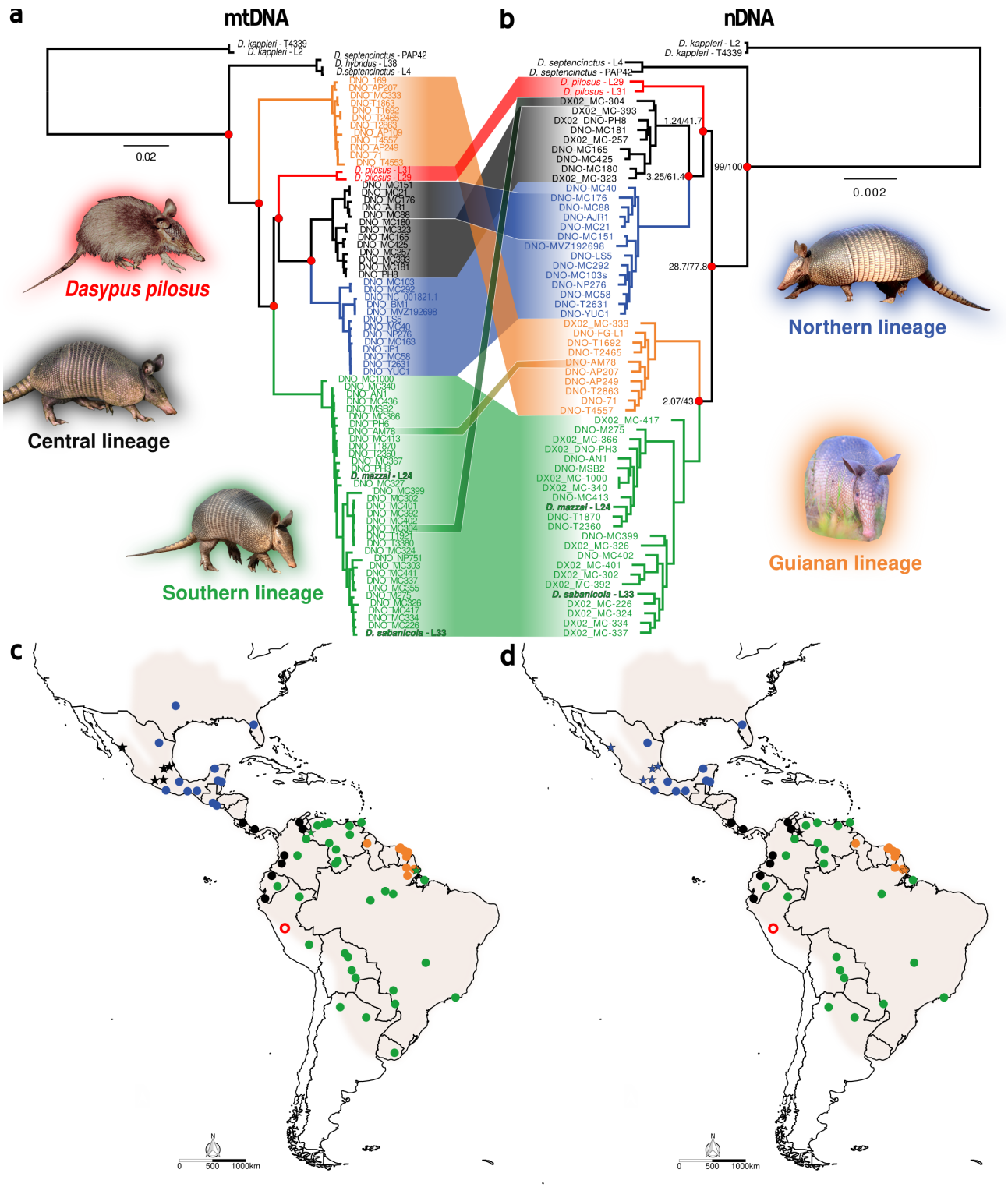


**Figure 1:** Phylogenetic relationships reconstructed by maximum likelihood and rooted using *Dasypus kappleri* from a) the mitogenomes of 81 individuals, and b) the 832 filtered nuclear loci of 62 individuals. Red circles at nodes indicate bootstrap support (BS = 100) for lineage inter-relationships and node labels represent the gene- and site-concordance factors (gCF/sCF). Maps represent the distribution of individuals according to their lineage and to c) mitogenomic and d) nuclear data. Individuals with discordant mito-nuclear lineages are represented by stars and *D. pilosus* by open circles. *D. novemcinctus* distribution is highlighted in gray. Photo credits: A. Baertschi (xenarthrans.org), K. Miller, Andresiade, A. Reed (iNaturalist.org), and Q. Martinez.

## Genetic structure and gene flow between lineages

In the analysis of genetic admixture, maximum statistical support was obtained for a model with four genetic clusters (error = 0.2769 for k = 4; Fig. S10), corresponding to the four *D. novemcinctus* lineages (Fig. 2). Similar to the previous results, the two samples of *D. mazzai* and *D. sabanicola* were consistently recovered as part of the Southern lineage. This analysis also identified eight individuals showing evidence of admixture, with assignation probabilities ranging 57%-98%: DNO-MC393, DNO-MC40, DNO-MC21, DNO-MC417, DNO-MC413, DNO-MC399, DNO-MC402, and DNO-MC392 (Fig. 2). Cases of admixture involved geographically adjacent lineages and were mainly located at contact zones.

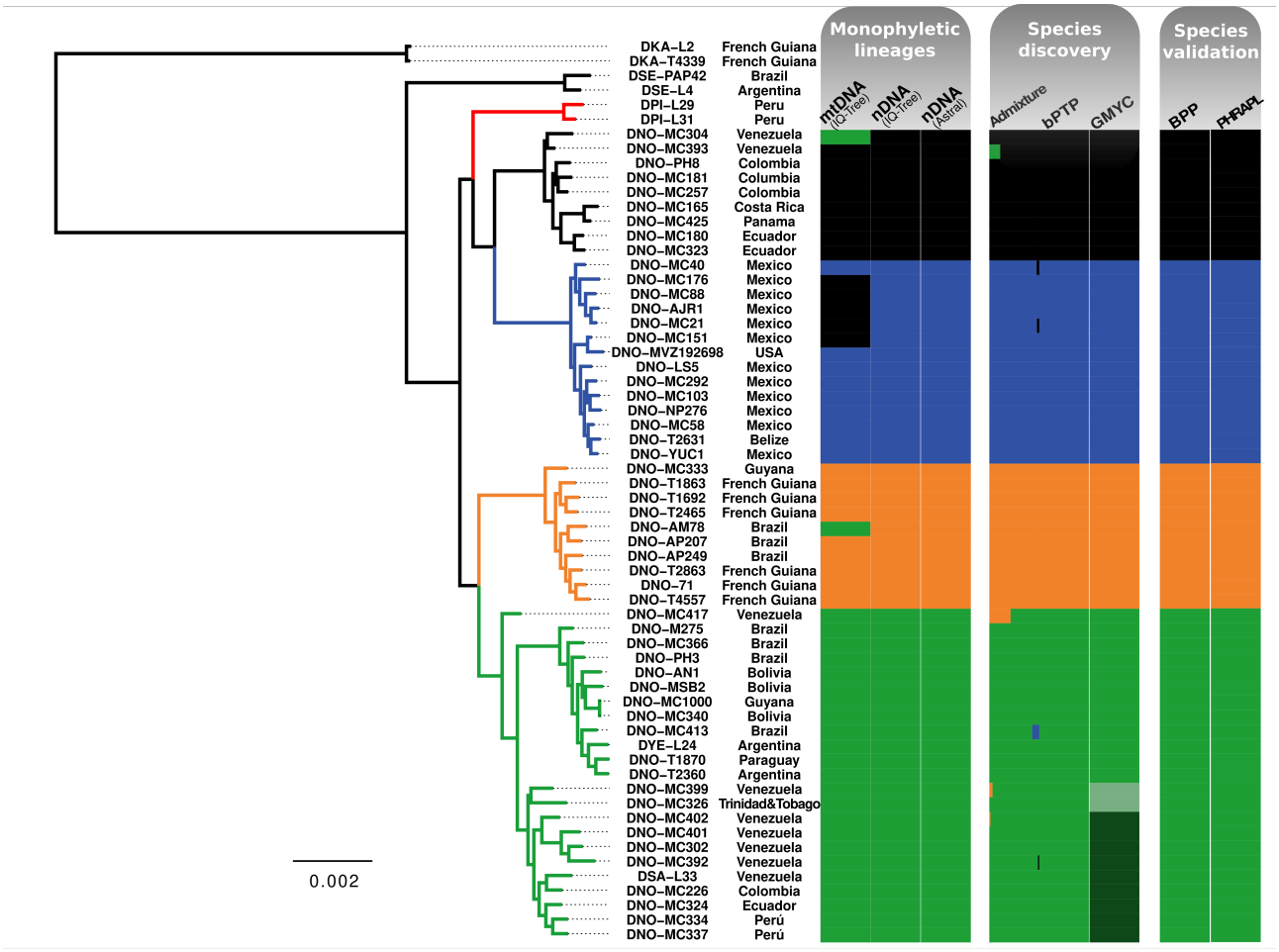


**Figure 2:** Phylogenetic relationships reconstructed by maximum likelihood from the 837 filtered nuclear loci and rooted using *Dasypus kappleri* and *D. septemcinctus* as outgroups. The color of the diagrams represents the assignment of individuals to lineages. For phylogenetic analyses (IQ-Tree and ASTRAL) barplot represents qualitative assignment to the monophyletic lineages reconstructed. For the Admixture analysis, barplot corresponds to the assignment probability. Finally, for phylogenetic delimitation barplot represents the species support.

The PCA of genetic variation confirmed the structure of these four lineages (Fig. 3). The PC1 axis distinguished the Guianan lineage from the others, while PC2 separated the Southern lineage from the Central and Northern lineages, the latter two being very close. The PC3 axis further distinguished *D. pilosus* from the four other lineages of the species complex. Two admixed individuals (DNO-MC393, DNO-MC417) had intermediate positions between their respective parental lineages. Furthermore, a PCA of the Southern lineage (Fig. S11), as well as admixture analyses with k = 5 (error = 0.2784) revealed two groups within the Southern lineage: one distributed in the north of South America (from Venezuela to the north of Peru) and the other one farther south (from Northern Brazil to north Argentina; Fig. S11).

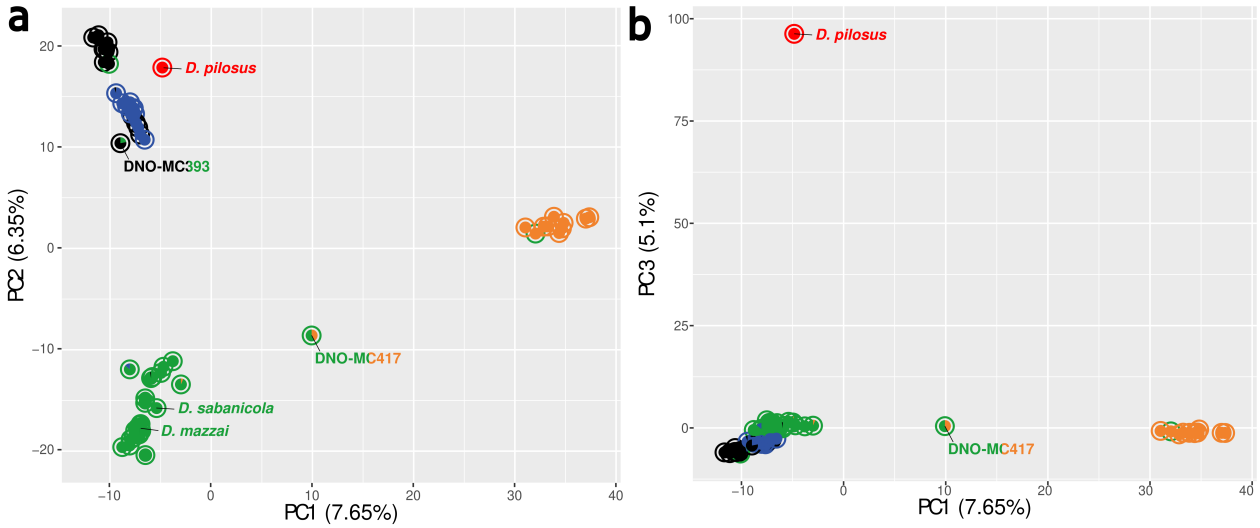


**Figure 3:** Principal Component Analysis of genetic variance (PCA) conducted on the 3,350 SNPs in 837 nuclear loci shared by all 56 individuals. a) Projection on first (PC1) and second (PC2) axes. b) Projection on first (PC1) and third (PC3) axes. The outer circles represent mitochondrial lineages and pie charts the nuclear lineages identified in admixture analyses. Individuals of *D. pilosus*, *D. sabanicola* and *D. mazzai* are labeled.

## Species delimitation

Species delimitation analyses with BPP and bPTP-h supported the four lineages as distinct species with high support (BPP posterior probability = 1.00; bPTP-h acceptance rate = 0.54; Fig. 2 and Fig. S12). This was also corroborated by PHRAPL that supported models with distinct species even if gene flow between lineages was estimated to occur and the genealogical divergence index (PHRAPL_gdi) was 0.3 (Table S3; Jackson et al. 2017). The GMYC analysis also supported Guianan, Central and Northern lineages, but split the Southern lineage into three species (Likelihood Ratio = 29.3; p-value $< 0.001$), with one species in Brazil, Bolivia, Paraguay, and Argentina, another

in Venezuela, Colombia, Ecuador, and Peru, and finally a last species composed of two individuals in Venezuela and Trinidad-Tobago (Fig. 2 and Fig. S13).

We used population genetic statistics to estimate genomic differentiation between lineages. Absolute divergence ($D_{xy}$) between the four lineages of the species complex was about 0.005 or 0.006 for all pairwise combinations (Fig. S14). These values were comparable or greater than those obtained in pairwise comparisons with the taxonomically unambiguous species *D. pilosus* ($D_{xy}$(*D. pilosus*) = 0.005), but lower than those of *D. s. septencinctus* and *D. kappleri* ($D_{xy}$(*D. septencinctus*) $>$ 0.007 ; $D_{xy}$(*D. kappleri*) $>$ 0.013). According to $D_a$ and GDI statistics, the divergence between the Southern and Guianan lineages was lower than between either lineages and *D. pilosus* ($D_a$(Southern - Guianan lineages) = 0.002 $<$ $D_a$(Southern lineage-*D.pilosus*) = 0.003; GDI(Southern - Guianan lineages) = 0.36 $<$ GDI(Southern lineage - *D.pilosus*) = 0.58). On the contrary, the Central and Northern lineages had similar levels of divergence to *D. pilosus* ($D_a$(Central - Northern lineages) = 0.003; GDI(Central - Northern lineages) = 0.58). Finally, both Southern-Guianan and Central-Northern lineages had less fixed differences than *D. pilosus* had with other lineages (Fixed diff.(Southern - Guianan lineages) = 0.003; Fixed diff.(Central - Northern lineages) = 0.003; Fixed diff.(*D. pilosus*-Southern lineage) = 0.004).

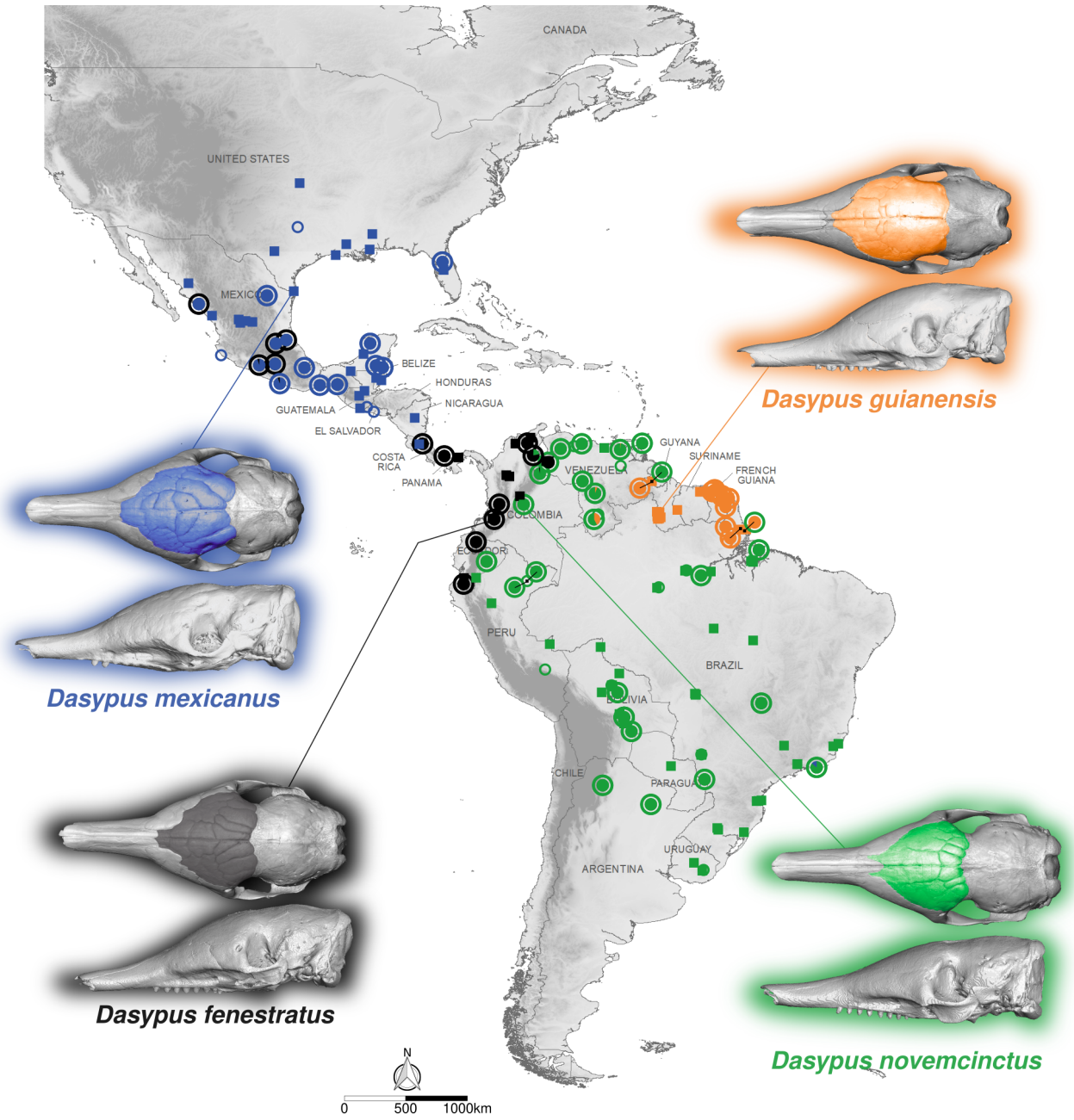


**Figure 4:** Distribution map and genetic composition of individuals of the four recognized species: *Dasypus mexicanus* (Northern lineage in blue), *Dasypus fenestratus* (Central lineage in black), D*asypus novemcinctus* (Southern lineage in green), and *Dasypus guianensis* sp. nov. (Guianan lineage in orange). Outer circles represent mitochondrial lineages and pie charts the assignment probability generated by admixture analyses based on nuclear data. Open circles indicate individuals for which only mitogenomic data is available. Squares represent individuals from Hautier et al. (2017) colored according to their morphogroup affiliation based on a Discriminant Analysis of skull shape using geometric morphometrics and for which no genetic data are available. Three-dimensional reconstructions of skulls obtained using microCT scans of representative specimens vouchers are presented for each species with frontal paranasal sinuses and recesses highlighted: USNM 33867 (*Dasypus mexicanus*), AMNH 32356 (*Dasypus fenestratus*), AMNH 136252 (*Dasypus novemcinctus*), and ROM 32868 (*Dasypus guianensis* sp. nov.).

# Discussion

## Disentangling cross-contamination and genotyping errors

Genotyping errors are inaccuracies in the determination of individual genotypes (Bonin et al. 2004) that can occur at multiple stages of a project such as: sampling (contamination and mislabelling), sample storage (postmortem mutation; Briggs et al. 2007), DNA extraction and library preparation (cross-contaminations; Ballenghien et al. 2017), amplification (PCR errors; Potapov and Ong 2017), sequencing (Heydari et al. 2019), data analysis (SNP calling errors; Hwang et al. 2015), and mapping of paralogs (Hohenlohe et al. 2011) (see Mastretta-Yanes et al., (2015) for an exhaustive review). In this study, we included several museum specimens that postmortem mutations and low levels of endogenous DNA make particularly prone to artifactual substitutions (Bi et al. 2013). Notably, historical DNA extracted from museum specimens is particularly sensitive to contamination from exogenous modern DNA (Raxworthy and Smith 2021).

We implemented methods to detect cross-contamination based on mitochondrial read depth of coverage, which have been proposed in the context of ancient DNA (Green et al. 2008, 2010). Indeed, when contaminated, sequence reads represent both endogenous and contaminant DNA, supporting two (or multiple) mitogenomes. Here, we used mitochondrial substitutions that contribute to differentiate lineages as diagnostic positions (DPs) to identify the lineage supported by the majority of mitochondrial reads in each individual sample (Fig. S3). Although we performed separate DNA extractions for museum and fresh samples, which differed significantly in their DNA content, we detected at least two cases of apparent cross-contamination between our samples (Fig. S4), which likely occurred at the library preparation step. These types of manipulation errors are common and are particularly difficult to detect in population genomic datasets including closely-related individuals (Ballenghien et al. 2017). Nevertheless, automated tools based on mitochondrial haplotype frequencies have been developed to detect cross-contaminations (Fiévet et al. 2019; Weissensteiner et al. 2021). This type of workbench error should therefore be more systematically searched and taken into account in subsequent evolutionary analyses.

Complementary to cross-contamination detection, data filtering is required. The most stringent strategy would be to exclude cross-

contaminated individuals from the dataset. However, when using rare samples, excluding the reads originating from contamination is preferable, but remains challenging. Such reads are expected to occur at lower frequency than those corresponding to endogenous DNA. Thus, they should have a neglectable effect on the reconstruction of the mitogenome as only the most frequent allele is called. However, for the nuclear genome, SNPs carried by the contaminant DNA or other genotyping errors can be misinterpreted as heterozygous positions. We minimized genotyping errors by using two filters. First, because genotyping errors are likely to have a read frequency that deviates from the expected 50/50 allele frequencies at heterozygous loci (Hansen et al. 2022), we retained only heterozygous positions supported by 30%-70% of the reads. It should be noted that this filter is only efficient with a low amount of errors or cross-contamination as they cannot be distinguished from endogenous DNA if they are equally represented. Second, the proportion of heterozygous individuals in a population for a given position is not expected to be higher than 50%, a deviation from Hardy-Weinberg equilibrium could be attributed to the mapping of paralogous sequences (Xu et al. 2002; Hosking et al. 2004; Hohenlohe et al. 2011). Therefore, we excluded 159 loci with at least 75% of individuals per lineage that were heterozygous for the same position. This happened even though the exons targeted for sequence capture were carefully selected from single-copy orthologous genes in the nine-banded armadillo genome. These two filters are still appropriate whatever the allele frequency in the population as they rely on characteristics that are not affected by it.

The efficiency of these filters was assessed by monitoring their effect on heterozygosity and inbreeding coefficient estimates. Because any genotyping error is expected to be present in a heterozygous state, they are expected to inflate heterozygosity and decrease inbreeding coefficient values (Jun et al. 2012; Petrou et al. 2019; Anderson et al. 2023). These two metrics enabled us to easily identify outlier individuals (Fig. S5). After applying these two filters, heterozygosity values and inbreeding coefficients became much more homogeneous among individuals of the same lineage, suggesting an efficient removal of genotyping errors. We thus confidently kept the two cross-contaminated individuals identified based on mitochondrial reads. As expected, the application of these filtering criteria had a particularly strong effect on admixture analyses (Fig. S15). Our results therefore call for a more systematic use of these two metrics to select appropriate filters in population genomics and species delimitation studies, as previously proposed (Hosking et al. 2004; Hansen et al. 2022).

## Species delimitation

This study represents the most comprehensive molecular assessment of the *Dasypus novemcinctus* complex. Our mitogenome tree and nuclear phylogenetic reconstructions based on both the concatenation of the 837 nuclear loci and the summary of gene trees (Fig. 1, Fig. S8) consistently recovered four major lineages within *D. novemcinctus*. These findings are largely corroborated by prior mitochondrial (Feijó et al. 2019; Arteaga et al. 2020) and morphological studies (Hautier et al. 2017). Nevertheless, we uncovered pervasive topological discordance in our data that may reflect a rapid diversification of the *Dasypus novemcinctus* complex. Ancient introgression seems to explain parts of this discordance, although the informativeness of our data is too limited to reconstruct these introgression events with high confidence (Supplementary Results, Fig. S9). On the other hand, two cases of mito-nuclear discordances concerning single individuals, and five individuals with admixed nuclear genomes, support ongoing gene flow. Most of them were located close to contact zones between adjacent lineages. Despite that, the four lineages were unanimously supported as distinct species in both discovery and validation methods. Species delimitation results should be considered carefully, because these models have been shown to misinterpret population structure as species boundaries (Sukumaran and Knowles 2017), and the outcome of validation methods depends on the pre-defined candidate species. However, we found that the four lineages showed similar levels of divergence and genetic structure compared to *D. pilosus*, further supporting their split into different species. Post-speciation gene flow is well known (Wang et al. 1997; Bull et al. 2006; Nosil 2008; Suo et al. 2008; Feder et al. 2012) and has been reported for several mammalian groups, especially when genome-scale methods are applied (Trigo et al. 2008, 2013; Kumar et al. 2017; Ge et al. 2023). Admixture across *Dasypus* lineages seems to be very limited spatially: admixed individuals come from the contact zones themselves, and most samples from range edges do not show any admixture. This phylogeographic pattern is consistent with distinct species maintaining narrow hybrid zones, in agreement with the other molecular lines of evidence.

## Integrative taxonomic support for four distinct species in the *Dasypus novemcinctus* complex

Based on the strong genetic integrity, diagnostic morphological differences, and the molecular delimitation results, we advocate for the elevation of the four lineages to species rank. The oldest available names for these lineages are: *Dasypus mexicanus* Peters,

1864 for the Northern lineage, *Dasypus fenestratus* Peters, 1864 for the Central lineage, and *Dasypus novemcinctus* Linnaeus, 1758 for the Southern lineage. The Guianan lineage lacks a binomial name and we describe it here as *Dasypus guianensis* sp. nov. (see Appendix I). The recognition of the Guiana Shield population as a new species accords with a bulk of morphological (Billet et al. 2017; Hautier et al. 2017; Feijó et al. 2018) and molecular (Huchon et al. 1999; this study; Gibb et al. 2016; Feijó et al. 2019; Arteaga et al. 2020) evidence. For example, studying the paranasal sinuses of nine-banded armadillos, Billet et al. (2017) recovered a distinct pattern shared only by specimens from the Guiana Shield (Fig. 5). Similarly, Hautier et al. (2017) recognized distinct skull morphology in nine-banded armadillos from that region. In the recent taxonomic revision of the genus, Feijó et al. (2018) also acknowledged a set of distinct cranial qualitative traits present in the Guianan long-nosed armadillos.

With this new classification, *D. fenestratus* is the species present in the western Andes from Ecuador, Colombia, Venezuela to Costa Rica (Fig. 4). *D. mexicanus* occurs from Costa Rica to the USA (Fig. 5). There is no longer uncertainty about its distribution as we discovered that individuals with *D. mexicanus* mitochondrial haplotypes found in South America (Arteaga et al. 2020) were likely the result of cross-contaminations (See Supplementary Results; Table S4; Fig. S16). *D. novemcinctus* is now limited to the eastern Andes, spanning from northern Argentina to eastern Peru, eastern Ecuador, eastern Colombia and Venezuela (Fig. 5). It is noteworthy that our analyses failed to recognize the two available individuals of *D. mazzai* and *D. sabanicola* as independent lineages, since they were consistently nested within *D. novemcinctus* (Fig. 1; Fig. S17). Additional individuals are still required to corroborate this result, but it should be noted that in this scenario, both species would be junior synonyms of *D. novemcinctus*. This finding is consistent with the similar paranasal sinus morphology shared between *D. sabanicola* and *D. novemcinctus* recovered by Billet et al. (2017). In contrast, Feijó et al. (2018) recognized the three species to be distinct based on carapace and morphological traits. The morphological differences might be related to allometric effects as *D. sabanicola* and *D. mazzai* are smaller than *D. novemcinctus*. Indeed, allometric effects are known to be strong on the skull of this group (Hautier et al. 2017; Le Verger et al. 2020, 2023) and nine-banded armadillos inhabiting open biomes (the habitats of *D. mazzai* and *D. sabanicola*) tend to be small (Feijó et al. 2020). The proportionally shorter muzzle and larger braincase in *D. mazzai* and *D. sabanicola* (Feijó et al. 2018: Fig. 5) fully match allometric expectations for the group (Le Verger et al. 2020) and thus

cannot be considered as strong evidence for specific distinction. Moreover, when reanalyzing the 16S rRNA sequences of Abba et al. (2018), we found *D. novemcinctus* individuals clustered with their *D. mazzai* 16S sequence, in agreement with our mitogenome analyses (Supplementary Results; Fig. S17). Additional morphological studies and new genomic samples of *D. sabanicola* and *D. mazzai* are still needed to properly clarify their taxonomic status.

In addition to the four major clades, we further identified two divergent sub-lineages within *D. novemcinctus* (Fig. S11): one including individuals from northern South America (Venezuela, Colombia, Ecuador, and Peru) and another from southern South America, including armadillos mostly sampled in Brazil, Bolivia, Argentina, and Paraguay. Although these lineages were supported as distinct species by the GMYC species delimitation method, they exhibit a shallow genetic divergence and were not recovered in any of the prior morphological studies (Billet et al. 2017; Hautier et al. 2017; Feijó et al. 2018). Interestingly, Feijó et al. (2018) reported a predominance of nine movable bands in northern South American armadillos, while those from the southern part exhibit mostly an eight-banded pattern. If these distinct patterns reflect some hidden osteological traits is still unclear. We also cannot rule out that this phylogeographic structure may be inflated by our sampling gaps and thus more individuals from Brazil, Bolivia, and Peru are required to better understand the genetic structure within *D. novemcinctus*.

Overall, focusing on the most widespread xenarthran species, we uncovered hidden genetic divergence leading to the recognition of four distinct species with parapatric distributions. Their distribution are limited by well-known geographic barriers (e.g. Andes, Guiana Shield) in South America and likely reflect large-scale biogeographic events driving speciation in other xenarthrans (Moraes-Barros and Arteaga 2015; Coimbra et al. 2017; Miranda et al. 2018; Feijó et al. 2019; Ruiz-García et al. 2020). Future genomic studies are required to fully understand the dynamics of the speciation process in xenarthrans, as previously done in South American felids (Trigo et al. 2013; Figueiró et al. 2017; Trindade et al. 2021; Ramirez et al. 2022) for instance.

## 5. Conclusion

Through the integration of museomic data and exon capture techniques, our study has yielded the most comprehensive molecular revision of the taxonomy of the *Dasypus* genus. Using phylogenetic reconstructions, multiple approaches to characterize genetic exchanges, and numerous methods to delimit lineages, our study provides a global view to support their taxonomic status. By placing our results in the context of previous molecular and morphological studies, we provide an integrative view that disentangles past discrepancies. This improved our understanding of speciation events and deciphered the relationships within the *Dasypus novemcinctus* complex. This integrative approach allowed us to recognize four species (*D. novemcinctus*, *D. fenestratus*, *D. mexicanus*, and *D. guianensis*) within a geographically widespread single taxon, including a new species from the Guiana Shield, which is the first new armadillo species described in the last 30 years. This new taxonomic arrangement raises the number of Xenarthra species to 42 but the validity of *D. mazzai* and *D. sabanicola* is strongly challenged. At the larger scale, further genomic studies are required to fully understand the diversification of xenarthrans.

## Acknowledgments

This paper is dedicated to the memory of François Catzeflis who passed away during the course of this study that he initiated in French Guiana back in the mid-1990s together with Dorothée Huchon and Emmanuel Douzery. We would like to thank Clara Belfiore, Malia Chevolot, and Benjamin Nigon for their contribution to this study as undergraduate students. This work would not have been possible without the generous help of the following individuals and institutions in accessing armadillo tissue samples through the years: Mariella Superina, Agustín Abba, Agustín Jiménez, Andy Noss, Rodolfo Rearte, Guillermo Pérez-Jimeno, Paulo Prodöhl, Danny Devillier, Luis Sigler, Bryant McAllister, Julio Perez, Vincent Bersot, Enrique Lessa, Cynthia Steiner, Omar Linares, Yuri Leite, Cuahutémoc Chávez, Horacio Bárcenas, Leonardo Lopez, Nathalie Delsuc, Sébastien Pascal, Sérgio Ferreira-Cardoso, Rémi Allio, Pierre-Henri Fabre, and Fabien Condamine; Philippe Gaucher, François Ouhoud-Renoux, Anne Lavergne, Roxanne Schaub, Lucile Dudoignon, Jean-François Mauffrey, and Jean-Christophe Vié

(French Guiana); Claudia Regina Silva (Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá, Macapá, Brazil); Maria Nazareth da Silva (Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil); Daniel Hernández (Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay); Sergio Vizcaíno (Museo de La Plata, La Plata, Argentina); Ross MacPhee (American Museum of Natural History, New York, USA); Jonathan Dunnum and Joseph Cook (Museum of Southwestern Biology, Albuquerque, USA); Jake Esselstyn, Donna Dittman, and Mark Hafner (Louisiana State University Museum of Natural Science, Baton Rouge, USA); Darrin Lunde (National Museum of Natural History, Washington, USA); Jim Patton (Museum of Vertebrate Zoology, Berkeley, USA); Géraldine Véron, Aurélie Verguin, and Céline Bens (Muséum national d'Histoire naturelle, Paris, France); Fernando Cervantes and Yolanda Hortelano (CNMA: Colección Nacional de Mamíferos, UNAM); Consuelo Lorenzo (Eco-SC-M: Colección Mastozoológica, Colegio de la Frontera Sur, México); Juan Carlos Vidal (ENCB-IPN: Colección Mastozoológica, Instituto Politécnico Nacional, México); Sol Menejes (HMAM: Colección Mastozoología Instituto Agropecuario de Hidalgo, México). We are also indebted to Alison Devault, Jennifer Klunk, Jake Enk, and Hendrik Poinar for sharing their expertise in exon capture bait design for museomics. The JAGUARS collection is supported by the Collectivité Territoriale de Guyane and the Direction Générale des Territoires et de la Mer Guyane. Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform. This work has been supported by grants from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257) and Investissements d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; CEMEB: ANR-10-LABX-0004). This is contribution ISEM 2024-XXX-SUD of the Institut des Sciences de l'Evolution de Montpellier.

# References

Abba A.M., Cassini G.H., Túnez J.I., Vizcaíno S.F. 2018. The enigma of the Yepes' armadillo: Dasypus mazzai, D. novemcinctus or D. yepesi? Rev. Mus. Argent. Cienc. Nat. 20:83–90.

Abba A.M., Superina M. 2010. The 2009/2010 armadillo red list assessment. Edentata. 11:135–184.

Alexander D.H., Novembre J., Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Allio R., Tilak M.-K., Scornavacca C., Avenant N.L., Kitchener A.C., Corre E., Nabholz B., Delsuc F. 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. Elife. 10:e63167.

Alves P.C., Melo-Ferreira J., Freitas H., Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. Philos. Trans. R. Soc. B Biol. Sci. 363:2831–2839.

Anderson G., Macdonald J.I., Potts J., Feutry P., Grewe P.M., Boutigny M., Davies C.R., Muir J.A., Roupsard F., Sanchez C. 2023. Evaluating DNA cross-contamination risk using different tissue sampling procedures on board fishing and research vessels. ICES J. Mar. Sci. 80:728–738.

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.

Arteaga M.C., Gasca-Pineda J., Bello-Bedoy R., Eguiarte L.E., Medellin R.A. 2020. Conservation genetics, demographic history, and climatic distribution of the nine-banded armadillo (Dasypus novemcinctus): an analysis of its mitochondrial lineages. Conservation Genetics in Mammals. Springer. p. 141–163.

Arteaga M.C., McCormack J.E., Eguiarte L.E., Medellín R.A. 2011. Genetic admixture in multidimensional environmental space: asymmetrical niche similarity promotes gene flow in armadillos (Dasypus novemcinctus). Evolution. 65:2470–2480.

Arteaga M.C., Piñero D., Eguiarte L.E., Gasca J., Medellín R.A. 2012. Genetic structure and diversity of the nine-banded armadillo in Mexico. J. Mammal. 93:547–559.

Auguie B., Antonov A. 2017. gridExtra: miscellaneous functions for "grid" graphics. R Package Version. 2:602.

Bailey V. 1905. Biological survey of Texas. North Am. Fauna. 25:1–222.

Ballenghien M., Faivre N., Galtier N. 2017. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. BMC Biol. 15:1–16.

Bi K., Linderoth T., Vanderpool D., Good J.M., Nielsen R., Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. Mol. Ecol. 22:6018–6032.

Billet G., Hautier L., De Thoisy B., Delsuc F. 2017. The hidden anatomy of paranasal sinuses reveals biogeographically distinct morphotypes in the nine-banded armadillo (Dasypus novemcinctus). PeerJ. 5:e3593.

Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C., Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. Mol. Ecol. 13:3261–3273.

Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4:e1660.

Boubli J.P., Byrne H., da Silva M.N., Silva-Júnior J., Araújo R.C., Bertuol F., Gonçalves J., de Melo F.R., Rylands A.B., Mittermeier R.A. 2019. On a new species of titi monkey (Primates: Plecturocebus Byrne et al., 2016), from Alta Floresta, southern Amazon, Brazil. Mol. Phylogenet. Evol. 132:117–137.

Briggs A.W., Stenzel U., Johnson P.L., Green R.E., Kelso J., Prüfer K., Meyer M., Krause J., Ronan M.T., Lachmann M. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. 104:14616–14621.

Brisson M.J. 1762. Regnum animale in classes IX distributum, sive synopsis methodica sistens classium, quadripedum scilicet & cetaceorum, particularum divisionem in ordines, sectiones, genera & species. Ed. Altera Auctior 2nd Ed Theodorum Haak Lunduni Batav. Leiden. 296.

Bull V., Beltrán M., Jiggins C.D., McMillan W.O., Bermingham E., Mallet J. 2006. Polyphyly and gene flow between non-sibling Heliconius species. BMC Biol. 4:1–17.

Burgin C.J., Colella J.P., Kahn P.L., Upham N.S. 2018. How many species of mammals are there? J. Mammal. 99:1–14.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Mol. Ecol. 22:4369–4383.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4:s13742-015-0047–8.

Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890.

Coimbra R.T.F., Miranda F.R., Lara C.C., Schetino M.A.A., Santos F.R. dos. 2017. Phylogeographic history of South American populations of the silky anteater Cyclopes didactylus (Pilosa: Cyclopedidae). Genet. Mol. Biol. 40:40–49.

Costa-Araújo R., De Melo F.R., Canale G.R., Hernández-Rangel S.M., Messias M.R., Rossi R.V., Silva F.E., Da Silva M.N.F., Nash S.D., Boubli J.P. 2019. The Munduruku marmoset: a new monkey species from southern Amazonia. PeerJ. 7:e7019.

Costa-Araújo R., Silva-Jr J.S., Boubli J.P., Rossi R.V., Canale G.R., Melo F.R., Bertuol F., Silva F.E., Silva D.A., Nash S.D. 2021. An integrative analysis uncovers a new, pseudocryptic species of Amazonian marmoset (Primates: Callitrichidae: Mico) from the arc of deforestation. Sci. Rep. 11:15665.

Coyne J.A., Orr H.A. 2004. Speciation. Sinauer Associates Sunderland, MA.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10:giab008.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Desmarest A.G. 1822. Mammalogie ou description des espèces de mammifères. Seconde partie, contenant les ordres de Rongeurs, des Édentés, des Pachydermes, des Ruminants et des Cétacés. Encyclopédie Méthodique. Paris: Veuve Agasse.

Dobzhansky T. 1982. Genetics and the Origin of Species. Columbia University Press.

Douzery E.J., 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.

Dres M., Mallet J. 2002. Host races in plant–feeding insects and their importance in sympatric speciation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 357:471–492.

Dufresnes C., Brelsford A., Crnobrnja-Isailović J., Tzankov N., Lymberakis P., Perrin N. 2015. Timeframe of speciation inferred from secondary contact zones in the European tree frog radiation (Hyla arborea group). BMC Evol. Biol. 15:1–8.

Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

Ezard T., Fujisawa T., Barraclough T.G. 2009. Splits: species' limits by threshold statistics. R Package Version. 1:r29.

Feder J.L., Egan S.P., Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Feijó A. in press. Xenarthrans of Brazilian Amazon: recent discoveries, knowledge gaps, and conservation concerns. Amazonian Mammals - Current Knowledge and Conservation Priorities. Springer.

Feijó A., Brandão M.V. 2022. Taxonomy as the first step towards conservation: an appraisal on the taxonomy of medium-and large-sized Neotropical mammals in the 21st century. SciELO Brasil.

Feijo A., Cordeiro-Estrela P. 2016. Taxonomic revision of the Dasypus kappleri complex, with revalidations of Dasypus pastasae (Thomas, 1901) and Dasypus beniensis Lönnberg, 1942 (Cingulata, Dasypodidae). Zootaxa. 4170:271–297.

Feijó A., Patterson B.D., Cordeiro-Estrela P. 2018. Taxonomic revision of the long-nosed armadillos, Genus Dasypus Linnaeus, 1758 (Mammalia, Cingulata). PLoS One. 13:e0195084.

Feijó A., Patterson B.D., Cordeiro-Estrela P. 2020. Phenotypic variability and environmental tolerance shed light on nine-banded armadillo Nearctic invasion. Biol. Invasions. 22:255–269.

Feijó A., Vilela J.F., Cheng J., Schetino M.A.A., Coimbra R.T., Bonvicino C.R., Santos F.R., Patterson B.D., Cordeiro-Estrela P. 2019. Phylogeny and molecular species delimitation of long-nosed armadillos (Dasypus: Cingulata) supports morphology-based taxonomy. Zool. J. Linn. Soc. 186:813–825.

Fiévet A., Bernard V., Tenreiro H., Dehainault C., Girard E., Deshaies V., Hupe P., Delattre O., Stern M.-H., Stoppa-Lyonnet D. 2019. ART-DeCo: easy tool for detection and characterization of cross-contamination of DNA samples in diagnostic next-generation sequencing analysis. Eur. J. Hum. Genet. 27:792–800.

Figueiró H.V., Li G., Trindade F.J., Assis J., Pais F., Fernandes G., Santos S.H., Hughes G.M., Komissarov A., Antunes A. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 3:e1700299.

Fraïsse C., Popovic I., Mazoyer C., Spataro B., Delmotte S., Romiguier J., Loire E., Simon A., Galtier N., Duret L. 2021. DILS: Demographic inferences with linked selection by using ABC. Mol. Ecol. Resour. 21:2629–2644.

Funk W.C., McKay J.K., Hohenlohe P.A., Allendorf F.W. 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27:489–496.

Galtier N. 2019. Delineating species in the speciation continuum: A proposal. Evol. Appl. 12:657–663.

Gardner A.L., Ramírez-Pulido J. 2020. Type localities of Mexican land mammals, with comments on taxonomy and nomenclature. Museum of Texas Tech University.

Garrison E., Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv Prepr. ArXiv12073907.

Ge D., Wen Z., Feijó A., Lissovsky A., Zhang W., Cheng J., Yan C., She H., Zhang D., Cheng Y. 2023. Genomic consequences of and demographic response to pervasive hybridization over time in climate-sensitive pikas. Mol. Biol. Evol. 40:msac274.

Gibb G.C., Condamine F.L., Kuch M., Enk J., Moraes-Barros N., Superina M., Poinar H.N., Delsuc F. 2016. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. Mol. Biol. Evol. 33:621–642.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.-Y. 2010. A draft sequence of the Neandertal genome. Science. 328:710–722.

Green R.E., Malaspinas A.-S., Krause J., Briggs A.W., Johnson P.L., Uhler C., Meyer M., Good J.M., Maricic T., Stenzel U. 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell. 134:416–426.

Gusmão A.C., Messias M.R., Carneiro J.C., Schneider H., de Alencar T.B., Calouro A.M., Dalponte J.C., de Souza Mattos F., Ferrari S.F., Buss G. 2019. A new species of titi monkey, Plecturocebus Byrne et al. 2016 (Primates, Pitheciidae), from southwestern Amazonia, Brazil. Primate Conserv. 33:1–15.

Hamilton N. 2018. ggtern: an extension to 'ggplot2', for the creation of ternary diagrams. R package v. 2.2. 2. .

Hansen M.H., Lang C.S., Abildgaard N., Nyvold C.G. 2022. Comparative evaluation of the heterozygous variant standard deviation as a quality measure for next-generation sequencing. J. Biomed. Inform. 135:104234.

Hao Z., Lv D., Ge Y., Shi J., Weijers D., Yu G., Chen J. 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. 6:e251.

Hautier L., Billet G., De Thoisy B., Delsuc F. 2017. Beyond the carapace: skull shape variation and morphological systematics of long-nosed armadillos (genus Dasypus). PeerJ. 5:e3650.

Helgen K.M., Pinto C.M., Kays R.W., Helgen L.E., Tsuchiya M.T., Quinn A., Wilson D.E., Maldonado J.E. 2013. Taxonomic revision of the olingos (Bassaricyon), with description of a new species, the Olinguito. Pensoft.

Herrmann D., Poncet B.N., Manel S., Rioux D., Gielly L., Taberlet P., Gugerli F. 2010. Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. Genome. 53:302–310.

Hey J., Pinho C. 2012. Population genetics and objectivity in species diagnosis. Evolution. 66:1413–1429.

Heydari M., Miclotte G., Van de Peer Y., Fostier J. 2019. Illumina error correction near highly repetitive DNA regions improves de novo genome assembly. BMC Bioinformatics. 20:1–13.

Hohenlohe P.A., Amish S.J., Catchen J.M., Allendorf F.W., Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. 11:117–122.

Hollister N. 1925. The systematic name of the Texas armadillo. J. Mammal. 6:60.

Hosking L., Lumsden S., Lewis K., Yeo A., McCarthy L., Bansal A., Riley J., Purvis I., Xu C.-F. 2004. Detection of genotyping errors by Hardy–Weinberg equilibrium testing. Eur. J. Hum. Genet. 12:395–399.

Huchon D., Delsuc F., Catzeflis F.M., Douzery E.J. 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.

Hwang S., Kim E., Lee I., Marcotte E.M. 2015. Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci. Rep. 5:17875.

ICZN. 1998. Opinion 1894. Regnum Animale..., ed. 2 (MJ Brisson, 1762): rejected for nomenclatural purposes, with the conservation of the mammalian generic names Philander (Marsupialia), Pteropus (Chiroptera), Glis, Cuniculus, and Hydrochoerus (Rodentia), Meles, Lutra and Hyaena (Carnivora), Tapirus (Perissodactyla), Tragulus and Giraffa (Artiodactyla). Bull. Zool. Nomencl. 55:64–71.

ICZN. 1999. International Code of Zoological Nomenclature: The Provisions of this Code Supersede Those of the Previous Editions with Effect from 1 January 2000. International Trust for Zoological Nomenclature.

Jackson N.D., Carstens B.C., Morales A.E., O'Meara B.C. 2017. Species delimitation with gene flow. Syst. Biol. 66:799–812.

Jun G., Flickinger M., Hetrick K.N., Romm J.M., Doheny K.F., Abecasis G.R., Boehnke M., Kang H.M. 2012. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. 91:839–848.

Kalyaanamoorthy S., Minh B.Q., Wong T.K., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Kassambara A. 2020. ggpubr:"ggplot2" based publication ready plots. R Package Version 04 0. 438.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kollár J., Poulíčková A., Dvořák P. 2022. On the relativity of species, or the probabilistic solution to the species problem. Mol. Ecol. 31:411–418.

Kumar V., Lammers F., Bidon T., Pfenninger M., Kolter L., Nilsson M.A., Janke A. 2017. The evolutionary history of bears is characterized by gene flow across species. Sci. Rep. 7:46487.

Le Verger K., Hautier L., Bardin J., Gerber S., Delsuc F., Billet G. 2020. Ontogenetic and static allometry in the skull and cranial units of nine-banded armadillos (Cingulata: Dasypodidae: Dasypus novemcinctus). Biol. J. Linn. Soc. 131:673–698.

Le Verger K., Hautier L., Gerber S., Bardin J., Delsuc F., González Ruiz L.R., Amson E., Billet G. 2023. Pervasive cranial allometry at different anatomical scales and variational levels in extant armadillos. Evolution.:qpad214.

Leaché A.D., Zhu T., Rannala B., Yang Z. 2019. The spectre of too many species. Syst. Biol. 68:168–181.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv Prepr. ArXiv13033997.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Luo A., Ling C., Ho S.Y., Zhu C.-D. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. Syst. Biol. 67:830–846.

Mallet J., Beltrán M., Neukirchen W., Linares M. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. BMC Evol. Biol. 7:1–16.

Marcgrave G. 1648. Historiae Naturalis Brasiliae. Haack Elzevier Leiden Amst.

Martin S.H., Van Belleghem S.M. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics. 206:429–438.

Mastretta-Yanes A., Arrigo N., Alvarez N., Jorgensen T.H., Piñero D., Emerson B.C. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol. Ecol. Resour. 15:28–41.

Mayr E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.

Melo-Ferreira J., Boursot P., Suchentrunk F., Ferrand N., Alves P.C. 2005. Invasion from the cold past: extensive introgression of mountain hare (Lepus timidus) mitochondrial DNA into three other hare species in northern Iberia. Mol. Ecol. 14:2459–2464.

Meyer M., Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010:pdb. prot5448.

Minh B.Q., Hahn M.W., Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R. 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.

Miranda F.R., Casali D.M., Perini F.A., Machado F.A., Santos F.R. 2018. Taxonomic review of the genus Cyclopes Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species. Zool. J. Linn. Soc. 183:687–721.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.

Moraes-Barros N., Arteaga M.C. 2015. Genetic diversity in Xenarthra and its relevance to patterns of neotropical biodiversity. J. Mammal. 96:690–702.

Nascimento F.O.D., Cheng J., Feijó A. 2021. Taxonomic revision of the pampas cat Leopardus colocola complex (Carnivora: Felidae): an integrative approach. Zool. J. Linn. Soc. 191:575–611.

Nosil P. 2008. Speciation with gene flow could be common. Mol. Ecol. 17:2101–2320.

Padial J.M., Miralles A., De la Riva I., Vences M. 2010. The integrative future of taxonomy. Front. Zool. 7:1–14.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Peters W. 1864. Ueber neue Arten der Säugethiergattungen Geomys, Haplodon und Dasypus. Monatsberichte K. Preuss. Akad. Wiss. Berl. 1865:177–181.

Petrou E.L., Drinan D.P., Kopperl R., Lepofsky D., Yang D., Moss M.L., Hauser L. 2019. Intraspecific DNA contamination distorts subtle population structure in a marine fish: Decontamination of herring samples before restriction-site associated sequencing and its effects on population genetic statistics. Mol. Ecol. Resour. 19:1131–1143.

Picard Toolkit. 2019. Broad institute, GitHub repository. .

Pompanon F., Bonin A., Bellemain E., Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6:847–859.

Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sumlin W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55:595–609.

Pontes A.R.M., Gadelha J.R., Melo E.R., De Sa F.B., Loss A.C., Júnior V.C., Costa L.P., Leite Y.L.R. 2013. A new species of porcupine, Genus Coendou (Rodentia, Erethizontidae) from the Atlantic forest of northeastern Brazil. Zootaxa. 3636:421–438.

Posit Team. 2022. RStudio: Integrated Development Environment for R; Posit Software, PBC: Boston, MA, USA,. .

Potapov V., Ong J.L. 2017. Examining sources of error in PCR by single-molecule sequencing. PLoS One. 12:e0169774.

Puechmaille S.J. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol. Ecol. Resour. 16:608–627.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., De Bakker P.I., Daly M.J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

de Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

R Core Team. 2018. R: A language and environment for statistical computing. .

Ramirez J.L., Lescroart J., Figueiró H.V., Torres-Florez J.P., Villela P.M., Coutinho L.L., Freitas P.D., Johnson W.E., Antunes A., Galetti Jr P.M. 2022. Genomic signatures of divergent ecological strategies in a recent radiation of Neotropical wild cats. Mol. Biol. Evol. 39:msac117.

Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 164:1645–1656.

Raxworthy C.J., Smith B.T. 2021. Mining museums for historical DNA: advances and challenges in museomics. Trends Ecol. Evol. 36:1049–1060.

Riesch R., Muschick M., Lindtke D., Villoutreix R., Comeault A.A., Farkas T.E., Lucek K., Hellen E., Soria-Carrasco V., Dennis S.R. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. Nat. Ecol. Evol. 1:1–13.

Robledo-Arnuncio J.J., Gaggiotti O.E. 2017. Estimating contemporary migration rates: effect and joint inference of inbreeding, null alleles and mistyping. J. Ecol. 105:49–62.

Rosel P.E., Hancock-hanser B.L., Archer F.I., Robertson K.M., Martien K.K., Leslie M.S., Berta A., Cipriano F., Viricel A., Viaud-Martinez K.A. 2017. Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. Mar. Mammal Sci. 33:76–100.

Roux C., Fraisse C., Romiguier J., Anciaux Y., Galtier N., Bierne N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 14:e2000234.

Ruedas L.A. 2017. A new species of cottontail rabbit (Lagomorpha: Leporidae: Sylvilagus) from Suriname, with comments on the taxonomy of allied taxa from northern South America. J. Mammal. 98:1042–1059.

Ruiz-García M., Chacón D., Plese T., Shostell J.M. 2020. Molecular phylogenetics of Bradypus (three-toed sloth, Pilosa: Bradypodidae, Mammalia) and phylogeography of Bradypus variegatus (brown-throated three-toed sloth) with mitochondrial gene sequences. J. Mamm. Evol. 27:461–482.

Sali A., Attali D. 2020. shinycssloaders: Add loading animations to a'shiny'output while it's recalculating. R package version 1.0. 0. CRAN R-Proj. Orgpackage Shinycssloaders.

Sievert C. 2020. Interactive web-based data visualization with R, plotly, and shiny. CRC Press.

Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12:e1005896.

Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298.

Stankowski S., Ravinet M. 2021. Defining the speciation continuum. Evolution. 75:1256–1273.

Stankowski S., Zagrodzka Z.B., Garlovsky M.D., Pal A., Shipilina D., Castillo D.G., Le Moan A., Leder E., Reeve J., Johannesson K. 2023. Selection on many loci drove the origin and spread of a key innovation. bioRxiv.:2023.02. 13.528213.

Sukumaran J., Holder M.T., Knowles L.L. 2021. Incorporating the speciation process into species delimitation. PLoS Comput. Biol. 17:e1008924.

Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. 114:1607–1612.

Suo Q., Ren-Chao Z., Yun-Qin L., Sonjai Havanond C.J., Su-Hua S. 2008. Molecular evidence for natural hybridization between Sonneratia alba and S. griffithii. J. Syst. Evol. 46:391–395.

Superina M., Pagnutti N., Abba A.M. 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.

Swofford D.L. 1998. Phylogenetic analysis using parsimony. Version 4 Sinauer Assoc. Sunderland Mass.

Tilak M.-K., Justy F., Debiais-Thibaud M., Botero-Castro F., Delsuc F., Douzery E.J. 2015. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. Conserv. Genet. Resour. 7:37–40.

Toews D.P., Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Trigo T.C., Freitas T.R.O., Kunzler G., Cardoso L., Silva J.C.R., Johnson W.E., O'Brien S.J., Bonatto S.L., Eizirik E. 2008. Inter-species hybridization among Neotropical cats of the genus Leopardus, and evidence for an introgressive hybrid zone between L. geoffroyi and L. tigrinus in southern Brazil. Mol. Ecol. 17:4317–4333.

Trigo T.C., Schneider A., de Oliveira T.G., Lehugeur L.M., Silveira L., Freitas T.R., Eizirik E. 2013. Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. Curr. Biol. 23:2528–2533.

Trindade F.J., Rodrigues M.R., Figueiró H.V., Li G., Murphy W.J., Eizirik E. 2021. Genome-wide SNPs clarify a complex radiation and support recognition of an additional cat species. Mol. Biol. Evol. 38:4987–4991.

Villanueva R.A.M., Chen Z.J. 2019. ggplot2: elegant graphics for data analysis. Taylor & Francis.

Wang R.L., Wakeley J., Hey J. 1997. Gene flow and natural selection in the origin of Drosophila pseudoobscura and close relatives. Genetics. 147:1091–1106.

Weissensteiner H., Forer L., Fendt L., Kheirkhah A., Salas A., Kronenberg F., Schoenherr S. 2021. Contamination detection in sequencing studies using the mitochondrial phylogeny. Genome Res. 31:309–316.

Wetzel R.M., Gardner A.L., Redford K.H., Eisenberg J.F. 2008. Order Cingulata. Mamm. S. Am. 1:128–156.

Wetzel R.M., Mondolfi E. 1979. The subgenera and species of long-nosed armadillos, genus Dasypus L. Vertebrate Ecology in the northern Neotropics (JF Eisenberg, ed.). Washington, DC, USA.: Smithsonian Institution Press.

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemund G., Hayes A., Henry L., Hester J. 2019. Welcome to the Tidyverse. J. Open Source Softw. 4:1686.

Wickham H., François R., Henry L., Müller K., Vaughan D. 2023. dplyr: A Grammar of Data Manipulation. R package version 1.1.4. Comput. Softw. HttpsCRAN R-Proj. Orgpackage Dplyr.

Wiens J.J., Kuczynski C.A., Stephens P.R. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. Biol. J. Linn. Soc. 99:445–461.

Wilke C.O. 2019. cowplot: streamlined plot theme and plot annotations for 'ggplot2.' R Package Version. 1.

Xu J., Turner A., Little J., Bleecker E.R., Meyers D.A. 2002. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? Hum. Genet. 111:573–574.

Yang Z. 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61:854–865.

Zachos F., Asher R. 2018. Mammalian Evolution, diversity and systematics. Walter de Gruyter GmbH & Co KG.

Zeileis A., Fisher J.C., Hornik K., Ihaka R., McWhite C.D., Murrell P., Stauffer R., Wilke C.O. 2019. colorspace: A toolbox for manipulating and assessing colors and palettes. ArXiv Prepr. ArXiv190306490.

Zhang H., Hare M.P. 2012. Identifying and reducing AFLP genotyping error: an example of tradeoffs when comparing population structure in broadcast spawning versus brooding oysters. Heredity. 108:616–625.

Zhang J., Kapli P., Pavlidis P., Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 29:2869–2876.

# Data accessibility

Data Accessibility: Raw Illumina reads have been submitted to the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and are available under BioProject number PRJNA949844. All datasets and other supplementary materials are available from the Zenodo public online repository (https://doi.org/10.5281/zenodo.7509863). Digital 3D models of the *Dasypus guianensis* holotype specimen are freely available through MorphoMuseuM (https://morphomuseum.com/specimenfiles/send-file-specimenfile/1200/4a65cc83; https://morphomuseum.com/specimenfiles/send-file-specimenfile/1201/7ed725a9).

# Supplementary Results

## Further comparison with literature

We extracted the partial 16S rRNA sequences of the four individuals of *D. s. septemcinctus*, *D. hybridus*, *D. novemcinctus* from French Guiana, *D. mazzai* and *D. sabanicola* analyzed in Abba et al. (2018) and combined them with our three mitochondrial sequences obtained for *D. novemcinctus* individuals from Argentina, Uruguay and Paraguay. *D. novemcinctus* individuals from Argentina, Uruguay, and Paraguay did not cluster with the reference mitogenome of *D. novemcinctus* from French Guiana but rather with *D. mazzai* and *D. sabanicola* (Fig. S17).

We also analyzed the partial D-loop sequences of 221 individuals from Arteaga et al. (2020) with the corresponding D-loop sequences extracted from our 19 individuals in common, allowing for direct sequence comparisons (Table S4). We found 13 individuals with incongruent sequences (Table S4). These South American individuals were all assigned to the Northern lineage in Arteaga et al. (2020) whereas the D-loop sequences that we obtained based on shotgun sequencing of the same individuals showed that they mostly belonged to the Southern lineage as expected from their geographical origin (Table S4, Fig. S16). This suggests that the original short D-loop fragment sequences obtained by PCR and Sanger sequencing in Arteaga et al. (2020) might have originated from cross-contamination with other samples of the Northern lineage.

## Mito-nuclear phylogenetic discordance and introgression in nine-banded armadillos

We used two complementary approaches to investigate whether introgression could explain the mito-nuclear discordances we recovered (Fig. 1a-b). First, we inferred phylogenetic networks with the nuclear data under a MSC model with introgression using the software SNaQ (Solís-Lemus and Ané 2016) implemented in the package PhyloNetworks (Solís-Lemus et al. 2017). *D. kappleri* was used as an outgroup, and *D. pilosus* was excluded because of its high amount of missing data. Concordance factors were computed in PhyloNetworks from the 832 gene trees previously used in the Astral analysis. SNaQ was then run with maximum numbers of reticulations (hmax) ranging from 0 to 3, with ten independent ML searches for each. This analysis supported the presence of introgression among *D. novemcinctus* lineages (Fig. S9a-b): introducing one or two reticulations improved the fit of the models compared to a strictly bifurcating tree (average loglik scores of 1.23 and 0.37 vs 2.45, respectively). Even when allowing three reticulations, a maximum of two were recovered. Because the fit of the best models with 1 and 2 reticulations were similar (loglik scores of 0.53 and 0.22, respectively), and increasing the number of parameters in the model might result in overfitting, we considered both as our best models. They recovered the same reticulation, indicating introgression between the ancestors of the Southern lineage and the "west Andes group" (*D. pilosus*, Central and Northern lineages). The major edge of the network supported the Southern lineage as sister to the Guianan lineage (inheritance values [$\gamma$] = 0.83 and 0.75, respectively). The network with two reticulations additionally supported introgression between Western Mexico populations and the Central lineage, with the major edge supporting the monophyly of the Mexican populations ($\gamma$ = 0.88).

To quantify the extent to which these reticulations are supported by our set of gene trees, we used a descriptive approach based on topologies. Because of the low resolution of the gene trees, we used a Topology Weighting (TW) approach to account for non-monophyly of the lineages (Martin and Van Belleghem 2017). We performed TW in two three-taxa sets: 1) Central, Southern and Guianan lineages, (to test for the reticulation in the first network); and 2) Western Mexico (Mexican individuals with Central lineage mitochondrial haplotypes), Eastern Mexico and Central lineages (to test for the additional reticulation in

the second network). In each case, we selected gene trees in which all focal individuals were represented (642 and 699 gene trees, respectively), and pruned them to keep the focal lineages and *D. kappleri* as an outgroup, using the ape R package (Paradis and Schliep 2019). TW was run using TWISST (Martin and Van Belleghem 2017). The results were first represented as the distribution of the weights for each of the three possible topologies. Furthermore, we followed Stankowski et al. (2023) and plotted the joint distribution of the three topologies in a ternary diagram using ggtern (Hamilton 2018). We investigated asymmetry in the ternary diagram as a signal of introgression: under neutral processes without gene flow, we expect that the two alternative topologies (i.e., those with the lowest support) would obtain similar weights, similarly to the assumption of the "ABBA-BABA" test (Green et al. 2010; Durand et al. 2011). We quantified asymmetry in the ternary diagrams using the DLR statistic of Stankowski et al. (2023): considering NT1 as the number of trees in the right half of the ternary diagram (i.e., the weight of the first alternative topology is higher than that of the second) and NT2 the number in the left half, then $DLR = \frac{NT1-(0.5\times(NT1+NT2))}{0.5\times(NT1+NT2}$. Significance was assessed using permutation tests with 100,000 permutations. The results of these analyses were concordant with the networks in that the gene tree topologies representing the edges of the network were more common than the alternative discordant topology (Fig. S9c-d). When investigating the position of the Guianan lineage with respect to the Southern and Central lineages, we found little difference between the nuclear and mitochondrial topologies (average weights of 0.37 and 0.35, respectively), while the third topology was slightly less common (average weight of 0.28). Consequently, the joint ternary distribution of the weights was significantly asymmetric towards the mitochondrial topology (DLR = 0.14, p-value = 0.0002). We also found that the gene trees were poorly resolved, with few lending full support (i.e., weight = 1) to any topology. The second analysis, focused on the Northern and Central lineages, recovered a different pattern. In this case, the nuclear topology (i.e., monophyletic Mexican populations) received high support (average weight of 0.69), while the discordant topologies had similar weights (average weights of 0.15 and 0.16 for the control and mitochondrial topologies, respectively). Nethertheless, we detected a significant asymmetry towards the mitochondrial topology (DLR = 0.20, p-value = 0.0002) but this signal is driven by topologies intermediate between the nuclear and mitochondrial topologies (i.e., in the upper half of the triangle) rather than strongly supporting the mitochondrial topology.

To conclude, these results show that two instances of mito-nulcear discordance are con-

sistent with signals of introgression in the nuclear data. However, the informativeness of our data is limited and the extent to which introgression affects the nuclear data is unclear, particularly in the case of the Western Mexican populations. Moreover, we focused only on two introgression events that coincided with mito-nuclear discordances, but introgression may affect other branches as well. A comprehensive understanding of ancient introgression in the *Dasypus* radiation will necessitate the use of larger genomic datasets (e.g., Whole Genome Sequencing).

The shifting position of *D. guianensis* between the mitochondrial and nuclear phylogenies may reflect incomplete lineage sorting (ILS) of mitochondrial haplotypes at the root of the radiation. However, we found strong support for the mitochondrial topology in the nuclear gene trees compared to expectations under stochastic coalescent processes (Fig. 2c). This result is in accord with our phylogenetic network analyses (Fig. 2a), which recovered a reticulation between the ancestors of *D. novemcinctus* and the "west Andes group" (*D. pilosus*, *D. fenestratus* and *D. mexicanus*; Fig. 2a). This would be consistent with mitochondrial transfer from the latter into the former at early stages of divergence, leading to the mitochondrial phylogenetic pattern we observed. This scenario would also conform with the biogeographic expectation, as the nuclear topology supports two clades separated by the Central and Northern Andes. Similarly, the presence of numerous *D. fenestratus* mitochondrial haplotypes in Western Mexico (Arteaga et al. 2012, 2020) is unlikely to be explained by ILS, which is not expected to generate geographically consistent patterns (Toews and Brelsford 2012). Accordingly, both gene tree distributions and phylogenetic networks supported introgression between *D. fenestratus* and the Western Mexico populations (Figure 2). A potential scenario underlying this pattern may have involved a southward expansion of *D. mexicanus* into the range of *D. fenestratus*, with *D. fenestratus* mitochondrial haplotypes persisting locally in *D. mexicanus* populations, as documented in Iberian hares for instance (Melo-Ferreira et al. 2005; Alves et al. 2008). Nevertheless, the power of our data remains limited in terms of introgression detection and further genomic analyses will be needed to understand the phylogeographic history of the *Dasypus* complex.

**Table S1:** List of biological samples with detailed information on their origin, nature, and use in this study. Samples of D. novemcinctus were classified as belonging to Southern, Guianan, Central, and Northern lineages following Hautier et al. (2017) and were colored accordingly.

| Samples | Taxon | Origin | Mito | Capture | Voucher | Collection date | DNA extraction |
|---|---|---|---|---|---|---|---|
| DKA-T4339 | *D. kappleri* | French Guiana | X | X | NA | 2003 | Gibb et al. 2016 |
| DKA-T3365 | *D. kappleri* | French Guiana | X | X | NA | 2001 | This study |
| DHY-ZVCM2010* | *D. septemcinctus hybridus* | Uruguay | X | Excluded | ZVC M2010 | 1976 | Gibb et al. 2016 |
| DSE-T3002 | *D. septemcinctus septemcinctus* | Argentina | X | X | NA | 2001 | Gibb et al. 2016 |
| DSE-PAP42 | *D. septemcinctus septemcinctus* | Brazil | X | X | NA | 1996 | This study |
| DPI-LSUMZ21888* | *D. pilosus* | Peru | X | X | LSUMZ Mammals 21888 | 1978 | Gibb et al. 2016 |
| DPI-MSB49990* | *D. pilosus* | Peru | X | X | MSB:Mamm:49990 | 1980 | Gibb et al. 2016 |
| DNO-T4553 | *D. novemcinctus novemcinctus* | French Guiana | X | NA | JAG M5909 | 2003 | This study |
| DNO-71 | *D. novemcinctus novemcinctus* | French Guiana | X | X | JAG M5361 | 1995 | This study |
| DNO-T4557 | *D. novemcinctus novemcinctus* | French Guiana | X | X | JAG M5910 | 2003 | This study |
| DNO-T2863 | *D. novemcinctus novemcinctus* | French Guiana | X | X | JAG M5911 | 2000 | This study |
| DNO-169 | *D. novemcinctus novemcinctus* | French Guiana | X | NA | JAG M5442 | 1995 | This study |
| DNO-T1863 | *D. novemcinctus novemcinctus* | French Guiana | X | X | JAG M5912 | 1997 | Gibb et al. 2016 |
| DNO-T1692 | *D. novemcinctus novemcinctus* | French Guiana | X | X | MNHN-ZM-MO-2000-226 | 1995 | This study |
| DNO-T2465 | *D. novemcinctus novemcinctus* | French Guiana | X | X | MNHN-ZM-MO-2001-1530 | 2000 | This study |
| DNO-MC333* | *D. novemcinctus novemcinctus* | Guyana | X | X | AMNH Mammals 42914 | 1920 | This study# |
| DNO-AP249 | *D. novemcinctus novemcinctus* | Brazil | X | X | IEPA 3102 | 2006 | This study |
| DNO-AP109 | *D. novemcinctus novemcinctus* | Brazil | X | Excluded | IEPA 2419 | 2005 | This study |
| DNO-AP207 | *D. novemcinctus novemcinctus* | Brazil | X | X | IEPA 2837 | 2006 | This study |
| DNO-MC151* | *D. novemcinctus mexicanus* | Mexico | X | X | HMAM532 | 2005 | Arteaga et al. 2012 |
| DNO-AJR1 | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2001 | This study |
| DNO-MC88* | *D. novemcinctus mexicanus* | Mexico | X | X | CNMA 37069 | 1992 | Arteaga et al. 2012 |
| DNO-MC21* | *D. novemcinctus davisi* | Mexico | X | X | ENCB-IPN 26577 | 1986 | Arteaga et al. 2012 |
| DNO-MC176* | *D. novemcinctus davisi* | Mexico | X | X | NA | 2007 | Arteaga et al. 2012 |
| DNO-MC165 | *D. novemcinctus fenestratus* | Costa Rica | X | X | NA | 2008 | Arteaga et al. 2012 |
| DNO-MC425* | *D. novemcinctus fenestratus* | Panama | X | X | USNM Mammals 578437 | 1990 | Arteaga et al. 2020 |
| DNO-MC393* | *D. novemcinctus fenestratus* | Venezuela | X | X | USNM Mammals 442790 | 1968 | This study# |
| DNO-MC323* | *D. novemcinctus aequatorialis* | Ecuador | X | X | AMNH Mammals 46555 | 1920 | This study# |
| DNO-MC180* | *D. novemcinctus aequatorialis* | Ecuador | X | X | NA | 2008 | Arteaga et al. 2020 |
| DNO-MC181* | *D. novemcinctus aequatorialis* | Colombia | X | X | NA | 2008 | Arteaga et al. 2012 |
| DNO-PH8* | *D. novemcinctus aequatorialis* | Colombia | X | X | AMNH Mammals 32356 | 1911 | This study |
| DNO-MC257* | *D. novemcinctus aequatorialis* | Colombia | X | X | USNM Mammals 281285 | 1943 | This study# |
| DNO-NC001821 | *D. novemcinctus mexicanus* | USA | X | X | NA | NA | Arnason et al. (1997) |
| DNO-MVZ192698 | *D. novemcinctus mexicanus* | USA | X | X | MVZ:Mamm:192698 | 1974 | This study |
| DNO-BM1 | *D. novemcinctus mexicanus* | USA | X | NA | NA | 2000 | This study |
| DNO-MC292* | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2009 | Arteaga et al. 2012 |
| DNO-MC103* | *D. novemcinctus mexicanus* | Mexico | X | X | Eco-SC-M 980 | 2000 | Arteaga et al. 2012 |
| DNO-NP276 | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2001 | This study |
| DNO-LS5 | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2000 | This study |
| DNO-MC40* | *D. novemcinctus mexicanus* | Mexico | X | Contam | NA | 2007 | Arteaga et al. 2012 |
| DNO-MC58* | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2003 | Arteaga et al. 2012 |
| DNO-YUC1 | *D. novemcinctus mexicanus* | Mexico | X | X | NA | 2001 | This study |
| DNO-MC163* | *D. novemcinctus mexicanus* | Guatemala | X | NA | NA | 2008 | Arteaga et al. 2012 |
| DNO-T2631 | *D. novemcinctus mexicanus* | Belize | X | X | NA | 2000 | This study |
| DNO-JP1 | *D. novemcinctus mexicanus* | El Salvador | X | NA | NA | 2000 | This study |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DNO-MC186* | *D. novemcinctus davisi* | Mexico | Excluded | NA | CNMA 16551 | 1974 | Arteaga et al. 2012 |
| DYE-MLP30.III.90.2 | *D. mazzai* | Argentina | X | X | MLP 30.III.90.2 | 1988 | Gibb et al. 2016 |
| DNO-PH6* | *D. novemcinctus novemcinctus* | Brazil | X | Excluded | AMNH Mammals 93116 | 1930 | This study |
| DNO-MC367* | *D. novemcinctus novemcinctus* | Brazil | X | Excluded | AMNH Mammals 133330 | 1937 | This study# |
| DNO-AM78 | *D. novemcinctus novemcinctus* | Brazil | X | X | IEPA 2560 | 2008 | This study |
| DNO-MC366* | *D. novemcinctus novemcinctus* | Brazil | X | X | AMNH Mammals 133274 | 1936 | This study# |
| DNO-MC413* | *D. novemcinctus novemcinctus* | Brazil | X | X | USNM Mammals 259485 | 1935 | Arteaga et al. 2020 |
| DNO-PH3* | *D. novemcinctus novemcinctus* | Brazil | X | X | AMNH Mammals 133266 | 1937 | This study |
| DNO-MC355* | *D. novemcinctus novemcinctus* | Brazil | X | Excluded | AMNH Mammals 91709 | 1930 | This study# |
| DNO-M275 | *D. novemcinctus novemcinctus* | Brazil | X | X | INPA SISTAP M275 | NA | This study |
| DNO-MC1000* | *D. novemcinctus novemcinctus* | Guyana | X | Contam | AMNH Mammals 42441 | 1919 | This study# |
| DNO-T2360 | *D. novemcinctus novemcinctus* | Argentina | X | X | NA | 2000 | This study |
| DNO-MC327* | *D. novemcinctus novemcinctus* | Uruguay | X | Excluded | AMNH Mammals 205727 | 1963 | This study# |
| DNO-T1870 | *D. novemcinctus novemcinctus* | Paraguay | X | X | NA | 1998 | This study |
| DNO-AN1 | *D. novemcinctus novemcinctus* | Bolivia | X | X | NA | 2000 | This study |
| DNO-MC340* | *D. novemcinctus novemcinctus* | Bolivia | X | X | AMNH Mammals247662 | 1980 | This study# |
| DNO-MSB2 | *D. novemcinctus novemcinctus* | Bolivia | X | X | MSB:Mamm:99077 | 2000 | This study |
| DNO-MC436* | *D. novemcinctus novemcinctus* | Bolivia | X | NA | AMNH Mammals 211674 | 1965 | Arteaga et al. 2020 |
| DNO-MC324* | *D. novemcinctus novemcinctus* | Ecuador | X | X | AMNH Mammals 67710 | 1924 | This study# |
| DNO-MC226* | *D. novemcinctus novemcinctus* | Colombia | X | X | AMNH Mammals 136252 | 1939 | This study# |
| DNO-MC334* | *D. novemcinctus novemcinctus* | Peru | X | X | AMNH Mammals 98462 | 1930 | Arteaga et al. 2020 |
| DNO-MC441* | *D. novemcinctus novemcinctus* | Peru | X | Excluded | AMNH Mammals 98463 | 1930 | Arteaga et al. 2020 |
| DNO-NP751 | *D. novemcinctus novemcinctus* | Peru | X | NA | NA | 2002 | This study |
| DNO-MC337* | *D. novemcinctus novemcinctus* | Peru | X | X | AMNH Mammals 98465 | 1927 | This study# |
| DNO-MC326* | *D. novemcinctus novemcinctus* | Trinidad & Tobago | X | X | AMNH Mammals 7549 | 1894 | This study# |
| DNO-MC401* | *D. novemcinctus novemcinctus* | Venezuela | X | X | USNM Mammals 442799 | 1968 | This study# |
| DNO-MC399* | *D. novemcinctus novemcinctus* | Venezuela | X | X | USNM Mammals 406706 | 1967 | Arteaga et al. 2020 |
| DNO-MC302* | *D. novemcinctus novemcinctus* | Venezuela | X | X | AMNH Mammals 144828 | 1946 | This study# |
| DNO-MC304* | *D. novemcinctus novemcinctus* | Venezuela | X | X | AMNH Mammals 33151 | 1906 | This study# |
| DNO-T3380 | *D. novemcinctus novemcinctus* | Venezuela | X | Excluded | NA | 2001 | This study |
| DNO-T1921 | *D. novemcinctus novemcinctus* | Venezuela | X | NA | NA | 1998 | This study |
| DNO-MC402* | *D. novemcinctus novemcinctus* | Venezuela | X | X | USNM Mammals 296615 | 1952 | Arteaga et al. 2020 |
| DNO-MC392* | *D. novemcinctus novemcinctus* | Venezuela | X | X | USNM Mammals 442787 | 1968 | This study# |
| DNO-MC417* | *D. novemcinctus novemcinctus* | Venezuela | X | X | USNM Mammals 406703 | 1967 | This study# |
| DNO-MC303* | *D. novemcinctus novemcinctus* | Venezuela | X | Excluded | AMNH Mammals 78517 | 1929 | This study# |
| DSA-USNM372834* | *D. sabanicola* | Venezuela | X | X | USNM Mammals 372834 | 1966 | Gibb et al. 2016 |

Notes: Taxon names are based on Wetzel et al. (2008) and Feijó et al. (2018). *: specimens sampled as dried skin pieces. #: museum specimens previously extracted in Arteaga et al. (2012; 2020) and re-extracted for this study. NA: Not Available. AMNH: American Museum of Natural History, New York, USA; CNMA: Colección Nacional de Mamíferos, Universidad Nacional Autónoma de México, México; Éco-SC-M: Colección Mastozoológica, Colegio de la Frontera Sur, México; ENCB-IPN: Colección Mastozoológica, Instituto Politécnico Nacional, México; HMAM: Colección Mastozoología Instituto Agropecuario de Hidalgo, México; IEPA: Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá, Macapá, Brazil; JAG: collection JAGUARS, ONG Kwata, Institut Pasteur (Cayenne, French Guiana); LSUMZ: Louisiana State University Museum of Natural Science, Baton Rouge, USA; MNHN: Muséum National d'Histoire Naturelle, Paris, France; MSB: Museum of Southwestern Biology, Albuquerque, USA; MLP: Museo de La Plata, La Plata, Argentina; MVZ: Museum of Vertebrate Zoology, Berkeley, USA; USNM: National Museum of Natural History, Washington, USA; ZVC: Colección de Vertebrados de la Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.

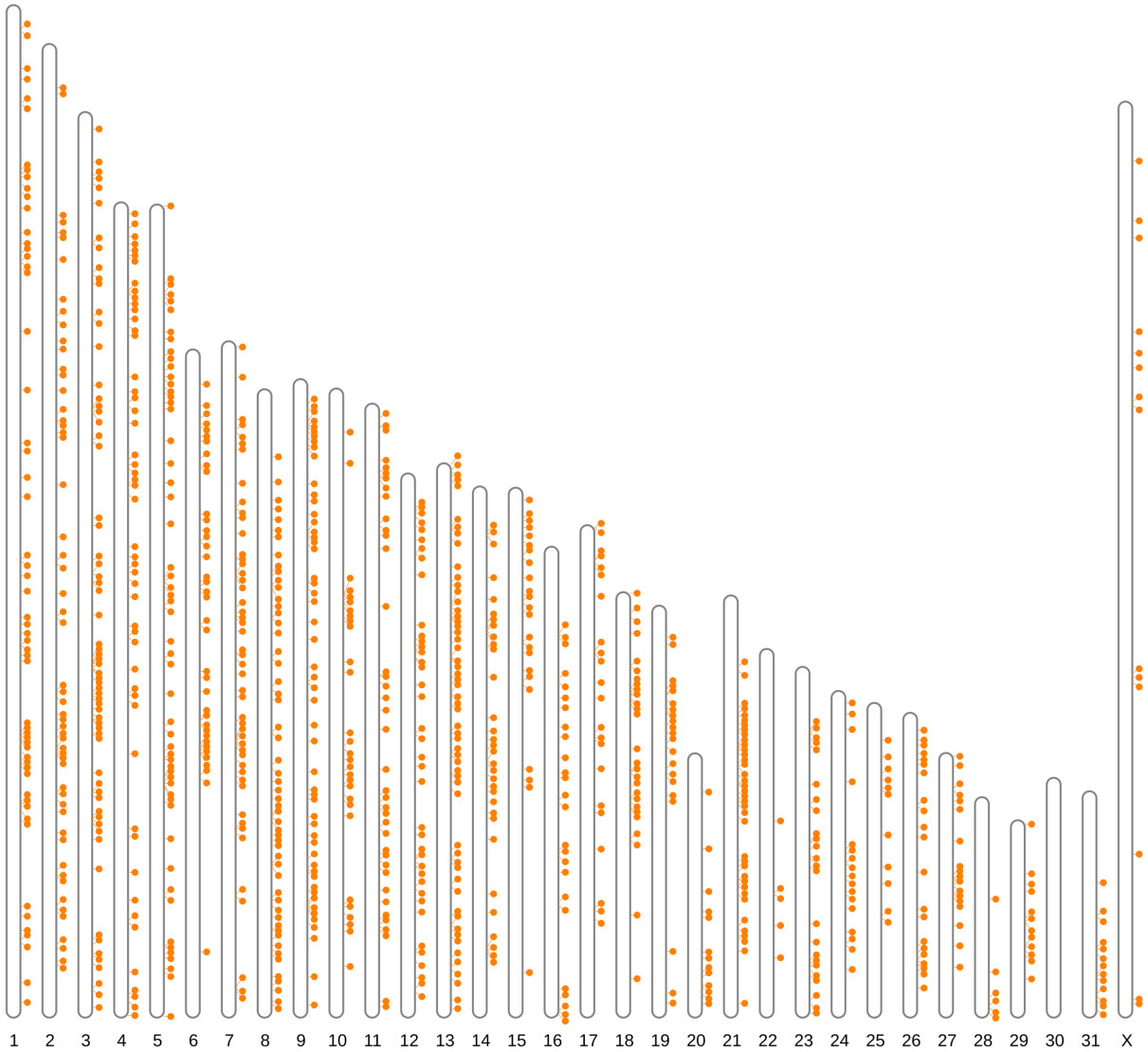


**Figure S1:** Distribution of 984 targeted nuclear loci (out of the 997 originally targeted) along the chromosome scale assembly of *Dasypus novemcinctus* (mDasNov1.hap2) realized with the R package RIdeogram (Hao et al. 2020).

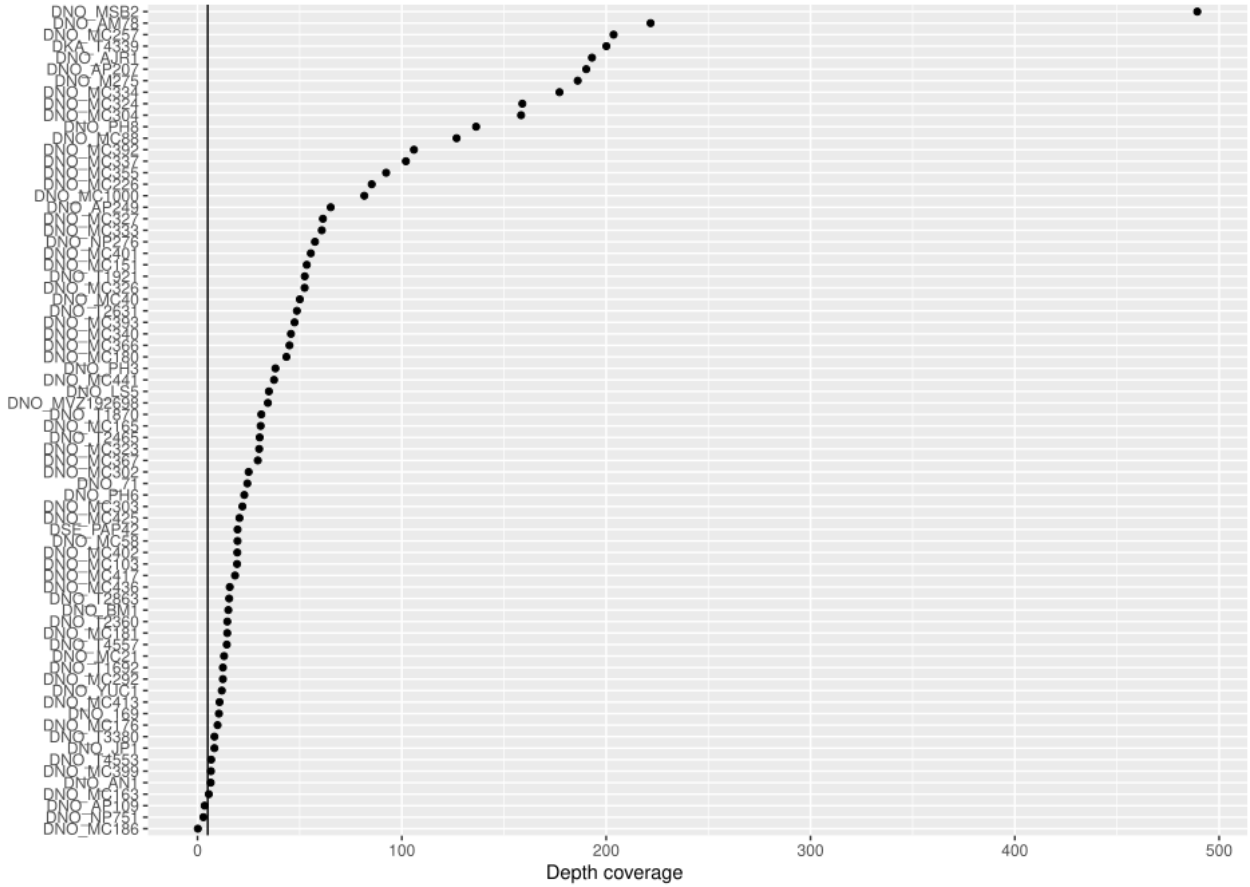


**Figure S2:** Mitochondrial genome depth of coverage for the 72 *Dasypus* individuals newly sequenced in this study. The black line corresponds to a 5x depth of coverage.

