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

Mathilde Barthe, Loïs Rancilhac, Maria C Arteaga, Anderson Feijó, Marie-Ka Tilak, Fabienne Justy, William J Loughry, Colleen M Mcdonough, Benoit de Thoisy, François Catzeflis, et al.

Spotlight Article

Syst. Biol. XX(XX):XX–XX, 2024

https://doi.org/10.1093/sysbio/syae027
Advance Access Publication June 22, 2024

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], William J. Loughry[7], Colleen M. McDonough[7], Benoit de Thoisy[8,9], François Catzeflis[1,†], Guillaume Billet[10], Lionel Hautier[1,11], Nabholz Benoit[1,12], and Frédéric Delsuc[1,*]

[1]*Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ. Montpellier, CNRS, IRD, Place E. Bataillon, 34095 Montpellier Cedex 05, France*
[2]*Animal Ecology, Department of Ecology and Genetics, Uppsala University, P.O. Box 256, SE-751 05 Uppsala, Sweden*
[3]*Department of biology, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus*
[4]*Department of Conservation Biology, CICESE, Carretera Ensenada, Tijuana No. 3918, Zona Playitas, CP. 22860, Ensenada, Baja California, México*
[5]*Negaunee Integrative Research Center, Field Museum of Natural History, 1400 S Lake Shore Dr, Chicago, IL 60605, United States*
[6]*Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China*
[7]*Department of Biology, Valdosta State University, 1500 North Patterson Street, Valdosta, GA 31698, United States*
[8]*Institut Pasteur de la Guyane, 23 Avenue Pasteur, BP 6010, Cayenne Cedex 97306, French Guiana*
[9]*Kwata NGO, 16 Avenue Pasteur, 97300 Cayenne, French Guiana*
[10]*Centre de Recherche en Paléontologie – Paris (CR2P), CNRS/MNHN/Sorbonne Université, Muséum national d'Histoire naturelle, 43 Rue Buffon, 75005 Paris, France*
[11]*Mammal Section, Life Sciences, Vertebrate Division, The Natural History Museum, Cromwell Road London, SW7 5BD, London, United Kingdom*
[12]*Institut universitaire de France, 1 Rue Descartes, 75231 Paris Cedex 05, 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
[*]*Correspondence to be sent to: Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ. Montpellier, CNRS, IRD, Place E. Bataillon, 34095 Montpellier Cedex 05, France; E-mail: mathilde.barthe.pro@gmail.com (M.B.); Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ. Montpellier, CNRS, IRD, Place E. Bataillon, 34095 Montpellier Cedex 05, France; E-mail: frederic.delsuc@umontpellier.fr.*

Received 10 July 2023; reviews returned 24 May 2024; accepted 19 June 2024
Associate Editor: Bryan Carstens

Abstract.—The nine-banded armadillo (*Dasypus novemcinctus*) is the most widespread xenarthran species across the Americas. Recent studies have suggested it is composed of 4 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 4 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 4 lineages within *D. novemcinctus* as 4 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 recognize 4 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 2 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. [Integrative taxonomy; mito-nuclear discordance; multilocus phylogeny; museomics; phylogeography; species delimitation.]

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 into 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 3 distinct species within the greater long-nosed armadillo (*Dasypus kappleri*) for instance (Feijó et al. 2018). Presently, 8 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 4 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 2 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 United States (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 2 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 4 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 and Superina, 2010; Superina et al., 2014; Feijó and Brandão 2022).

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 discordance (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) (de Queiroz, 2007; Degnan & Rosenberg, 2009; Wiens et al., 2010; Toews & Brelsford, 2012). 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 4 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; Robledo-Arnuncio and Gaggiotti, 2017; Raxworthy and Smith, 2021). 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; Ewart et al. 2019; 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 AND METHODS

Biological Sampling

A total of 80 armadillo tissue samples were obtained over the years for this study through multiple sources (Supplementary Table S1). Our sampling notably includes 46 individuals sampled in the form of dried skin pieces. Among these, 38 were obtained from museum specimens collected between 1894 and 2000 and stored in 12 different collections. The other 8 were received in the form of dried ear biopsies not associated with museum vouchers. The remaining samples were fresh tissue biopsies stored in 95% ethanol. 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 de la 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, and 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 µL instead of 100 µl. 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 2-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) (Supplementary 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 (Supplementary 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 2 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 (Supplementary Material on Dryad, https://doi.org/10.5061/dryad.95x69p8sz).

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 4 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 (Supplementary 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 (Supplementary 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 = proportion of DPs × read frequency; Supplementary 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 2 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 coverage under 10× 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 0.3–0.7, 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 3541 sequences (57 loci per individual on average) with more than 40% missing data (Supplementary Table S2) and 9 individuals with more than 55% average missing data across loci (see Supplementary Table S1 for details). We also excluded one locus containing less than 3 individuals as this is the minimal number required to reconstruct a phylogenetic tree. Note however that only 7 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 given nuclear 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 2 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 2 loci without any SNP and 3 loci represented by only 3 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 and 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 4 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 2 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 2 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 4 lineages of the *D. novemcinctus* complex and the 2 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, 3350 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 populations based on the transition in the substitution rate expected to follow 2 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 Ma 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 4 *D. novemcinctus* lineages defined in phylogenetic analyses (Fig. 1, see results for more details) and the 3 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 τ and θ, 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 4 configurations, 3 individuals within lineages A and B, and 2 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 10,000 simulated trees covering 48 demographic scenarios and 3 delimitation hypotheses (i.e., A, B, C represent 1 , 2 , or 3 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 Fraïsse et al. 2021; October 2022) was used to estimate population genomics statistics including mean pairwise divergence between lineages (Dxy) and net divergence (computed as $Da = \frac{Dxy-(Pi1+Pi2\ )}{2}$, where Pi_1 and Pi_2 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 2 diploid genomes (Allio et al. 2021). Heterozygosity was computed for each individual (Pi_1, Pi_2), and total nucleotide diversity was computed as the mean pairwise divergence between all chromosomes, within and between individuals (Pi_{Tot}) (https://github.com/benoitnabholz/seq_stat_2pop). Then, the Genetic Differentiation Index, which is conceptually similar to Fst (Allio et al. 2021), was estimated as $1 - \frac{(Pi1,Pi2)/2}{PiTot}$.

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 (Supplementary Fig. S4). Only 2 individuals had reads supporting 2 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 2 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 (Supplementary Fig. S4). For these 2 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

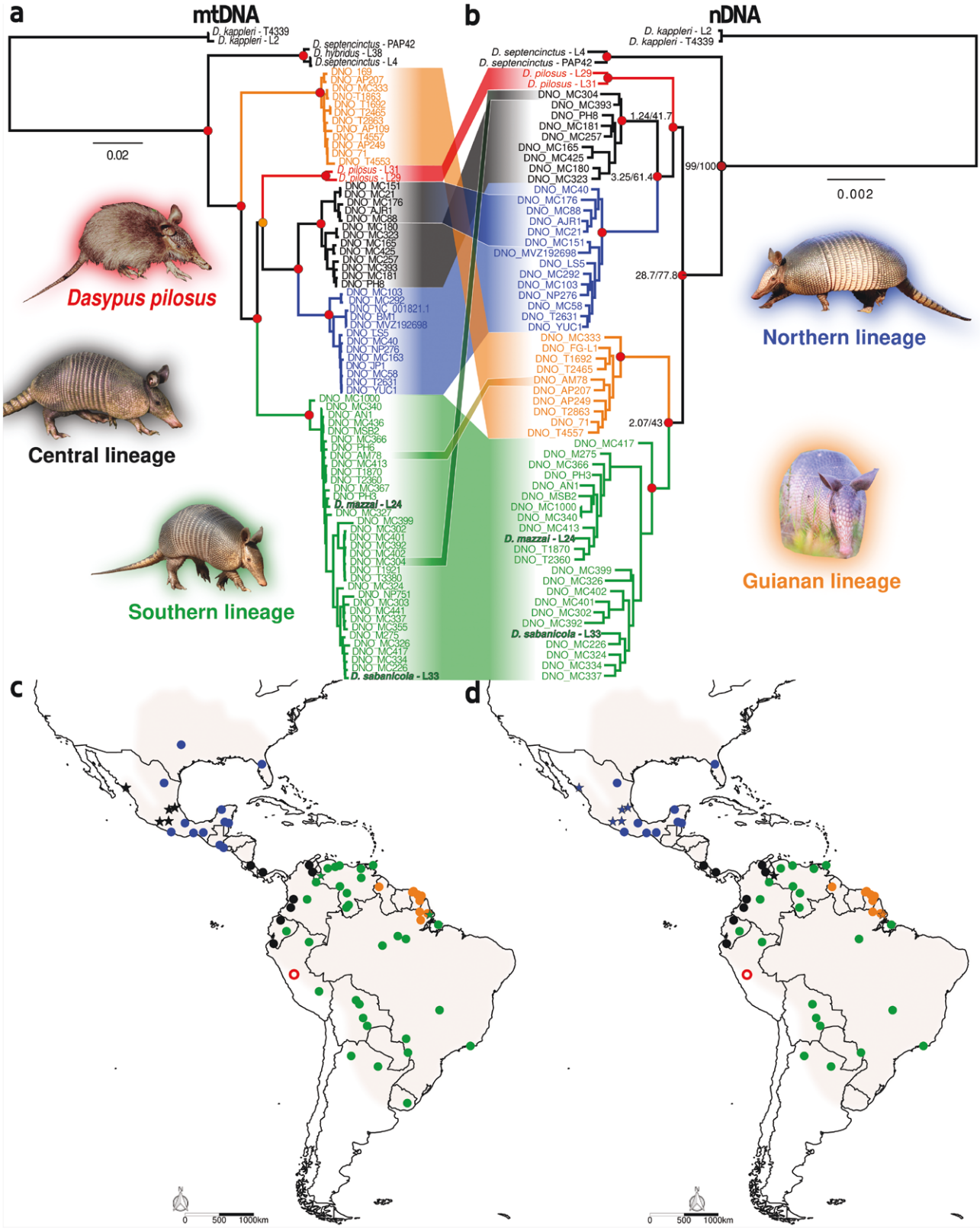

FIGURE 1. Phylogenetic relationships reconstructed by maximum likelihood and rooted using *Dasypus kappleri* from a) the mitogenomes of 81 individuals, and b) the 832 filtered nuclear loci of 62 individuals. Circles indicate bootstrap support (BS > 94) for the main nodes and node labels represent gene- 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. The *D. novemcinctus* distribution is highlighted. Photo credits: A. Baertschi (xenarthrans.org), K. Miller, Andresiade, A. Reed (iNaturalist.org), and Q. Martinez.

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 (Supplementary 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 2 filters turned unexpected negative inbreeding coefficient values to positive and led to a more even distribution of heterozygosity values among individuals (Supplementary 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 (Supplementary 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 (Supplementary Table S2). Of the 837 loci, 94% had less than 20% missing data and 98% were represented by at least 20 individuals (Supplementary 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 (Supplementary Fig. S8), all supported 4 biogeographical lineages within *D. novemcinctus*, while the 2 individuals of *D. sabanicola* and *D. mazzai* fell within the Southern lineage (Fig. 1). Although receiving strong Bootstrap Support (BS = 100), the composition of these 4 lineages differed between the mitochondrial and nuclear trees. A major case of mito-nuclear discordance concerned populations from Western Mexico, which carried 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.

Moreover, the inter-relationships of the 4 biogeographical lineages also differed between the nuclear and mitochondrial datasets (Fig. 1a and 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 2 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-groups, splitting the *D. novemcinctus* complex into 2 clades separated by the Andes.

Both topological conflicts and low gene- and site-concordance factors (respectively gCF and sCF, Fig. 1b) indicated high levels of discordance in our dataset. When accounting for ILS, the topologies recovered were similar to concatenation analyses (Supplementary Fig. S8). However, we also found that the 2 main mito-nuclear discordances (changing the 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).

Genetic Structure and Gene Flow Between Lineages

In the analysis of genetic admixture, maximum statistical support was obtained for a model with 4 genetic clusters (error = 0.2769 for $K = 4$; Supplementary Fig. S10), corresponding to the 4 *D. novemcinctus* lineages (Fig. 2). Similar to the previous results, the 2 samples of *D. mazzai* and *D. sabanicola* were consistently recovered as part of the Southern lineage. This analysis also identified 8 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.