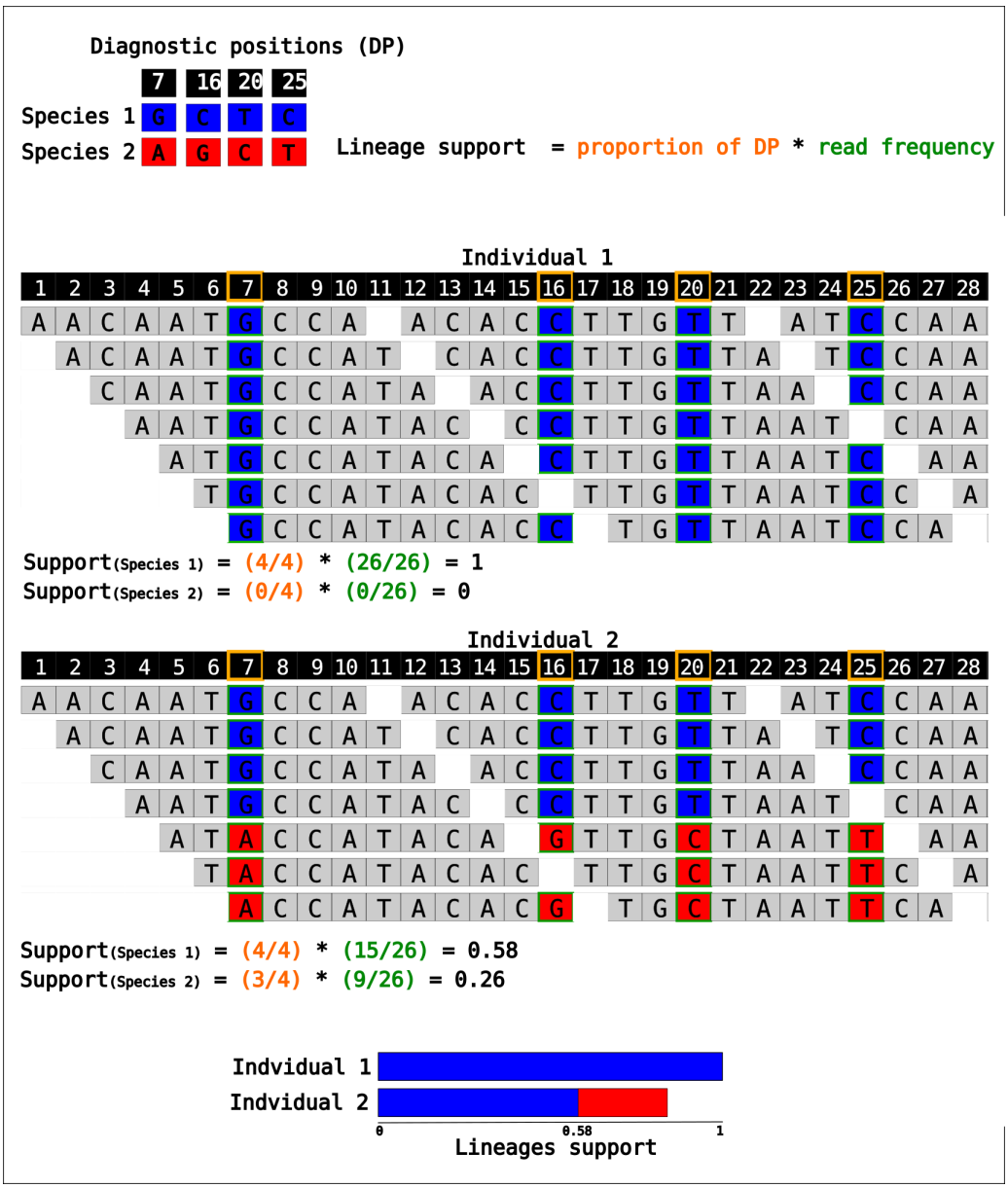


**Figure S3:** Calculation of mitochondrial lineage support for detecting contamination. Positions 7, 16, 20 and 25 are diagnostic positions (DP) that distinguish species 1 from species 2. Lineage support is estimated as the product of proportion of DP and read frequency. Individual 1 supports all four DP of species 1 (proportion of DP = 4/4) and each position is supported by all reads (total read frequency = 26/26). In individual 2, reads support the four DP of species 1. However, only three of the four DP support species 2 as the position 16 is excluded (supported by less than three reads). However, DP of species 1 are more strongly supported by reads (total reads frequency (species 1) = 15/26; total reads frequency (species 2) = 9/26). This pattern can be interpreted as cross-contamination. In some cases, reads could support a different haplotype in a DP position (e.g. individual variation, genotyping errors or contamination with something other than the lineages tested), such reads are not recorded, participating to a total lineage support lower than 1.

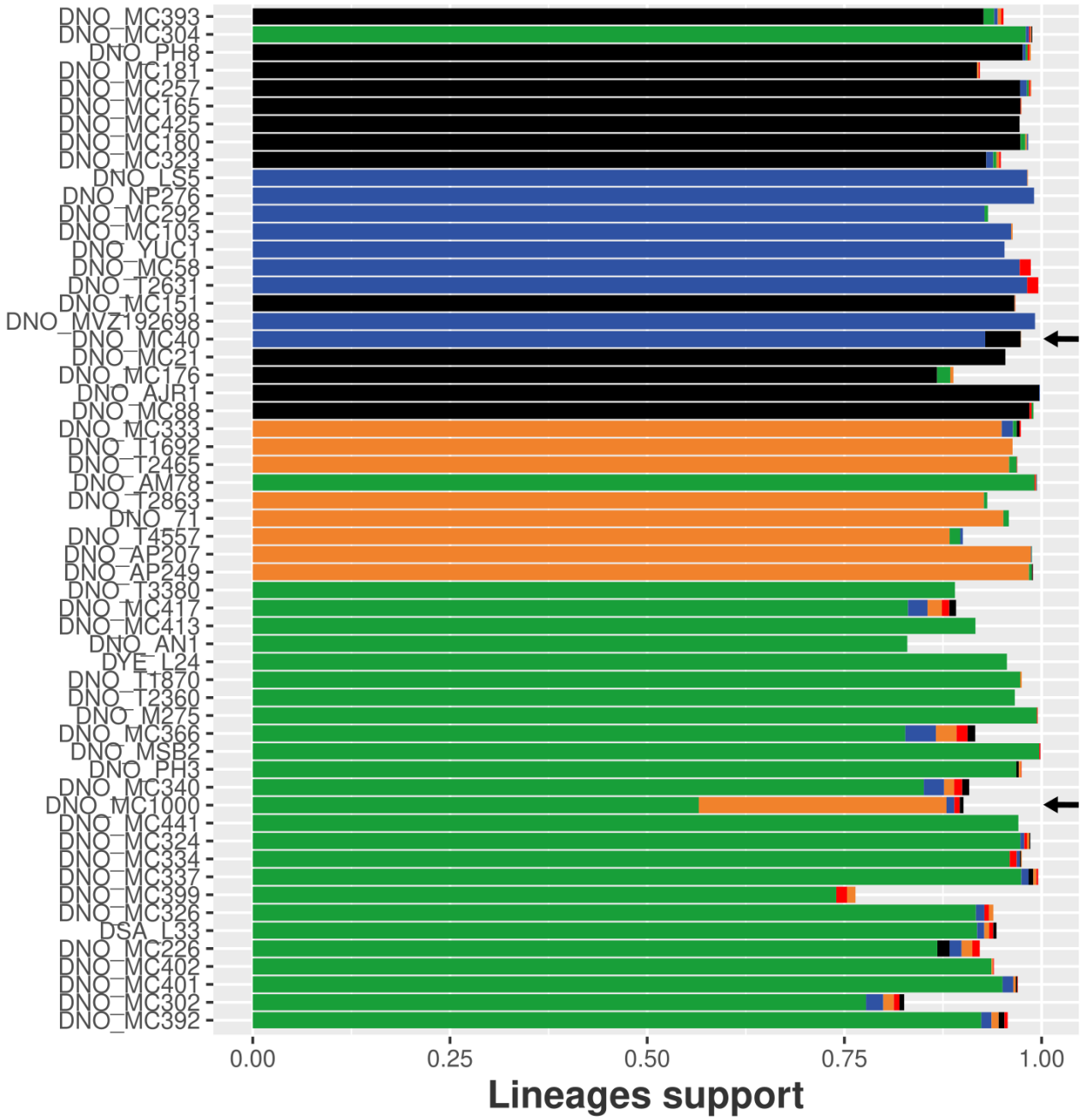


**Figure S4:** Mitochondrial lineage support for each individual (see Fig. S3 for the calculation method). Reads supporting synapomorphies of the Northern lineage are in blue, Central lineage in black, Southern lineage in green, Guianan lineage in orange, and *Dasypus pilosus* in red. Reads supporting DP of *D. pilosus* (in red) within individuals of *D. novemcinctus* complex cannot be explained by cross-contamination as DNA extractions and library preparations for these individuals were performed separately from those of the *D. novemcinctus* complex. They can therefore serve as a control for the proportion of shared genotyping errors expected between individuals that have not cross-contaminated each other. Arrows indicate two individuals (DNO-MC1000 and DNO-MC40) suspected to be cross-contaminated.

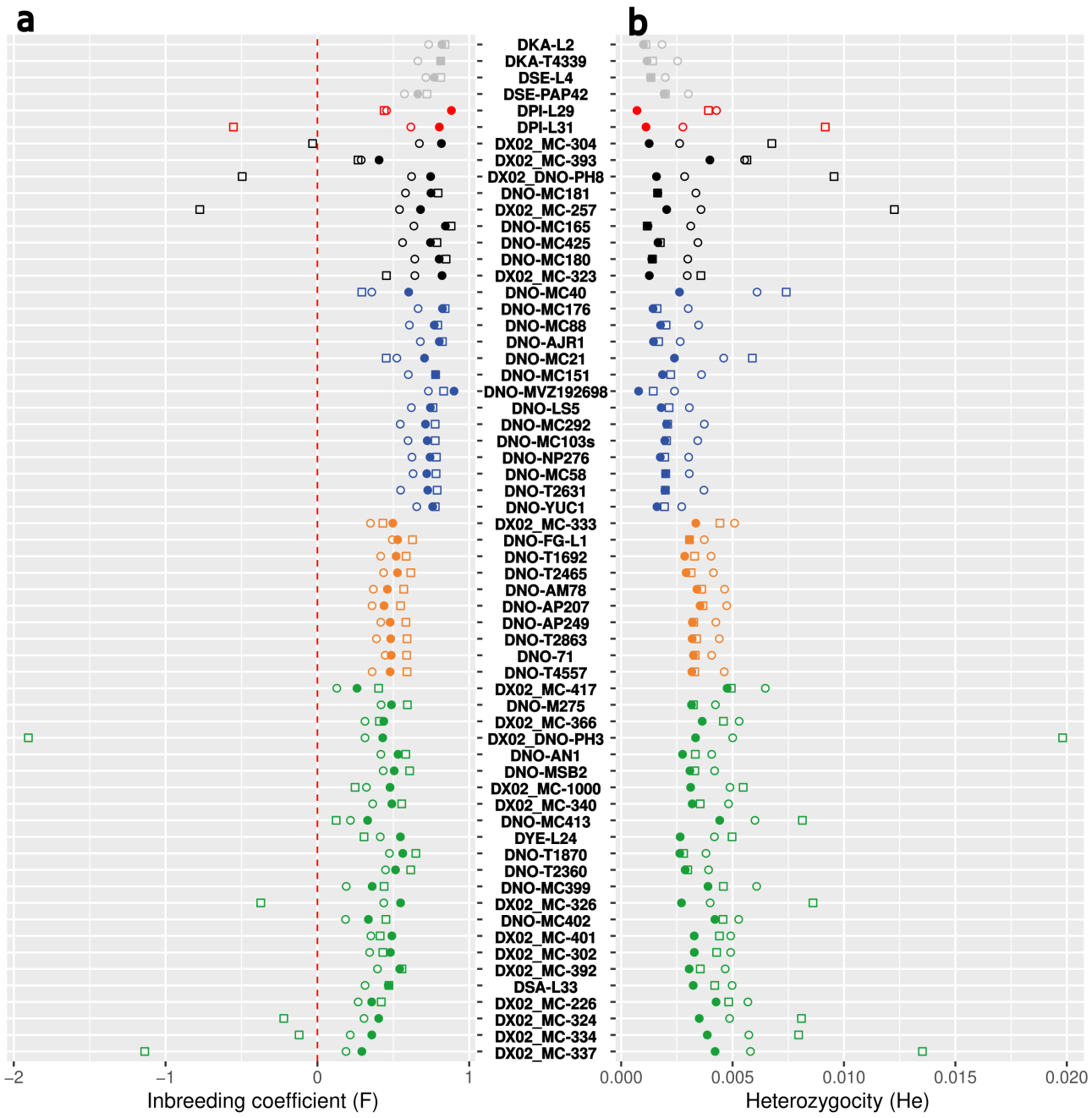


**Figure S5:** a) Inbreeding coefficient and b) heterozygosity estimate for individuals according to cleaning steps. Squares correspond to initial data, open circles to data corrected to ensure that heterozygous positions are supported by a proportion of reads between 0.3 and 0.7, and filled circles to final data in which 159 potential paralogous loci were excluded.

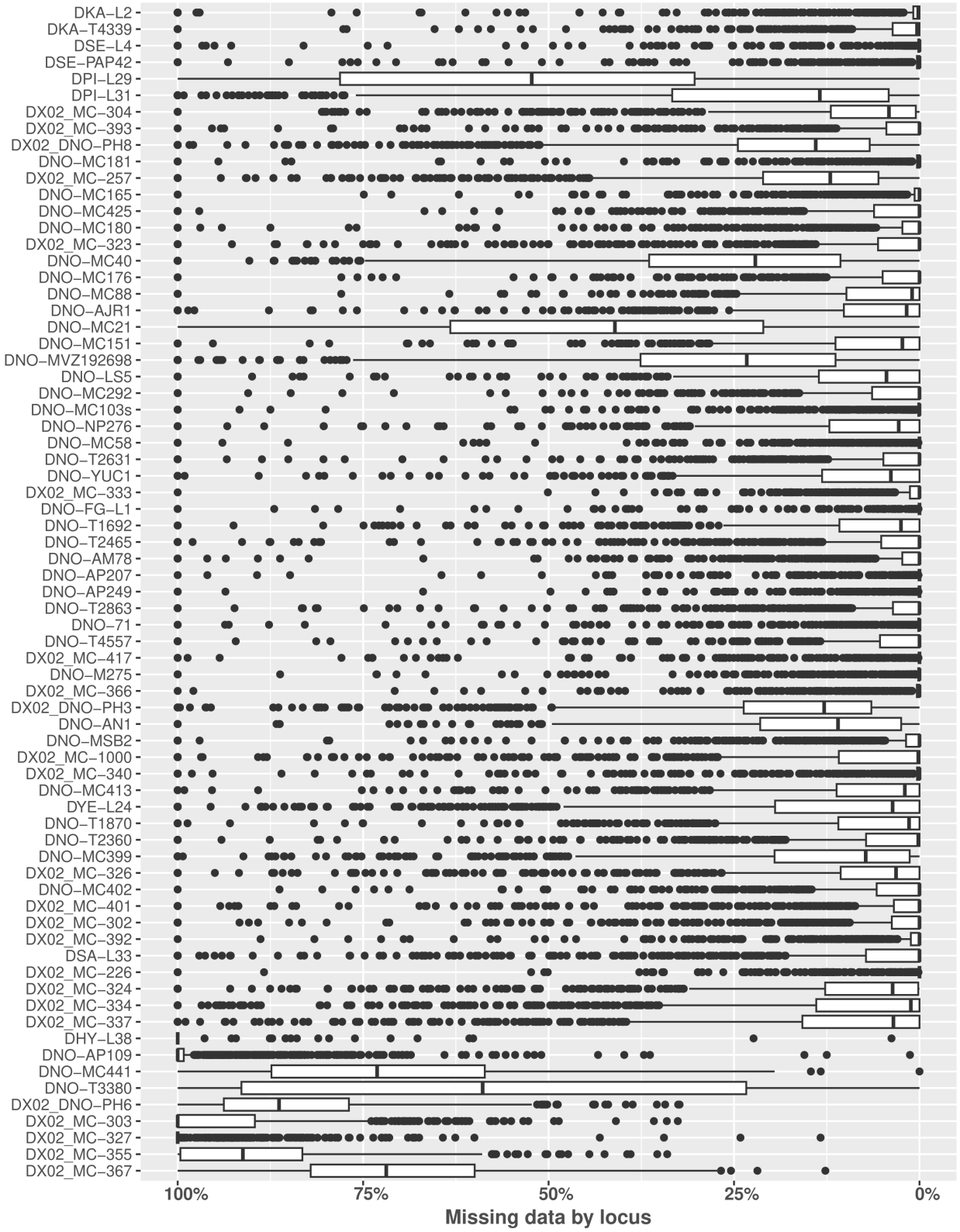


**Figure S6:** Percentage of missing data by captured locus for each *Dasypus* individual. Loci with more than 40% average missing and nine individuals with more than 55% average missing data across loci were excluded based on this analysis.

**Table S2:** Quality statistics a) by locus and b) by individuals after filtering out loci with more than 60% average missing data, for the nine individuals with more than 45% average missing data across loci and the 159 loci with high deviation from Hardy-Weinberg expectation, respectively. a) Table extract (see `https://zenodo.org/deposit/7509864` for the full table)

| Individuals | Mean depth coverage | Mean missing data | Number of loci |
|---|---|---|---|
| DKA-L2 | 145.47216196319 | 0.019545135226994 | 815 |
| DKA-T4339 | 128.282704257908 | 0.0337373971411189 | 822 |
| DNO-71 | 164.064652141983 | 0.0188200833047739 | 817 |
| DNO-AJR1 | 74.4479456140351 | 0.0584562478571429 | 798 |
| DNO-AM78 | 125.471584615385 | 0.0325786748962149 | 819 |
| DNO-AN1 | 45.8460666666667 | 0.118296186221374 | 786 |
| DNO-AP207 | 172.251841425121 | 0.0137938810507249 | 828 |
| DNO-AP249 | 289.19898686747 | 0.0151699203493979 | 830 |
| DNO-FG-L1 | 248.097744363636 | 0.0108809179151519 | 825 |

b) Table extract (see `https://zenodo.org/deposit/7509864` for the full table)

| Locus | Mean depth coverage | Mean missing data | Number of individuals |
|---|---|---|---|
| Dasypus_novemcinctus_ABCA1 | 154.10323 | 0.0119297385 | 60 |
| Dasypus_novemcinctus_ABCA12 | 112.0884 | 0.0325755145901639 | 61 |
| Dasypus_novemcinctus_ABCA3 | 18.2854066666667 | 0.3316624 | 15 |
| Dasypus_novemcinctus_ABCA4 | 162.991165573771 | 0.00678754475409793 | 61 |
| Dasypus_novemcinctus_ABCA9 | 152.206083870968 | 0.0256872637096769 | 62 |
| Dasypus_novemcinctus_ABCB10 | 50.3231 | 0.342132666666667 | 3 |
| Dasypus_novemcinctus_ABCB4 | 142.327971666667 | 0.00945504799999997 | 60 |
| Dasypus_novemcinctus_ABCB5 | 111.291563333333 | 0.0890815195 | 60 |
| Dasypus_novemcinctus_ABCC12 | 80.6293703703704 | 0.0384984033333329 | 54 |

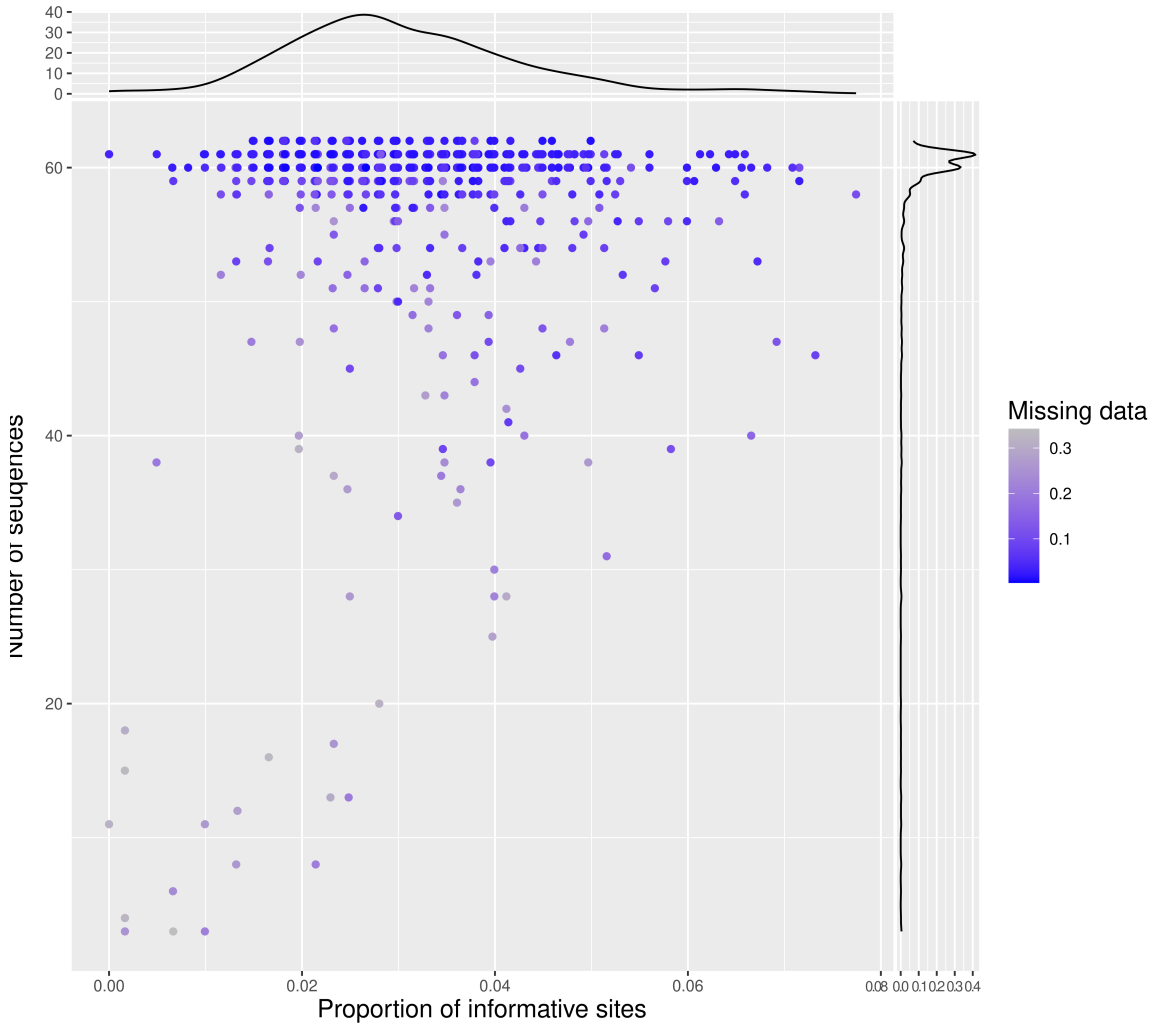


**Figure S7:** Summary information of the 837 cleaned loci representing the number of sequences and the proportion of variable sites according to the percentage of missing data.

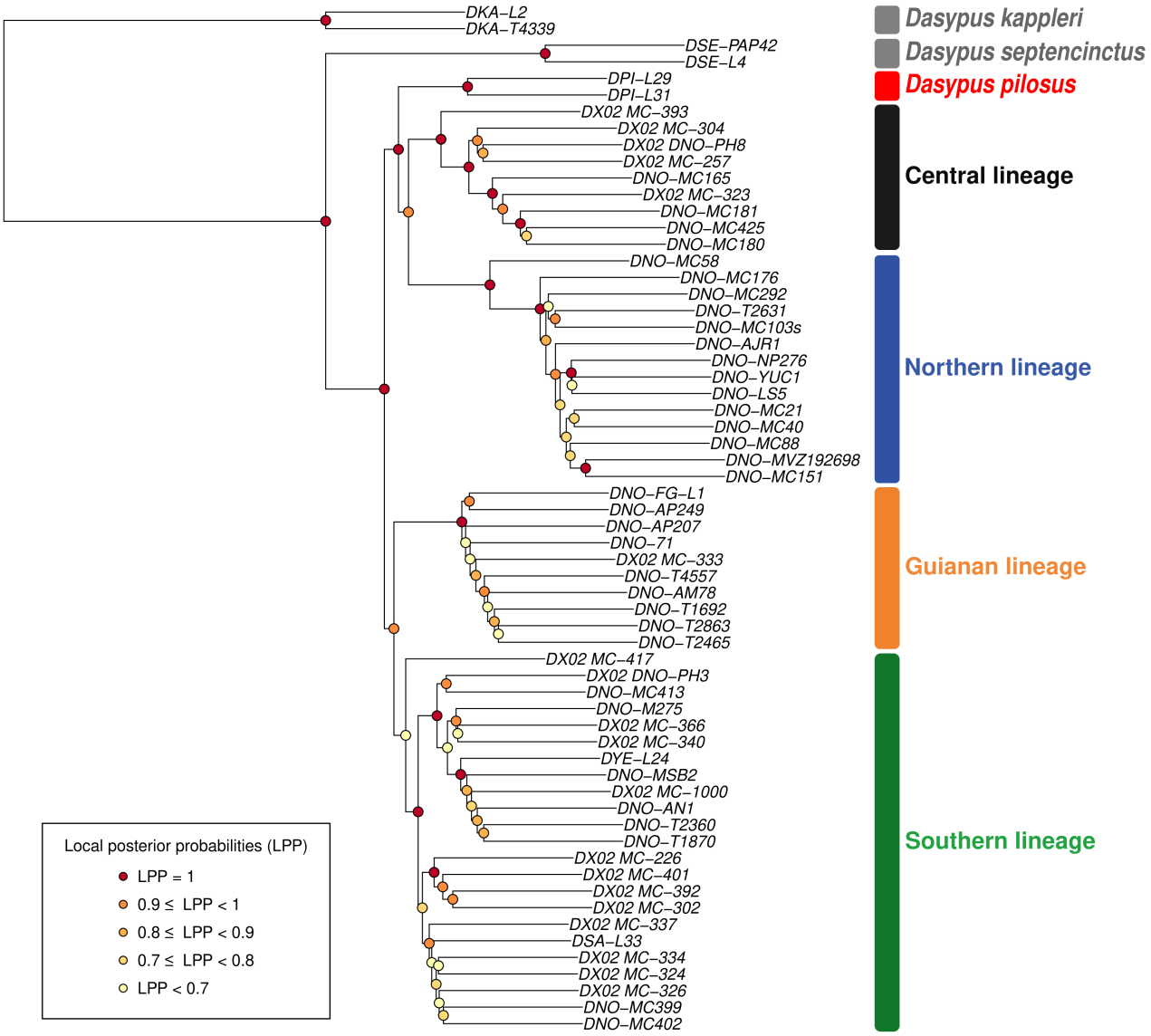


**Figure S8:** Phylogenetic relationships of the 62 *Dasypus* individuals obtained using Astral on the 832 ML gene trees from the captured nuclear loci reconstructed with IQ-Tree and ModelFinder. Node circles are coloured according to the local posterior probabilities (LPP).

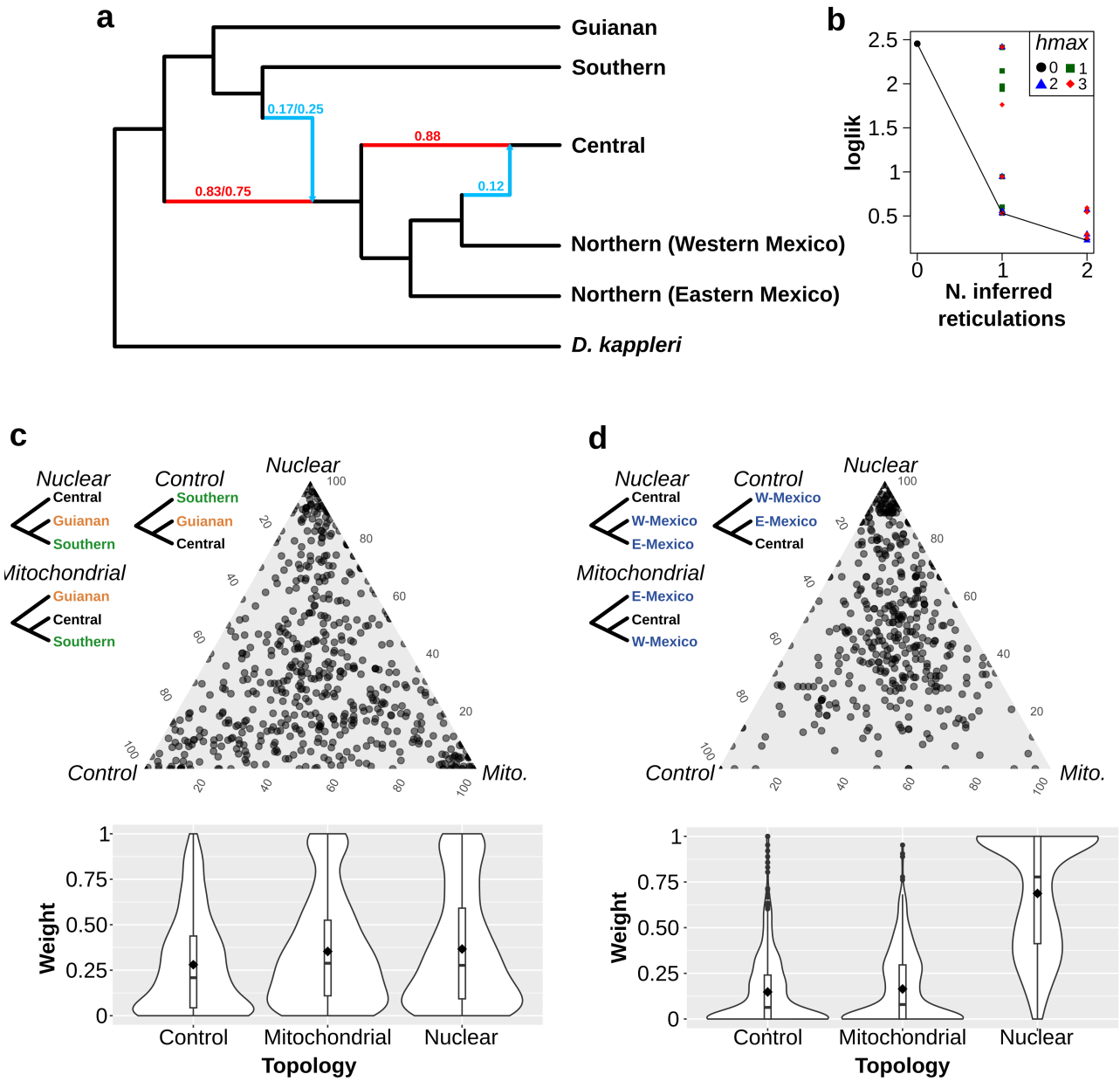


**Figure S9:** Results of analyses to detect introgression. a) Best phylogenetic network inferred with SNaQ, with two reticulations. Number at hybrid edges denote inheritance values ($\gamma$). Values at the deepest reticulation indicate $\gamma$ estimated in the second best (1 reticulation) and best networks, respectively. Major edges are indicated in red and minor edges in blue. b) Likelihood scores of the inferred networks as a function of the number of reticulations. The color and shape of the dots denote the maximum number of reticulations allowed (hmax), and the line shows the likelihood score (loglik) of the best model for 0, 1 and 2 reticulations. c-d) Results of topology weighting analyses focusing on the position of the Guianan lineages and the Western Mexico populations, respectively. Top panels show the joint distribution of the weights at the three possible topologies in a ternary diagram. Bottom panels show violin plots of the distribution of each topology individually, with diamonds indicating the average weights.

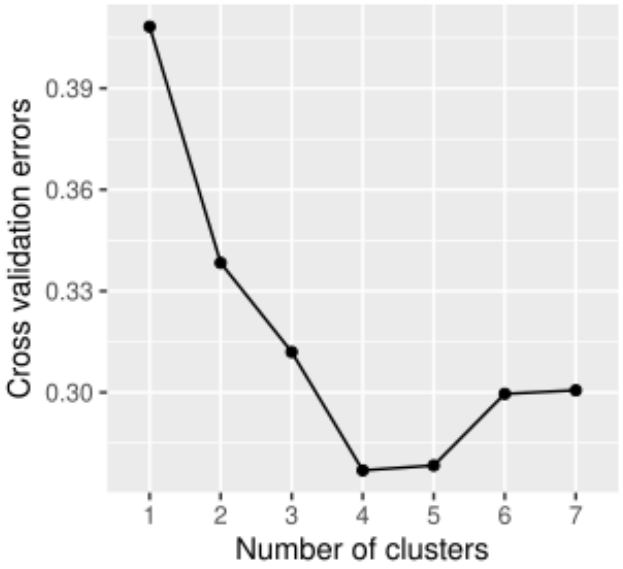


**Figure S10:** Cross validation errors according to the number of clusters (K) investigated. The lower the error value, the higher the support for the cluster number.

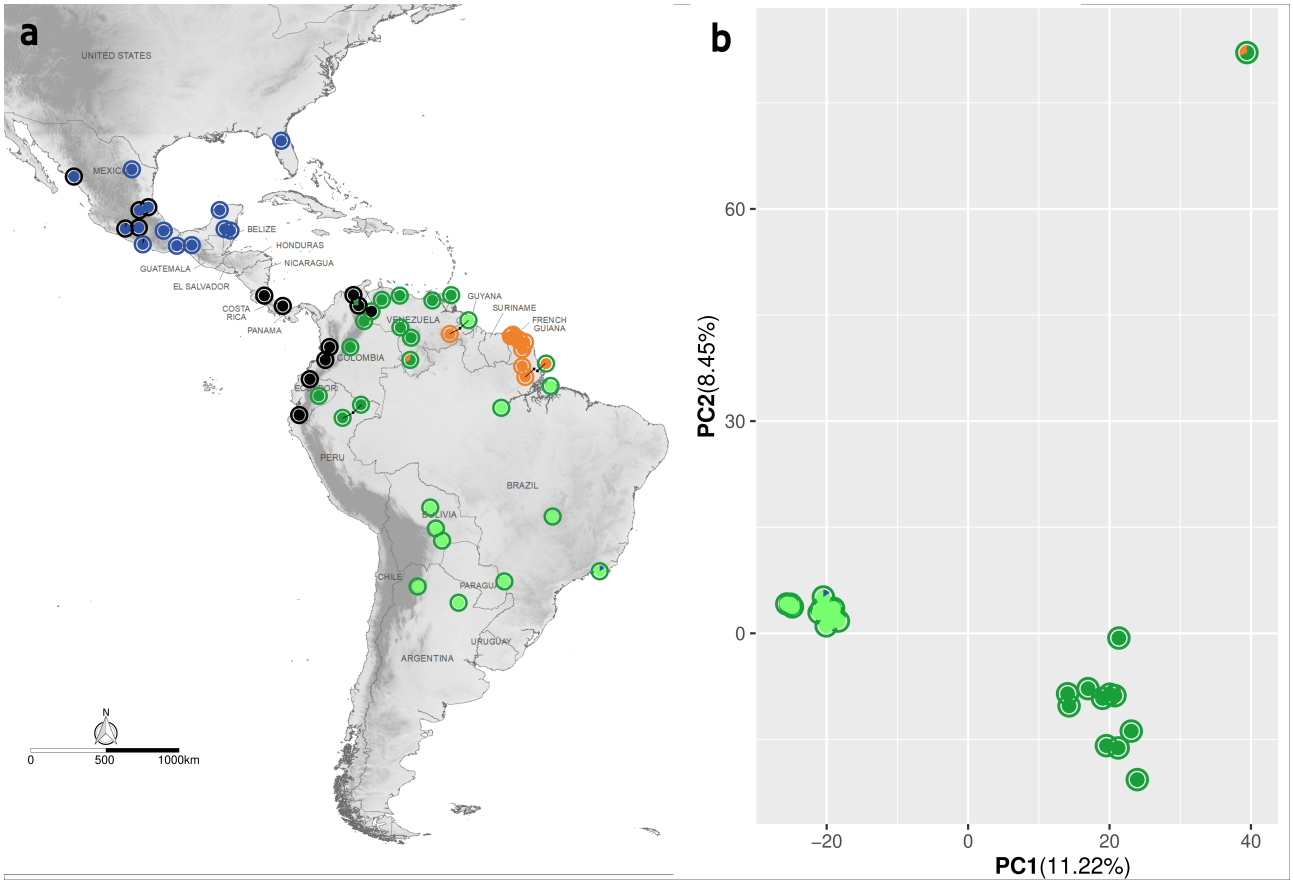


**Figure S11:** Detailed analysis of the sub-structure within the newly recognized *D. novemcinctus* (Southern lineage). a) Distribution and genome composition detailed for all individuals. Outer circles represent mitochondrial lineages and the pie charts represent the assignment probability estimated by the Admixture analysis with K = 5. b) PCA of genetic variance conducted on 4182 SNPs for the 24 individuals composing the Southern lineage.

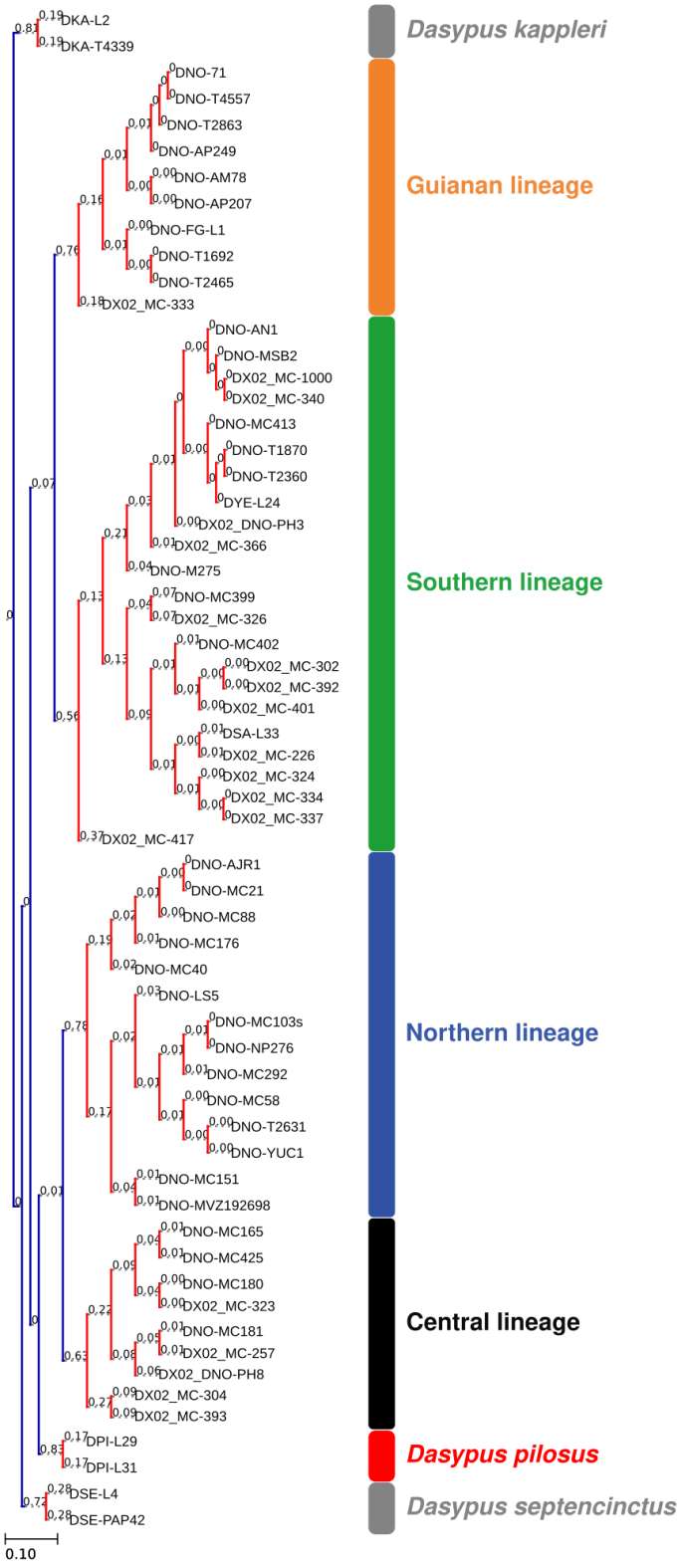


**Figure S12:** Species delimitation estimated by bPTP-h. The species tree reconstructed using IQ-TREE and ModelFinder from the 837 concatenated nuclear loci illustrates the most supported partition found by bPTP. Clades in red correspond to taxa identified as distinct species by bPTP.

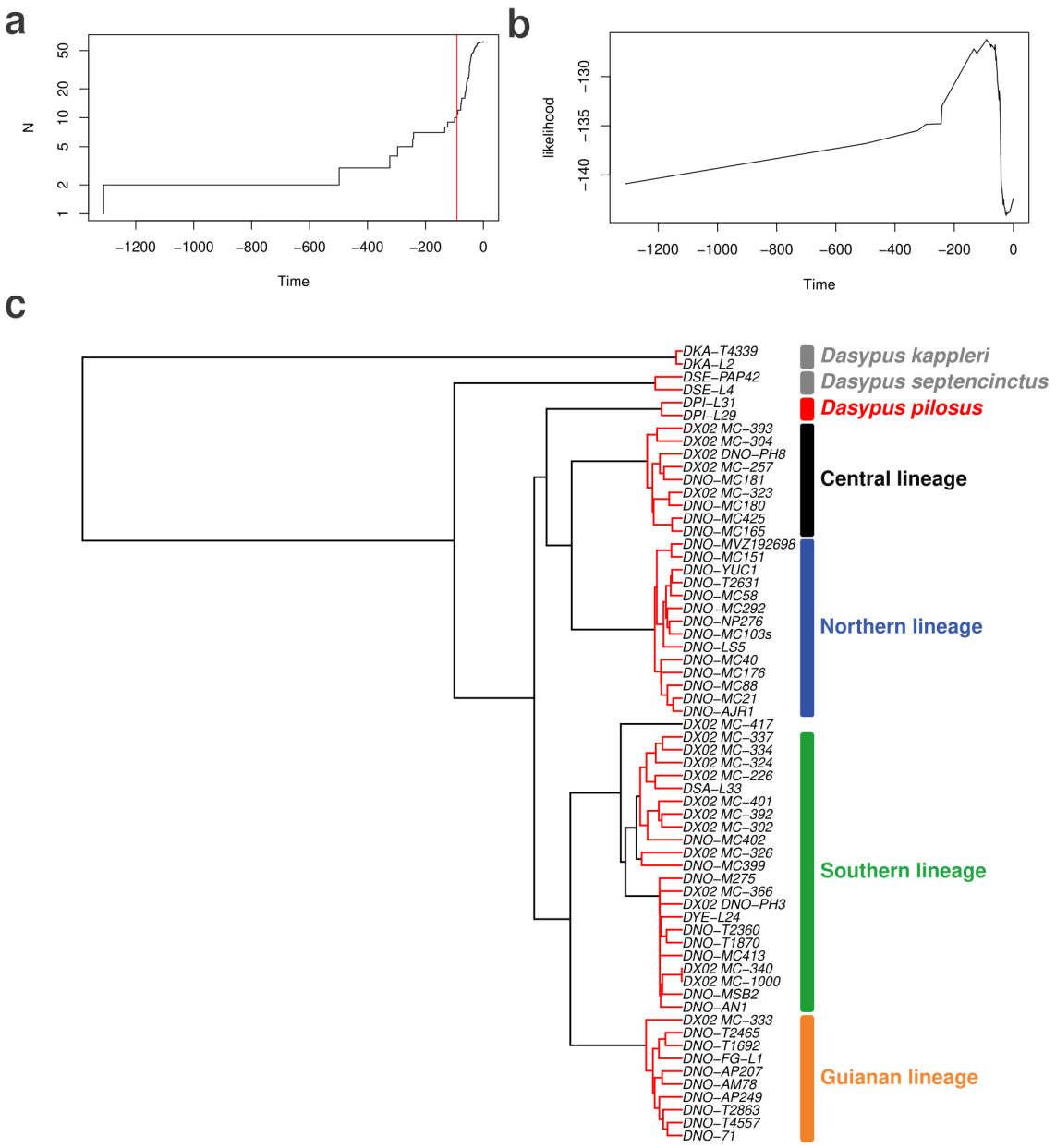


**Figure S13:** Species delimitation estimated using GMYC. a) Number of lineages through time. The red line corresponds to the threshold estimated by GMYC. b) Likelihood profile through time. c) Ultrametric reconstruction of the concatenated sequences of the 837 nuclear loci using IQ-TREE and ModelFinder for each partition. Clades in red represent taxa supported as distinct species by GMYC.

**Table S3:** The full table is available at `https://zenodo.org/deposit/7509864\`

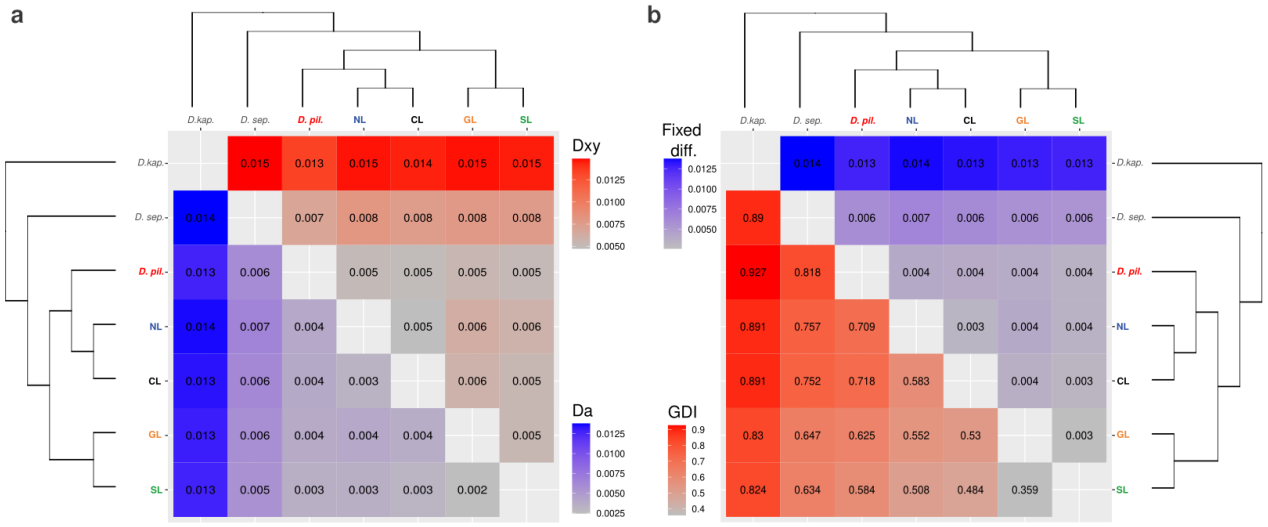


**Figure S14:** Heatmaps of pairwise genetic indexes between lineages. a) Absolute measure of genetic divergence ($D_{xy}$; red) and the net measure of genetic divergence ($D_a$; blue). b) Rate of fixed differences (blue) and the Genomic Differentiation Index (GDI; red).

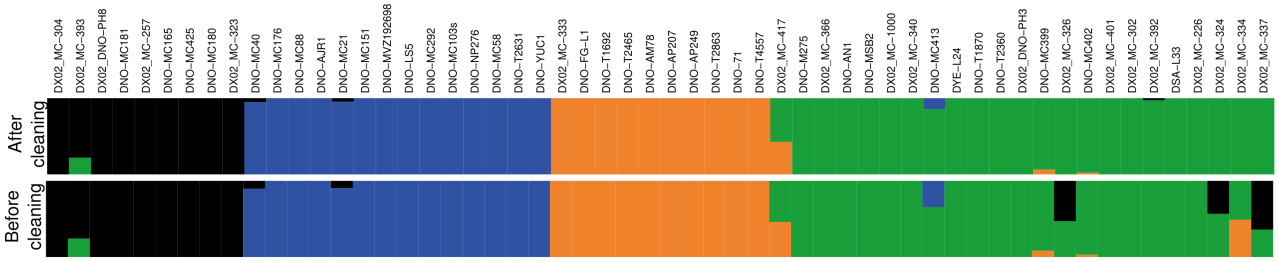


**Figure S15:** Effect of filters on admixture results.

**Table S4:** Nineteen individuals from Arteaga et al. (2020) were re-extracted and resequenced using shotgun sequencing in our study. The comparison revealed several individuals with a different D-loop sequence. This table details the lineage to which these individuals belong according to Arteaga et al. (2020) and to our mitogenomic results, their sampling locality, and their corresponding haplotype in Arteaga et al. (2020), with the total number of individuals associated with each haplotype in parentheses.

| Individuals | Lineage according to Arteaga et al. (2020) | Lineage according to this study | Locality | Haplotype from Arteaga et al. (2020) |
|---|---|---|---|---|
| MC257 | Northern lineage | Central Lineage | Colombia | H54 (26) |
| MC323 | Northern lineage | Central Lineage | Ecuador | Dn33 (1) |
| MC333 | Northern lineage | Guianan Lineage | Guyana | Dn32 (1) |
| MC326 | Northern lineage | Southern Lineage | Trinidad & Tobago | Dn26 (2) |
| MC302 | Northern lineage | Southern Lineage | Venezuela | H54 (26) |
| MC304 | Northern lineage | Southern Lineage | Venezuela | H54 (26) |
| MC303 | Northern lineage | Southern Lineage | Venezuela | H54 (26) |
| MC324 | Northern lineage | Southern Lineage | Ecuador | Dn30 (6) |
| MC367 | Northern lineage | Southern Lineage | Brazil | Dn30 (6) |
| MC355 | Northern lineage | Southern Lineage | Brazil | Dn35 (4) |
| MC366 | Northern lineage | Southern Lineage | Brazil | Dn34 (1) |
| MC327 | Northern lineage | Southern Lineage | Uruguay | Dn35 (4) |
| MC337 | Northern lineage | Southern Lineage | Peru | Dn36 (2) |
| MC226 | Northern lineage | Southern Lineage | Colombia | Dn30 (6) |
| MC340 | Southern Lineage | Southern Lineage | Bolivia | Dn13 (1) |
| MC392 | Southern Lineage | Southern Lineage | Venezuela | Dn22 (1) |
| MC401 | Southern Lineage | Southern Lineage | Venezuela | Dn21 (1) |
| MC417 | Southern Lineage | Southern Lineage | Venezuela | Dn7 (1) |
| MC393 | Central Lineage | Central Lineage | Venezuela | Dn3 (1) |

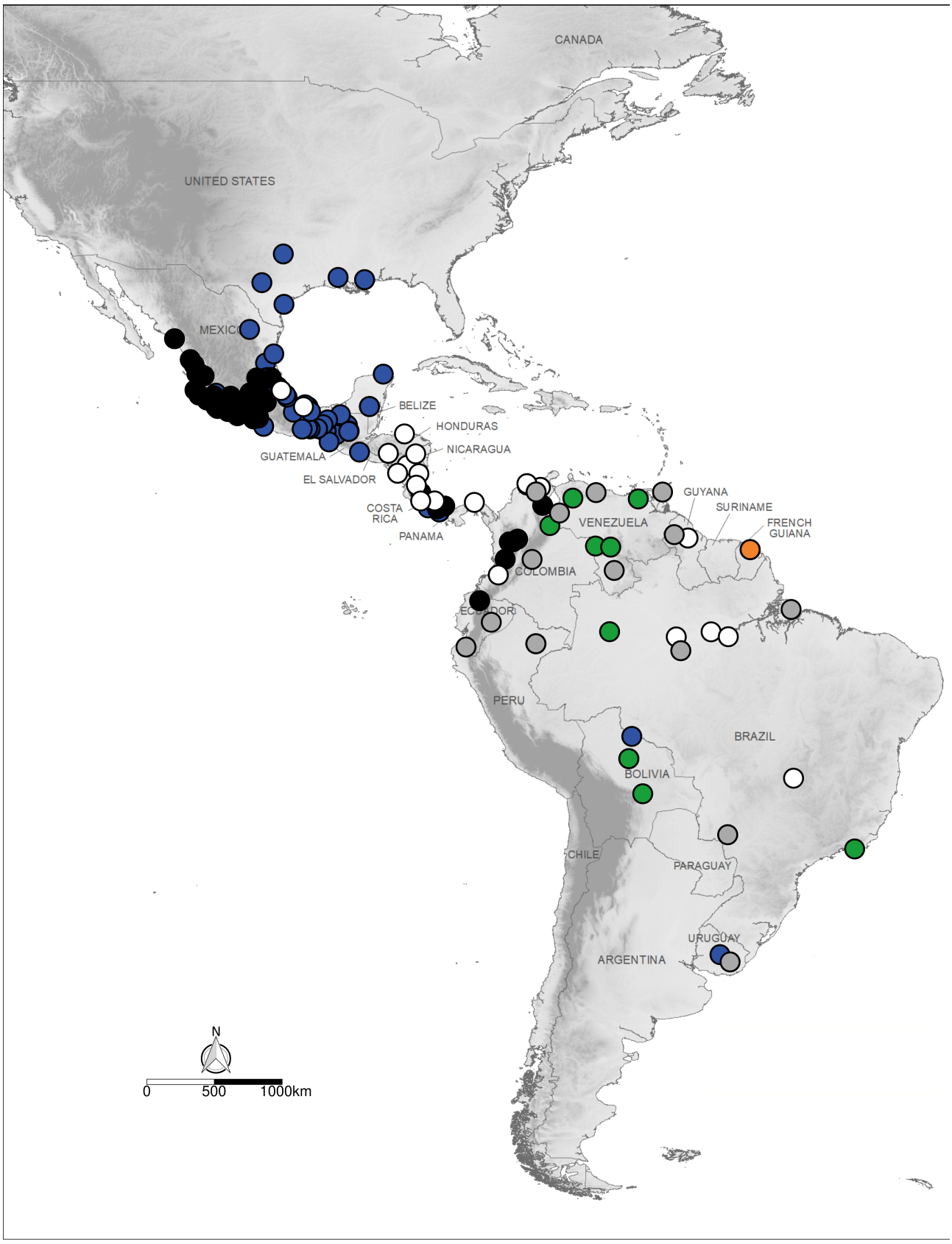


**Figure S16:** Updated map from Arteaga et al. (2020). Coloured dots correspond to the lineages as identified in the original study. Gray dots are individuals re-extracted and re-sequenced in our study that were likely misassigned. White dots correspond to individuals from a haplotype that contains at least one individual likely to have been misassigned.

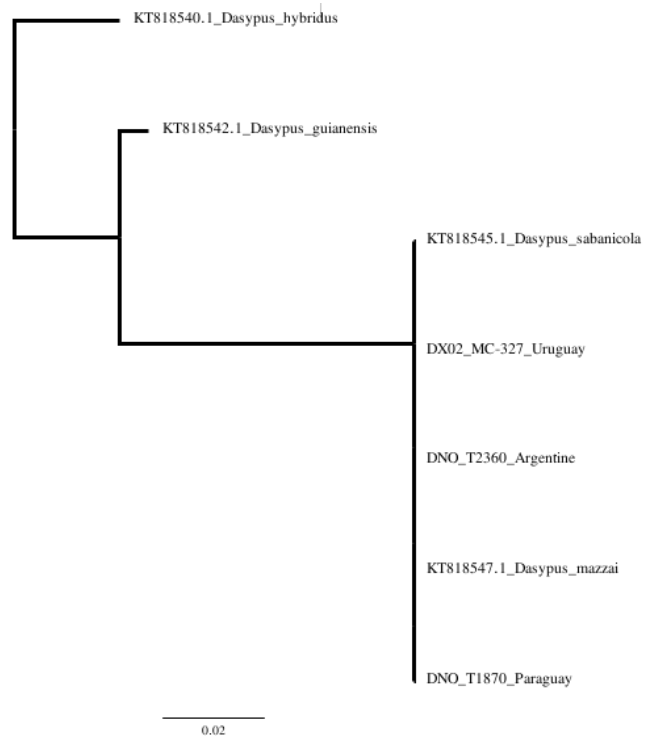


**Figure S17:** Maximum likelihood phylogenetic tree of 212 pb of the 16s ribosomal RNA (same as analyzed in Abba et al., (2018)) reconstructed with IQ-Tree. Individuals of *D. hybridus*, *D. guianensis* sp. nov. , *D. mazzai*, and *D. sabanicola* are the ones used in their study. Our *D. novemcinctus* individuals from Argentina, Paraguay and Uruguay group with *D. mazzai* and *D. sabanicola* instead of *D. guianensis* as reported in their study.

**Table S5:** Adult cranial measurements (in millimeters) of the four Dasypus species recognized in this study following Feijó & Cordeiro-Estrela (2016). Measurements are given as Mean ± standard deviation, minimum and maximum values (number of specimens) are based on specimens examined by Feijó et al. (2018).

| *D. mexicanus* | *D. fenestratus* | *D. guianensis* | *D. novemcinctus* | |
|---|---|---|---|---|
| greatest length of skull | 97.91 ± 3.62 , 85.78 - 104.86 (59) | 98.46 ± 6.2 , 86.47 - 107.61 (19) | 106.40 ± 7.27 , 96.96–128.31 (26) | 96.58 ± 4.92 , 81.23 - 108.61 (533) |
| condylobasal length | 89.45 ± 3.58 , 76.64 - 96.30 (59) | 89.2 ± 5.73 , 77.54 - 96.41 (19) | 98.26 ± 6.79 , 88.74–117.45 (26) | 88.68 ± 4.69 , 73.63 - 99.86 (533) |
| anterior palatal length | 21.63 ± 1.30 , 18.74 - 24.62 (59) | 22.12 ± 1.85 , 19.12 - 25.6 (19) | 23.92 ± 1.74 , 20.54–28.21 (26) | 21.81 ± 1.74 , 14.37 - 27.05 (533) |
| palatal length | 64.23 ± 2.94 , 53.37 - 69.78 (59) | 63.56 ± 4.37 , 54.41 - 69.68 (19) | 71.84 ± 5.86 , 63.78–90.16 (26) | 63.95 ± 3.76 , 52.54 - 74.68 (533) |
| maxilla length | 38.10 ± 2.31 , 30.39 - 42.89 (59) | 40.99 ± 3.2 , 35.91 - 47.04 (19) | 42.39 ± 4.48 , 36.65–54.23 (26) | 37.03 ± 2.67 , 26.68 - 44.92 (533) |
| palatine Length | 16.84 ± 1.47 , 13.21 - 19.96 (59) | 13.91 ± 1.75 , 10.7 - 17.6 (19) | 19.32 ± 3 , 15.33–28 (26) | 18.51 ± 1.93 , 11.37 - 25.44 (531) |
| infraorbital canal length | 9.42 ± 1.24 , 6.52 - 12.26 (57) | 6.58 ± 0.84 , 5.38 - 8.14 (18) | 6.47 ± 1.05 , 3.61–8.40 (26) | 7.31 ± 1.16 , 3.18 - 10.49 (531) |
| maxillary toothrow length | 23.66 ± 1.60 , 19.81 - 27.57 (59) | 22.52 ± 2.34 , 16.83 - 25.01 (19) | 25.07 ± 1.90 , 22.20–30.88 (26) | 23.33 ± 1.70 , 18.25 - 29.12 (533) |
| nasal length | 34.29 ± 2.29 , 26.02 - 38.82 (59) | 34.92 ± 2.71 , 28.79 - 38.8 (19) | 36.53 ± 3.11 , 30.75–42.70 (26) | 31.63 ± 2.77 , 22.86 - 39.03 (529) |
| lacrimal length | 10.64 ± 1.42 , 6.62 - 13.67 (59) | 10.3 ± 1.72 , 7.03 - 12.76 (19) | 13.03 ± 1.5 , 10.38–17.38 (26) | 10.39 ± 1.27 , 7.22 - 14.56 (532) |
| rostral length | 59.93 ± 2.50 , 52.36 - 66.00 (59) | 60.64 ± 4.99 , 50.37 - 67.75 (19) | 66.32 ± 5.54 , 57.85–82.71 (26) | 58.09 ± 3.78 , 47.15 - 68.17 (532) |
| anteorbital breadth | 32.96 ± 1.58 , 29.13 - 37.05 (59) | 33.66 ± 2.38 , 29.69 - 38.37 (19) | 36.20 ± 2.39 , 30.5–40.61 (26) | 32.16 ± 1.90 , 23.43 - 37.98 (533) |
| palatal breadth | 15.31 ± 0.91 , 12.95 - 17.39 (59) | 15.47 ± 2.09 , 13.13 - 22.07 (19) | 16.12 ± 2.25 , 11.05–20.15 (26) | 14.23 ± 2.18 , 8.15 - 18.75 (532) |
| palatine breadth | 13.41 ± 0.91 , 11.43 - 16.21 (59) | 14.2 ± 1.35 , 12.02 - 16.98 (19) | 15.24 ± 1.91 , 12.24–20.59 (26) | 14.56 ± 0.98 , 10.07 - 17.53 (533) |
| postorbital constriction | 23.76 ± 1.24 , 20.77 - 26.53 (59) | 24.19 ± 1.38 , 22.12 - 26.42 (19) | 25.96 ± 1.39 , 22.76–28.71 (26) | 22.90 ± 1.19 , 17.27 - 34.91 (533) |
| braincase breadth | 31.05 ± 1.73 , 28.64 - 41.54 (59) | 31.33 ± 1.69 , 28.68 - 34.28 (19) | 33.57 ± 1.76 , 30.24–37.99 (26) | 31.04 ± 1.35 , 19.32 - 39.1 (533) |
| zygomatic breadth | 41.26 ± 2.21 , 35.58 - 45.69 (59) | 41.6 ± 2.83 , 37.15 - 45.41 (19) | 44.95 ± 2.84 , 39.86–54.04 (26) | 40.55 ± 2.47 , 26.6 - 46.89 (533) |
| mastoid breadth | 27.77 ± 1.11 , 25.05 - 30.86 (59) | 28.43 ± 1.7 , 24.57 - 31.62 (19) | 29.14 ± 2.14 , 25.86–35.72 (26) | 27.32 ± 1.38 , 20.56 - 36.74 (533) |
| height of jugal bone | 6.33 ± 0.91 , 4.97 - 9.36 (59) | 7.39 ± 0.98 , 5.61 - 8.78 (19) | 8.13 ± 0.9 , 6.21–10.16 (26) | 6.64 ± 0.97 , 4.11 - 10.45 (532) |
| mandible length | 77.62 ± 3.43 , 65.71 - 83.49 (59) | 77.99 ± 4.38 , 71.29 - 83.93 (18) | 84.68 ± 5.97 , 77.23–101.46 (26) | 76.43 ± 4.41 , 60.69 - 88.97 (532) |
| mandibular toothrow length | 25.13 ± 1.83 , 20.97 - 30.35 (59) | 24.16 ± 2.17 , 17.48 - 26.4 (19) | 26.61 ± 2.11 , 22.04–33.01 (26) | 24.82 ± 1.61 , 19.05 - 29.27 (533) |

**Table S6:** Adult external measurements (in millimeters) of the four Dasypus species recognized in this study. External and carapace measurements following Feijó et al. (2018). Measurements are given as Mean ± standard deviation, minimum and maximum values (number of specimens) are based on specimens examined by Feijó et al. (2018).

| *D. mexicanus* | *D. fenestratus* | *D. guianensis* | *D. novemcinctus* | |
|---|---|---|---|---|
| total length | 807.6 ± 57.1 , 680 - 883 (9) | 780.4 ± 64.02 , 642 - 850 (10) | 859 ± 56.39 , 790 – 945 (5) | 733.3 ± 90.1 , 402 - 950 (176) |
| tail length | 385.6 ± 37.3 , 305 - 430 (8) | 397.7 ± 70 , 262 - 550 (10) | 395.5 ± 50.74 , 339 – 460 (6) | 312.6 ± 66.3 , 115 - 450 (170) |
| hindfoot length | 96.7 ± 19.4 , 51 - 120 (9) | 81.4 ± 21.4 , 40 - 99 (7) | 95 ± 14.72 , 75 – 110 (4) | 90.1 ± 10.9 , 44 - 112 (171) |
| ear length | 39.5 ± 2.4 , 35 - 43 (8) | 41.8 ± 2.4 , 40 - 45 (6) | 49.25 ± 4.35 , 45 – 55 (4) | 41.8 ± 4.6 , 20 - 56 (157) |
| weight (in g): | - | 3285.8 ± 844.5 , 2173 - 4256 (5) | 6075 ± 0.67 , 5600 – 6500 (2) | 4271.7 ± 1000 , 1600 - 6550 (146) |
| dorsal length of the cephalic shield | 97.49 ± 5.66 , 82.75 - 109 (28) | 86.94 ± 10.60 , 73.24 - 103 (10) | 107.85 ± 7.58 , 95.11 - 116 (6) | 95.54 ± 11.27 , 68.92 - 137 (122) |
| dorsal length of the scapular shield | 96.65 ± 9.41 , 75 - 112 (29) | 86.6 ± 11.63 , 69 - 110 (15) | 88.59 ± 10.20 , 69 - 99 (9) | 92.25 ± 8.76 , 69 - 115 (169) |
| dorsal length of the pelvic shield | 131.46 ± 13.05 , 103 - 157 (28) | 123.06 ± 14.22 , 105 - 151 (15) | 149.57 ± 13.68 , 124 - 168 (7) | 133.15 ± 13.19 , 90 - 165 (171) |
| dorsal length of the caudal sheath with rings | 244.04 ± 22.71 , 204 - 284 (23) | 243.81 ± 32.74 , 200 - 324 (11) | 271 ± 21.55 , 243 - 294 (6) | 220.13 ± 25.82 , 164.13 - 287 (111) |
| dorsal length of the caudal sheath without rings | 142.17 ± 30.57 , 97 - 215 (17) | 135.28 ± 22.72 , 114 - 172 (7) | 150.5 ± 18.30 , 124 - 180 (6) | 124.17 ± 22.09 , 59 - 170 (90) |

# CHAPTER II - XENARTHRAN EVOLUTIONARY HISTORY UNRAVELED BY WHOLE-GENOME SEQUENCING AT THE SPECIES LEVEL


## Abstract

Xenarthrans (armadillos, anteaters, and sloths) are the least diverse of the four major clades of placental mammals, with only 42 currently valid species. However, some hidden diversity probably remains to be described as some genera have not been evaluated for decades. We sequenced 19 new mitogenomes and 72 whole genomes to constitute the most exhaustive phylogenomic datasets of Xenarthra assembled to date with 261 mitogenomes covering 37 extant and 7 extinct species and 94 whole-genome sequencing encompassing 36 distinct species. Discovery species delimitation methods based on mitochondrial phylogenetic relationships provided a species hypothesis that was further evaluated through comparisons of the genomic differentiation between lineages. A taxonomy consistent with equivalent genome wide divergece within xenarthrans would require revalidating *B. ephippiger* and *B. griseus*, lump *B. pygmaeus* with *B. griseus*, and supported a new species of pygmy anteater. Genome wide divergence also suggest that species of the Euphractinae subfamily should all belong to the *Euphractus* genus. The evaluation of the demographic history of xenarthran species revealed an overall decline of effective population size and the alarmingly low genomic diversity observed in four species make them a priority for assessing their conservation status. Based on this revised taxonomic framework, we reconstructed the most comprehensive time-calibrated phylogeny of Xenarthra based on whole genomes. Furthermore,

disentangling incomplete lineage sorting and gene flow revealed that both have contributed to discordant topological signals in parts of the xenarthran phylogeny. Overall, our results shed light on the evolutionary history of xenarthrans and contributed to identify directions for future taxonomic investigations and conservation status assessments.

**Keywords**: Species delimitation, taxonomic revision, whole genomes, conservation genomics, phylogenomics, genetree discordances, demography, mammals.

# Introduction

Third-generation sequencing technologies have significantly improved the production of mammalian genomes (Larsen and Matocq 2019). By sequencing longer DNA read fragments, these technologies have enabled to overcome the complexity of large and repeat-rich mammalian genomes, thus producing more contiguous genome reconstructions (Lee et al. 2016). In conjunction with cost reduction, these technical breakthroughs sparked large-scale genome sequencing initiatives such as the Zoonomia consortium (Zoonomia Consortium 2020), the DNA Zoo (Dudchenko et al. 2017), the Vertebrate Genome Project (Rhie et al. 2021), and the Darwin Tree of Life Project (Darwin Tree Of Life Project Consortium 2022), which already produced numerous high quality de novo genome assembly for mammals. Collectively, these sequencing efforts have resulted in 675 available genomes assemblies that have considerably improved our understanding of mammalian relationships and their evolutionary history (Zoonomia Consortium 2020; Foley et al. 2023; Upham and Landis 2023). Indeed, analyzing whole genomes offers a global and more accurate vision of the processes that have shaped mammalian genome evolution (Christmas et al. 2023; Kaplow et al. 2023; Kuderna et al. 2023; Osmanski et al. 2023) and genetic diversity (Wilder et al. 2023). This also notably allowed revealing the importance of genome-wide phylogenetic discordances generated by gene flow and/or Incomplete Lineage Sorting (ILS; Degnan and Rosenberg 2009; Vanderpool et al. 2020; Foley et al. 2023). Moreover, combining reference genomes with large-scale resequencing of additional species/individuals has allowed reconstructing the evolutionary history of entire mammalian orders such as pangolins (Heighton et al. 2023) or primates (Shao et al. 2023). Nevertheless, available genomes unevenly and non-exhaustively cover the ˜6,500 currently recognized mammalian species (Burgin et al. 2018; Upham and Landis 2023).

Xenarthrans (armadillos, anteaters, and sloths), despite constituting one of the four major clades of placental mammals, are particularly poorly represented with only 10 species with available genomes (Burgin et al. 2018; Feijó and Brandão 2022) including three high-quality chromosome-scale genomes. Diversifying for the last 65 Myr during South America splendid isolation (Delsuc et al. 2004; Gibb et al. 2016), they evolved striking morphological adaptations with arboreal sloths (Folivora), myrmecophagous anteaters (Vermilingua), and armored armadillos (Cingulata), which today constitute a restricted diversity of 42 extant recognized species endemic to the neotropics (Burgin et al. 2018; Feijó and Brandão 2022). Sloths count only seven species distributed in two distinct genera: *Bradypus* and *Choloepus* (Gardner 2005, 2007; Miranda et al. 2023). Anteaters are represented by 10 species within the *Myrmecophaga*, *Tamandua* and *Cyclopes* genera (Gardner 2005, 2007; Miranda et al. 2018). Finally, armadillos are the most speciose group with 25 species classified within the families Dasypodidae and Chlamyphoridae (Gardner 2005; Wetzel et al. 2008; Gibb et al. 2016; Feijó et al. 2018, 2019; Feijo and Anacleto 2021; Barthe et al. (Accepted)). However, 71.4% of xenarthran genera have not been the subject of a taxonomic study in 50 years (Feijó and Brandão 2022). Feijo and Brandão (2022) suspected that such taxonomic stability could be explained by a lack of taxonomic studies rather than a good understanding of their real species diversity. Indeed, when considering the recently studied genera, important taxonomic changes have been implemented at the family, genus, subgenus, species, and subspecies levels (Gibb et al. 2016; Feijó et al. 2018, 2019; Miranda et al. 2018, 2023; Feijo and Anacleto 2021). This recent taxonomic activity suggests that a lot of work remains to be done and suggests also an underestimation of the current species diversity with potential repercussions regarding their conservation status (Feijó and Brandão 2022).

Potential hidden species diversity is even more conceivable as some currently established xenarthran species have wide geographical distributions, encompassing major geological formations such as the Andes, the Guiana Shield, or restricted pathways such as the Isthmus of Panama. Those topological structures have been identified as reproductive biogeographic barriers for numerous species (Cortés-Ortiz et al. 2003; Redondo et al. 2008; Weir and Price 2011; Fouquet et al. 2012; Gutiérrez-García and Vázquez-Domínguez 2013; Esquerré et al. 2019), and might thus have been involved in the speciation process within xenarthrans. Some studies have notably identified diverging lineages separated by such formations within *Dasypus novemcinctus* (Gibb et al. 2016; Billet et al. 2017; Hautier et al. 2017; Feijó et al. 2018, 2019; Arteaga et al. 2020), *Bradypus var-*

*iegatus* (Ruiz-García et al. 2020), and *Cyclopes didactylus* (Coimbra et al. 2017; Miranda et al. 2018). However, neither the taxonomic status nor the implication of these biogeographical barriers have been explored at the genomic scale so far.

In addition to these taxonomic uncertainties, some phylogenetic relationships remain difficult to elucidate. Estimating the phylogenetic relationships of the xenarthrans have long been challenging because of their peculiar morphology (Delsuc and Douzery 2008). The emergence of molecular studies provided unprecedented advances in our understanding of xenarthran phylogeny (Delsuc et al. 2001, 2002, 2003). However, since Delsuc et al. (2003), the internal relationships within the two subfamilies Tolypeutinae and Euphractinae appeared unresolved (short internal nodes and low support values) even with different types of molecular data (e.g., with non-coding retroposon flanking sequences (Möller-Krull et al. 2007) and the concatenation of nuclear exons and two mitochondrial genes (Delsuc et al. 2012)). This result was somewhat expected in Euphractinae, because the three genera (*Euphractus*, *Zaedyus*, *Chaetophractus*) are anatomically very similar and previous morphological studies found support for conflicting groupings (Engelmann 1985; Patterson et al. 1989; Abrantes and Bergqvist 2006; Gaudin and Wible 2006). However, it was more surprising regarding Tolypeutinae in which morphological analyses have consistently supported the grouping of *Priodontes* and *Cabassous* into Priodontini, excluding *Tolypeutes* (Engelmann 1985; McKenna and Bell 1997; Cetica et al. 1998; Abrantes and Bergqvist 2006; Gaudin and Wible 2006). Gibb et al. (2016) provided the first phylogenetic study including all xenarthran species described at the time based on complete mitogenomes. This mitogenomic study strongly supported the paraphyly of Priodontini by grouping *Cabassous* and *Tolypeutes* to the exclusion of *Priodontes* but did not allow deciphering the relationships within Euphractinae. The incongruence with morphological studies and discordance among molecular markers have been suspected to be induced by ancient gene flow, ILS in relation with their fast diversification, or morphological convergence (Delsuc et al. 2002, 2003; Gibb et al. 2016). However, until now, phylogenetic studies (Möller-Krull et al. 2007; Delsuc and Douzery 2008; Delsuc et al. 2012; Abba et al. 2015; Gibb et al. 2016) have been widely restricted to mitogenomes and few nuclear molecular markers that incompletely reflect evolutionary history.

Here we provide whole genome sequencing for 72 xenarthrans, combined with 22 available genomes. Thus, we cover 37 of the 42 xenarthran species currently described, missing only the newly described *Cyclopes*

*xinguensis*, *C. rufus*, and the recently elevated *Dasypus beniensis*, *D. pastasae* and *Cabassous squamicaudis* (Miranda et al. 2018; Feijó et al. 2019; Feijo and Anacleto 2021). To provide a homogeneous taxonomic revision among xenarthran species, we estimated genomic differentiation indices between closely related species of xenarthrans. In addition, we evaluated their genetic diversity and inbreeding which contribute to determine which species should be prioritized for conservation assessment. We also explored the origin of phylogenetic discordances using coalescent-based methods. Finally, considering this comprehensive framework, we provided a time-scaled phylogeny of xenarthrans that allowed us to further explore the influence of biogeography on the xenarthran diversification.

# Results

## Mitochondrial phylogeny

The complete mitogenomes of 261 xenarthran individuals were gathered, combining 158 available on Genbank and 103 newly generated (Table S1). This data set includes 37 species out of the 42 extant recognised species but also seven extinct species (*Mylodon darwinii*, *Acratocnus ye*, *Parocnus serus*, *Megatherium americanum*, *Megalonyx jeffersonii*, *Nothrotheriops shastensis*, *Doedicurus* sp.) for which mitogenomes have recently been assessed (Delsuc et al. 2016, 2018, 2019; Mitchell et al. 2016). Maximum likelihood (ML) allowed reconstructing the most complete mitogenomic phylogeny of xenarthran species to date with globally robust bootstrap support (i.e. only six interspecific nodes had a bootstrap support of less than 100, Figure S1, Figure 1). This phylogenetic reconstruction was consistent with previous studies (Gibb et al. 2016; Delsuc et al. 2019; Tejada et al. 2023) except within *Bradypus variegatus* where two lineages, separating individuals from the southern Andes and those from Central America (hereafter designated as *Bradypus variegatus* S and *Bradypus variegatus* C respectively), were paraphyletic due to *Bradypus tridactylus*, which appeared closer to *Bradypus variegatus* S than *Bradypus variegatus* C. In addition, three individuals *Bradypus variegatus* NMB3, *Choloepus hoffmanni* NMB11 and *Euphractus sexcinctus* PAP76 clustered with a different species than expected, possibly due to potential cross-contamination or simply misidentification. For this reason these individuals have been excluded from further analyses (see Table S1). In addition, several individuals whose mitogenomes were labeled as "unverified" in GenBank grouped together and were highly divergent from other individuals of the same species suggesting sequencing or assembly errors, so we decided to exclude all "unverified" sequences (see Table S1). Finally, this mitogenomic framework at the population

level (in average six individuals per extant species) highlighted multiple genetically differentiated lineages notably within *Bradypus variegatus* (Figure 1). In addition, the insular *Bradypus pygmaeus* appeared to have a very low divergence from other continental individuals of *Bradypus variegatus* (Patristic distance = 0.003 substitution per nucleotide site, for comparison, the two subspecies of *Dasypus septemcinctus septemcinctus* and *D. sept. hybridus* have a Patristic distance = 0.005 substitution per nucleotide site).

To formally evaluate these mitochondrial lineages, we used two species discovery methods that propose a phylogenetic species delimitation based on substitution rate (bPTP) or branching rate of a time calibrated tree (GMYC). These two methods supported the delimitation of 52 and 48 putative species respectively (Figure 1, Figure S2, SGMYC). Within armadillos (Cingulata), these species delimitation methods generally agreed with the current taxonomy, and only two additional species were proposed (splitting *Cabassous unicinctus* and *Euphractus sexcinctus*; Figure 1). Within Pilosa, these methods supported much more partitions than is recognized by the current taxonomy. Notably, they supported splitting the five species of pygmy anteaters (*Cyclopes*) represented in our dataset into nine (bPTP) or seven (GMYC), and the brown-throated three-toed sloth (*Bradypus variegatus*) into 12 (bPTP) or 10 (GMYC) distinct species. These methods conversely supported *Bradypus variegatus* and *Bradypus pygmaeus* as the same species. However, the mitochondrial genome does not recombine and therefore represents a genealogical history that could be different from other loci. Considering this, additional supports were required to properly evaluate species boundaries in xenarthran using genome-wide nuclear markers.

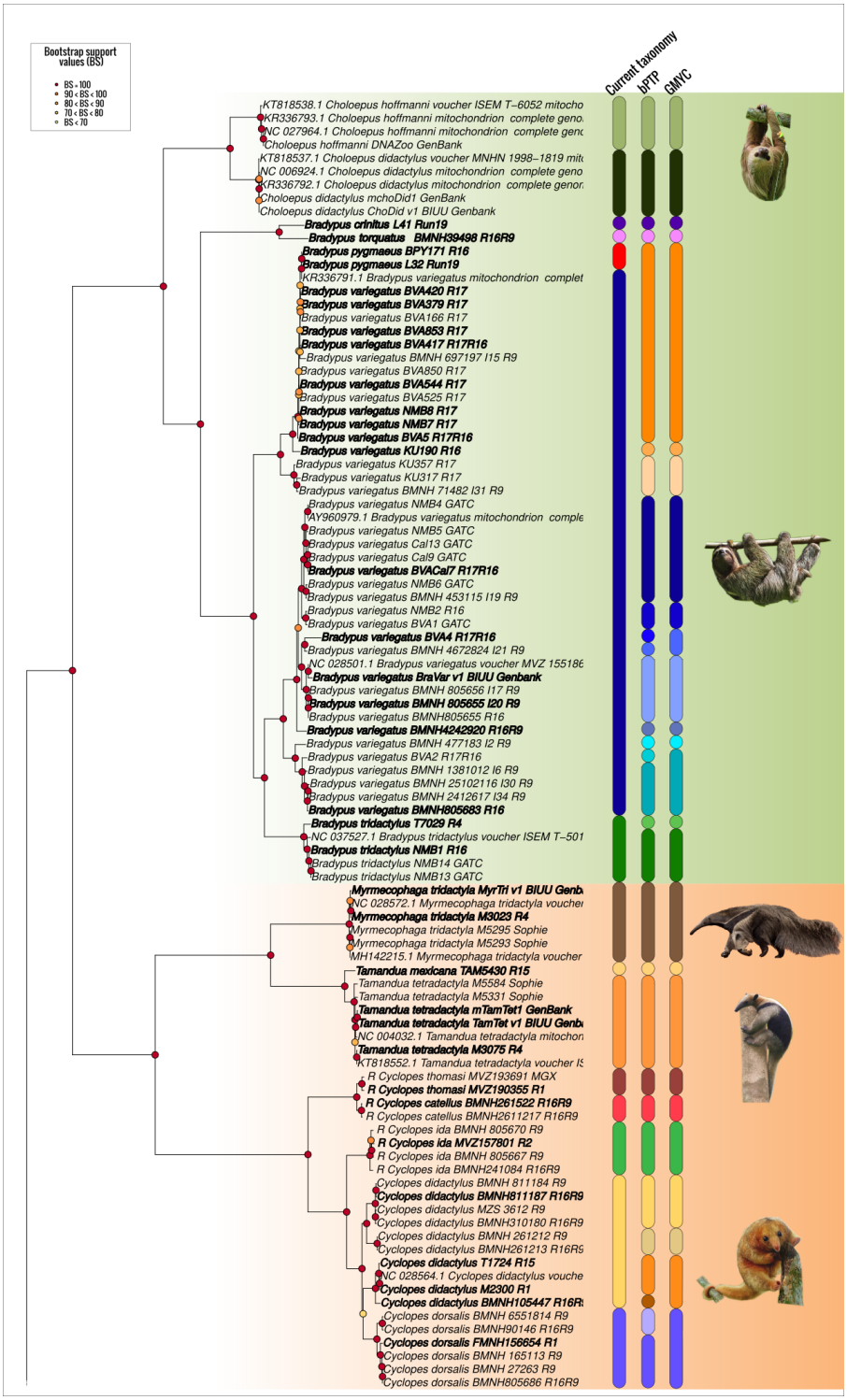


**Figure 1:** Phylogenetic relationships reconstructed by maximum likelihood from the partition of 15 mitochondrial genes for 222 individuals representing 37 species of the current xenarthran taxonomy. Node circles indicate bootstrap support (BS), the redder the color the higher the BS. Diagrams represent the assignment of individuals to distinct species according to the current taxonomy or to the partition supported by the phylogenetic delimitation methods bPTP and GMYC. Bold highlight the 73 individuals successfully sequenced in the whole genome. Photo credits: derekz65, benkelly, joao_andriola, lucaboscain, anthony2005, andresiade, silviolamothe, preli, Bradley Davis sclateria (iNaturalist.org), Mariella Superina, Quentin Martinez.

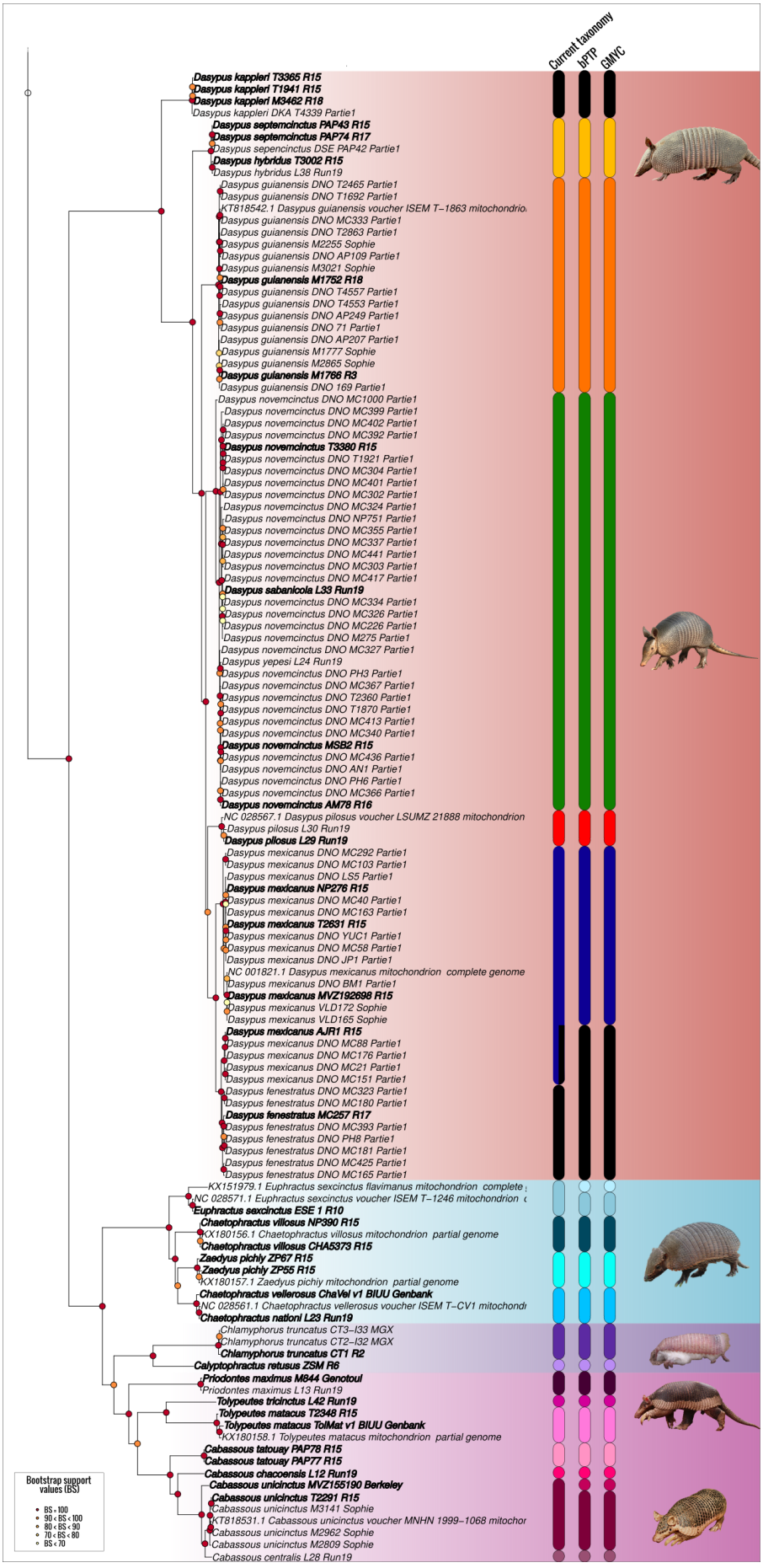


**Figure 1:** Continued.

## Whole genome sequencing and genotyping

The whole genome of 72 xenarthran individuals (including 24 museum samples; Table S2) were sequenced using short reads with the Illumina technology. Twenty-two already available genomes completed this dataset, allowing to consider a total of 94 individuals representing 37 xenarthran species out of the 42 currently described (Table S2). Genome assemblies of 11 species were used as reference genomes for mapping reads of closely-related resequenced individuals (Table S3). A proportion of 79% (74 individuals) obtained a mean depth of coverage higher than 5X with 43% (41 individuals) higher than 10X with less than 30% of missing data (Figure S4, Table S2). Thus, the 20 individuals (including the unique representative of *Cabassous centralis*) with less than 5X depth of coverage were excluded from downstream analyses, only the unique representative of *Calyptophractus retusus* was kept for phylogenetic reconstructions despite its low depth of coverage (4.40X). The individual *Euphractus sexcinctus* PAP76 was also excluded considering that its mitogenome suggested contamination/misidentification. For these 74 mapped individuals, we extracted orthologous genes of the Mammalia OrthoDB10 by localizing them in their respective reference genomes (see Figure S5 for their detailed completeness) to constitute a first phylogenomic dataset (hereafter *Busco_dataset*). A second dataset representing random genomic regions (hereafter *Random_dataset*) has been constituted by randomly sampling 1000 regions of 100kb from autosomal chromosomes.

## Genetic structure and genomic divergence

To further evaluate species partitions supported by species delimitation discovery methods based on mitochondrial sequences, we first performed a Principal Component Analysis (PCA) of genetic variance based on the multi-locus nuclear data using the *Random_dataset* (Figure PCA). Within three-toed sloths (*Bradypus* spp.), individuals from *Bradypus variegatus* C plus *B. pygmaeus* formed a cluster, which was distinct from individuals belonging to *Bradypus variegatus* S. These two genetic clusters were equidistant from *Bradypus tridactylus*. For pygmy anteaters (*Cyclopes* spp.), the genome-wide PCA confirmed the genetic differentiation of *C. thomasi*, *C. catellus*, *C. ida*, *C. dorsalis* and *C. didactylus*. When focussing on *Cyclopes didactylus* and *Cyclopes dorsalis* individuals (Figure S6), the individual BMNH 811187 is separated from other individuals of the same species (*Cyclopes didactylus*) by second axis (PCA2). Finally, within long-nosed armadillos (*Dasypus* spp.), the PCA also confirmed the genetic distinctiveness of *D. pilo-*

*sus, D. mexicanus, D. fenestratus, D. novemcinctus* and *D. guianensis*.