The PCA of genetic variation confirmed the structure of these 4 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 2 being very close. The PC3 axis further distinguished *D. pilosus* from the 4 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 (Supplementary Fig. S11), as well as admixture analyses with $k = 5$ (error = 0.2784) revealed 2 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).

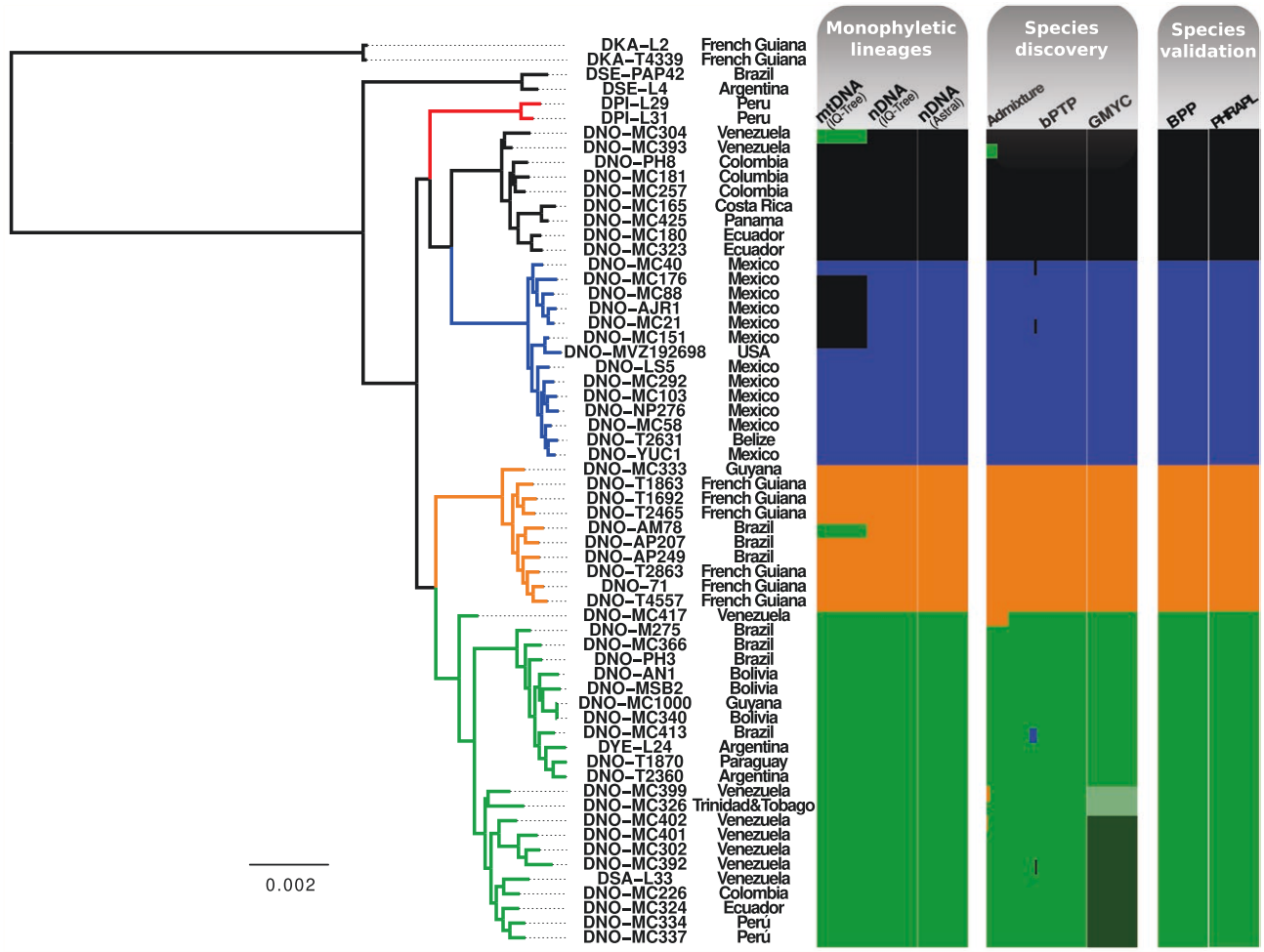

FIGURE 2. Phylogenetic relationships reconstructed by maximum likelihood from the 837 filtered nuclear loci and rooted using *Dasypus kappleri* and *D. septemcinctus* as outgroups. The color of the diagrams represents the assignment of individuals to lineages. For phylogenetic analyses (IQ-Tree and ASTRAL) 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.

Species Delimitation

Species delimitation analyses with BPP and bPTP-h supported the 4 lineages as distinct species with high support (BPP posterior probability = 1.00; bPTP-h acceptance rate = 0.54; Fig. 2 and Supplementary Fig. S12). This was also corroborated by PHRAPL that supported models with distinct species (Supplementary Fig. S13) 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 3 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 2 individuals in Venezuela and Trinidad-Tobago (Fig. 2 and Supplementary Fig. S14).

We used population genetic statistics to estimate genomic differentiation between lineages. Absolute divergence (Dxy) between the 4 lineages of the species complex was about 0.005 or 0.006 for all pairwise combinations (Supplementary Fig. S15). These values were comparable or greater than those obtained in pairwise comparisons with the taxonomically unambiguous species *D. pilosus* ($Dxy_{(D.\,pilosus)} = 0.005$), but lower than those of *D. s. septencinctus* and *D. kappleri* ($Dxy_{(D.\,septencinctus)} > 0.007$; $Dxy_{(D.\,kappleri)} > 0.013$). According to Da and GDI statistics, the divergence between the Southern and Guianan lineages was lower than between either lineages and *D. pilosus* ($Da_{(\text{Southern—Guianan lineages})} = 0.002 < Da_{(\text{Southern lineage-}D.pilosus)} = 0.003$; $GDI_{(\text{Southern—Guianan lineages})} = 0.36 < GDI_{(\text{Southern lineage-}D.pilosus)} = 0.58$). On the contrary, the Central and Northern lineages had similar levels of divergence to *D. pilosus* ($Da_{(\text{Central—Northern lineages})} = 0.003$; $GDI_{(\text{Central—Northern lineages})} = 0.58$). Finally, both Southern-Guianan and Central-Northern lineages had less fixed differences than *D. pilosus* had with other lineages (Fixed diff.$_{(\text{Southern—Guianan lineages})} = 0.003$; Fixed diff.$_{(\text{Central—Northern lineages})} = 0.003$; Fixed diff.$_{(D.\,pilosus\text{-Southern lineage})} = 0.004$).

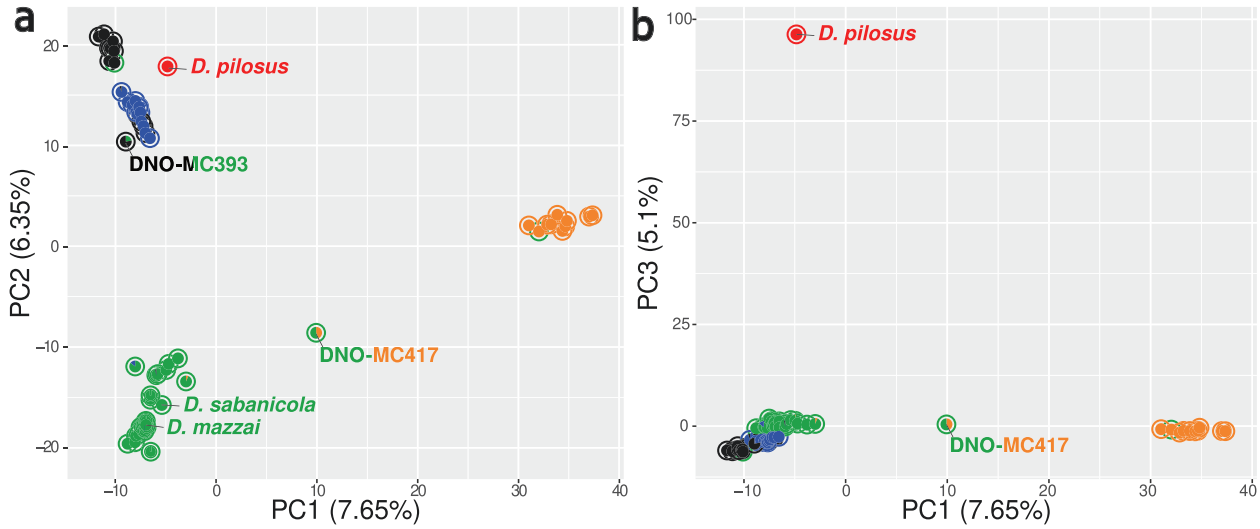

FIGURE 3. Principal component analysis of genetic variance (PCA) conducted on the 3350 SNPs in 837 nuclear loci shared by all 56 individuals. a) Projection on first (PC1) and second (PC2) axes. b) Projection on first (PC1) and third (PC3) axes. The outer circles represent mitochondrial lineages and pie charts the nuclear lineages identified in admixture analyses. Individuals of *D. pilosus*, *D. sabanicola* and *D. mazzai* are labeled.

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 mislabeling), 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 2 (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 (Supplementary Fig. S3). Although we performed separate DNA extractions for museum and fresh samples, which differed significantly in their DNA content, we detected at least 2 cases of apparent cross-contamination between our samples (Supplementary 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 a 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 2 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 2 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 2 metrics enabled us to easily identify outlier individuals (Supplementary Fig. S5). After applying these 2 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 2 cross-contaminated individuals identified based on mitochondrial reads. As expected, the application of these filtering criteria had a particularly strong effect on admixture analyses (Supplementary Fig. S16). Our results therefore call for a more systematic use of these 2 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 Based on Genomic Data

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, Supplementary Fig. S8) consistently recovered 4 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, 2 cases of mito-nuclear discordances concerning single individuals, and 5 individuals with admixed nuclear genomes, support ongoing gene flow. Most of them were located close to contact zones between adjacent lineages. Despite that, the 4 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 4 lineages showed similar levels of divergence and genetic structure compared to *D. pilosus*, further supporting their split into different species. Postspeciation 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 4 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 4 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. 4). 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. 4). 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; Supplementary Table S4; Supplementary Fig. S17). *D. novemcinctus* is now limited to the eastern Andes, spanning from northern Argentina to eastern Peru, eastern Ecuador, eastern Colombia and Venezuela (Fig. 4). It is noteworthy that our analyses