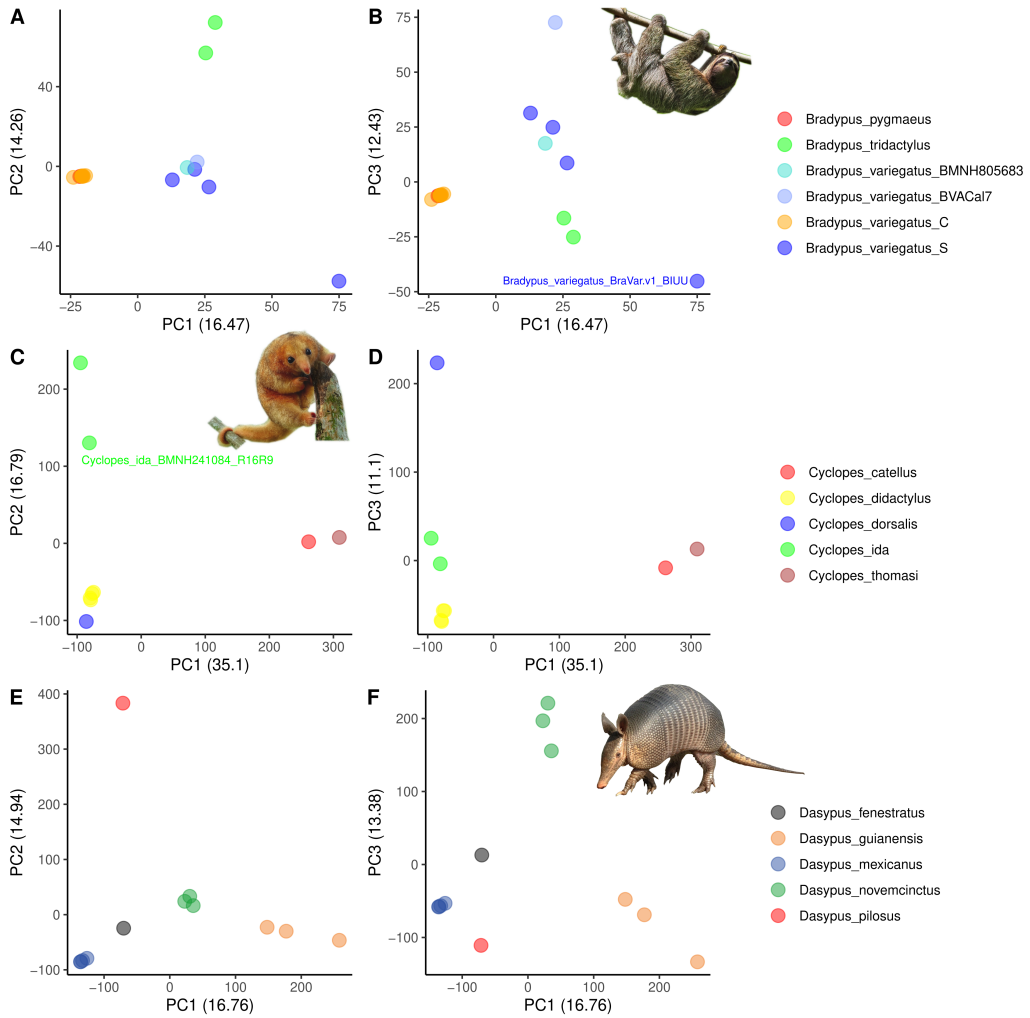


**Figure 2:** Principal Component Analysis of genetic variance (PCA) conducted on: A,B) the 4,502 SNPs shared by 19 *Bradypus* spp. individuals from 200 nuclear loci; C,D) the 66,687 SNPs shared by 9 *Cyclopes* spp. individuals from 400 nuclear loci; E,F) the 100,850 SNPs shared by 12 *Dasypus* spp. individuals from 200 nuclear loci. Loci were selected from the *Random_dataset* a) Projection on first (PC1) and second (PC2) axes. b) Projection on first (PC1) and third (PC3) axes.

The restricted number of individuals representing partitioned species limited the usage of population genomics methods to identify species boundaries. We thus decided to apply a genome-wide comparative approach using other xenarthran species with consensual taxonomic status as references to compare with. We quantified the pairwise genetic differentiation between closely related species within all xenarthran genera using the *Random_dataset*. We evaluated the genetic structure using the Genetic Differentiation Index (Allio et al. 2021), which corresponds to the FST (Nei 1983) applied to two diploid individuals. We also estimated divergence using Da and Dxy measures. The five pygmy anteater individuals representing recently recognized *Cyclopes* species ob-

tained a GDI $>$ 0.45 which is comparable or even greater than between *T. mexicana* and *T. tetradactyla* (GDI = 0.45; Figure 3). However, only *C. ida*, *C. dorsalis*, and *C. didactylus* had comparable pairwise divergence (Dxy $>$ 0.0064; Da $>$ 0.4) with *T. mexicana* and *T. tetradactyla* (Dxy = 0.0066; Da = 0.0036) as *C. catellus* and *C. thomasi* showed a Dxy = 0.0046 and Da = 0.0026 (Figure 3, Figure S7, Figure S8). In addition, further partition of *C. didactylus* were evaluated: individual BMNH 105447 showed genetic differentiation and divergence from other individuals of *C. didactylus* comparable to intraspecific values, contrary to *C. didactylus* BMNH 811187, which showed higher genetic differentiation (GDI = 0.466) than *Tamandua* species and greater divergence (Dxy = 0.0053; Da = 0.003) than *C. catellus* and *C. thomasi* (Figure 3, Figure S7, Figure S8). Within three-toed sloths of the *Bradypus* genus, the two lineages *Bradypus variegatus* C and *Bradypus variegatus* S showed comparable genetic differentiation and divergence (GDI = 0.64; Dxy = 0.0089; Da = 0.0065) than *Choloepus didactylus* and *C. hoffmanni* (GDI = 0.65; Dxy = 0.0082; Da = 0.0067; Figure 3, Figure S7, Figure S8). However, the pygmy sloth (*Bradypus pygmaeus*) appeared weakly differentiated from the lineage *Bradypus variegatus* C, with differentiation and divergence comparable to intraspecific values (GDI = 0.35; Dxy = 0.0018; Da = 0.0008; Figure 3, Figure S7, Figure S8). Within the *Bradypus variegatus* S lineage, additional partitioned species were evaluated but neither individuals BVA4 nor BMNH 4242920 showed genetic differentiation comparable to interspecific values (GDI $<$ 0.33; Dxy $<$ 0.0054; Da $<$ 0.0022), contrary to individuals BVACal7 and BMNH 805683, which were more differentiated (GDI = 0.548; Dxy $>$ 0.0062; Da $>$ 0.004) than *T. mexicana* and *T. tetradactyla* (Figure 3, Figure S7, Figure S8). Within armadillos, *Dasypus fenestratus* and *Dasypus mexicanus* were more differentiated (GDI = 0.63; Da = 0.0052) than *Chaetophractus velerosus* and *Zaedyus pichiy* (GDI = 0.52; Da = 0.0049), however, for *D. guianensis* and *D. novemcinctus*, despite a genomic differentiation (GDI = 0.35) similar to the subspecies *D. septemcinctus hybridus* and *D. septemcinctus septemcinctus* (GDI = 0.34), they showed a comparable divergence (Dxy = 0.0077; Da = 0.0048) than *C. vellerosus* and *Z. pichiy* (Dxy = 0.0078; Da = 0.0049). Concerning *D. sabanicola*, its genomic differentiation with *D. novemcinctus* (GDI = 0.24; Dxy = 0.0053; Da = 0.0026) was comparable to intraspecific differentiation of *D. novemcinctus* (GDI = 0.24; Dxy = 0.0057; Da = 0.0027). The Euphractinae subfamily is composed of the three genera *Euphractus*, *Zaedyus*, and *Chaetophractus*. However, their genetic structure and divergence (0.52 $<$ GDI $<$ 0.805; 0.008 $<$ Dxy $<$ 0.01; 0.005 $<$ Da $<$ 0.01) was comparable to intrageneric values of naked-tailed armadillos of the *Cabassous* genus (0.46 $<$ GDI $<$ 0.75; 0.009 $<$ Dxy $<$ 0.01; 0.005 $<$ Da $<$

0.01) and lower than between the two three-banded armadillos species *Tolypeutes matacus* and *T. tricinctus* (GDI = 0.8; Dxy = 0.02; Da = 0.02; Figure 3, Figure S7, Figure S8). Finally, the two individuals of southern naked-tailed armadillo (*Cabassous unicinctus*) showed genetic differentiation (GDI = 0.46) comparable to tamanduas species, despite lower divergence values (Dxy = 0.0032; Da = 0.0019; Figure 3, Figure S7, Figure S8).

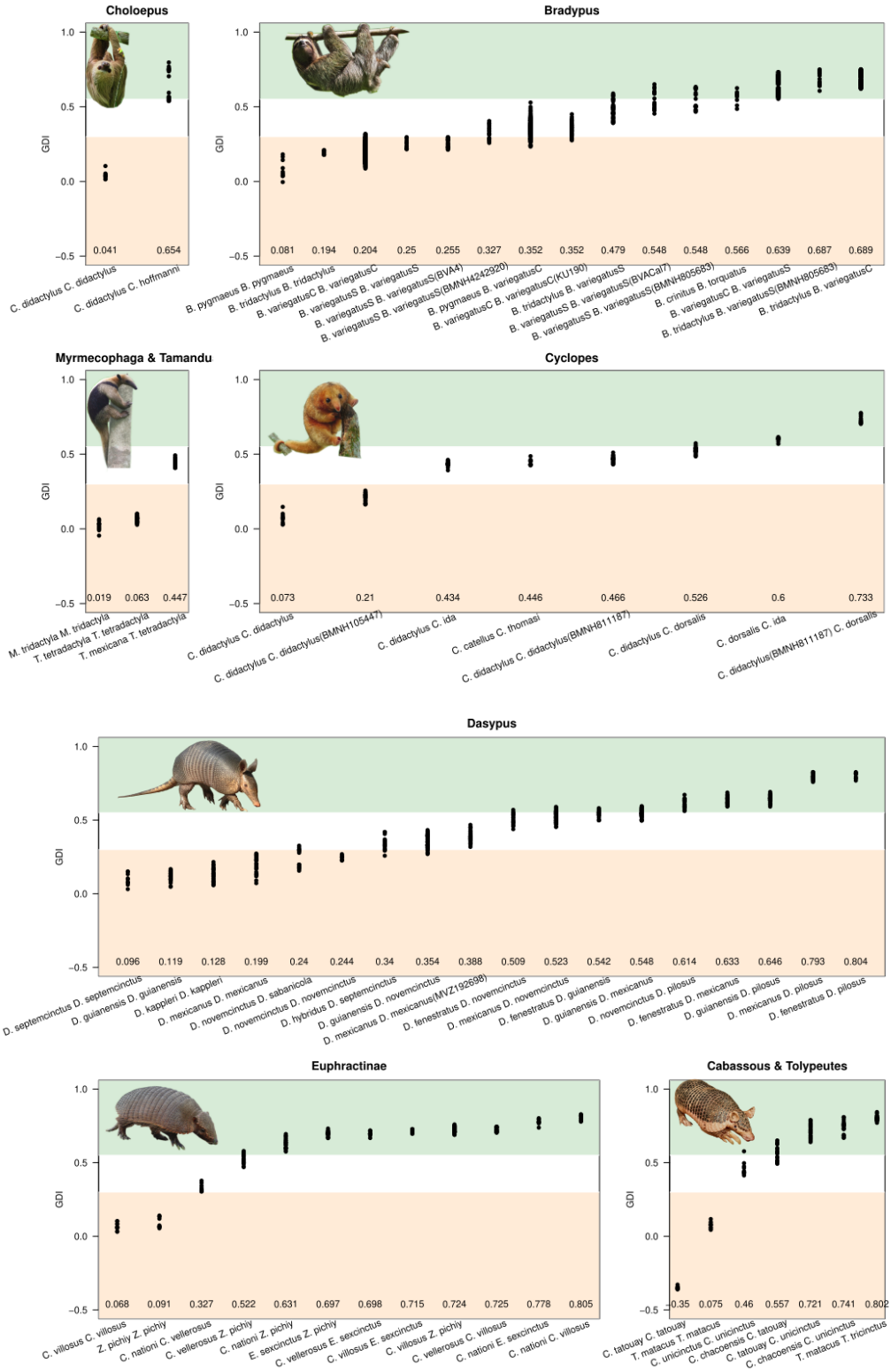


**Figure 3:** Pairwise genetic differentiation between xenarthran species estimated with the Genetic Differentiation Index (GDI; Allio et al. 2021). Each point represents the mean GDI estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. The mean GDI value is indicated below each graph. The higher the GDI, the greater the genetic differentiation. Color stripes are arbitrary and are intended to facilitate comparisons between the different genera.

## Genetic diversity, inbreeding, and demography

The genome-wide diversity has been evaluated using the expected proportion of heterozygous positions (i.e heterozygosity, He). Among xenarthran species, heterozygosity varied from 0.0008 heterozygous sites per nucleotide for *Chlamyphorus truncatus* to 0.0048 for one individual of *Bradypus variegatus* S, which is a comparable variation magnitude to other mammals (Figure 4, Figure S9). Three species (*Choloepus hoffmanni*, *Dasypus pilosus* and *Chlamyphorus truncatus*) and individual BMNH 805683 from the *Bradypus variegatus* S lineage, had a lower He value than the critically endangered *Bradypus pygmaeus* (He $<$ 0.0011). These values were comparable to *Nasalis larvatus* (He = 0.0013), *Rhinopithecus roxellana* (He = 0.0013), which are considered as endangered by the IUCN (Figure S9). In addition, three individuals (*Dasypus mexicanus* MVZ192698, *Cyclopes didactylus* BMNH105447, and *Choloepus didactylus* mchoDid1) showed lower He values than other individuals of their respective species, which could potentially reflect local or recent interbreeding (Figure 4). To further evaluate inbreeding levels, we screened genomes to detect Runs of Homozygosity (RoHs). For most xenarthran species, cumulative size of homozygous fragments was lower than 15 Mb. Only the three individuals from the two species of three-banded armadillos (genus *Tolypeutes*) presented a significant amount of RoHs with fragment size higher than 10Mb.

The variation in effective population size (Ne) over the last 10 million years (Myr) has been investigated using pairwise sequential Markovian coalescent (PSMC) based on the distribution of heterozygous positions along the genome. Remarkably, the trajectory of individuals of the same species seemed quite well synchronized (see the eight individuals of *B. variegatus* C, the three individuals of *T. tetradactyla*, *D. kappleri*, and *D. mexicanus*, and the two individuals of *M. tridactyla*, *D. septemcinctus*, *Cyclopes didactylus*, *Chaetophractus villosus*, *Zaedyus pichiy*, and *Tolypeutes matacus*). However, some individuals showed markedly distinct demographic histories from others of the same species or lineage (*Choloepus didactylus*, *Bradypus tridactylus*, lineages of *Bradypus variegatus* S, and *D. guianensis*).

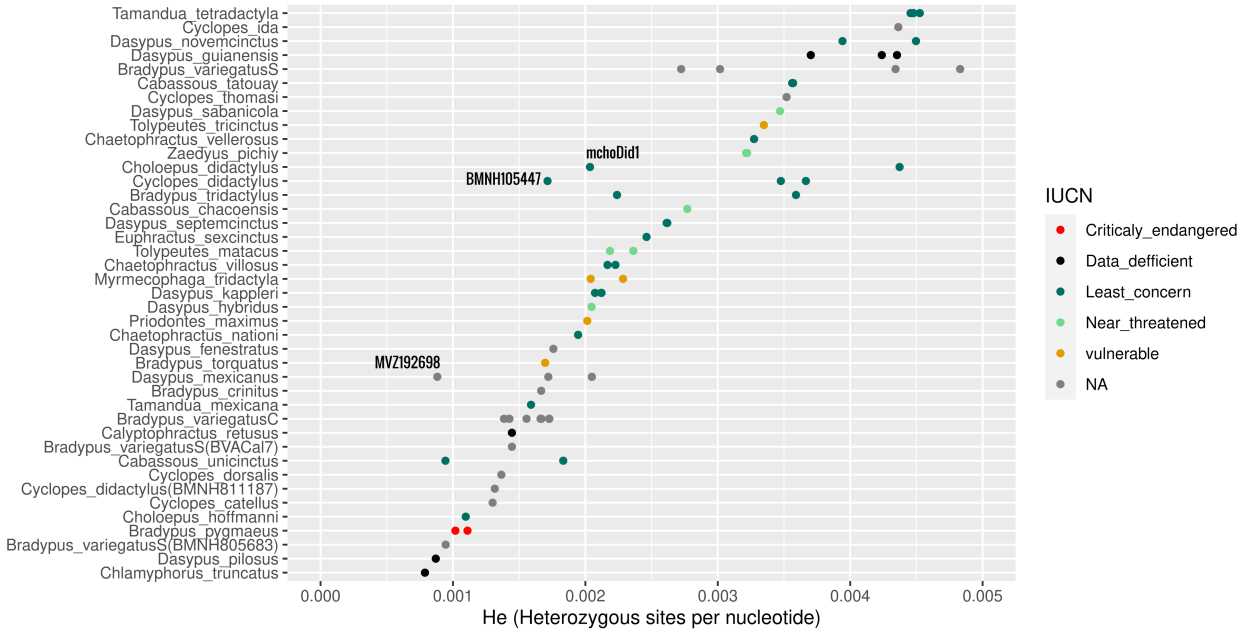


**Figure 4:** Genome-wide heterozygosity (He) estimated from 1000 random regions of 100kb from the *Random_dataset*. Each point represents an individual with circle colors corresponding to the IUCN status of the species to which it belongs. Outlier individuals are labeled.

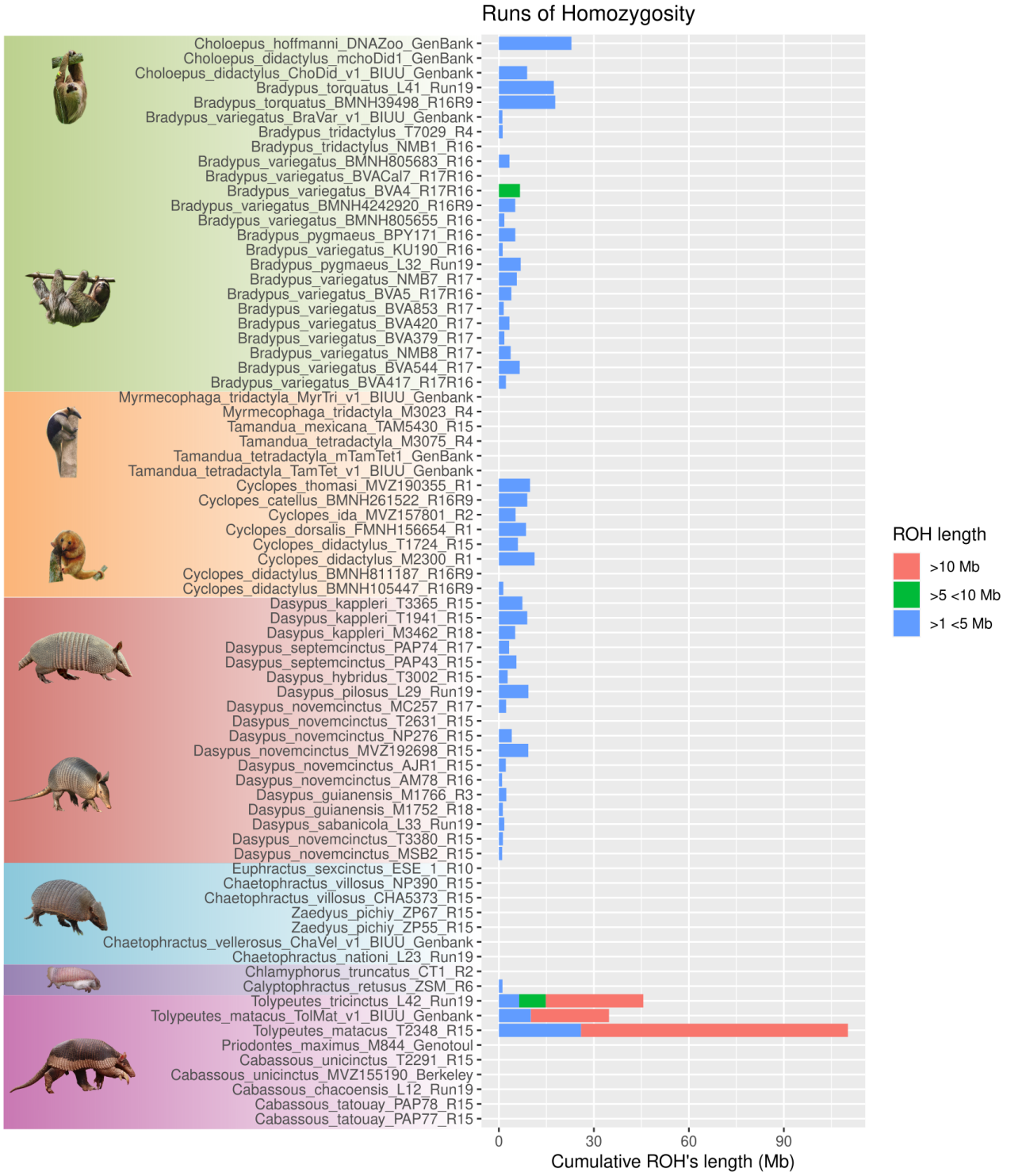


**Figure 5:** Runs of Homozygosity (ROH) in the genomes of xenarthran species. Bars represent cumulative RoH's length (Mb) for three sizes of homozygous fragments between 1 and 5 Mb in blue, between 5 and 10 Mb in green, and more than 10 Mb in red.

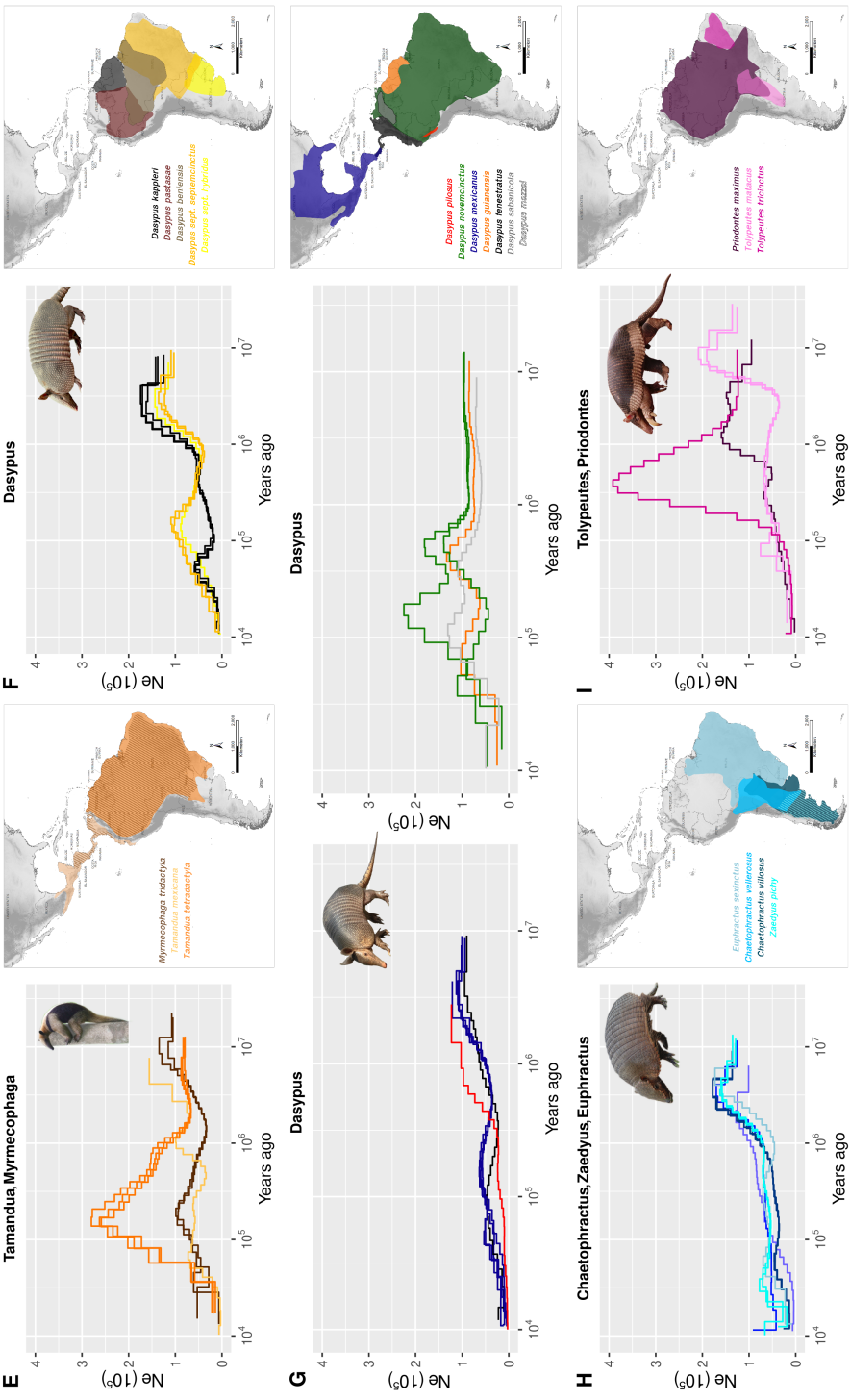


**Figure 6:** Demographic histories of xenarthran species or lineages estimated by pairwise sequential Markovian coalescent (PSMC) using parameters '4+30*2+4+6+10' (Nadachowska-Brzyska et al., 2013). Evolution of the effective population size ($Ne * 10^5$, Y-axis) was investigated over the last 10 million years (X-axis). Mutation rate ($\mu$) was set to 10-8 substitution per nucleotides per generation and generation time (g) of each species is detailed in Table S4). For each PSMC graph, the geographical distributions of the considered species are provided. A) *Choloepus* spp., B,C) *Bradypus* spp., D) *Cyclopes* spp., E) *Myrmecophaga tridactyla* and *Tamandua* spp., F, G) *Dasypus* spp. H) *Chaetophractus* spp., *Zaedyus pichiy*, *Euphractus sexcinctus*, I) *Tolypeutes* spp. and *Priodontes maximus*.

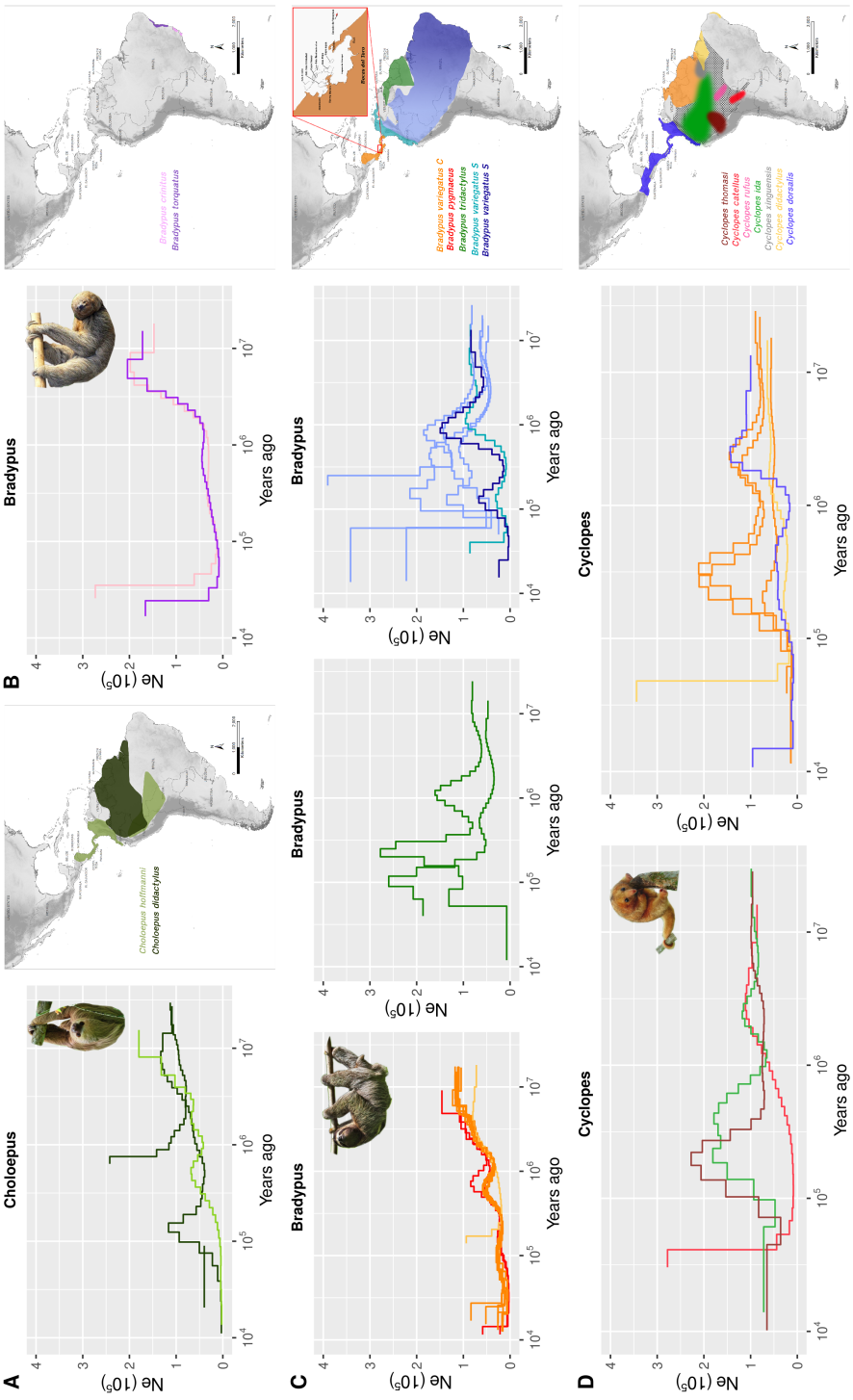


**Figure 6:** Continued.

## Phylogenomic analyses

The phylogenetic relationships of xenarthran species have been inferred from the 40 best quality individuals representing all but seven currently described xenarthran species (*Cyclopes rufus*, *Cyclopes xinguensis*, *Dasypus pastasae*, *Dasypus beniensis*, *Dasypus mazzai*, *Cabassous squamicaudis*, *Cabassous centralis*; Table S5, Table S6). Three afrotherian species have been included as outgroups (*Dugong dugon, Orycteropus afer, Loxodonta africana*) and 1,908 BUSCO genes (representing a total of 4,032,759 aligned sites whose 855,243 were parsimony informative) have been used to reconstruct the species tree using the ML approach (Table S7). The species tree inferred from the concatenation under a codon partition model (Figure 7) presented a congruent topology with the one inferred from the quartet consensus of gene trees (Figure S10). This nuclear phylogeny was consistent with the mitogenomic tree reconstructed by Gibb et al., (2016), while including seven recently recognized species (*Bradypus crinitus*, *Cyclopes catellus*, *Cyclopes thomasi*, *Cyclopes ida*, *Cyclopes dorsalis*, *Dasypus fenestratus*, and *Dasypus novemcinctus*) and four new lineages potentially corresponding to undescribed species (*Bradypus variegatus* C, *Bradypus variegatus* S, *Bradypus variegatus* S BMNH 805683, and *Cyclopes didactylus* BMNH 811187). This is the first genome-wide species level phylogenetic reconstruction for Xenarthra, with all nodes supported by bootstrap (BS = 100), except for one node within *Dasypus* with an extremely short branch (Figure 7).

Although family and subfamily relationships were consistent with mitochondrial findings, there were mito-nuclear discordances in Euphractinae inducing a different grouping of *Chaetophractus villosus*. In addition, some topological conflicts were detected using gene- and site- concordance factors (respectively gCF and sCF) by comparing the proportion of genes and sites supporting alternative topologies (Figure 7, Figure S11). Topological conflicts were observed within the species complexes of *Bradypus* spp., *Cyclopes* spp., and *Dasypus* spp.. As suggested by the mito-nuclear discordance involving *Chaetophractus villosus*, a discordant signal was also found within the nuclear genomes. In all these cases, the species tree was supported by at least 39% of informative sites even if a discordant signal was found between gene trees (Figure 7, Figure S11). Remarkably, this was not the case for two quartets that obtained an equivalent site support of the three possible topologies (Figure 7, Figure S11). As these discordants topologies can be induced by both incomplete lineage sorting (ILS) and/or gene flow (GF), we disentangled these factors using the Aphid method that compares coalescence times of genes trees as GF is expected to produce shorter genealogies than

ILS (Galtier 2023). Thus, for *Cyclopes* spp. (nodes 77 & 78 in Figure S11) about 20% of discordant topologies were associated with ILS. (Figure S12). For the triplet *C. dorsalis* and the two other lineages of *C. didactylus* (node 78 in Figure S11), 41.7% of these discordant topologies were estimated as due to GF, with an imbalance of 60% toward *C. dorsalis* and *C. didactylus* M2300. When considering *C. ida* (node 77 in Figure S11), less than 11% were detected as due to GF and mainly directed between *C. ida* and the two lineages of *C. didactylus* (for at least 80%, Figure S12). Within *Bradypus* spp., about 27% of discordant topologies of the node 83 were attributed to GF between the *B. tridactylus* and *B. variegatus* S lineages, preferentially with individual BVA4 (about 60%). When including *B. variegatus* C, 30% of GF was detected, well balanced between *B. tridactylus* and *Bradypus variegatus* S (BMNH 805683), while only 13.7% was detected when including the other lineage of *Bradypus variegatus* S (BVA4), which was almost exclusively between this individual and *B. variegatus* C. For the five species of long-nosed armadillos (i.e. *Dasypus pilosus*, *D. novemcinctus*, *D. guianensis*, *D. mexicanus*, and *D. fenestratus*), between 28% and 51% of discordant signal were induced by GF. The GF involving *D. pilosus* was well balanced between *D. novemcinctus* - *D. guianensis* and *D. mexicanus* - *D. fenestratus*, contrary to south American lineages (*D. novemcinctus* and *D. guianensis*), which have preferentially interbred with *D. fenestratus* rather than *D. mexicanus* (imbalance $>$ 63%). Within Euphractinae, about 30% of GF led to the discordant signal of node 62, and it was particularly balanced between species (Figure S12). About 50% of discordant topologies were explained by GF within Tolypeutinae, and affected preferentially (about 55%) *Cabassous* spp. and *Priodontes maximus* (Figure S12). Finally, we had also evaluated the relationships among the Euphractinae, Chlamyphorinae, and Tolypeutinae subfamilies (node 52 in Figure S11), which revealed 46% of GF imbalance between Chlamyphorinae and Euphractinae. Surprisingly, imbalanced ILS was detected with an excess of topologies supporting the grouping of Euphractinae and Chlamyphorinae to the exclusion of *Cabassous* spp..

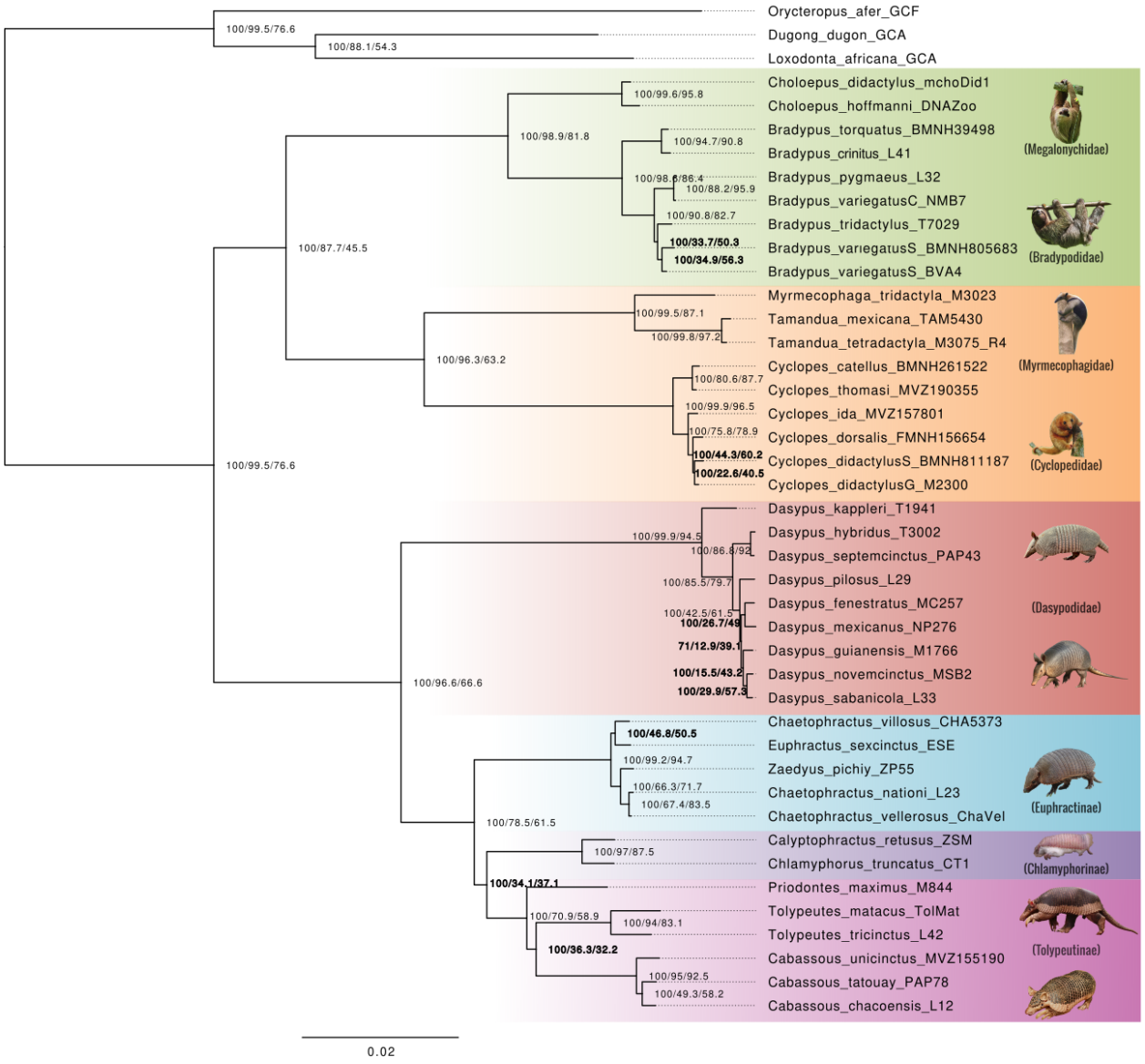


**Figure 7:** Phylogenetic relationships of 40 xenarthran species or subspecies reconstructed by maximum likelihood based on 1,908 BUSCO genes and rooted using three afrotherian outgroups. Node labels indicate bootstrap support (BS), and gene- and site-concordance factors (gCF/sCF). Node labels in bold highlight topological conflicts between genes and sites. Photo credits: derekz65, benkelly, lucaboscain, anthony2005, andresiade, silviolamothe, preli, Bradley Davis sclateria (iNaturalist.org), Mariella Superina, Quentin Martinez.

## Divergence time estimation

Divergence times were estimated based on the 100 more clock-like genes using a Bayesian approach based on a relaxed molecular clock with autocorrelated rates. Despite the fact that all model parameters and divergence time estimates have converged (ESS >> 200; Figure S13) , the posterior distributions of folivoran node ages showed bimodal distributions (Figure 8). The crown age of Folivora was estimated at 29.8 Mya (Table 1). Similarly, ancestral nodes of Xenarthra, Pilosa, Cingulata, and Vermilingua were estimated at 65.9 Mya,

59.2 Mya, 47.8 Mya, and 46.1 Mya, respectively.(Table 1). The three most diverse xenarthran genera (i.e. *Bradypus*, *Cyclopes*, and *Dasypus*) all originated around 11 Mya (Late Miocene) and diversified between 5 and 1.8 Mya (Pliocene and Pleistocene; Table 1). These estimations are anterior to the separation of the two *Tamandua* species (1.57 Mya; Table 1). Within Chlamyphoridae, the separation between Euphractinae and Chlamyphorinae/Tolypeutinae was estimated at 33.8 Mya, followed by the separation of the latter at 26.3Mya. The separation between *Tolypeutes* and *Cabassous* was estimated at 24.4 Mya based on these nuclear data (Table 1). Finally, within Chlamyphorinae, the divergence between *Chlamyphorus* and *Calyptophractus* around 9.69 Mya was similar to the one between the two species of *Tolypeutes* estimated at 9.26 Mya.

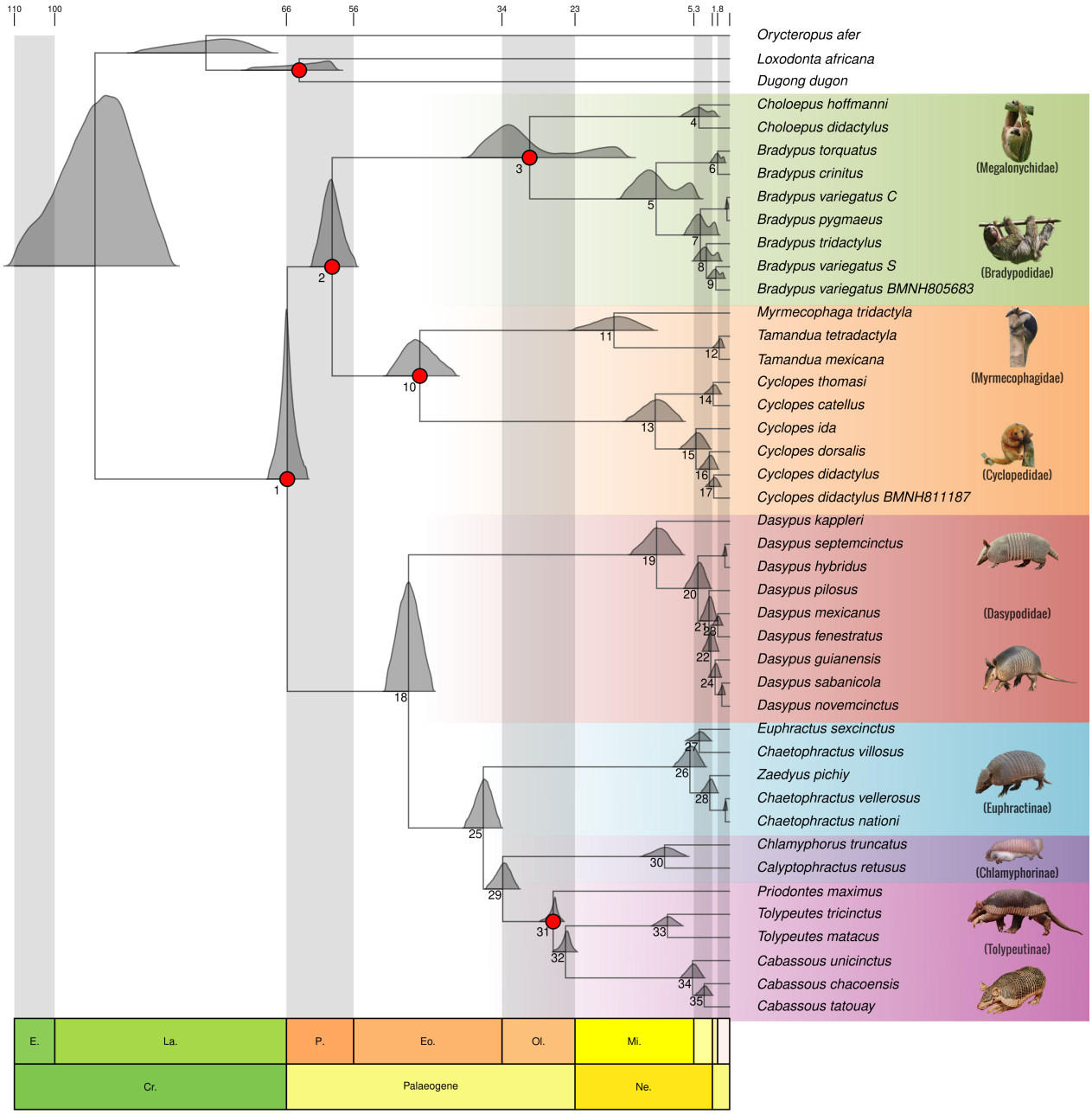


**Figure 8:** Divergence time tree of 40 xenarthran species or subspecies based on 100 most clock-like BUSCO genes and reconstruct using MCMCTree with relaxed clock with autocorrelated-rate and the the HKY + Γ substitution model of sequence evolution. Node distribution represents the 95% credibility intervals around mean age estimates. Red circles represent nodes used as calibration constraints. Node numbers refer to the Table 1 detailing divergence time estimates.

**Table 1:** Divergence time estimated with MCMCTree with relaxed clock with autocorrelated-rate and the HKY + Γ substitution model of sequence evolution based on 100 most clock-like BUSCO genes representing 40 xenarthran species or subspecies. For 36 clades, this table details their node number according to Figure 8 and compares mean posterior times estimated in Myr in this study, in Gibb et al. 2016, Delsuc et al. 2012 and 2004. The 95% credibility intervals are indicated in brackets. Clades in bold serve as calibration constraints. MND: Mito-Nuclear discordance.

| **Clades** | **Nodes** | **This study** | **Gibb et al. (2016)** | **Delsuc et al. (2012)** | **Delsuc et al. (2004)** |
|---|---|---|---|---|---|
| Xenarthra | 1 | 65.87 [62.9-68.7] | 67.7 ± 3.0 [60.4–71.6] | 67.8 ± 3.4 [61.3–74.7] | 64.7 ± 4.9 [55.3–74.6] |
| Pilosa (anteaters + sloths) | 2 | 59.19 [55.7-62.1] | 58.4 ± 4.1 [48.6–64.7] | 60.1 ± 3.6 [53.1–67.2] | 55.2 ± 4.9 [45.8–65.2] |
| Folivora (sloths) | 3 | 29.78 [15.8-38.1] | 29.9 ± 6.5 [16.5–39.6] | 28.3 ± 3.4 [22.0–35.2] | 20.8 ± 3.3 [15.0–27.8] |
| Megalonychidae (two-toed sloths) | 4 | 4.58 [1.87-7.12] | 9.2 ± 3.5 [3.5–16.7] | – | – |
| Bradypodidae (three-toed sloths) | 5 | 10.95 [4.86-15.94] | 19.0 ± 4.7 [9.6–27.0] | – | – |
| *B. torquatus* / *B. crinitus* | 6 | 1.79 [0.71-2.82] | – | – | – |
| *B. griseus* / others | 7 | 4.36 [1.81-6.63] | 7.7 ± 2.4 [3.6–12.6] | – | – |
| *B. tridactylus* / *B. variegatus* | 8 | 3.48 [1.38-5.3] | 5.7 ± 1.8 [2.6–9.5] | – | – |
| *B. variegatus* / *B. ephippiger* | 9 | 2.06 [0.81-3.21] | – | – | – |
| Vermilingua (anteaters) | 10 | 46.13 [40.95-50.85] | 37.8 ± 4.9 [26.9–46.2] | 45.5 ± 3.7 [38.4–52.8] | 40.0 ± 4.4 [31.8–49.0] |
| *Myrmecophaga* / *Tamandua* | 11 | 17.23 [11.57-23.17] | 12.7 ± 3.3 [7.0–19.8] | 13.6 ± 2.1 [9.9–18.2] | 10.1 ± 1.8 [6.9–14.1] |
| *T. mexicana* / *T. tetradactyla* | 12 | 1.57 [0.82-2.47] | 1.0 ± 0.4 [0.4–2.0] | – | – |
| *Cyclopes* | 13 | 11.09 [7.17-15.37] | – | – | – |
| *C. thomasi* / *C. catellus* | 14 | 2.46 [1.34-3.71] | – | – | – |
| *C. ida* / others | 15 | 5.02 [3-7.26] | – | – | – |
| *C. dorsalis* / others | 16 | 3.03 [1.78-4.41] | – | – | – |
| *C. didactylus* / *C.* sp. | 17 | 2.37 [1.37-3.49] | – | – | – |
| Cingulata (armadillos) | 18 | 47.82 [44.31-51.14] | 44.9 ± 3.5 [38.3–52.1] | 42.3 ± 3.8 [35.1–50.0] | 39.7 ± 4.5 [31.3–49.1] |
| Dasypodinae (long-nosed armadillos) | 19 | 10.88 [7.21-14.6] | 12.4 ± 3.4 [7.2–20.4] | 11.2 ± 2.0 [7.8–15.6] | 7.3 ± 1.6 [4.6–10.9] |
| *D. septemcinctus* / others | 20 | 4.74 [2.93-6.6] | 5.1 ± 1.7 [2.7–9.2] | – | – |
| *D. pilosus* / others | 21 | 3.07 [1.89-4.35] | MND | – | – |
| *D. mexicanus* / *D. fenestratus* / | 22 | 2.81 | MND | – | – |

| | | | | | |
|---|---|---|---|---|---|
| | | [1.7-3.95] | | | |
| *D. guianensis / D. novemcinctus* | | | | | |
| *D. mexicanus / D. fenestratus* | 23 | 1.79 [1.07-2.51] | – | – | – |
| *D. guianensis / D. novemcinctus* | 24 | 2.17 [1.3-3.08] | MND | – | – |
| Chlamyphoridae | 25 | 36.68 [34.15-39.38] | 37.2 ± 3.4 [31.5–44.7] | 34.5 ± 3.6 [27.8–41.9] | 32.9 ± 4.1 [25.2–41.5] |
| Euphractinae (hairy armadillos) | 26 | 5.92 [3.72-8.2] | 11.0 ± 2.8 [6.8–17.8] | 8.3 ± 1.6 [5.5–11.8] | 6.2 ± 1.4 [3.8–9.3] |
| *C. villosus / E. sexcinctus* | 27 | 4.53 [2.86-6.39] | MND | – | – |
| *Zaedyus pichiy* / others | 28 | 2.96 [1.82-4.2] | 8.2 ± 2.3 [4.9–13.7] | – | – |
| Chlamyphorinae / Tolypeutinae | 29 | 33.79 [31.48-36.21] | 32.6 ± 3.1 [27.9–40.0] | 32.9 ± 3.6 [26.3–40.2] | – |
| Chlamyphorinae (fairy armadillos) | 30 | 9.69 [6.65-13.04] | 19.4 ± 2.7 [15.2–25.9] | 17.3 ± 2.7 [12.4–23.0] | – |
| Tolypeutinae | 31 | 26.28 [24.76-28.18] | 25.7 ± 2.7 [22.4–32.7] | 26.1 ± 3.2 [20.2–32.9] | 21.8 ± 3.3 [15.8–28.9] |
| *Tolypeutes / Cabassous* | 32 | 24.44 [22.81-26.23] | 22.5 ± 2.6 [19.0–29.0] | 24.2 ± 3.1 [18.5–30.7] | 20.5 ± 3.2 [14.7–27.3] |
| *Tolypeutes* | 33 | 9.26 [7.17-11.42] | 14.1 ± 2.0 [11.0–19.1] | – | – |
| *Cabassous* | 34 | 5.52 [3.88-7.19] | 10.9 ± 1.9 [8.0–15.5] | – | – |
| *C. chacoensis / C. tatouay* | 35 | 3.8 [2.63-5.04] | MND | – | – |
| *C. centralis / C. unicinctus* | – | – | 1.3 ± 0.3 [0.8–2.1] | – | – |

# Discussion

## Genomic evidence for new xenarthran species

In this study, we reconstructed the phylogenetic relationships of 261 xenarthran mitogenomes covering 88% of the recognized species plus seven extinct species. This population scale phylogenetic reconstruction allowed us to identify multiple lineages suggested to represent distinct species by species discovery methods (Figure 1). We used this species delimitation partition as species hypotheses to evaluate further using nuclear genomes. Thus, we analyzed the whole genome of 73 individuals, representing 86% (36/42) of the recognized species and >85% (44/52 bPTP; 43/48 GMYC) of the partitioned species delimitation hypotheses.

We first evaluated the genome-wide divergence and genetic differentiation between 36 currently recognized xenarthran species. This provided a remarkable framework to

evaluate the taxonomic status of recently split species complexes or lineages with uncertain status through a comparative approach including currently recognized xenarthran species. Individuals of the five pygmy anteaters species (*Cyclopes* spp.) evaluated in this study, as well as individuals of northern (*Bradypus torquatus*) and southern (*Bradypus crinitus*) maned three-toed sloths, showed comparable genomic differentiations than the one between northern (*Tamandua mexicana*) and southern (*Tamandua tetradactyla*) tamanduas. This genome-wide evaluation therefore corroborates their species status, which was recently proposed based on morphology and only five molecular markers (Coimbra et al. 2017; Miranda et al. 2018, 2023). In addition, within pygmy anteaters, *Cyclopes dorsalis* was nested within *Cyclopes didactylus* rendering them paraphyletic (Figure 1). Analyzing their nuclear genomes recovered the monophyly of *Cyclopes didactylus* but revealed a very short internal node with high amount of discordant signal for an alternative grouping with *C. dorsalis* (Figure 7, Figure S11). These two *Cyclopes didactylus* lineages emerged 2.4 Myr ago (Figure 8; Table 1) and are distributed in the North East of South America, with a lineage being restricted to the Guiana Shield and the other with a disjunct distribution south of the Amazon river in the Brazilian Northeast (Figure S14). These two lineages were previously identified by Coimbra et al. (2017) using five molecular markers. However, considering their limited morphological differences, Miranda et al. (2018) remained conservative and did not formally describe them as separate species. Here, by evaluating their genome-wide differentiation, these two lineages of *Cyclopes didactylus* were demonstrated to show comparable genetic differentiation values to other *Cyclopes* species and the comparison with other xenarthran species supported their status as distinct species (Figure 3, Figure S7, Figure S8). In addition, evolution of their effective population size also supported their independent evolution these last million years (Figure 6). According to the type locality from Suriname, the epithet of *didactylus* belongs to the population from the Guiana Shield. Finally, a formal description is require for the Brazilian Northeast lineage, as there is no taxa currently in synonymy.

Within the brown-throated three-toed sloth (*Bradypus variegatus*), both mitochondrial and nuclear phylogenetic reconstructions revealed paraphyly :*Bradypus variegatus* C and *Bradypus variegatus* S are separated by the pale-throated three-toed sloth (*Bradypus tridactylus*) (Figure 1, Figure 7). Despite conflictual genetic signals at this topological node (Figure 7, Figure S11), the two lineages were more differentiated than many recognized xenarthran species and showed comparable values to Linnaeus's (*Choloepus*

*didactylus*) and Hoffmann's (*Choloepus hoffmanni*) two-toed sloths, which supports their elevation to the species level (Figure 3, Figure S7, Figure S8). These two lineages separated around 4.4 Myr ago (Figure 8; Table 1) and have distinct distributions, with *Bradypus variegatus* C located in Central America (from Panama to Nicaragua), while *Bradypus variegatus* S is found in South America on the western side of the Andes and throughout the Amazonian basin on the eastern side (Figure S15). The type specimen of *Bradypus variegatus* (Schinz, 1825) comes from Bahia (Brazil) thus the southern lineage must conserve its denomination. The Central American lineage should inherit the epithet *griseus* (Gray, 1871) considering that its type specimen from Costa Rica represents the older description of a Central American three-toed sloth taxon. Surprisingly, the insular pygmy sloth species (*Bradypus pygmaeus*) was found nested within *Bradypus griseus* in both mitochondrial and nuclear phylogenetic reconstructions (Figure 1, Figure 7). Despite its striking morphological differentiation from the mainland populations, genetic differentiation comparisons with other xenarthran species would support this insular population as a subspecies of *Bradypus griseus* rather than a distinct species (Figure 3, Figure S7, Figure S8). In addition, mitochondrial species discovery methods suggested an additional delimitation partition in this Central American three-toed sloth species by distinguishing individuals from Nicaragua from others of Central America (Figure S15). This distribution corresponds to the previously described taxa *B. castaneiceps* (Gray, 1871), however, we unfortunately did not succeed in sequencing whole genomes with sufficient quality from these individuals, so their taxonomic evaluation is still required. Within the South American lineage of *Bradypus variegatus*, mitochondrial species discovery methods suggested additional partitions that were further evaluated. Indeed, PSMC of South American individuals revealed numerous distinct demographic trajectories in the last million years (Figure 6). Genome-wide differentiation comparisons revealed that individuals BVACal7 (Para, Brazil) and BMNH 805683 (Venezuela) were as genetically differentiated from other individuals of *Bradypus variegatus* than were the two tamanduas species (Figure 3, Figure S7, Figure S8). Divergence time of individual BMNH 805683 with other *Bradypus variegatus* has been estimated at 2 Myr (Figure 8; Table 1). Individuals of the same mitochondrial haplogroup are distributed in the north-western side of the Andes (Figure S15), which corresponds to *Bradypus ephippiger* (Philippi,1870), first described in the northern Andes of Colombia by Thomas (1917) and then restricted to the Atrato River in Colombia by Cabrera (1958). Ruiz-García et al. (2020) evaluated the mitochondrial genomes of 77 individuals of *B. variegatus* and also recovered this

North-western Andean mitotype, however their restricted dataset did not allow them to evaluate its taxonomic status. The two other lineages found within *Bradypus variegatus* are located on the eastern side of the Andes (Figure S15). Individuals of the same mitochondrial haplogroup than BVACal7 (Para, Brazil) are located in eastern Brazil (in the Atlantic forest and Pará state; Figure S15), which corresponds to the type locality of *Bradypus variegatus* (Schinz, 1825) and should thus conserve this denomination. Finally, the distribution of the last lineage is more fuzzy as individuals are found in Ecuador, Bolivia and Brazil, so their exact connection with *Bradypus variegatus* still needs to be assessed to properly define its epithet (Figure S15).

Within Cingulata, the genome-wide evaluations of *D. novemcinctus*, *D. guianensis*, *D. fenestratus* and *D. mexicanus* confirmed their species status as recently assessed based on exon capture data (Barthe et al.in revision). Moreover, in agreement with Barthe et al. (Accepted), we did not find evidence for supporting *D. sabanicola* as a distinct species from *D. novemcinctus*. Divergence time for the crown age of *D. novemcinctus*, *D. guianensis*, *D. fenestratus*, and *D. mexicanus* was estimated at 2.8 Mya and successive rapid speciation events led to these four species with a mean divergence time of 1.8 Mya for *D. fenestratus* and *D. mexicanus*, and 2.2 Mya for *D. novemcinctus* and *D. guianensis*. Among the Euphractinae subfamily, three genera are currently recognized: *Euphractus*, *Chaetophractus*, and *Zaedyus*. However, considering the demonstrated paraphyly of the genus *Chaetophractus* and the restricted genomic divergence among members of this subfamily (lower than between *Tolypeutes* spp. or *Bradypus* spp.; Figure 3, Figure S7, Figure S8), the four species might better be placed in the *Euphractus* (Wagler, 1830) genus to be consistent with the rest of the xenarthran taxonomy.

Finally, mitochondrial phylogenetic reconstruction found the southern naked-tailed armadillo (*Cabassous unicinctus*) to be paraphyletic due to the exclusion of the MVZ155190 individual and the grouping of the northern naked-tailed armadillo (*Cabassous centralis*) with other *Cabassous unicinctus* individuals. By evaluating genome-wide divergence between specimen MVZ155190 and another *Cabassous unicinctus* individual from French Guiana, we obtained values comparable to other xenarthran species (e.g. *Tamandua* spp. and *Cyclopes* spp.). The MVZ155190 individual was sampled in Peru, while all others *Cabassous unicinctus* of our dataset come from French Guiana. Despite the high divergence between these individuals with distant geographical locations, we remained conservative about their potential status as distinct species considering the lack of information on the distribution of these two lineages.

## Conservation implications for xenarthran species

Conservation efforts are essential considering xenarthrans face numerous threats including hunting, habitat degradation, and illegal pet trade (Superina and Abba 2020). This group is the least diverse of the four major placental mammal clades with only 42 currently recognized species (Burgin et al. 2018; Feijó and Brandão 2022; Barthe et al, in revision). Nevertheless, only 11 of them are listed as Least Concern in the IUCN Red List of Threatened Species (Table 2). Recent taxonomic revisions within Xenarthra have been proposed and had significant implications for conservation efforts (Coimbra et al. 2017; Miranda et al. 2018, 2023; Feijo and Anacleto 2021). Notably the monospecific species of pygmy anteater (*Cyclopes didactylus*) that was considered as Least Concern has been split into seven distinct species, which considerably restricted their geographical distribution (Coimbra et al. 2017; Miranda et al. 2018). Given these recent taxonomic changes, their conservation status needs to be reassessed. Eighteen xenarthran species that have been revised since 2018 are in this situation and are still awaiting the evaluation of their conservation status by the IUCN Anteater, Sloth and Armadillo Specialist Group (ASASG). Here, by proposing a new species of *Cyclopes* and revalidating three new *Bradypus* species, we increased this number to 22. In addition, five xenarthran species are still considered as Data Deficient in the IUCN Red List (Table 2). Indeed, many xenarthran species are difficult to study in the wild notably because of their ecology (e.g. fossorial, arboreal) and the lack of researchers working on these species (Superina and Abba 2020). Thus, evaluating threats and their precise distribution can be challenging and require fieldwork, which can hinder conservation measures (Feijó et al. 2022). Although these fieldwork informations are essential, genetics estimation of their population size and inbreeding can constitute an initial approach to target priority species (Theissinger et al. 2023). Here, by reconstructing the demographic evolution over the last million years we highlight an overall reduction in population size of xenarthran species (Figure 6). However, by estimating heterozygosity and distribution of ROHs, we suggest that *Choloepus hoffmanni*, *Dasypus pilosus*, *Chlamyphorus truncatus* and *Bradypus ephippiger* are a priority concern for assessing their conservation status based on their reduce genetic diversity (Figure 4, Figure 5).

Finally, by recognizing *Bradypus pygmaeus* as a subspecies of *Bradypus griseus*, we hope not to hinder conservation of this unique island population currently considered as Crit-

ically Endangered in the IUCN Red List. Despite this taxonomical change, this population, distinguished by its remarkable size reduction due to insular dwarfism, remains under severe threat due to its limited range confined to the island Escudo de Veraguas and ongoing habitat degradation (Voirin 2015).

**Table 2:** IUCN Red List categories for xenarthran species.

| **Species** | **IUCN Red List category** |
|---|---|
| *Choloepus didactylus* | Least concern |
| *Choloepus hoffmanni* | Least concern |
| *Bradypus torquatus* | Re-evaluation needs |
| *Bradypus crinitus* | – |
| *Bradypus griseus* | – |
| *Bradypus tridactylus* | Least concern |
| *Bradypus variegatus* | Least concern $\geq$ re-evaluation needs |
| *Bradypus ephippiger* | – |
| *Bradypus* spp. | – |
| *Cyclopes catellus* | – |
| *Cyclopes thomasi* | – |
| *Cyclopes ida* | – |
| *Cyclopes rufus* | – |
| *Cyclopes xinguensis* | – |
| *Cyclopes dorsalis* | – |
| *Cyclopes didactylus* | Re-evaluation needs |
| *Cyclopes* sp. | – |
| *Myrmecophaga tridactyla* | Vulnerable |
| *Tamandua mexicana* | Least concern |
| *Tamandua tetradactyla* | Least concern |
| *Dasypus kappleri* | Re-evaluation needs |
| *Dasypus beniensis* | – |
| *Dasypus pastasae* | – |
| *Dasypus septemcinctus* | Least concern |
| *Dasypus pilosus* | Data deficient |
| *Dasypus sabanicola* | Near threatened |
| *Dasypus mazzai* | Data deficient |
| *Dasypus novemcinctus* | Re-evaluation needs |
| *Dasypus guianensis* | – |
| *Dasypus fenestratus* | – |
| *Dasypus mexicanus* | – |
| *Euphractus sexcinctus* | Least concern |
| *Chaetophractus villosus* | Least concern |
| *Zaedyus pichiy* | Near threatened |
| *Chaetophractus vellerosus* | Least concern |
| *Chlamyphorus truncatus* | Data deficient |
| *Calyptophractus retusus* | Data deficient |
| *Priodontes maximus* | Vulnerable |
| *Tolypeutes matacus* | Near threatened |
| *Tolypeutes tricinctus* | Vulnerable |
| *Cabassous tatouay* | Least concern |
| *Cabassous chacoensis* | Near threatened |
| *Cabassous centralis* | Data deficient |
| *Cabassous unicinctus* | Re-evaluation needs |
| *Cabassous squamicaudis* | – |

## Phylogenomic conflicts within xenarthrans

We reconstructed the more complete species level phylogeny of Xenarthra representing 36 of the 42 currently recognized species plus *Cyclopes* sp., *Bradypus ephippiger* and *Bradypus griseus* elevated as distinct species in this study (Figure 7). This genome-wide framework was concordant with previous mitochondrial reconstruction (Möller-Krull et al. 2007; Delsuc et al. 2012; Gibb et al. 2016) but revealed multiple nodes with conflictual signals between genes and sites (Figure 7, Figure S11). Notably, conflictual resolution of Tolypeutinae and Euphractinae highlighted by Delsuc et al. (2003) that remained controversial in later studies (Möller-Krull et al. 2007; Delsuc et al. 2012) appeared highly discordant with nearly equivalent support of the three topologies (Figure S11). Similarly, the crown of Chlamyphoridae also supported almost equivalently the three topologies (Figure S11). Additional nodes within *Cyclopes* spp., *Bradypus* spp. and *Dasypus* spp. supported discordant signals despite supporting preferentially a topology. Such phylogenetic discordances are expected when successive speciation events occur before complete lineage sorting, leading to shared ancestral polymorphism by multiple lineages (i.e incomplete lineage sorting; ILS) that will be sorted independently in each lineage, creating phylogenetic conflicts (Maddison 1997). Discordances can also result from gene flow (GF) when reproductive isolation is not complete. Disentangling these two processes in xenarthran nodes with highly discordant signals revealed that both GF and ILS had contributed to discordant genes trees. Notably, between 9.5% and 30% of discordant signals were attributed to ILS, with the highest values observed were within *Cyclopes* spp. and within *Bradypus griseus* and other South American species. The faster the successive speciation events and the larger the ancestral effective population size, the higher the values of ILS. Thus, it would be interesting to evaluate if this proportion of ILS corroborates differences between divergence time estimates (Table 1) and effective population size estimated by PSMC (Figure 6). Furthermore, nodes with highly discordant signals also supported GF, from 9.8% up to 50.9%. This corroborates evidence of GF previously found in the *Dasypus* genus (Barthe et al, in revision). In addition, the excess of topologies atributed to ILS supporting ((Chlamyphorinae, Euphractinae), *Cabassous*) suggest GF between *Cabassous* and a ghost lineage (Tricou et al. 2022a, 2022b). Together, this evidence for hybridization highlights the prevalence of post-speciation GF in xenarthrans and suggests speciation events that occurred with contact zones or secondary contacts. This is in line with other studies that also reported GF during speciation of mammalian species (Trigo et al. 2008,

2013; Kumar et al. 2017; Ge et al. 2023). Finally, considering gene flow can induce underestimation of speciation time we estimated divergence time based on genes minimizing topological conflict with the species tree (Smith et al. 2018).

## Biogeography

The Andean mountain range is the result of multiple pulses of tectonic uplift that occurred throughout the Cenozoic, which created a physical barrier between populations by reducing migration and also promoted local adaptation by inducing habitat differentiation notably along an altitudinal gradient (Hoorn et al. 2010). Its impact on speciation has been widely recognized in South American mammals (Cortés-Ortiz et al. 2003; Redondo et al. 2008), birds (Weir and Price 2011), and reptiles (Esquerré et al. 2019). The northern Andes of Venezuela, Colombia, and Ecuador delimit the distributions of several xenarthran species pairs such as *Tamandua tetradactyla / T. mexicana* (Moraes-Barros and Arteaga 2015; Ruiz-García et al. 2021), *Cabassous unicinctus / C. centralis* (Moraes-Barros and Arteaga 2015; Feijo and Anacleto 2021), *Cyclopes didactylus / C. dorsalis* (Coimbra et al. 2017; Miranda et al. 2018), and *Dasypus novemcinctus / D. fenestratus* (Barthe et al. accepted) or the two divergent mitochondrial lineages of Hoffmann's two-toed sloth (*Choloepus hoffmanni*; Moraes-Barros and Arteaga 2015). Here we have also recognized a distinct three-toed sloth species (*Bradypus ephippiger*) distributed north of the Andes and separated from southern individuals by this mountain range. These numerous cases of xenarthran species separated by the Andes suggest that this biogeographic barrier has played an important role in their speciation and in the maintenance of their genetic differentiation.

The Guiana Shield is an ancient craton located in northern South America that includes parts of eastern Venezuela, Guyana, Suriname, French Guiana, and northeastern Brazil, bounded by the Amazon to the south and the Orinoco to the west (Hammond 2005; Lujan et al. 2011). Its ancient rock formations have been eroded over millions of years, creating a landscape of mountains and plateaus that supports among the most extensive tracts of relatively pristine lowland tropical rainforests. Its ancient history, relative geographic isolation, and ecological singularity make it an area of high biodiversity in South America with numerous endemic species of plants, birds, amphibians and mammals (Kelloff and Funk 2004; Hollowell and Reynolds 2005; Fouquet et al. 2012). The Guiana Shield seems to have influenced the distribution of several xenarthran species, including the pale-throated three-toed sloth (*Bradypus tridactylus*), the greater long-nosed armadillo (*Dasy-*

*pus kappleri*; Feijo and Cordeiro-Estrela 2016; Feijó et al. 2019) and the guianan long-nosed armadillo (*Dasypus guianensis*; Barthe et al. accepted). In this study, we have also recognized the common silky anteater (*Cyclopes didactylus*) as being restricted to the Guiana Shield. This growing body of evidence of xenarthran species with a restricted distribution in the Guiana Shield suggests that this biogeographic region could have significantly influenced xenarthran diversification.

The presence of similar distributions among the three major xenarthran orders suggests the existence of common biogeographic barriers that have influenced their speciation. Thus, it provides important insights to better understand the factors that promote and maintain the biodiversity of this clade endemic to the Neotropics. Our revision of their taxonomy including genome-wide markers will allow performing more accurate diversification analyses to characterize the potential drivers of speciation within xenarthran species complexes. Previous studies have successfully addressed these aspects in other neotropical mammals, such as South American felids (Trigo et al. 2013; Figueiró et al. 2017; Trindade et al. 2021; Ramirez et al. 2022), canids (Chavez et al. 2022), marsupials (Giarla and Jansa 2014), and rodents (Vallejos-Garrido et al. 2023). Future genomic studies of this kind are required to fully understand the dynamics of the speciation process in xenarthrans.

# Conclusion

This study encompasses a nearly exhaustive whole-genome sequencing dataset of Xenarthra which has allowed to conduct the most comprehensive molecular revision of their species diversity. To be consistent with the current taxonomy, we performed a comparison of genomic differentiation between all xenarthran species that revalidate *B. ephippiger* and *B. griseus*, lump the Critically Endangered *B. pygmaeus* with *B. griseus* and suggests the description of a new species of anteater, *Cyclopes* sp., that requires a formal morphological description to assert its validity. This raises the number of xenarthran species to 44, but four species (*D. mazzai*, *D. sabanicola*s, *B. variegatus* and *C. unicinctus*) still require further investigation. In addition, species of Ephractinae should all belong to the *Euphractus* genus. Based on this taxonomic revision, this study provided a comprehensive assessment of the evolutionary history of Xenarthra notably by evaluating demographic history and genomic diversity, disentangling factors that have induced phylogenetic discordances and providing a time scale reference phylogenetic framework.

# Materials and Methods

## Biological sampling

Tissues samples of 103 xenarthran individuals were collected for this study. Thirty-one of them were sampled as dried skin pieces from museum specimens in the British Museum of Natural History (BMNH). Tissue samples of the 72 other individuals were obtained over the years through multiple sources (F. Delsuc, pers. comm.)

## DNA extractions and sequencing

Genomic DNA extractions were carried out using fresh tissue biopsies preserved in 95% ethanol using the DNeasy Blood & Tissue Kit from QIAGEN, following the manufacturer's instructions. For museum dried skin samples, a meticulous process was adopted: each sample was handled sequentially under a UV hood, with thorough cleaning in between to prevent any potential cross-contamination. DNA extractions were then conducted in batches of 12 samples, including a blank extraction control, utilizing the same DNeasy Blood & Tissue Kit in a clean room environment under UV. A minor adjustment was made to the elution volume, reducing it to 70 $\mu$l from the standard 100 $\mu$l. Subsequently, Illumina libraries were prepared from the DNA extracts, employing a cost-effective adaptation of the Meyer and Kircher (2010) protocol as suggested by Tilak et al. (2015). Shotgun Illumina sequencing was employed to obtain the mitogenomes by sequencing 10 million paired-end reads of 150 bp by Novogene for each sample. Among these individuals, additional sequencing was performed for 72 of them to obtain the whole genome at targeted depth of coverage of 15x.

## Mitochondrial dataset

We gathered the raw reads of the 103 individuals obtained by shotgun sequencing with raw reads of 98 publicly available genomes (including 22 sequences as whole genomes). For the 94 individuals sequenced as whole genomes, we used seqtk sample v1.3 to obtain a sub-sample of reads representing at least 15X of mitochondrial depth coverage. FastP v0.21.0 (Chen et al. 2018) was used to clean these 201 raw data. The reference mitogenome of *Cyclopes dorsalis*, *C. ida* and *C. thomasi* (MVZ190355) were generated using Mitofinder v1.4 (Allio et al. 2020). These reference mitogenomes and those already available (detailed in Table S3), were used to map reads of each individual with bwa mem v.0.7.17 (Li 2013) with default parameters according to the type of data (single-end or paired-end). Mapping files were converted, ordered and indexed according to their position on the reference mitogenome using Samtools v1.9 (Li et al. 2009) and Picard v2.25.5 (Picard Toolkit 2019). Duplicate reads were

marked using MarkDuplicates v2.25.5 (Picard Toolkit 2019). Freebayes v1.3.1 (Garrison and Marth 2012) had performed variant calling with adequate options for haploid data. The depth coverage was estimated by Samtools depth v1.9 (Li et al. 2009) with a mapping quality threshold of 30, a base quality threshold of 30 and the option « -a ». Thus, positions with less than 5x depth coverage were masked. In addition, we used bcftools filter to mask positions that were not of type "snp" or "ref", and also "complex" positions that induce a frame shift. Finally, bcftools consensus v1.14 (Danecek et al. 2021) converted the vcf file to fasta format using haploid parameters and masking positions that did not pass our thresholds. We complete this dataset with 60 publicly available mitochondrial sequences (see Table S1 for details) downloaded using eDirect tools.

## Mitochondrial phylogenetic reconstruction

For each reference mitogenomes, we defined a partition file corresponding to the 15 mitochondrial CDS and rRNA. Using AMAS split (Borowiec 2016) we extracted these regions from the mitochondrial sequences. These genes were aligned separately with mafft (Katoh and Standley 2013) with default parameters. Finally, for the 261 individuals, these 15 aligned genes were concatenated with AMAS concat (Borowiec 2016) and the mitogenomic tree was reconstructed by maximum likelihood (ML) with IQ-TREE v2.1.4 (Minh et al. 2020) with a partitioned model ("-spp" option) and using ModelFinder ("-m TESTNEW" option; Kalyaanamoorthy et al. 2017) to identified the best-fitting model.

## Mitochondrial phylogenetic delimitation

We reconstructed the ML tree for 222 individuals (excluding fossiles, unverified sequences from GenBank and 3 individuals (NMB3, PAP76 and NMB11) that are suspected to have been contaminated) to conduct species delimitation based on substitution rates using the Bayesian version of PTP (bPTP-ML) v0.51 (Zhang et al. 2013). Then we reconstruct an ultrametric tree by calibrating the tree using IQ-TREE with the "–date-tip 0" option and the same other options previously used, and setting the calibrated nodes ("–date" option) to 67.7 Myr for the crown of Xenarthra, 58.4 Myr for the crown of Pilosa, 29.9Myr for the crown of Folivora, and 37.8 Myr for Vermilingua following Gibb et al.(2016). The Generalized Mixed Yule Coalescent method (GMYC; Pons et al. 2006) from the R package splits v1 (Ezard et al. 2009) was used with default parameters.

## Whole genomes sequencing dataset

For the 72 individuals sequenced as whole genome plus the 22 publicly available whole genome, we used FastP v0.21.0 (Chen et al. 2018) to clean the raw data. We mapped reads with bwa mem v0.7.17 (Li 2013) on reference genomes (detailed in Table S3) using adequate parameters for paired-end data, and all the others parameters as default. We converted, ordered and indexed mapping files according to their position on the reference genome using Samtools v1.9 (Li et al. 2009) and Picard v2.25.5 (Picard Toolkit 2019). Duplicate reads were marked using MarkDuplicates v2.25.5 (Picard Toolkit 2019). We performed variant calling for diploid data using Freebayes v1.3.1(Garrison and Marth 2012). Samtools depth v1.9 (Li et al. 2009) with a mapping quality threshold of 30, a base quality threshold of 30 and the option « -a » estimated depth coverage in order to mask position with less than 6x depth coverage. VCF2FastaFreebayes was used to generate diploid fasta sequences and masked additional positions with quality inferior as 300 and heterozygous positions with a read frequency of the minor allele inferior to 0.3 (deviating from the 0.5 expectation). Nineteen individuals with less than 5X depth coverage plus individual PAP76, suspected to have been contaminated, have been discarded from downstream analyses.

Extraction of BUSCO regions - We identified the orthologous regions of the Mammalia OrthoDB10 BUSCO gene set of our 14 reference genomes using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.7 (Waterhouse et al., 2018). For each reference genome, numbers of genes found are summarized in the Figure S5. We extracted those genes in the diploid fasta sequences for the 74 individuals using exon coordinates provided by BUSCO analysis. Diploid sequences were converted in IUPAC format. Sequences with more than 50% of missing data were excluded. This dataset is referred as *Busco_dataset* Forty individuals with the lower rate of missing data were selected to represent all Xenarthrans species or lineages (Table S5). Then, 2006 BUSCO genes shared by at least 37 individuals (out of the 40 best representatives) were selected (see Tables S6, S7). These sequences were aligned using OMM_Macse V12.01 with the option "no_prefiltering".

Extract Random regions - To exclude potential contaminant scaffold of reference genome we focussed on those assigned as Chordata by the blastn v2.12.0+ and BlobTools v1.1.1 analyses (Supplementary files). We also excluded sexual chromosomes identified as scaffold with at least 50% aligning with the sexual chromosomes of *Choloepus didactylus* (NC_051334.1, NC_051335.1) using Lastz v1.04.22 and the script sum_seq_lastz_chrX.py. We sampled, with shuf, 100 regions of 100kb generated

by fasta_generate_regions.py and replicate this operation 10 times for each reference genome. These regions were extracted for all individuals using a home made script to constitute the *Random_dataset*.

## Phylogenetic reconstruction

The 2006 Busco genes from the *Busco_dataset* were aligned separately with OMM_Macse V12.01 and the option "no_prefiltering". Gene trees were reconstructed by ML with IQ-TREE v2.2.0. Phylter v0.9.11 identified 98 gene trees containing anormal branch length that we filtered out from the dataset. Then aligned sequences of the 1908 Busco genes were concatenated using Amas and the species tree was reconstructed by ML-with IQ-TREE v2.2.0 (Minh et al. 2020) with a partitioned model ("-spp" option) using ModelFinder ("-m TESTNEW" option; Kalyaanamoorthy et al. 2017) to identify the best-fitting based on the Bayesian Information Criteria (BIC). Node support was estimated using 1000 ultrafast bootstrap replicates and gene- and site-concordance factors (Minh et al. 2020).

## Divergence time estimation

We performed a Bayesian inference of divergence time with the MCMCTree program implemented in PAML v.4.10.7 (Yang 2007), using a Markov Chain Monte Carlo (MCMC) to generate the posterior distribution of divergence times. This method is time and computationally expensive when applied on large datasets (Smith et al. 2018). For this reason, we used the SortaDate approach to select 100 genes based on three criteria : i) clock-likeness estimated from the root-to-tip variance, ii) reasonable tree length, and iii) minimizing topological conflict with the species tree. A partition by codon positions was used given the higher likelihood value estimated by IQ-Tree for this tree rather than for the tree with no partition. The topology for MCMCTrees inferences was fixed according to the species tree obtained with the dataset composed by 1908 BUSCO genes. The phylogeny was constrained by six calibration points : i) for the node between Loxodonta and Dugong the minimum age was set to 72.3 Myr and maximum to 59.2 Myr, ii) between *Dasypus* and *Bradypus* to $<$ 66 Myr and $>$47.8 Myr iii) between *Cyclopes* and *Tamandua* to $<$ 56 Myr and $>$ 15.97 Myr, iv) between *Bradypus* and *Choloepus* to $<$ 40.6 Myr and $>$ 15.97 Myr v) between *Cyclopes* and *Bradypus* to $>$ 66 Myr and $>$ 31.5 Myr v) between *Cabassous* and *Priodontes* to $<$ 36 Myr and $<$ 26 Myr (Gibb et al., 2016; Foley et al. 2023). Finally, the root was constrained by an upper bound set at 131.5 Myr (Foley et al. 2023).

Approximate likelihood estimation - First, the maximum likelihood estimate of branch lengths were approximated by the Taylor expansion proposed by Thorne et al

and implemented in MCMCTree to improve the speed of the MCMC. Thus, the gradients (g) and the Hessian matrices (H) for the branch length were estimated (usedata = 2) using the HKY + Γ substitution model of sequence evolution (HKY85, model =4). Prior distribution for the node ages without calibration constraints were uniformly distributed (BDparas = 1 1 0.1). The prior of the substitution rates followed a discrete gamma distribution with alpha set to 0.5 and the number of categories in discrete gamma to 4. Following Gibb et al. (2016) we used a relaxed clock with autocorrelated-rate (the geometric Brownian motion model; GBM)(Rannala and Yang, 2007). In the second place, the branch lengths estimation were used to start the MCMC analysis to estimate divergence times (usedata=2). The MCMC chain ran for 5,000,000 generations and sampled every 100 generations. The first 500,000 generations were discarded as burn-in. This analysis was run twice to verify the MCMC had successfully converged.

## Species delimitations

Species discovery methods based on genetic variation of species complexes (i.e *Bradypus*, *Cyclopes*, *Dasypus*) was performed through a principal component analysis (PCA) on the randomly selected diploid sequences from the *Random_dataset* using the program popPhyl_PCA (`https://github.com/popgenomics/popPhyl_PCA`). Individuals with a high rate of missing data were excluded from this analysis to increase the number of SNPs represented by all individuals. Thus, 4,502, 66,687, and 100,850 SNPs shared by the individuals of *Bradypus*, *Cyclopes*, *Dasypus* respectively. Population statistics - We used the script ABCstat_global.txt from the DILSmcsnp program (last accessed 1st October 2022; Fraïsse et al. 2021) to estimate multiple population genomics statistics. Thus, the mean pairwise divergence (Dxy) and net divergence (computed as c, where $Pi1$ and $Pi2$ are the pairwise nucleotide diversity of populations 1 and 2, respectively) were evaluated between lineages or close relative species. We also used the script seq_stat_2pop_2N to estimate complementary population genetic statistics based on two diploid genomes (Allio et al. 2021). For each of the two individuals ($Pi1$, $Pi2$), the heterozygosity was assessed and the total nucleotide diversity was computed as the mean pairwise divergence between all chromosomes, within and between individuals ($PiTot$) (`https://github.com/benoitnabholz/seq_stat_2pop`). Finally, a FST-like statistic, the Genetic Differentiation Index (Allio et al. 2021), was estimated as $1 - \frac{(Pi1,Pi2)/2}{PiTot}$.

### Genetic diversity and inbreeding

The genetic diversity of the 94 individuals was estimated using two metrics : i) heterozygocsity (He) and ii) Run of Homozygosity (RoHs). The mean He was calculated as the rate of heterozygous position estimated from the random regions dataset using a homemade script. Then, RoHs were estimated with bcftools v.1.14 using the option "roh" with the allele frequency fixed at 0.1 as there is only one sample in the vcf file and the option "-M" set to 1e-3. Then RoH were gathered into three size categories in order to estimate the cumulative size of fragments from i) $>$ 1 and $<$ 5Mb, ii) $>$ 5 and $<$ 10Mb and iii)$>$10Mb.

### Demographic analysis

The effective population size (Ne) trajectory of 74 individuals was estimated using the Pairwise Sequentially Markovian Coalescent (PSMC v.0.6.5-r67; Li and Durbin 2011) approach. We converted whole genome fasta sequences in PSMC format using the script Fasta2PSMCFasta. Then PSMC ran for 30 iterations, using a maximum 2N0 coalescent time of 15 and an initial theta/rho ratio of 4. We test different pattern of parameters '$4 + 30 \times 2 + 4 + 6 + 10$' (Nadachowska-Brzyska et al., 2013) and $-p$ '$4 + 25 \times 2 + 4 + 6$' (Kim et al., 2016) but also $-p$ '$4 + 10 \times 3 + 4$', $-p$ '$4 + 20 \times 2 + 4 + 6 + 10$'. Finally, PSMC were scaled with the mammalian mutation rate of $\mu = 10^{-5}$ mutation/site/generation (Ekblom et al., 2018; Gopalakrishnan et al., 2017) and a generation time (g) detailed in Table S4). Plots were generated using psmc.results' (Liu and Hansen, 2017) via R v. 4.1.2 (R Core Team 2018) and packages ggplot2 (Wickham 2016) and cowplot (Wilke 2019).

## References

Abba A.M., Cassini G.H., Valverde G., Tilak M.-K., Vizcaino S.F., 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.

Abrantes E.A.L., Bergqvist L.P. 2006. Proposta filogenética para os Dasypodidae (Mammalia: Cingulata). Paleontol. Vertebr. Gd. Temas E Contrib. Científicas V Gallo PM Brito HMA Silva FJ Figueiredo Eds Interciência Ltda Rio Jan. Braz.:261–274.

Allio R., Schomaker-Bastos A., Romiguier J., Prosdocimi F., Nabholz B., Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour. 20:892–905.

Allio R., Tilak M.-K., Scornavacca C., Avenant N.L., Kitchener A.C., Corre E., Nabholz B., Delsuc F. 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. Elife. 10:e63167.

Álvarez-Carretero S., Tamuri A.U., Battini M., Nascimento F.F., Carlisle E., Asher R.J., Yang Z., Donoghue P.C., Dos Reis M. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. Nature. 602:263–267.

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.

Arteaga M.C., Gasca-Pineda J., Bello-Bedoy R., Eguiarte L.E., Medellin R.A. 2020. Conservation genetics, demographic history, and climatic distribution of the nine-banded armadillo (*Dasypus novemcinctus*): an analysis of its mitochondrial lineages. Conserv. Genet. Mamm. Integr. Res. Using Nov. Approaches.:141–163.

Barthe, M. et al. (n.d.). "Exon capture museomics deciphers the nine-banded armadillo species complex and identifies a new species endemic to the Guiana Shield". In: Systematic Biology. In revision.

Billet G., Hautier L., De Thoisy B., Delsuc F. 2017. The hidden anatomy of paranasal sinuses reveals biogeographically distinct morphotypes in the nine-banded armadillo (*Dasypus novemcinctus*). PeerJ. 5:e3593.

Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4:e1660.

Burgin C.J., Colella J.P., Kahn P.L., Upham N.S. 2018. How many species of mammals are there? J. Mammal. 99:1–14.

Cetica P.D., Solari A.J., Merani M.S., De R.J., Burgos M.H. 1998. Evolutionary sperm morphology and morphometry in armadillos. J. Submicrosc. Cytol. Pathol. 30:309–314.