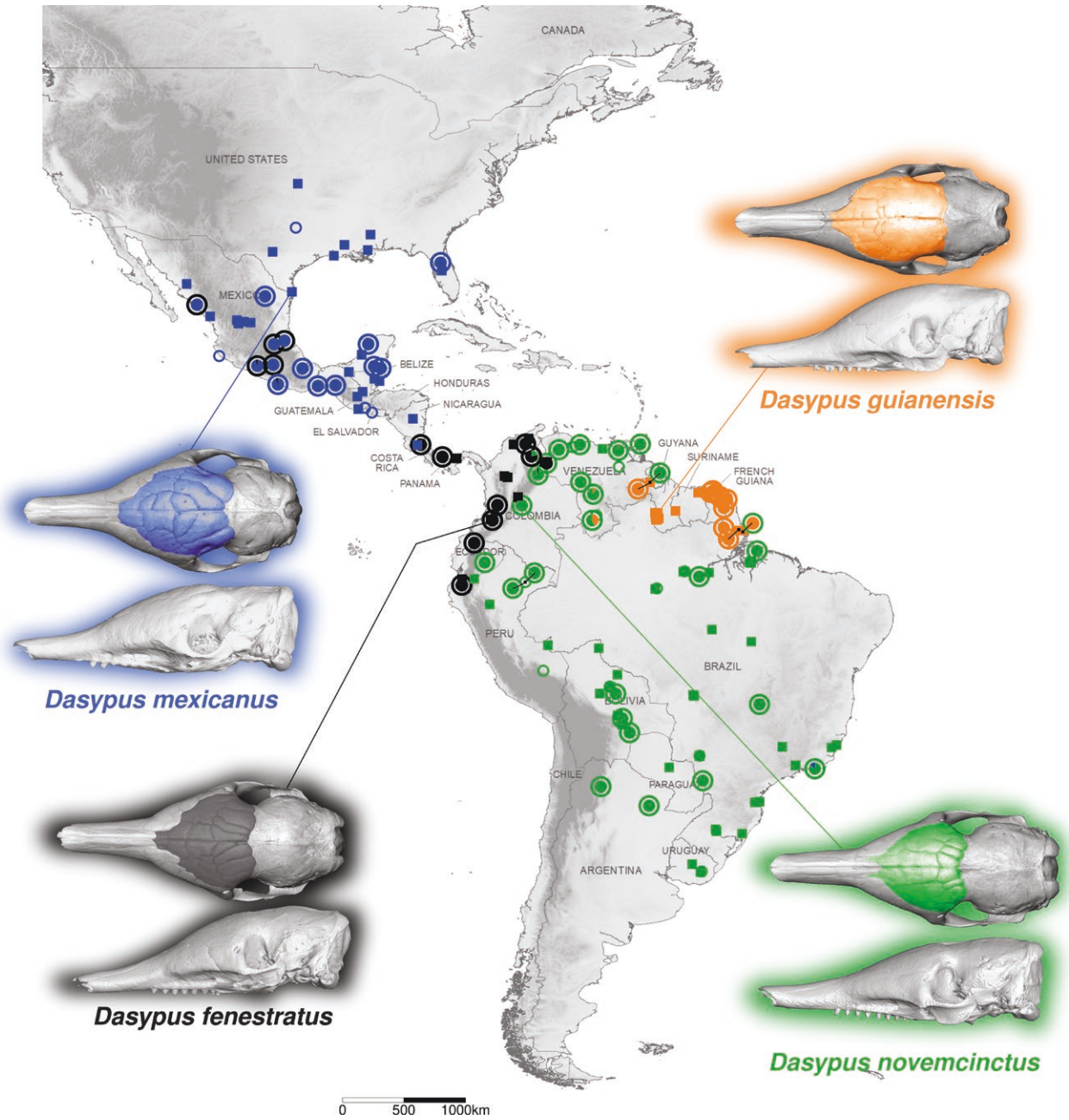

FIGURE 4. Distribution map and genetic composition of individuals of the four recognized species: *Dasypus mexicanus* (Northern lineage), *Dasypus fenestratus* (Central lineage), *Dasypus novemcinctus* (Southern lineage), and *Dasypus guianensis* sp. nov. (Guianan lineage). Outer circles represent mitochondrial 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) 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 specimen vouchers are presented for each species with frontal paranasal sinuses and recesses highlighted: USNM Mammals 33867 (*Dasypus mexicanus*), AMNH Mammals M-32356 (*Dasypus fenestratus*), AMNH Mammals M-136252 (*Dasypus novemcinctus*), and ROM Mammals 32868 (*Dasypus guianensis* sp. nov.).

failed to recognize the 2 available individuals of *D. mazzai* and *D. sabanicola* as independent lineages, since they were consistently nested within *D. novemcinctus* (Fig. 1; Supplementary Fig. S18). 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 3 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, 2024) 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. 4) 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; Supplementary Fig. S18). 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 4 major clades, we further identified 2 divergent sub-lineages within *D. novemcinctus* (Supplementary 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 9 movable bands in northern South American armadillos, while those from the southern part exhibit mostly an 8-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 4 distinct species with parapatric distributions. Their distribution is 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.

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 4 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.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.95x69p8sz

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. We also like to thank the editors and 3 anonymous referees for insightful comments. This is contribution ISEM 2024-102 SUD of the Institut des Sciences de l'Evolution de Montpellier.

CONFLICT OF INTEREST

None declared.

FUNDING

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). Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform. The JAGUARS collection is supported by the Collectivité Territoriale de Guyane and the Direction Générale des Territoires et de la Mer Guyane.

DATA AVAILABILITY

All datasets and supplementary materials are available from Zenodo (https://doi.org/ 10.5281/zenodo.7509863) and Dryad (https://doi.org/10.5061/dryad.95x69p8sz) public online repositories. 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. 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).

APPENDIX 1: TAXONOMIC ACCOUNT

In this section we provide the formal description of the new *Dasypus* species from the Guiana Shield and accounts for the other three species of the *D. novemcinctus* complex, including synonyms, type material, type locality, diagnosis, and geographic distribution. External and cranial measurements are available in Tables S5, S6 and are based on specimens examined by Feijó et al. (2018). Synonyms are based on Wetzel (2008) and Feijó et al. (2018).

Dasypus guianensis Barthe, Feijó, de Thoisy, Catzeflis, Billet, Hautier & Delsuc, sp. nov.

lsid:zoobank.org:act:D82875C3-B21D-4472-AA2B-0DF9443C2B38

Armadillo guianensis Brisson, 1762: 27; unavailable name (ICZN 1998: Opinion 1894).

Dasypus peba Desmarest, 1822:368, part

Common name: Guianan long-nosed armadillo.

Holotype: MNHN-ZM-MO-2001-1317, a female stored as a whole body specimen in the wet collection of the Muséum National d'Histoire Naturelle, Paris, France. This specimen was collected in 1998 by Jean-Christophe Vié during the Faune Sauvage rescue operation at the Petit-Saut dam in French Guiana. Specimen figured in Billet et al. (2017; Fig. 4O). Digital 3D models of the *Dasypus guianensis* holotype specimen are freely available through MorphoMuseuM (Barthe et al. 2024; https://doi.org/10.18563/journal.m3.204).

Type locality: Petit-Saut, Sinnamary, French Guiana.

Paratypes: MNHN-ZM-MO-2001-1530, a juvenile male preserved as a whole body specimen deposited at the Muséum National d'Histoire Naturelle, Paris, France. This specimen was collected in Roura, French Guiana in January 2000 by Philippe Gaucher. Its mitogenome is deposited in GenBank (accession number PP916004). JAG M5328, a juvenile male preserved as a whole body specimen deposited in the JAGUARS collection managed by Kwata NGO and hosted at the Institut Pasteur de la Guyane, Cayenne, French Guiana. This specimen was collected in Macouria, French Guiana on the 27 February 2021 by Benoit de Thoisy. FMNH Mammals 95452, an adult female preserved as skin and skull deposited at the Field Museum of Natural History, Chicago, United States. This specimen was collected in Loksie Hattie, Saramacca, Brokopondo, Suriname on 27 December 1961 by Philip Hershkovitz. IEPA 2837, also figured in Billet et al. (2017; Fig. 4P), an adult male preserved as a skull and tissue specimen deposited at the Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá, Macapá, Brazil. This specimen was collected in Fazenda Chamflora, Tartarugalzinho, Amapá, Brazil on 30 January 2006 by Claudia Regina Silva as part of Expedições Corredor da Biodiversidade. AMNH Mammals M-42914, an adult female deposited at the American Museum of Natural History, New York,

USA. This specimen was collected in Kartabo Point, Mazaruni Region, Cuyuni, Guyana on 1 August 1920.