Chavez D.E., Gronau I., Hains T., Dikow R.B., Frandsen P.B., Figueiró H.V., Garcez F.S., Tchaicka L., de Paula R.C., Rodrigues F.H. 2022. Comparative genomics uncovers the evolutionary history, demography, and molecular adaptations of South American canids. Proc. Natl. Acad. Sci. 119:e2205986119.

Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890.

Christmas M.J., Kaplow I.M., Genereux D.P., Dong M.X., Hughes G.M., Li X., Sullivan P.F., Hindle A.G., Andrews G., Armstrong J.C. 2023. Evolutionary constraint and innovation across hundreds of placental mammals. Science. 380:eabn3943.

Coimbra R.T.F., Miranda F.R., Lara C.C., Schetino M.A.A., Santos F.R. dos. 2017. Phylogeographic history of South American populations of the silky anteater *Cyclopes didactylus* (Pilosa: Cyclopedidae). Genet. Mol. Biol. 40:40–49.

Cortés-Ortiz L., Bermingham E., Rico C., Rodrigúez-Luna E., Sampaio I., Ruiz-Garcià M. 2003. Molecular systematics and biogeography of the Neotropical monkey genus, Alouatta. Mol. Phylogenet. Evol. 26:64–81.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10:giab008.

Darwin Tree Of Life Project Consortium D.T. of L.P. 2022. Sequence locally, think globally: the Darwin Tree of Life Project. Proc. Natl. Acad. Sci. 119:e2115642118.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Delsuc F., Ctzeflis 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. Proc. R. Soc. Lond. B Biol. Sci. 268:1605–1615.

Delsuc F., Douzery E.J. 2008. Recent advances and future prospects in xenarthran molecular phylogenetics. Biol. Xenarthra. 11.

Delsuc F., Gibb G.C., Kuch M., Billet G., Hautier L., Southon J., Rouillard J.-M., Fernicola J.C., Vizcaíno S.F., MacPhee R.D. 2016. The phylogenetic affinities of the extinct glyptodonts. Curr. Biol. 26:R155–R156.

Delsuc F., Kuch M., Gibb G.C., Hughes J., Szpak P., Southon J., Enk J., Duggan A.T., Poinar H.N. 2018. Resolving the phylogenetic position of Darwin's extinct ground sloth (*My-*

*lodon darwinii*) using mitogenomic and nuclear exon data. Proc. R. Soc. B Biol. Sci. 285:20180214.

Delsuc F., Kuch M., Gibb G.C., Karpinski E., Hackenberger D., Szpak P., Martínez J.G., Mead J.I., McDonald H.G., MacPhee R.D. 2019. Ancient mitogenomes reveal the evolutionary history and biogeography of sloths. Curr. Biol. 29:2031-2042. e6.

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. Mol. Biol. Evol. 19:1656–1671.

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. Mol. Phylogenet. Evol. 28:261–275.

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. Mol. Phylogenet. Evol. 62:673–680.

Delsuc F., Vizcaíno S.F., Douzery E.J. 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. BMC Evol. Biol. 4:1–13.

Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K., Hoeger M., Durand N.C., Shamim M.S., Machol I., Lander E.S., Aiden A.P. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95.

Engelmann G. 1985. The phylogeny of Xenarthra. Evol. Ecol. Armadillos Sloths Vermilinguas.:51–64.

Esquerré D., Brennan I.G., Catullo R.A., Torres-Pérez F., Keogh J.S. 2019. How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae). Evolution. 73:214–230.

Ezard T., Fujisawa T., Barraclough T.G. 2009. Splits: species' limits by threshold statistics. R Package Version. 1:r29.

Feijo A., Anacleto T.C. 2021. Taxonomic revision of the genus *Cabassous* McMurtrie, 1831 (Cingulata: Chlamyphoridae), with revalidation of *Cabassous squamicaudis* (Lund, 1845). Zootaxa. 4974:47-78-47–78.

Feijó A., Brandão M.V. 2022. Taxonomy as the first step towards conservation: an appraisal on the taxonomy of medium-and large-sized Neotropical mammals in the 21 st century. SciELO Brasil.

Feijo A., Cordeiro-Estrela P. 2016. Taxonomic revision of the *Dasypus kappleri* complex, with revalidations of *Dasypus pastasae* (Thomas, 1901) and *Dasypus beniensis* Lönnberg, 1942 (Cingulata, Dasypodidae). Zootaxa. 4170:271–297.

Feijó A., Ge D., Wen Z., Xia L., Yang Q. 2022. Identifying hotspots and priority areas for xenarthran research and conservation. Divers. Distrib. 28:2778–2790.

Feijó A., Patterson B.D., Cordeiro-Estrela P. 2018. Taxonomic revision of the long-nosed armadillos, Genus *Dasypus* Linnaeus, 1758 (Mammalia, Cingulata). PLoS One. 13:e0195084.

Feijó A., Vilela J.F., Cheng J., Schetino M.A.A., Coimbra R.T., Bonvicino C.R., Santos F.R., Patterson B.D., Cordeiro-Estrela P. 2019. Phylogeny and molecular species delimitation of long-nosed armadillos (*Dasypus*: Cingulata) supports morphology-based taxonomy. Zool. J. Linn. Soc. 186:813–825.

Figueiró H.V., Li G., Trindade F.J., Assis J., Pais F., Fernandes G., Santos S.H., Hughes G.M., Komissarov A., Antunes A. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 3:e1700299.

Foley N.M., Mason V.C., Harris A.J., Bredemeyer K.R., Damas J., Lewin H.A., Eizirik E., Gatesy J., Karlsson E.K., Lindblad-Toh K. 2023. A genomic timescale for placental mammal evolution. Science. 380:eabl8189.

Fouquet A., Noonan B.P., Rodrigues M.T., Pech N., Gilles A., Gemmell N.J. 2012. Multiple quaternary refugia in the eastern Guiana Shield revealed by comparative phylogeography of 12 frog species. Syst. Biol. 61:461.

Fraïsse C., Popovic I., Mazoyer C., Spataro B., Delmotte S., Romiguier J., Loire E., Simon A., Galtier N., Duret L. 2021. DILS: Demographic inferences with linked selection by using ABC. Mol. Ecol. Resour. 21:2629–2644.

Galtier N. 2023. Phylogenetic conflicts: distinguishing gene flow from incomplete lineage sorting. bioRxiv.:2023.07. 06.547897.

Gardner A.L. 2005. Mammal Species of the World: a taxonomic and geographic reference. The Johns Hopkins University Press, Baltimore, Estados Unidos,. .

Gardner A.L. 2007. Mammals of South America, volume 1: marsupials, xenarthrans, shrews, and bats. University of Chicago Press.

Garrison E., Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv Prepr. ArXiv12073907.

Gaudin T.J., Wible J.R. 2006. The Phylogeny of Living and Extinct Armadillos (Mammalia, Xenarthra, Cingulata): A Craniodental Analysis. Amniote paleobiology: perspectives on the evolution of mammals, birds and reptiles,. University of Chicago Press. p. 153–198.

Ge D., Wen Z., Feijó A., Lissovsky A., Zhang W., Cheng J., Yan C., She H., Zhang D., Cheng Y. 2023. Genomic consequences of and demographic response to pervasive hybridization over time in climate-sensitive pikas. Mol. Biol. Evol. 40:msac274.

Giarla T.C., Jansa S.A. 2014. The role of physical geography and habitat type in shaping the biogeographical history of a recent radiation of Neotropical marsupials (Thylamys: Didelphidae). J. Biogeogr. 41:1547–1558.

Gibb G.C., Condamine F.L., Kuch M., Enk J., Moraes-Barros N., Superina M., Poinar H.N., Delsuc F. 2016. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. Mol. Biol. Evol. 33:621–642.

Gutiérrez-García T.A., Vázquez-Domínguez E. 2013. Consensus between genes and stones in the biogeographic and evolutionary history of Central America. Quat. Res. 79:311–324.

Hammond D.S. 2005. Biophysical features of the Guiana shield. Trop. For. Guiana Shield Anc. For. Mod. World.:15–194.

Hautier L., Billet G., De Thoisy B., Delsuc F. 2017. Beyond the carapace: skull shape variation and morphological systematics of long-nosed armadillos (genus *Dasypus*). PeerJ. 5:e3650.

Heighton S.P., Allio R., Murienne J., Salmona J., Meng H., Scornavacca C., Bastos A.D., Njiokou F., Pietersen D.W., Tilak M.-K. 2023. Pangolin genomes offer key insights and resources for the world's most trafficked wild mammals. bioRxiv.:2023.02. 16.528682.

Hollowell T., Reynolds R.P. 2005. Checklist of the terrestrial vertebrates of the Guiana Shield. Biological Society of Washington, National Museum of Natural History.

Hoorn C., Wesselingh F.P., Ter Steege H., Bermudez M.A., Mora A., Sevink J., Sanmartín I., Sanchez-Meseguer A., Anderson C.L., Figueiredo J.P. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. science. 330:927–931.

Johnson J., Muren E., Swofford R., Turner-Maier J., Marinescue V.D., Genereux D.P., Lindblad-Toh K. 2018. The 200 mammals project: sequencing genomes by a novel cost-effective method, yielding a high-resolution annotation of the human genome. Unpubl.

Kalyaanamoorthy S., Minh B.Q., Wong T.K., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Kaplow I.M., Lawler A.J., Schäffer D.E., Srinivasan C., Sestili H.H., Wirthlin M.E., Phan B.N., Prasad K., Brown A.R., Zhang X. 2023. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. Science. 380:eabm7993.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kelloff C.L., Funk V.A. 2004. Phytogeography of the Kaieteur Falls, Potaro Plateau, Guyana: floral distributions and affinities. J. Biogeogr. 31:501–513.

Kuderna L.F., Gao H., Janiak M.C., Kuhlwilm M., Orkin J.D., Bataillon T., Manu S., Valenzuela A., Bergman J., Rousselle M. 2023. A global catalog of whole-genome diversity from 233 primate species. Science. 380:906–913.

Kumar V., Lammers F., Bidon T., Pfenninger M., Kolter L., Nilsson M.A., Janke A. 2017. The evolutionary history of bears is characterized by gene flow across species. Sci. Rep. 7:46487.

Larsen P.A., Matocq M.D. 2019. Emerging genomic applications in mammalian ecology, evolution, and conservation. J. Mammal. 100:786–801.

Lee H., Gurtowski J., Yoo S., Nattestad M., Marcus S., Goodwin S., Richard McCombie W., Schatz M.C. 2016. Third-generation sequencing and the future of genomics. BioRxiv.:048603.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv Prepr. ArXiv13033997.

Li H., Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature. 475:493–496.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Lujan N.K., Armbruster J.W., Albert J.S., Reis R.E. 2011. The guiana shield. Hist. Biogeogr. Neotropical Freshw. Fishes. 211:224.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

McKenna M.C., Bell S.K. 1997. Classification of mammals: above the species level. Columbia University Press.

Meyer M., Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010:pdb. prot5448.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.

Miranda F.R., Casali D.M., Perini F.A., Machado F.A., Santos F.R. 2018. Taxonomic review of the genus *Cyclopes* Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species. Zool. J. Linn. Soc. 183:687–721.

Miranda F.R., Garbino G.S., Machado F.A., Perini F.A., Santos F.R., Casali D.M. 2023. Taxonomic revision of maned sloths, subgenus *Bradypus* (Scaeopus), Pilosa, Bradypodidae, with revalidation of *Bradypus crinitus* Gray, 1850. J. Mammal. 104:86–103.

Mitchell K.J., Scanferla A., Soibelzon E., Bonini R., Ochoa J., Cooper A. 2016. Ancient DNA from the extinct South American giant glyptodont *Doedicurus* sp.(Xenarthra: Glyptodontidae) reveals that glyptodonts evolved from Eocene armadillos. Mol. Ecol. 25:3499–3508.

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). Mol. Biol. Evol. 24:2573–2582.

Moraes-Barros N., Arteaga M.C. 2015. Genetic diversity in Xenarthra and its relevance to patterns of neotropical biodiversity. J. Mammal. 96:690–702.

Osmanski A.B., Paulat N.S., Korstian J., Grimshaw J.R., Halsey M., Sullivan K.A., Moreno-Santillán D.D., Crookshanks C., Roberts J., Garcia C. 2023. Insights into mammalian TE diversity through the curation of 248 genome assemblies. Science. 380:eabn1430.

Pacifici M., Santini L., Di Marco M., Baisero D., Francucci L., Marasini G.G., Visconti P., Rondinini C. 2013. Generation length for mammals. Nat. Conserv. 5:89–94.

Patterson B., Segall W., Turnbull W.D. 1989. ear region in Xenarthrans (= Edentata: Mammalia). .

Picard Toolkit. 2019. Broad institute, GitHub repository. .

Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sumlin W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55:595–609.

R Core Team. 2018. R: A language and environment for statistical computing. .

Ramirez J.L., Lescroart J., Figueiró H.V., Torres-Florez J.P., Villela P.M., Coutinho L.L., Freitas P.D., Johnson W.E., Antunes A., Galetti Jr P.M. 2022. Genomic signatures of divergent ecological strategies in a recent radiation of Neotropical wild cats. Mol. Biol. Evol. 39:msac117.

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.

dos Reis M.D., Gunnell G.F., Barba-Montoya J., Wilkins A., Yang Z., Yoder A.D. 2018. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. Syst. Biol. 67:594–615.

Rhie A., McCarthy S.A., Fedrigo O., Damas J., Formenti G., Koren S., Uliano-Silva M., Chow W., Fungtammasan A., Kim J. 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 592:737–746.

Ruiz-García M., Chacón D., Plese T., Shostell J.M. 2020. Molecular phylogenetics of *Bradypus* (three-toed sloth, Pilosa: Bradypodidae, Mammalia) and phylogeography of *Bradypus variegatus* (brown-throated three-toed sloth) with mitochondrial gene sequences. J. Mamm. Evol. 27:461–482.

Ruiz-García M., Pinilla-Beltrán D., Murillo-García O.E., Pinto C.M., Brito J., Shostell J.M. 2021. Comparative mitogenome phylogeography of two anteater genera (*Tamandua* and *Myrmecophaga*; Myrmecophagidae, Xenarthra): Evidence of discrepant evolutionary traits. Zool. Res. 42:525.

Shao Y., Zhou L., Li F., Zhao L., Zhang B.-L., Shao F., Chen J.-W., Chen C.-Y., Bi X., Zhuang X.-L. 2023. Phylogenomic analyses provide insights into primate evolution. Science. 380:913–924.

Smith S.A., Brown J.W., Walker J.F. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PloS One. 13:e0197433.

Superina M., Abba A.M. 2020. Conservation perspectives for a highly disparate lineage of mammals: the Xenarthra. Mastozool. Neotropical. 27:48–67.

Tejada J.V., Antoine P.-O., Münch P., Billet G., Hautier L., Delsuc F., Condamine F.L. 2023. Bayesian total-evidence dating revisits sloth phylogeny and biogeography: a cautionary tale on morphological clock analyses. Syst. Biol.:syad069.

Theissinger K., Fernandes C., Formenti G., Bista I., Berg P.R., Bleidorn C., Bombarely A., Crottini A., Gallo G.R., Godoy J.A. 2023. How genomics can help biodiversity conservation. Trends Genet. 39:545–559.

Tilak M.-K., Justy F., Debiais-Thibaud M., Botero-Castro F., Delsuc F., Douzery E.J. 2015. A cost-effective straightforward protocol for shotgun Illumina libraries designed to

assemble complete mitogenomes from non-model species. Conserv. Genet. Resour. 7:37–40.

Tricou T., Tannier E., de Vienne D.M. 2022a. Ghost lineages highly influence the interpretation of introgression tests. Syst. Biol. 71:1147–1158.

Tricou T., Tannier E., de Vienne D.M. 2022b. Ghost lineages can invalidate or even reverse findings regarding gene flow. PLoS Biol. 20:e3001776.

Trigo T.C., Freitas T.R.O., Kunzler G., Cardoso L., Silva J.C.R., Johnson W.E., O'Brien S.J., Bonatto S.L., Eizirik E. 2008. Inter-species hybridization among Neotropical cats of the genus Leopardus, and evidence for an introgressive hybrid zone between L. geoffroyi and L. tigrinus in southern Brazil. Mol. Ecol. 17:4317–4333.

Trigo T.C., Schneider A., de Oliveira T.G., Lehugeur L.M., Silveira L., Freitas T.R., Eizirik E. 2013. Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. Curr. Biol. 23:2528–2533.

Trindade F.J., Rodrigues M.R., Figueiró H.V., Li G., Murphy W.J., Eizirik E. 2021. Genome-wide SNPs clarify a complex radiation and support recognition of an additional cat species. Mol. Biol. Evol. 38:4987–4991.

Upham N.S., Landis M.J. 2023. Genomics expands the mammalverse. Science. 380:358–359.

Vallejos-Garrido P., Pino K., Espinoza-Aravena N., Pari A., Inostroza-Michael O., Toledo-Muñoz M., Castillo-Ravanal B., Romero-Alarcón V., Hernández C.E., Palma R.E. 2023. The importance of the Andes in the evolutionary radiation of Sigmodontinae (Rodentia, Cricetidae), the most diverse group of mammals in the Neotropics. Sci. Rep. 13:2207.

Vanderpool D., Minh B.Q., Lanfear R., Hughes D., Murali S., Harris R.A., Raveendran M., Muzny D.M., Hibbins M.S., Williamson R.J. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol. 18:e3000954.

Voirin B. 2015. Biology and conservation of the pygmy sloth, *Bradypus pygmaeus*. J. Mammal. 96:703–707.

Weir J.T., Price M. 2011. Andean uplift promotes lowland speciation through vicariance and dispersal in Dendrocincla woodcreepers. Mol. Ecol. 20:4550–4563.

Wetzel R.M., Gardner A.L., Redford K.H., Eisenberg J.F. 2008. Order Cingulata. Mamm. S. Am. 1:128–156.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. .

Wilder A.P., Supple M.A., Subramanian A., Mudide A., Swofford R., Serres-Armero A., Steiner C., Koepfli K.-P., Genereux D.P., Karlsson E.K. 2023. The contribution of historical processes to contemporary extinction risk in placental mammals. Science. 380:eabn5856.

Wilke C.O. 2019. cowplot: streamlined plot theme and plot annotations for 'ggplot2.' R Package Version. 1.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Zhang J., Kapli P., Pavlidis P., Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 29:2869–2876.

Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. Nature. 587:240–245.

# Supplementary

**Table S1:** List of mitochondrial sequences used, detailing information on the depth of coverage for individuals mapped, type of sequences, their specimen vouchers, and reference. Mitochondrial assemblies used as reference mitogenomes for mapping are represented in bold. Cov. : Coverage; UNVERIF. : UNVERIFIED; unpub.: unpublished; In rev. : In revision.

| Species | ID | Coverage | Type | Voucher/Origin | Reference |
|---|---|---|---|---|---|
| *Acratocnus ye* | NC_042752.1 | NA | SUBFOSSIL | UF 76365 | Delsuc et al. 2019 |
| *Megalonyx jeffersonii* | NC_042736.1 | NA | SUBFOSSIL | PMA P98.6.28 | Delsuc et al. 2019 |
| *Megatherium americanum* | NC_042737.1 | NA | SUBFOSSIL | C.2C_Layer 2 | Delsuc et al. 2019 |
| *Megatherium americanum* | MK903497.1 | NA | SUBFOSSIL | C.2E_Layer 4_1 | Delsuc et al. 2019 |
| *Megatherium americanum* | MK903498.1 | NA | SUBFOSSIL | C.2E_Layer 4_2 | Delsuc et al. 2019 |
| *Megatherium americanum* | MK903499.1 | NA | SUBFOSSIL | MAPB4R 3965 | Delsuc et al. 2019 |
| *Mylodon darwinii* | NC_037941.1 | NA | SUBFOSSIL | NHMUK:PV M8758 | Delsuc et al. 2018 |
| *Mylodon darwinii* | MK903500.1 | NA | SUBFOSSIL | MNHN 1905-4 | Delsuc et al. 2019 |
| *Mylodon darwinii* | KR336794 | NA | SUBFOSSIL | NA | Slater et al. 2016 |
| *Nothrotheriops shastensis* | NC_042753.1 | NA | SUBFOSSIL | RC L12 #1 | Delsuc et al. 2019 |
| *Parocnus serus* | NC_042754.1 | NA | SUBFOSSIL | UF 75452 | Delsuc et al. 2019 |
| *Choloepus hoffmanni* | DNAZoo | 51.86 | WGS | NA | NA |
| *Choloepus hoffmanni* | NMB11 | 157.88 | Mito | ChMT | This study |
| *Choloepus hoffmanni* | T6052 | NA | Mito | UM ISEM T-6052 | Gibb et al. 2016 |
| *Choloepus hoffmanni* | SLA03 | NA | Mito | KR336793 | Slater et al. 2016 |
| *Choloepus hoffmanni* | GB03 | NA | Mito | NC_027964 | Song et al. 2015 |
| *Choloepus hoffmanni* | MF616393.1 | NA | UNVERIFIED | Panama | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616394.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616395.1 | NA | UNVERIFIED | Panama | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616396.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616397.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616398.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616399.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Choloepus hoffmanni* | MF616400.1 | NA | UNVERIFIED | Costa Rica | Ruiz-Garcia et al. 2017 |
| *Choloepus didactylus* | ChoDid_v1_BIUU | 628.48 | WGS | NA | Johnson et al. 2018 |
| *Choloepus didactylus* | mchoDid1 | 45.04 | WGS | CM026720.1 | NA |
| *Choloepus didactylus* | GB02 | NA | Mito | NC_006924 | McLenachan & Penny (unpublished) |
| *Choloepus didactylus* | MF616392.1 | NA | UNVERIFIED | Venezuela | Ruiz-Garcia et al. 2017 |
| *Choloepus didactylus* | MF616391.1 | NA | UNVERIFIED | Suriname | Ruiz-Garcia et al. 2017 |
| *Choloepus didactylus* | SLA02 | NA | Mito | KR336792 | Slater et al. 2016 |
| *Choloepus didactylus* | T1722 | NA | Mito | MNHN 1998-1819 | Gibb et al. 2016 |
| *Bradypus torquatus* | MF582364.1 | NA | UNVERIFIED | NA | Ruiz-Garcia et al. 2017 |
| *Bradypus torquatus* | BMNH39498 | 36.80 | WGS | BMNH 3.9.4.98 | This study |
| *Bradypus torquatus* | L41 | 65.75 | WGS | Tube 449/11 | This study |
| *Bradypus pygmaeus* | MF582362.1 | NA | UNVERIFIED | NA | Ruiz-Garcia et al. 2017 |
| *Bradypus pygmaeus* | BPY171 | 123.95 | WGS | USNM 579171 | This study |
| *Bradypus pygmaeus* | L32 | 490.98 | WGS | USNM 579179 | This study |
| *Bradypus tridactylus* | MF582363.1 | NA | UNVERIFIED | NA | Ruiz-Garcia et al. 2017 |
| *Bradypus tridactylus* | NMB1 | 122.17 | WGS | AP98 | This study |
| *Bradypus tridactylus* | T7029 | 441.21 | WGS | JAG M1664 | Allio 2021 |
| *Bradypus tridactylus* | NMB13 | 72.80 | Mito | BT2903 | This study |
| *Bradypus tridactylus* | NMB14 | 127.57 | Mito | BT0103 | This study |
| *Bradypus tridactylus* | ISEM T5013 | NA | Mito | G-764 | Allio 2021 |
| *Bradypus variegatus* | MF582353.1 | NA | UNVERIFIED | Peru | Ruiz-Garcia et al. 2017 |

| | | | | | |
|---|---|---|---|---|---|
| *Bradypus variegatus* | MF582354.1 | NA | UNVERIFIED | Venezuela | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582355.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582356.1 | NA | UNVERIFIED | Colombia | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582357.1 | NA | UNVERIFIED | Brazil | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582358.1 | NA | UNVERIFIED | Brazil | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582359.1 | NA | UNVERIFIED | Brazil | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582360.1 | NA | UNVERIFIED | Brazil | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | MF582361.1 | NA | UNVERIFIED | Brazil | Ruiz-Garcia et al. 2017 |
| *Bradypus variegatus* | BVA166 | 1224.42 | WGS | USNM 579166 | This study |
| *Bradypus variegatus* | BVA2 | 165.11 | WGS | BVARII | This study |
| *Bradypus variegatus* | BVA379 | 1721.71 | WGS | USNM 575379 | This study |
| *Bradypus variegatus* | BVA417 | 426.52 | WGS | USNM 578417 | This study |
| *Bradypus variegatus* | BVA420 | 900.63 | WGS | USNM 578420 | This study |
| *Bradypus variegatus* | BVA4 | 67.17 | WGS | ACR02111 | This study |
| *Bradypus variegatus* | BVA525 | 6766.25 | WGS | USNM 449525 | This study |
| *Bradypus variegatus* | BVA544 | 477.74 | WGS | USNM 449544 | This study |
| *Bradypus variegatus* | BVA5 | 36.61 | WGS | PNB4 | This study |
| *Bradypus variegatus* | BVA850 | 6699.61 | WGS | USNM 464850 | This study |
| *Bradypus variegatus* | BVA853 | 836.62 | WGS | USNM 464853 | This study |
| *Bradypus variegatus* | BVACal7 | 19.37 | WGS | BMNH 1923.8.10.11 | This study |
| *Bradypus variegatus* | KU317 | 22.43 | WGS | KU 106317 | This study |
| *Bradypus variegatus* | KU357 | 38.33 | WGS | KU 96357 | This study |
| *Bradypus variegatus* | NMB7 | 734.38 | WGS | PNM2 | This study |
| *Bradypus variegatus* | NMB8 | 782.20 | WGS | PNM5 | This study |
| *Bradypus variegatus* | BMNH4242920 | 43.83 | WGS | BMNH 42.4.29.20 | This study |
| *Bradypus variegatus* | BMNH805655 | 21.14 | WGS | BMNH 80.5.6.55 | This study |
| *Bradypus variegatus* | KU190 | 38.10 | WGS | KU 145190 | This study |
| *Bradypus variegatus* | BMNH805683 | 24.63 | WGS | BMNH 80.5.6.83 | This study |
| *Bradypus variegatus* | NMB2 | 33.09 | WGS | DPV 02008 | This study |
| *Bradypus variegatus* | NMB3 | 202.98 | WGS | DPV 02104 | This study |
| *Bradypus variegatus* | BraVar_v1_BIUU | 458.99 | WGS | NA | Johnson et al. 2018 |
| *Bradypus variegatus* | BMNH_1381012_I6 | 171.44 | Mito | BMNH 13.8.10.12 | This study |
| *Bradypus variegatus* | BMNH_2412617_I34 | 549.19 | Mito | BMNH 24.12.6.17 | This study |
| *Bradypus variegatus* | BMNH_25102116_I30 | 13.86 | Mito | BMNH 25.10.21.16 | This study |
| *Bradypus variegatus* | BMNH_4242920_I32 | 299.13 | Mito | BMNH 42.4.29.20 | This study |
| *Bradypus variegatus* | BMNH_453115_I19 | 144.72 | Mito | BMNH 45.3.11.5 | This study |
| *Bradypus variegatus* | BMNH_4672824_I21 | 144.10 | Mito | BMNH 46.7.28.24 | This study |
| *Bradypus variegatus* | BMNH_477183_I2 | 29.43 | Mito | BMNH 47.7.18.3 | This study |
| *Bradypus variegatus* | BMNH_697197_I15 | 01/06/40 | Mito | BMNH 69.7.19.7 | This study |
| *Bradypus variegatus* | BMNH_71482_I31 | 360.31 | Mito | BMNH 71.4.8.2 | This study |
| *Bradypus variegatus* | BMNH_805655_I20 | 79.88 | Mito | BMNH 80.5.6.55 | This study |
| *Bradypus variegatus* | BMNH_805656_I17 | 167.06 | Mito | BMNH 80.5.6.56 | This study |
| *Bradypus variegatus* | BVA1 | 49.99 | Mito | 2085 | This study |
| *Bradypus variegatus* | NMB4 | 36.65 | Mito | DPV 02117 | This study |
| *Bradypus variegatus* | NMB5 | 25/10/24 | Mito | AC109 | This study |
| *Bradypus variegatus* | NMB6 | 21.46 | Mito | DPV 02153 | This study |
| *Bradypus variegatus* | BvaCal9 | 30.55 | Mito | BMNH 1923.8.10.13 | This study |
| *Bradypus variegatus* | BvaCal13 | 49.70 | Mito | BMNH 1923.8.10.21 | This study |
| *Bradypus variegatus* | T2999 | NA | Mito | MVZ 155186 | Gibb et al. 2016 |
| *Bradypus variegatus* | SLA01 | NA | Mito | KR336791 | Slater et al. 2016 |
| *Bradypus variegatus* | AY960979 | NA | Mito | GB01 | McLenachan & Penny (unpublished) |

| | | | | | |
|---|---|---|---|---|---|
| *Cyclopes catellus* | BMNH2611217 | 24.13 | WGS | BMNH 26.1.12.17 | This study |
| *Cyclopes catellus* | BMNH261522 | 21.42 | WGS | BMNH 26.1.5.22 | This study |
| *Cyclopes didactylus* | BMNH105447 | 76.73 | WGS | BMNH 10.5.4.47 | This study |
| *Cyclopes didactylus* | BMNH261213 | 38.23 | WGS | BMNH 26.1.2.13 | This study |
| *Cyclopes didactylus* | BMNH310180 | 109.85 | WGS | BMNH 3.10.1.80 | This study |
| *Cyclopes didactylus* | BMNH811187 | 36.95 | WGS | BMNH 8.11.18.7 | This study |
| *Cyclopes didactylus* | T1724 | 56.10 | WGS | UM ISEM T-1724 | This study |
| *Cyclopes didactylus* | M2300 | 57.18 | WGS | JAG M2300 | Allio 2021 |
| *Cyclopes didactylus* | BMNH_261212 | 84.85 | Mito | BMNH 26.1.2.12 | This study |
| *Cyclopes didactylus* | BMNH_811184 | 41.50 | Mito | BMNH 8.11.18.4 | This study |
| *Cyclopes didactylus* | MZS_03612 | 4877.74 | Mito | MZS 03612 | This study |
| *Cyclopes didactylus* | T2073 | NA | Mito | MNHN_1998-234 | Gibb et al. 2016 |
| *Cyclopes dorsalis* | BMNH90146 | 22.34 | WGS | BMNH 90.1.4.6 | This study |
| *Cyclopes dorsalis* | BMNH805686 | 16.91 | WGS | BMNH 80.5.6.86 | This study |
| *Cyclopes dorsalis* | FMNH156654 | 119.47 | WGS | FMNH 156654 | Allio 2021 |
| *Cyclopes dorsalis* | BMNH_165113 | 22.73 | Mito | BMNH 16.5.11.3 | This study |
| *Cyclopes dorsalis* | BMNH_27263 | 3099.11 | Mito | BMNH 2.7.26.3 | This study |
| *Cyclopes dorsalis* | BMNH_6551814 | 117.74 | Mito | BMNH 65.5.18.14 | This study |
| *Cyclopes ida* | BMNH241084 | 32.21 | WGS | BMNH 24.10.8.4 | This study |
| *Cyclopes ida* | MVZ157801 | 56.07 | WGS | MVZ 157801 | This study |
| *Cyclopes ida* | BMNH_805667 | 139.54 | Mito | BMNH 80.5.6.67 | This study |
| *Cyclopes ida* | BMNH_805670 | 200.32 | Mito | BMNH 80.5.6.70 | This study |
| *Cyclopes thomasi* | MVZ190355 | 33.76 | WGS | MVZ 190355 | Allio 2021 |
| *Cyclopes thomasi* | MVZ193691 | 164.63 | Mito | MVZ 193691 | This study |
| *Myrmecophaga tridactyla* | MH142215.1 | NA | Mito | M002146 | Lemos Queiroz et al. 2018 |
| *Myrmecophaga tridactyla* | M3023 | 19.89 | WGS | JAG M3023 | Allio 2021 |
| *Myrmecophaga tridactyla* | MyrTri_v1_BIUU | 664.31 | WGS | NA | Johnson et al. 2018 |
| *Myrmecophaga tridactyla* | M5293 | 224.41 | Mito | JAG M5293 | Teullet et al. 2023 |
| *Myrmecophaga tridactyla* | T2862 | NA | Mito | NC_028572 | Gibb et al. 2016 |
| *Myrmecophaga tridactyla* | M5295 | 278.73 | Mito | JAG M5295 | Teullet et al. 2023 |
| *Tamandua mexicana* | TAM5430 | 89.68 | WGS | NC_028574 | Allio 2021 |
| *Tamandua tetradactyla* | TamTet_v1_BIUU | 74.10 | WGS | NA | Johnson et al. 2018 |
| *Tamandua tetradactyla* | M3075 | 100.35 | WGS | JAG M3075 | Allio 2021 |
| *Tamandua tetradactyla* | mTamTet1 | 15.96 | WGS | CM043432.1 | Uliano-Silva et al. (unpublished) |
| *Tamandua tetradactyla* | M5331 | 433.96 | Mito | JAG M5331 | Teullet et al. 2023 |
| *Tamandua tetradactyla* | M5584 | 288.80 | Mito | JAG M5584 | Teullet et al. 2023 |
| *Tamandua tetradactyla* | KT818552 | NA | Mito | UM ISEM T-6054 | Gibb et al. 2016 |
| *Tamandua tetradactyla* | NC_004032.1 | NA | Mito | NA | Arnason et al. 2002 |
| *Dasypus kappleri* | T1941 | 22.23 | WGS | UM ISEM T-1941 | This study |
| *Dasypus kappleri* | T3365 | 114.47 | WGS | UM ISEM T-3365 | This study |
| *Dasypus kappleri* | M3462 | 65.63 | WGS | JAG M3462 | This study |
| *Dasypus kappleri* | DKA_T4339 | 200.09 | Mito | UM ISEM T-4339 | Barthe et al. (Accepted) |
| *Dasypus septemcinctus* | PAP74 | 27.34 | WGS | UM ISEM T-PAP74 | This study |
| *Dasypus septemcinctus* | PAP43 | 115.47 | WGS | UM ISEM T-PAP43 | This study |
| *Dasypus sepencinctus* | DSE_PAP42 | 19.59 | Mito | UM ISEM T-PAP42 | Barthe et al. (Accepted) |
| *Dasypus septemcinctus* | T3002 | 103.32 | WGS | UM ISEM T-3002 | This study |
| *Dasypus hybridus* | L38 | 57.09 | WGS | ZVC M2010 | This study |
| *Dasypus guianensis* | M1752 | 94.28 | WGS | JAG M1752 | This study |
| *Dasypus guianensis* | M1766 | 124.34 | WGS | JAG M1766 | This study |
| *Dasypus guianensis* | M1777 | 01/06/48 | Mito | JAG M1777 | Teullet et al. 2023 |
| *Dasypus guianensis* | M2255 | 93.47 | Mito | JAG M2255 | Teullet et al. 2023 |

| | | | | | |
|---|---|---|---|---|---|
| *Dasypus guianensis* | M2865 | 43.25 | Mito | JAG M2865 | Teullet et al. 2023 |
| *Dasypus guianensis* | M3021 | 309.93 | Mito | JAG M3021 | Teullet et al. 2023 |
| *Dasypus guianensis* | T1863 | NA | Mito | KT818542/JAG M5912 | Gibb et al. 2016 |
| *Dasypus guianensis* | DNO_MC333 | 60.82 | Mito | AMNH 42914 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_T1692 | 01/12/42 | Mito | MNHN 2000-226 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_T2465 | 30.41 | Mito | MNHN 2001-1530 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_T2863 | 15.48 | Mito | JAG M5911 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_71 | 24.37 | Mito | JAG M5361 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_T4557 | 14.31 | Mito | JAG M5910 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_AP207 | 190.25 | Mito | IEPA 2837 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_AP249 | 65.17 | Mito | IEPA 3102 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_AP109 | 01/03/49 | Mito | IEPA 2419 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_169 | 01/10/47 | Mito | JAG M5442 | Barthe et al. (Accepted) |
| *Dasypus guianensis* | DNO_T4553 | 01/06/68 | Mito | JAG M5909 | Barthe et al. (Accepted) |
| *Dasypus mazzai* | L24 | 122.71 | WGS | MLP 30.III.90.2 | This study |
| *Dasypus sabanicola* | L33 | 30.49 | WGS | USNM 372834 | This study |
| *Dasypus novemcinctus* | NC_001821.1 | NA | Mito | NA | Arnason et al. 1997 |
| *Dasypus novemcinctus* | MC257 | 35.66 | WGS | USNM 281285 | This study |
| *Dasypus novemcinctus* | AM78 | 46.82 | WGS | IEPA 2560 | This study |
| *Dasypus novemcinctus* | NP276 | 97.85 | WGS | UM ISEM T-NP276 | This study |
| *Dasypus novemcinctus* | T2631 | 74.08 | WGS | UM ISEM T-2631 | This study |
| *Dasypus novemcinctus* | T3380 | 20.94 | WGS | UM ISEM T-3380 | This study |
| *Dasypus novemcinctus* | AJR1 | 82.04 | WGS | UM ISEM T-AJR1 | This study |
| *Dasypus novemcinctus* | MSB2 | 231.02 | WGS | MSB 99077 | This study |
| *Dasypus novemcinctus* | MVZ192698 | 111.71 | WGS | MVZ 192698 | This study |
| *Dasypus novemcinctus* | VLD165 | 728.96 | Mito | UM ISEM T-VLD165 | Teullet et al. 2023 |
| *Dasypus novemcinctus* | VLD172 | 93.94 | Mito | UM ISEM T-VLD172 | Teullet et al. 2023 |
| *Dasypus novemcinctus* | DNO_MC304 | 158.37 | Mito | AMNH 33151 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC417 | 18.35 | Mito | USNM 406703 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC413 | 01/10/77 | Mito | USNM 259485 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_AN1 | 01/06/44 | Mito | UM ISEM T-AN1 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_T1870 | 31.21 | Mito | UM ISEM T-1870 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_T2360 | 14.57 | Mito | UM ISEM T-2360 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_M275 | 186.06 | Mito | INPA SISTAP M275 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC366 | 45.00 | Mito | AMNH 133274 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_PH3 | 38.12 | Mito | AMNH 133266 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC340 | 45.73 | Mito | AMNH 247662 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC1000 | 81.62 | Mito | AMNH 42441 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC441 | 37.44 | Mito | AMNH 98463 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC324 | 158.96 | Mito | AMNH 67710 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC334 | 177.14 | Mito | AMNH 98462 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC337 | 101.97 | Mito | AMNH 98465 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC399 | 01/06/54 | Mito | USNM 406706 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC326 | 52.40 | Mito | AMNH 7549 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC226 | 85.26 | Mito | AMNH 136252 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC402 | 19.45 | Mito | USNM 296615 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC401 | 55.45 | Mito | USNM 442799 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC302 | 25/01/24 | Mito | AMNH 144828 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC392 | 105.92 | Mito | USNM 442787 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_NP751 | 01/02/89 | Mito | UM ISEM T-NP751 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC303 | 21.97 | Mito | AMNH 78517 | Barthe et al. (Accepted) |

| | | | | | |
|---|---|---|---|---|---|
| *Dasypus novemcinctus* | DNO_MC327 | 61.36 | Mito | AMNH 205727 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC355 | 92.31 | Mito | AMNH 91709 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC367 | 29.49 | Mito | AMNH 133330 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_MC436 | 15.77 | Mito | AMNH 211674 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_PH6 | 22.89 | Mito | AMNH 93116 | Barthe et al. (Accepted) |
| *Dasypus novemcinctus* | DNO_T1921 | 52.48 | Mito | UM ISEM T-1921 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_LS5 | 34.93 | Mito | UM ISEM T-LS5 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_MC292 | 01/12/41 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_MC103 | 19.34 | Mito | Eco-SC-M 980 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_YUC1 | 01/11/87 | Mito | UM ISEM T-YUC1 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_MC58 | 19.57 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_MC40 | 50.07 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_BM1 | 15/07/24 | Mito | UM ISEM T-BM1 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_JP1 | 01/08/25 | Mito | UM ISEM T-JP1 | Barthe et al. (Accepted) |
| *Dasypus mexicanus* | DNO_MC163 | 01/05/51 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC151 | 53.46 | Mito | HMAM 532 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC21 | 01/12/98 | Mito | ENCB-IPN 26577 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC176 | 01/09/80 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC88 | 126.78 | Mito | CNMA 37069 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC393 | 47.49 | Mito | USNM 442790 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_PH8 | 136.34 | Mito | AMNH 32356 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC181 | 14.54 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC165 | 30.90 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC425 | 20.52 | Mito | USNM 578437 | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC180 | 43.49 | Mito | NA | Barthe et al. (Accepted) |
| *Dasypus fenestratus* | DNO_MC323 | 30.17 | Mito | AMNH 46555 | Barthe et al. (Accepted) |
| *Dasypus pilosus* | LSUMZ21888 | NA | Mito | LSUMZ 21888 | Gibb et al. 2016 |
| *Dasypus pilosus* | L29 | 55.94 | WGS | MSB 49990 | This study |
| *Dasypus pilosus* | L30 | 01/02/87 | WGS | LSUMZ 19240 | This study |
| *Doedicurus sp.* | KU517659 | NA | FOSSIL | FD-2016a | Delsuc et al. 2016 |
| *Doedicurus sp.* | KX098449 | NA | FOSSIL | NA | Mitchell et al. 2016 |
| *Euphractus sexcinctus* | PAP76 | 70.60 | WGS | UM ISEM T-PAP76 | This study |
| *Euphractus sexcinctus* | KX151979 | NA | Mito | NA | Mitchell et al. 2016 |
| *Euphractus sexcinctus* | ESE_1 | 22.57 | WGS | UM ISEM ZOO_MTP | Allio 2021 |
| *Euphractus sexcinctus* | T1246 | NA | Mito | UM ISEM T-1246 | Gibb et al. 2016 |
| *Chaetophractus villosus* | CHA5373 | 112.85 | WGS | UM ISEM T-2351 | This study |
| *Chaetophractus villosus* | NP390 | 20.51 | WGS | UM ISEM T-NP390 | This study |
| *Chaetophractus villosus* | KX180156 | NA | Mito | NA | Mitchell et al. 2016 |
| *Chaetophractus vellerosus* | NC_028561 | NA | Mito | UM ISEM T-CV1 | Gibb et al. 2016 |
| *Chaetophractus vellerosus* | ChaVel_v1_BIUU_Genbank | 60.28 | WGS | NA | Johnson et al. 2018 |
| *Chaetophractus nationi* | L23 | 44.81 | WGS | UM ISEM T-LP1 | This study |
| *Zaedyus pichiy* | ZP55 | 265.13 | WGS | UM ISEM T-6055 | This study |
| *Zaedyus pichiy* | ZP67 | 54.67 | WGS | UM ISEM T-6060 | This study |
| *Zaedyus pichiy* | KX180157 | NA | Mito | NA | Mitchell et al. 2016 |
| *Chlamyphorus truncatus* | CT1 | 78.41 | WGS | UM ISEM T-CT1 | Allio 2021 |
| *Chlamyphorus truncatus* | CtrCT2-I32 | 25/09/24 | Mito | UM ISEM T-CT2 | This study |
| *Chlamyphorus truncatus* | CtrCT3-I33 | 15.27 | Mito | UM ISEM T-CT3 | This study |
| *Calyptophractus retusus* | ZSM | 71.29 | WGS | ZSM T-Bret | Allio 2021 |
| *Priodontes maximus* | M844 | 27/03/24 | WGS | JAG M844 | Allio 2021 |
| *Priodontes maximus* | L13 | 37.03 | WGS | UM ISEM T-2353 | This study |
| *Tolypeutes tricinctus* | L42 | 30.71 | WGS | JB21 | Gibb et al. 2016 |

| | | | | | |
|---|---|---|---|---|---|
| *Tolypeutes matacus* | T2348 | 24.99 | WGS | UM ISEM T-2348 | This study |
| *Tolypeutes matacus* | TolMat_v1_BIUU | 18.15 | WGS | NA | Johnson et al. 2018 |
| *Tolypeutes matacus* | KX180158 | NA | Mito | NA | Mitchell et al. 2016 |
| *Cabassous tatouay* | PAP77 | 149.89 | WGS | UM ISEM T-PAP77 | This study |
| *Cabassous tatouay* | PAP78 | 30.47 | WGS | UM ISEM T-PAP78 | This study |
| *Cabassous chacoensis* | L12 | 79.89 | WGS | UM ISEM T-2350 | This study |
| *Cabassous centralis* | L28 | 35.55 | WGS | AMNH MO-10752 | This study |
| *Cabassous unicinctus* | T2291 | 18.63 | WGS | UM ISEM T-2291 | This study |
| *Cabassous unicinctus* | KT818531 | NA | Mito | MNHN 1999-1068 | Gibb et al. 2016 |
| *Cabassous unicinctus* | MVZ155190 | 311.78 | WGS | MVZ 155190 | (Allio 2021) |
| *Cabassous unicinctus* | M2809 | 214.49 | Mito | JAG M2809 | Teullet et al. 2023 |
| *Cabassous unicinctus* | M2962 | 95.34 | Mito | JAG M2962 | Teullet et al. 2023 |
| *Cabassous unicinctus* | M3141 | 729.79 | Mito | JAG M3141 | Teullet et al. 2023 |

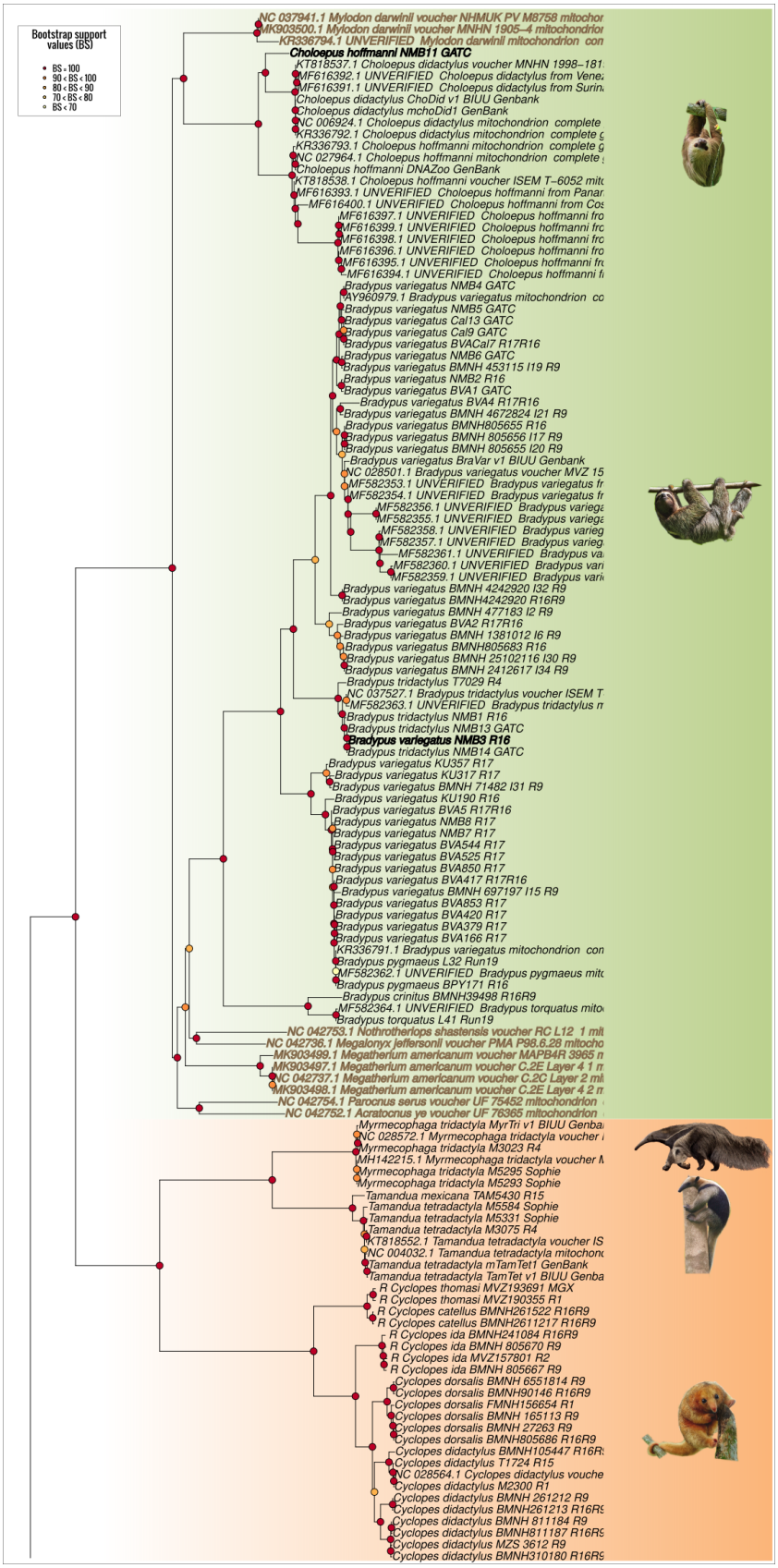


**Figure S1:** Phylogenetic relationships reconstructed by maximum likelihood for the 15 mitochondrial genes partitioned of 261 xenarthran individuals covering 37 extant species and 7 extinct species (highlight in brown). Node circles indicate bootstrap support (BS), the redder the color the higher the BS. Three potential contaminated samples are in bold.

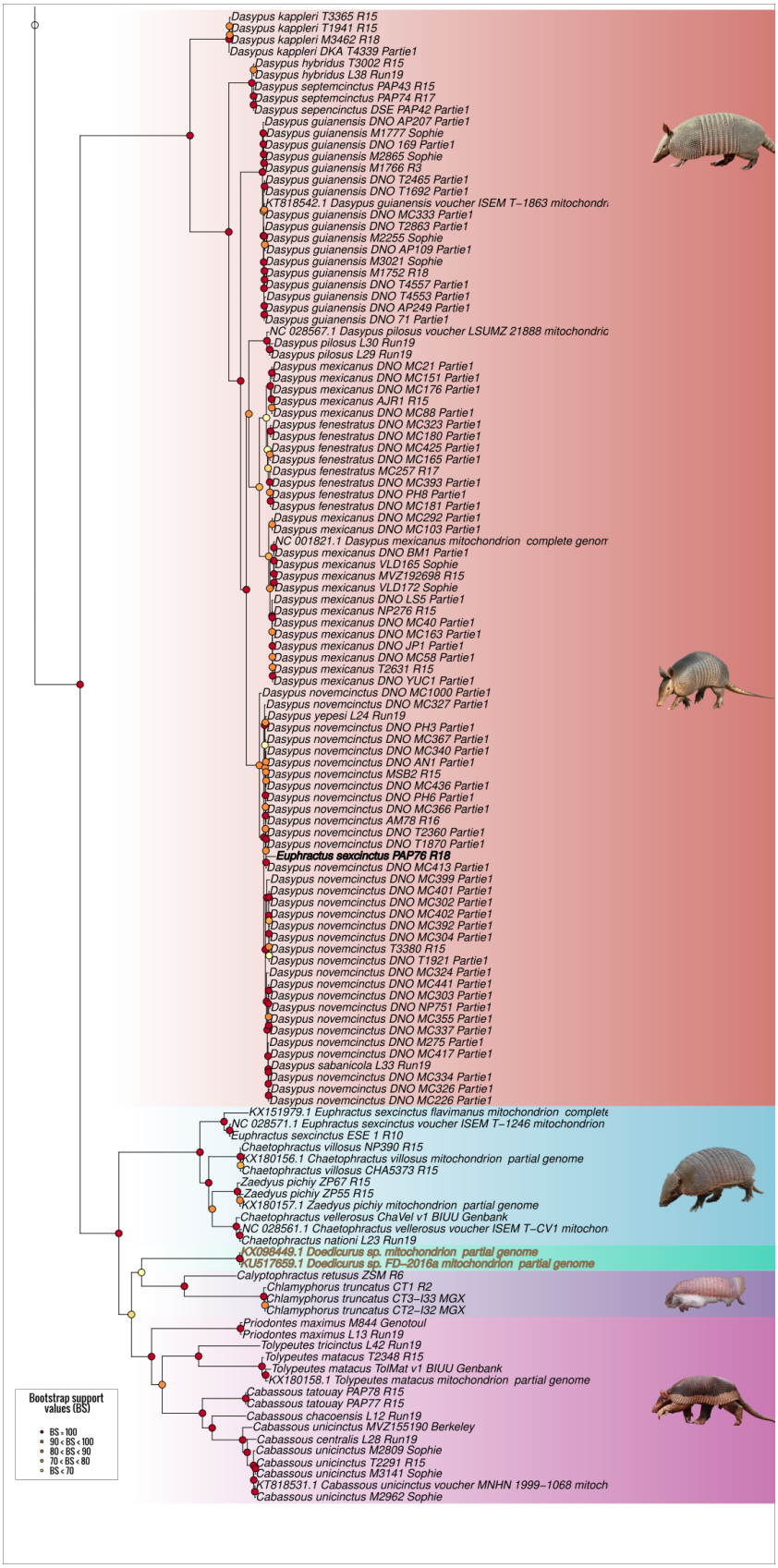


**Figure S1:** Continued.

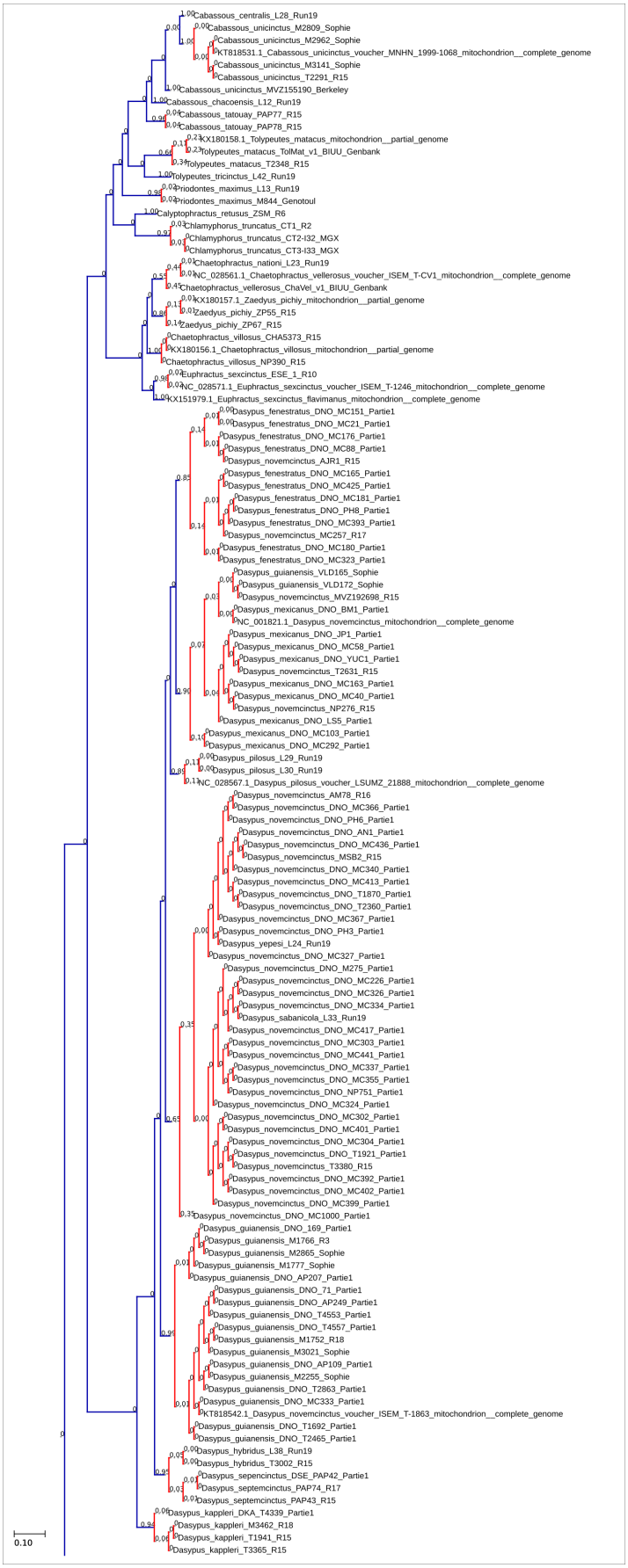


**Figure S2:** Species delimitation estimated by bPTP-h. The species tree reconstructed from the 15 concatenated mitochondrial genes using IQ-TREE and ModelFinder illustrates the most supported partition found by bPTP. The 52 red clades represent taxa identified as distinct species by bPTP.

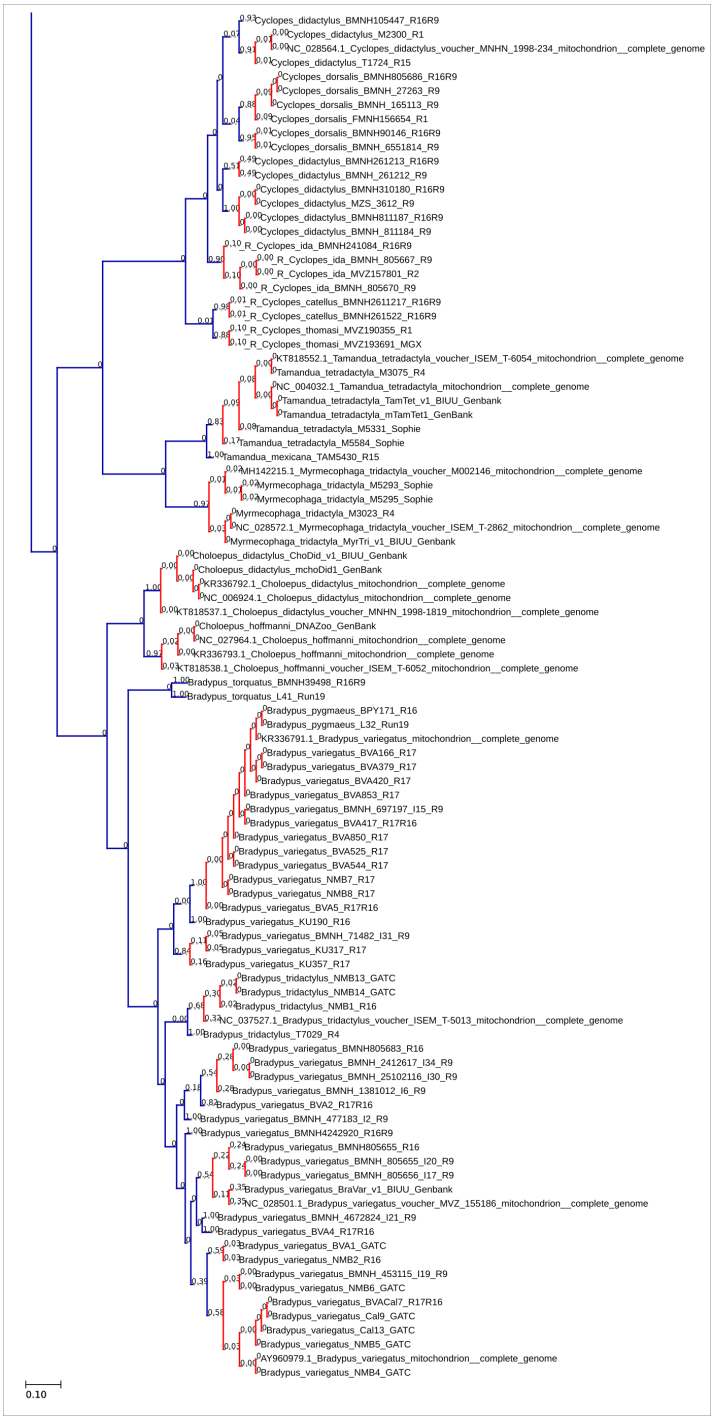


**Figure S2:** Continued.

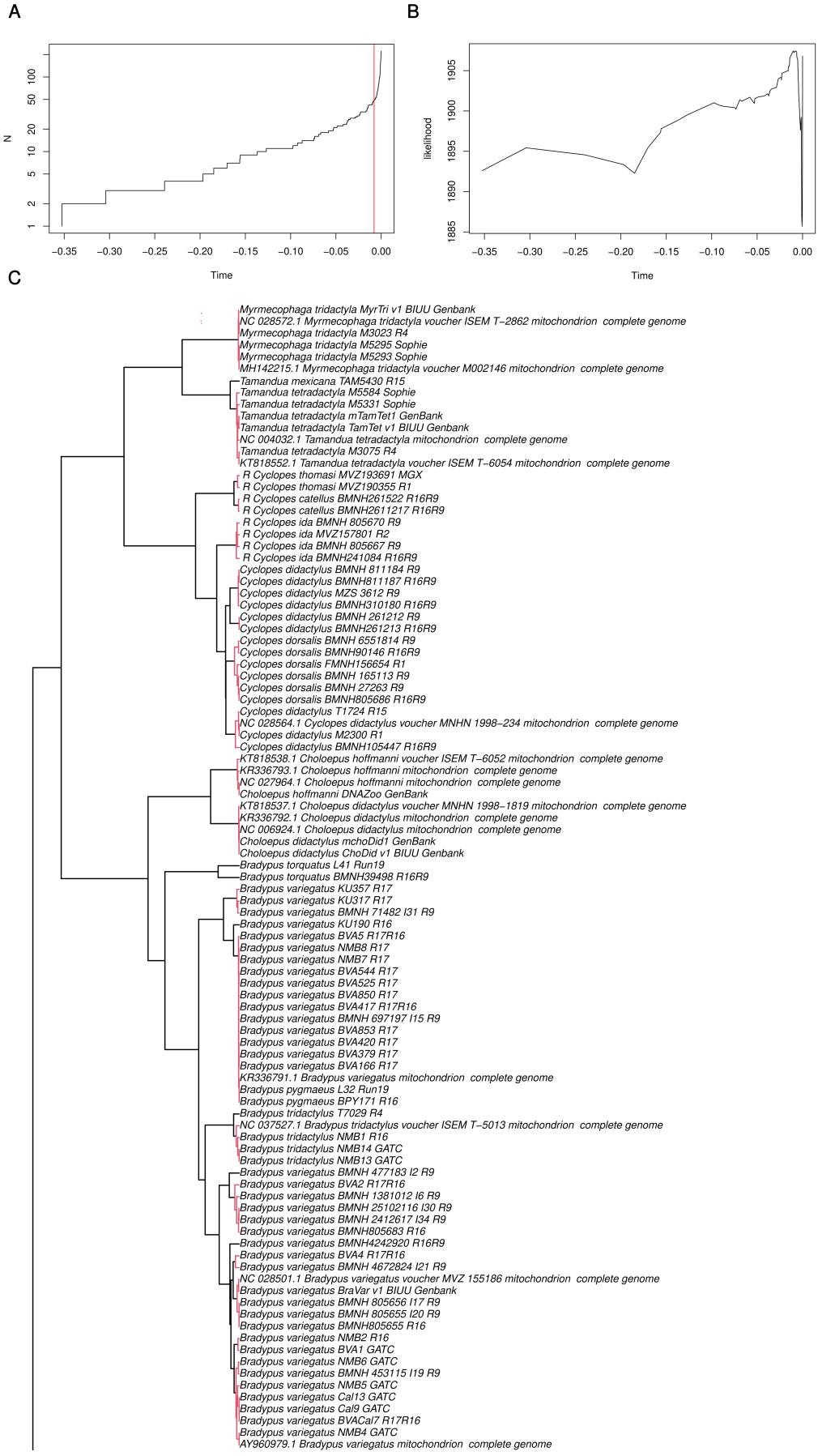


**Figure S3:** Species delimitation estimated using GMYC. A) Number of lineages (N) through time. The red line corresponds to the threshold estimated by GMYC. B) Likelihood profile through time. C) Ultrametric reconstruction of the concatenated 15 mitochondrial genes for 222 individuals using IQ-TREE. The 48 clades in red represent taxa supported as distinct species by GMYC.

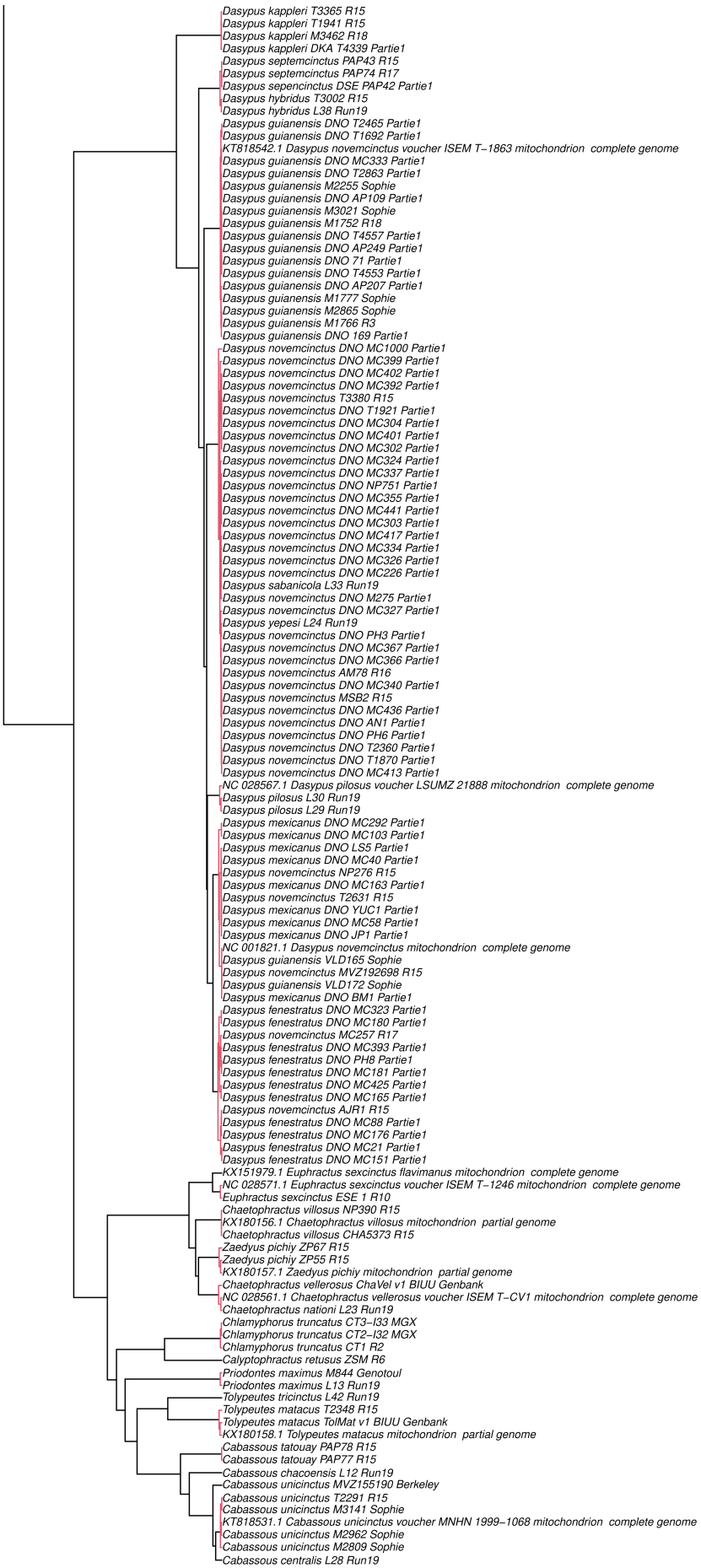


**Figure S3:** Continued.

**Table S2:** List of whole genome sequences used, detailing information on the depth of coverage, proportion of missing data, type of tissue sampled, and reference. Individuals highlighted in red have been excluded because of their depth of coverage $< 5X$, only *Calyptophractus retusus*, in bold in this table, has been kept despite its low depth of coverage.

| **Species** | **ID** | **Depth of coverage** | **Proportion of missing data** | **Tissue type** | **Reference** |
|---|---|---|---|---|---|
| *Choloepus hoffmanni* | DNAZoo | 13.6 | 0.17 | Fresh | GenBank |
| *Choloepus didactylus* | mchoDid1 | 12.95 | 0.12 | Fresh | GenBank |
| *Choloepus didactylus* | ChoDid,v1 | 10.48 | 0.17 | Fresh | Genbank |
| *Bradypus crinitus* | L41 | 12.13 | 0.2 | Fresh | This study |
| *Bradypus torquatus* | BMNH39498 | 8.4 | 0.31 | Museum | This study |
| *Bradypus variegatus* | BraVar.v1 | 9.77 | 0.44 | Fresh | Genbank |
| *Bradypus tridactylus* | T7029 | 17.3 | 0.05 | Fresh | (Allio 2021) |
| *Bradypus variegatus* | KU317 | 0.17 | 1 | Museum | This study |
| *Bradypus variegatus* | NMB3 | 3.57 | 0.72 | Fresh | This study |
| *Bradypus tridactylus* | NMB1 | 7.07 | 0.5 | Fresh | This study |
| *Bradypus variegatus* | BVA2 | 3.25 | 0.8 | Fresh | This study |
| *Bradypus variegatus* | BMNH805683 | 5.66 | 0.53 | Museum | This study |
| *Bradypus variegatus* | BVACal7 | 9.59 | 0.26 | Museum | This study |
| *Bradypus variegatus* | BVA4 | 10.72 | 0.21 | Fresh | This study |
| *Bradypus variegatus* | BMNH4242920 | 6.65 | 0.43 | Museum | This study |
| *Bradypus variegatus* | NMB2 | 3.5 | 0.73 | Fresh | This study |
| *Bradypus variegatus* | BMNH805655 | 5.98 | 0.49 | Museum | This study |
| *Bradypus variegatus* | KU357 | 3.64 | 0.74 | Museum | This study |
| *Bradypus pygmaeus* | BPY171 | 6.63 | 0.54 | Fresh | This study |
| *Bradypus variegatus* | KU190 | 5.57 | 0.6 | Museum | This study |
| *Bradypus pygmaeus* | L32 | 18.24 | 0.11 | Fresh | This study |
| *Bradypus variegatus* | NMB7 | 9.74 | 0.2 | Fresh | This study |
| *Bradypus variegatus* | BVA5 | 10.6 | 0.2 | Fresh | This study |
| *Bradypus variegatus* | BVA853 | 9.74 | 0.24 | Fresh | This study |
| *Bradypus variegatus* | BVA420 | 7.8 | 0.32 | Fresh | This study |
| *Bradypus variegatus* | BVA379 | 8.68 | 0.28 | Fresh | This study |
| *Bradypus variegatus* | NMB8 | 8.5 | 0.31 | Fresh | This study |
| *Bradypus variegatus* | BVA544 | 9.92 | 0.25 | Fresh | This study |
| *Bradypus variegatus* | BVA417 | 10.16 | 0.23 | Fresh | This study |
| *Bradypus variegatus* | BVA525 | 2.47 | 0.89 | Fresh | This study |

| | | | | | |
|---|---|---|---|---|---|
| *Bradypus variegatus* | BVA850 | 1.71 | 0.95 | Fresh | This study |
| *Bradypus variegatus* | BVA166 | 3.93 | 0.74 | Fresh | This study |
| *Myrmecophaga tridactyla* | MyrTri.v1 | 19.51 | 0.03 | Fresh | Genbank |
| *Myrmecophaga tridactyla* | M3023 | 17.07 | 0.03 | Fresh | (Allio 2021) |
| *Tamandua mexicana* | TAM5430 | 14.53 | 0.09 | Fresh | (Allio 2021) |
| *Tamandua tetradactyla* | M3075 | 21.12 | 0.05 | Fresh | (Allio 2021) |
| *Tamandua tetradactyla* | mTamTet1 | 30.55 | 0.06 | Fresh | GenBank |
| *Tamandua tetradactyla* | TamTet.V1 | 21.01 | 0.06 | Fresh | Genbank |
| *Cyclopes didactylus* | BMNH310180 | 2.98 | 0.83 | Museum | This study |
| *Cyclopes didactylus* | BMNH261213 | 3.34 | 0.8 | Museum | This study |
| *Cyclopes thomasi* | MVZ190355 | 19.62 | 0.14 | Fresh | (Allio 2021) |
| *Cyclopes catellus* | BMNH261522 | 6.78 | 0.49 | Museum | This study |
| *Cyclopes catellus* | BMNH2611217 | 4.47 | 0.7 | Museum | This study |
| *Cyclopes ida* | MVZ157801 | 22.66 | 0.09 | Fresh | (Allio 2021) |
| *Cyclopes ida* | BMNH241084 | 4.98 | 0.6 | Museum | This study |
| *Cyclopes dorsalis* | FMNH156654 | 19.97 | 0.09 | Fresh | (Allio 2021) |
| *Cyclopes dorsalis* | BMNH90146 | 3.08 | 0.79 | Museum | This study |
| *Cyclopes dorsalis* | BMNH805686 | 4.33 | 0.66 | Museum | This study |
| *Cyclopes didactylus* | T1724 | 10.54 | 0.14 | Fresh | This study |
| *Cyclopes didactylus* | M2300 | 21.21 | 0.05 | Fresh | (Allio 2021) |
| *Cyclopes didactylus* | BMNH811187 | 6.17 | 0.47 | Museum | This study |
| *Cyclopes didactylus* | BMNH105447 | 5.25 | 0.58 | Museum | This study |
| *Dasypus kappleri* | T3365 | 10.85 | 0.2 | Fresh | This study |
| *Dasypus kappleri* | T1941 | 12.42 | 0.19 | Fresh | This study |
| *Dasypus kappleri* | M3462 | 29.05 | 0.22 | Fresh | This study |
| *Dasypus septemcinctus* | PAP74 | 13.84 | 0.15 | Fresh | This study |
| *Dasypus septemcinctus* | PAP43 | 13.58 | 0.15 | Fresh | This study |
| *Dasypus hybridus* | T3002 | 15.05 | 0.19 | Fresh | This study |
| *Dasypus hybridus* | L38 | 4.74 | 0.59 | Museum | This study |
| *Dasypus pilosus* | L29 | 7.99 | 0.34 | Museum | This study |
| *Dasypus pilosus* | L30 | 0.41 | 0.99 | Museum | This study |
| *Dasypus fenestratus* | MC257 | 6.49 | 0.45 | Fresh | This study |
| *Dasypus mexicanus* | T2631 | 10.82 | 0.13 | Fresh | This study |
| *Dasypus mexicanus* | NP276 | 10.95 | 0.13 | Fresh | This study |
| *Dasypus mexicanus* | MVZ192698 | 11.31 | 0.11 | Fresh | This study |
| *Dasypus mexicanus* | AJR1 | 10.18 | 0.15 | Fresh | This study |

| | | | | | |
|---|---|---|---|---|---|
| *Dasypus novemcinctus* | AM78 | 9.87 | 0.23 | Fresh | This study |
| *Dasypus guianensis* | M1766 | 69.5 | 0.11 | Fresh | This study |
| *Dasypus guianensis* | M1752 | 23.1 | 0.23 | Fresh | This study |
| *Dasypus sabanicola* | L33 | 8.96 | 0.28 | Museum | This study |
| *Dasypus novemcinctus* | T3380 | 10.43 | 0.17 | Fresh | This study |
| *Dasypus mazzai* | L24 | 2.71 | 0.84 | Museum | This study |
| *Dasypus novemcinctus* | MSB2 | 10.67 | 0.17 | Fresh | This study |
| *Euphractus sexcinctus* | PAP76 | 5.32 | 0.59 | Fresh | This study |
| *Euphractus sexcinctus* | ESE1 | 44.63 | 0.04 | Fresh | (Allio 2021) |
| *Chaetophractus villosus* | NP390 | 9.36 | 0.22 | Fresh | This study |
| *Chaetophractus villosus* | CHA5373 | 9.78 | 0.2 | Fresh | This study |
| *Zaedyus pichiy* | ZP67 | 8.98 | 0.4 | Fresh | This study |
| *Zaedyus pichiy* | ZP55 | 9.79 | 0.31 | Fresh | This study |
| *Chaetophractus vellerosus* | ChaVel.v1 | 12.17 | 0.22 | Fresh | Genbank |
| *Chaetophractus nationi* | L23 | 10.53 | 0.31 | Fresh | This study |
| *Chlamyphorus truncatus* | CT1 | 19.98 | 0.03 | Fresh | (Allio 2021) |
| ***Calyptophractus retusus*** | ZSM | 4.41 | 0.67 | Museum | (Allio 2021) |
| *Tolypeutes tricinctus* | L42 | 9.01 | 0.45 | Fresh | This study |
| *Tolypeutes matacus* | TolMat.v1 | 12.93 | 0.1 | Fresh | Genbank |
| *Tolypeutes matacus* | T2348 | 9.79 | 0.16 | Fresh | This study |
| *Priodontes maximus* | M844 | 13.58 | 0.13 | Fresh | (Allio 2021) |
| *Priodontes maximus* | L13 | 1.95 | 0.93 | Fresh | This study |
| *Cabassous centralis* | L28 | 1.91 | 0.91 | Museum | This study |
| *Cabassous unicinctus* | T2291 | 7.29 | 0.34 | Fresh | This study |
| *Cabassous unicinctus* | MVZ155190 | 29.44 | 0 | Fresh | (Allio 2021) |
| *Cabassous chacoensis* | L12 | 5.37 | 0.62 | Fresh | This study |
| *Cabassous tatouay* | PAP78 | 7.59 | 0.36 | Fresh | This study |
| *Cabassous tatouay* | PAP77 | 7.69 | 0.36 | Fresh | This study |

**Table S3:** List and details of the reference genome used to mapped resequenced individuals. VGP: Vertebrate Genomes Project, BI: Broad Institute.

| **Species** | **ID** | **Accession** | **Reference** |
|---|---|---|---|
| *Dugong dugon* | mDugDug1.hap1 | GCA_030035585.1 | VGP |
| *Orycteropus afer* | OryAfe1.0 | GCF_000298275.1 | BI |
| *Loxodonta africana* | mLoxAfr1.hap1 | GCA_030020305.1 | VGP |
| *Choloepus didactylus* | mchoDid1.pri | GCF_015220235.1 | VGP |
| *Bradypus tridactylus* | ISEM_T7029_MaSuRCA | | Allio 2021 |
| *Cyclopes didactylus* | ISEM_M2300_MaSuRCA | | Allio 2021 |
| *Myrmecophaga tridactyla* | ISEM_M3023_MaSuRCA_HiC | | Allio 2021 |
| *Tamandua tetradactyla* | ISEM_M3075_MaSuRCA_HiC | | Allio 2021 |
| *Dasypus novemcinctus* | mDasNov1.hap2 | GCF_030445035.1 | VGP |
| *Euphractus sexcinctus* | ISEM_ESE1_MaSuRCA | | Allio 2021 |
| *Chlamyphorus truncatus* | ISEM_CT1_MaSuRCA | | Allio 2021 |
| *Priodontes maximus* | ISEM_M844_MaSuRCA | | Allio 2021 |
| *Tolypeutes matacus* | TolMat_v1_BIUU | GCA_004025125.1 | BI |
| *Cabassous unicinctus* | ISEM_MVZ155190_Discovar | | Allio 2021 |

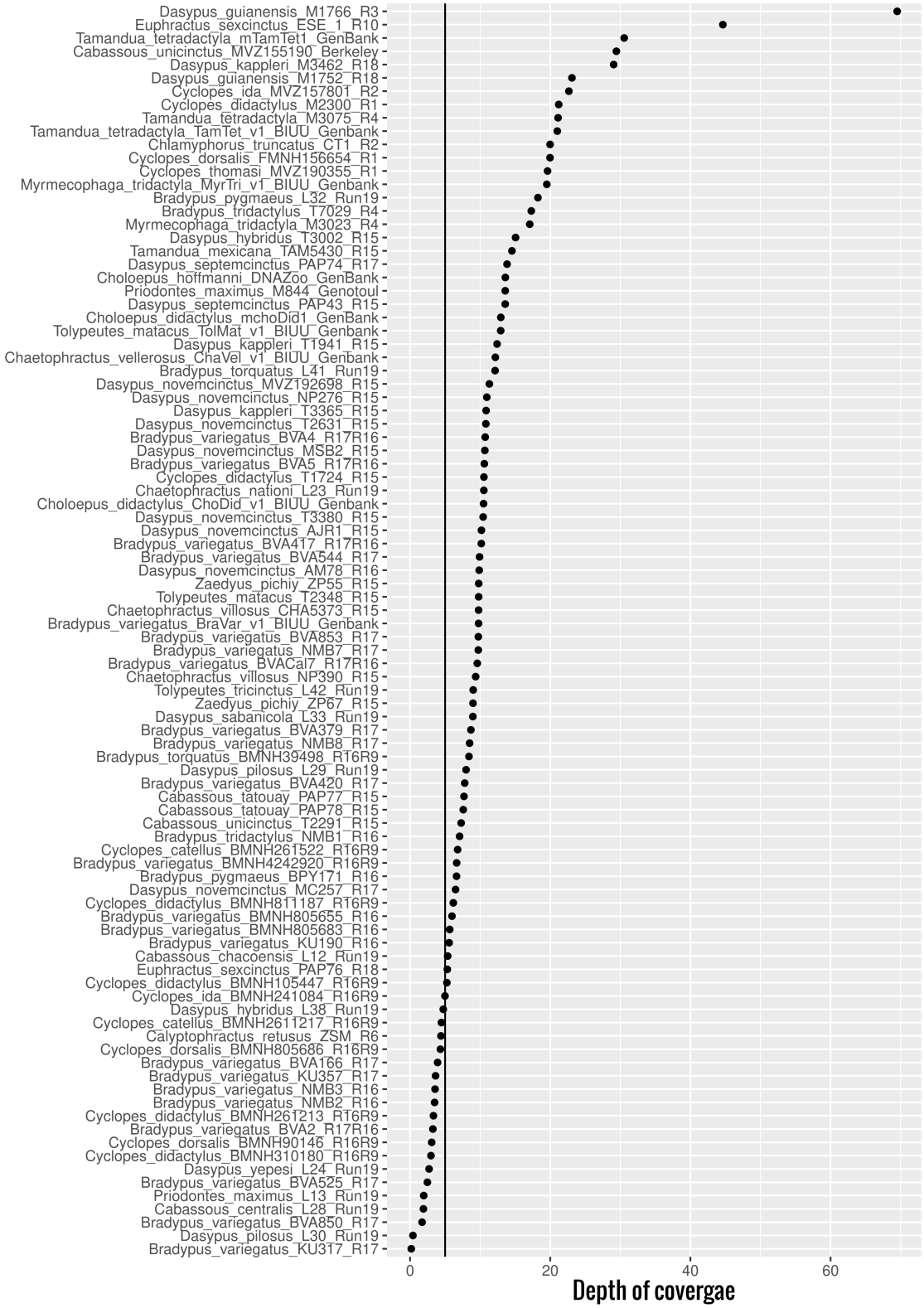


**Figure S4:** Genome-wide depth of coverage estimated from the *Random_dataset* for the 94 individuals used in this study. The black line corresponds to a 5x depth of coverage. Except for the unique representative of *Calyptophractus retusus*, individuals with a depth coverage below this 5X threshold have been excluded.

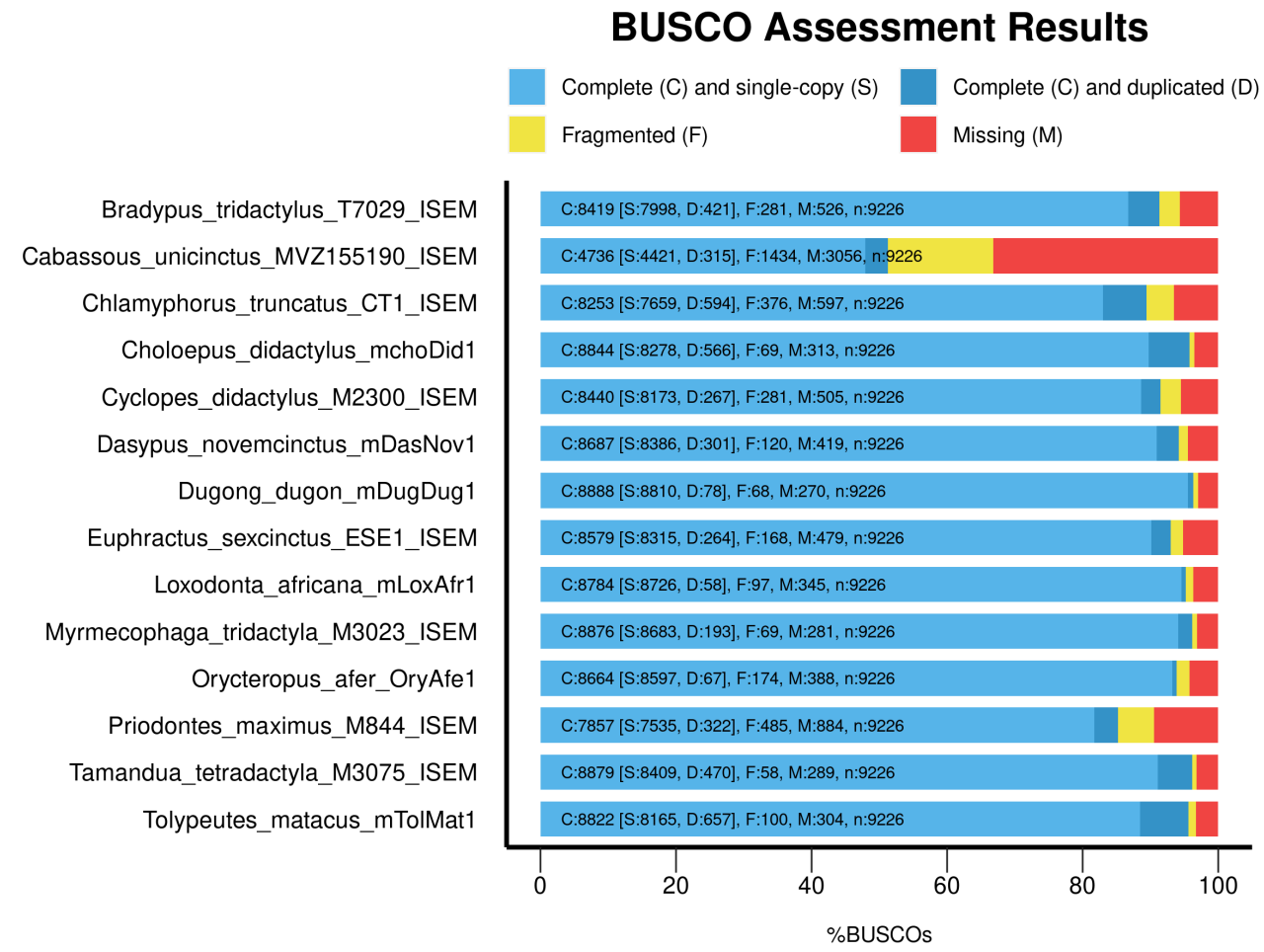


**Figure S5:** Completeness of orthologous genes of the Mammalia OrthoDB10 BUSCO gene set extracted for 14 reference genome assemblies. Bars represent the percentage of complete single-copy (light blue), complete duplicated (dark blue), fragmented (yellow), and missing (red) genes. The number of genes are detailed for all these categories according to these abbreviations : C : complete genes, S: complete single-copy, D: complete duplicated, F: fragmented, M: missing (red) and n: total.

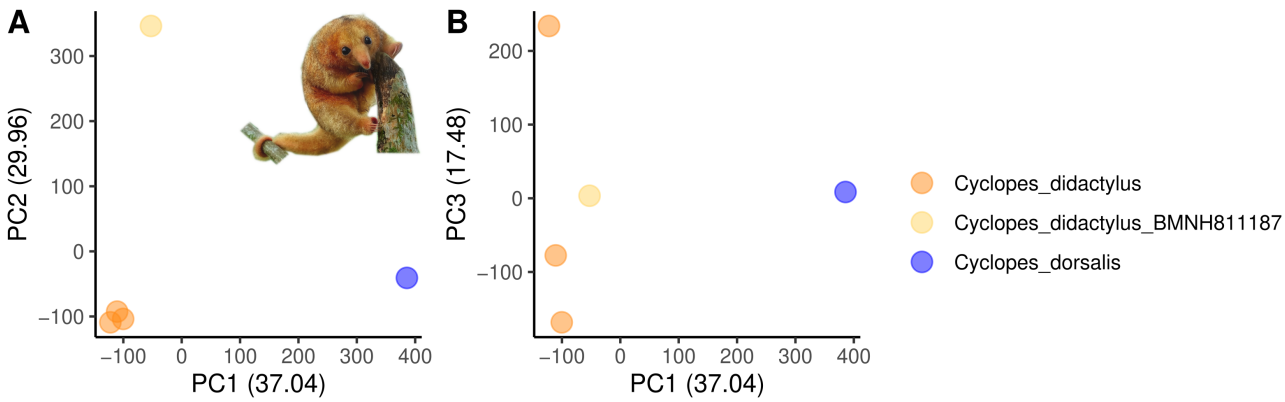


**Figure S6:** Principal Component Analysis of genetic variance (PCA) conducted on 102,079 SNPs shared by 5 *Cyclopes* spp. individuals from 400 nuclear loci. A) PC1 and PC2 B) PC1 and PC3.

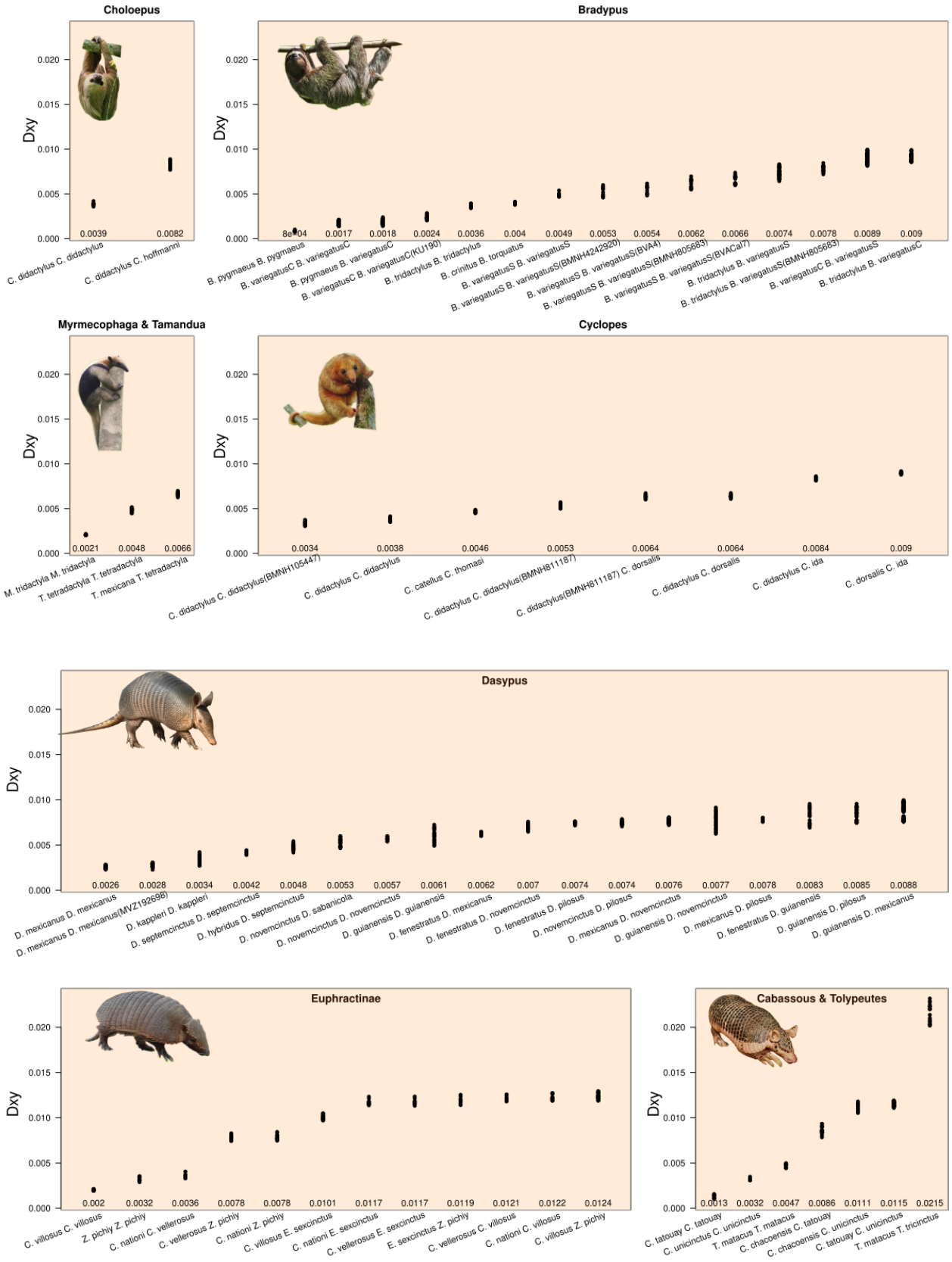


**Figure S7:** Pairwise genetic divergence (Dxy). Each point represents the mean Dxy estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. The mean Dxy value is indicated below the graph.

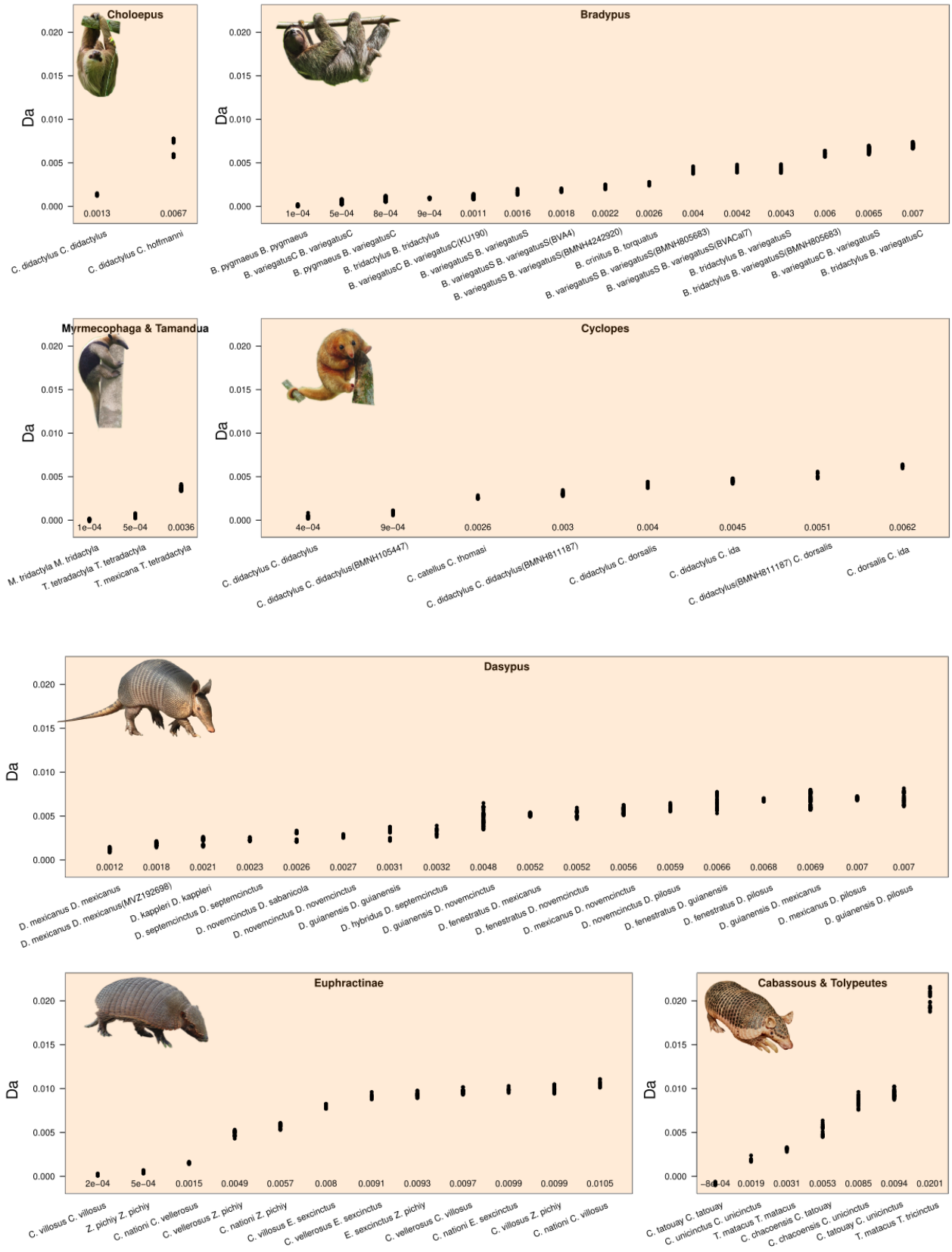


**Figure S8:** Pairwise genetic divergence (Da). Each point represents the mean Da estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. The mean Dxy value is indicated below the graph.

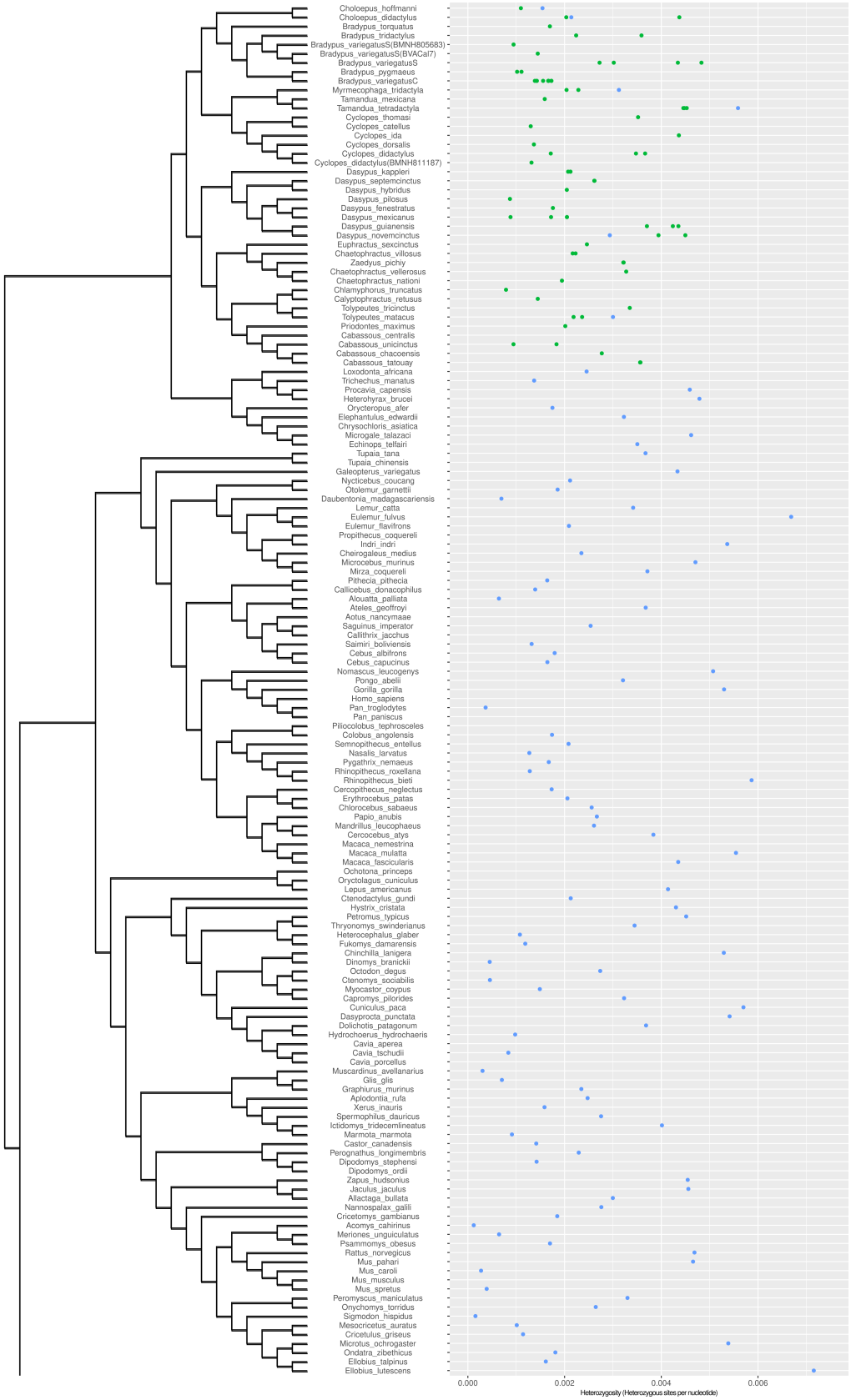


**Figure S9:** Genome-wide heterozygosity (He) estimated for 274 mammalian species. Blue points have been taken from the Zoonomia project (Wilder et al. 2023), green points have been estimated in this study.

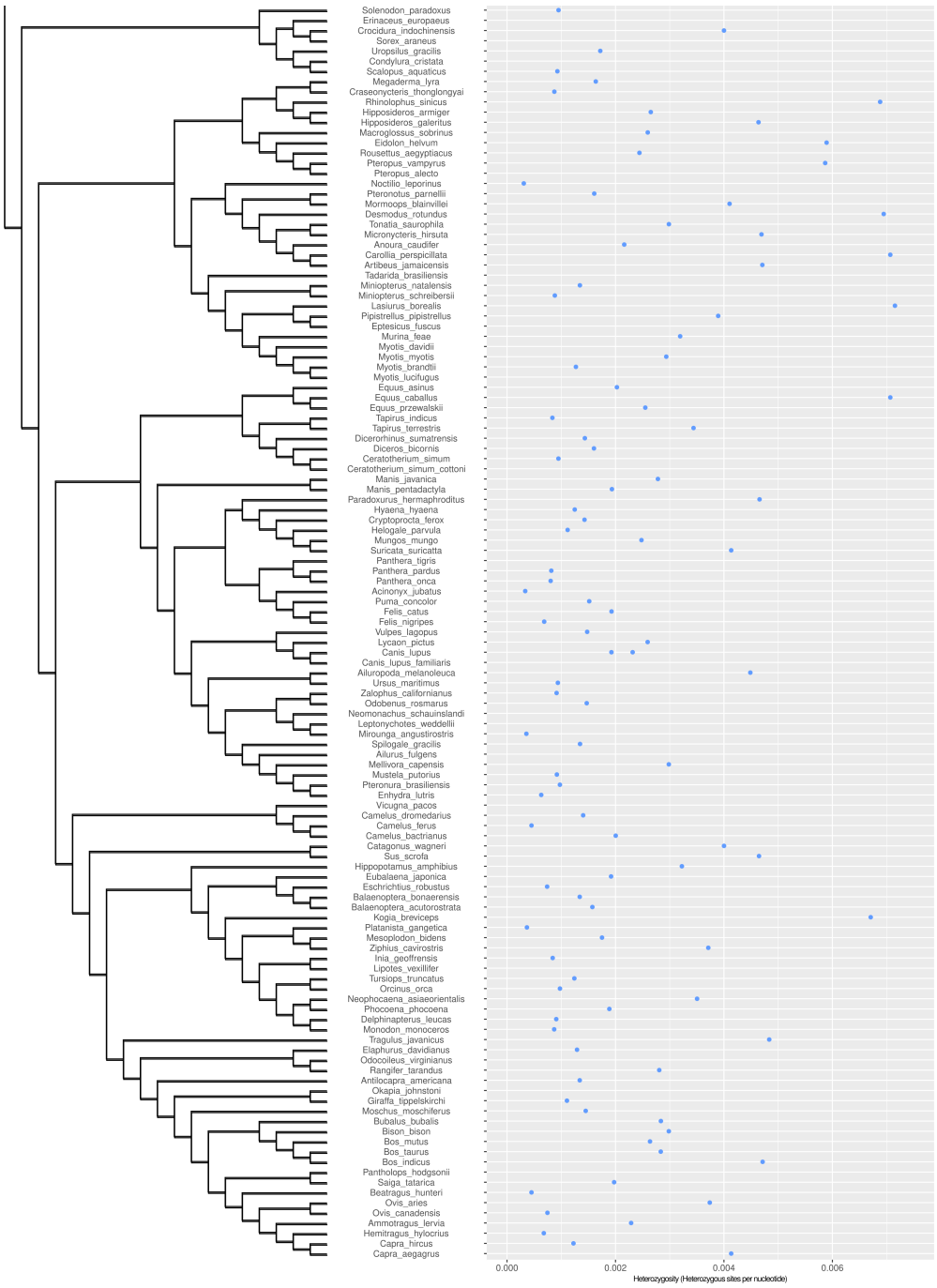


**Figure S9:** Continued.

**Table S4:** Generation time of xenarthran species used in PSMC analysis and estimated by Pacifici et al. (2013). For the 8 species in bold generation time were not available, so the value of the closest species have been used.

| Species | Generation time (years) |
|---|---|
| *Choloepus didactylus* | 12.8 |
| *Choloepus hoffmanni* | 14 |
| *Bradypus pygmaeus* | 10.6 |
| *Bradypus torquatus* | 10.6 |
| ***Bradypus crinitus*** | 10.6 |
| *Bradypus tridactylus* | 10.6 |
| *Bradypus variegatus* | 10.6 |
| *Cyclopes didactylus* | 10.8 |
| ***Cyclopes ida*** | 10.8 |
| ***Cyclopes dorsalis*** | 10.8 |
| ***Cyclopes thomasi*** | 10.8 |
| ***Cyclopes catellus*** | 10.8 |
| *Myrmecophaga tridactyla* | 11.3 |
| *Tamandua mexicana* | 5.1 |
| *Tamandua tetradactyla* | 5.1 |
| *Dasypus kappleri* | 4.5 |
| *Dasypus septemcinctus septemcinctus* | 4.5 |
| *Dasypus septemcinctus hybridus* | 4 |
| *Dasypus novemcinctus* | 5 |
| ***Dasypus guianensis*** | 5 |
| ***Dasypus mexicanus*** | 5 |
| ***Dasypus fenestratus*** | 5 |
| *Dasypus pilosus* | 4.5 |
| *Dasypus sabanicola* | 4.5 |
| *Dasypus mazzai* | 4.5 |
| *Euphractus sexcinctus* | 5 |
| *Chaetophractus vellerosus nationi* | 4 |
| *Chaetophractus vellerosus vellerosus* | 4 |
| *Chaetophractus villosus* | 4 |
| *Zaedyus pichiy* | 4 |
| *Calyptophractus retusus* | 4 |
| *Chlamyphorus truncatus* | 5 |
| *Priodontes maximus* | 5.4 |
| *Tolypeutes matacus* | 11.7 |
| *Tolypeutes tricinctus* | 3 |
| *Cabassous chacoensis* | 5.1 |
| *Cabassous tatouay* | 5.1 |
| *Cabassous unicinctus* | 5.1 |
| *Cabassous centralis* | 5.1 |

**Table S5:** Completeness of Busco sequences for the 94 xenarthran individuals.

| Genus | Species | ID | Rate of missing data |
|---|---|---|---|
| *Bradypus* | *ephippiger* | BMNH805683 | 0.43 |
| *Bradypus* | *ephippiger* | BVA2 | 0.80 |
| *Bradypus* | Exclude | BMNH4242920 | 0.30 |
| *Bradypus* | Exclude | BraVar | 0.12 |
| *Bradypus* | Exclude | BVACal7 | 0.27 |
| *Bradypus* | Exclude | NMB3 | 0.87 |
| *Bradypus* | *pygmaeus* | BPY171 | 0.80 |
| *Bradypus* | *pygmaeus* | L32 | 0.01 |
| *Bradypus* | *torquatus*A | BMNH39498 | 0.26 |
| *Bradypus* | *torquatus*B | L41 | 0.06 |
| *Bradypus* | *tridactylus* | NMB1 | 0.77 |
| *Bradypus* | *tridactylus* | T7029 | 0.01 |
| *Bradypus* | *variegatus*C | BVA166 | 0.76 |
| *Bradypus* | *variegatus*C | BVA379 | 0.30 |
| *Bradypus* | *variegatus*C | BVA417 | 0.33 |
| *Bradypus* | *variegatus*C | BVA420 | 0.38 |
| *Bradypus* | *variegatus*C | BVA5 | 0.18 |
| *Bradypus* | *variegatus*C | BVA525 | 0.86 |
| *Bradypus* | *variegatus*C | BVA544 | 0.35 |
| *Bradypus* | *variegatus*C | BVA853 | 0.28 |
| *Bradypus* | *variegatus*C | KU190 | 0.82 |
| *Bradypus* | *variegatus*C | KU317 | 0.99 |
| *Bradypus* | *variegatus*C | KU357 | 0.74 |
| *Bradypus* | *variegatus*C | NMB7 | 0.11 |
| *Bradypus* | *variegatus*C | NMB8 | 0.50 |
| *Bradypus* | *variegatus*S | BMNH805655 | 0.48 |
| *Bradypus* | *variegatus*S | BVA4 | 0.30 |
| *Bradypus* | *variegatus*S | NMB2 | 0.90 |
| *Cabassous* | *chacoensis* | L12 | 0.39 |
| *Cabassous* | Exclude | L28 | 0.98 |
| *Cabassous* | *tatouay* | PAP77 | 0.23 |
| *Cabassous* | *tatouay* | PAP78 | 0.14 |
| *Cabassous* | *unicinctus* | MVZ155190 | 0.00 |
| *Cabassous* | *unicinctus* | T2291 | 0.24 |
| *Calyptophractus* | *retusus* | ZSM | 0.42 |
| *Chaetophractus* | *nationi* | L23 | 0.23 |
| *Chaetophractus* | *vellerosus* | ChaVel | 0.19 |
| *Chaetophractus* | *villosus* | CHA5373 | 0.05 |
| *Chaetophractus* | *villosus* | NP390 | 0.09 |
| *Chlamyphorus* | *truncatus* | CT1 | 0.01 |
| *Choloepus* | *didactylus* | ChoDid | 0.10 |
| *Choloepus* | *didactylus* | mchoDid1 | 0.06 |
| *Choloepus* | *hoffmanni* | DNAZoo | 0.24 |
| *Cyclopes* | *catellus* | BMNH2611217 | 0.74 |
| *Cyclopes* | *catellus* | BMNH261522 | 0.45 |
| *Cyclopes* | *didactylus*G | BMNH105447 | 0.67 |
| *Cyclopes* | *didactylus*G | M2300 | 0.02 |
| *Cyclopes* | *didactylus*G | T1724 | 0.08 |
| *Cyclopes* | *didactylus*S | BMNH811187 | 0.41 |
| *Cyclopes* | *dorsalis* | BMNH805686 | 0.68 |
| *Cyclopes* | *dorsalis* | BMNH90146 | 0.88 |

| | | | |
|---|---|---|---|
| *Cyclopes* | *dorsalis* | FMNH156654 | 0.01 |
| *Cyclopes* | Exclude | BMNH261213 | 0.98 |
| *Cyclopes* | Exclude | BMNH310180 | 0.88 |
| *Cyclopes* | *ida* | BMNH241084 | 0.66 |
| *Cyclopes* | *ida* | MVZ157801 | 0.01 |
| *Cyclopes* | *thomasi* | MVZ190355 | 0.02 |
| *Dasypus* | Exclude | AJR1 | 0.04 |
| *Dasypus* | Exclude | AM78 | 0.09 |
| *Dasypus* | Exclude | L24 | 0.83 |
| *Dasypus* | *fenestratus* | MC257 | 0.39 |
| *Dasypus* | *guianensis* | M1752 | 0.11 |
| *Dasypus* | *guianensis* | M1766 | 0.01 |
| *Dasypus* | *hybridus* | L38 | 0.49 |
| *Dasypus* | *hybridus* | T3002 | 0.03 |
| *Dasypus* | *kappleri* | M3462 | 0.05 |
| *Dasypus* | *kappleri* | T1941 | 0.02 |
| *Dasypus* | *kappleri* | T3365 | 0.03 |
| *Dasypus* | *mexicanus* | MVZ192698 | 0.03 |
| *Dasypus* | *mexicanus* | NP276 | 0.02 |
| *Dasypus* | *mexicanus* | T2631 | 0.03 |
| *Dasypus* | *novemcinctus* | MSB2 | 0.03 |
| *Dasypus* | *novemcinctus* | T3380 | 0.05 |
| *Dasypus* | *pilosus* | L29 | 0.45 |
| *Dasypus* | *sabanicola* | L33 | 0.18 |
| *Dasypus* | *septemcinctus* | PAP43 | 0.01 |
| *Dasypus* | *septemcinctus* | PAP74 | 0.06 |
| *Euphractus* | *sexcinctus* | ESE | 0.01 |
| *Euphractus* | *sexcinctus* | PAP76 | 0.09 |
| *Myrmecophaga* | *tridactyla* | M3023 | 0.01 |
| *Myrmecophaga* | *tridactyla* | MyrTri | 0.02 |
| *Priodontes* | *maximus* | L13 | 0.97 |
| *Priodontes* | *maximus* | M844 | 0.19 |
| *Tamandua* | *mexicana* | TAM5430 | 0.01 |
| *Tamandua* | *tetradactyla* | M3075 | 0.01 |
| *Tamandua* | *tetradactyla* | mTamTet1 | 0.01 |
| *Tamandua* | *tetradactyla* | TamTet | 0.01 |
| *Tolypeutes* | *matacus* | T2348 | 0.08 |
| *Tolypeutes* | *matacus* | TolMat | 0.04 |
| *Tolypeutes* | *tricinctus* | L42 | 0.32 |
| *Zaedyus* | *pichiy* | ZP55 | 0.34 |
| *Zaedyus* | *pichiy* | ZP67 | 0.46 |

**Table S6:** Quality of the 40 best xenarthran representatives for the 2006 Busco genes. For each individual, its mean sequence completeness and its frequency of Busco genes represented are detailed.

| ID | Genus | Species | Rate of missing data | Rate of Busco genes |
|---|---|---|---|---|
| BMNH261522 | *Cyclopes* | *catellus* | 0.37 | 83.80 |
| BMNH39498 | *Bradypus* | *torquatus* | 0.17 | 99.50 |
| BMNH805683 | *Bradypus* | *ephippiger* | 0.35 | 88.93 |
| BMNH811187 | *Cyclopes* | *didactylus* | 0.33 | 92.47 |
| BVA4 | *Bradypus* | *variegatus*S | 0.20 | 98.31 |
| CHA5373 | *Chaetophractus* | *villosus* | 0.04 | 100.00 |
| ChaVel | *Chaetophractus* | *vellerosus* | 0.16 | 97.86 |
| CT1 | *Chlamyphorus* | *truncatus* | 0.00 | 97.51 |
| DNAZoo | *Choloepus* | *hoffmanni* | 0.13 | 97.01 |
| ESE | *Euphractus* | *sexcinctus* | 0.00 | 100.00 |
| FMNH156654 | *Cyclopes* | *dorsalis* | 0.00 | 100.00 |
| L12 | *Cabassous* | *chacoensis* | 0.32 | 67.15 |
| L23 | *Chaetophractus* | *nationi* | 0.16 | 97.76 |
| L29 | *Dasypus* | *pilosus* | 0.27 | 64.31 |
| L32 | *Bradypus* | *pygmaeus* | 0.01 | 100.00 |
| L33 | *Dasypus* | *sabanicola* | 0.13 | 99.80 |
| L41 | *Bradypus* | *crinitus* | 0.04 | 100.00 |
| L42 | *Tolypeutes* | *tricinctus* | 0.29 | 94.52 |
| M1766 | *Dasypus* | *guianensis* | 0.00 | 100.00 |
| M2300 | *Cyclopes* | *didactylus* | 0.01 | 100.00 |
| M3023 | *Myrmecophaga* | *tridactyla* | 0.00 | 99.30 |
| M3075 | *Tamandua* | *tetradactyla* | 0.00 | 99.30 |
| M844 | *Priodontes* | *maximus* | 0.16 | 96.76 |
| MC257 | *Dasypus* | *fenestratus* | 0.29 | 87.54 |
| mchoDid1 | *Choloepus* | *didactylus* | 0.04 | 99.05 |
| MSB2 | *Dasypus* | *novemcinctus* | 0.02 | 100.00 |
| MVZ155190 | *Cabassous* | *unicinctus* | 0.00 | 84.75 |
| MVZ157801 | *Cyclopes* | *ida* | 0.00 | 100.00 |
| MVZ190355 | *Cyclopes* | *thomasi* | 0.01 | 100.00 |
| NMB7 | *Bradypus* | *variegatus*C | 0.08 | 99.95 |
| NP276 | *Dasypus* | *mexicanus* | 0.02 | 100.00 |
| PAP43 | *Dasypus* | *septemcinctus* | 0.00 | 100.00 |

| | | | | |
|---|---|---|---|---|
| PAP78 | *Cabassous* | *tatouay* | 0.12 | 83.75 |
| T1941 | *Dasypus* | *kappleri* | 0.01 | 100.00 |
| T3002 | *Dasypus* | *hybridus* | 0.01 | 99.95 |
| T7029 | *Bradypus* | *tridactylus* | 0.00 | 100.00 |
| TAM5430 | *Tamandua* | *mexicana* | 0.01 | 99.30 |
| TolMat | *Tolypeutes* | *matacus* | 0.04 | 98.80 |
| ZP55 | *Zaedyus* | *pichiy* | 0.28 | 82.45 |
| ZSM | *Calyptophractus* | *retusus* | 0.34 | 84.15 |

**Table S7:** Extract of the quality for 9 of the 2006 Busco genes represented by at least 37 individuals. For each gene the mean sequences completeness and the number of individuals are detailed.

| **Busco genes** | **Rate of missing data** | **Number of individuals** |
|---|---|---|
| 1000at40674 | 0.11 | 37 |
| 100137at40674 | 0.09 | 37 |
| 10017at40674 | 0.11 | 38 |
| 100265at40674 | 0.13 | 40 |
| 100296at40674 | 0.09 | 37 |
| 10035at40674 | 0.08 | 37 |
| 100716at40674 | 0.10 | 40 |
| 100721at40674 | 0.12 | 40 |
| 100812at40674 | 0.09 | 39 |
| ... | ... | ... |

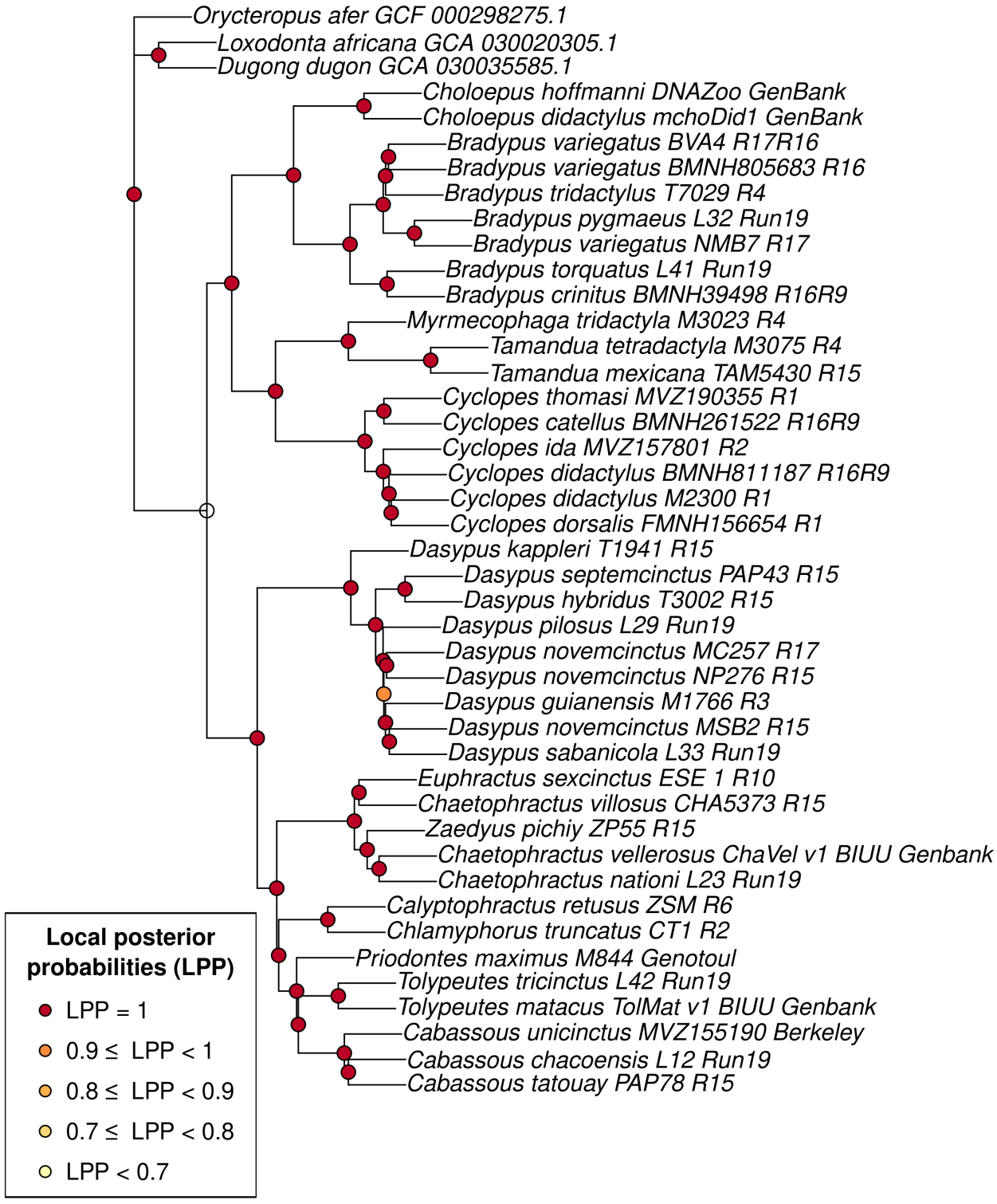


**Figure S10:** Phylogenetic relationships of the 40 best representatives of xenarthran species plus 3 outgroup (Dugong dugon, Orycteropus afer, Loxodonta africana) obtained using Astral on the 1908 Busco gene trees reconstructed with IQ-Tree and ModelFinder. Node circles are coloured according to the local posterior probabilities (LPP).

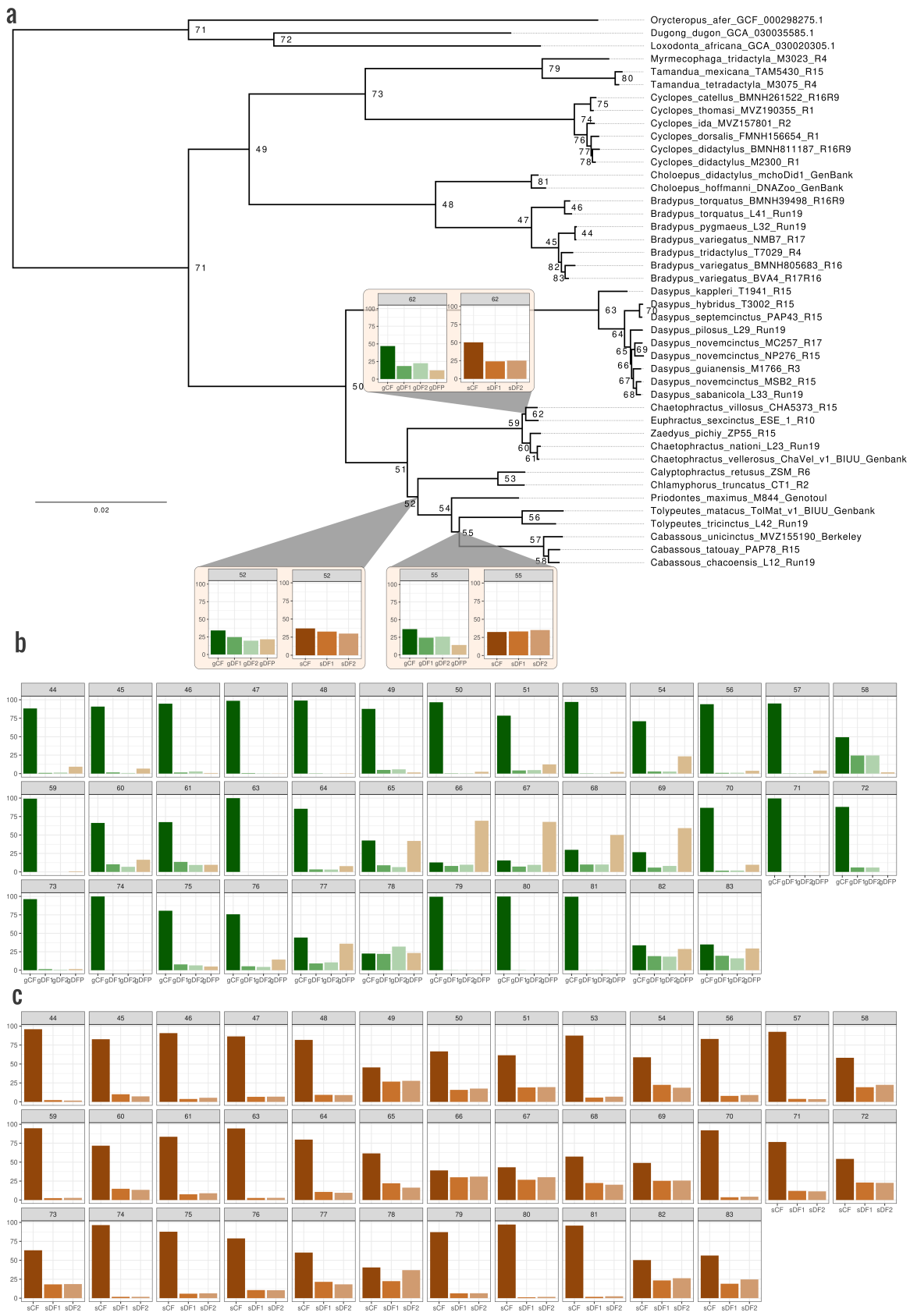


**Figure S11:** Topological conflicts estimated from 1908 Busco sequences and gene trees. For each node, numbered in the phylogenetic reconstruction (a), the concordance factor is estimated by comparing the proportion of genes supporting the species tree (gCF), the first or second alternative topology (gDF1, gDF2 respectively) or a paraphyletic topology (gDFP). Bar plots (b) illustrate relative frequency of gCF, gDF1, gDF2, gDFP. Similarly, the proportion of sites supporting the species tree (sCF), the first or the second alternative topology (sDF1, sDF2 respectively) are estimated and illustrated by the bar plots (c).

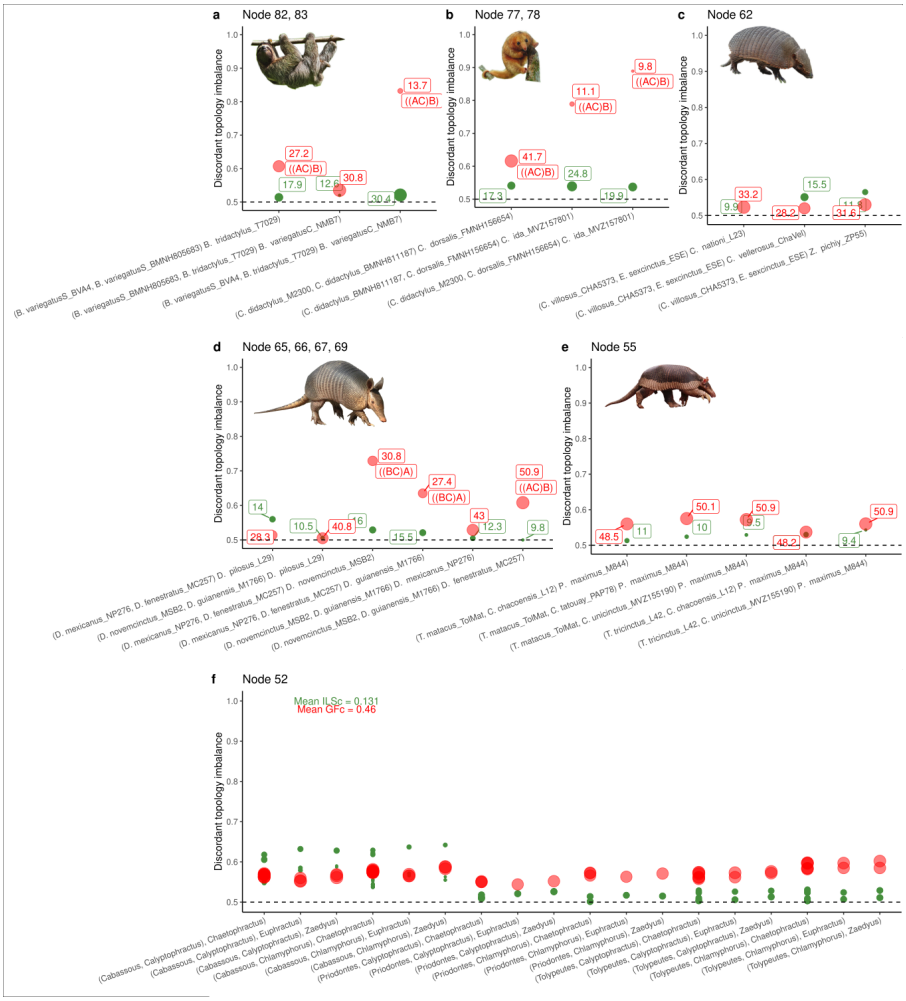


**Figure S12:** Disentangling incomplete lineage sorting (ILS) in red and gene flow (GF) in green using Aphid for nodes with topological conflicts: a) *Bradypus* spp. b) *Cyclopes* spp., c) Euphractinae, d) *Dasypus* spp., e) Tolypeutinae, f) Chlamyphoridae. Dot size represents the proportion of ILS and GF in phylogenetic conflicts (labels details these values) and the position on the Y-axis illustrates the imbalance proportion of one discordant topology (((A,C),B) or ((B,C),A)). Node numbers are illustrated in Figure S11 a.

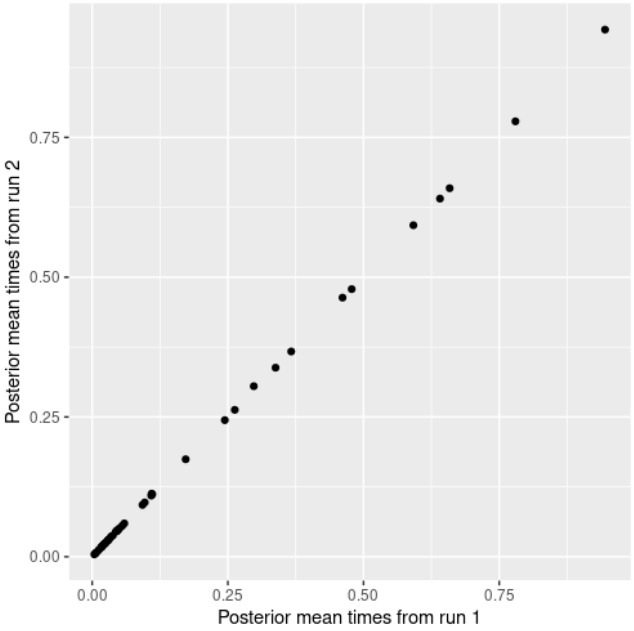


**Figure S13:** Posterior mean times of the different runs of approximations analyses. Similar posterior distributions between run 1 and 2 means the two approximations analyses successfully converged.

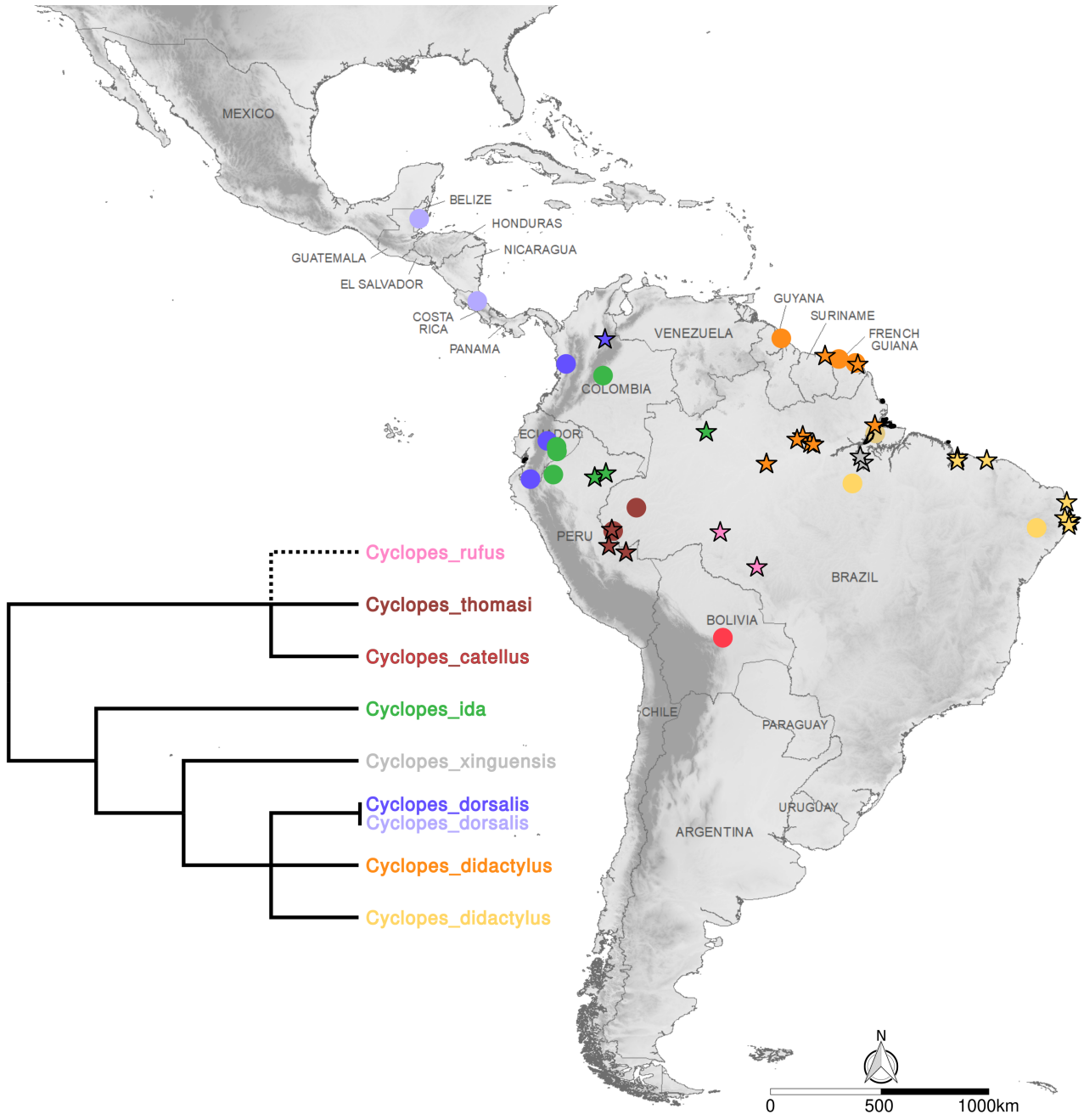


**Figure S14:** Distribution of 23 *Cyclopes* spp. evaluated in this study symbolized by dots and 33 from Coimbra et al (2017) symbolized by stars. Colors represent different species or lineages. The phylogenetics relationship within this genus is detailed by the cladogram; dotted line represent the uncertain position of *C. rufus*.

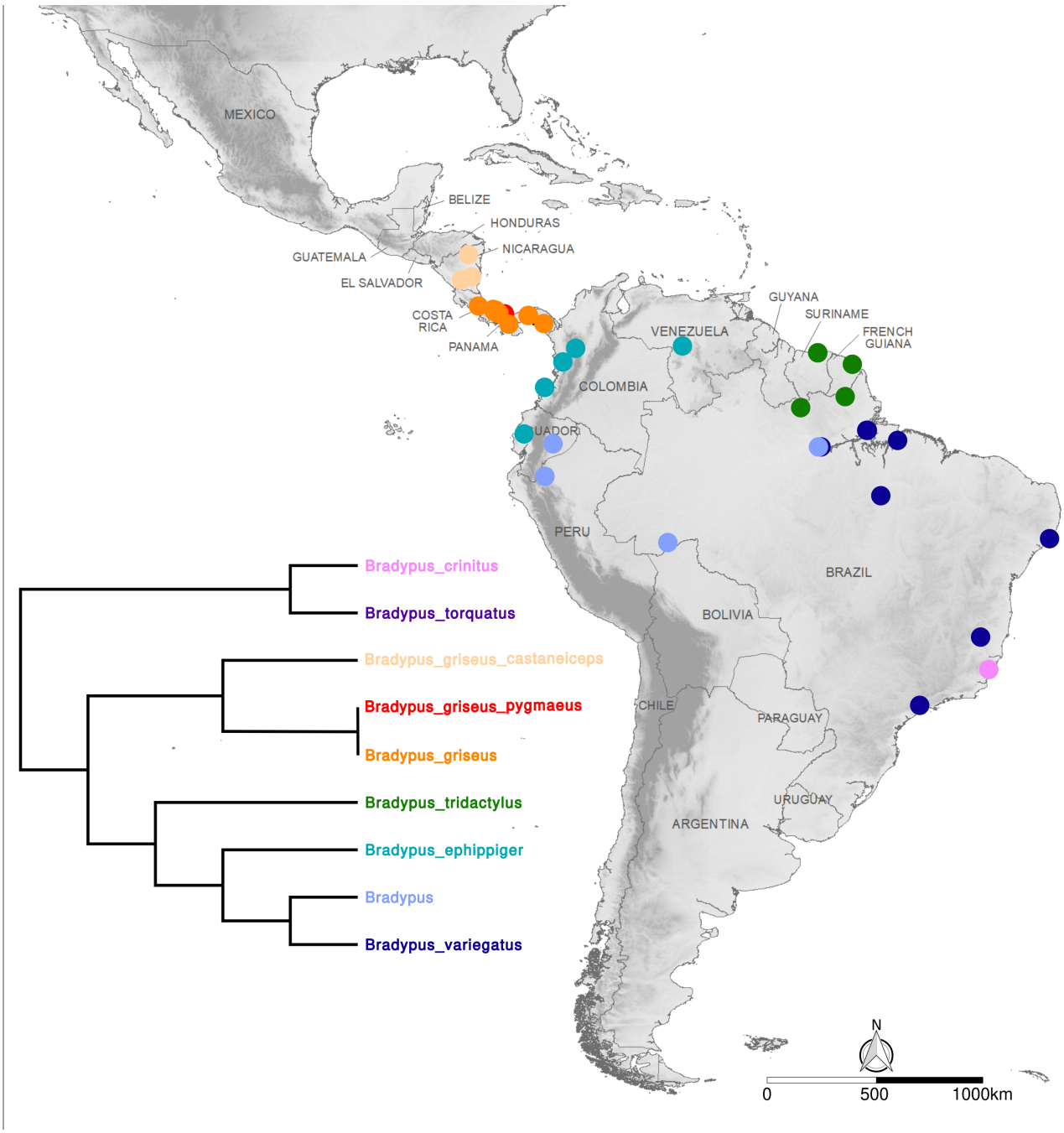


**Figure S15:** Distribution of 48 *Bradypus* spp. Colors represent different species or lineages. The phylogenetics relationship within this genus is detailed by the cladogram.

# DISCUSSION AND PERSPECTIVES

Throughout this thesis, we have delved into the taxonomic conundrum of xenarthrans and unraveled some complex phylogenetic relationships. Chapters 1 and 2 shed new light on the species diversity and evolutionary history of this mammalian group. In this final section, I would like to further discuss the implications and perspectives of these findings for understanding xenarthran evolution. I will start by discussing a more general topic: sampling strategies for species delimitation, emphasizing methods that minimize the impact on wild populations while addressing the challenges we encountered with such roadkill and museum samples. Subsequently, I will synthesize the taxonomic revisions made within xenarthrans as highlighted in this thesis, suggesting future directions for taxonomic investigations, and discussing the conservation implications of these results. In addition, I will discuss how genome-based phylogenies have enhanced our comprehension of xenarthran evolution. Finally, in the last section, I will raise two broader questions: i) What factors (biogeography, local adaptation, demography) have contributed to speciation and diversification in this mammalian group?; and ii) Is the species taxonomic status characterized by homogeneous genetic differentiation among mammals? Preliminary results further expanded in the Appendix are also presented in support of this discussion and perspectives section.

## Sampling and sequencing strategies for species delimitation: dealing with non-invasive samples

In this phase of declining biodiversity, describing living organisms and understanding the drivers of their speciation is a key to designing adequate conservation measures. In this context, minimizing the impact on populations is a crucial element in producing ethical science, and even more when focussing on threatened species (Winker et al., 2010). However, providing an accurate species de-

limitation hypothesis is often based on population genetics methods that require sampling of a large number of individuals. For this reason, non invasive methods such as roadkill or museum specimens hosted in natural history collections offer great potential.

Unfortunately, several practical issues complicate the use of these types of samples, which have generally been stored or found in non-optimal conditions for preserving DNA (e.g. ambient temperature, formalin (Camacho-Sanchez et al., 2013; Raxworthy and Smith, 2021). Indeed, after the death of the organism, DNA repair mechanisms cease working and damage caused by enzymes that are naturally present in cells, microbial activity, or environmental factors (e.g. UV radiations) induce DNA fragmentation and postmortem mutations (Eglinton and Logan, 1991; Lindahl, 1993). For roadkill samples, their storage at the adequate temperature (e.g. cryopreservation) and/or buffers (e.g. 95% ethanol, NAP buffer, RNAlater) is key to preserve DNA integrity by limiting biological and chemical pathways (Camacho-Sanchez et al., 2013; Wong et al., 2012). As degradation after death increases over time (Johnson and Ferris, 2002), the faster the sample is stored the better the DNA. For museum samples, storage techniques initially intended to preserve specimens, could have detrimental effects on DNA (Do and Dobrovic, 2015; Hykin et al., 2015; Zimmermann et al., 2008). The highly degraded nature of DNA from museum samples generally yields to restricted quantities of endogenous DNA contributing to a high failure rate in sequencing (Ewart et al., 2019). In addition, there is higher susceptibility to contamination especially when museum/old samples are treated simultaneously with more recent samples containing high amounts of DNA, which generates additional sequencing costs to obtain enough endogenous DNA reads.

So, even if museum specimens and roadkill samples provide a great biodiversity source for generating genomic data, the difficulty associated with sequencing, contamination, and DNA damage that could prevent accurate downstream analyses, may have restricted their more widespread use. This highlights the need to find adequate sequencing and bioinformatic solutions to treat those samples. During my PhD project, I have extracted DNA from 11 roadkill and 68 museum specimens and attempted to face the challenges inherent to these samples. In this section, I wanted to share our experience through our successes but also our failures.

*Extraction & libraries* - Contrary to roadkill, museum samples do not always contain the DNA quantity required for sequencing. Indeed, we mainly used dried skin pieces, sometimes no larger than a few millimeters to minimize damage to collections. Thus, five out of the 68 museum samples extracted during my thesis had too low DNA content to construct libraries. Factors influencing DNA preservation of museum specimens (e.g. time in collection, tissue type, storage technique employed) are still poorly understood but could provide precious information to select tissues or specimens as optimal sources of DNA. For example, Casas-Marce et al. (2010) and McDonough et al. (2018) evaluated the quality of different types of tissue sampled from museum specimens and

found a greater DNA quality of keratinous tissues such as claws. However, a recent special issue on museomic advances published by Frontiers in Ecology and Evolution, regrouping multiple articles that evaluated DNA preservation of museum samples, found contrasting results on the effect of taxon, tissue type, age, and preservation history (but see Fong et al. (2023) and references therein). This highlights our need to further explore DNA preservation factors to increase sequencing success.

*Short reads versus long reads* - With their highly fragmented DNA, roadkill and museum samples are well adapted to short read sequencing. In our case, we used both paired-end (PE) and single-end (SE) sequencing through various reads sizes (i.e 75pb SE, 100pb SE, 100pb PE, 150pb PE). We also tried the fastidious experience of sequencing roadkill with long reads. Not less than 11 individuals were extracted. The four least fragmented individuals with enough material were sequenced using Oxford Nanopore Technology on a GridION device. Despite this selection, we encountered two main problems. First, the sequenced fragments were very small. In the first flow cell we ran, the N50 value was of only 1.3kb (good quality samples could have N50 ten times higher). Second, the flow cell lost many pores (clogged or dead) and therefore generated a very small amount of data. We supposed this was due to the poor purity of our samples. In our case, the best solution was to perform beads purification that fixed the longest DNA fragments to magnetic beads. Using a magnetic support, the beads are plated on one side of the tube allowing to clean the eluent and shortest fragments of DNA. Using this technique, we reached a N50 of 2.1kb for the *Dasypus guianensis* M1752 individual (Figure 1).

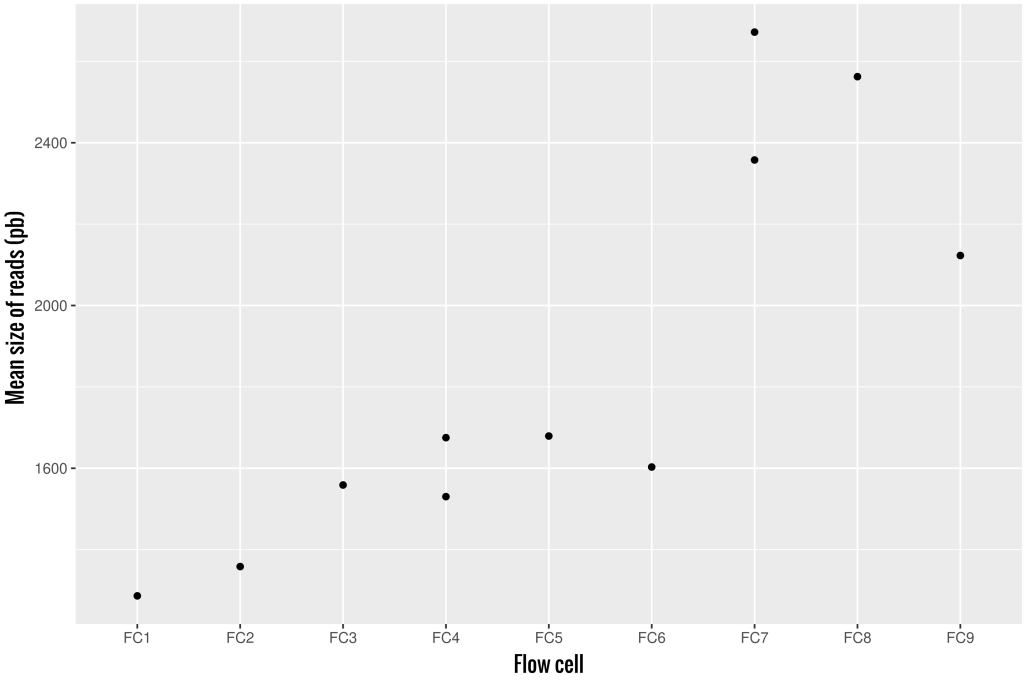


**Figure 1:** Mean size of the long reads fragments for the nine flow cells (FC) used.

*Shotgun versus capture* - Shotgun sequencing and sequencing targeted regions by capture methods are both technically feasible with these samples. However, capture presents the considerable advantage of targeting specific regions and thus limits sequencing of potential contaminant DNA from divergent organisms (e.g., fungi, bacteria). For this reason, when we used shotgun sequencing to obtain the whole genome of museum samples, we first performed an initial sequencing of 10 million reads to estimate the proportion of endogenous reads (Figure 2). We then performed a megablast against the nt_database using blastn v2.11.0 to identify reads attributed to xenarthran species (Chen et al., 2015). Indeed, sequencing highly contaminated samples increases the cost as more sequenced reads are required to produce enough data to reconstruct endogenous genomes. In our case, four individuals had less than 10% reads associated with xenarthrans and have thus been excluded from our selection of individuals for whole-genome sequencing (Figure 2).

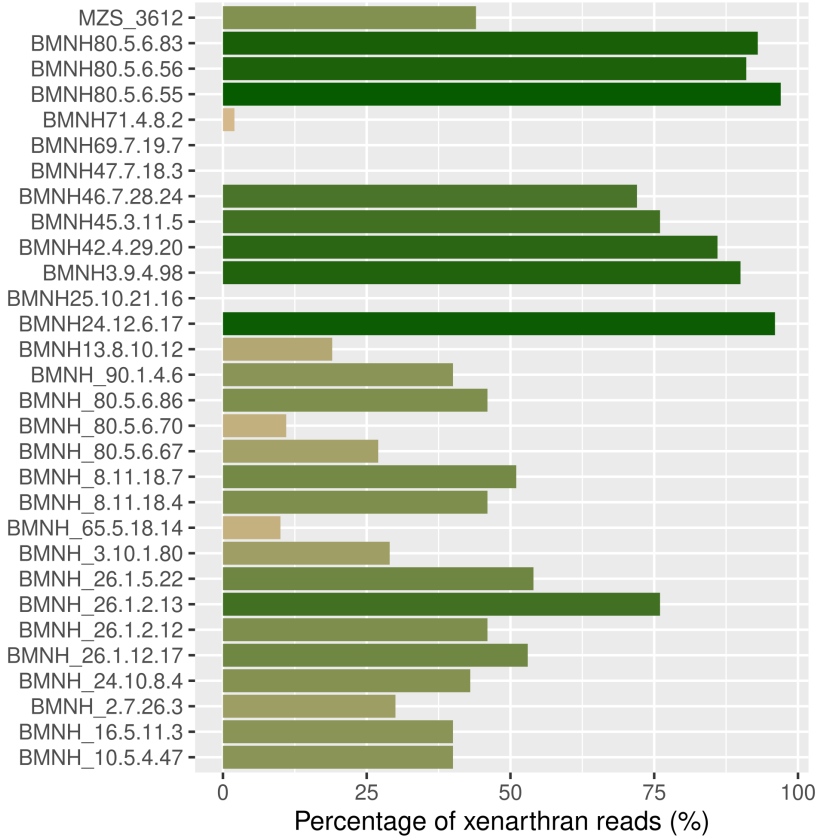


**Figure 2:** Percentage of reads corresponding to xenarthran species estimated by blast from 10 million reads sequenced for 30 individuals of *Bradypus* spp. and *Cyclopes* spp.

*Mapping versus assembly* - Mapping sequencing reads to a reference genome presents many advantages for these degraded of samples. First, it eliminates reads too far from the reference that can be attributed to contamination. Second, it allows reconstructing more contiguous genomes from these fragmented samples than de novo assembly. The problems with mapping are that: 1) reads can be excluded due to high error rate caused by postmortem mutations, and 2) it requires high-

quality genome assembly of a close relative to serve as reference genome. Even though numerous xenarthran reference genomes were already available, for *Dasypus*, only a moderate quality (46,379 scaffolds) reference genome of *D. novemcinctus* (Dasnov3.2) was published at the beginning of my PhD project. Therefore, we wanted to fill this gap by generating de novo assembly using long reads from roadkill samples of *D. guianensis* and *D. kappleri* collected in French Guiana. However, both hybrid assemblers (MaSuRCA v.4.1.0 (Zimin et al., 2013), HASLR v.0.8a1 (Haghshenas et al., 2020)) and long read assemblers (Flye v.2.9.1-b1781 (Kolmogorov et al., 2019), NextDenovo v.2.5.2 (Hu et al., 2023)) failed to assemble their 3.6 Gb genomes. We also tried using long reads to scaffold the short read assembly reconstructed with Megahit v.1.2.9 (Li et al., 2016) using SAMBA (Zimin and Salzberg, 2022). Despite our efforts to sequence the longest fragments (see previous section "Short reads versus long reads"), the distribution of read length was too small to assemble the repeat-reach genomes of these two long-nosed armadillo species.

*Errors versus biological signals* - Even though captured loci and mapping approaches limit contamination from divergent DNA sources, sequencing reads contaminated from close relatives will not be excluded by these techniques. In our case, we sequenced multiple individuals of the same species complexes, so potential cross-contamination could really be problematic. Indeed, as discussed in the first chapter, cross-contamination can artefactually be called as heterozygous positions, thus mimicking gene flow. Identifying contamination from close relatives and disentangling these errors from biological signals is still really difficult (Irestedt et al., 2022). For this reason, the best practice is minimizing contamination risk upstream of sequencing. In our case, we performed DNA extractions in controlled conditions suitable for degraded DNA: we manipulated samples one by one under a hood, cleaning with UV between each new sample. Despite these precautionary measures, we estimated individual supports for multiple mitochondrial haplotypes corresponding to those of the distinct lineages of our species complex. This method was really similar to the one employed by Green et al. (2008, 2010) who used mitochondrial diagnostic markers, which were attributed to extant human or Neanderthal and examined their support in mitochondrial alignments. This method allowed us to identify two samples with evidence of cross-contamination.

When working with degraded DNA such as museum or roadkill samples, every step is challenging and requires effort to optimize the chances of success, so what could be more frustrating than finding evidence for cross-contamination? In many studies, samples that have been cross-contaminated are discarded from subsequent analyses despite the loss of data that is induced. However, during my project, we employed filters on heterozygous positions based on read frequency. Indeed, both alleles of heterozygous SNPs are expected to be supported by ~50% of reads covering the position contrary to contaminated reads that could be expected at lower read frequency than endogenous DNA. Museum samples are also more prone to genotyping errors than

fresh tissues (e.g. postmortem mutations, sequencing errors, etc…; Ewart et al., 2019). Multiple approaches allowed estimating the proportion of genotyping, but here, by filtering out SNPs with low read frequency, we also minimized their potential bias for genotyping errors.

Even though improvements are still required to maximize the success of extraction, sequencing, bioinformatics, and cleaning (Raxworthy and Smith, 2021), these types of degraded DNA samples have already allowed successful species delimitation (Helgen et al., 2013; Mason et al., 2016). My thesis project has contributed to illustrate the power of non-invasive methods to delimit non-model species, and also proposed a cross-contamination detection approach and some filters to minimize genotyping errors that can increase accuracy of results based on museum and roadkill samples.

# Systematic revision of xenarthran species

## Taxonomic revision

During my PhD project, we analyzed 261 individual mitogenomes, 71 individuals with genome-wide exon capture data, and 93 individuals with whole genome sequences. Together, this represents 34 distinct xenarthran species out of the 39 described at the beginning of my project. In Chapter 1, we mainly employed population genetics analyses to delimit species within the long-nosed armadillo complex (*Dasypus* spp.) as 56 individuals were representing the four morpho-groups identified by Hautier et al. (2017) and Billet et al. (2017). Phylogenetic delimitation methods and gene flow analyses support these four lineages as distinct species. In addition, we also compared pairwise genetic differentiation between individuals of these lineages with a taxon that we defined as the reference for the taxonomic status of species (*D. pilosus*). In our case, employing this approach successfully corroborated results obtained with other species delimitation methods and thus supported the revalidation of *D. mexicanus, D. fenestratus* and the description of a new species endemic to the Guiana Shield, *D. guianensis*. This comparative approach has been proposed by Galtier (2019) and allows delimitation based on a more objective threshold of genome-wide divergence and encourages homogeneous delimitation between taxa. In addition, this method can be applied even with a single individual representative of each lineage. Indeed, in a panmictic population, the two alleles constituting an individual are inherited from non-related parents and thus, the heterozygosity estimated from these alleles should reflect the genetic diversity of the population. In addition, as a result of recombination, different regions of the genome are expected to be independent and thus could serve as replicates for the estimation of genetic diversity. Considering this, we used this method at the xenarthran scale in Chapter 2 by considering whole genomes of 36 out of the updated 42 xenarthran species with a restricted number of individuals by species. Species with a

consensual taxonomic status were used as references and provided a comparative framework to delimit the three-toed sloth (*Bradypus* spp). and pygmy anteater (*Cyclopes* spp.) species complexes. This allowed the revalidation of *Bradypus griseus* and *B. eppiphiger* as distinct species and suggested the existence of an additional new species of pygmy anteater in Northeast Brazil which have to be describe. Moreover, other potentially distinct lineages of *Bradypus* (in South America), and *Cabassous* (in Peru) require further exploration, notably to determine their exact geographical distribution and taxonomic status. Indeed, by evaluating taxonomic rank based on a unique representative of each lineage, the comparative method provided only restricted information on the geographical distribution of these lineages. Thus, even though this approach can make use of limited sampling of individuals from wild populations, some cases require a more representative sampling across their geographical distribution. On the other hand, for three species, *Bradypus pygmaeus*, *Dasypus sabanicola*, and *Dasypus mazzai*, our results suggest little genetic differentiation with *Bradypus griseus* in Central America and *Dasypus novemcinctus* in South America, respectively. For the pygmy sloth (*B. pygmaeus*), we directly used the species holotype and a paratype, so validity of these samples as representants of the species cannot be questioned, contrary to *Dasypus sabanicola* and *Dasypus mazzai*. Holotypes must be evaluated to potentially confirm these species as junior synonyms of *D. novemcinctus*. Together, Chapter 1 and 2 have contributed to improve our understanding of the species richness in Xenarthra by increasing the number of xenarthran species from 39 to 43 with five species requiring further investigations (*Dasypus sabanicola, Dasypus mazzai, Bradypus variegatus, Cyclopes didactylus* and *Cabassous unicinctus*).

Furthermore, taxonomic revision significantly influences conservation status, particularly in the case of splitting widely distributed species. For this reason, in Chapter 2, we have evaluated the genetic diversity based on genome-wide heterozygosity and runs of homozygosity (ROH). These results suggest prioritizing the evaluation of the conservation status of *Choloepus hoffmanni, Dasypus pilosus, Chlamyphorus truncatus*, and *Bradypus ephippiger* considering their low genetic diversity. Indeed, many xenarthran species lack crucial information to evaluate adequate conservation measures such as their main threats or precise geographic distributions. This is mainly due to their peculiar and sometimes elusive ecology (e.g. fossorial, arboreal), which makes them difficult to study in the field. Thus, such genetic diversity information as performed can helpfully contribute and guide conservation reassessments. In addition, the insular population of the pygmy three-toed sloth (*Bradypus pygmaeus*) considered as Critically Endangered in the IUCN Red List has been recognized in Chapter 2 as an insular subspecies of *Bradypus griseus*. We hope this taxonomical change will not hinder its conservation as its occurrence in the island Escudo de Veraguas is severely threatened due to its limited range and ongoing habitat degradation (Voirin, 2015).

## Phylogenetic reconstruction

Following these taxonomic revisions, this thesis encompassed the most complete genome-wide datasets assembled for *Dasypus* spp. (Chapter 1) and Xenarthra (Chapter 2), which allowed us to provide comprehensive phylogenetic reconstructions of their species diversity. Our results mainly corroborate previous studies based on mitochondrial markers (Gibb et al., 2016) and/or some nuclear markers (Abba et al., 2015; Delsuc et al., 2003, 2002, 2012; Möller-Krull et al., 2007), but they also allowed deciphering remaining uncertainties on the evolutionary history of this major placental clade. Notably, in Chapter 1, we clarified a case of mitochondrial introgression in a western Mexico population of *Dasypus* with a mitochondrial genome corresponding to *D. fenestratus* and a nuclear genome to *D. mexicanus*, which had previously misled the interpretation of the distribution of these populations. In addition, our whole genome-based phylogenetic framework highlighted cases of topological conflicts between nuclear genes and sites reconstructions at the xenarthran scale (Chapter 2). Notably, the relationships within Tolypeutinae and Euphractinae revealed highly supported alternative topologies (Chapter 2), which explains why previous studies had great difficulty in resolving their internal relationships (Abba et al., 2015; Delsuc et al., 2003). Moreover, the inter-relationships of Euphractinae, Tolypeutinae, and Chlamyphorinae showed almost equivalent support for the three possible alternative topologies, therefore representing this node by a strict bifurcation would be misleading and a polytomy might better reflect their evolutionary history. This situation is similar to Paenungulata, where the inter-relationships of elephants (*Proboscidea*), hyraxes (*Hyracoidea*), and sea cows (*Sirenia*) showed an equivalent support for the three possible alternative topologies, which must be considered as a real polytomy (Bowman et al., 2023; but see Liu et al., 2024). Further disentangling factors that have led to the discordant signals at these nodes but also within *Cyclopes* spp., *Bradypus* spp. and *Dasypus* spp. revealed that both gene flow and incomplete lineage sorting have contributed to these discordant signals. In many cases, gene flow was almost well balanced between lineages, thus it would have been impossible to detect based on classical ABBA/BABA-like tests that only consider deviation from balanced support for alternative topologies (expected in case of ILS) as evidence of GF. Here, we used the recently proposed Aphid method that takes advantage of differences in coalescence times of gene trees (Galtier, 2023), as GF is expected to produce shorter genealogies than ILS, which successfully allowed detecting balanced GF between xenarthran lineages. Consequently, this evidence for GF suggests that for multiple xenarthran species, speciation has occurred through contact zone or secondary contact.

Even though our genome-wide phylogenetic reconstruction at the species level provided a better understanding of the evolutionary history of extant xenarthran species, this is just the tip of the iceberg, as extant species represent only a fraction of the diversity of xenarthrans during their evolu-

tionary history. Indeed, the extinction event that happened during the Pleistocene led to the disappearance of numerous species from the megafauna and notably 90% of sloth diversity (Lyons et al., 2004; Steadman et al., 2005). The recent extinction of this past diversity has eroded our capacity to understand the evolutionary history of xenarthrans using molecular data from extant species. Notably, estimating the paleodiversity of xenarthrans using such a restricted number of extant species provided unrealistic estimations (Appendix 1). Nevertheless, studying gene discordances in Chapter 2 suggested at least a case of gene flow from a ghost lineage within the *Cabassous* genus. In addition, unprecedented advances in ancient DNA sequencing have recently made it possible to decipher their relationships including extinct species that morphology alone did not clearly establish. Notably, Delsuc et al. (2019) completely revisited the phylogenetic relationships of six extinct sloth species using ancient mitogenomics and highlighted important disagreements with previous morphological reconstruction. Furthermore, including extinct species have important implications in our understanding of phenotypic evolution of this clade. For example, the extinct giant ground sloth, *Mylodon darwinii*, was recovered as the sister-group of extant two-fingered sloths (*Choloepus* spp.), implying the convergent evolution of arboreality in the extant *Bradypus* and *Choloepus* genera from a terrestrial ancestor (Delsuc et al., 2018, 2019). More recently, combining morphological and molecular characters, Tejada et al. (2023) reconstructed the phylogenetic relationships of both extant and extinct sloth species through a total-evidence method. This allowed them to infer more precise ancestral states, diversification dynamics, and historical biogeography but also revealed an important influence of taxon sampling due to a bias in fossil preservation in different regions. Therefore, despite considerable progress made in our understanding of xenarthran evolutionary history in the last decades, there is still much to be done, notably by including their past diversity.

## Environment, demography, adaptation: what factors have driven speciation in xenarthran species?

As exposed in the Introduction of this PhD thesis, environmental heterogeneity is a crucial driver of speciation by directly reducing gene flow (through isolation mechanisms that reduce the probability of matting) or indirectly reducing gene flow (by allowing fixation of neutral or adaptative genetic barriers for reproductive isolation). For this reason, biogeography, by exploring the distribution of the biodiversity and associated environmental patterns, is essential to understanding factors that have promoted speciation or contributed to the maintenance of species. In Chapter 2, we have notably discussed the potential impact of the Andean mountain range on speciation and maintenance of some xenarthran species barriers, as it delimits the distributions of *Tamandua tetradactyla* / *T. mexicana* (Moraes-Barros and Arteaga, 2015; Ruiz-García et al., 2021), *Cabassous unicinctus* / *C.*

*centralis* (Feijo and Anacleto, 2021; Moraes-Barros and Arteaga, 2015), *Cyclopes didactylus / C. dorsalis* (Coimbra et al., 2017; Miranda et al., 2018), *Dasypus novemcinctus / D. fenestratus* (Chapter 1; Barthe et al., n.d. ), *Bradypus ephippiger / B. variegatus* (Chapter 2) and the two divergent mitochondrial lineages of Hoffmann's two-toed sloth (*Choloepus hoffmanni*). To further explore this, we evaluated if the Andean elevation throughout the Cenozoic has influenced xenarthran diversification through diversification models and notably environmental dependent models (Appendix 1). These results suggest that northern and north-central Andes (as defined in Boschman and Condamine (2022)) did not significantly influence xenarthran diversification. However, the elevation of Colombian and Venezuelan Andes was positively correlated with the speciation rate of xenarthrans (Appendix 1). In these two regions, the Andes elevated more recently, starting during the Miocene and their build-up could have interrupted gene flow between populations on both sides therefore favoring speciation of the aforementioned species pairs. It should thus be interesting to evaluate if Andean elevation has increased speciation by reducing mating probability (i.e. geographic isolation) or by creating a new local environment (i.e. reproductive isolation).

Evaluating factors that have influenced the diversification of a clade (e.g. environment, temperature, diversity, etc…) can reveal potential drivers of speciation. However, as these diversification analyses seem restricted to correlations, a microevolutionary scale can allow evaluating molecular mechanisms that have led to the divergence of lineages. Notably, two main scenarios are generally proposed to explain how environmental changes influence divergence: i) heterogeneous environments favor local adaptation, and ii) environmental barriers split ancestral populations into smaller ones that are then more sensitive to drift. Understanding the relative implications of both these forces (i.e. adaptation and drift) in the speciation process remains challenging and has to be assessed. In this context, the case of the insular population of the pygmy three-toed sloth (*Bradypus griseus pygmaeus*) is particularly interesting (Appendix 2). Indeed, this population is significantly smaller in body size than the continental population (*Bradypus griseus*). This morphological change is frequently described for insular species, thus, this convergent insular dwarfism has been included as part of the island syndrome (Adler and Levins, 1994). Since Van Valen (1965) and Lomolino (1985), insular dwarfism has been explained by adaptive forces as a compromise between resource availability and predation pressures. However, restricted effective population size on islands (here, the population of *Bradypus griseus pygmaeus* is estimated at about 500 individuals) suggests an important effect of drift. We tried to disentangle the relative implication of adaptation and drift in this morphological change, unfortunately, our results, based on only two individuals of *Bradypus griseus pygmaeus*, are ambiguous (Appendix 2). Sequencing additional individuals will be key to properly assess this question.

# Is genetic differentiation between species homogeneous across mammals?

During the speciation process, geneflow progressively descreases and genetic differantiation increases. In the absence of a consensus for a species delimitation threshold, it has been proposed to use the genetic differentiation between recognized taxa as a threshold that would be less arbitrary (Galtier, 2019). Thus, in Chapters 1 and 2, we have evaluated genetic differentiation between pairs of individuals from different xenarthran species to serve as a taxonomic reference. This comparative approach provided a remarkable framework to delimit lineages with uncertain taxonomic status. Applying this method to a restricted number of individuals per lineages in Chapter 2 allowed to successfully corroborate results obtained with other species delimitation methods conducted at the population scale (Chapter 1) or by previous studies based on morphology (Billet et al., 2017; Hautier et al., 2017; Miranda et al., 2018; Miranda et al., 2023).

However, this comparative approach relies on two assumptions: i) taxa used as references must represent true species, and ii) their genetic differentiation (divergence, genetic structure) must be a good proxy for taxonomic status that can then be transposed to an independent species complex. The first assumption is still difficult to evaluate since we have no consensual species concept. To evaluate the veracity of the second assumption, it is necessary to examine how genetic differentiation (divergence, genetic structure) varies with taxonomic status. Rosel et al. (2017) have explored this question through populations, subspecies, and species pairs of cetaceans. Overall, their results suggest an overlapping distribution of different population genomic measures (Fst, fixed differences, $\phi$ST) between taxonomic status, the only exception was for the Da (i.e. net synonymous divergence), which allowed discerning the taxonomic rank: a value above 2% corresponds to distinct species, while a value below 0.05% indicates populations. Even though these results suggest the Da as a good predictor of taxonomic rank, their results are restricted to mitochondrial data representing only cetaceans species/populations. Thus, it is still required to assess the distribution of this genetic differentiation at a broader taxonomic scale and also to estimate the potential influence of the genetic regions used (i.e. coding and non-coding) using whole genome data. We explored this with Matthieu Chombart, a master student by gathering 69 publicly available whole genomes representing height mammalian genera. Following the same pipeline as in Chapter 2, we estimated pairwise divergence (i.e. Da and Dxy) and genomic differentiation (Genomic Differentiation Index, GDI sensu Allio et al., 2021).

Preliminary results revealed lower values of divergence (Da and Dxy) for coding regions than randomly sampled regions (Figure A3.2, A3.3 in Appendix 3). This result is expected considering

that purifying selection acting on coding sequences reduces genetic diversity at these loci by counter selecting non-synonymous mutations with deleterious effect. This reduced within-population diversity has induced Cruickshank and Hahn (2014) to argue for higher FST values in these coding regions than randomly sampled regions. However, our results do not support this hypothesis as GDI estimations (i.e. FST (Nei and Chesser, 1983) applied to two diploid individuals) obtained for coding regions were similar to those of randomly sampled genomic regions (Figure A3.1 Appendix 3). These results are rather in line with those of Matthey-Doret and Whitlock (2019) who found no effect of background selection on FST estimations contrary to genetic diversity or Dxy. The variation in divergence estimates between the genetic regions used (i.e. coding or random) calls for caution when comparing studies based on coding or genome-wide data, unlike GDI whose estimates seem more equivalent. In addition, these preliminary results revealed higher values of GDI, Da, and Dxy for pairs of species than for pairs of lower taxonomic rank (Figure A3.4 Appendix 3). This suggests these metrics could approximate the taxonomic rank and usefully contribute to delimit species. However, we unfortunately have not yet been able to conduct a regression analysis (e.g. linear model) to properly evaluate the effects of all these parameters on genetic differentiation. Indeed, the genetic differentiation estimation (i.e. the response variable) does not satisfy the assumption of statistical independence due to our pairwise comparison approach. To correct for this non-independence, we plan to follow (Monnet, 2023) who faced a similar issue of non-independence since pairs can have individuals in common. To figure it out, they included two family effects to account for the possible covariance induced by each member of a pair and sampled pairs making sure an individual is not in the two distinct family effects at the same time.

Despite this, based on general tendency, we observed variations between different clades, for example the two species of gorilla (i.e. *Gorilla gorilla* and *Gorilla beringei*; GDI = 0.44, Da = 0.0023, Dxy = 0.003) have genetic differentiation estimates comparable to the two subspecies of white rhinoceros (*Ceratotherium simum simum* and *Ceratotherium simum cottoni*; GDI = 0.42, Da = 0.0016, Dxy = 0.0019; (Figure A3.1, A3.2, A3.3 Appendix 3). This difference could be biological (i.e. some lineage pairs can no longer interbreed at a certain value of genetic differentiation when others still can) if similar reproductive isolation is associated with different values of genetic differentiation. However, Roux et al. (2016) identified an interruption of gene flow above 2% of net synonymous divergence between various metazoan pairs of populations/species, this shared threshold rather suggesting homogeneous species boundaries between taxa. This raises two questions: i) do we want taxonomy to reflect homogeneous divergence between taxa?, and ii) do we want the taxonomic rank of species to correspond to a complete cessation of gene flow?

## Personal reflection on taxonomy

To end on a more reflective note, I wanted to share a few lines on my personal point of view concerning the discrete vision of taxonomy. This aspect has long seemed paradoxical to me, given that the process of speciation is continuous. Taxonomy was developed at a time when we really thought living organisms as discrete and fixed entities. Even if we no longer have this vision, we still conserve this discrete framework today. The main reason is that it simplifies the complexity of the living world, facilitating scientific communication and the mathematical modeling of biodiversity. Unfortunately, retaining this framework, which does not formally reflect evolution, has a considerable impact on the way we think about the living world, and contributes to the difficulty of understanding evolution by students (Manikas et al., 2023). I believe that the comparative approach proposed by Galtier (2019) brings more nuance to species delimitation, which could contribute to a better representation of the ongoing process of speciation.

## General conclusion

This thesis has conducted the most comprehensive molecular systematic study of xenarthrans illustrating the power of genomics to reveal cryptic species diversity. Encompassing a nearly exhaustive whole genome dataset of Xenarthra, we suggest revalidated the species *Dasypus mexicanus*, *Dasypus fenestratus*, *Bradypus ephippiger*, and *Bradypus griseus*, synonymise the Critically Endangered *B. pygmaeus* with *B. griseus*, and reveal new species of long-nosed armadillo (*Dasypus guianensis*) and pygmy anteater which require to be described. Through this revised taxonomic framework, we reconstructed the most comprehensive time-calibrated phylogeny of Xenarthra based on whole genome data that shed light on their evolutionary history. By disentangling the respective contribution of incomplete lineage sorting and gene flow in discordant topological signals across the xenarthran phylogeny, we highlighted multiple events of gene flow suggesting speciation with contacts between lineages. Further exploring factors influencing speciation in xenarthrans suggested that Colombian and Venezuelan Andes have promoted their speciation by separating several species pairs. Finally, our results have contributed to identifying future directions for further taxonomic investigations and conservation status assessment of xenarthran species.

## References

Abba, A. M. et al. (2015). "Systematics of hairy armadillos and the taxonomic status of the Andean hairy armadillo (Chaetophractus nationi)". In: *Journal of Mammalogy* 96.4. ISBN: 1545-1542 Publisher: Oxford University Press US, pp. 673–689.

Adler, G. H. and R. Levins (1994). "The island syndrome in rodent populations". In: *The Quarterly review of biology* 69.4. ISBN: 0033-5770 Publisher: University of Chicago Press, pp. 473–490.

Allio, R. et al. (2021). "High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox". In: *Elife* 10. Publisher: eLife Sciences Publications Limited, e63167. ISSN: 2050-084X.

Barthe, M. et al. (n.d.). "Exon capture museomics deciphers the nine-banded armadillo species complex and identifies a new species endemic to the Guiana Shield". In: *Systematic Biology* (). Accepted.

Billet, G. et al. (2017). "The hidden anatomy of paranasal sinuses reveals biogeographically distinct morphotypes in the nine-banded armadillo (Dasypus novemcinctus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3593. ISSN: 2167-8359.

Boschman, L. M. and F. L. Condamine (2022). "Mountain radiations are not only rapid and recent: Ancient diversification of South American frog and lizard families related to Paleogene Andean orogeny and Cenozoic climate variations". In: *Global and Planetary Change* 208. ISBN: 0921-8181 Publisher: Elsevier, p. 103704.

Bowman, J., D. Enard, and V. J. Lynch (2023). "Phylogenomics reveals an almost perfect polytomy among the almost ungulates (Paenungulata)". In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 2023.12. 07.570590.

Camacho-Sanchez, M. et al. (2013). "Preservation of RNA and DNA from mammal samples under field conditions". In: *Molecular ecology resources* 13.4. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 663–673.

Casas-Marce, M., E. Revilla, and J. A. Godoy (2010). "Searching for DNA in museum specimens: a comparison of sources in a mammal species". In: *Molecular Ecology Resources* 10.3. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 502–507.

Chen, Y. et al. (2015). "High speed BLASTN: an accelerated MegaBLAST search tool". In: *Nucleic acids research* 43.16. ISBN: 0305-1048 Publisher: Oxford University Press, pp. 7762–7768.

Coimbra, R. T. F. et al. (2017). "Phylogeographic history of South American populations of the silky anteater *Cyclopes didactylus* (Pilosa: Cyclopedidae)". In: *Genetics and Molecular*

*Biology* 40. Publisher: Sociedade Brasileira de Genética, pp. 40–49. ISSN: 1415-4757, 1678-4685. DOI: 10.1590/1678-4685-GMB-2016-0040.

Cruickshank, T. E. and M. W. Hahn (2014). "Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow". In: *Molecular ecology* 23.13. Publisher: Wiley Online Library, pp. 3133–3157. ISSN: 0962-1083.

Delsuc, F., M. J. Stanhope, and E. J. P. Douzery (2003). "Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes". In: *Molecular Phylogenetics and Evolution* 28.2, pp. 261–275. ISSN: 1055-7903. DOI: 10.1016/S1055-7903(03)00111-8.

Delsuc, F. et al. (2002). "Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting". In: *Molecular Biology and Evolution* 19.10, pp. 1656–1671. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a003989.

Delsuc, F. et al. (2012). "Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos". In: *Molecular Phylogenetics and Evolution* 62.2, pp. 673–680. ISSN: 1055-7903. DOI: https://doi.org/10.1016/j.ympev.2011.11.008.

Delsuc, F. et al. (2018). "Resolving the phylogenetic position of Darwin's extinct ground sloth (Mylodon darwinii) using mitogenomic and nuclear exon data". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1878. ISBN: 0962-8452 Publisher: The Royal Society, p. 20180214.