Distribution: *Dasypus guianensis* is distributed in the Guiana Shield region, extending from the north of the lower Amazon and east of the Rio Negro rivers in northern Brazil (with confirmed records to Amazonas, Amapá, Pará, and Roraima states) to French Guiana, Suriname, Guyana, and east of the Orinoco river in Venezuela.

Diagnosis: *D. guianensis* is recognized by a combination of traits: number of movable bands ranging from 8 to 9, 11 thoracic vertebrae, poorly developed scales at the knee, four digits on forefoot, non-hairy carapace, very robust dome-shaped skull, with the frontal bones well-developed and markedly vaulted resulting in a distinct sigmoidal dorsal profile in lateral view. The anterior tip of the snout is wide; the premaxilla is long. The maxilla is well-developed and expanded laterally. The interorbital width is wide. The lacrimal bone is large. The palate is wide and rectangular with robust teeth. The anterior border of the palatine bone is located well behind the posterior end of the tooth row; the pterygoid process is short. The premaxillary-maxillary suture is located beyond the incisive foramina and the jugal is short, robust and long vertically (Hautier et al. 2017; Feijó et al. 2018). The paranasal sinuses of *D. guianensis* are unique among long-nosed armadillos, characterized by a long and strongly inflated caudal frontal sinus 2 (see Billet et al., (2017)). *D. guianensis* is the largest species of the *D. novemcinctus* complex (Tables S5, S6).

Description: *D. guianensis* is a medium-sized armadillo (average total length: 860 mm; average weight: 6 kg). As in other species of the genus, the head is conical with a long and tubular rostrum and conical-shaped ears placed at the dorsal of the head. In *D. guianensis*, the pelvic shield is about 35% longer than the scapular shield. The number of movable bands range from 8 to 9, but most of the examined specimens have nine. The osteoderms in the caudal sheath are arranged in 13-14 concentric rings that cover 40-50% of the tail. The forefeet have four fingers and the hindfeet has five toes. The carapace is hairless and its color is overall dark brown in the dorsum and markedly yellowish in the lateral; the yellowish portion can cover up to two-thirds of the carapace.

The skull of *D. guianensis* is conical and robust, with an elongate rostrum (about 40% of greatest skull length). The maxilla is large and expanded laterally along its entire length. The frontal is well developed and markedly vaulted, giving a strong sigmoidal profile to the skull in lateral view. This trait reflects the inner paranasal sinuses configuration, which is characterized by strong inflation of the posterior portion of the caudal frontal sinuses (CFS) 2. The CFS2 is the largest frontal sinus and occupies most of the pneumatized frontal area. The CFS 3 and 4 and the anterior part of CFS 2 are narrow. When present, the CFS0 is small, elongated transversally and slightly shifted anteriorly. The jugal is robust and long vertically. The palate is long, large, and flat, extending to the level of the squamosal process of the zygomatic arch. The lateral border of the palatine is rounded. Permanent teeth are molariforms, euhypsodont and restricted to the maxillary bone. The mandible is slender, slightly curved, and dorsoventrally shallow. The coronoid process is very long, posteriorly inclined, and sharpened. The condylar process is very short and the angular process is abrupt and separated from the condyloid by a short, shallow curve.

Measurements: Adult measurements (in millimeters) are given as the mean ± standard deviation, minimum–maximum (n). External measurements: total length: 859 ± 56.39, 790–945 (n=5); tail length: 395.5 ± 50.74, 339–460 (n=6); hindfoot length: 95 ± 14.72, 75–110 (n=4); ear length: 49.25 ± 4.35, 45–55 (n=4); weight (in kg): 6.075 ± 0.67, 5.6–6.5 (n=2). Carapace measurements following Feijó et al. (2018): dorsal length of the cephalic shield: 104.3 ± 11, 93–115 (n=3); dorsal length of the scapular shield: 92.1 ± 9.5, 75.4–98 (n=5); dorsal length of the pelvic shield: 150.5 ± 14.4, 136–168 (n= 4); dorsal length of the caudal sheath with rings: 268.3 ± 21.6, 248–291 (n=3); dorsal length of the caudal sheath without rings: 149 ± 9.9, 142–156 (n=2). Cranial measurements from 26 adult individuals following Feijó and Cordeiro-Estrela (2016): greatest length of skull: 106.40 ± 7.27, 96.96–128.31; condylobasal length: 98.26 ± 6.79, 88.74–117.45; anterior palatal length: 23.92 ± 1.74, 20.54–28.21; palatal length: 71.84 ± 5.86, 63.78–90.16; maxilla length: 42.39 ± 4.48, 36.65–54.23; palatine length: 19.32 ± 3, 15.33–28; infraorbital canal length: 6.47 ± 1.05, 3.61–8.40; maxillary toothrow length: 25.07 ± 1.90, 22.20–30.88; nasal length: 36.53 ± 3.11, 30.75–42.70; lacrimal length: 13.03 ± 1.5, 10.38–17.38; rostral length: 66.32 ± 5.54, 57.85–82.71; anteorbital breadth: 36.20 ± 2.39, 30.5–40.61; palatal breadth: 16.12 ± 2.25, 11.05–20.15; palatine breadth: 15.24 ± 1.91, 12.24–20.59; postorbital constriction: 25.96 ± 1.39, 22.76–28.71; braincase breadth: 33.57 ± 1.76, 30.24–37.99; zygomatic breadth: 44.95 ± 2.84, 39.86–54.04; mastoid breadth: 29.14 ± 2.14, 25.86–35.72; height of jugal bone: 8.13 ± 0.9, 6.21–10.16; mandible length: 84.68 ± 5.97, 77.23–101.46; mandibular toothrow length: 26.61 ± 2.11, 22.04–33.01 (Tables S5, S6).