Delsuc, F. et al. (2019). "Ancient mitogenomes reveal the evolutionary history and biogeography of sloths". In: *Current Biology* 29.12. ISBN: 0960-9822 Publisher: Elsevier, 2031–2042. e6.

Do, H. and A. Dobrovic (2015). "Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization". In: *Clinical chemistry* 61.1. ISBN: 0009-9147 Publisher: Oxford University Press, pp. 64–71.

Eglinton, G. and G. A. Logan (1991). "Molecular preservation". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 333.1268. ISBN: 0962-8436 Publisher: The Royal Society London, pp. 315–328.

Ewart, K. M. et al. (2019). "Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species". In: *Molecular Ecology Resources* 19.6. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 1578–1592.

Feijo, A. and T. C. Anacleto (2021). "Taxonomic revision of the genus Cabassous McMurtrie, 1831 (Cingulata: Chlamyphoridae), with revalidation of Cabassous squamicaudis (Lund, 1845)". In: *Zootaxa* 4974.1. ISBN: 1175-5334, pp. 47–78–47–78.

Fong, J. J. et al. (2023). "Recent advances in museomics: revolutionizing biodiversity research". In: *Frontiers in Ecology and Evolution* 11. ISBN: 2296-701X Publisher: Frontiers, p. 1188172.

Galtier, N. (2019). "Delineating species in the speciation continuum: A proposal". In: *Evolutionary Applications* 12.4. Publisher: Wiley Online Library, pp. 657–663. ISSN: 1752-4571.

— (2023). "Phylogenetic conflicts: distinguishing gene flow from incomplete lineage sorting". In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 2023.07. 06.547897.

Gibb, G. C. et al. (2016). "Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans". In: *Molecular Biology and Evolution* 33.3, pp. 621–642. ISSN: 0737-4038. DOI: 10.1093/molbev/msv250.

Green, R. E. et al. (2008). "A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing". In: *Cell* 134.3. ISBN: 0092-8674 Publisher: Elsevier, pp. 416–426.

Green, R. E. et al. (2010). "A draft sequence of the Neandertal genome". In: *Science* 328.5979. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 710–722.

Haghshenas, E. et al. (2020). "HASLR: fast hybrid assembly of long reads". In: *Iscience* 23.8. ISBN: 2589-0042 Publisher: Elsevier.

Hautier, L. et al. (2017). "Beyond the carapace: skull shape variation and morphological systematics of long-nosed armadillos (genus Dasypus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3650. ISSN: 2167-8359.

Helgen, K. M. et al. (2013). *Taxonomic revision of the olingos (Bassaricyon), with description of a new species, the Olinguito*. ZooKeys. Vol. 324. Pensoft. 1-83. ISBN: 954-642-695-4.

Hu, J. et al. (2023). "An efficient error correction and accurate assembly tool for noisy long reads". In: *Biorxiv*. Publisher: Cold Spring Harbor Laboratory, p. 2023.03. 09.531669.

Hykin, S. M., K. Bi, and J. A. McGuire (2015). "Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing". In: *PloS one* 10.10. ISBN: 1932-6203 Publisher: Public Library of Science San Francisco, CA USA, e0141579.

Irestedt, M. et al. (2022). "A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins". In: *Molecular Ecology Resources* 22.7. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 2672–2684.

Johnson, L. A. and J. A. Ferris (2002). "Analysis of postmortem DNA degradation by single-cell gel electrophoresis". In: *Forensic science international* 126.1. ISBN: 0379-0738 Publisher: Elsevier, pp. 43–47.

Kolmogorov, M. et al. (2019). "Assembly of long, error-prone reads using repeat graphs". In: *Nature biotechnology* 37.5. ISBN: 1087-0156 Publisher: Nature Publishing Group US New York, pp. 540–546.

Li, D. et al. (2016). "MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices". In: *Methods* 102. ISBN: 1046-2023 Publisher: Elsevier, pp. 3–11.

Lindahl, T. (1993). "Instability and decay of the primary structure of DNA". In: *nature* 362.6422. ISBN: 0028-0836 Publisher: Nature Publishing Group UK London, pp. 709–715.

Liu, G. et al. (2024). "Phylogenomics of Afrotherian mammals and improved resolution of extant Paenungulata". In: *Molecular Phylogenetics and Evolution* 195. ISBN: 1055-7903 Publisher: Elsevier, p. 108047.

Lomolino, M. V. (1985). "Body size of mammals on islands: the island rule reexamined". In: *The American Naturalist* 125.2. ISBN: 0003-0147 Publisher: University of Chicago Press, pp. 310–316.

Lyons, S. K., F. A. Smith, and J. H. Brown (2004). "Of mice, mastodons and men: human-mediated extinctions on four continents". In: *Evolutionary Ecology Research* 6.3. ISBN: 1522-0613 Publisher: Evolutionary Ecology, Ltd., pp. 339–358.

Manikas, M. et al. (2023). “The species problem in evolution education”. In: *Journal of Mathematics and Science Teacher* 3.1.

Mason, V. C. et al. (2016). “Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates”. In: *Science advances* 2.8. ISBN: 2375-2548 Publisher: American Association for the Advancement of Science, e1600633.

Matthey-Doret, R. and M. C. Whitlock (2019). “Background selection and FST: consequences for detecting local adaptation”. In: *Molecular ecology* 28.17. ISBN: 0962-1083 Publisher: Wiley Online Library, pp. 3902–3914.

McDonough, M. M. et al. (2018). “Performance of commonly requested destructive museum samples for mammalian genomic studies”. In: *Journal of Mammalogy* 99.4. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 789–802.

Miranda, F. R. et al. (2018). “Taxonomic review of the genus Cyclopes Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species”. In: *Zoological Journal of the Linnean Society* 183.3, pp. 687–721. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlx079.

Miranda, F. R. et al. (2023). “Taxonomic revision of maned sloths, subgenus Bradypus (Scaeopus), Pilosa, Bradypodidae, with revalidation of Bradypus crinitus Gray, 1850”. In: *Journal of Mammalogy* 104.1. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 86–103.

Möller-Krull, M. et al. (2007). “Retroposed Elements and Their Flanking Regions Resolve the Evolutionary History of Xenarthran Mammals (Armadillos, Anteaters, and Sloths)”. In: *Molecular Biology and Evolution* 24.11, pp. 2573–2582. ISSN: 0737-4038. DOI: 10.1093/molbev/msm201.

Monnet, F. (2023). “Chapitre II: Test of two factors susceptible to influence the dynamics of speciation in plants: Selfing rate and life form” in Speciation dynamics, contrast between plants and animals. Thèse de doctorat, sous la direction de Xavier Vekemans, Camille Roux, Yves Van de Peer. Université de Lille.” In.

Moraes-Barros, N. and M. C. Arteaga (2015). “Genetic diversity in Xenarthra and its relevance to patterns of neotropical biodiversity”. In: *Journal of Mammalogy* 96.4, pp. 690–702. ISSN: 0022-2372. DOI: 10.1093/jmammal/gyv077.

Nei, M. and R. K. Chesser (1983). "Estimation of fixation indices and gene diversities". In: *Annals of human genetics* 47.3. ISBN: 0003-4800 Publisher: Wiley Online Library, pp. 253–259.

Raxworthy, C. J. and B. T. Smith (2021). "Mining museums for historical DNA: advances and challenges in museomics". In: *Trends in Ecology & Evolution* 36.11. ISBN: 0169-5347 Publisher: Elsevier, pp. 1049–1060.

Rosel, P. E. et al. (2017). "Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences". In: *Marine Mammal Science* 33 (S1). Publisher: Wiley Online Library, pp. 76–100. ISSN: 0824-0469.

Roux, C. et al. (2016). "Shedding light on the grey zone of speciation along a continuum of genomic divergence". In: *PLoS Biology* 14.12. Publisher: Public Library of Science San Francisco, CA USA, e2000234. ISSN: 1544-9173.

Ruiz-García, M. et al. (2021). "Comparative mitogenome phylogeography of two anteater genera (Tamandua and Myrmecophaga; Myrmecophagidae, Xenarthra): Evidence of discrepant evolutionary traits". In: *Zoological Research* 42.5. Publisher: Editorial Office of Zoological Research, Kunming Institute of Zoology, The …, p. 525.

Steadman, D. W. et al. (2005). "Asynchronous extinction of late Quaternary sloths on continents and islands". In: *Proceedings of the National Academy of Sciences* 102.33. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 11763–11768.

Tejada, J. V. et al. (2023). "Bayesian total-evidence dating revisits sloth phylogeny and biogeography: a cautionary tale on morphological clock analyses". In: *Systematic Biology*. ISBN: 1063-5157 Publisher: Oxford University Press US, syad069.

Van Valen, L. (1965). "Morphological variation and width of ecological niche". In: *The American Naturalist* 99.908. ISBN: 0003-0147 Publisher: Science Press, pp. 377–390.

Voirin, B. (2015). "Biology and conservation of the pygmy sloth, Bradypus pygmaeus". In: *Journal of Mammalogy* 96.4. ISBN: 1545-1542 Publisher: Oxford University Press US, pp. 703–707.

Winker, K. et al. (2010). "The importance, effects, and ethics of bird collecting". In: *The Auk* 127.3. ISBN: 0004-8038 Publisher: Oxford University Press, pp. 690–695.

Wong, P. B. et al. (2012). “Tissue sampling methods and standards for vertebrate genomics”. In: *Gigascience* 1.1. ISBN: 2047-217X Publisher: Oxford University Press, pp. 2047–217X–1–8.

Zimin, A. V. and S. L. Salzberg (2022). “The SAMBA tool uses long reads to improve the contiguity of genome assemblies”. In: *PLoS computational biology* 18.2. ISBN: 1553-734X Publisher: Public Library of Science San Francisco, CA USA, e1009860.

Zimin, A. V. et al. (2013). “The MaSuRCA genome assembler”. In: *Bioinformatics* 29.21. ISBN: 1367-4803 Publisher: Oxford University Press, pp. 2669–2677.

Zimmermann, J. et al. (2008). “DNA damage in preserved specimens and tissue samples: a molecular assessment”. In: *Frontiers in Zoology* 5. Publisher: Springer, pp. 1–13.

# APPENDIX 1 - EXPLORING THE INFLUENCE OF ANDEAN ELEVATION ON XENARTHRAN DIVERSIFICATION

## Introduction

The Neotropics, comprising South America and Central America, is a region of high diversity whose biogeographic history has long been considered a determining factor in the diversification of species that underlies its unique biodiversity (Antonelli et al., 2018; Hoorn et al., 2010; Meseguer et al., 2022). After splitting from the supercontinent Pangea in the Late Cretaceous (˜90 Mya; Simpson 1980; Figure A1.1), South America experienced progressive Andean uplift. The elevation started during Late Cretaceous (˜70 Mya) in: i) the northern Andes, in Ecuador (Montes et al., 2019), and ii) in the central Andes along the Pacific coast (Horton 2018). During the Paleogene (˜50 to ˜30 Mya), the uplift of these two regions has notably extended eastward and northward to Ecuador (Figure A1.1). In the middle Miocene (˜10 Mya), the elevation spread further eastward, reaching Venezuela in the northern part and covering a bigger part of Bolivia and Argentina (but see Boschman (2021), Boschman and Condamine (2022) and references therein). This Venezuelan Andean uplift induced hydrological rearrangements by shifting from the lacustrine Pebas system to the fluvial Amazonian river system ˜9 Mya (Hoorn et al., 2010, 2017; Figure A1.1). Overall, the Andean elevation had a major impact throughout South American landscapes and climate by acting as a barrier to westward air masses. By cooling down, these air masses cause heavy precipitations (Garreaud et al., 2009). In addition, this last northward elevation generated the closure of the Isthmus of Panama (3.5 Myr;Antonelli and Sanmartín, 2011; Hoorn et al., 2010), which promoted migration of numerous taxa known as the Great American Biotic Interchange (GABI) and led to the extinction of numerous species endemic to South America (Carrillo et al. 2020). Finally, the Quaternary was characterized by glacial cycles

that influenced the distribution of the Amazon landscape (Cheng et al., 2013) and the sea level (Val et al., 2021). All these factors influenced habitat fragmentation and have been invoked to explain the numerous speciation events that make the Neotropics a highly diverse region (Hoorn et al., 2022, 2010; Rull, 2011). As an ancient South American endemic clade, xenarthrans have evolved in isolation for most of the Cenozoic until the Great American Biological Interchange, and therefore represent a perfect model for studying the effect of these paleoenvironmental changes on their diversification (Delsuc et al., 2004; Gibb et al., 2016).

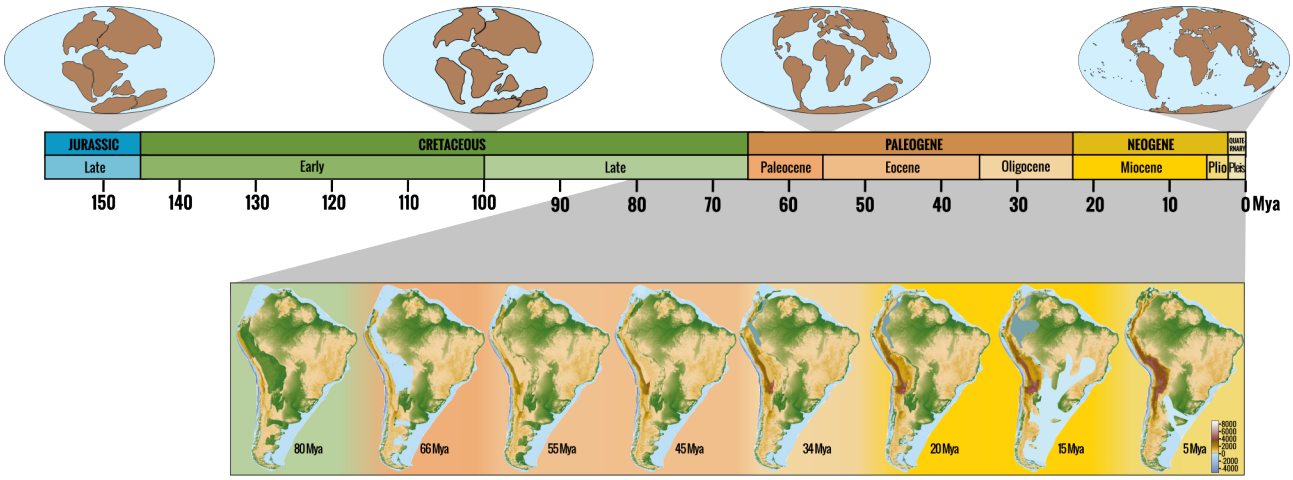


**Figure A1.1:** Paleogeographic maps of Earth (modified from Encyclopaedia Britannica and Meseguer and Condamine (2017)) along the geological timescale back to 150 Mya. Evolution of the South America landscape elevation is represented back to 80 Mya (modified from Boschman and Condamine (2022)). Marine incursions are colored in light blue and lakes and wetlands in gray-blue.

# Materials & Methods

## Estimating xenarthran diversification rates

Diversification rates of xenarthrans on the timescale phylogeny reconstructed in Chapter 2 have been estimated using Morlon et al. (2011)'s model automatized by Mazet et al. (2023) in the R package RPANDA 2.0 (Morlon et al., 2016)). This method supports heterogeneous diversification rates along the phylogeny by allowing subclades to follow distinct birth-death models (i.e. diversification shifts). The sampling fractions of five subclades (*Bradypus*, *Cyclopes*, *Dasypus*, Euphractinae, and Tolypeutinae) have been evaluated using the function get.sampling.fraction. Thirty-two combinations of these subclades were estimated by comb.shift function and were compared to a clade with no shift. For all these combinations, five diversification models were evaluated: i) BCST, constant speciation rate with no extinction rate corresponding to a Yule model; ii) BCST_DCST, constant speciation and extinction rates, corresponding to a constant birth-death model; iii) BVAR, variable speciation rate through time, with no extinction rate; iv) BVAR_DCST, variable speciation rate through time with a con-

stant extinction rate; and v) BCST_DVAR, constant speciation rate with variable extinction rate. The shift.estimate function estimated the best fitting model for all subclades and backbones and allowed to identify the best shift combination based on the AICc criterion. We compared diversification rates estimated without constraints, or constrained by the maximum rate set to 2 (rmax=2) or by the maximum number of species set to 10000 (nmax=10000). Finally, according to the best shift combination and diversification model, we estimated the paleodiversity dynamics of xenarthrans using the function apply_prob_dtt, which estimates a confidence interval using a probabilistic approach (Billaud et al., 2020).

### Environment-dependent models

The effect of Andean elevation was evaluated on a clade homogeneous (i.e. without diversification rate shifts) by comparing eight models using the function div.models. The eight models evaluated correspond to the five listed in the previous section (i.e. BCST, BCST_DCST, BVAR, BVAR_DCST, BCST_DVAR), plus three environment-dependent diversification models (i.e. BenvVAR, BenvVAR_DCST, BCST_DenvVAR). For the environment-dependent models, speciation or extinction rates followed an exponential dependency to the environmental factor. Here, we evaluated the effect of Andean elevation corresponding to three regions: i) the mean northern and central north Andes elevation estimated by Boschman and Condamine (2022), ii) the mean elevation of locations Santa Marta massif, Maracaibo, Garzon, Perija Santander, middle Magdalena valley basin, and iii) the mean elevation of Venezuelan coastal ranges and Merida venezuelan. For these three regions, the corresponding curve of elevation was estimated using a spline interpolation with degree of freedom of 3.67, 40, and 40 respectively. Finally, the best fitting model was selected using the AICc criterion.

## Results & Discussion

### Diversification rates and paleodiversity of xenarthrans

Phylogenetic birth-death models can be employed to determine whether a clade is most likely to have evolved with a constant or variable diversification rate over time. Thus, Gibb et al. (2016) reconstructed a molecular timescale phylogenetic tree at the species level based on their complete mitogenomes to further explore how xenarthrans diversification had evolved through time. They found a constant diversification rate through the Cenozoic (speciation rate ($\lambda$) and extinction rate ($\mu$) were both constant). However, these results are not concordant with paleontological data indicating that the late Pleistocene extinction event affected many xenarthran species (Lyons et al., 2004).

Since the work of Gibb et al. (2016), xenarthran taxonomy has been revised leading to an increased in the number of species from 31 in 2016 to 39 in 2023 (Feijo and Anacleto, 2021; Feijó et al., 2018, 2019; Gibb et al., 2016; Miranda et al., 2018; Miranda et al., 2023). In addition, considering the revisions conducted during this PhD project, we now enumerate 46 species. Taxon sampling is known to influence diversification estimation, for this reason, we actualised these diversification analyses using the time calibrated tree based on the whole genome of 40 xenarthran species (sampling fraction of 87%) estimated in Chapter 2. In addition, important technical improvements have emerged in the macroevolutionary field as Mazet et al. (2023) allowed the automatisation of Morlon et al. (2011) model and implemented it in the RPANDA 2.0 R package (Morlon et al., 2016). Indeed, this model allows diversification rates to vary not just over time but also across clades. This method works by evaluating different changes in diversification rates along the phylogeny (shifts) and estimates the birth-death model that fits the best for each subclade and backbone identified. In our case, we evaluated 32 different combinations of shifts and the best combination included five shifts located on the ancestral branches of *Bradypus*, *Cyclopes*, *Dasypus*, Euphractinae, and Tolypeutinae (Table A1.1, Figure A1.2). The best diversification model for the backbone included a constant speciation rate and a variable extinction rate increasing over time (as specified by the negative Beta value of -0.256; Table A1.2; Figure A1.2). For all subclades, the best model was a constant speciation rate (Figure A1.2). These results differ from those obtained by Gibb et al. (2016) and appear more congruent with the late Pleistocene extinction event during which numerous species from the megafauna became extinct, and notably 90% of sloth diversity (Lyons et al., 2004; Steadman et al., 2005). However, this estimation is anterior to this extinction event estimated to be of Pleistocene age by paleontological records.

In addition, paleodiversity dynamics reconstructed from this diversification rate estimate produces an estimation of the ancestral diversity of xenarthran that reached a peak of diversity higher than 230 000 species (Figure A1.3). Even though this clade is known to have been more diverse in the past, this estimation is far too unrealistic. Unfortunately, constraining these rates by maximum rate value (rate.max=2) or maximum number of species (n.max=10000) did not allow us to identify more realistic estimates, as rates reached the constraint values. Estimating xenarthran paleodiversity based only on molecular data seems challenging considering the restricted number of extant species in this clade, and even more considering that the diversification rates of the backbone were estimated from only seven species in our case. Thus, it should be more appropriate to include fossils to more accurately estimate their diversification.

**Table A1.1:** Support for the five best shift combinations estimated from the shift.estimate function detailing the number of parameters, the log likelihood, the AICc and the ΔAICc. Numbers in the Combination column correspond to the crown or subclades with 53 for *Bradypus*; 46: *Cyclopes*; 60: *Dasypus*; 69: EUPHRACTINAE; 75: TOLYPEUTINAE

| Combination | Parameters | logL | AICc | ΔAICc |
|---|---|---|---|---|
| 46.53.60.69.75 | 13 | -109.24 | 271.48 | 0.00 |
| whole_tree | 2 | -135.28 | 274.89 | 3.41 |
| 60 | 4 | -132.77 | 275.97 | 4.49 |
| 75 | 4 | -131.84 | 276.07 | 4.60 |
| 60.75 | 7 | -127.83 | 276.81 | 5.33 |

**Table A1.2:** Diversification models evaluated for the backbone of xenarthrans (see Materials & Methods for details on models). For each model, this table details the number of parameters in the model (Param.), the estimated log-likelihood (logL), the corrected Akaike information criterion (AICc), the difference with the lower AICc (ΔAICc). Estimations of model parameters are also detailed with Lambda corresponding to the speciation rate at present, Alpha the dependency of speciation rate on time, Mu the extinction rate at present, and Beta the dependency of extinction rate on time.

| Models | Param. | logL | AICc | Lambda | Alpha | Mu | Beta | ΔAICc |
|---|---|---|---|---|---|---|---|---|
| BCST_DVAR | 3 | -27.35 | 68.70 | 0.26 | – | 3.54 | -0.26 | 0.00 |
| BVAR_DCST | 3 | -35.28 | 84.56 | 0.96 | 0.01 | 1.17 | – | 15.86 |
| BCST_DCST | 2 | -42.87 | 92.74 | 0.21 | – | 0.21 | – | 24.05 |
| BCST | 1 | -45.64 | 94.08 | 0.03 | – | – | – | 25.38 |
| BVAR | 2 | -45.61 | 98.22 | 0.03 | 0.01 | – | – | 29.52 |

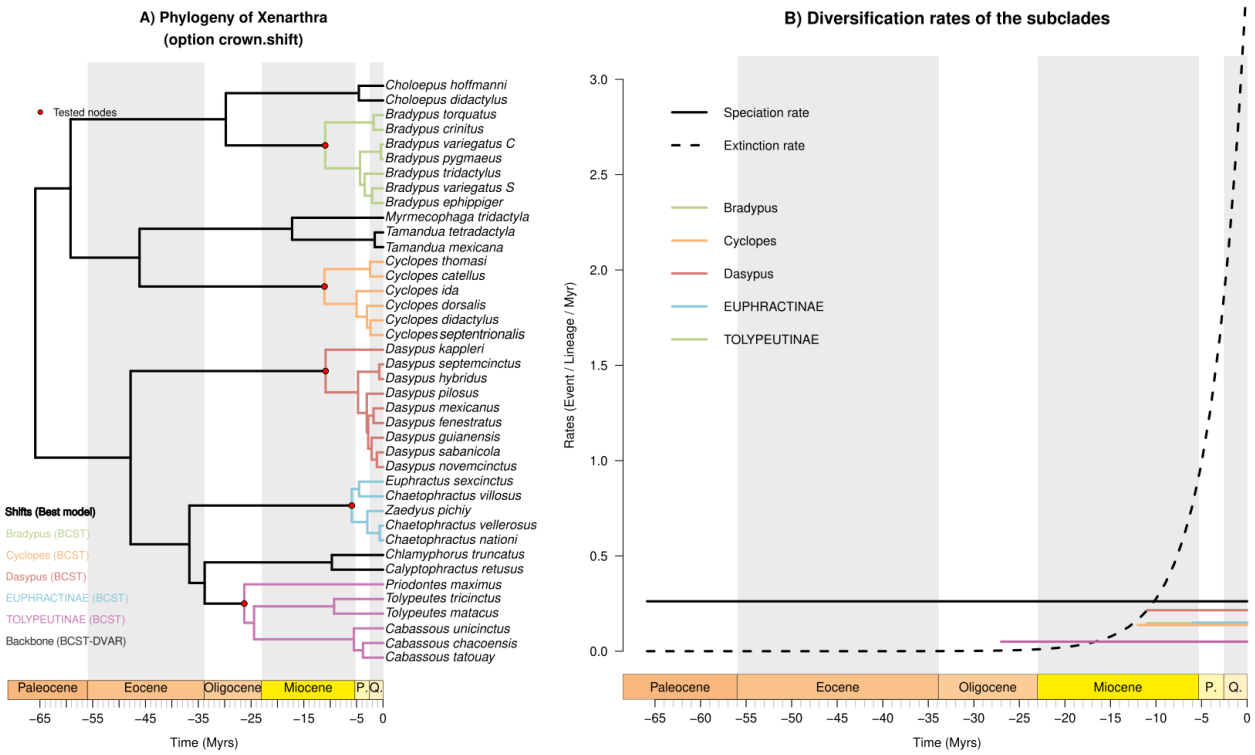


**Figure A1.2:** Diversification rates estimated for the xenarthran clade following the heterogeneous diversification rate model proposed by Morlon et al. (2011) and Mazet et al. (2023). A) Phylogeny of Xenarthra, with the five colored subclades corresponding to the best shift model. The red circles represent nodes tested for crown shift. For each subclade, the legend indicates the best model between parentheses. B) Estimation of diversification rates over time. The continuous line represents speciation rate, the dotted line represents extinction rate, and colors correspond to each subclade.

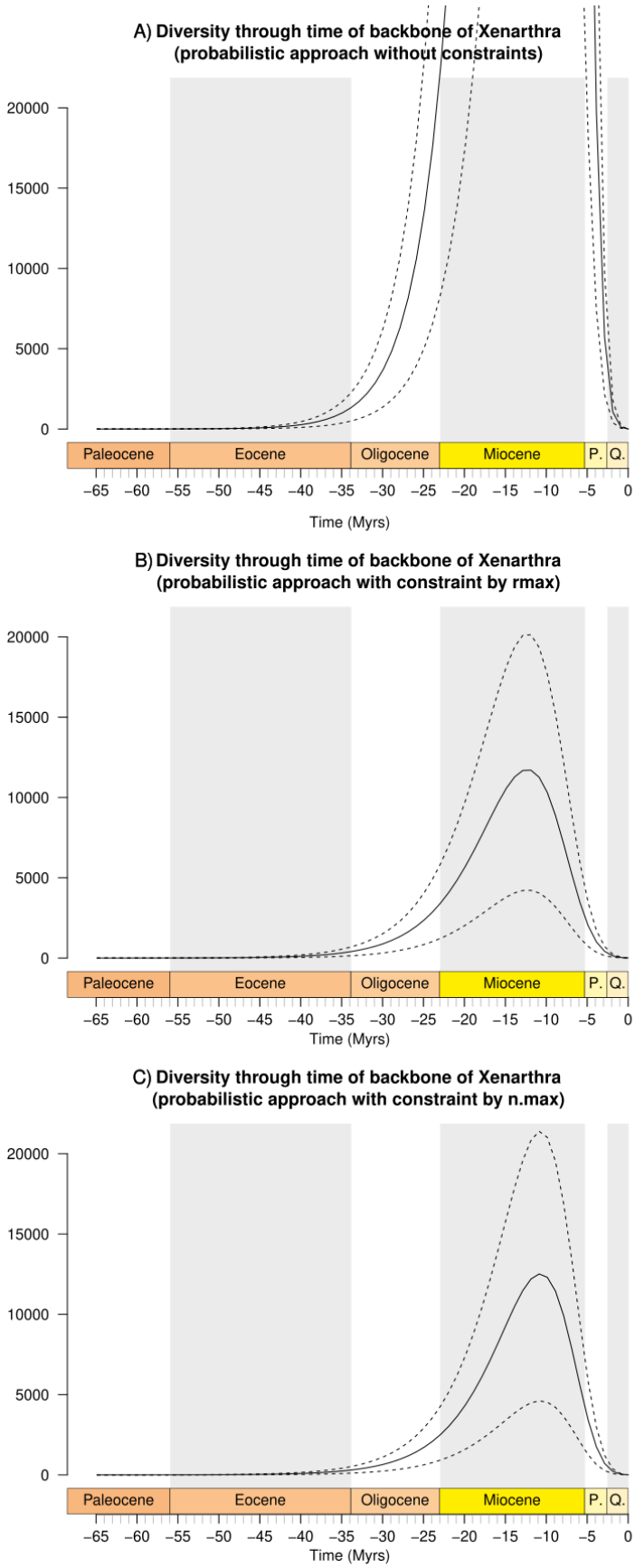


**Figure A1.3:** Paleodiversity reconstructions of the backbone of the xenarthran clade using the probabilistic approach. Dotted lines indicate the 95% confidence interval.

## Environmental factors driving diversification

Andean mountain range, resulting from multiple pulses of tectonic uplift throughout the Cenozoic, is a biogeographic barrier delimiting the distribution of multiple xenarthran species (e.g. *Dasypus fenestratus* and *D. novemcinctus*; *Bradypus ephippiger* and *B. variegatus*; *Cyclopes dorsalis* and *C. ida*). For this reason, Moraes-Barros and Arteaga (2015) and this Chapter 2 suggested that Andean uplift could have driven the diversification and distribution of extant xenarthran species. To evaluate if indeed Andean elevation has influenced xenarthran diversification, we estimated the likelihoods and the corrected Akaike information criterion (AICc) for different models, including environment-dependent models (i.e. varying according to an environmental variable, here Andean elevation) using the function div.models (Mazet, pers. comm., Condamine et al., 2013). First, we evaluated the impact of the northern and north-central Andes as defined in Boschman and Condamine (2022) on xenarthran diversification and on three subclades for which the Andes delimit the distribution of species pairs (i.e. *Dasypus*, *Bradypus*, *Cyclopes*). Despite close AICc values, the Andean elevation dependent model obtained lower support than the model with constant speciation and extinction rates ($AICc_{(BCST\_DCST)}$ = 274.9; $AICc_{(BAndesVAR)}$ = 276.5; Table A1.3). However, this Andean region includes three geomorphologic domains with distinct elevation histories (Boschman and Condamine, 2022). The first domain corresponds to the western part of the central and northern Andes (see Figure A1.4, and Boschman and Condamine (2022) for the original figure) and started its elevation in the central part during the Paleogene, progressively reaching the north and east. The domain 2 corresponds to the Eastern Cordillera and Garzon Massif (eastern part of the Columbian Andes), and domain 3 to the Merida Andes (a part of Venezuelan Andes). For these two domains, their elevation occurred more recently, starting during the Miocene. Thus, their elevation could have interrupted gene flow between populations on both sides and favored speciation. We evaluated this hypothesis, using the same methods based on the mean elevation of these domains (Figure A1.4). The domains 2 and 3 elevation dependent models best fitted xenarthran diversification, and revealed a positive correlation between their elevation and the speciation rate of the whole xenarthran phylogeny (Tables A1.4, A1.5). However, when evaluating subclades, only *Dasypus* revealed a significant effect of domain 3 elevation on its speciation, as for all other subclades, the constant speciation rate model offered the best fitting to their diversification.

Despite restricted statistical power due to the low number of species in this clade, xenarthran diversification appears to be correlated to the northern east elevation of the Andes. It would thus be interesting to evaluate if the Andean elevation can be a direct cause of increasing speciation as an environmental barrier reducing mating probability or by creating new local environments. Noteworthy, this environmental variable can have no influence on speciation and only be correlated to

another variable directly driving speciation. In addition, multiple factors can influence speciation, and it would be interesting to evaluate other environment-dependent, temperature-dependent, or diversity-dependent models on xenarthran diversification.

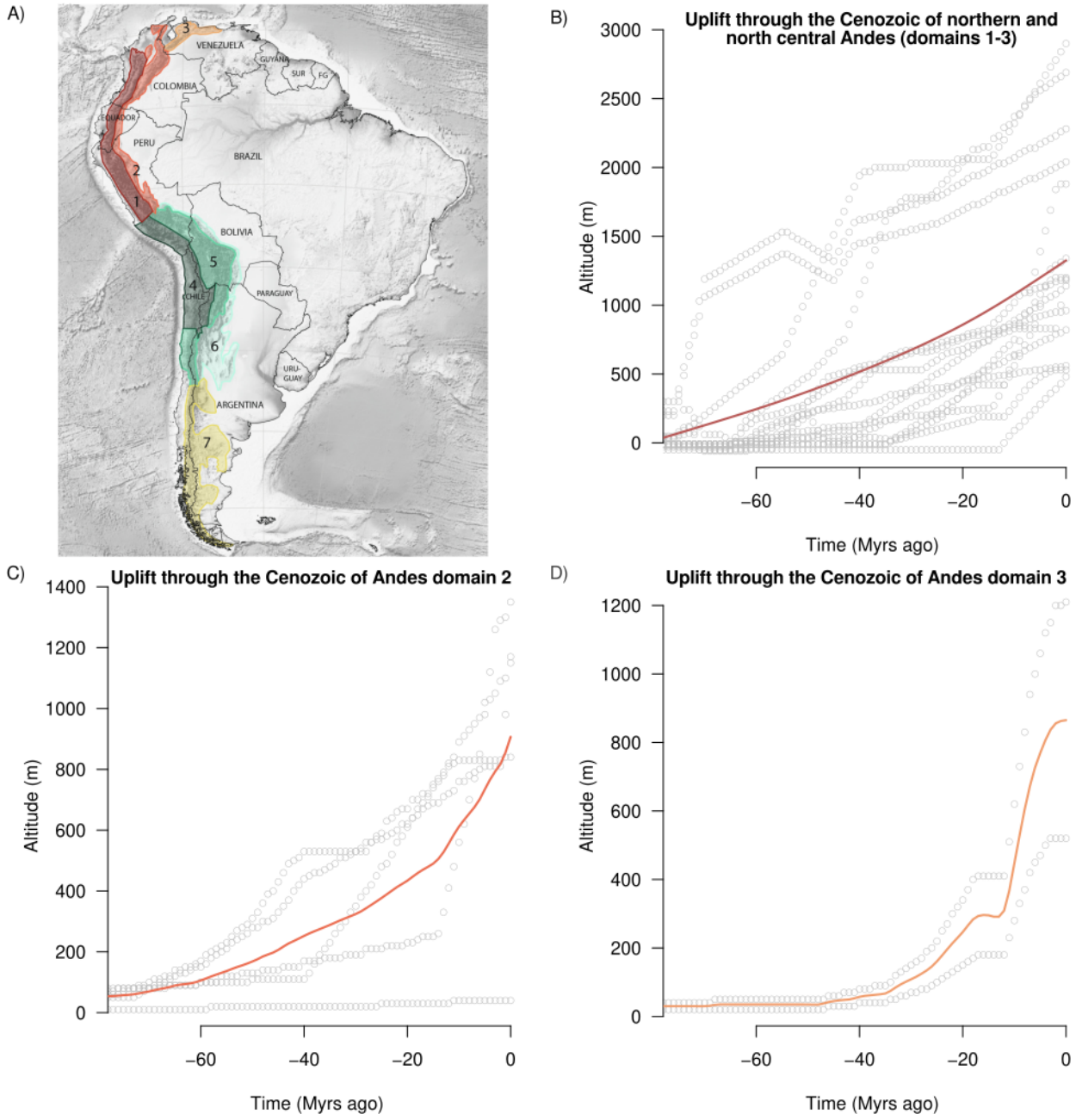


**Figure A1.4:** A) Map of South America with the seven geomorphological domains from Boschman and Condamine (2022). Paleoelevation of B) the northern and central Andes (red domains 1-3), C) the domain 2 (corresponding to the Santa Marta massif, Maracaibo, Garzon, Perija Santander, middle Magdalena valley basin locations), and D) domain 3 (corresponding to Venezuelan coastal ranges and Merida). Dots represent multiple locations and the curve corresponds to mean elevation estimated by the spline interpolation with degree of freedom of 3.67, 40 and 40 respectively.

**Table A1.3:** Diversification models including three environment-dependent models evaluated for xenarthrans, *Dasypus*, *Cyclopes* and *Bradypus* genera (see Materials & Methods for details on models). Environment-dependent models include an exponential dependency to the northern and central Andes elevation estimated by Boschman and Condamine, 2022 (see Figure A1.4 a-b). For each model, this table details the number of parameters in the model (Param.), the estimated log-likelihood (logL), the corrected Akaike information criterion (AICc), the difference with the lower AICc (ΔAICc). Estimations of model parameters are also detailed with Lambda corresponding to the speciation rate at present, Alpha the dependency of speciation rate on time, Mu the extinction rate at present, and Beta the dependency of extinction rate on time.

| Models | Param. | logL | AICc | Lambda | Alpha | Mu | Beta | ΔAICc |
|---|---|---|---|---|---|---|---|---|
| **Xenarthra** | | | | | | | | |
| BCST_DCST | 2 | -135.28 | 274.89 | 0.18 | – | 0.16 | – | 0.00 |
| BAndesVAR | 2 | -136.08 | 276.48 | 0.01 | 0.00 | – | – | 1.59 |
| BVAR_DCST | 3 | -134.99 | 276.64 | 0.20 | 0.01 | 0.22 | – | 1.75 |
| BCST_DVAR | 3 | -134.99 | 276.65 | 0.20 | – | 0.22 | -0.01 | 1.77 |
| BAndesVAR_DCST | 3 | -135.00 | 276.67 | 0.30 | 0.00 | 0.22 | – | 1.78 |
| BCST_DAndesVAR | 3 | -135.02 | 276.70 | 0.20 | – | 0.14 | 0.00 | 1.82 |
| BVAR | 2 | -136.85 | 278.03 | 0.13 | -0.04 | – | – | 3.14 |
| BCST | 1 | -142.50 | 287.10 | 0.07 | – | – | – | 12.22 |
| | | | | | | | | |
| ***Dasypus*** | | | | | | | | |
| BCST | 1 | -19.08 | 40.73 | 0.22 | – | – | – | 0.00 |
| BVAR | 2 | -17.95 | 41.90 | 0.41 | -0.21 | – | – | 1.17 |
| BAndesVAR | 2 | -17.96 | 41.92 | 0.00 | 0.01 | – | – | 1.18 |
| BCST_DCST | 2 | -18.46 | 42.93 | 0.39 | – | 0.37 | – | 2.20 |
| BVAR_DCST | 3 | -17.95 | 46.70 | 0.41 | -0.21 | 0.00 | – | 5.97 |
| BAndesVAR_DCST | 3 | -18.04 | 46.87 | 0.00 | 0.01 | 0.16 | – | 6.14 |
| BCST_DVAR | 3 | -18.33 | 47.46 | 0.34 | – | 0.21 | 0.05 | 6.73 |
| BCST_DAndesVAR | 3 | -19.08 | 48.96 | 0.22 | – | -0.01 | -0.12 | 8.23 |
| | | | | | | | | |
| ***Cyclopes*** | | | | | | | | |
| BCST | 1 | -12.73 | 28.47 | 0.14 | – | – | – | 0.00 |
| BVAR | 2 | -12.43 | 32.86 | 0.22 | -0.13 | – | – | 4.39 |
| BAndesVAR | 2 | -12.44 | 32.87 | 0.00 | 0.01 | – | – | 4.41 |
| BCST_DCST | 2 | -12.62 | 33.24 | 0.20 | – | 0.15 | – | 4.77 |
| BVAR_DCST | 3 | -12.43 | 42.86 | 0.22 | -0.13 | 0.00 | – | 14.39 |
| BAndesVAR_DCST | 3 | -12.44 | 42.88 | 0.00 | 0.01 | 0.00 | – | 14.41 |
| BCST_DVAR | 3 | -12.50 | 43.00 | 0.18 | – | 0.06 | 0.10 | 14.53 |
| BCST_DAndesVAR | 3 | -12.73 | 43.47 | 0.14 | – | 0.01 | -0.05 | 15.00 |
| | | | | | | | | |
| ***Bradypus*** | | | | | | | | |
| BCST | 1 | -14.76 | 32.32 | 0.17 | – | – | – | 0.00 |
| BVAR | 2 | -13.72 | 34.44 | 0.37 | -0.24 | – | – | 2.11 |
| BAndesVAR | 2 | -13.72 | 34.45 | 0.00 | 0.01 | – | – | 2.13 |
| BCST_DCST | 2 | -14.05 | 35.11 | 0.35 | – | 0.35 | – | 2.79 |
| BVAR_DCST | 3 | -13.72 | 41.44 | 0.37 | -0.24 | 0.00 | – | 9.11 |
| BAndesVAR_DCST | 3 | -13.73 | 41.46 | 0.00 | 0.01 | 0.05 | – | 9.14 |
| BCST_DVAR | 3 | -14.05 | 42.10 | 0.33 | – | 0.31 | 0.01 | 9.78 |
| BCST_DAndesVAR | 3 | -14.76 | 43.52 | 0.17 | – | -0.04 | -0.03 | 11.20 |

**Table A1.4:** Diversification models including three environment-dependent models evaluated for xenarthrans, *Dasypus*, *Cyclopes* and *Bradypus* genera (see Materials & Methods for details on models). Environment-dependent models include an exponential dependency to the domain 2 elevation of the Andes estimated by Boschman and Condamine, 2022 (see Figure A1.4 a-c). For each model, this table details the number of parameters in the model (Param.), the estimated log-likelihood (logL), the corrected Akaike information criterion (AICc), the difference with the lower AICc (ΔAICc). Estimations of model parameters are also detailed with Lambda corresponding to the speciation rate at present, Alpha the dependency of speciation rate on time, Mu the extinction rate at present, and Beta the dependency of extinction rate on time.

| Models | Param. | logL | AICc | Lambda | Alpha | Mu | Beta | ΔAICc |
|---|---|---|---|---|---|---|---|---|
| **Xenarthra** | | | | | | | | |
| BAndesVAR | 2 | -135.17 | 274.66 | 0.01 | 0.00 | – | – | 0.00 |
| BCST_DCST | 2 | -135.28 | 274.89 | 0.18 | – | 0.16 | – | 0.23 |
| BVAR_DCST | 3 | -134.99 | 276.64 | 0.20 | 0.01 | 0.22 | – | 1.98 |
| BCST_DVAR | 3 | -134.99 | 276.65 | 0.20 | – | 0.22 | -0.01 | 1.99 |
| BAndesVAR_DCST | 3 | -135.03 | 276.72 | 0.29 | 0.00 | 0.23 | – | 2.06 |
| BCST_DAndesVAR | 3 | -135.05 | 276.76 | 0.20 | – | 0.15 | 0.00 | 2.10 |
| BVAR | 2 | -136.85 | 278.03 | 0.13 | -0.04 | – | – | 3.36 |
| BCST | 1 | -142.50 | 287.10 | 0.07 | – | – | – | 12.44 |
| ***Dasypus*** | | | | | | | | |
| BCST | 1 | -19.08 | 40.73 | 0.22 | – | – | – | 0.00 |
| BVAR | 2 | -17.95 | 41.90 | 0.41 | -0.21 | – | – | 1.17 |
| BAndesVAR | 2 | -18.19 | 42.38 | 0.00 | 0.01 | – | – | 1.65 |
| BCST_DCST | 2 | -18.46 | 42.93 | 0.39 | – | 0.37 | – | 2.20 |
| BVAR_DCST | 3 | -17.95 | 46.70 | 0.41 | -0.21 | 0.00 | – | 5.97 |
| BAndesVAR_DCST | 3 | -17.98 | 46.76 | 0.00 | 0.01 | 0.02 | – | 6.03 |
| BCST_DVAR | 3 | -18.33 | 47.46 | 0.34 | – | 0.21 | 0.05 | 6.73 |
| BCST_DAndesVAR | 3 | -19.08 | 48.96 | 0.22 | – | -0.06 | -0.05 | 8.23 |
| ***Cyclopes*** | | | | | | | | |
| BCST | 1 | -12.73 | 28.47 | 0.14 | – | – | – | 0.00 |
| BVAR | 2 | -12.43 | 32.86 | 0.22 | -0.13 | – | – | 4.39 |
| BAndesVAR | 2 | -12.54 | 33.07 | 0.01 | 0.00 | – | – | 4.61 |
| BCST_DCST | 2 | -12.62 | 33.24 | 0.20 | – | 0.15 | – | 4.77 |
| BVAR_DCST | 3 | -12.43 | 42.86 | 0.22 | -0.13 | 0.00 | – | 14.39 |
| BAndesVAR_DCST | 3 | -12.44 | 42.88 | 0.00 | 0.01 | 0.00 | – | 14.42 |
| BCST_DVAR | 3 | -12.50 | 43.00 | 0.18 | – | 0.06 | 0.10 | 14.53 |
| BCST_DAndesVAR | 3 | -12.73 | 43.47 | 0.14 | – | 0.01 | -0.05 | 15.00 |
| ***Bradypus*** | | | | | | | | |
| BCST | 1 | -14.76 | 32.32 | 0.17 | – | – | – | 0.00 |
| BVAR | 2 | -13.72 | 34.44 | 0.37 | -0.24 | – | – | 2.11 |
| BAndesVAR | 2 | -13.85 | 34.69 | 0.00 | 0.01 | – | – | 2.37 |
| BCST_DCST | 2 | -14.05 | 35.11 | 0.35 | – | 0.35 | – | 2.79 |
| BVAR_DCST | 3 | -13.72 | 41.44 | 0.37 | -0.24 | 0.00 | – | 9.11 |
| BAndesVAR_DCST | 3 | -13.74 | 41.48 | 0.00 | 0.01 | 0.05 | – | 9.16 |
| BCST_DVAR | 3 | -14.05 | 42.10 | 0.33 | – | 0.31 | 0.01 | 9.78 |
| BCST_DAndesVAR | 3 | -14.76 | 43.52 | 0.17 | – | -0.04 | -0.03 | 11.20 |

**Table A1.5:** Diversification models including three environment-dependent models evaluated for xenarthrans, *Dasypus*, *Cyclopes* and *Bradypus* genera (see Materials & Methods for details on models). Environment-dependent models include an exponential dependency to the domain 3 elevation of the Andes estimated by Boschman and Condamine, 2022 (see Figure A1.4 a-d). For each model, this table details the number of parameters in the model (Param.), the estimated log-likelihood (logL), the corrected Akaike information criterion (AICc), the difference with the lower AICc (ΔAICc). Estimations of model parameters are also detailed with Lambda corresponding to the speciation rate at present, Alpha the dependency of speciation rate on time, Mu the extinction rate at present, and Beta the dependency of extinction rate on time.

| Models | Param. | logL | AICc | Lambda | Alpha | Mu | Beta | ΔAICc |
|---|---|---|---|---|---|---|---|---|
| **Xenarthra** | | | | | | | | |
| BAndesVAR | 2 | -134.13 | 272.59 | 0.02 | 0.00 | – | – | 0.00 |
| BCST_DCST | 2 | -135.28 | 274.89 | 0.18 | – | 0.16 | – | 2.30 |
| BAndesVAR_DCST | 3 | -134.53 | 275.72 | 0.03 | 0.00 | 0.00 | – | 3.13 |
| BVAR_DCST | 3 | -134.99 | 276.64 | 0.20 | 0.01 | 0.22 | – | 4.05 |
| BCST_DVAR | 3 | -134.99 | 276.65 | 0.20 | – | 0.22 | -0.01 | 4.07 |
| BCST_DAndesVAR | 3 | -135.15 | 276.96 | 0.20 | – | 0.17 | 0.00 | 4.38 |
| BVAR | 2 | -136.85 | 278.03 | 0.13 | -0.04 | – | – | 5.44 |
| BCST | 1 | -142.50 | 287.10 | 0.07 | – | – | – | 14.52 |
| | | | | | | | | |
| ***Dasypus*** | | | | | | | | |
| BAndesVAR | 2 | -17.12 | 40.24 | 0.00 | 0.01 | – | – | 0.00 |
| BCST | 1 | -19.08 | 40.73 | 0.22 | – | – | – | 0.50 |
| BVAR | 2 | -17.95 | 41.90 | 0.41 | -0.21 | – | – | 1.66 |
| BCST_DCST | 2 | -18.46 | 42.93 | 0.39 | – | 0.37 | – | 2.69 |
| BVAR_DCST | 3 | -17.95 | 46.70 | 0.41 | -0.21 | 0.00 | – | 6.46 |
| BAndesVAR_DCST | 3 | -17.96 | 46.71 | 0.00 | 0.01 | 0.00 | – | 6.47 |
| BCST_DVAR | 3 | -18.33 | 47.46 | 0.34 | – | 0.21 | 0.05 | 7.23 |
| BCST_DAndesVAR | 3 | -19.08 | 48.96 | 0.22 | – | -0.01 | -0.12 | 8.72 |
| | | | | | | | | |
| ***Cyclopes*** | | | | | | | | |
| BCST | 1 | -12.73 | 28.47 | 0.14 | – | – | – | 0.00 |
| BAndesVAR | 2 | -11.88 | 31.76 | 0.00 | 0.01 | – | – | 3.29 |
| BVAR | 2 | -12.43 | 32.86 | 0.22 | -0.13 | – | – | 4.39 |
| BCST_DCST | 2 | -12.62 | 33.24 | 0.20 | – | 0.15 | – | 4.77 |
| BVAR_DCST | 3 | -12.43 | 42.86 | 0.22 | -0.13 | 0.00 | – | 14.39 |
| BAndesVAR_DCST | 3 | -12.44 | 42.87 | 0.01 | 0.00 | 0.00 | – | 14.40 |
| BCST_DVAR | 3 | -12.50 | 43.00 | 0.18 | – | 0.06 | 0.10 | 14.53 |
| BCST_DAndesVAR | 3 | -12.73 | 43.47 | 0.14 | – | 0.01 | -0.05 | 15.00 |
| | | | | | | | | |
| ***Bradypus*** | | | | | | | | |
| BCST | 1 | -14.76 | 32.32 | 0.17 | – | – | – | 0.00 |
| BAndesVAR | 2 | -13.16 | 33.33 | 0.00 | 0.01 | – | – | 1.01 |
| BVAR | 2 | -13.72 | 34.44 | 0.37 | -0.24 | – | – | 2.11 |
| BCST_DCST | 2 | -14.05 | 35.11 | 0.35 | – | 0.35 | – | 2.79 |
| BVAR_DCST | 3 | -13.72 | 41.44 | 0.37 | -0.24 | 0.00 | – | 9.11 |
| BAndesVAR_DCST | 3 | -13.72 | 41.45 | 0.00 | 0.01 | 0.00 | – | 9.12 |
| BCST_DVAR | 3 | -14.05 | 42.10 | 0.33 | – | 0.31 | 0.01 | 9.78 |
| BCST_DAndesVAR | 3 | -14.76 | 43.52 | 0.17 | – | -0.04 | -0.03 | 11.20 |

# References

Antonelli, A. and I. Sanmartín (2011). "Why are there so many plant species in the Neotropics?" In: *Taxon* 60.2. ISBN: 0040-0262 Publisher: Wiley Online Library, pp. 403–414.

Antonelli, A. et al. (2018). "Amazonia is the primary source of Neotropical biodiversity". In: *Proceedings of the National Academy of Sciences* 115.23. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 6034–6039.

Billaud, O. et al. (2020). "Estimating diversity through time using molecular phylogenies: Old and species-poor frog families are the remnants of a diverse past". In: *Systematic Biology* 69.2. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 363–383.

Boschman, L. M. (2021). "Andean mountain building since the Late Cretaceous: A paleoelevation reconstruction". In: *Earth-Science Reviews* 220. ISBN: 0012-8252 Publisher: Elsevier, p. 103640.

Boschman, L. M. and F. L. Condamine (2022). "Mountain radiations are not only rapid and recent: Ancient diversification of South American frog and lizard families related to Paleogene Andean orogeny and Cenozoic climate variations". In: *Global and Planetary Change* 208. ISBN: 0921-8181 Publisher: Elsevier, p. 103704.

Cheng, H. et al. (2013). "Climate change patterns in Amazonia and biodiversity". In: *Nature communications* 4.1. ISBN: 2041-1723 Publisher: Nature Publishing Group UK London, p. 1411.

Condamine, F. L., J. Rolland, and H. Morlon (2013). "Macroevolutionary perspectives to environmental change". In: *Ecology letters* 16. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 72–85.

Delsuc, F., S. F. Vizcaíno, and E. J. Douzery (2004). "Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans". In: *BMC Evolutionary Biology* 4.1. ISBN: 1471-2148 Publisher: BioMed Central, pp. 1–13.

Feijo, A. and T. C. Anacleto (2021). "Taxonomic revision of the genus Cabassous McMurtrie, 1831 (Cingulata: Chlamyphoridae), with revalidation of Cabassous squamicaudis (Lund, 1845)". In: *Zootaxa* 4974.1. ISBN: 1175-5334, pp. 47–78–47–78.

Feijó, A., B. D. Patterson, and P. Cordeiro-Estrela (2018). "Taxonomic revision of the long-nosed armadillos, Genus Dasypus Linnaeus, 1758 (Mammalia, Cingulata)". In: *PLoS One* 13.4. Publisher: Public Library of Science San Francisco, CA USA, e0195084. ISSN: 1932-6203.

Feijó, A. et al. (2019). "Phylogeny and molecular species delimitation of long-nosed armadillos (Dasypus: Cingulata) supports morphology-based taxonomy". In: *Zoological Journal of the Linnean Society* 186.3. Publisher: Oxford University Press UK, pp. 813–825. ISSN: 0024-4082.

Garreaud, R. D. et al. (2009). "Present-day south american climate". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 281.3. ISBN: 0031-0182 Publisher: Elsevier, pp. 180–195.

Gibb, G. C. et al. (2016). "Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans". In: *Molecular Biology and Evolution* 33.3, pp. 621–642. ISSN: 0737-4038. DOI: 10.1093/molbev/msv250.

Hoorn, C., L. Palazzesi, and D. Silvestro (2022). *Editorial Preface to Special Issue: Exploring the impact of Andean uplift and climate on life evolution and landscape modification: From Amazonia to Patagonia*. Vol. 211. Pages: 103759 Publication Title: Global and Planetary Change. Elsevier. ISBN: 0921-8181.

Hoorn, C. et al. (2010). "Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity". In: *science* 330.6006. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 927–931.

Hoorn, C. et al. (2017). "The Amazon at sea: Onset and stages of the Amazon River from a marine record, with special reference to Neogene plant turnover in the drainage basin". In: *Global and Planetary Change* 153. ISBN: 0921-8181 Publisher: Elsevier, pp. 51–65.

Lyons, S. K., F. A. Smith, and J. H. Brown (2004). "Of mice, mastodons and men: human-mediated extinctions on four continents". In: *Evolutionary Ecology Research* 6.3. ISBN: 1522-0613 Publisher: Evolutionary Ecology, Ltd., pp. 339–358.

Mazet, N. et al. (2023). "Estimating clade-specific diversification rates and palaeodiversity dynamics from reconstructed phylogenies". In: *Methods in Ecology and Evolution* 14.10. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 2575–2591.

Meseguer, A. S. and F. L. Condamine (2017). "Ancient tropical extinctions contributed to the latitudinal diversity gradient". In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 236646.

Meseguer, A. S. et al. (2022). "Diversification dynamics in the Neotropics through time, clades, and biogeographic regions". In: *Elife* 11. ISBN: 2050-084X Publisher: eLife Sciences Publications Limited, e74503.

Miranda, F. R. et al. (2018). "Taxonomic review of the genus Cyclopes Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species". In: *Zoological Journal of the Linnean Society* 183.3, pp. 687–721. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlx079.

Miranda, F. R. et al. (2023). "Taxonomic revision of maned sloths, subgenus Bradypus (Scaeopus), Pilosa, Bradypodidae, with revalidation of Bradypus crinitus Gray, 1850". In: *Journal of Mammalogy* 104.1. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 86–103.

Montes, C. et al. (2019). "Continental margin response to multiple arc-continent collisions: The northern Andes-Caribbean margin". In: *Earth-Science Reviews* 198. ISBN: 0012-8252 Publisher: Elsevier, p. 102903.

Moraes-Barros, N. and M. C. Arteaga (2015). "Genetic diversity in Xenarthra and its relevance to patterns of neotropical biodiversity". In: *Journal of Mammalogy* 96.4, pp. 690–702. ISSN: 0022-2372. DOI: 10.1093/jmammal/gyv077.

Morlon, H., T. L. Parsons, and J. B. Plotkin (2011). "Reconciling molecular phylogenies with the fossil record". In: *Proceedings of the National Academy of Sciences* 108.39. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 16327–16332.

Morlon, H. et al. (2016). "RPANDA: an R package for macroevolutionary analyses on phylogenetic trees". In: *Methods in Ecology and Evolution* 7.5. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 589–597.

Rull, V. (2011). "Neotropical biodiversity: timing and potential drivers". In: *Trends in ecology & evolution* 26.10. ISBN: 0169-5347 Publisher: Elsevier, pp. 508–513.

Steadman, D. W. et al. (2005). "Asynchronous extinction of late Quaternary sloths on continents and islands". In: *Proceedings of the National Academy of Sciences* 102.33. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 11763–11768.

Val, P. et al. (2021). “Geological history and geodiversity of the Amazon”. In: *Amazon Assessment Report*.

# APPENDIX 2 - IS ISLAND SPECIES DIVERGENCE DRIVEN BY LOCAL ADAPTATION OR DEMOGRAPHY?

## Introduction

The Bocas del Toro archipelago, off the coast of Panama, is made up of several islands formed by rising sea levels during the Holocene (Anderson and Handley, 2001; Summers et al., 1997). A total of four isolation events from the mainland have been described, leading to the formation of islands over the last 10,000 years: i) Escudo de Veraguas Island, 8,900 years ago, ii) Colon Island, 5,200 years ago, iii) Cayo Nancy and Bastimentos Islands, 4,700 years ago, and iv) Cayo Agua Island, 3,400 years ago (Figure A2.1; Anderson and Handley, 2001; Summers et al., 1997). Different populations were isolated from mainland populations by vicariance, and they are characterized by major morphological changes with their mainland close relatives. Of the nine mammal species present on the oldest island, Escudo de Veraguas, there are four cases of island gigantism (*Artibeus incomitatus*, *Glossophaga soricina*, *Micronycteris megalotis*, *Hoplomys gymnurus*), three cases of island dwarfism (*Caluromys derhianus*, *Carollia brevicauda*, and *Bradypus variegatus*), and two species (*Saccopteryx leptura* and *Myotis riparius*) show no morphological distinction from the mainland (Kalko, 1994). Numerous cases of morphological changes in body size have been described for island species around the world, and thus, these convergent phenotypic changes have been included as part of the island syndrome (Adler and Levins, 1994).

Since Van Valen (1965) and Lomolino (1985), genetic changes leading to island dwarfism have been mainly explained by adaptive forces. It has been argued that this reduced body size corresponds to the adaptive optimum: a compromise between resource availability and predation

pressures (Lomolino, 1985). However, the founder effect and reduced effective population size of the insular population lead to a reduction in genetic diversity and an increase in genetic drift. These demographic factors reduce the effectiveness of selection on islands (Kimura, 1962). Indeed, in a work I performed during my Master 2 internship, we highlighted that reduced immune functions in island passerine species could be explained by increased genetic burden due to increased drift on islands (see Barthe et al., 2022 in Appendix 4).

Thus, if such morphological changes are adaptive, they should be characterized by a selection coefficient sufficiently high to counteract these demographic factors (Ohta, 1992). Faced with this paradox, Biddick and Burns (2021a) proposed an alternative model in which drift would be at the origin of body size changes on islands (see also Biddick and Burns, 2021b; Diniz-Filho et al., 2021). To our knowledge, only Prentice et al. (2017) have attempted to characterize the forces (adaptive and demographic) driving island dwarfism in lynxes. Unfortunately, their study is based on a single growth factor gene, so their results could be impacted by a very small number of SNPs.

The three-toed sloths from Escudo de Veraguas are significantly smaller than the mainland species, *Bradypus griseus* (previously *Bradypus variegatus*; mean size 50.5 cm (±1.1) vs. 60.8 cm (±1.6)). This convinced Anderson and Handley (2001) to elevate this insular population to the species taxonomic rank under the name *Bradypus pygmaeus*. However, our results obtained in Chapter 2 based on the whole genome of nine individuals of *Bradypus griseus* and two *Bradypus pygmaeus*, revealed a strong genetic similarity (Da = 0.0008; Dxy = 0.0018, as a comparison the two subspecies of *Dayspus septemcinctus septemcinctus* and *D. sept. hybridus* were more divergent with Da = 0.032 and Dxy = 0.0048; see Chapter 2). These molecular results thus suggest that these island and mainland populations may still belong to the same species, despite their morphological distinction as discussed in Chapter 2. Considering their low genomic divergence, morphological differentiation appears very fast and despite their restricted population size (actually estimated at about 500 individuals). A question remains though: what evolutionary force(s) led to this fast morphological differentiation?

To explore demographic and adaptive factors that may underlie the morphological differentiation of *Bradypus pygmaeus,* we gathered whole genome sequencing for eight *Bradypus griseus* individuals, the mainland species including islands close to the coast, and for two *Bradypus pygmaeus* individuals (Figure A2.1). We assessed the effect of drift through randomly selected genes by comparing Pn/Ps between insular and mainland populations. These genes are not expected to have been involved in adaptation and thus can serve as control. Then, genes linked with dwarfism, as listed in the platform GeneCard, have been used to compare Pn/Ps between insular and mainland populations. A higher Pn/Ps difference between mainland and island population for genes involved in dwarfism than control genes might be attributed to a change in selection pressure (Figure A2.3).


**Figure A2.1:** Location of the eight individuals of *B. griseus griseus* (in orange) and the two individuals of *B. griseus pygmaeus* (red) sampled.

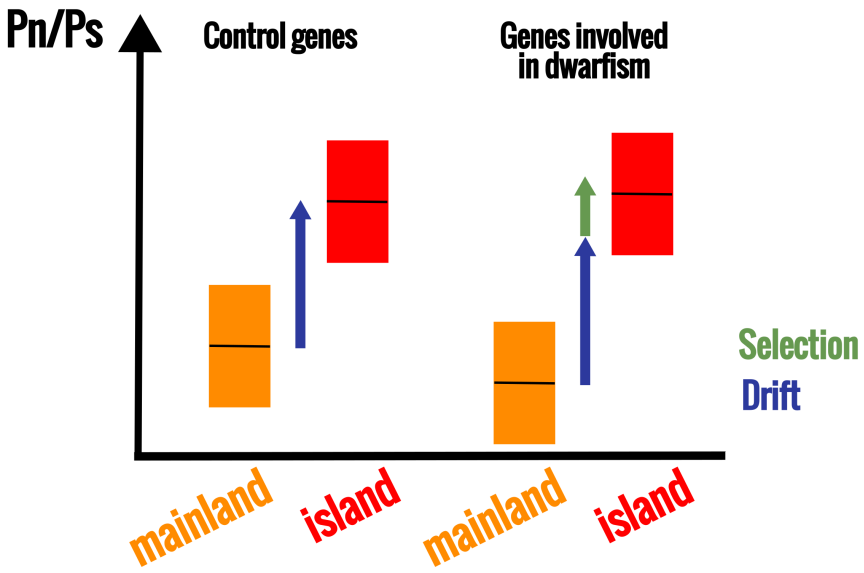


**Figure A2.2:** Conceptual diagram illustrating the expected effect of evolutionary forces (i.e. drift, blue arrow and selection, green arrow) on Pn/Ps of two populations with different effective population sizes, Ne (i.e. mainland, with large population size in orange, and island with reduced population size in red). Control genes randomly sampled in the genome are generally under purifying selection inducing a low Pn/Ps. For populations with a lower Ne, drift must reduce the efficacy of selection, inducing the fixation of weakly deleterious mutations increasing the Pn/Ps. Thus, the difference of Pn/Ps between mainland and island populations would allow us to quantify the effect of drift. Then, if genes involved in dwarfism underwent positive selection, beneficial mutations will contribute to increase Pn/Ps. Thus, a higher Pn/Ps difference between mainland and island populations for genes involved in dwarfism will evidence positive selection.

# Materials & Methods

## Dataset assembly

The raw Illumina reads of 16 individuals of *Bradypus griseus*, including two *Bradypus griseus pygmaeus*, analyzed in Chapter 2 were reemployed for this study. Raw reads obtained by 150PE Illumina sequencing were cleaned using FastP v0.21.0 (Chen et al., 2018). The genome of *Bradypus tridactylus* V3450 served as a reference to map reads of each individual using bwa mem v.0.7.17 (Li, 2013) with default parameters. Mapping files were converted, then reads were ordered according to their position on the reference genome, and finally mapping files were indexed using Samtools v1.9 (Li et al., 2009) and Picard v2.25.5 (Toolkit, 2019). Duplicate reads were marked with MarkDuplicates v2.25.5 (Toolkit, 2019). Variants were called using Freebayes v1.3.1 (Garrison and Marth, 2012) with adequate options for diploid data. Depth of coverage was estimated with Samtools depth v1.9 (Li et al., 2009) by setting parameters to a mapping quality threshold of 30, a base quality threshold of 30, and the option "-a". Diploid fasta sequences were generated using VCF2FastaFreebayes and positions

with less than 6x depth of coverage were masked, as well as positions with a quality less than300 and heterozygous positions with a read frequency of the minor allele less than 0.3 (deviating from the 0.5 expectation). Six individuals with less than 6x depth of coverage were excluded (KU190, KU317, BVA850, BVA525, KU357, BVA166).

## Extracting dwarfism and random genes

Genes potentially involved in dwarfism were identified by screening the GeneCard database with the search terms "short stature" OR "growth impairment" OR "height" OR "dwarfism" OR "dwarf" OR "growth restriction" OR "growth retardation" (search performed on the 31/08/2023). The 1000 genes with the highest functionality score were then selected. The exon coordinates of these genes were extracted from the annotation file of the reference genome *Bradypus tridactylus* V3450. In addition, the exon coordinates of 1000 randomly sampled genes were also extracted from the annotation file of the reference genome *Bradypus tridactylus* V3450 to serve as control. These coordinates enabled the extraction of the corresponding regions in the 11 mapped individuals using a custom script. For each gene, sequences were aligned, cleaned, based reading frames with the OMM_Macse V12.01 pipeline (Ranwez et al., 2021) using the option "no_prefiltering". Finally, stop codons were excluded. Respectively, 81 control genes and 54 dwarfism genes including a heterozygous position shared by at least 75% of individuals were excluded as this deviation from Hardy-Weiberg equilibrium can reflect the mapping of paralogs.

## Pn/Ps estimation and statistical analysis

We estimated synonymous (Ps) and non-synonymous (Pn) nucleotide diversities using the script seq_stat_coding based on the Bio++ library (Guéguen et al., 2013). The mean Pn/Ps was calculated as the sum of Pn over the sum of Ps (Wolf et al., 2009) for the concatenation of 60 genes, and replicated 100 times. The Pn/Ps ratio was used as the dependent variable of a linear mixed model with the mainland or insular origin and the category of genes (control or involved in dwarfism) as explanatory variables. This linear model analysis was conducted with packages lme4 and lmerTest (Bates et al., 2012; Kuznetsova et al., 2017). We evaluated five linear mixed models: i) model including origin and gene category parameters and also the interaction effect, ii) model using both origin and gene category parameters, iii) model with only the gene category parameter, iv) model with only the origin parameter, and finally v) null model. To select the best fitting model we used two distinct approaches: i) the difference in corrected Akaike Information Criterion (AICc) estimated using the qpcR package (Spiess and Spiess, 2018), and ii) an ANOVA test between models following a model simplification approach.

# Results & Discussion

First, we evaluated the effect of the number of individuals on Pn/Ps estimations, considering the distinct number of individuals constituting these two subspecies. Based on control genes of the eight mainland individuals of *B. griseus*, we detected bias on the mean Pn/Ps estimated from a reduced number of alleles (Figure A2.3). A similar pattern has been identified in Gayral et al. (2013), but the reason was unclear. Unfortunately, as only two individuals of insular *B. griseus pygmaeus* have been sequenced, we randomly sampled four alleles of mainland individuals of *B. griseus* to perform comparisons. As expected, the insular population had a lower genetic diversity than the mainland population (Ps(island) = 0.001, Ps(mainland) = 0.002; Figure A2.4). The Pn/Ps estimated from control genes showed higher values for the mainland (mean Pn/Ps = 0.33) than the island population (mean Pn/Ps = 0.29; Figure A2.4). This surprising result is the opposite of our expectations as higher drift on the island should increase the load of weakly deleterious mutations (Barthe et al., 2022; Leroy et al., 2021). Similarly for genes involved in dwarfism, the Pn/Ps of the mainland population was higher than the island population (mean Pn/Ps = $0.22 > 0.10$). Model selection based on AICc as well as on model simplification with ANOVA identified the model with interaction (Pn/Ps ~1+ category +origin+category *origin) as the best fitting model (Table A2.1). This model includes an effect of the gene category with dwarfism associated with a lower Pn/Ps of -0.13 than control genes ($p < 0.001$). In addition, island populations were associated with a reduction in Pn/Ps of -0.05 ($p < 0.001$), however this effect is the opposite of our expectation. This could be due to sampling bias as this analysis only relies on two individuals per population. Finally, an additional reduction was attributed to genes involved in dwarfism on the island population (estimate = -0.04, $p < 0.001$). Considering the uncertainty of these results, we plan on sequencing five additional individuals from the *B. griseus pygmaeus* island population to explore the effect of sampling bias on Pn/Ps more accurately. Furthermore, evaluating adaptation through the McDonald–Kreitman test could provide complementary information by taking into account divergence. Finally, performing genome scans would also contribute to identify highly differentiated genetic regions between the mainland and island populations in order to further understand what drove their morphological divergence.

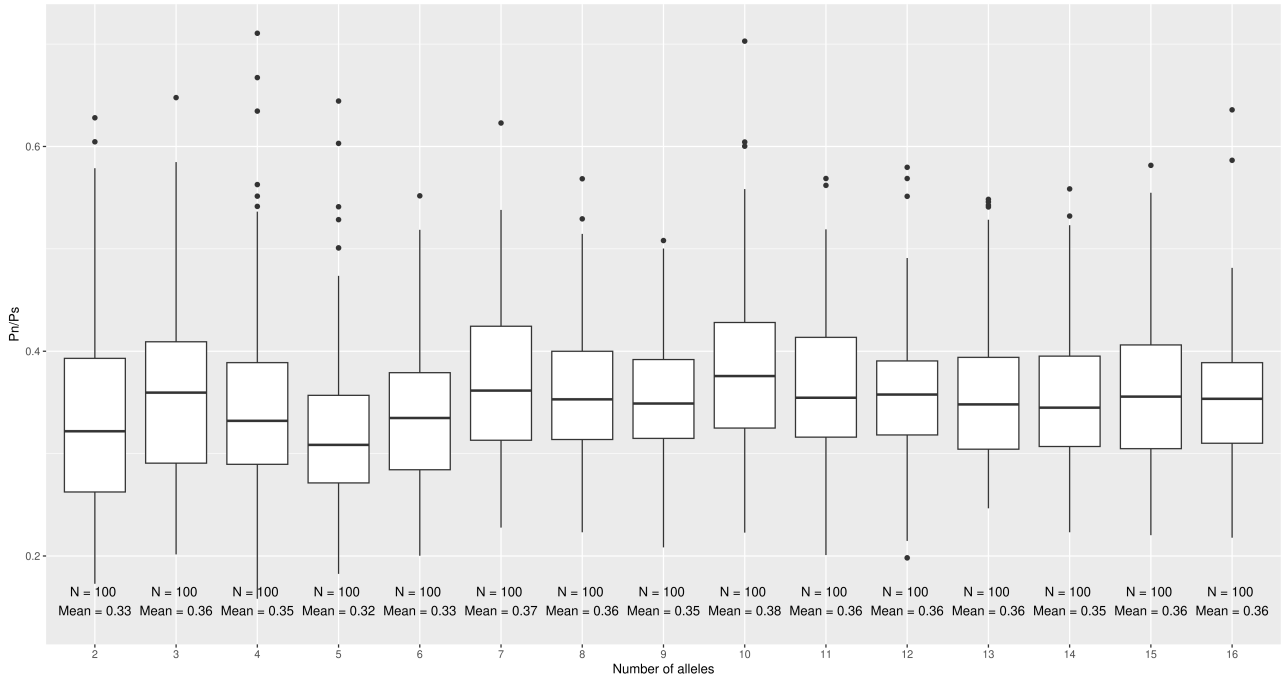


**Figure A2.3:** Effect of the number of alleles on Pn/Ps estimations. Each point corresponds to the Pn/Ps value estimated from the concatenation of 60 control genes based on randomly sampled alleles. For each category of allele, 100 replicates (N= 100) have been performed and the mean Pn/Ps was estimated.

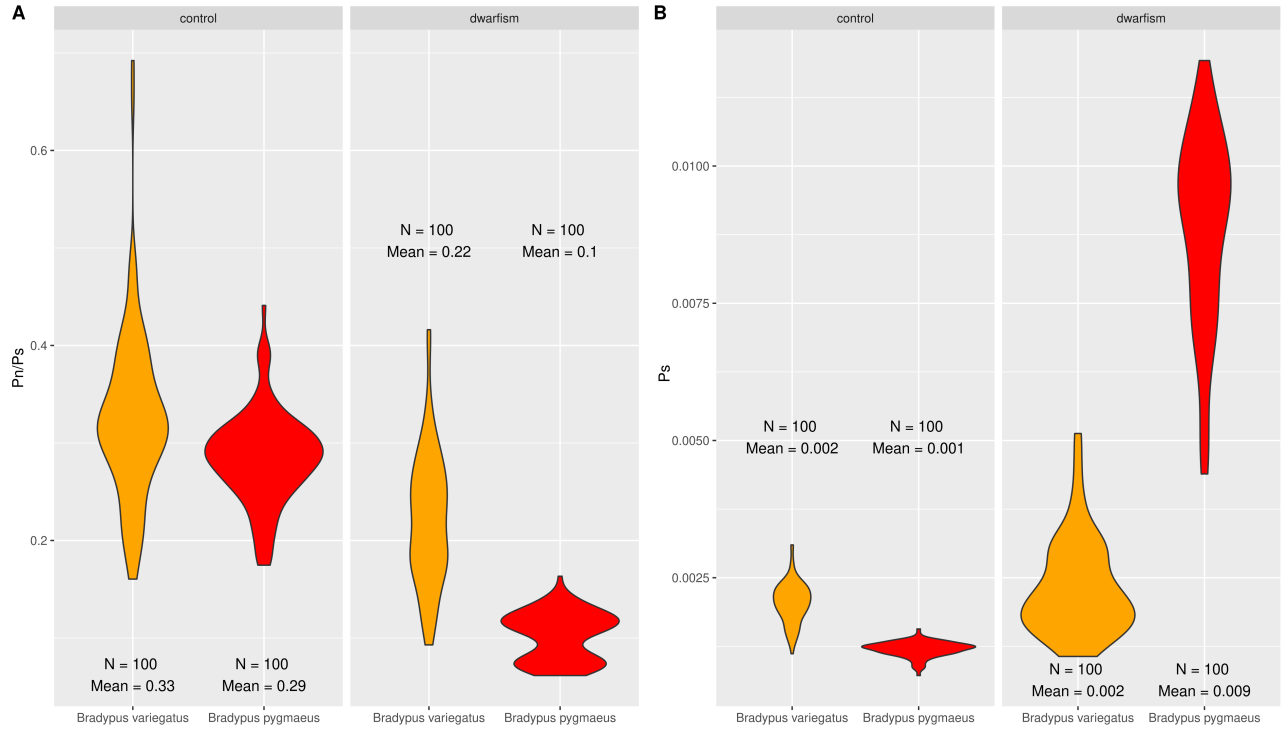


**Figure A2.4:** A) Pn/Ps and A) Ps of the mainland population (B. griseus) in orange and island population (B. griseus pygmaeus) in red for control genes (left panel) and genes involved in dwarfism (right panel). The distribution of Pn/Ps is based on 100 replicates (N=100) each based on the concatenation of 60 genes.

**Table A2.1:** Statistical model explaining the Pn/Ps variation. For each model, this table details the corrected Akaike information criterion (AICc), the difference with the lower AICc (ΔAICc), the estimated Likelihood, and ANOVA test. For the ANOVA test only p-values of comparisons with model n°1 are presented as the more complex model (n°1) explains a larger proportion of the variance.

| **Models** | | **Model selection by AIC** | | | A**NOVA test** |
|---|---|---|---|---|---|
| n° | Details | AICc | ΔAICc | Likelihood | n°1 |
| 1 | Pn/Ps~1+ category +origin+ category *origin | -1,187.28 | 0.00 | 1.00E+00 | |
| 2 | Pn/Ps~1+ category +origin | -1,168.67 | 18.61 | 9.08E-05 | 6.21E-06 |
| 3 | Pn/Ps~1+ category | -1,014.47 | 172.81 | 2.98E-38 | 2.20E-16 |
| 4 | Pn/Ps~1+ origin | -721.59 | 465.70 | 7.51E-102 | 2.20E-16 |
| 5 | Pn/Ps~1 | -665.80 | 521.48 | 5.78E-114 | |

**Table A2.2:** Summary of model n°1, best statistical model selected using AICc and ANOVA test. * indicates significant values : * $< 0.05$; ** $< 0.01$; *** $< 0.001$

| **Model** | **Parameters** | | | | | |
|---|---|---|---|---|---|---|
| | | Origine | Category | Estimate | p.value | |
| Pn/Ps ~1 + category + origin + category*origin | Intercept | Mainland | Control genes | 0.34 | <2E-016 | *** |
| | | Island | | -0.05 | 5.55E-11 | *** |
| | | | Dwarfism | -0.14 | <2E-016 | *** |
| | | Island | Dwarfism | -0.05 | 6.21E-06 | *** |

# References


Adler, G. H. and R. Levins (1994). "The island syndrome in rodent populations". In: *The Quarterly review of biology* 69.4. ISBN: 0033-5770 Publisher: University of Chicago Press, pp. 473–490.

Anderson, R. P. and C. O. Handley (2001). "A new species of three-toed sloth (Mammalia: Xenarthra) from Panama, with a review of the genus Bradypus". In: *Proceedings-Biological Society of Washington* 114.1. Publisher: BIOLOGICAL SOCIETY OF WASHINGTON, pp. 1–33. ISSN: 0006-324X.

Barthe, M. et al. (2022). "Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome". In: *Peer Community Journal* 2. ISBN: 2804-3871.

Bates, D. M. et al. (2012). "Package 'lme4'". In: *CRAN R Found. Stat. Comput.*

Biddick, M. and K. C. Burns (2021a). "A simple null model predicts the island rule". In: *Ecology Letters* 24.8. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 1646–1654.

Biddick, M. and K. C. Burns (2021b). "Minimal models provide maximally parsimonious explanations". In: *Ecology Letters* 24.11. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 2524–2525.

Chen, S. et al. (2018). "fastp: an ultra-fast all-in-one FASTQ preprocessor". In: *Bioinformatics* 34:i884–i890. DOI: https://doi.org/10.1093/bioinformatics/bty560.

Diniz-Filho, J. A. F. et al. (2021). "Too simple models may predict the island rule for the wrong reasons". In: *Ecology Letters* 24.11. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 2521–2523.

Garrison, E. and G. Marth (2012). "Haplotype-based variant detection from short-read sequencing". In: *arXiv preprint arXiv:1207.3907*.

Gayral, P. et al. (2013). "Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap". In: *PLoS genetics* 9.4. ISBN: 1553-7390 Publisher: Public Library of Science San Francisco, USA, e1003457.

Guéguen, L. et al. (2013). "Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution". In: *Molecular Biology and Evolution* 30.8, pp. 1745–1750. ISSN: 0737-4038, 1537-1719. DOI: https://doi.org/10.1093/molbev/mst097.

Kalko, E. K. (1994). "Evolution, biogeography, and description of a new species of fruit-eating bat, genus Artibeus Leach (1821), from Panamá". In: *Z. Säugetierkd.* 59, p. 257.

Kimura, M. (1962). "On the probability of fixation of mutant genes in a population". In: *Genetics* 47.6. Publisher: Oxford University Press, p. 713.

Kuznetsova, A., P. Brockhoff, and R. Christensen (2017). "lmerTest package: tests in linear mixed effects models". In: *Journal of statistical software* 82, pp. 1–26. DOI: https://doi.org/10.18637/jss.v082.i13.

Leroy et al. (2021). "Island songbirds as windows into evolution in small populations". In: *Current Biology* 31.6. ISBN: 0960-9822 Publisher: Elsevier, 1303–1310. e4. DOI: https://doi.org/10.1016/j.cub.2020.12.040.

Li, H. (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv:1303.3997*.

Li, H. et al. (2009). "The sequence alignment/map format and SAMtools". In: *Bioinformatics* 25.16. ISBN: 1367-4803 Publisher: Oxford University Press, pp. 2078–2079.

Lomolino, M. V. (1985). "Body size of mammals on islands: the island rule reexamined". In: *The American Naturalist* 125.2. ISBN: 0003-0147 Publisher: University of Chicago Press, pp. 310–316.

Ohta, T. (1992). "The nearly neutral theory of molecular evolution". In: *Annual review of ecology and systematics* 23.1. ISBN: 0066-4162 Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 263–286.

Prentice, M. B. et al. (2017). "Selection and drift influence genetic differentiation of insular Canada lynx (Lynx canadensis) on Newfoundland and Cape Breton Island". In: *Ecology and Evolution* 7.9. ISBN: 2045-7758 Publisher: Wiley Online Library, pp. 3281–3294.

Ranwez, V., N. Chantret, and F. Delsuc (2021). "Aligning protein-coding nucleotide sequences with MACSE". In: *Multiple Sequence Alignment: Methods and Protocols*. ISBN: 107161035X Publisher: Springer, pp. 51–70.

Spiess, A.-N. and M.-N. Spiess (2018). "Package 'qpcR". In: *Model. Anal. Real-Time PCRdata Httpscran R-Proj*. OrgwebpackagesqpcRqpcR Pdf.

Summers, K. et al. (1997). "Phenotypic and genetic divergence in three species of dart-poison frogs with contrasting parental behavior". In: *Journal of Heredity* 88.1. ISBN: 1465-7333 Publisher: Oxford University Press, pp. 8–13.

Toolkit, P. (2019). *Broad institute, GitHub repository*. https://broadinstitute.github.io/picard/.

Van Valen, L. (1965). "Morphological variation and width of ecological niche". In: *The American Naturalist* 99.908. ISBN: 0003-0147 Publisher: Science Press, pp. 377–390.

Wolf, J. et al. (2009). "Nonlinear Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects Inference of Selection". In: *Genome Biol. Evol* 1, pp. 308–319. DOI: https://doi.org/10.1093/gbe/evp030.

# APPENDIX 3 - IS GENETIC DIFFERENTIATION BETWEEN SPECIES HOMOGENEOUS ACROSS MAMMALS?

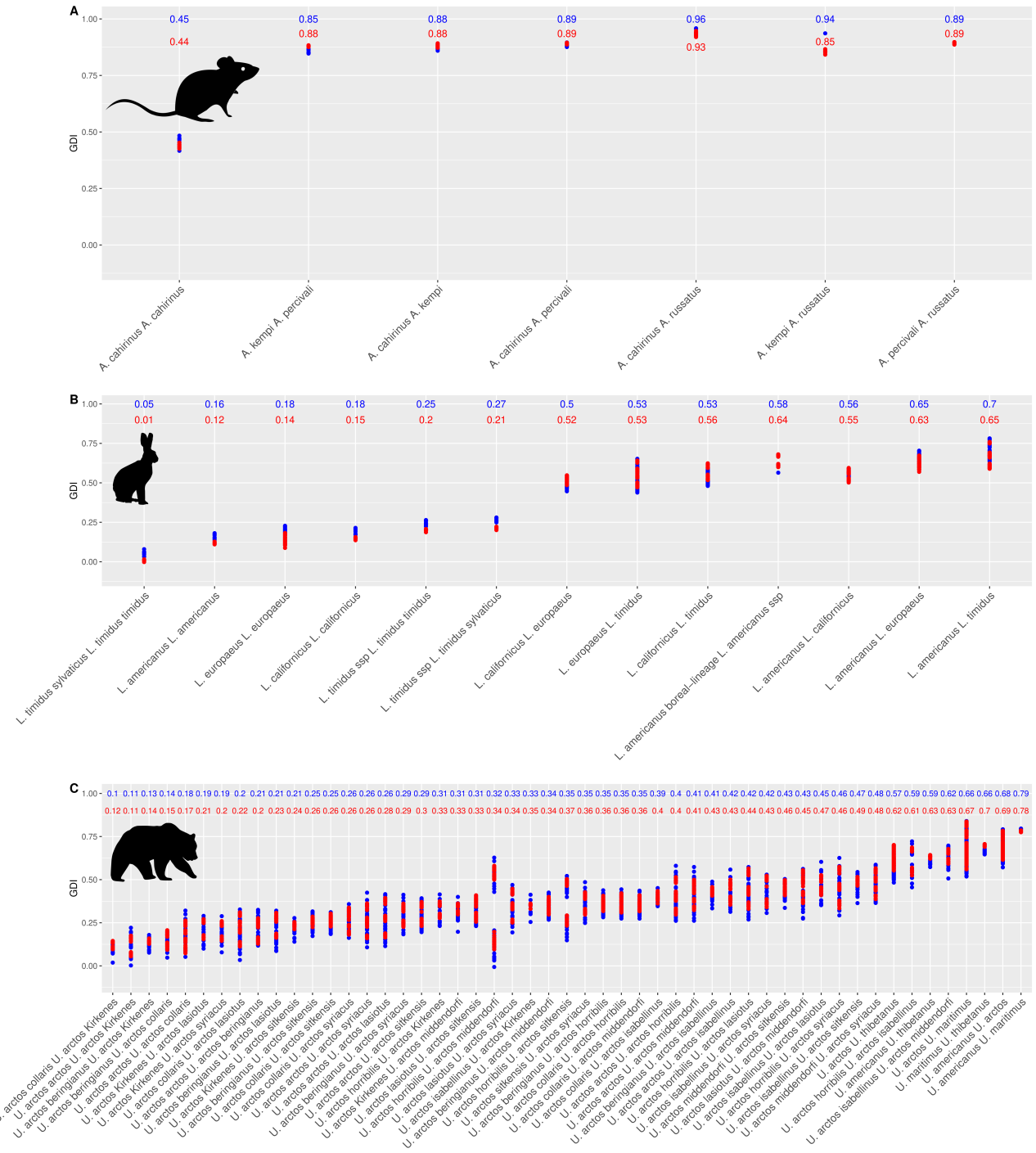


**Figure A3.1:** Pairwise genetic differentiation between mammalian species from eight genera estimated with the Genetic Differentiation Index (GDI; Allio et al., 2021). Blue points represent the mean GDI estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. Red points represent the mean GDI estimated from 10 replicates of 50 BUSCO genes (~100kb). The mean GDI value is indicated above each graph in the corresponding color. The higher the GDI, the greater the genetic differentiation. Each panel corresponds to a genus: A) *Acomys*, B) *Lepus*, C) *Ursus*, D) *Balaenoptera*, E) *Manis*, F) *Homo*, G) *Gorilla*, and H) *Ceratotherium*. Silhouettes are from phylopic.org.