Etymology: The epithet *guianensis* refers to the Guiana Shield region, an important endemism area in northern South America, where the species occurs.

Taxonomic remarks: Brisson (1762) described *Armadillo guianensis* but this name is unavailable for nomenclatural purposes (ICZN 1998: Opinion 1894). Desmarest (1822) described *Dasypus peba* including Brazil, Guyana, and Paraguay as its type locality and listed ten names as synonyms based on six sources. All specimens which form the bases of these synonyms have equal status as syntypes (ICZN 1999: Article 73.2.1). Because the names listed by Desmarest relate to more than one taxon, *D. peba* represents a composite species. To preserve the long-standing view of *D. peba* as a synonym of *D. novemcinctus* Linnaeus, 1758 (Wetzel (2008); Feijó et al. 2018), here we select the specimen

illustrated by Marcgrave (1648; page 231) as the lectotype of *Dasypus peba* Desmarest, 1822, which was the first reference used by Desmarest to found its species. Because this specimen was previously selected as the lectotype of *D. novemcinctus* by Feijó et al. (2018), *D. peba* becomes an objective junior synonym of *D. novemcinctus* Linnaeus, 1758.

Dasypus mexicanus Peters, 1864

Dasypus cucurbitinus Lichtenstein, 1830:100; type locality "Jautepec," Mexico; *nomen oblitum* (see Gardner and Ramírez-Pulido, (2020)).

Dasypus novemcinctus var. Mexicanus Peters, 1864:180; original description

Dasypus mexicanus: Fitzinger, 1871:332; name combination.

Tatusia mexicana: Gray, 1873:14; name combination.

Tatusia leptorhynchus Gray, 1873:15; type locality "Guatemala."

Tatu nomencinctum texanum Bailey, 1905:52; type locality "Brownsville, Texas."

Dasypus novemcinctus davisi Russell, 1953:21; type locality "Huitzilac, 8500 feet, Morelos, Mexico".

Common name: Mexican long-nosed armadillo.

Holotype: ZMB 2743, a well-preserved mounted specimen deposited at the Museum für Naturkunde, Berlin, Germany.

Type locality: Originally Mexico, but later restricted to Colima, Mexico by Bailey (1905:52) and redefined by Hollister (1925:60) as Matamoras, Tamaulipas, Mexico. Gardner and Ramírez-Pulido (2020), however, argued that the holotype of *mexicanus* collected by Uhde came from western Mexico and therefore Colima should be reinstated as its type locality.

Distribution: *Dasypus mexicanus* occurs from Costa Rica to the center-eastern United States. In the USA, it is considered an invasive species. Available records confirm the sympatry of this species with *D. fenestratus* in Costa Rica.

Diagnosis: Compared with the four species of the *D. novemcinctus* complex, *D. mexicanus* has an intermediate size between *D. fenestratus*/*D.novemcinctus* and *D. guianensis* (see Tables S5, S6). The dorsal length of the scapular shield is markedly longer in *D. mexicanus* (average length 96.6 mm) than in the other three species. Its skull exhibits moderately developed frontal sinuses convergent toward the midline, long premaxillary bones and infraorbital canal, a long and slender jugal, an anterior border of the palatine located at the level of the posterior end of the tooth row, short pterygoid processes, and a basicranium slightly above the palatal plane (Hautier et al. 2017). The main diagnostic features of the paranal sinuses that differentiate *D. mexicanus* from *D. novemcinctus* and *D. guianensis* are the anteroposterior elongation of the CFS 2 to 5, the left and right CFS2, which are obliquely oriented and contact each other posterior to the CFS1, and the subdivision and the relative shortening of the rostral frontal recesses (RFR) 1 (Billet et al. 2017).

Dasypus fenestratus Peters, 1864

Dasypus fenestratus Peters, 1864:180; original description.

Tatusia granadiana Gray, 1873:14; type locality "Concordia," Antioquia, Colombia.

[Tatusia (Tatusia)] granadiana: Trouessart, 1898:1140; name combination.

[Tatus (Tatus)] granadianus: Trouessart, 1905:814; name combination.

Dasypus novemcinctus aequatorialis Lönnberg, 1913:34; type locality "Peruchu, altitude 7–9,000 feet," Pichincha, Ecuador.

Common name: Western Andean long-nosed armadillo.

Type: Lectotype (ZMB 3175) designated by Wetzel and Mondolfi (1979:50). It consists of an adult male preserved as skin and skull deposited at the Museum für Naturkunde, Berlin, Germany.

Type locality: "Costa Rica"; restricted to San José, Costa Rica by Wetzel and Mondolfi (1979:50).

Distribution: *Dasypus fenestratus* is distributed from the western slopes of the Andes in Ecuador, Colombia, and Venezuela to Costa Rica.

Diagnosis: Compared with the other species of the *D. novemcinctus* complex, *D. fenestratus* is smaller than *D. guianensis* and *D.mexicanus*, although it overlaps to varying degrees in some of the measurements (see Tables S5, S6). *D. fenestratus* has the smallest carapace in the complex. The measurements with the lowest overlap to *D. mexicanus*, the parapatric species, are total body-tail length (average 807 mm in *mexicanus* and 780 mm in *fenestratus*), hindfoot length (average 96.7 mm in *mexicanus* and 81.4 mm in *fenestratus*), dorsal length of the scapular shield (average 96.6 mm in *mexicanus* and 86.6 mm in *fenestratus*), dorsal length of the pelvic shield (average 131 mm in *mexicanus* and 123 mm in *fenestratus*), palatine length (average 17 mm in *mexicanus* and 14 mm in *fenestratus*), and infraorbital canal length (average 9.4 mm in *mexicanus* and 6.5 mm in *fenestratus*). The skull of *D. fenestratus* is high and short, and characterized by moderately developed frontal sinuses (generally resembling those of *D. mexicanus*), short palatine (an anterior border of the palatine located well behind the posterior end of the tooth row) and maxillary toothrow, a well-developed anterior part of the zygomatic arch (short and high jugal) that is much larger than the posterior part, and a basicranium well above the palatal plane.