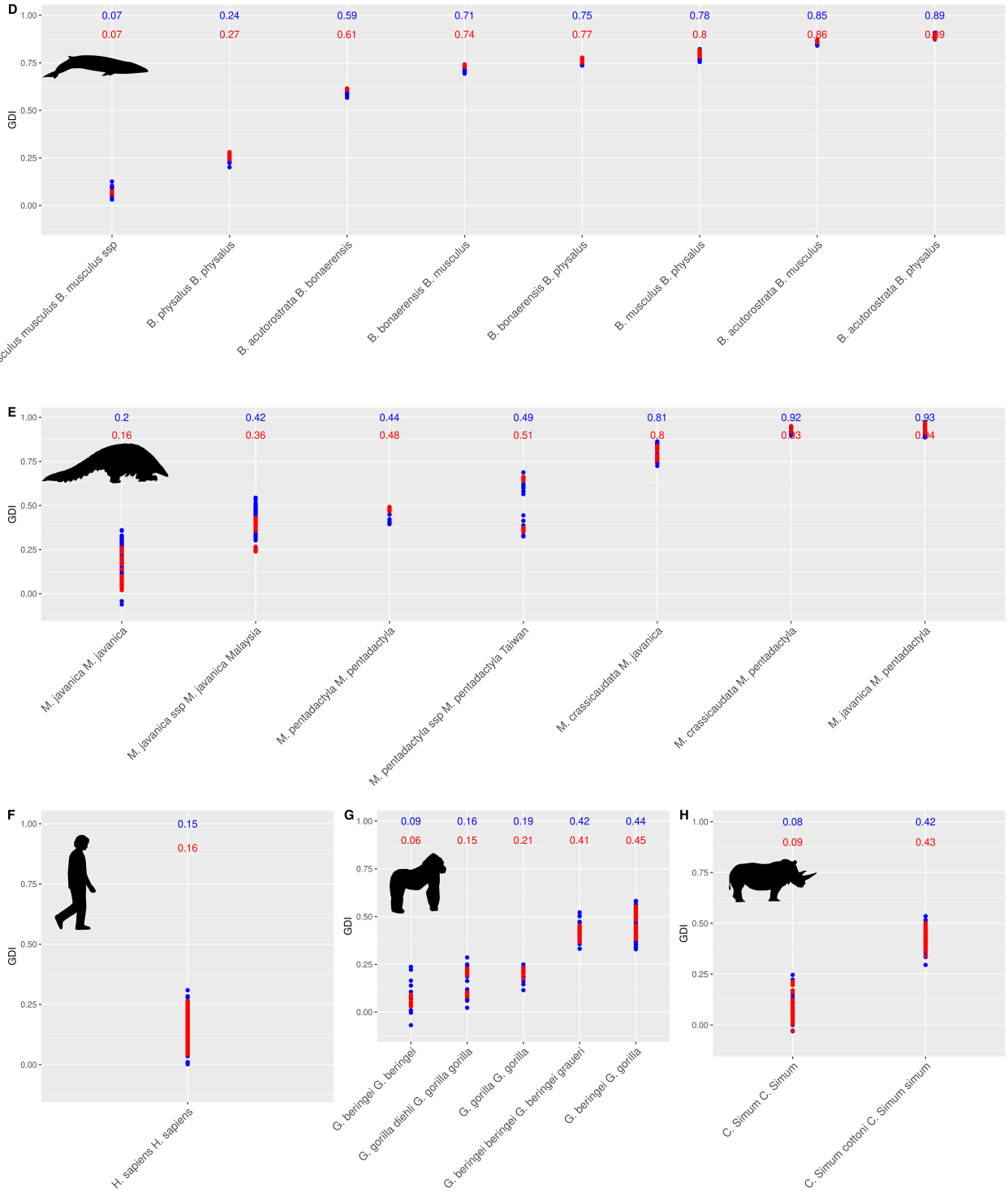


**Figure A3.1:** Continued.

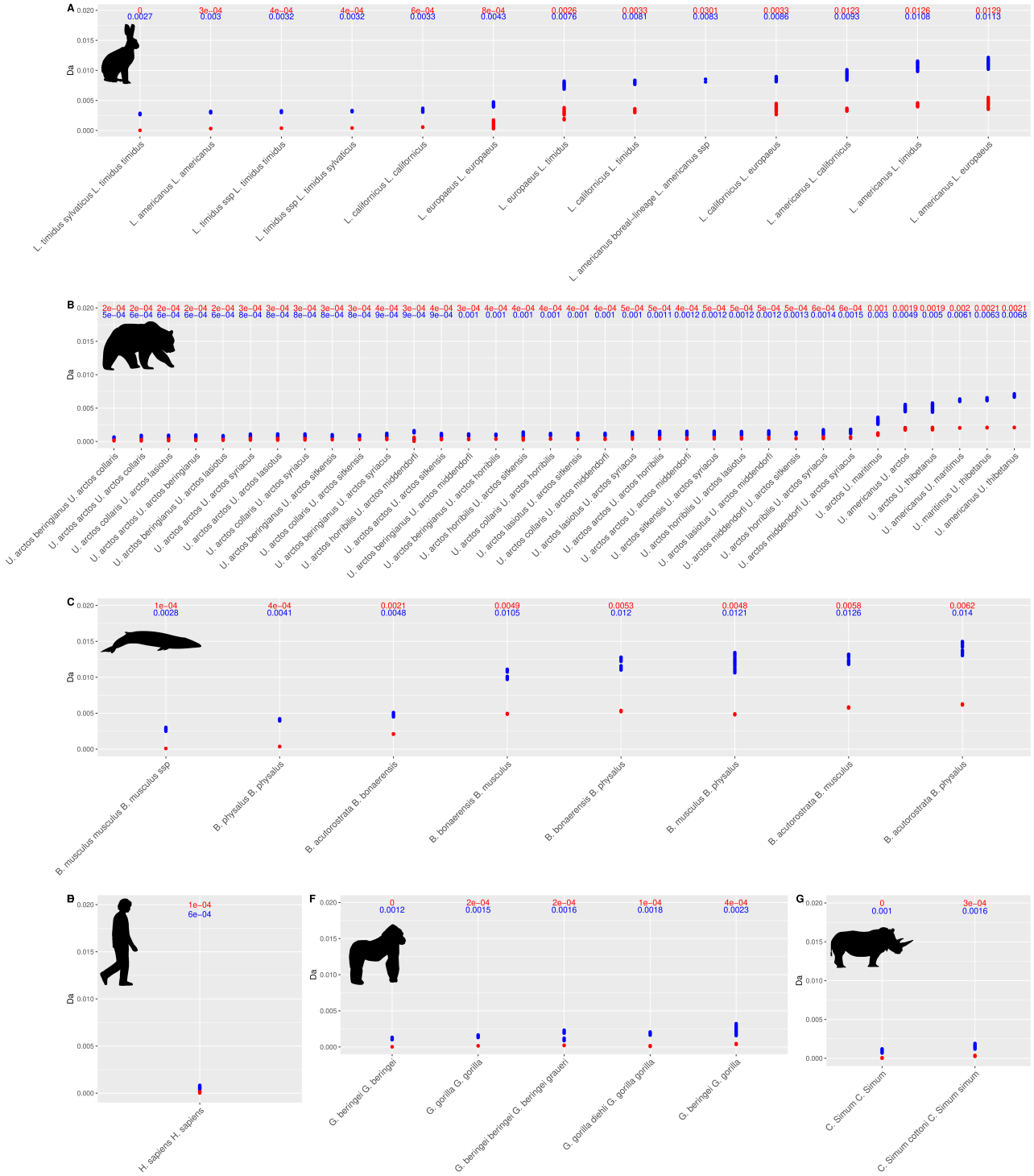


**Figure A3.2:** Pairwise genetic differentiation between mammalian species from six genera estimated with the net synonymous divergence (Da). Blue points represent the mean Da estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. Red points represent the mean Da estimated from 10 replicates of 50 BUSCO genes (~100kb). The mean Da value is indicated above each graph in the corresponding color. Each panel corresponds to a genus: A) *Lepus*, B) *Ursus*, C) *Balaenoptera*, D) *Homo*, E) *Gorilla*, and F) *Ceratotherium*. Silhouettes are from phylopic.org.

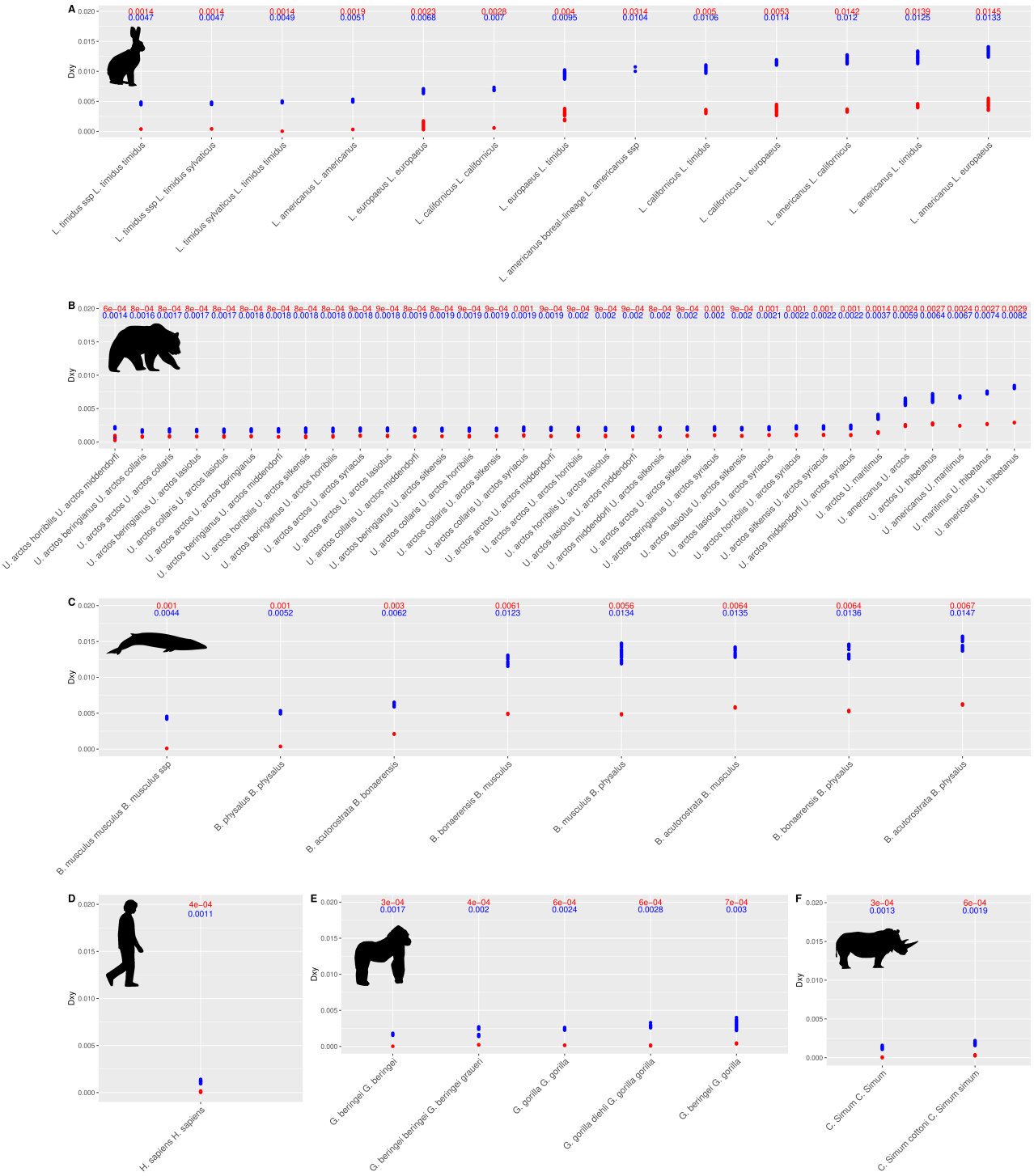


**Figure A3.3:** Pairwise genetic divergence (Dxy) between mammalian species from six genera. Blue points represent the mean Dxy estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals. Red points represent the mean Dxy estimated from 10 replicates of 50 BUSCO genes (~100kb). The mean Dxy value is indicated above each graph in the corresponding color. Each panel corresponds to a genus: A) *Lepus*, B) *Ursus*, C) *Balaenoptera*, D) *Homo*, E) *Gorilla*, and F) *Ceratotherium*. Silhouettes are from phylopic.org.

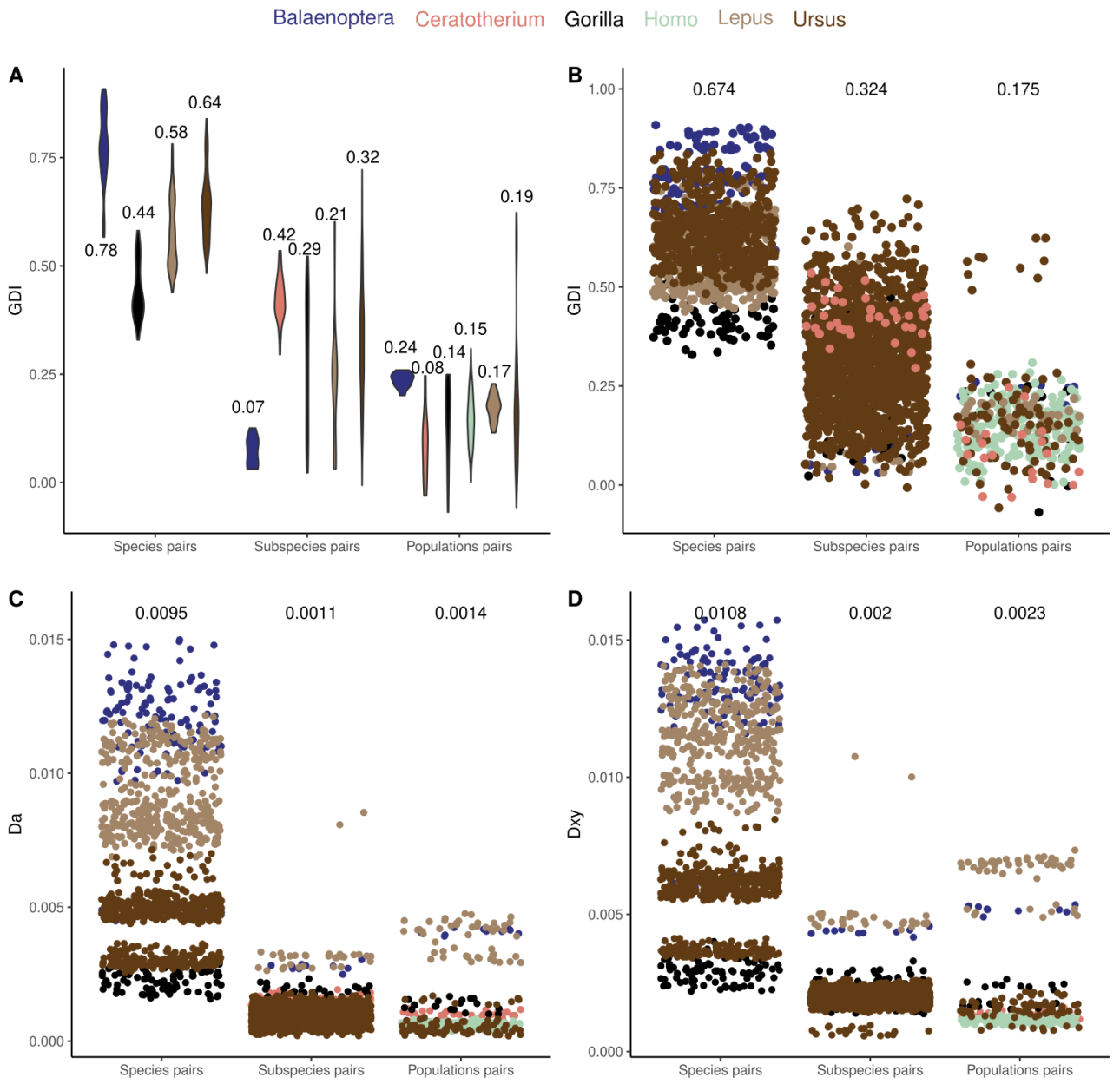


**Figure A3.4:** Distribution of the genomic differentiation of randomly sampled genetic regions between pairs of mammal individuals belonging to eight genera according to their taxonomic status. A) Distribution of GDI for each genus. B, C, D) Points represent the mean GDI (B), Da (C), and Dxy (D) as estimated from 10 regions of 100kb randomly sampled in the genome of a pair of individuals.

# References

Allio, R. et al. (2021). "High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox". In: *Elife* 10. Publisher: eLife Sciences Publications Limited, e63167. ISSN: 2050-084X.

# APPENDIX 4 - EVOLUTION OF IMMUNE GENES IN ISLAND BIRDS: REDUCTION IN POPULATION SIZES CAN EXPLAIN ISLAND SYNDROME

Peer Community Journal

Section: Evolutionary Biology

RESEARCH ARTICLE

**Published**
2022-11-07

**Cite as**
Mathilde Barthe, Claire Doutrelant, Rita Covas, Martim Melo, Juan Carlos Illera, Marie-Ka Tilak, Constance Colombier, Thibault Leroy, Claire Loiseau and Benoit Nabholz (2022) *Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome*, Peer Community Journal, 2: e67.

**Correspondence**
mathilde.barthe.pro@gmail.com

**Peer-review**
Peer reviewed and recommended by PCI Evolutionary Biology, https://doi.org/10.24072/pci.evolbiol.100153



# Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome

Mathilde Barthe[1], Claire Doutrelant[2], Rita Covas[3,4], Martim Melo[3,4,5,6], Juan Carlos Illera[7], Marie-Ka Tilak[1], Constance Colombier[1], Thibault Leroy[1,8], Claire Loiseau[2,3,5], and Benoit Nabholz[1,9]

Volume **2** (2022), article e67

https://doi.org/10.24072/pcjournal.186

## Abstract

Shared ecological conditions encountered by species that colonize islands often lead to the evolution of convergent phenotypes, commonly referred to as “island syndrome”. Reduced immune functions have been previously proposed to be part of this syndrome, as a consequence of the reduced diversity of pathogens on island ecosystems. According to this hypothesis, immune genes are expected to exhibit genomic signatures of relaxed selection pressure in island species. In this study, we used comparative genomic methods to study immune genes in island species (N = 20) and their mainland relatives (N = 14). We gathered public data as well as generated new data on innate (TLR: Toll-Like Receptors, BD: Beta Defensins) and acquired immune genes (MHC: Major Histocompatibility Complex classes I and II), but also on hundreds of genes with various immune functions. As a control, we used a set of 97 genes, not known to be involved in immune functions based on the literature, to account for the increased drift effects of the lower effective population sizes in island species. We used synonymous and non-synonymous variants to estimate the selection pressure acting on immune genes. We found that BDs and TLRs have higher ratios of non-synonymous over synonymous polymorphisms (Pn/Ps) than randomly selected control genes, suggesting that they evolve under a different selection regime. However, simulations show that this is unlikely to be explained by ongoing positive selection or balancing selection. For the MHC genes, which evolve under balancing selection, we used simulations to estimate the impact of population size variation. We found a significant effect of drift on immune genes of island species leading to a reduction in genetic diversity and efficacy of selection. However, the intensity of relaxed selection was not significantly different from control genes, except for MHC class II genes. These genes exhibit a significantly higher level of non-synonymous loss of polymorphism than expected assuming only drift and evolution under frequency dependent selection, possibly due to a reduction of extracellular parasite communities on islands. Overall, our results showed that demographic effects lead to a decrease in the immune functions of island species, but the relaxed selection that is expected to be caused by a reduced parasite pressure may only occur in some categories of immune genes.

[1]ISEM, Univ Montpellier, CNRS, IRD – Montpellier, France, [2]CEFE, CNRS, Univ Montpellier, EPHE, IRD – Montpellier, France, [3]CIBIO-InBio, Research Center in Biodiversity and Genetic Resources, Associated Laboratory, Campus Agrário de Vairão – Vairão, Portugal, [4]FitzPatrick Institute, University of Cape Town – Rondebosch, South Africa, [5]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 – Vairão, Portugal, [6]MHNC-UP, Natural History and Science Museum of the University of Porto – Porto, Portugal, [7]Biodiversity Research Institute (CSIC-Oviedo University-Principality of Asturias), Oviedo University – Mieres, Spain, [8]IRHS-UMR1345, Université d’Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 – Beaucouzé, France, [9]Institut universitaire de France, Paris

## Introduction

Island colonizers face new communities of competitors, predators and parasites in a small area with limited resources, which generally results in high extinction rates of colonizers (Losos and Ricklefs, 2009). Oceanic island faunas are characterized by a low species richness, coupled with high population densities for each species (MacArthur and Wilson, 1967; Warren et al., 2015) - which translates into communities with, on average, lower levels of inter-specific interactions and higher levels of intra-specific competition (but see Rando et al., 2010 for an example of character displacement due to competition among island finch species). These shared island characteristics are thought to underlie the evolution of convergent phenotypes, in what is called the 'island syndrome' (Baeckens and Van Damme, 2020). Convergence has been documented in multiple traits, such as size modification (dwarfism or gigantism; Lomolino, 2005), reduction of dispersal (Baeckens and Van Damme, 2020), shift towards K life-history strategies (Boyce, 1984; Covas, 2012; MacArthur and Wilson, 1967), evolution of generalist traits (Blondel, 2000; Warren et al., 2015), or changes in color and acoustic signals (Doutrelant et al., 2016; Grant, 1965).

Reduced immune function has also been hypothesized to be an island syndrome trait, directly linked to reduced parasite pressure on islands (Lobato et al., 2017; Matson and Beadell, 2010; Wikelski et al., 2004). Island parasite communities are i) less diverse (Beadell et al., 2006; Illera et al., 2015; Loiseau et al., 2017; Maria et al., 2009; Pérez-Rodríguez et al., 2013), and ii) could be less virulent due to the expansion of the ecological niche expected by the theory of island biogeography. In fact, island parasites are often more generalist than their mainland counterparts, which could lead to a reduced virulence due to the trade-off between replication capacity and resistance against host immune defenses (Garamszegi, 2006; Hochberg and Møller, 2001; Pérez-Rodríguez et al., 2013). Overall, a reduction of parasitic pressure should lead to a weakening of the immune system due to the costs of maintaining efficient immune functions (Lindström et al., 2004; Matson and Beadell, 2010; Wikelski et al., 2004). Such reduction may have important implications for the ability of these populations to resist or tolerate novel pathogens. The introduction of avian malaria in the Hawaiian archipelago, and the subsequent extinctions and population declines of many endemic species is the most emblematic example (Van Riper III et al., 1986; Wikelski et al., 2004).

Immunological parameters, such as blood leukocyte concentration, antibodies or other immune proteins (e.g. haptoglobin), hemolysis, and hemagglutination (Lee et al., 2006; Matson and Beadell, 2010) may serve as proxies to determine population immune functions. To date, the majority of studies that focused on island avifauna have found ambiguous results, with either no support for a reduced immune response on island species (Beadell et al., 2007; Matson, 2006), or contrasting results, such as a lower humoral component (total immunoglobulins) on islands, but a similar innate component (haptoglobin levels) between island and mainland species (Lobato et al., 2017). The use of immune parameters as proxies of immune function is fraught with difficulties (Lobato et al., 2017). The study of molecular evolution of immune genes therefore represents an alternative strategy to tackle this question. However, it is necessary to distinguish neutral effects (i.e. the demographic effects resulting from island colonization) from selective ones, the potential relaxation of selection pressures due to the changes in the pathogen community.

The bottleneck experienced by species during island colonization leads to a decrease in genetic variability (Frankham, 1997). A reduced genetic diversity at loci involved in immunity should have a direct implication on immune functions (Hale and Briskie, 2007 but see ; Hawley et al., 2005; Spurgin et al., 2011). Also, small population sizes increase genetic drift, which may counteract the effect of natural selection on weakly deleterious mutations (Ohta, 1992). Several recent studies found a greater load of deleterious mutations in island species (Kutschera et al., 2020; Leroy et al., 2021b; Loire et al., 2013; Robinson et al., 2016; Rogers and Slatkin, 2017). Finally, it is necessary to differentiate genes involved in the innate versus the acquired immune response. The innate immune response is the first line of defense and is composed of phagocytes, macrophages and dendritic cells. These cells allow non-specific recognition of pathogens (Akira, 2003; Alberts et al., 2002). For example, Toll-Like Receptors (TLR; transmembrane proteins) trigger a chain reaction leading to the production of various substances, including antimicrobial peptides such as beta-defensins (BD) that have active properties in pathogen cell lysis (Velová et al., 2018). On the other hand, the acquired immune system allows a specific response,

characterized by immune memory. Major Histocompatibility Complex (MHC) genes code for surface glycoproteins that bind to antigenic peptides, and present them to the cells of the immune system; class I and II genes ensure the presentation of a broad spectrum of intra- and extracellular-derived peptides, respectively (Klein, 1986). Although all these genes are directly involved in the identification and neutralization of pathogens, previous studies found that they evolve under different selection regimes: TLRs and BDs are under purifying selection which usually results in the selective removal of deleterious alleles and stabilizing selection (Grueber et al., 2014; van Dijk et al., 2008), whereas MHC genes are under balancing selection (Bernatchez and Landry, 2003).

Recent studies on birds (Gonzalez-Quevedo et al., 2015a, 2015b), amphibians (Belasen et al., 2019), and lizards (Santonastaso et al., 2017) found that the demographic history of island populations led to the loss of genetic variants at immune genes involved in pathogen recognition, such as TLRs and MHC. For example, Santonastaso et al., (2017) revealed that the polymorphism pattern in MHC genes and microsatellites covary positively with island area in *Podarcis* lizards, suggesting a dominant role for genetic drift in driving the evolution of the MHC. Gonzalez-Quevedo, et al. (2015a) found a similar pattern comparing TLR and microsatellite polymorphism in the Berthelot pipit, *Anthus berthelotii*, an endemic species from Macaronesia, supporting a predominant role of genetic drift in TLR evolution. However, these studies did not explicitly test the hypothesis of a relaxed selection pressure on islands imposed by an impoverished parasite community. All other things being equal, it is expected that the polymorphism of a coding sequence decreases with population size (Buffalo, 2021; Leroy et al., 2021b). Therefore, a decrease in polymorphism with population size could not be taken as a proof of a relaxation in the selection pressure.

To be able to demonstrate a change in natural selection, a traditional approach is to contrast polymorphism of synonymous sites (Ps) with polymorphism of non-synonymous sites (Pn). Synonymous mutations do not change amino acid sequences, whereas non-synonymous mutations do. Thus, synonymous mutations are expected to be neutral while non-synonymous could be subject to selection.

Following population genetic theory, in a diploid population, Ps = 4 $Ne\ \mu$ and Pn = 4 $Ne\ \mu\ f$, where $Ne$ is the effective population size, $\mu$ is the mutation rate and $f$ is a function that integrates the probability of an allele to segregate at a given frequency. $f$ depends on the distribution of the fitness effect (DFE) of mutations (Eyre-Walker and Keightley, 2007). This distribution scales with $Ne$ as the fitness effect is dependent on $Ne$ multiplied by the coefficient of selection $s$ (Kimura, 1962). The nearly-neutral theory predicts that the DFE includes a large proportion of mutations with a $Ne*s$ close to 0 (Ohta, 1992). As a consequence, an increase of $Ne$ will lead to an increase of the fitness effect of weakly deleterious mutations, in such a way that these mutations will be more easily removed from the population by natural selection, therefore reducing Pn relative to Ps, leading to a negative correlation between Pn/Ps and Ps (through Ne; Welch et al., 2008). The presence of linked mutations, that are positively selected, does not change this relationship qualitatively (Castellano et al., 2018; Chen et al., 2020 and our simulations below).

Shifts in the parasitic community on islands are expected to have an impact on the Pn/Ps ratio of immune genes. However, the fixation probability depends on the product $Ne*s$, and variation in $Ne$ is also expected to impact the efficacy of selection and thus the Pn/Ps ratio across the entire transcriptome, particularly in the presence of slightly deleterious mutations (Charlesworth and Eyre-Walker, 2008; Leroy et al., 2021b; Loire et al., 2013; Ohta, 1992). In addition, due to their lower population sizes, island birds compared to continental species exhibit a genome-wide reduction in genetic diversity and efficacy of selection (Kutschera et al., 2020; Leroy et al., 2021b). Therefore, we expect a similar reduction in immune genes' diversity even without any change in the parasite pressure.

To disentangle the effect of population size from a change in parasite pressure and estimate the impact of demography on the efficacy of selection, we studied a dataset of 34 bird species (20 insular and 14 mainland species; Figure 1) combining the 24 species of Leroy et al. (2021b) and 10 newly generated by targeted-capture sequencing (Table 1). We randomly selected protein-coding genes (i.e., control genes) involved in various biological functions (Fijarczyk et al., 2016; Leroy et al., 2021b). The selection pressure acting on the randomly selected control genes is expected to be similar between island and mainland bird species. Therefore, the variation of Pn/Ps of the control genes is only dependent on the variation of $Ne$. In contrast, if a reduced parasite pressure on islands directly impacts the evolution of immune genes, the Pn/Ps of immune genes is expected to show a larger variation between island and continental species

than the control genes. More specifically, for genes under purifying selection, non-synonymous weakly deleterious mutations, normally eliminated under strong selection, would be maintained, leading to an increase of Pn/Ps. By contrast, for genes under balancing selection, non-synonymous advantageous mutations, normally maintained in the polymorphism under strong selection, would be fixed or eliminated leading to a decrease of Pn/Ps (Figure 2).

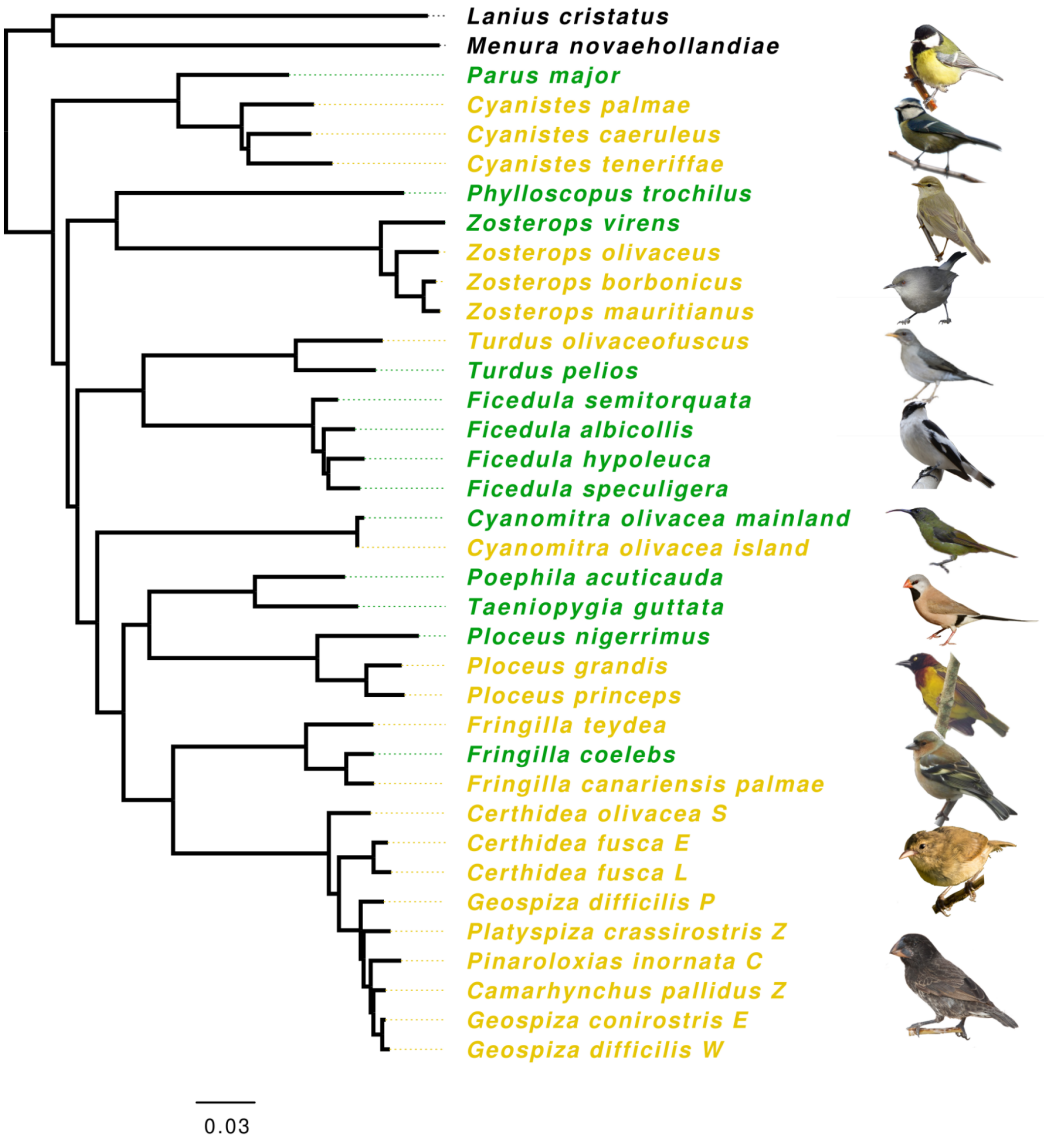


**Figure 1:** Phylogeny based on mitochondrial genes of species from the dataset reconstructed by maximum likelihood method (IQTREE model GTR+Gamma). Species names in yellow indicate island species, and in green, mainland species. Ultrafast bootstrap values are provided in the supplementary methods. Some relationships are poorly supported. Bird representations are not to scale. Photos from top to bottom : *P. major, C. caeruleus, P. trochilus, Z. borbonicus, T. pelios, F. albicollis, C. olivacea, P. acuticauda, P. grandis, F. coelebs, C. fusca, G. conirostris*. Photo credits: A. Chudý, F. Desmoulins, E. Giacone, G. Lasley, Lianaj, Y. Lyubchenko, B. Nabholz, J.D. Reynolds, K. Samodurov, A. Sarkisyan, Wimvz, Birdpics, T. Aronson, G. Lasley, P. Vos (iNaturalist.org); M. Gabrielli (*Zosterops borbonicus*).

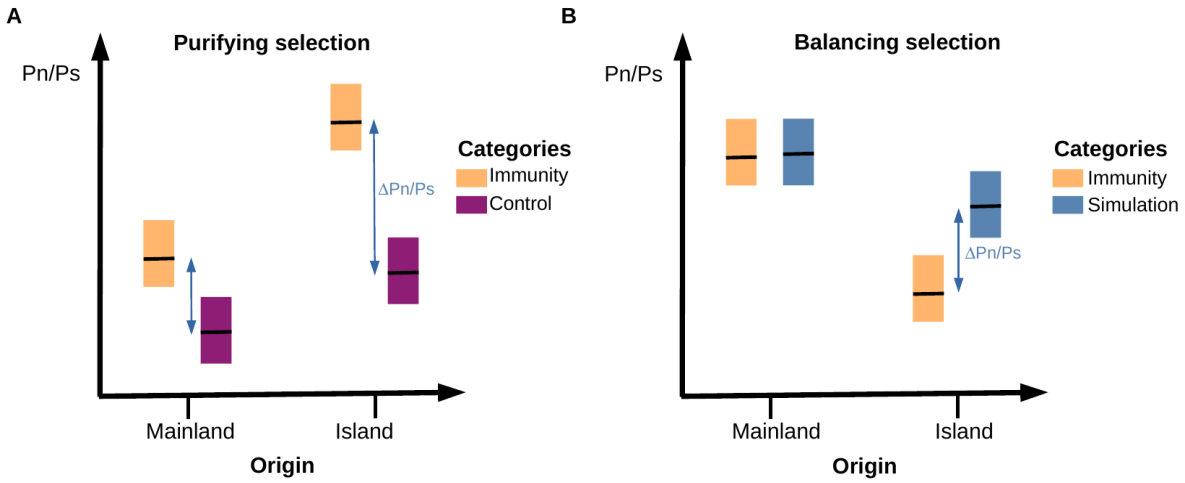


**Figure 2:** Conceptual diagram showing the expected results under the hypothesis of a relaxation in the selection pressure of the immune genes in island species due to a change in the parasitic community. A) Genes evolving under purifying selection where control genes are randomly selected protein-coding genes. B) Genes evolving under balancing selection where controls are obtained from SLiM simulations of genes evolving under the same balancing selection but different population size. Under the hypothesis of a relaxed selection as a consequence of the reduced diversity of pathogens on island ecosystems, the difference in Pn/Ps between categories (ΔPn/Ps) is expected to be different between species' origin, leading to a statistical interaction between gene categories and origin.

## Methods

### Dataset

Alignments of Coding DNA Sequences (CDS) of individuals from 24 species were obtained from Leroy et al. (2021b). In addition, data for ten other species (six and four from islands and mainland, respectively) were newly generated for this study by targeted-capture sequencing. Blood samples and subsequent DNA extractions were performed by different research teams. The complete dataset consisted of 34 bird species (20 and 14 insular and mainland species respectively; Table 1; Figure 1). We filtered alignments in order to retain only files containing a minimum of five diploid individuals per site (Table 1).

Sequence enrichment was performed using MYBaits Custom Target Capture Kit targeting 21 immune genes: 10 Toll-Like receptors (TLR), 9 Beta Defensins (BD), 2 Major Histocompatibility Complex (MHC) and 97 control genes (see below). We followed the manufacturer's protocol (Rohland and Reich, 2012). Illumina high-throughput sequencing, using a paired-end 150 bp strategy, was performed by Novogene (Cambridge, UK).

**Table 1:** List of species and sampling localities, along with the type of data obtained and the number of individuals (N).

| Species | Origin | Island/Country | N | Reference genome | Reference for population genomics data | Type of data |
|---|---|---|---|---|---|---|
| *Cyanistes teneriffae palmae* | Island | La Palma | 15 | *Cyanistes caeruleus* (This study) | (Mueller et al., 2016) | Capture |
| *Cyanistes teneriffae teneriffae* | Island | Tenerife | 14 | | | |
| *Cyanistes caeruleus* | Mainland | France | 15 | | | |
| *Parus major* | Mainland | Europe | 10 | *Parus major* (Laine et al., 2016) | (Corcoran et al., 2017) | Whole genome |
| *Phylloscopus trochilus* | Mainland | Europe | 9 | *Phylloscopus trochilus* (Lundberg et al., 2017) | (Lundberg et al., 2017) | Whole genome |
| *Zosterops virens* | Mainland | South Africa | 7 | *Zosterops borbonicus* (Leroy et al., 2021a) | (Leroy et al., 2021b) | Whole genome |
| *Zosterops olivaceus* | Island | Réunion | 15 | | | |
| *Zosterops mauritianus* | Island | Mauritius | 9 | | | |
| *Zosterops borbonicus* | Island | Réunion | 25 | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Ficedula semitorquata* | Mainland | Europe | 20 | *Ficedula albicollis* (Ellegren et al., 2012) | (Ellegren et al., 2012) | Whole genome |
| *Ficedula albicollis* | Mainland | Europe | 20 | | | |
| *Ficedula speculigera* | Mainland | Nord Africa | 20 | | | |
| *Ficedula hypoleuca* | Mainland | Europe | 20 | | | |
| *Turdus olivaceofuscus* | Island | São Tomé | 15 | *Turdus pelios* (This study) | This study | Capture |
| *Turdus pelios* | Mainland | Gabon | 15 | | | |
| *Cyanomitra olivacea* | Island | Príncipe | 15 | *Cyanomitra olivacea* (This study) | This study | Capture |
| *Cyanomitra olivacea* | Mainland | Gabon | 15 | | | |
| *Ploceus grandis* | Island | São Tomé | 13 | *Ploceus cucullatus* (This study) | This study | Capture |
| *Ploceus princeps* | Island | Príncipe | 13 | | | |
| *Ploceus nigerrimus* | Mainland | Cameroon Gabon | 14 | | | |
| *Poephila acuticauda acuticauda* | Mainland | Australia | 10 | *Taeniopygia guttata* (Warren et al., 2010) | (Singhal et al., 2015) | Whole genome |
| *Taeniopygia guttata castanotis* | Mainland | Australia | 19 | | | |
| *Fringilla teydea* | Island | Tenerife | 10 | *Fringilla coelebs* (Recuerda et al., 2021) | (Leroy et al., 2021b) | Whole genome |
| *Fringilla canariensis palmae* | Island | La Palma | 15 | | | |
| *Fringilla coelebs* | Mainland | Spain | 9 | | | |
| *Certhidea olivacea* | Island | Santiago (Galápagos) | 5 | *Geospiza fortis* (Zhang et al., 2012) | (Lamichhaney et al., 2015) | Whole genome |
| *Certhidea fusca* | Island | San Cristobal (Galápagos) | 10 | | | |
| *Certhidea fusca* | Island | Española (Galápagos) | 10 | | | |
| *Geospiza difficilis* | Island | Pinta(Galápagos) | 10 | | | |
| *Platyspiza crassirostris* | Island | Santa Cruz (Galápagos) | 5 | | | |
| *Pinaroloxias inornata* | Island | Coco (Galápagos) | 8 | | | |
| *Camarhynchus pallidus* | Island | Santa Cruz (Galápagos) | 5 | | | |
| *Geospiza difficilis* | Island | Wolf (Galápagos) | 8 | | | |
| *Geospiza conirostris* | Island | Española (Galápagos) | 10 | | | |

*Newly generated draft genome sequence*

We generated whole genome sequences at moderate coverage (~40X) for *Turdus pelios*, *Ploceus cucullatus* and *Cyanomitra olivacea* (from Gabon). Library preparation from blood DNA samples and Illumina high-throughput sequencing using a paired-end 150 bp strategy were performed at Novogene (Cambridge, UK). Raw reads were cleaned using FastP (vers. 0.20.0; Chen et al., 2018). Genomes assemblies were performed using SOAPdenovo (vers. 2.04) and Gapcloser (v1.10) (Luo et al., 2012) with parameters "-d 1 -D 2" and a kmers size of 33. Protein annotation was performed by homology detection using genBlastG (She et al., 2011; http://genome.sfu.ca/genblast/download.html) and the transcriptome of the collared flycatcher (*Ficedula albicollis*; assembly FicAlb1.5; Ellegren et al., 2012) as reference.

*Capture data processing*

Reads from targeted-capture sequencing were cleaned with FastP (vers. 0.20.0; Chen et al., 2018). Reads of each individual were mapped respectively to the nearest available reference genomes using bwa mem (vers. 0.7.17; Li, 2013; Table 1), with default parameters. Samtools (vers. 1.3.1; Li et al., 2009) and Picard (vers. 1.4.2; Picard Toolkit 2019) were used to convert the mapping files, order and index reads according to their position on the chromosomes (or scaffolds) of the reference genomes or on the draft genomes generated in this study for *Ploceus*, *Cyanomitra* and *Turdus*. Duplicate reads were marked using MarkDuplicates (vers. 1.140; Picard Toolkit 2019). SNP calling was performed with Freebayes (vers. 1.3.1; Garrison and Marth, 2012). Freebayes output file (VCF file) was converted to a fasta file by filtering out sites with a minimum quality of 40 and a sequencing depth between 10 and 1000X (sites outside these thresholds were treated as missing data, i.e., 'N'). CDS were then extracted from the alignments using the coordinates of the annotations (gff files). CDS were aligned using MACSE (vers. 2.03; Ranwez et al., 2011) to prevent frameshift mutation errors and GNU-parallel (Tange, 2018) was used to parallelise the computation.

*Selection and identification of immune and control genes*

We defined several groups of immune genes to compare with the control genes. The control group consisted of 97 protein-coding genes randomly selected in the genome of *Zosterops borbonicus* (Leroy et al., 2021a). These control genes allowed the estimation of the average selection pressure that a gene, not involved in the immune response, undergoes in the genome under a given effective population size. These genes were single copy (absence of paralogue) and had a variable GC content representative of the whole transcriptome.

For the immune genes, we selected three sets of genes from i) a limited set of genes (Core Group) where functions are unambiguously related to immunity, and ii) two larger sets of genes (Database-group & Sma3s-group), obtained through an automatic annotation pipeline.

The Core Group included MHC class I and class II genes, 10 Toll-Like Receptors (TLRs; Velová et al., 2018) and 9 Beta Defensins (BD; Chapman et al., 2016). The Database group included genes identified by Immunome Knowledge Base (Ortutay and Vihinen, 2009, http://structure.bmc.lu.se/idbase/IKB/; last access 04/02/2020) and InnateDB (Breuer et al., 2013, http://www.innatedb.com ; last access 04/02/2020). We also added a set of genes for which the genetic ontology indicated a role in immune functions. To do so, we used the chicken (*Gallus gallus*) annotation (assembly GRCg6a downloaded from Ensembl database in March 2020; https://www.ensembl.org/). We identified genes with the terms "immun*" or "pathogen*" in their Gene Ontology identifiers description (directory obtained from http://geneontology.org/). This set included 2605 genes considered to be involved in immunity, although some may be only indirectly involved in immunity or have a small impact on immune functions. Finally, the third set of genes (Sma3s-group) has been built up through the Sma3s-group program (vers. 2; Munoz-Mérida et al., 2014). This program annotated sequences in order to be associated with biological functions through gene ontology identifiers. The annotation of the genome of *F. albicollis* allowed us to identify 3136 genes associated with the genetic ontology "immune system processes''. Like for the Database group, this set may include genes with various functions in the immune response. It should be noted that Sma3s-group and Database-group were not mutually exclusive, and some genes were present in both groups. An analysis was performed to identify and exclude genes under balancing selection from Database-group and Sma3s-group sets using BetaScan (vers. 2; Siewert and Voight, 2020), due to the potentially antagonistic responses of these genes. Very few genes (only 2 and 3 genes from Database-group and Sma3s-group sets) were identified and removed from the analysis (see Detection of genes under balancing selection in Supplementary Methods).

*Test for contamination and population structure*

We used the program CroCo (vers. 1.1; Simion et al., 2018) to identify candidates for cross-species contamination (see supplementary materials for details). Overall, we did not detect a clear case of cross-species contamination in our dataset (Figure S1; Barthe and Nabholz, 2022). Contigs identified as potential contamination always involve a pair of species belonging to the same genus. In this case, contamination could be difficult to identify due to the low genetic divergence between species.

For the newly sequenced species, we also performed PCA analyses using allele frequencies of control genes. We used the function dudi.pca of adegenet R package (Jombart and Ahmed, 2011). This analysis aims to check for population structure and to detect potentially problematic individuals (i.e., contaminated individuals). This analysis led to the exclusion of 4 individuals (*Ploceus princeps* P6-174; *P. grandis* ST10_094; *P. nigerrimus* G3_016; *C. teneriffae* TF57) for which we suspected contamination. Otherwise, no extra population structure was detected (Figure S2-S4; Barthe and Nabholz, 2022).

*Hidden paralogy*

We computed the statistic $F_{IS} = 1-H_0/H_e$ where $H_0$ is the average number of heterozygous individuals observed ($H_0$ = #heterozygous / n ; where n is the sample size) and $H_e$ is the expected number of heterozygous individuals at Hardy-Weinberg (HW) equilibrium ($H_e$ = (n/(n-1) 2 * p * (1-p))*n where n is the sample size and p the allele frequency of a randomly chosen allele). $F_{IS}$ varies between -1 and 1 with positive value representing excess of homozygous individuals and negative value representing excess of heterozygous individuals compared to the HW proportions. Gene with high value of nucleotide diversity (Pi) and negative value of $F_{IS}$ could represent a potential case where hidden paralogous sequences have not been separated and where all the individuals present heterozygous sites in the positions where a substitution occurred between the paralogous copies. Five sequences corresponding to the TLR21 genes appeared problematic (Pi > 0.01 and $F_{IS}$ < -0.5; Figure S5; Barthe and Nabholz, 2022) and were excluded from further analyses.

The MHC genes were more difficult to analyse. Indeed, heterozygosity could be comparable to divergence under balancing selection. This made the identification of orthologs very difficult. We identified a variable number of genes among species (from 1 to 10 genes for MHC class I and MHC class II). We checked the sequence similarity for the 10 copies of the MHC class II in *F. albicollis* and the 7 copies of the MHC class I genes in *C. caeruleus* using cd-hit (Fu et al., 2012). For MHC class II, sequence

divergences were always higher than 15% indicating that reads were likely correctly assigned to their corresponding gene copy. For MHC class I, sequence similarity could be as high as 95%. In this case, we relied on the fact that the reads from very similar paralogous copies were not be confidently assigned to a gene copy sequence by the mapping software. This should lead to a low mapping score quality and were likely to be discarded during the genotype calling procedure. For example, 3 out of 7 of the *Cyanistes* MHC class I genes were not correctly genotyped and were missing from our final dataset.

**Data Analysis**

*SLiM simulations*

We used SLiM (vers. 3.3.2; Haller and Messer, 2017) to estimate the impact of demographic changes on polymorphism patterns under various selection regimes. The following parameters were used in all simulations. Sequences of 30kb with a mutation rate of $4.6e^{-9}$ substitutions/site/generation were simulated (Smeds et al., 2016). Recombination was set to be equal to mutation rate. Introns/exons pattern was reproduced by simulating fragments of 3kb separated by one bp with a very high recombination rate of 0.1 rec./site/generation. We chose 3kb because TLR CDS were typically single-exon sequences of 2-3kb (Velová et al., 2018). Five types of mutations were possible: i) neutral synonymous mutations, ii) codominant non-synonymous mutations with a Distribution of Fitness Effect (DFE) following a gamma law of mean = -0.025 and shape = 0.3, which corresponds to the DFE estimated in Passerines by Rousselle et al. (2020), iii) codominant non-synonymous mutations positively selected with s = 0.1, iv) non-synonymous mutations under balancing selection with an effect on fitness initially set at 0.01 but re-estimated by the program at each generation according to the mutation frequency in the population, thus including a frequency-dependent effect and v) non-synonymous mutations under overdominance with a dominance coefficient of 1.2.

We simulated a coding sequence organization where positions one and two of the codons were considered as non-degenerated sites, with the non-synonymous types of mutations previously described were possible in various proportions. The third position was considered as completely neutral where only synonymous mutations could appear.

In the absence of control genes evolving under balancing selection, we used SLiM to generate a set of control genes for this category. We simulated two populations of 270,000 and 110,000 individuals, representing mainland and island effective population size respectively.

We also explored the effect of positive and balancing selection on the pattern of Ps and Pn/Ps in a population of size 50,000, 110,000, 270,000 and 500,000. In order to speed up the computational time, we reduced the population size by a factor 100 and rescaled mutation rate, recombination rate and selection coefficient accordingly running 10 replicates per simulation.

All the details of the simulation parameters, calculations of non-synonymous polymorphism rate (Pn) and synonymous polymorphism rate (Ps) of simulated sequences, as well as SLiM command lines are provided in Supplementary Methods and Materials.

*Polymorphism analyses*

Synonymous (Ps) and non-synonymous (Pn) nucleotide diversities were estimated from seq_stat_coding written from the Bio++ library (Available as Supplementary data; Guéguen et al., 2013). The mean Pn/Ps was computed as the sum of Pn over the sum of Ps (Wolf et al., 2009). Ps of concatenated sequences of control genes were estimated for each species of our dataset. For the whole-genome sequenced species, we compared the Pn/Ps and Ps estimated from the 97 control genes with the values from Leroy et al., (2021b; ~5000 genes used in their study). Pn/Ps and Ps correlations showed a $R^2$ of 0.6 and 0.95 respectively (Figure S6; Barthe and Nabholz, 2022). Thus, the 97 control genes used in our study were representative of the larger set of genes from Leroy et al (2021b). This allowed us to identify *Phylloscopus trochilus* as an outlier. Unlike for all other species (e.g. *Fringilla coelebs*, Figure S7; Barthe and Nabholz, 2022), synonymous polymorphism level was correlated to the amount of missing data in *P. trochilus* alignments (Figure S7; Barthe and Nabholz, 2022). As such, we excluded *P. trochilus* from further analysis.

The mean Pn/Ps, calculated from the concatenated sequences of genes from the same gene class (control genes; BD; TLR; MHC I; MHC II; Database-group; Sma3s-group), was estimated for each bird species. Alternative transcripts were identified based on the genomic position in the GFF file. If several

transcripts were available, one transcript was randomly selected. Pn/Ps estimates based on less than four polymorphic sites were excluded from the analysis, as were those with no polymorphic non-synonymous sites.

*Statistical analyses*

To estimate the impact of demographic history on genome-wide polymorphism of island species and the potentially reduced constraints on their immune genes, we computed the ratio of non-synonymous nucleotide diversity over synonymous nucleotide diversity (Pn/Ps). A linear mixed model was performed, using the Pn/Ps ratio as dependent variable and, as explanatory variables, the mainland or insular origin of species as well as the category of genes (packages lme4 and lmerTest (Bates et al., 2012; Kuznetsova et al., 2017)). In order to take into account the phylogenetic effect, the taxonomic rank "family" was included as a random effect in the model. We also used a generalized linear mixed model (using the function glmer of the package lme4) with the family "Gamma(link="log")" which led to the same results (Figure S15 to S24; Barthe and Nabholz, 2022). Five linear mixed models were defined i) model including origin and gene category parameters and also the interaction effect ii) model using both origin and gene category parameters, iii) model with only the gene category parameter, iv) model with only the origin parameter, and finally v) null model. In some cases, the phylogenetic effect was difficult to estimate because the number of species per family was reduced to one. In that case, we choose to reduce the number of families by grouping Turdidae with Muscicapidae, Nectariniidae, and Estrildidae with Ploceidae and Fringillidae within Thraupidae. The results obtained with these family groupings were similar to the original model (Table S1; Barthe and Nabholz, 2022), except when stated. The categories Database-group and Sma3s-group were tested separately from the Core group because they contained hundreds of genes annotated using the automatic pipeline that were only available for species with genome wide data. Database-group and Sma3s-group were not analysed simultaneously because they contained a partially overlapping set of genes. Finally, genes evolving under purifying selection and genes evolving under balancing selection were also analysed separately. Model selection was based on two methods. First, we used the difference in corrected Akaike Information Criterion (ΔAICc) calculated using the qpcR package (Spiess and Spiess, 2018). Second, a model simplification using an ANOVA between models was also performed.

We also tested an alternative model using the difference between Pn/Ps of immune genes and control genes (ΔPn/Ps) as dependent variable, and species origin as explanatory variable. Under the hypothesis of a relaxation in selection pressure on islands due to a change in the parasite community, we expected the ΔPn/Ps to be higher on island species compared to the mainland ones and, therefore, the species origin (i.e., mainland or island) to be significant. In this model, we used the Phylogenetic Generalized Least Squares model (PGLS; implemented in the "nlme" packages; Pinheiro et al., 2017). This model assumed that the covariance between species follows a Brownian motion evolution process along the phylogeny (implemented using the "corBrownian" function from the '"ape"' package; Paradis and Schliep, 2019). The species phylogeny was estimated using mitochondrial genes and a maximum likelihood inference implemented in IQTREE (model GTR+Gamma and ultrafast bootstrap; Nguyen et al., 2014; median of 11,134 bp analysed per species). The phylogeny with the bootstrap support is provided as supplementary material.

All the statistical analyses were performed using R (R Core Team, 2018), and dplyr package (Wickham, 2016). Graphical representations were done using ggplot2, ggrepel, ggpubr and ggpmisc (Aphalo, 2020; Kassambara, 2018; Slowikowski et al., 2018; Wickham, 2016).

## Results

For the 150 individuals (10 species with 15 individuals each) for which we generated new data by targeted capture sequencing, an average of 3.3 million paired-ends reads per individual was generated (Table S1; Barthe and Nabholz, 2022). Additionally, we generated three new draft assemblies using 40x pair-end illumina data for the species for which no closely related reference genomes were available. N50 and total size were 1.11 Gb and 27.9 kb for *Cyanomitra olivaceus;* 1.10 Gb and 31.7 kb for *Ploceus cucullatus* and 1.13 Gb and 14.3 kb for *Turdus pelios*. After mapping, genotyping and cleaning, we analysed 86 control and 16 immune genes on average per species, out of the 141 targeted genes (120

control and 21 immune related genes; Table S4; Barthe and Nabholz, 2022). For the species with whole-genome sequences, we analysed 106 control and 20 immune genes on average per species, out of the 141 targeted genes, and 875 and 688 genes on average in the Database-group and Sma3s-group respectively (Table S4; Barthe and Nabholz, 2022).

For the species for which full genome sequences were available, the Ps and Pn/Ps estimated using the control genes reflect the Ps and Pn/Ps of the whole transcriptome (Figure S6; Barthe and Nabholz, 2022).

**Population genetics of BD and TLR immune genes**

In order to characterize the selection regimes shaping the BD and TLR polymorphisms (Figure 3), we first analyzed the variation of Pn/Ps ratios among gene categories using a linear mixed model.

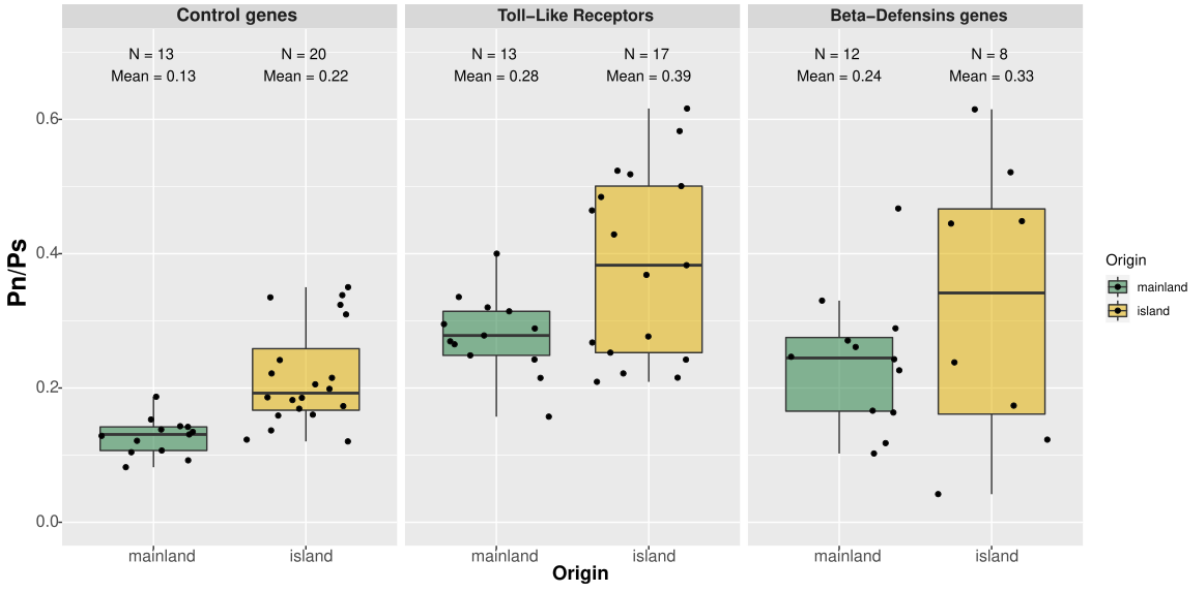


**Figure 3:** Pn/Ps according to species origin (mainland in green and insular in orange) for different gene categories under purifying selection. The number of species (N), and the mean Pn/Ps are shown for each modality.

Model selection based on AICc as well as model selection approach based on simplification with ANOVA identified the model n° 2, including the origin (i.e., mainland or island) and gene category without interaction (Table 2). In this model, island origin of species is associated with a greater Pn/Ps (0.14 vs. 0.10; Table 3; $p < 0.01$). Gene categories corresponding to TLRs and BDs showed a significantly higher Pn/Ps than control genes (Table 3; $p < 0.001$). Our statistical analysis confirmed that island birds have a higher Pn/Ps ratio than mainland relatives, in agreement with the nearly-neutral theory of evolution. It also reveals that immune genes have a higher Pn/Ps than randomly selected control genes suggesting that BD and TLR evolve under a different selection regime than non-immune related genes.

Next, we investigated the cause of the higher Pn/Ps of immune genes by testing three hypotheses. First, we excluded a bias due to a lower number of immune genes, and therefore higher variance in the estimation of Pn/Ps in immune genes. Immune genes still had significantly higher Pn/Ps compared to a random subsample of control genes of comparable size (Figure S8 & S9; Barthe and Nabholz, 2022). Second, the Pn/Ps of immune genes could be inflated by positive selection. It is well known that immune genes are subject to frequent adaptation due to arms race evolution with pathogens (Enard et al., 2016; Shultz and Sackton, 2019; Velová et al., 2018). We evaluated the effect of positively selected genes on the Pn/Ps using SLiM simulations with both positively and negatively selected mutations. The presence of recurrent positive selection could increase the Pn/Ps leading to a higher Pn/Ps in immune genes if this category was more prone to adaptive evolution (Figure 4A). However, positive selection always led to a drastic decrease in Ps due to genetic sweep effect at linked sites (Figure 4B). BDs and TLRs had a slightly higher or similar Ps than control genes (Figure S9, mean Ps = 0.007, 0.004 and 0.003 for BDs, TLRs and control genes respectively, effect of gene category $p < 0.1$; Barthe and Nabholz, 2022) and, as a consequence, even if positive selection is likely to have impacted the evolution of immune genes, it is not

the cause of the higher Pn/Ps observed here. Third, balancing selection could be present, at least temporarily, in the evolution of BDs and TLRs genes (Kloch et al., 2018; Levy et al., 2020). Simulation analyses confirmed that balancing selection causes an increase of Ps and Pn/Ps (Figure 4C & 4D). However, a change in effective population size had an opposite effect on the Pn/Ps according to whether selection was negative or balancing. In the presence of slightly deleterious mutations, Pn/Ps decreases with *Ne* whereas it increases in the presence of balancing selection. Island birds had higher Pn/Ps ratios than mainland birds for BDs and TLRs. Therefore, we can rule out balancing selection as the main factor explaining the high Pn/Ps of immune genes because, in this case, Pn/Ps of island birds should be lower. Another possible explanation is a relaxed selection of immune genes. It is likely that immune genes are overall less constrained than the control genes. It has been shown that evolutionary constraints are more related to gene expression than to function (Drummond et al., 2005; Drummond and Wilke, 2008) and therefore, functionally important genes could still have a high Pn/Ps.

Overall, our analyses do not support a strong impact of ongoing adaptive mutation or balancing selection on BDs and TLRs. However, these immune genes do not evolve as random genes (not involved in immune functions) and present a significantly higher Pn/Ps of 0.20 ($p < 0.001$ ; Table 3).

**No evidence of a reduced impact of the parasite communities on the polymorphism pattern of immunes genes in island birds**

For BDs and TLRs, the best model selected includes the origin (i.e., mainland or island) and gene category without interaction, corresponding to model n°2 (see above and Table 2). This model has no interaction between origin and gene categories invalidating the hypothesis of a reduced parasite communities on islands (Figure 2).

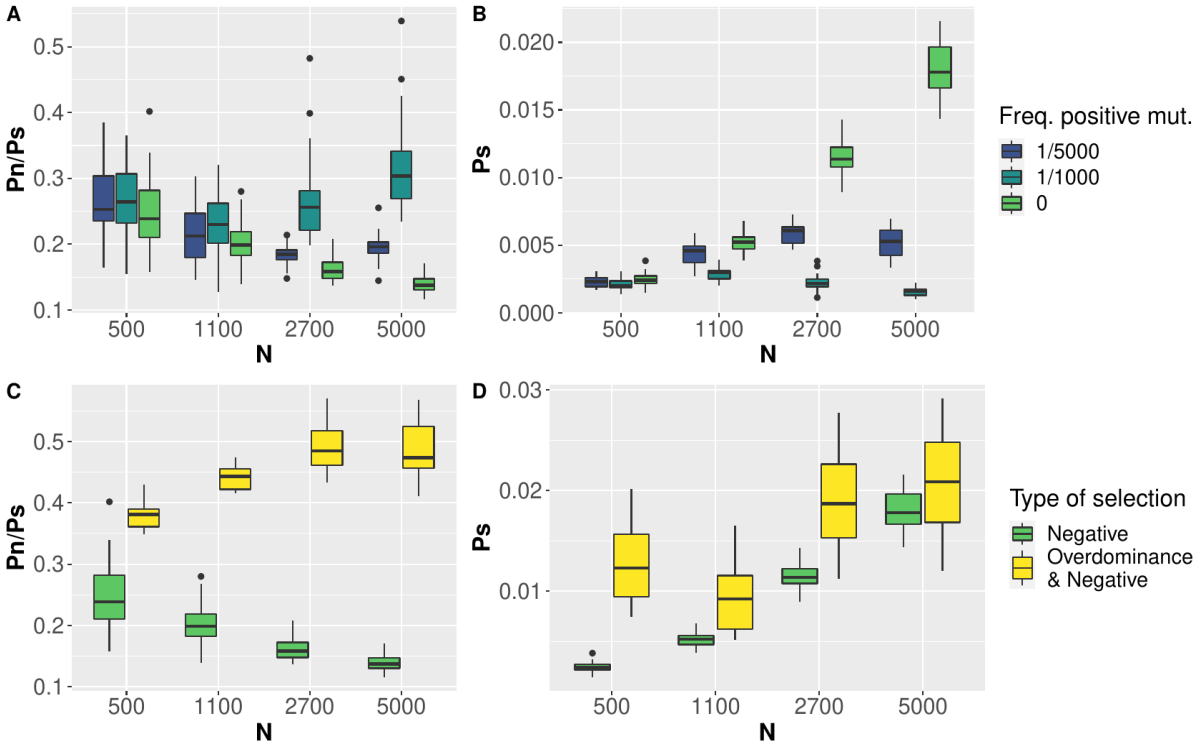


**Figure 4:** Neutral polymorphism (Ps) and ratio of selected over neutral polymorphism (Pn/Ps) estimated from SLiM simulations. A) Pn/Ps as a function of population size, N and B) Ps as a function of N. In both A and B, color indicates the frequency of positively selected mutations compared to deleterious mutations. C) Pn/Ps as a function of N and D) Ps as a function of N. In both C and D, yellow indicates simulations with overdominance mutation (h = 1.2) and negatively selected mutations and green indicates simulations with only negatively selected mutations.

**Table 2:** Statistical model explaining Pn/Ps variation of Toll-Like Receptors, Beta-Defensins genes, and control genes. The p-values of ANOVA test between simpler models are not reported if a more complex model explains a larger proportion of the variance.

| | Model | Model selection by AIC | | | ANOVA test | | | |
|---|---|---|---|---|---|---|---|---|
| n° | Details | AICc | ΔAICc | Likelihood | n° 1 | 2 | 3 | 4 |
| 1 | Pn/Ps~ 1+ category +origin+ category *origin | -5.39 | 8.83 | 0.01 | | 0.63 | | |
| 2 | Pn/Ps~ 1+ category +origin | -14.22 | 0 | 1 | | | 0.002 | 3.71E-05 |
| 3 | Pn/Ps~1+ category | -11.8 | 2.42 | 0.3 | | | | |
| 4 | Pn/Ps~1+ origin | -6.83 | 7.39 | 0.02 | | | | |
| 5 | Pn/Ps~1 | -6.44 | 7.78 | 0.02 | | | | |

**Table 3:** Summary of the model n°2, best statistical model selected using AICc explaining variation in Pn/Ps in control genes, Toll-Like receptors and Beta-Defensins genes under purifying selection with origin, gene category parameters. * indicates significant values : * < 0.05; ** < 0.01; *** < 0.001.

| Model | *Parameters* | | | *Estimate* | *P.value* | |
|---|---|---|---|---|---|---|
| | | *Origin* | *Category* | | | |
| *Origin and Gene category (n°2)* | *Intercept* | *mainland* | *Control genes* | *0.10* | *2.65E-02* | * |
| | | *island* | | *0.14* | *4.56E-03* | ** |
| | | | *Toll-Like Receptors* | *0.20* | *7.43E-05* | *** |
| | | | *Beta-Defensins genes* | *0.20* | *3.16E-04* | *** |

For larger sets of genes, identified using an automatic pipeline and gene annotation, model selection based on AICc and simplification with ANOVA (Table S5, S8; Barthe and Nabholz, 2022) identified models n°4 that included origine parameters which associated a higher Pn/Ps of at least 0.07 for island species ($p < 0.001$; Table S6, S7, S9, S10; Barthe and Nabholz, 2022, Figure 5). Model selection by simplification with ANOVA identified models n°1 with interaction effect between origin and gene category associated with a reduced Pn/Ps for TLR and BD genes of island species that invalidate our hypothesis (Table S7, S10; Barthe and Nabholz, 2022).

The alternative statistical approach using the difference between Pn/Ps of immune genes and control genes (ΔPn/Ps) as dependent variable, and species origin as explanatory variable under a PGLS framework lead to similar results. Island was never associated to a statistically higher ΔPn/Ps (Table S2; Barthe and Nabholz, 2022) providing no support for an increased relaxed selection of immune genes in island species.

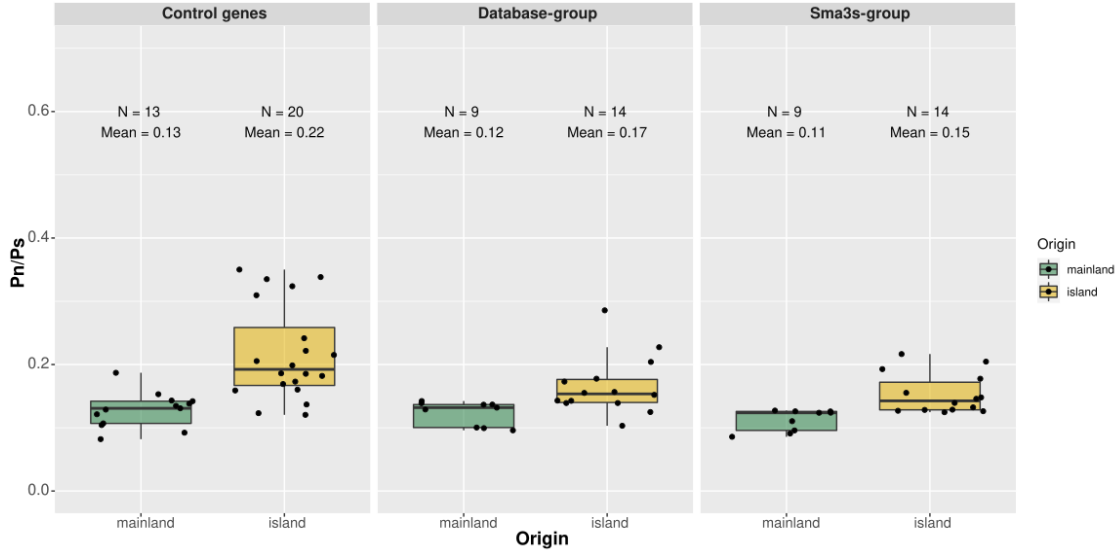


**Figure 5:** Boxplot of Pn/Ps according to species origin (mainland in green and insular in orange) for different gene categories under purifying selection. The number of individuals (N), and the mean Pn/Ps are shown for each modality.

**Genes under balancing selection**

First, we estimated the effect of population size variation on the Pn/Ps of the genes evolving under balancing selection by simulating sequences under frequency dependent or overdominance selection using SLiM (see Methods and Supplementary Methods). The simulation under frequency dependent selection revealed an average Pn/Ps equal to 0.8 for island species and 1.2 for mainland species (Figure 6). Under overdominance, simulated sequences from island and mainland populations respectively have an average Pn/Ps equal to 0.54 and 1.03 (Figure 6).

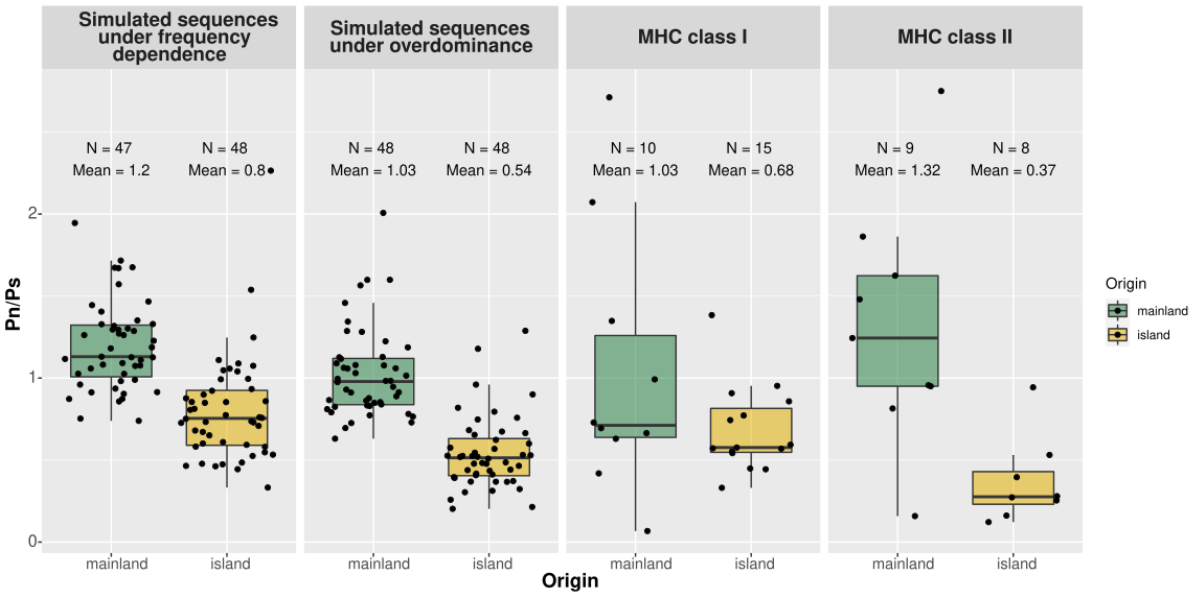


**Figure 6:** Boxplot of Pn/Ps according to species origin (mainland in green and insular in orange) for different gene categories under balancing selection. The number of species (N), and the mean Pn/Ps are shown for each modality. The control groups correspond to the results obtained from simulated sequence via SLiM (see Methods and Supplementary Methods Simulation of control genes under balancing selection).

Using simulations under frequency dependent selection as well as simulations under the overdominance, model selection by AIC identifies the model n°4 with origin, contrary to the method by simplification with ANOVA which identified the full model (model n°1) therefore including significant interaction between origin and genes category (Table 4). This interaction effect is significant for the MHC II ($p < 0.05$, Table S12; Barthe and Nabholz, 2022) but not for MHC I. As expected, island species have a significantly lower Pn/Ps in MHC genes compared to mainland species ($p < 0.01$; except for the full model based on control genes evolving under overdominance Table S12; Barthe and Nabholz, 2022).

**Table 4:** Statistical model explaining Pn/Ps variation of genes under balancing selection (i.e MHC class I and II), and simulated sequences under i) frequency dependent or ii) overdominance. The p-values of ANOVA test between simpler models are not reported if a more complex model explains a larger proportion of the variance.

| | Model | | Model selection by AIC | | | ANOVA test | | | |
|---|---|---|---|---|---|---|---|---|---|
| Type of balancing selection | n° | Details | AICc | ΔAICc | Likelihood | n°1 | 2 | 3 | 4 |
| Frequency dependent | 1 | Pn/Ps~1+ category +origin+ category *origin | 157.17 | 5.62 | 0.06 | | 0.019 | | |
| | 2 | Pn/Ps~1+ category +origin | 157.85 | 6.31 | 0.04 | | | | |
| | 3 | Pn/Ps~1+ category | 187.58 | 36.04 | 0.00 | | | | |
| | 4 | Pn/Ps~1+ origin | 151.54 | 0.00 | 1.00 | | | | |
| | 5 | Pn/Ps~1 | 180.52 | 28.97 | 0.00 | | | | |
| Overdominance | 1 | Pn/Ps~1+ category +origin+ category *origin | 140,56 | 8,50 | 0,01 | | 0.024 | | |
| | 2 | Pn/Ps~1+ category +origin | 140,56 | 8,50 | 0,01 | | | | |
| | 3 | Pn/Ps~1+ category | 185,91 | 53,85 | 0,00 | | | | |
| | 4 | Pn/Ps~1+ origin | 132,05 | 0,00 | 1,00 | | | | |
| | 5 | Pn/Ps~1 | 177,54 | 45,49 | 0,00 | | | | |

## Discussion

On oceanic islands, the depauperate parasite community is expected to lead to a relaxation of selection on the immune system. In this study, we found support for such an effect, but only on MHC class II genes and using simulated sequences under balancing selection as control. No effect was detected for MHC class I genes nor for innate immune genes (TLRs and BDs), evolving under purifying selection. On these sets of genes, increased drift effects on island populations limit the efficacy of selection in accordance with the nearly-neutral theory (Ohta, 1992). The ability to distinguish between the selective and nearly-neutral processes (relaxed selection due to environmental change vs. drift) could only be achieved by our approach of using random genes (i.e., “control genes'') to estimate the genome-wide effect of potential variation in effective population size between populations.

### Effects of effective population size variation

Our results support the nearly-neutral theory of evolution for those genes under purifying selection, whereby strong genetic drift acting on small island populations reduces the efficacy of natural selection, leading to an increase in non-synonymous nucleotide diversity compared to the mostly neutral, synonymous nucleotide diversity (i.e., Pn/Ps; Ohta, 1992). This is materialized by a genome-wide increase in frequency of weakly deleterious mutations (Kutschera et al., 2020; Leroy et al., 2021b; Loire et al., 2013; Robinson et al., 2016; Rogers and Slatkin, 2017).

For genes evolving under balancing selection, we performed simulations under the hypotheses of overdominance (heterozygote advantage) or frequency-dependent (rare allele advantage). Our results showed reduced Pn/Ps for smaller population sizes (Figure 6, S10, S11). This simulation confirmed our expectations (Figure 2) that a reduction in the efficacy of selection results in a decrease in the frequency of non-synonymous polymorphism, as, under normal circumstances, selection maintains those mutations at intermediate frequencies. It also matches what we obtained for the empirical results, where both MHC classes I and II had a reduced Pn/Ps in island birds. This result supports that the fitness effect of having non-synonymous polymorphisms segregating at high frequencies is not strong enough to counteract entirely the effect of genetic drift on islands.

### Effects of selection on immune genes

For immune genes, we tried to characterize the nature of the selection acting on BDs and TLRs genes. Comparing those genes with control genes and using simulations, we were able to rule out that directional positive selection and balancing selection had a major impact shaping the polymorphism of these immune genes. In contrast, the pattern of Pn/Ps between island and mainland populations is in line with the effect of purifying selection in the presence of slightly deleterious mutations. However, no effect was detected on insular species, beyond what could be attributed to genetic drift. This is in line with the result of Gonzalez-Quevedo et al. (2015b) and Grueber et al. (2013) who found that TLR genetic diversity was mostly influenced by genetic drift. At first sight, this result seems not in line with the fact that island parasite communities are less diverse (Beadell et al., 2006; Loiseau et al., 2017; Maria et al., 2009; Pérez-Rodríguez et al., 2013; but see Illera et al., 2015). However, a reduced number of pathogens has also been found to be associated with a higher prevalence in birds and reptiles from the Macaronesian archipelago (Illera and Perera, 2020). Therefore, these two patterns, i.e. a less diverse pathogen's community on islands with a higher prevalence, could still imply a strong selection pressure on immune genes.

In contrast, for MHC genes that unambiguously evolve under balancing selection, MHC class II genes presented a reduction in non-synonymous polymorphism larger than the effects of drift alone, when simulated sequences are used as control. This was the only case where a role for relaxed selection pressures in the molecular evolution of immune genes could be invoked.

Our results are in accordance with the hypothesis of Lee (2006), which proposes that innate and acquired immunity may exhibit distinct responses to changes in pressures due to different costs and benefits. However, they contrast with the study of Santonastaso et al. (2017) that identified no change in selection pressures on MHC II genes in a lizard species and concluded that their evolution was mostly governed by drift. Similarly, Agudo et al. (2011) also found a prominent role for genetic drift over selection in the evolution of MHC II genes in the Egyption vulture (*Neophron percnopterus*).

Our results rely on simulations that may be affected by the choice of the parameter values. First, we performed simulations using a fixed effective population size (*Ne*) estimated from the polymorphism data. Using others values of *Ne* had a weak impact on the relative difference between island and mainland species for the overdominance type of selection (Figure S10, S11; Barthe and Nabholz, 2022). Secondly, we simulated two types of selection, namely overdominance (Doherty and Zinkernagel, 1975) and frequency-dependent (Slade and McCallum, 1992), but it has been argued that the maintenance of MHC polymorphism could be the result of fluctuating selection (Hill, 1991). Additionally, recombination has also been put forward as a mechanism responsible for generating diversity (Spurgin et al., 2011). Therefore, our results for the MHC II genes, which is based on the relative difference between Pn/Ps of island and mainland species comparing empirical and simulated data, should be taken cautiously as their significance can be dependent on the specific parameters that we used, although we did our best to select a realistic range of parameters.

The observed difference between MHC class I and II could be explained by their different pathogen targets: MHC class I genes are primarily involved in the recognition of intracellular pathogens (Kappes and Strominger, 1988), while MHC class II genes are directly involved in the recognition of extracellular pathogens (Bjorkman and Parham, 1990). These differences could lead to variable selection pressures depending on the extracellular versus intracellular parasite communities present on islands. In addition, the relaxed selection pressures on MHC II genes from adaptive immunity is in line with a reduction in acquired immunity parameters found by Lobato et al. (2017).

Future work should take into account that there is an extensive variation in the number of MHC gene copies across the avian phylogeny (Minias et al., 2019; O'Connor et al., 2020). Particularly, it was recently discovered that Passerines have a very dynamic evolution of duplication/loss events compared to other birds (Minias et al., 2019). Here, we used the two copies of MHC gene I and II currently annotated in the collared flycatcher genome as target sequences for our targeted-capture sequencing. The future improvement of genome assembly, resulting from the development of long-reads technology (Peona et al., 2021, 2018), should help to annotate with increased precision all MHC copies and to study the whole repertoire of MHC genes.

**Consequences of drift and selection on immunity**

The potential relaxation of the natural selection acting on immune genes in island species is expected to reduce immune functions and increase susceptibility of island populations to pathogens. This is true even if this relaxation is only the consequence of a reduction in the effective population size and not caused by a reduction of the pressure exerted by the parasitic community. This is in line with the results of Hawley et al. (2005) and Belasen et al. (2019) who showed that a decrease in diversity of immune loci (MHC II or through immune proxy) was associated with a reduction in immune functions. It should be noted that even if migration rate is reduced on islands, sedentary and endemic island species are not completely free from the exposure of exogen pathogens through migratory birds (Levin et al., 2013).

As a final remark, we would like to stress that more research is needed (i) to ascertain both selection pressures on innate and adaptive immune responses and the load of deleterious mutations due to drift, also identified by an increasing body of work (Loire et al., 2013; Robinson et al., 2016; Rogers and Slatkin, 2017; Kutschera et al., 2020; Leroy et al., 2021b), and (ii) to better describe island parasite communities. To date, most of the studies investigated intracellular parasite communities on islands, and more specifically haemosporidian parasites, avian pox and coccidian parasites (Cornuault et al., 2012; Illera et al., 2015, 2008; Ishtiaq et al., 2010; Loiseau et al., 2017; Martinez et al., 2015; Padilla et al., 2017; Pérez-Rodríguez et al., 2013; Silva-Iturriza et al., 2012), whereas very few evaluated the extracellular parasite diversity, such as helminths (Nieberding et al., 2006, but see the review of Illera and Perera 2020 for reptiles). Metabarcoding of parasites is a new technique to evaluate at the same time both communities of intracellular and extracellular parasites (Bourret et al., 2021) and might therefore be a promising approach to compare their communities in island and mainland populations.

## Conclusion

Our comparative population genomics study has investigated the combined effects of drift and selection on immune genes from island and mainland passerines. The study of synonymous and

non-synonymous polymorphism of these genes confirmed that island species, with smaller population sizes than their mainland counterparts, were more impacted by drift, which induces a load of weakly deleterious mutations in their genome. Indeed most of the genes studied here involved in the immune response do not show a statistically different pattern from control genes. Only MHC II genes, involved in the recognition of extracellular pathogens, showed a reduction in their non-synonymous polymorphism in island species. This response, which may be attributed to reduced selection pressures on these genes, could be associated with the suspected reduced parasitic communities on islands. The increased load of deleterious mutations as well as the potential relaxed selection pressures on MHC II support the reduced immune functions of island species, which could be added to the list of other convergent responses of the island syndrome.

## Acknowledgements

In Gabon, we thank the Director and the guides of the Lekedi Park, Marie Charpentier for her help in organizing the expedition, and Elisa Lobato and Alexandre Vaz for fieldwork assistance and outreach work. In São Tomé and Príncipe, we thank the Directorate of the Environment and the Department for Nature Conservation, its directors—Arlindo Carvalho and Victor Bonfim—Guilhermino, the Association Monte Pico, its president Luis Mário, and its members. Elisa Lobato, Philippe Perret, Octávio Veiga, Bikegila, and Yelli provided invaluable assistance in the field. Permissions for fieldwork were given by the authorities of São Tomé and Príncipe and Gabon (CENAREST authorization No. AR0053/12/MENESTFPRSCJS/CENAREST/CG/CST/CSAR). Permits for the Canary Islands were provided by the Regional Government (Ref.: 2012/0710), and the Cabildo of La Palma and Tenerife. In Montpellier, we thank the blue tit team (https://oreme.org/observation/ecopop/mesanges/) for the capture of the individuals used in this study. The analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform services. This research was conducted in the scope of the international twin-lab “LIA – Biodiversity and Evolution” between CIBIO (Portugal) and ISEM and CEFE-CNRS (France). This is ISEM publication n° ISEM 2022-223. Preprint version 4 of this article has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (https://doi.org/10.24072/pci.evolbiol.100153)

## Data, scripts, code, and supplementary information availability

Datasets, scripts, supplementary figures and texts are available on figshare (Barthe and Nabholz, 2022). The reads newly generated for this study have been deposited in the NCBI Sequence Read Archive under the bioproject PRJNA724656.

## Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article. BN is recommender for PCI evolutionary biology.

## Funding

This research was funded by the Labex CeMEB (project ISLAND IMMUNITY) for BN, CL and CD, the ANR (BirdIslandGenomic project, ANR-14-CE02-0002) for MB and BN, the National Geographic Society (Grant/Award Number:W251-12), the British Ecological Society (Grant/Award Number: 369/4558) to Elisa Lobato, RC and CD, the Portuguese Foundation for Science and Technology under the PTDC/BIA-EVL/29390/2017 “DEEP” Research Project for MM, RC and CL, and as a provider of structural funding to CIBIO (UIDB/50027/2021), the Spanish Ministry of Science, Innovation and Universities, the European Regional Development Fund (Ref.: PGC2018-097575-B-I00) for JCI, and European Union’s Horizon 2020 research and innovation programme under grant agreement 854248 for MM.

## References


Agudo R, Alcaide M, Rico C, Lemus JA, Blanco G, Hiraldo F, Donázar JA (2011) Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. *Molecular Ecology* **20**, 2329–2340. https://doi.org/10.1111/j.1365-294X.2011.05107.x.

Akira S (2003) Toll-like receptor signaling. *Journal of Biological Chemistry* **278**, 38105–38108. https://doi.org/10.1074/jbc.R300028200.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Innate immunity. In: *Molecular Biology of the Cell*. Garland Science, New York.

Aphalo PJ (2020) ggpmisc: Miscellaneous Extensions to "ggplot2"(R package version 0.3. 6).

Baeckens S, Van Damme R (2020) The island syndrome. *Current Biology* **30**, R338–R339. https://doi.org/10.1016/j.cub.2020.03.029

Barthe M, Doutrelant C, Covas R, Melo M, Illera JC, Tilak M-K, Colombier C, Leroy T, Loiseau C, Nabholz B (2022) Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome. bioRxiv, 2021.11.21.469450, ver. 4 peer-reviewed and recommended by Peer Community in Evolutionary Biology. https://doi.org/10.1101/2021.11.21.469450

Barthe M, Nabholz B (2022) Supplementary materials for "Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome". *Figshare*. https://doi.org/10.6084/m9.figshare.16954921.v7

Bates DM, Maechler M, Bolker B, Walker S (2012) Package 'lme4'. *CRAN R Found Stat Comput*.

Beadell JS, Atkins C, Cashion E, Jonker M, Fleischer RC (2007) Immunological change in a parasite-impoverished environment: divergent signals from four island taxa. *PLoS One* **2:e896**,. https://doi.org/10.1371/journal.pone.0000896.

Beadell JS, Ishtiaq F, Covas R, Melo M, Warren BH, Atkinson CT, Bensch S, Graves GR, Jhala YV, Peirce MA (2006) Global phylogeographic limits of Hawaii's avian malaria. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2935–2944. https://doi.org/10.1098/rspb.2006.3671.

Belasen AM, Bletz MC, S LD, Toledo LF, James TY (2019) Long-term habitat fragmentation is associated with reduced MHC IIB diversity and increased infections in amphibian hosts. *Frontiers in Ecology and Evolution* **6**,. https://doi.org/10.3389/fevo.2018.00236.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16**, 363–377. https://doi.org/10.1046/j.1420-9101.2003.00531.x.

Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual Review of Biochemistry* **59**, 253–288. https://doi.org/10.1146/annurev.bi.59.070190.001345.

Blondel J (2000) Evolution and ecology of birds on islands: trends and prospects. *Vie et Milieu/Life & Environment* 205–220. https://hal.sorbonne-universite.fr/hal-03186916

Bourret V, Gutiérrez López R, Melo M, Loiseau C (2021) Metabarcoding options to study eukaryotic endoparasites of birds. *Ecology and Evolution* **11**, 10821–10833. https://doi.org/10.1002/ece3.7748.

Boyce MS (1984) Restitution of gamma-and k-selection as a model of density-dependent natural selection. *Annual Review of Ecology and Systematics* **15**, 427–447. https://doi.org/10.1146/annurev.es.15.110184.002235

Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ (2013) InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Research* **41:D1228–D1233**,. https://doi.org/10.1093/nar/gks1147.

Buffalo V (2021) Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox (G Sella, Ed.). *ELife* **10**, e67509. https://doi.org/10.7554/eLife.67509.

Castellano D, James J, Eyre-Walker A (2018) Nearly neutral evolution across the Drosophila melanogaster genome. *Molecular Biology and Evolution* **35**, 2685–2694. https://doi.org/10.1093/molbev/msy164.

Chapman H JR, O H, AS K, RH C, RL W, J. (2016) The evolution of innate immune genes: purifying and balancing selection on β-defensins in waterfowl. *Molecular Biology and Evolution* **33**, 3075–3087. https://doi.org/10.1093/molbev/msw167.

Charlesworth J, Eyre-Walker A (2008) The McDonald–Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution* **25**, 1007–1015. https://doi.org/10.1093/molbev/msn005.

Chen J, Glémin S, Lascoux M (2020) From drift to draft: how much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? *Genetics* **214**, 1005–1018. https://doi.org/10.1534/genetics.119.302869.

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34:i884–i890**,. https://doi.org/10.1093/bioinformatics/bty560.

Corcoran P, Gossmann TI, Barton HJ, Slate J, Zeng K (2017) Determinants of the Efficacy of Natural Selection on Coding and Noncoding Variability in Two Passerine Species. *Genome Biology and Evolution* **9**, 2987–3007. https://doi.org/10.1093/gbe/evx213.

Cornuault J, Bataillard A, Warren BH, Lootvoet A, Mirleau P, Duval T, Milá B, Thébaud C, Heeb P (2012) The role of immigration and in-situ radiation in explaining blood parasite assemblages in an island bird clade. *Molecular Ecology* **21**, 1438–1452. https://doi.org/10.1111/j.1365-294X.2012.05483.x

Covas R (2012) Evolution of reproductive life histories in island birds worldwide. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1531–1537. https://doi.org/10.1098/rspb.2011.1785.

van Dijk A, Veldhuizen EJ, Haagsman HP (2008) Avian defensins. *Veterinary Immunology and Immunopathology* **124**, 1–18. https://doi.org/10.1016/j.vetimm.2007.12.006.

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**, 50–52. https://doi.org/10.1038/256050a0.

Doutrelant C, Paquet M, Renoult JP, Grégoire A, Crochet P-A, Covas R (2016) Worldwide patterns of bird colouration on islands. *Ecology Letters* **19**, 537–545. https://doi.org/10.1111/ele.12588.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences* **102**, 14338–14343. https://doi.org/10.1073/pnas.0504070102.

Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352. https://doi.org/10.1016/j.cell.2008.05.042.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756–760. https://doi.org/10.1038/nature11584.

Enard D, Cai L, Gwennap C, Petrov DA (2016) Viruses are a dominant driver of protein adaptation in mammals. *Elife* **5**, e12469. https://doi.org/10.7554/eLife.12469.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**, 610–618. https://doi.org/10.1038/nrg2146.

Fijarczyk A, Dudek K, Babik W (2016) Selective Landscapes in newt Immune Genes Inferred from Patterns of Nucleotide Variation. *Genome Biology and Evolution* **8**, 3417–3432. https://doi.org/10.1093/gbe/evw236.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity* **78**, 311–327. https://doi.org/10.1038/hdy.1997.46.

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. https://doi.org/10.1093/bioinformatics/bts565.

Garamszegi LZ (2006) The evolution of virulence and host specialization in malaria parasites of primates. *Ecology Letters* **9**, 933–940. https://doi.org/10.1111/j.1461-0248.2006.00936.x.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint* arXiv:1207.3907. https://doi.org/10.48550/arXiv.1207.3907.

Gonzalez-Quevedo C, Phillips KP, Spurgin LG, Richardson DS (2015) 454 screening of individual MHC variation in an endemic island passerine. *Immunogenetics* **67**, 149–162. https://doi.org/10.1007/s00251-014-0822-1.

Gonzalez-Quevedo C, Spurgin LG, Illera JC, Richardson DS (2015) Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology* **24**, 5852–5863. https://doi.org/10.1111/mec.13437.

Grant PR (1965) The adaptive significance of some size trends in island birds. *Evolution* 355–367.

https://doi.org/10.2307/2406446.
Grueber CE, Wallis GP, Jamieson IG (2013) Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology* **22**, 4470–4482. https://doi.org/10.1111/mec.12404.
Grueber CE, Wallis GP, Jamieson IG (2014) Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PloS One* **9**, e89632. https://doi.org/10.1371/journal.pone.0089632.
Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, Bernard A, Scornavacca C, Nabholz B, Haudry A, Dachary L, Galtier N, Belkhir K, Dutheil JY (2013) Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution. *Molecular Biology and Evolution* **30**, 1745–1750. https://doi.org/10.1093/molbev/mst097.
Hale KA, Briskie JV (2007) Decreased immunocompetence in a severely bottlenecked population of an endemic New Zealand bird. *Animal Conservation* **10**, 2–10. https://doi.org/10.1111/j.1469-1795.2006.00059.x.
Haller BC, Messer PW (2017) SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution* **34**, 230–240. https://doi.org/10.1093/molbev/msw211.
Hawley DM, Sydenstricker KV, Kollias GV, Dhondt AA (2005) Genetic diversity predicts pathogen resistance and cell-mediated immunocompetence in house finches. *Biology Letters* **1**, 326–329. https://doi.org/10.1098/rsbl.2005.0303.
Hill AV (1991) HLA associations with malaria in Africa: some implications for MHC evolution. 'Mol. Evol. Major Histocompat. Complex'. pp. 403–420. (Springer: Berlin, Heidelberg) https://doi.org/10.1007/978-3-642-84622-9_33.
Hochberg ME, Møller AP (2001) Insularity and adaptation in coupled victim–enemy associations. *Journal of Evolutionary Biology* **14**, 539–551. https://doi.org/10.1046/j.1420-9101.2001.00312.x.
Illera JC, Emerson BC, Richardson DS (2008) Genetic characterization, distribution and prevalence of avian pox and avian malaria in the Berthelot's pipit (Anthus berthelotii) in Macaronesia. *Parasitology Research* **103**, 1435–1443. https://doi.org/10.1007/s00436-008-1153-7.
Illera JC, Fernández-Álvarez Á, Hernández-Flores CN, Foronda P (2015) Unforeseen biogeographical patterns in a multiple parasite system in Macaronesia. *Journal of Biogeography* **42**, 1858–1870. https://doi.org/10.1111/jbi.12548.
Illera JC, Perera A (2020) Where are we in the host-parasite relationships of native land vertebrates in Macaronesia? *Ecosistemas* **29**, 1971. https://doi.org/10.7818/ECOS.1971
Institute B (2019) "Picard Toolkit", Broad institute, GitHub repository. *Picard Toolkit*.
Ishtiaq F, Clegg SM, Phillimore AB, Black RA, Owens IP, Sheldon BC (2010) Biogeographical patterns of blood parasite lineage diversity in avian hosts from southern Melanesian islands. *Journal of Biogeography* **37**, 120–132. https://doi.org/10.1111/j.1365-2699.2009.02189.x.
Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.
Kappes D, Strominger JL (1988) Human class II major histocompatibility complex genes and proteins. *Annual Review of Biochemistry* **57**, 991–1028. https://doi.org/10.1146/annurev.bi.57.070188.005015.
Kassambara A (2018) ggpubr:"ggplot2" based publication ready plots. *R Package Version* **01**, 7.
Kimura M (1962) On the Probability of Fixation of Mutant Genes in a Population. *Genetics* **47**, 713–719. https://doi.org/10.1093%2Fgenetics%2F47.6.713.
Klein J (1986) 'Natural history of the major histocompatibility complex.' (Wiley)
Kloch A, Wenzel MA, Laetsch DR, Michalski O, Bajer A, Behnke JM, Welc-Falęciak R, Piertney SB (2018) Signatures of balancing selection in toll-like receptor (TLRs) genes–novel insights from a free-living rodent. *Scientific Reports* **8**, 1–10. https://doi.org/10.1038/s41598-018-26672-2.
Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell N, Hunt GR, Ritchie MG, Rutz C, Wiberg RAW, Wolf JBW (2020) Purifying Selection in Corvids Is Less Efficient on Islands. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz233.
Kuznetsova A, Brockhoff PB, Christensen RH (2017) lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* **82**, 1–26. https://doi.org/10.18637/jss.v082.i13.
Laine VN, Gossmann TI, Schachtschneider KM, Garroway CJ, Madsen O, Verhoeven KJ, De Jager V, Megens

H-J, Warren WC, Minx P (2016) Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications* **7**, 1–9. https://doi.org/10.1038/ncomms10474.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N (2015) Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375. https://doi.org/10.1038/nature14181.

Lee KA (2006) Linking immune defenses and life history at the levels of the individual and the species. *Integrative and Comparative Biology* **46**, 1000–1015. https://doi.org/10.1093/icb/icl049.

Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443**, 50–55. https://doi.org/10.1038/nature05096.

Leroy T, Anselmetti Y, Tilak M-K, Bérard S, Csukonyi L, Gabrielli M, Scornavacca C, Milá B, Thébaud C, Nabholz B (2021) A bird's white-eye view on avian sex chromosome evolution. *Peer Community Journal* **1**, e63. https://doi.org/10.24072/pcjournal.70.

Leroy T, Rousselle M, Tilak M-K, Caizergues AE, Scornavacca C, Recuerda M, Fuchs J, Illera JC, De Swardt DH, Blanco G (2021) Island songbirds as windows into evolution in small populations. *Current Biology* **31**, 1303-1310. e4. https://doi.org/10.1016/j.cub.2020.12.040.

Levin II, Zwiers P, Deem SL, Geest EA, Higashiguchi JM, Iezhova TA, Jiménez-Uzcátegui G, Kim DH, Morton JP, Perlut NG, Renfrew RB, Sari EHR, Valkiunas G, Parker PG (2013) Multiple Lineages of Avian Malaria Parasites (Plasmodium) in the Galapagos Islands and Evidence for Arrival via Migratory Birds. *Conservation Biology* **27**, 1366–1377. https://doi.org/10.1111/cobi.12127.

Levy H, Fiddaman SR, Vianna JA, Noll D, Clucas GV, Sidhu JK, Polito MJ, Bost CA, Phillips RA, Crofts S (2020) Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Molecular Biology and Evolution* **37**, 1708–1726. https://doi.org/10.1093/molbev/msaa040.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:13033997*. https://doi.org/10.48550/arXiv.1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl* **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lindström KM, Foufopoulos J, Pärn H, Wikelski M (2004) Immunological investments reflect parasite abundance in island populations of Darwin's finches. *Proceedings of the Royal Society of London Series B: Biological Sciences* **271**, 1513–1519. https://doi.org/10.1098/rspb.2004.2752.

Lobato E, Doutrelant C, Melo M, Reis S, Covas R (2017) Insularity effects on bird immune parameters: A comparison between island and mainland populations in West Africa. *Ecology and Evolution* **7**, 3645–3656. https://doi.org/10.1002/ece3.2788.

Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, Lourenço JM, Galtier N (2013) Population genomics of the endangered giant Galapagos tortoise. *Genome Biology* **14**, R136. https://doi.org/10.1186/gb-2013-14-12-r136.

Loiseau C, Melo M, Lobato E, Beadell JS, Fleischer RC, Reis S, Doutrelant C, Covas R (2017) Insularity effects on the assemblage of the blood parasite community of the birds from the Gulf of Guinea. *Journal of Biogeography* **44**, 2607–2617. https://doi.org/10.1111/jbi.13060.

Lomolino MV (2005) Body size evolution in insular vertebrates: generality of the island rule. *Journal of Biogeography* **32**, 1683–1699. https://doi.org/10.1111/j.1365-2699.2005.01314.x.

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* **457**, 830–836. https://doi.org/10.1038/nature07893.

Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, Åkesson S, Bensch S (2017) Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1**, 155–168. https://doi.org/10.1002/evl3.15.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**,. https://doi.org/10.1186/2047-217X-1-18.

MacArthur RH, Wilson EO (1967) The theory of island biogeography. 'Theory Isl. Biogeogr.' (Princeton university press)

Maria L, Svensson E, Ricklefs RE (2009) Low diversity and high intra-island variation in prevalence of avian Haemoproteus parasites on Barbados, Lesser Antilles. *Parasitology* **136**, 1121–1131. https://doi.org/10.1017/S0031182009990497.

Martinez J, Vasquez RA, Venegas C, Merino S (2015) Molecular characterisation of haemoparasites in forest birds from Robinson Crusoe Island: is the Austral Thrush a potential threat to endemic birds? *Bird Conservation International* **25**, 139–152. https://doi.org/10.1017/S0959270914000227.

Matson KD (2006) Are there differences in immune function between continental and insular birds? *Proceedings Biological Sciences / The Royal Society* **273**, 2267–2274. https://doi.org/10.1098/rspb.2006.3590.

Matson KD, Beadell JS (2010) Infection, immunity, and island adaptation in birds.

Minias P, Pikus E, Whittingham LA, Dunn PO (2019) Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution* **11**, 17–28. https://doi.org/10.1093/gbe/evy253.

Mueller JC, Kuhl H, Timmermann B, Kempenaers B (2016) Characterization of the genome and transcriptome of the blue tit Cyanistes caeruleus: polymorphisms, sex-biased expression and selection signals. *Molecular Ecology Resources* **16**, 549–561. https://doi.org/10.1111/j.1944-9720.1986.tb01032.x.

Munoz-Mérida A, Viguera E, Claros MG, Trelles O, Pérez-Pulido AJ (2014) Sma3s: a three-step modular annotator for large sequence datasets. *DNA Research* **21**, 341–353. https://doi.org/10.1093/dnares/dsu001.

Nguyen L-T, Schmidt HA, Haeseler A, Minh BQ (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274. https://doi.org/10.1093/molbev/msu300.

Nieberding C, Morand S, Libois R, Michaux J (2006) Parasites and the island syndrome: the colonization of the western Mediterranean islands by Heligmosomoides polygyrus (Dujardin, 1845. *Journal of Biogeography* **33**, 1212–1222. https://doi.org/10.1111/j.1365-2699.2006.01503.x.

O'Connor EA, Hasselquist D, Nilsson J-Å, Westerdahl H, Cornwallis CK (2020) Wetter climates select for higher immune gene diversity in resident, but not migratory, songbirds. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20192675. https://doi.org/10.1098/rspb.2019.2675.

Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**, 263–286. http://www.jstor.org/stable/2097289.

Ortutay C, Vihinen M (2009) Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies. *Nucleic Acids Research* **37**, 622–628. https://doi.org/10.1093/nar/gkn982.

Padilla DP, Illera JC, Gonzalez-Quevedo C, Villalba M, Richardson DS (2017) Factors affecting the distribution of haemosporidian parasites within an oceanic island. *International Journal for Parasitology* **47**, 225–235. https://doi.org/10.1016/j.ijpara.2016.11.008.

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jønsson KA, Zhou Q (2021) Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular ecology resources* **21**, 263–286. https://doi.org/10.1111/1755-0998.13252.

Peona V, Weissensteiner MH, Suh A (2018) How complete are "complete" genome assemblies?—An avian perspective. *Molecular ecology resources* **18**, 1188–1195. https://doi.org/10.1111/1755-0998.12933.

Pérez-Rodríguez A, Ramírez Á, Richardson DS, Pérez-Tris J (2013) Evolution of parasite island syndromes without long-term host population isolation: Parasite dynamics in Macaronesian blackcaps Sylvia atricapilla. *Global Ecology and Biogeography* **22**, 1272–1281. https://doi.org/10.1111/geb.12084.

Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Willigen B, Maintainer R (2017) Package 'nlme. *Linear Nonlinear Mix Eff Models Version* **3**,. https://svn.r-project.org/R-packages/trunk/nlme/.

R Core Team (2018) R: A language and environment for statistical computing.

Rando JC, Alcover JA, Illera JC (2010) Disentangling Ancient Interactions: A New Extinct Passerine Provides

Insights on Character Displacement among Extinct and Extant Island Finches. *PLoS One* **5:e12956**,. https://doi.org/10.1371/journal.pone.0012956.

Ranwez V, Harispe S, Delsuc F, Douzery EJ (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* **6:e22594**,. https://doi.org/10.1371/journal.pone.0022594.

Recuerda M, Vizueta J, Cuevas-Caballé C, Blanco G, Rozas J, Milá B (2021) Chromosome-level genome assembly of the common chaffinch (Aves: Fringilla coelebs): a valuable resource for evolutionary biology. *Genome Biology and Evolution* **13**, evab034. https://doi.org/10.1093/gbe/evab034.

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, Marsden CD, Lohmueller KE, Wayne RK (2016) Genomic flatlining in the endangered island fox. *Current Biology* **26**, 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062.

Rogers RL, Slatkin M (2017) Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet* **13:e1006601**,. https://doi.org/10.1371/journal.pgen.1006601.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**, 939–946. https://doi.org/10.1101/gr.128124.111.

Rousselle M, Simion P, Tilak M-K, Figuet E, Nabholz B, Galtier N (2020) Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genetics* **16**, e1008668. https://doi.org/10.1371/journal.pgen.1008668.

Santonastaso T, Lighten J, Oosterhout C, Jones KL, Foufopoulos J, Anthony NM (2017) The effects of historical fragmentation on major histocompatibility complex class II β and microsatellite variation in the Aegean island reptile, Podarcis erhardii. *Ecology and Evolution* **7**, 4568–4581. https://doi.org/10.1002/ece3.3022.

She R, Chu JS-C, Uyar B, Wang J, Wang K, Chen N (2011) genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics* **27**, 2141–2143. https://doi.org/10.1093/bioinformatics/btr342.

Shultz AJ, Sackton TB (2019) Immune genes are hotspots of shared positive selection across birds and mammals. *Elife* **8**, e41815. https://doi.org/10.7554/eLife.41815.

Siewert KM, Voight BF (2020) BetaScan2: Standardized Statistics to Detect Balancing Selection Utilizing Substitution Data. *Genome Biology and Evolution* **12**, 3873–3877. https://doi.org/10.1093/gbe/evaa013.

Silva-Iturriza A, Ketmaier V, Tiedemann R (2012) Prevalence of avian haemosporidian parasites and their host fidelity in the central Philippine islands. *Parasitology International* **61**, 650–657. https://doi.org/10.1016/j.parint.2012.07.003.

Simion P, Belkhir K, François C, Veyssier J, Rink JC, Manuel M, Philippe H, Telford MJ (2018) A software tool 'CroCo'detects pervasive cross-species contamination in next generation sequencing data. *BMC Biology* **16**, 1–9. https://doi.org/10.1186/s12915-018-0486-7.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN (2015) Stable recombination hotspots in birds. *Science* **350**, 928–932. https://doi.org/10.1126/science.aad0843.

Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* **132**,. https://doi.org/10.1093%2Fgenetics%2F132.3.861.

Slowikowski K, Schep A, Hughes S, Lukauskas S, Irisson J-O, Kamvar ZN, Ryan T, Christophe D, Hiroaki Y, Gramme P (2018) Package ggrepel. *Autom Position Non-Overlapping Text Labels* **ggplot2**,.

Smeds L, Qvarnstrom A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. *Genome Research* gr-204669. https://doi.org/10.1101/gr.204669.116.

Spiess A-N, Spiess MA-N (2018) Package 'qpcR. 'Model Anal Real-Time PCRdata Httpscran R-Proj'. (OrgwebpackagesqpcRqpcR Pdf)

Spurgin LG, Van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC, Richardson DS (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**, 5213–5225. https://doi.org/10.1111/j.1365-294X.2011.05367.x.

Tange O (2018) GNU parallel 2018. https://doi.org/10.5281/zenodo.1146014

Van Riper III C, Van Riper SG, Goff ML, Laird M (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs* **56**, 327–344.

https://doi.org/10.2307/1942550.
Velová H, Gutowska-Ding MW, Burt DW, Vinkler M (2018) Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. *Mol Biol Evol* **35**, 2170–2184. https://doi.org/10.1093/molbev/msy119
Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S (2010) The genome of a songbird. *Nature* **464**, 757–762. https://doi.org/10.1038/nature08819.
Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J (2015) Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters* **18**, 200–217. https://doi.org/10.1111/ele.12398.
Welch JJ, Eyre-Walker A, Waxman D (2008) Divergence and Polymorphism Under the Nearly Neutral Theory of Molecular Evolution. *Journal of Molecular Evolution* **67**, 418–426. https://doi.org/10.1007/s00239-008-9146-9.
Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. https://doi.org/10.1007/978-3-319-24277-4_9.
Wikelski M, Foufopoulos J, Vargas H, Snell H (2004) Galápagos birds and diseases: invasive pathogens as threats for island species. *Ecology and Society* **9**,. http://www.jstor.org/stable/26267654.
Wolf JBW, Künstner A, Nam K, Jakobsson M, Ellegren H (2009) Nonlinear Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects Inference of Selection. *Genome Biol Evol* **1**, 308–319. https://doi.org/10.1093/gbe/evp030.
Zhang G, Parker P, Li B, Li H, Wang J (2012) The genome of Darwin's Finch (Geospiza fortis). GigaScience. https://doi.org/10.5524/100040.

# FRENCH SUMMARY

## Contexte général de la thèse

Ma thèse s'inscrit dans un contexte où notre compréhension du processus de spéciation s'améliore, mais où le concept d'espèce est encore débattu (Queiroz, 2007; Stankowski and Ravinet, 2021b). Même si le processus de spéciation est continu (Stankowski and Ravinet, 2021a), la taxonomie a hérité d'une vision discrète, et l'absence de consensus sur un seuil le long de ce continuum de spéciation contribue à la difficulté de délimiter les espèces taxonomiques (Zachos, 2018). Malgré cela, les espèces sont l'unité de base de la conservation et des études taxonomiques sont nécessaires de toute urgence, en particulier pour les grands mammifères, qui sont plus sensibles à l'extinction (Feijó and Brandão, 2022).

Les Xénarthres sont l'un des quatre principaux clades de mammifères placentaires, avec Afrotheres, Laurasiatheres et Euarchontoglires (Murphy et al., 2001a,b). Ils se sont diversifiés lorsque l'Amérique du Sud était géographiquement isolée des autres continents, et ont atteint plus de 200 genres décrits (Figure 1, McKenna and Bell, 1997). Au cours du pliocène, l'Amérique du Sud a été reconnectée à l'Amérique du Nord par l'émergence de l'isthme de Panama. Cela a permis des événements de dispersion entre les deux continents, connus sous le nom de Great American Biotic Interchange (GABI) et, notamment, les xénarthres ont pu coloniser avec succès l'Amérique centrale et l'Amérique du Nord (Figure 1, MacDonald et al., 2007; Patterson and Pascual, 1968. À la fin du Pléistocène, il y a environ 11 000 ans, une extinction majeure a touché de nombreux groupes de mammifères. La plupart des xénarthres se sont éteints, principalement les plus grandes formes terrestres (paresseux géants et glyptodontes ; Lyons et al., 2004). De cette grande diversité, il ne reste aujourd'hui que 14 genres et 39 espèces vivantes (Abba et al., 2015; Feijó et al., 2018; Miranda et al.,

2018; Wetzel et al., 2008).

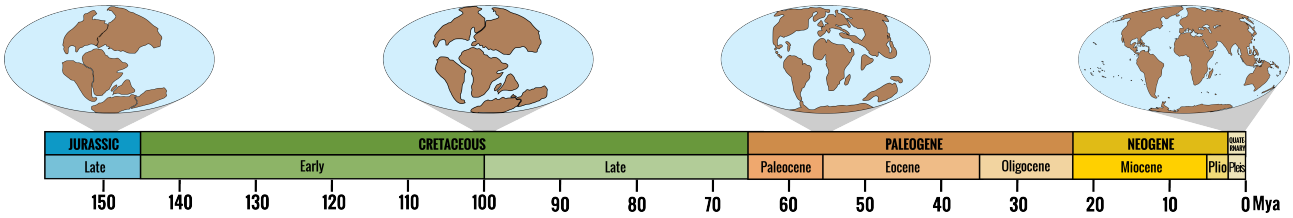


**Figure 1:** Cartes paléogéographiques de la terre (modifiées d'après Encyclopaedia Britannica et Meseguer et Condamine (2017)) sur l'échelle des temps géologiques remontant jusqu'au Jurassique il y a 150 Mya.

Ces 39 espèces, endémiques des régions néotropicales, sont le seul témoin vivant de la diversité passée des xénarthrans. Certaines d'entre elles ont une large distribution, incluant des formations géologiques ou des structures topologiques telles que les Andes, le bouclier guyanais ou l'Amérique centrale. Ces formations ont été identifiées comme des barrières biogéographiques reproductives pour de nombreuses espèces (Cortés-Ortiz et al., 2003; Esquerré et al., 2019; Fouquet et al., 2012; Gutiérrez-García and Vázquez-Domínguez, 2013; Redondo et al., 2008; Weir and Price, 2011), et pourraient également avoir été impliquées dans le processus de spéciation au sein des xénarthres. Certaines études ont notamment identifié des lignées divergentes au statut taxonomique incertain séparées par de telles formations au sein des tatous à neuf bandes, *Dasypus novemcinctus* (Arteaga et al., 2020; Billet et al., 2017; Feijó et al., 2018, 2019; Gibb et al., 2016; Hautier et al., 2017), paresseux à gorge brune, *Bradypus variegatus* (Ruiz-García et al., 2020), et les fourmiliers nains, *Cyclopes didactylus* (Coimbra et al., 2017; Miranda et al., 2018). Cette potentielle diversité cryptique est d'autant plus envisageable que 71,4 % des genres de xénarthres n'ont pas fait l'objet d'une étude taxonomique depuis 50 ans (Feijó and Brandão, 2022). Feijó and Brandão, 2022 suggèrent que la stabilité taxonomique pourrait s'expliquer par un manque d'études taxonomiques plutôt que par une bonne compréhension de leur diversité. En effet, si l'on considère les genres récemment étudiés, d'importants changements taxonomiques ont été mis en œuvre au niveau de la famille, du genre, du sous-genre, de l'espèce et de la sous-espèce (Feijó et al., 2018, 2019; Gibb et al., 2016; Miranda et al., 2018; Miranda et al., 2023). Cette activité taxonomique récente suggère qu'il reste beaucoup de travail à faire dans ce clade. La perception erronée des limites taxonomiques au sein de ce clade pourrait avoir induit une sous-estimation de la diversité actuelle de ce groupe, et de potentielles mauvaises appréciations quant à leur statut de conservation.

Outre ces incertitudes taxonomiques, certaines relations phylogénétiques sont encore difficiles à élucider. En effet, les relations phylogénétiques des xénarthres a longtemps été un défi en raison de leur morphologie particulière (Delsuc and Douzery, 2008). L'émergence des études moléculaires a permis une avancée sans précédent dans notre compréhension des relations entre les xénarthres.

Notamment, dans leurs premières études phylogénétiques moléculaires, Delsuc et al. (2001, 2002, 2003) ont résolu les relations intra-ordinales et confirmé les résultats morphologiques antérieurs. Möller-Krull et al., 2007 et Delsuc et al., 2012 ont inclus davantage d'espèces de xénarthres et ont donc représenté tous les genres reconnus. Toutefois, comme l'ont souligné Delsuc et al., 2003, les relations internes au sein des deux sous-familles Tolypeutinae et Euphractinae ne semblaient toujours pas résolues (nœuds internes courts et faibles valeurs de soutien), même avec différents types de données moléculaires (respectivement avec des séquences non codantes flanquantes de rétrotransposons et la concaténation d'exons nucléaires et de deux gènes mitochondriaux). Ce résultat était quelque peu attendu chez les Euphractinae, car les trois genres (*Euphractus*, *Zaedyus*, *Chaetophractus*) semblent très similaires et des études morphologiques antérieures ont soutenu plusieurs topologies possibles (Abrantes and Bergqvist, 2006; Engelmann, 1985; Gaudin and Wible, 2006; Patterson et al., 1989). Cependant, ceci est plus surprenant en ce qui concerne les Tolypeutinae pour lesquels les analyses morphologiques soutiennent fortement le regroupement des *Priodontes* et des *Cabassous* formant la tribu des Priodontini (Abrantes and Bergqvist, 2006; Cetica et al., 1998; Engelmann, 1985; Gaudin and Wible, 2006; McKenna and Bell, 1997). Gibb et al., 2016 ont fourni la première étude phylogénétique mitochondriale complète incluant toutes les espèces de xénarthres décrites à l'époque. Cette étude mitogénomique a clairement soutenu la paraphylie des Priodontini en regroupant *Cabassous* avec *Tolypeutes*, mais n'a pas permis de décrypter les relations au sein des Euphractinae. Les discordances avec les études morphologiques et entre les marqueurs moléculaires ont été suspectées d'être induites par d'anciens flux de gènes, un tri incomplet des lignées (ILS pour Incomplete Lineage Sorting) ou une convergence morphologique. Cependant, jusqu'à présent, les études phylogénétiques ont été largement limitées aux mitogénomes et à quelques marqueurs moléculaires nucléaires qui reflètent de manière incomplète l'histoire de l'évolution.

Grâce aux progrès récents et à la baisse des coûts du séquençage de l'ADN, il est désormais possible de produire des données génomiques afin d'examiner plus en détail la taxonomie et la phylogénie des xénarthres. Ainsi, en utilisant les génomes disponibles et les données génomiques de reséquençage récemment générées, ma thèse s'est focalisée sur les défis suivants : i) quel est le statut taxonomique des lignées/espèces de xénarthres récemment mises en évidence ? ii) La prise en compte des flux de gènes et de l'ILS permet-elle de mieux comprendre l'histoire évolutive des espèces de xénarthres ? iii) Quels facteurs (biogéographie, adaptation locale, démographie) ont pu contribuer à la spéciation et à la diversification de ce groupe ? iv) Le statut taxonomique des espèces est-il caractérisé par une différenciation génétique homogène parmi les mammifères ?

Pour répondre à ces questions, nous devions avoir accès aux données génétiques des espèces potentiellement menacées. Nous avons estimé qu'il était essentiel d'utiliser l'échantillonnage le

moins invasif pour les populations sauvages possible en recourant à la muséomique, aux animaux tués sur la route et aux données en libre accès, et nous avons relevé les défis que ces données représentaient.

# Chapitre 1 : La muséomique et la capture d'exons permettent de déchiffrer le complexe d'espèces du tatou à neuf bandes (*Dasypus novemcinctus*)

Dans la première partie, nous nous sommes concentrés sur l'espèce de xénarthre la plus répandue, le tatou à neuf bandes (*Dasypus novemcinctus*). Des études récentes ont suggéré une diversité cryptique avec des morphotypes distincts dont le statut taxonomique et la distribution géographique sont incertains (Arteaga et al., 2020; Billet et al., 2017; Feijó et al., 2018, 2019; Gibb et al., 2016; Hautier et al., 2017). Nous avons majoritairement utilisé des spécimens de musée, et séquencé 997 loci par capture d'exons. Cela nous a permis d'analyser 81 spécimens représentant l'ensemble de la distribution de ce complexe d'espèces (Figure 2c,d). Comme notre ensemble de données repose en grande partie sur des spécimens de musée, nous avons pris soin d'identifier et de corriger la contamination croisée potentielle et les erreurs de séquençage omniprésentes dans l'ADN dégradé. Nettoyer ces erreurs est un défi qui pourrait permettrait d'améliorer les délimitations des espèces. En effet, ces erreurs peuvent imiter le flux de gènes entre les lignées et donc brouiller les frontières entre les espèces. Inspirés par les méthodes d'analyse de l'ADN ancien (Green et al., 2008, 2010), nous avons soigneusement contrôlé nos librairies pour éviter toute contamination croisée. Pour cela, nous avons déterminé si l'ADN mitochondrial des individus séquencés soutenait des positions diagnostiques correspondant a plusieurs lignées distinctes. Enfin, nous avons mis en œuvre des méthodes de filtrage pour nettoyer efficacement l'ADN nucléaire basé sur la faible fréquence attendue par les lectures issues d'ADN contaminant ou supportant une erreur de génotypage.

Sur la base de mitogénomes complets et d'un millier de loci nucléaires capturés, nous avons reconstruit les relations phylogénétiques et utilisé de multiples approches pour délimiter les espèces et évaluer les échanges génétiques (Figure 2). Nous avons identifié des cas de discordance mito-nucléaire entre quatre lignées distinctes, mais nous avons constaté un flux génétique limité à leurs zones de contact (Figures 2 et 3). Une approche comparative a révélé une différenciation génétique comparable au sein du complexe du tatou à long nez par rapport aux espèces de *Dasypus* bien reconnues, ce qui confirme leur statut d'espèces distinctes. En accord avec les études moléculaires et

morphologiques précédentes, nos résultats fournissent une vue d'ensemble qui soutient fortement quatre espèces distinctes. Ceci inclut la première nouvelle espèce de tatou décrite au cours des 30 dernières années (*Dasypus guianensis* sp. nov.) endémique du bouclier guyanais.

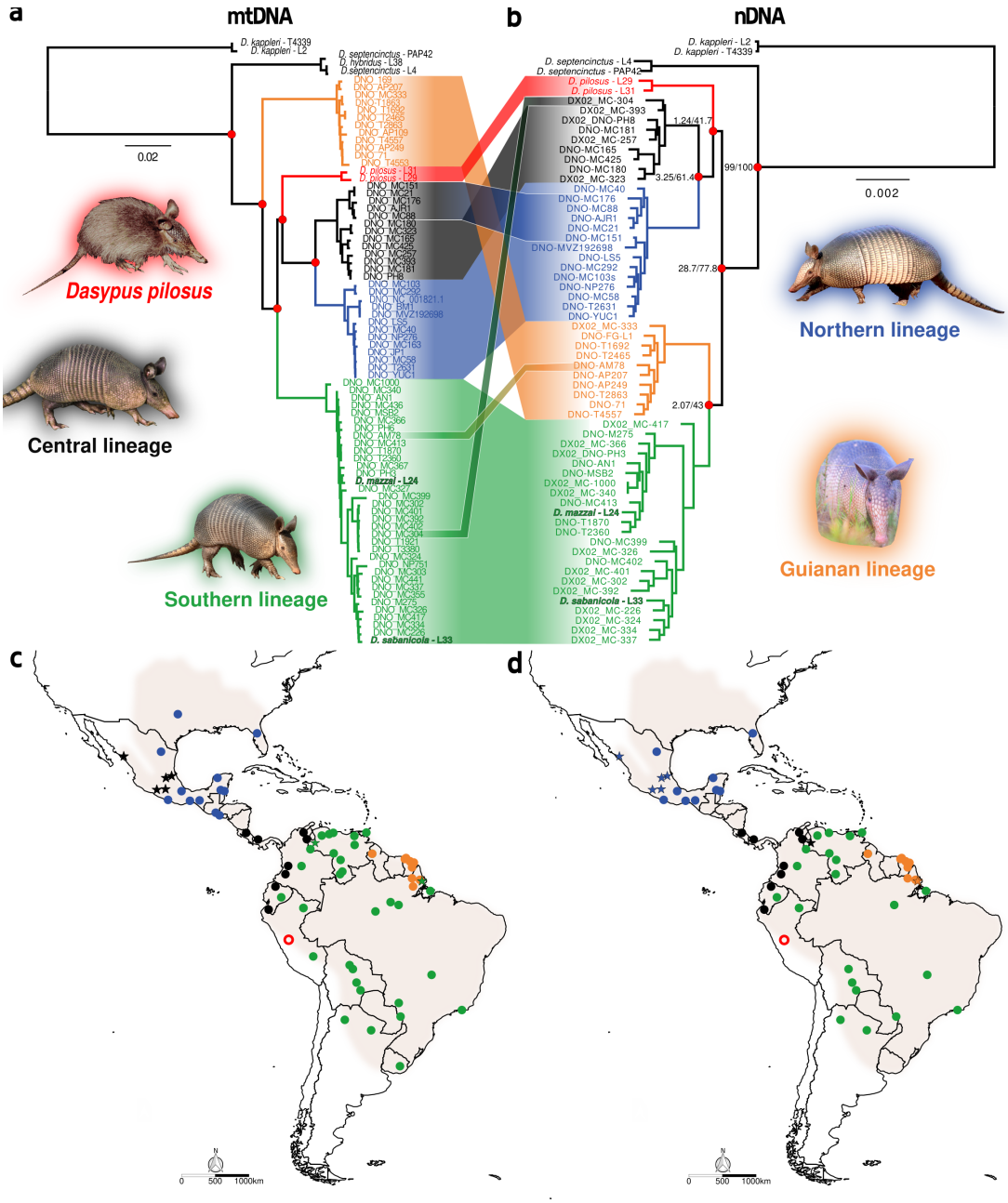


**Figure 2:** Relations phylogénétiques reconstruites par maximum de vraisemblance et enracinées en utilisant *Dasypus kappleri* à partir a) des mitogénomes de 81 individus, et b) des 832 loci nucléaires filtrés de 62 individus. Les cercles rouges aux nœuds indiquent le soutien bootstrap (BS = 100) pour les interrelations entre lignées et les étiquettes des nœuds représentent les facteurs de concordance des gènes et des sites (gCF/sCF). Les cartes représentent la distribution des individus en fonction de leur lignée et des données c) mitogénomiques et d) nucléaires. Les individus dont les lignées mito-nucléaires sont discordantes sont représentés par des étoiles et *D. pilosus* par des cercles ouverts. La distribution de *D. novemcinctus* est surlignée en gris. Crédits photos : A. Baertschi(xenarthrans.org), K. Miller, Andresiade, A. Reed (iNaturalist.org), et Q. Martinez.

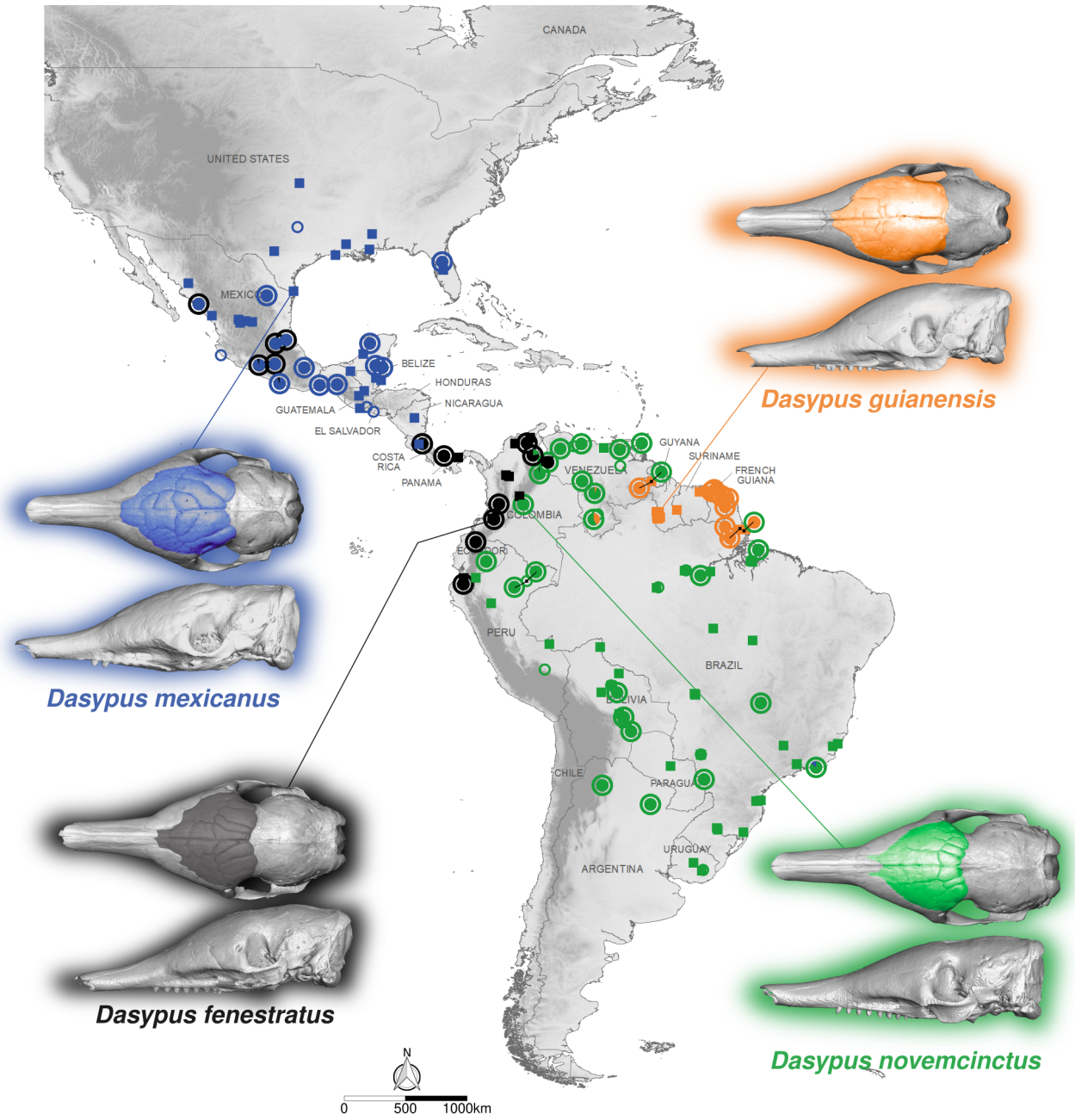


**Figure 3:** Carte de distribution et composition génétique des individus des quatre espèces reconnues : *Dasypus mexicanus* (lignée nord en bleu), *Dasypus fenestratus* (lignée centrale en noir), *Dasypus novemcinctus* (lignée sud en vert), et *Dasypus guianensis* sp. nov. (lignée guyanaise en orange). Les cercles extérieurs représentent les lignées mitochondriales et les diagrammes circulaires la probabilité d'affectation générée par les analyses de mélange basées sur les données nucléaires. Les cercles ouverts indiquent les individus pour lesquels seules les données mitogénomiques sont disponibles. Les carrés représentent les individus de Hautier et al., 2017 colorés en fonction de leur appartenance à un morphogroupe sur la base d'une Analyse Discriminante de la forme du crâne utilisant la morphométrie géométrique et pour lesquels aucune donnée génétique n'est disponible. Des reconstructions tridimensionnelles de crânes obtenues à partir de scans microCT de spécimens vouchers représentatifs sont présentées pour chaque espèce avec les sinus frontaux paranasaux et les évidements mis en évidence : USNM 33867(*Dasypus mexicanus*), AMNH 32356(*Dasypus fenestratus*), AMNH 136252(*Dasypus novemcinctus*), et ROM 32868(*Dasypus guianensis* sp. nov.).

# Chapitre 2 : Le séquençage du génome complet pour résoudre les incertitudes taxonomiques et phylogénétiques au sein des xénarthres

Les données moléculaires ont considérablement amélioré notre compréhension de l'histoire évolutive des xénarthres (Delsuc and Douzery, 2008). Cependant, un certain nombre de statuts taxonomiques sont encore incertains et certaines relations phylogénétiques ne sont toujours pas résolues. En particulier, des études antérieures ont révélé des lignées distinctes au sein de *Bradypus variegatus* et de *Cyclopes didactylus* dont le statut taxonomique reste à confirmer (Coimbra et al., 2017; Miranda et al., 2018; Ruiz-García et al., 2020). D'autre part, des discordances phylogénétiques ont été constatées au sein des Tolypeutinae et des Euphractinae (Abba et al., 2015; Delsuc and Douzery, 2008; Delsuc et al., 2003, 2012; Gibb et al., 2016; Möller-Krull et al., 2007). Toutefois, la contribution de l'ILS et/ou du flux de gènes doivent encore être évalués.

Nous avons reséquencé 19 mitogénomes et 72 génomes complets qui ont contribué à constituer les ensembles de données les plus exhaustifs de Xenarthra, composés de 261 mitogénomes couvrant 37 espèces actuelles et 7 espèces éteintes, et de 94 génomes complets couvrant 36 espèces actuelles sur les 42 espèces de xénarthres actuellement reconnues. Dans un premier temps, nous avons utilisé des méthodes de découverte d'espèces basées sur les relations phylogénétiques mitochondriales (GMYC et bPTP). Celles ci ont fourni une hypothèse d'espèce qui a été évaluée de manière plus approfondie à partir des données génomiques. Nous avons examiné la différenciation génétique (GDI, Allio et al., 2021) et la divergence (Da, Dxy) par paire à l'échelle du génome entre les espèces étroitement apparentées au sein de tous les genres de xénarthres. Cela nous a permis de fournir un cadre comparatif pour délimiter les lignées au statut taxonomique incertain (Figure 4). Ainsi, pour être cohérent avec la taxonomie actuelle, cette approche comparative à l'échelle du génome suggère de revalider *B. ephippiger* et *B. griseus*, de regrouper *B. pygmaeus* avec *B. griseus* et de décrire une nouvelle espèce de fourmilier, *Cyclopes* sp. (Figure 4). Cette approche suggère également que les espèces d'Ephractinae devraient toutes appartenir au genre *Euphractus* (Figure 3). Nous avons évalué la diversité génétique et la consanguinité de ces espèces qui ont alarmé sur la faible diversité génomique de quatre espèces de *Choloepus hoffmanni*, *Dasypus pilosus*, *Chlamyphorus truncatus* et *Bradypus ephippiger*. Ces espèces doivent être une préoccupation majeure pour l'évaluation de leur statut de conservation. D'autre part, l'évaluation de l'histoire démographique des xénarthres révèle un déclin général de la taille efficace des populations ces derniers millions d'années.

Sur la base de cette révision taxonomique, nous avons reconstruit la phylogénie des xénarthres,

calibrée dans le temps, la plus complète à ce jour qui corrobore les résultats précédents (Gibb et al., 2016, Figure 5). Enfin, nous avons exploré la contribution du trie de lignées incomplet et du flux de gènes dans les discordances de topologies de gènes chez les xénarthres. En utilisant une approche basée sur les différences de temps de coalescence attendues par ces deux mécanismes (i.e le flux de gène produit des temps de coalescence plus courts) nous avons pu détecter que l'ILS et le flux de gènes ont contribué à des topologiques discordantes dans la phylogénie des xénarthres (Figure 6). Ces preuves d'hybridation soulignent la prévalence du flux de gènes post-spéciation chez les xénarthres et suggèrent que ces événements de spéciation se sont produits avec des zones de contact.

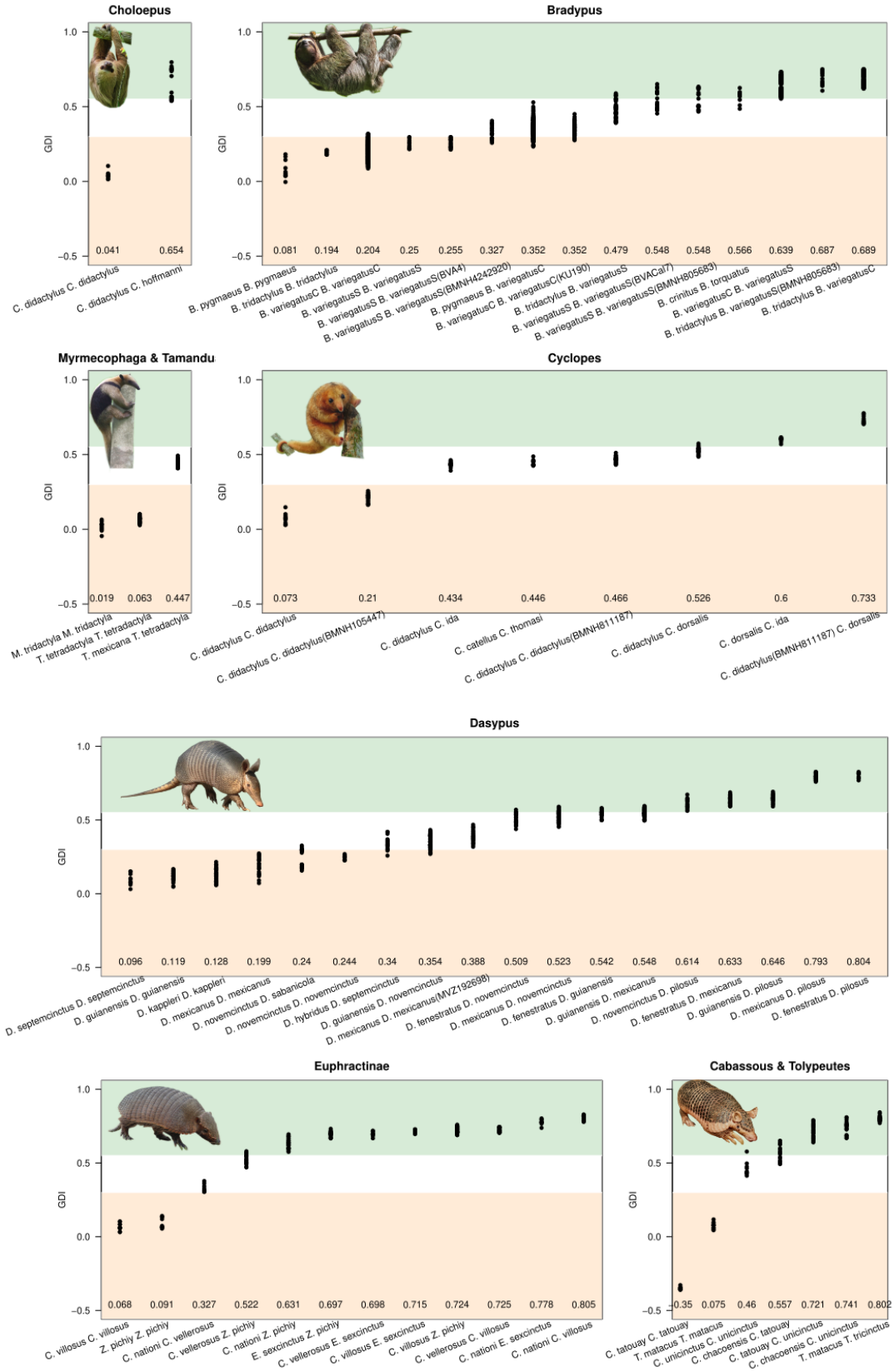


**Figure 4:** Différenciation génétique par paire entre les espèces de xénarthres estimée à l'aide de l'indice de différenciation génétique (GDI ; Allio et al., 2021). Chaque point représente le GDI moyen estimé à partir de 10 régions de 100kb échantillonnées au hasard dans le génome d'une paire d'individus. La valeur moyenne du GDI est indiquée sous chaque graphique. Plus le GDI est élevé, plus la différenciation génétique est importante. Les bandes de couleur sont arbitraires et visent à faciliter les comparaisons entre les différents genres.

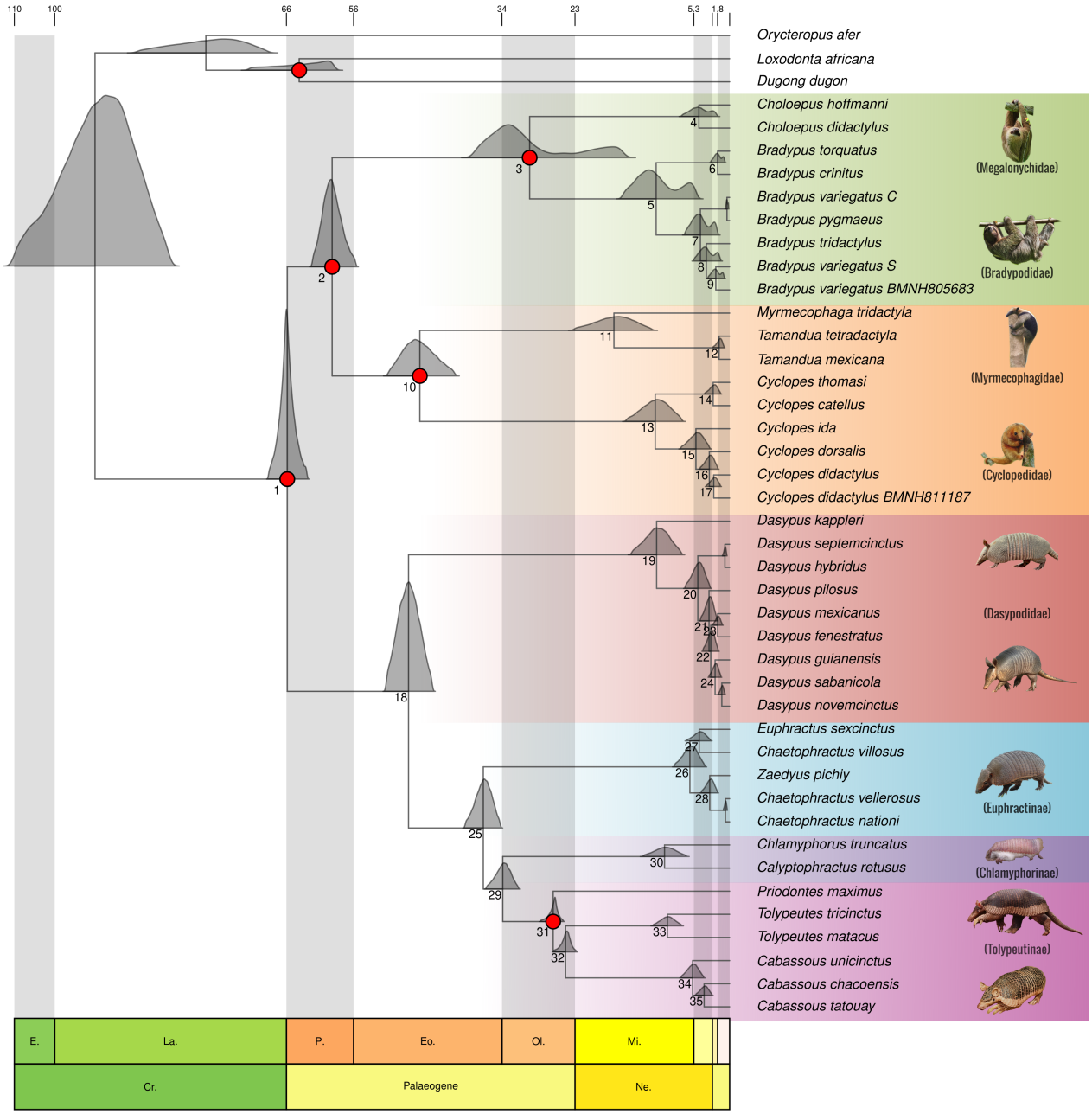


**Figure 5:** Arbre des temps de divergence de 40 espèces ou sous-espèces de xénarthres basé sur 100 gènes BUSCO et reconstruit à l'aide de MCMCTree avec une horloge relachée avec un taux autocorrélé et le modèle de substitution HKY + Γ. La distribution des nœuds représente les intervalles de crédibilité à 95% autour des estimations de l'âge moyen. Les cercles rouges représentent les nœuds utilisés comme contraintes d'étalonnage. Les numéros de nœuds renvoient au tableau Node_estimation détaillant les estimations du temps de divergence.

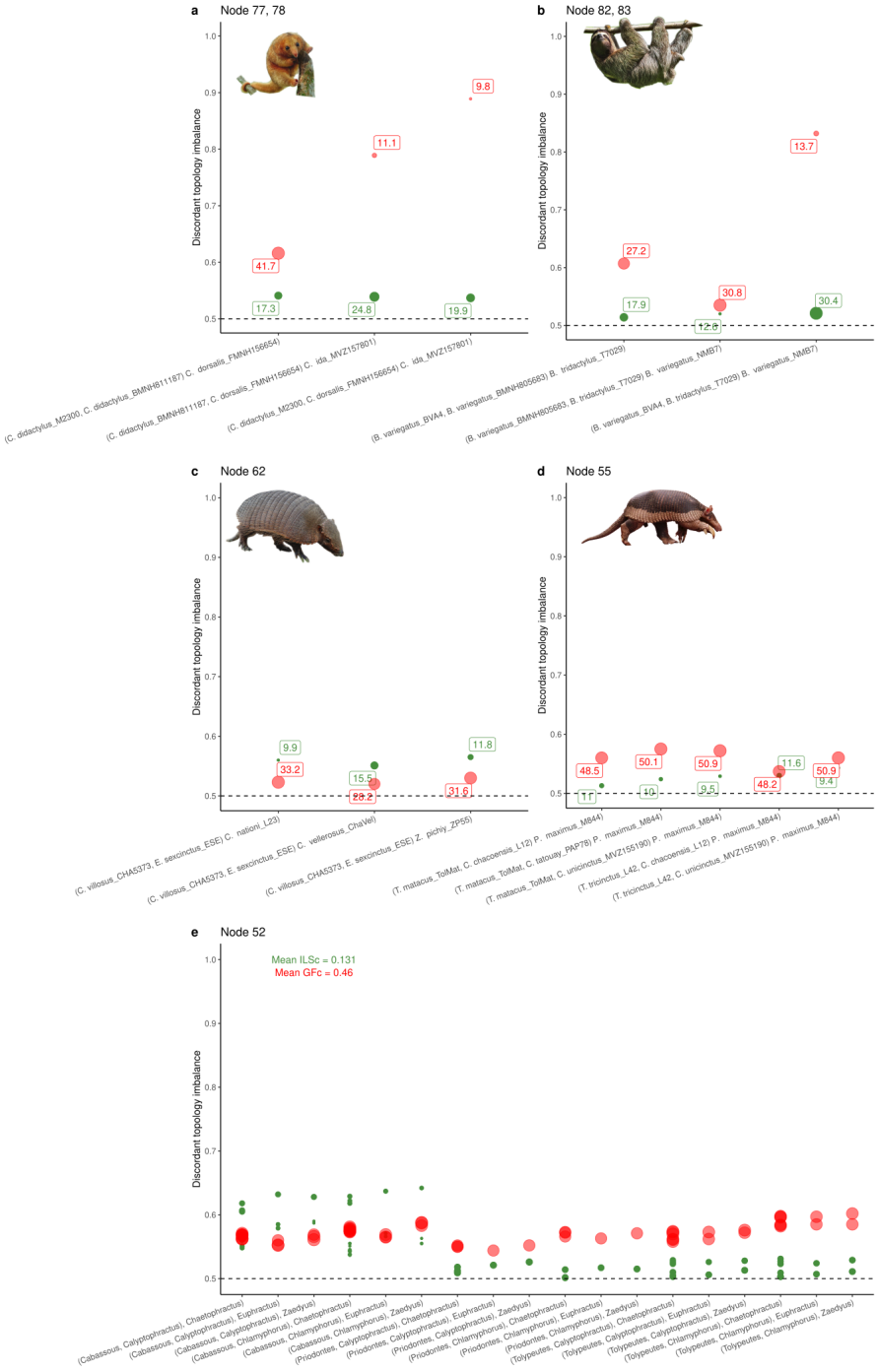


**Figure 6:** Démêler le triage incomplet des lignées (ILS) en rouge et le flux de gènes (GF) en vert en utilisant Aphid pour les nœuds présentant des conflits topologiques : a) *Bradypus* spp. b) *Cyclopes* spp. c) Euphractinae, d) *Dasypus* spp. e) Tolypeutinae, f) Chlamyphoridae. La taille des points représente la proportion d'ILS et de GF dans les conflits phylogénétiques (les étiquettes détaillent ces valeurs) et la position sur l'axe Y illustre la proportion de déséquilibre d'une topologie discordante (((A,C),B) ou ((B,C),A)). Les numéros de nœuds sont illustrés dans la figure S11 a.

## Synthèse générale et perspectives

Cette thèse a mené la révision moléculaire la plus complète des xénarthres, illustrant la puissance de la muséomique pour révéler la diversité des espèces cryptiques. En englobant un ensemble presque exhaustif de données génomiques de Xenarthra, nous avons revalidé *Dasypus mexicanus*, *Dasypus fenestratus*, *Bradypus ephippiger* et *Bradypus griseus*, regroupé *B. pygmaeus*, en danger critique d'extinction, avec *B. griseus*, et révélé de nouvelles espèces de tatou, *Dasypus guianensis*, et de fourmilier nain, *Cyclopes* sp. Grâce à ce cadre taxonomique révisé, nous avons reconstruit la phylogénie calibrée dans le temps la plus complète des Xenarthra, qui met en lumière leur histoire évolutive. En démêlant la contribution du tri de lignées incomplet et et du flux de gènes dans les signaux de topologies discordantes de la phylogénie des xénarthres, nous avons mis en évidence de multiples événements de flux de gènes suggérant une spéciation avec des contacts entre les lignées. En discussion, nous avons exploré plus en détail les facteurs influençant la spéciation chez les xénarthres. Nos résultats suggèrent notamment que les Andes colombiennes et vénézuéliennes ont favorisé la spéciation de ce clade en séparant de nombreuses paires d'espèces de xénarthres. Enfin, nos résultats ont contribué à identifier les directions futures des études taxonomiques et a suggéré, sur la base de leur faible diversité génétique, que l'évaluation de l'état de conservation de *Choloepus hoffmanni*, *Dasypus pilosus*, *Chlamyphorus truncatus*, et *Bradypus ephippiger* soit une priorité.

## References

Abba, A. M. et al. (2015). "Systematics of hairy armadillos and the taxonomic status of the Andean hairy armadillo (Chaetophractus nationi)". In: *Journal of Mammalogy* 96.4. ISBN: 1545-1542 Publisher: Oxford University Press US, pp. 673–689.

Abrantes, E. A. L. and L. P. Bergqvist (2006). "Proposta filogenética para os Dasypodidae (Mammalia: Cingulata)". In: *Paleontologia de vertebrados: grandes temas e contribuições científicas (V. Gallo, PM Brito, HMA Silva, and FJ Figueiredo, eds.). Interciência Ltda., Rio de Janeiro, Brazil*, pp. 261–274.

Allio, R. et al. (2021). "High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox". In: *Elife* 10. Publisher: eLife Sciences Publications Limited, e63167. ISSN: 2050-084X.

Arteaga, M. C. et al. (2020). "Conservation genetics, demographic history, and climatic distribution of the nine-banded armadillo (Dasypus novemcinctus): an analysis of its mi-

tochondrial lineages". In: *Conservation genetics in mammals: integrative research using novel approaches*. ISBN: 3030333337 Publisher: Springer, pp. 141–163.

Billet, G. et al. (2017). "The hidden anatomy of paranasal sinuses reveals biogeographically distinct morphotypes in the nine-banded armadillo (Dasypus novemcinctus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3593. ISSN: 2167-8359.

Cetica, P. D. et al. (1998). "Evolutionary sperm morphology and morphometry in armadillos". In: *Journal of submicroscopic cytology and pathology* 30.2, pp. 309–314. ISSN: 1122-9497.

Coimbra, R. T. F. et al. (2017). "Phylogeographic history of South American populations of the silky anteater *Cyclopes didactylus* (Pilosa: Cyclopedidae)". In: *Genetics and Molecular Biology* 40. Publisher: Sociedade Brasileira de Genética, pp. 40–49. ISSN: 1415-4757, 1678-4685. DOI: 10.1590/1678-4685-GMB-2016-0040.

Cortés-Ortiz, L. et al. (2003). "Molecular systematics and biogeography of the Neotropical monkey genus, Alouatta". In: *Molecular Phylogenetics and Evolution* 26.1, pp. 64–81. ISSN: 1055-7903. DOI: 10.1016/S1055-7903(02)00308-1.

Delsuc, F. and E. J. Douzery (2008). "Recent advances and future prospects in xenarthran molecular phylogenetics". In: *The biology of the Xenarthra* 11. Publisher: University Press of Florida Gainesville.

Delsuc, F., M. J. Stanhope, and E. J. P. Douzery (2003). "Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes". In: *Molecular Phylogenetics and Evolution* 28.2, pp. 261–275. ISSN: 1055-7903. DOI: 10.1016/S1055-7903(03)00111-8.

Delsuc, F. et al. (2012). "Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos". In: *Molecular Phylogenetics and Evolution* 62.2, pp. 673–680. ISSN: 1055-7903. DOI: https://doi.org/10.1016/j.ympev.2011.11.008.

Engelmann, G. (1985). "The phylogeny of Xenarthra". In: *The evolution and ecology of armadillos, sloths, and vermilinguas*. Publisher: Smithsonian Institution Press, pp. 51–64.

Esquerré, D. et al. (2019). "How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae)". In: *Evolution* 73.2. _eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.13657, pp. 214–230. ISSN: 1558-5646. DOI: 10.1111/evo.13657.

Feijó, A. and M. V. Brandão (2022). *Taxonomy as the first step towards conservation: an appraisal on the taxonomy of medium-and large-sized Neotropical mammals in the 21st century*. Vol. 39. Publication Title: Zoologia (Curitiba). SciELO Brasil. ISBN: 1984-4670.

Feijó, A., B. D. Patterson, and P. Cordeiro-Estrela (2018). "Taxonomic revision of the long-nosed armadillos, Genus Dasypus Linnaeus, 1758 (Mammalia, Cingulata)". In: *PLoS One* 13.4. Publisher: Public Library of Science San Francisco, CA USA, e0195084. ISSN: 1932-6203.

Feijó, A. et al. (2019). "Phylogeny and molecular species delimitation of long-nosed armadillos (Dasypus: Cingulata) supports morphology-based taxonomy". In: *Zoological Journal of the Linnean Society* 186.3. Publisher: Oxford University Press UK, pp. 813–825. ISSN: 0024-4082.

Fouquet, A. et al. (2012). "Multiple quaternary refugia in the eastern Guiana Shield revealed by comparative phylogeography of 12 frog species". In: *Systematic Biology* 61.3. ISBN: 1076-836X Publisher: Oxford University Press, p. 461.

Gaudin, T. J. and J. R. Wible (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*, Chicago (IL). University of Chicago Press, pp. 153–198.

Gibb, G. C. et al. (2016). "Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans". In: *Molecular Biology and Evolution* 33.3, pp. 621–642. ISSN: 0737-4038. DOI: 10.1093/molbev/msv250.

Green, R. E. et al. (2008). "A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing". In: *Cell* 134.3. ISBN: 0092-8674 Publisher: Elsevier, pp. 416–426.

Green, R. E. et al. (2010). "A draft sequence of the Neandertal genome". In: *Science* 328.5979. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 710–722.

Gutiérrez-García, T. A. and E. Vázquez-Domínguez (2013). "Consensus between genes and stones in the biogeographic and evolutionary history of Central America". In: *Quaternary Research* 79.3. ISBN: 0033-5894 Publisher: Cambridge University Press, pp. 311–324.

Hautier, L. et al. (2017). "Beyond the carapace: skull shape variation and morphological systematics of long-nosed armadillos (genus Dasypus)". In: *PeerJ* 5. Publisher: PeerJ Inc., e3650. ISSN: 2167-8359.

Lyons, S. K., F. A. Smith, and J. H. Brown (2004). "Of mice, mastodons and men: human-mediated extinctions on four continents". In: *Evolutionary Ecology Research* 6.3. ISBN: 1522-0613 Publisher: Evolutionary Ecology, Ltd., pp. 339–358.

MacDonald, G. H., S. F. Vizcaíno, and M. S. Bargo (2007). "Skeletal anatomy and the fossil history of the Vermilingua". In: *The Biology of the Xenarthra/Vizcaíno, Sergio Fabián; Loughry, WJ*.

McKenna, M. C. and S. K. Bell (1997). *Classification of mammals: above the species level*. Columbia University Press. ISBN: 0-231-52853-1.

Miranda, F. R. et al. (2018). "Taxonomic review of the genus Cyclopes Gray, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species". In: *Zoological Journal of the Linnean Society* 183.3, pp. 687–721. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlx079.

Miranda, F. R. et al. (2023). "Taxonomic revision of maned sloths, subgenus Bradypus (Scaeopus), Pilosa, Bradypodidae, with revalidation of Bradypus crinitus Gray, 1850". In: *Journal of Mammalogy* 104.1. ISBN: 0022-2372 Publisher: Oxford University Press US, pp. 86–103.

Möller-Krull, M. et al. (2007). "Retroposed Elements and Their Flanking Regions Resolve the Evolutionary History of Xenarthran Mammals (Armadillos, Anteaters, and Sloths)". In: *Molecular Biology and Evolution* 24.11, pp. 2573–2582. ISSN: 0737-4038. DOI: 10.1093/molbev/msm201.

Murphy, W. J. et al. (2001a). "Molecular phylogenetics and the origins of placental mammals". In: *Nature* 409.6820. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6820 Primary_atype: Research Publisher: Nature Publishing Group, pp. 614–618. ISSN: 1476-4687. DOI: 10.1038/35054550.

Murphy, W. J. et al. (2001b). “Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics”. In: *Science* 294.5550. Publisher: American Association for the Advancement of Science Section: Report, pp. 2348–2351. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1067179.

Patterson, B. and R. Pascual (1968). “The fossil mammal fauna of South America”. In: *The Quarterly Review of Biology* 43.4. ISBN: 0033-5770 Publisher: Stony Brook Foundation, Inc., pp. 409–451.

Patterson, B., W. Segall, and W. D. Turnbull (1989). “ear region in Xenarthrans (= Edentata: Mammalia)”. In: Publisher: Field Museum of Natural History.

Queiroz, K. de (2007). “Species concepts and species delimitation”. In: *Systematic Biology* 56.6. Publisher: Society of Systematic Zoology, pp. 879–886. ISSN: 1076-836X.

Redondo, R. A. F. et al. (2008). “Molecular systematics of the genus Artibeus (Chiroptera: Phyllostomidae)”. In: *Molecular Phylogenetics and Evolution* 49.1, pp. 44–58. ISSN: 1055-7903. DOI: 10.1016/j.ympev.2008.07.001.

Ruiz-García, M. et al. (2020). “Molecular phylogenetics of Bradypus (three-toed sloth, Pilosa: Bradypodidae, Mammalia) and phylogeography of Bradypus variegatus (brown-throated three-toed sloth) with mitochondrial gene sequences”. In: *Journal of Mammalian Evolution* 27.3. Publisher: Springer, pp. 461–482. ISSN: 1573-7055.

Stankowski, S. and M. Ravinet (2021a). “Defining the speciation continuum”. In: *Evolution* 75.6. ISBN: 0014-3820 Publisher: Wiley Online Library, pp. 1256–1273.

— (2021b). “Quantifying the use of species concepts”. In: *Current Biology* 31.9. ISBN: 0960-9822 Publisher: Elsevier, R428–R429.

Weir, J. T. and M. Price (2011). “Andean uplift promotes lowland speciation through vicariance and dispersal in Dendrocincla woodcreepers”. In: *Molecular Ecology* 20.21. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-294X.2011.05294.x, pp. 4550–4563. ISSN: 1365-294X. DOI: 10.1111/j.1365-294X.2011.05294.x.

Wetzel, R. M. et al. (2008). “Order Cingulata”. In: *Mammals of South America* 1. Publisher: Chicago University of Chicago Press, pp. 128–156.

Zachos, F. E. (2018). “Mammals and meaningful taxonomic units: the debate about species concepts and conservation”. In: *Mammal Review* 48.3. Publisher: Wiley Online Library, pp. 153–159. ISSN: 0305-1838.

# REMERCIEMENTS

La science repose avant tout sur le fruit d'échanges, de brainstorming, de transmissions de connaissances et de matériel, et d'apprentissage au travers d'une diversité de personnes et de rencontres qui génèrent ce magnifique réseau de *connaissance* (où les deux sens de ce terme s'appliquent merveilleusement bien). J'aimerais donc prendre le temps de remercier toutes les personnes qui ont nourri mes réflexions et qui ont permis à ce projet de thèse de se concrétiser.

**Merci pour cette aventure scientifique,**

Fred, c'est avec toi que j'ai découvert le travail dans le monde de la génomique il y a 5 ans. Tu m'as transmis le goût pour ce modèle biologique fascinant que tu connais tant. Merci d'avoir accepté de te lancer dans l'encadrement de ce projet de thèse lorsque je suis venu toquer en 2020 pour te demander de poursuivre en thèse et de m'avoir proposé de défendre devant le CEBA ce projet hybride (que tu avais ajusté à mes thématiques d'intérêts). À ce moment-là, j'étais loin de m'imaginer que je me prendrais autant de passion pour cette "deuxième partie" que tu m'avais proposé, merci de m'avoir mis sur le chemin qui a initié mon engouement pour comprendre ce processus de spéciation. Merci pour ta confiance et pour la liberté avec laquelle tu m'as laissé m'approprier ce sujet (sur lequel tu travaillais pourtant depuis 20 ans).

Il y a certains enseignants dans notre parcours qui nous marquent particulièrement tant ils nous ont inspirés et passionnés. Benoît, tu m'as transmis le virus pour la génomique évolutive à travers tes cours captivants, qui m'ont orienté vers cette voie scientifique qui me fascine tant. Merci pour ta générosité intellectuelle qui m'a tant apporté depuis le master, pour ton enthousiasme, ta curiosité scientifique et ta bienveillance indéfectible. Merci de toujours prendre le temps (même quand tu ne l'as pas) pour discuter, expliquer, brainstormer. Qu'il est agréable de ressortir de ton bureau avec

des questions en moins et tellement d'idées en plus.

Je voudrais remercier ceux qui ont contribué à financer ce projet de thèse. Merci au LabEx du Centre d'Étude de la Biodiversité Amazonienne ainsi qu'au Conseil européen de la recherche ( European Research Council; ERC), au travers du projet ConvergeAnt porté par Frédéric Delsuc, d'avoir soutenu ce projet de thèse et de m'avoir fait confiance pour le mener à bien. Merci Fred d'avoir soutenu financièrement l'ensemble des manip' de biologie moléculaire que j'ai pu effectuer, et d'avoir permis le séquençage d'un tel jeu de données.

Merci à ceux qui ont contribué à la matière première de ma thèse, des échantillons collectés durant les 130 dernières années par de nombreux scientifiques/collaborateurs passionnés qui ont, pour certains, dévoué leur carrière à ce travail. Je suis aussi reconnaissante de toutes les personnes qui ont géré ces collections depuis tout ce temps et dont l'organisation irréprochable nous permet de valoriser ces tissus. Merci à toutes ces personnes d'avoir permis que toutes ces données soient générées et de nous permettre d'un peu mieux comprendre ce groupe fascinant et encore mystérieux sur de nombreux points.

Merci à ceux qui ont permis de générer les données de séquençage issues de ces échantillons, Amandine, Fabby, Rémi, Marie-Ka, Fred, ainsi que le personnel des différentes plateformes de séquençage (MGX, Novogene, Arbore Bio Science). Marie-Ka, merci de m'avoir encadrée, guidée, conseillée avec autant de positivité et de confiance dans cette aventure qu'a représenté le séquençage de ces échantillons dégradés. Merci pour ta passion communicative pour la biologie moléculaire et de m'avoir généreusement transmis le savoir issu de tes années d'expérience.

Merci à tous ceux qui ont nourri les réflexions scientifiques autour de ce projet, et qui m'ont guidé à la réalisation des analyses qui composent ce travail. Notamment, merci Rémi de m'avoir initié à la bioinfo' il y a maintenant 5 ans, et guidée pour mes premiers scriptes, ta maîtrise de cet outil m'a énormément motivé et inspiré. Je remercie également les membres de l'équipe Phylogénie et Évolution Moléculaires et plus largement l'ISEM pour ce cadre scientifique stimulant et enrichissant. Merci aux co-auteurs ainsi qu'aux reviewers qui se sont impliqués dans ce projet et dont les remarques pertinentes ont permis de produire un travail plus complet et rigoureux.

Matthieu, merci pour le travail que tu as réalisé pendant ce stage et de m'avoir permis de découvrir le plaisir d'encadrer, de transmettre et conseiller.

Je voudrais également remercier les membres de mes différents comités de suivi de thèse, Marie Sémon, Pierre-Olivier Antoine, Vincent Ranwez, Tristan Lefébure, Pierre-Alexandre Gagnaire,

Mélanie Debiais-Thibaud et Carole Smadja de m'avoir accompagné au cours de ce projet de thèse, pour ces échanges stimulants, vos conseils éclairés, ainsi que pour votre soutien scientifique et personnel.

Enfin, j'aimerais remercier les membres de mon jury, Antoine Fouquet, Violaine Nicolas-Colin, Carole Smadja, Christelle Fraisse et Nicolas Galtier d'avoir accepté d'évaluer mon travail. J'espère que ce manuscrit vous aura intéressé et qu'il suscitera de plus amples discussions.

**Merci à ceux qui m'ont inspiré et m'inspirent encore,**

J'aimerais commencer par remercier les enseignants, Didier Forcioli, Pierrick Labbé, Mélanie Debiais-Thibaud, Benoit Nabholz et Emmanuel Douzery, qui tout au long de mon parcours universitaire m'ont motivée et encouragée à suivre cette voie qui m'a emmenée à cette thèse.

Merci à tous ceux qui contribuent à faire de la science un monde meilleur, aux inspirants acteurs (militants ?) d'une science plus éthique (PCI, PCJ, Camille Noûs, Daphnee), plus verte (Labo 1.5, les groupes de Transition Écologique qui voient le jour dans de plus en plus de labos) et toujours plus ouverte au partage des données. Merci à ceux qui osent dire que ça ne va pas et encore plus à ceux qui œuvrent pour des solutions. Pour ces raisons, je voudrais remercier l'ISEM d'avoir encouragé de telles démarches et qui m' a permis de m'épanouir dans un cadre qui soutient et partage mes convictions.

Dans cette continuité, j'aimerais remercier la direction et l'ISEM plus généralement d'être aussi soucieux du bien-être au travail. Carolle, PierrO, merci pour votre soutien et encouragements durant ma thèse.

Merci à Femme & Sciences, de soutenir et valoriser les femmes scientifiques, et de m'avoir aidé a prendre confiance en moi et à me sentir légitime d'être là. Et merci Julien, mon mentor, de m'avoir suivi depuis le début de ma thèse, soutenue, conseillée, guidée.

**Sur un plan un peu plus personnel,**

Merci à ceux qui avant moi sont passés par là, Rémi, Nathan, Sophie, Eliette, François, de m'avoir motivé à suivre vos traces, de m'avoir permis de voir, avant mon heure, la "traversée émotionnelle" de la fin de thèse, et de m'avoir montré que cette étape est surmontable. Merci aussi à Laura Mayer et Pierre Barry pour vos thèses que j'aurais aimé lire plus tôt.

Parce que la thèse construit des amitiés plus fortes que nos simples liens professionnels. Merci aux non-permanents de l'ISEM pour l'ambiance si chaleureuse et bienveillante que vous avez su créer (Alba, Alexis, Alex, Arthur, Bérénice, Elodie, Eliette, Gwen, Lucas, Jean-Loup, Mannue, Marie, Marie,

Nathan, Noémie, Quentin, Rémi, Sophie et tous ceux que je n'ai pas cités). Merci aux permanents de l'équipe avec qui j'ai apprécié partager nos déjeuners et divers apéros. Fabien, merci pour ta positivité et ta bienveillance. Marie-Ka, merci pour ton optimisme général et contagieux, ainsi que pour ton soutien dans tous les moments de ma thèse.

Merci à mes co-bureaux, Eliette, Nathan, Arthur, d'avoir contribué à cette atmosphère où il a fait bon travailler, brainstormer, débattre, mais aussi (et surtout !!) goûter. Nathan, merci de m'avoir initié à la macro', et de m'avoir montré qu'il est possible de passer des journées sur R sans (trop) râler. Arthur, merci pour tous nos midis, pour nos débats entrecoupés de science, de politique et de développement personnel.

Merci également au labo de Lille et au bureau 206 pour votre accueil si chaleureux, pour ces discussions et bières partagées et à ces amitiés crées. Emilie, Flavia, Francois, vivement la prochaine édition d' "écriture dans la pâture", je ne sais pas ce qu'on écrira, mais normalement, plus nos thèses !!

Merci à tous les copains qui ont réussi à me sortir l'esprit du travail pour des randos, treks, vides greniers, sorties naturalistes, jeux de sociétés ect... Notamment, Gaspar, Vanina, Romain, Robert, Baptiste, Rémy, Pierre, May, et les autres, merci pour toutes ces sorties spéléo, canyon, escalade et toutes les aventures que l'on a partagé ces trois années, qui m'ont aidé à prendre confiance en moi, et à prendre de la hauteur (littéralement) par rapport à ma thèse. Et mine de rien, l'effort fourni pour cette fin de thèse et le dépassement de soi qu'il demande m'a souvent rappelé le Berger (l'humidité en moins !).

Bryan, Noël, la petite ortie, Bjorn, Barnabé, les Jérômes, merci de m'avoir changé les idées et donné envie d'apprendre dans tant d'autres domaines.

Cedric, malgré les années et la distance, tu es toujours là pour me soutenir, merci.

Parce que la thèse est aussi une aventure dont les hauts et les bas forgent des liens solides, tant amicaux que familiaux, j'aimerais finir par remercier les personnes dont le soutien indéfectible m'a permis d'aboutir à ce travail.

Sophie, je suis heureuse que nos thèses aient permis à nos chemins de se croiser et que nous ayons été aussi soudées dans cette aventure. Merci pour la force que tu m'as transmise tout au long de ma thèse, merci de toujours avoir eu les mots justes, réconfortants et motivants.

Eliette, tu as été présente dans tant d'aspects de ma vie ces dernières années, en étant à la fois co-bureau, colloc, mais aussi copine. Merci qu'il ait toujours été si simple de travailler avec toi, de vivre avec toi, ou simplement d'Etre avec toi. J'admire ta passion débordante pour toutes les petites bêtes de notre monde qui ont le don de m'émerveiller à chaque nouvelle que tu me présentes (même

si pour les reines, c'est plutôt moi qui les déniche !).

La colloc, merci pour cet environnement qui m'a permis de m'épanouir pendant ma thèse, pour ces 3 années riches en débats scientifiques, politiques ou perso' aussi intéressants que surprenant, pour ces séances de jeux de société ou de sport et pour toutes ces répétitions d'oraux. R1, merci pour ces nocturnes à travailler (ou à papoter, aussi, ~~parfois~~ souvent), pour ces cours de stats et ces débats à essayer de modéliser la vie aussi bien que la science.

Jean Loup, merci de m'avoir accueilli pour finir ma thèse dans ce cadre si serein et paisible.
Aux Darwin, merci d'avoir fait de ces années de master mes plus belles années d'étude et d'avoir préservé ce groupe aussi soudé et bienveillant qu'à ses débuts malgré notre diversité de personnalité et de centre d'intérêts.

François, même si tu sais déjà tout, je suis heureuse de pouvoir te remercier ici pour tout ce que tu fais pour moi, depuis ces années, tu n'as cessé de me soutenir, m'encourager et de m'apprendre tant de choses. Merci pour tous ces débats sur la spéciation, ces cours de genet' des pop', et ces debugging de R (de tous ces bugs débiles dont tu arrives pourtant à trouver une logique). Merci de me pousser à me dépasser, et d'être toujours avec moi pour en assumer les conséquences (que ce soit pour me récupérer trempée, à vélo, au fin fond des Alpes, ou pour m'épauler après m'avoir initié au latex 72h avant ma fin de thèse). Merci pour tous ces moments de joie et l'équilibre que tu m'apportes. Pour tous ces moments passés et tous ceux qu' il nous reste, merci.

Enfin, je voudrais remercier ma famille. Maman, Papa, merci de m'avoir toujours encouragée à suivre la voie qui me rendrait heureuse, et soutenue quoi que je fasse. Maman, merci pour cette connexion qui fait de la distance quelque chose de presque imperceptible. Renou, Stéphanie, merci pour vos conseils toujours éclairés. Camille, tu t'es toujours battue pour arriver là où tu es, et je te remercie pour l'inspiration que ta force me procure. Désolé d'avoir été tant absente, mais promis, maintenant, je rentre bientôt.

# Abstract

Xenarthrans (armadillos, anteaters, and sloths) are the least diverse of the four major clades of placental mammals, with only 39 currently recognized species. However, this emblematic Neotropical clade includes species complexes with genetically distinct lineages of uncertain taxonomic status. To address this issue, we conducted a comprehensive genomic analysis, encompassing 261 individual mitogenomes, an exon capture dataset for 71 individuals, and 94 whole genomes, representing 34 distinct xenarthran species. As we included numerous museum samples, we carefully cleaned up potential genotyping errors and cross contaminations that could blur species boundaries by mimicking gene flow. This nearly exhaustive genome-wide dataset allowed us to revise Xenarthra taxonomy employing multiple lines of evidence (species delimitation methods, estimation of gene flow, genomic differentiation, morphology, demography) raising the number of xenarthran species recognized to 43. This revised taxonomic framework enabled the reconstruction of the most comprehensive time-calibrated phylogeny of Xenarthra based on whole genomes, shedding new light on their evolutionary history. By disentangling incomplete lineage sorting and gene flow in discordant parts of the xenarthran phylogeny, we identified multiple events of gene flow suggesting the maintenance of contact zones during speciation events. Further exploring factors promoting speciation within xenarthrans pointed to the potential influence of Colombian and Venezuelan Andes in their diversification. Overall, this work contributes to a deeper understanding of xenarthran evolution but also provides insights into future directions for taxonomic investigations and conservation status assessment within this fascinating mammalian group.



*Keywords:* Species delimitation, molecular systematics, taxonomy, whole-genome, museomics, speciation, conservation genomics, phylogenomics, gene-tree discordance, neotropics, biogeography, mammals.



# Résumé

Les xénarthres (tatous, fourmiliers et paresseux) sont les moins diversifiés des quatre principaux clades de mammifères placentaires, avec seulement 39 espèces actuellement reconnues. Cependant, ce clade néotropical emblématique inclut des complexes d'espèces avec des lignées génétiquement distinctes dont le statut taxonomique reste incertain. Pour les évaluer, nous avons analysé des données génomiques, regroupant 261 mitogénomes, des données issues de la capture d'exons pour 71 individus, et 94 génomes complets, représentant 34 espèces de xénarthres. Comme nous avons inclus de nombreux échantillons de musées, nous avons soigneusement nettoyé les erreurs potentielles de génotypage et les contaminations croisées qui pouvaient brouiller les délimitations d'espèces en imitant le flux de gènes. Ainsi, grâce à un jeu de données quasi exhaustif de données génomiques pour les xénarthres, nous avons révisé leur taxonomie en utilisant de multiples éléments de soutient (méthodes de délimitation d'espèces, estimation du flux de gènes, différenciation génomique, morphologie, démographie), ce qui a permis d'augmenter le nombre d'espèces de xénarthres reconnues à 43. Ce cadre taxonomique révisé a permis de reconstruire la phylogénie calibrée dans le temps la plus complète des xénarthres basée sur des génomes complets, révélant plus précisément leur histoire évolutive. En démêlant le tri de lignées incomplet et le flux de gènes des arbres de gènes discordants au sein des xénarthres, nous avons identifié de multiples événements de flux de gènes, ce qui suggère le maintien de zones de contact au cours de certains événements de spéciation au sein de ce clade. Enfin, en étudiant les facteurs favorisant la spéciation des xénarthres, nous suggérons l'influence des Andes colombiennes et vénézuéliennes sur leur diversification. En conclusion, cette thèse a contribué à une meilleure compréhension de l'évolution des xénarthres, mais elle donne également des pistes potentielles pour orienter les futures études taxonomiques et l'évaluation du statut de conservation des espèces de ce groupe fascinant de mammifères.



*Mots-clés:* Délimitation d'espèces, systématique moléculaire, taxonomie, génomes complets, muséomique, spéciation, génomique de la conservation, phylogénomique, discordance des arbres de gènes, néotropiques, biogéographie, mammifères.