Remarks: Peters (1864) described *D. mexicanus* from Mexico and *D. fenestratus* from Costa Rica, but we found evidence that both species occur in sympatry in northern Costa Rica. By examining the lectotype of *D. fenestratus*, we confirmed its skull and carapace traits fit the Central lineage population and thus we consider it the

oldest available name for this group. Future molecular data from the lectotype can confirm its identity.

Dasypus novemcinctus Linnaeus, 1758

[Dasypus] novemcinctus Linnaeus, 1758:51; original description.

[Tatu] Novemcincta: Blumenbach, 1779:74; name combination.

Dasypus novenxinctus Peale and Palisot de Beauvois, 1796:18; incorrect subsequent spelling of *Dasypus novemcinctus* Linnaeus.

Dasypus longicaudatus Daudin in Lacépède, 1802:173; no locality given; based on Buffon's "Le tatou à longue queue" (Lacépède, 1802; p. 168, pl. 22) of unknown province; preoccupied by *Dasypus longicaudatus* Kerr.

[Dasypus] serratus Fischer, 1814:128; type localities "Paraquaia, inprimis in provincial Buenos-Ayres (Boni Aëris)."

Dasypus decumanus Illiger, 1815:108; nomen nudum.

Dasypus decumanus Olfers, 1818:219; nomen nudum.

T[atus]. niger Olfers, 1818:220; type localities "Paraguay, Brasilien"; preoccupied by *Loricatus niger* Desmarest.

Dasypus niger Lichtenstein, 1818:20; type locality not given; based on *D. novemcinctus* Linnaeus; therefore, the type locality is "Pernambuco, Brazil" (lectotype designated by Feijó et al. 2018); preoccupied by *Loricatus niger* Desmarest.

Dasypus peba Desmarest, 1822:368, part; type localities "Le Brésil, le Guyane, le Paraguay…On ne le trouve pas dans la province de Buenos-Ayres."; junior objective synonym of *D. novemcinctus* Linnaeus (lectotype designated here); therefore, the type locality is "Pernambuco, Brazil."

Dasypus longicaudus Schinz, 1824:253; unavailable name.

D[asypus]. longicaudus Wied-Neuwied, 1826:531; type locality "In den Waldern am Mucuri"; identified by Ávila-Pires (1965; p. 12) as Morro d'Arara, Rio Mucuri, Bahia, Brazil.

Tatusia peba: Lesson, 1827:311; name combination

Dasypus [(Cachicamus)] novemcinctus: McMurtrie, 1831:163; name combination.

Dasypus uroceras Lund, 1839[1841]:pl. 12, Fig 5; type locality "Rio das Velhas, Floddal" (p. 73), Lagoa Santa, Minas Gerais, Brazil.

D[asypus] uroceros Burmeister, 1848:199; incorrect subsequent spelling of *Dasypus uroceras* Lund.

Praopus longicaudatus Burmeister, 1854:298; name combination

Cachicamus novemcinctus Degland, 1854:125; name combination

Dasypus pepa Krauss, 1862:19; incorrect subsequent spelling of *Dasypus peba* Desmarest.

D[asypus]. longicaudatus Peters, 1864:179; incorrect subsequent spelling of *Dasypus longicaudus* Wied-Neuwied; not *D. longicaudatus* Kerr.

Dasypus lundii Fitzinger, 1871:340; type locality "Brasilien."

Tatusia platycercus Hensel, 1872:105; type locality "Urwald von Rio Grande do Sul," Brazil.

Tatusia brevirostris Gray, 1873:15; type localities "Rio de Janeiro," Brazil, and "Bolivia"; Wetzel (2008) selected the specimen from Rio de Janeiro (skin BM 44.3.7.2; skull: BM 46.5.13.16) as the lectotype.

Tatusia leptocephala Gray, 1873:16; type locality "Brazils."

Tatusia boliviensis Gray, 1873:16; type locality "Bolivia."

T[atusia]. leptorhinus Gray, 1874:246; incorrect subsequent spelling of *Tatusia leptorhynchus* Gray.

Praopus 9-cinctus Burmeister, 1879:434; name combination

Tatusia novemcincta: Thomas, 1880:402; name combination.

Tatusia longicaudatus: Allen, 1895:187; name combination.

[Tatusia (Tatusia)] novem-cincta: Trouessart, 1898:1139; name combination

[Tatusia (Tatusia)] platycercus: Trouessart, 1898:1140; name combination.

[Tatusia (Tatusia)] brevirostris: Trouessart, 1898:1140; name combination.

[Tatusia (Tatusia)] leptocephala: Trouessart, 1898:1140; name combination.

[Tatusia (Tatusia)] boliviensis: Trouessart, 1898:1140; name combination.

[Tatusia (Tatusia)] granadiana: Trouessart, 1898:1140; name combination.

Tatua novemcincta: Robinson and Lyon, 1901:161; name combination.

[Tatus (Tatus)] novem-cinctus: Trouessart, 1905:814; name combination.

[Tatus (Tatus)] platycercus: Trouessart, 1905:814; name combination.

[Tatus (Tatus)] brevirostris: Trouessart, 1905:814; name combination.

[Tatus (Tatus)] leptocephalus: Trouessart, 1905:814; name combination.

[Tatus (Tatus)] boliviensis: Trouessart, 1905:814; name combination.

[Tatus (Tatus)] granadianus: Trouessart, 1905:814; name combination.

Tatusia novemcincta var. *mexianae* Hagmann, 1908:29; type locality "Insel Mexiana," Pará, Brazil.

Dasypus boliviensis: Grandidier and Neveu-Lemaire, 1908:5; type locality "environs d'Uyuni," Potosí, Bolivia; preoccupied by *Tatusia boliviensis* Gray.

Dasypus novemcinctus hoplites Allen, 1911:195; type locality "hills back of Gouyave, island of Grenada", Lesser Antilles.

D[asypus]. longi-cauda Larrãnaga, 1923:343; type locality "provincial paracuarensi"; based on Azara's (1802; p. 144) "Negro"; a junior synonym and homonym of *Dasypus longicaudus* Wied-Neuwied.

D[asypus]. brevirostris: Yepes, 1933:230; name combination.

Common name: Nine-banded armadillo.

Type: Feijó et al. (2018) designated the specimen illustrated by Marcgrave (1648; p. 231) as the lectotype of *D. novemcinctus* Linnaeus.

Type locality: Pernambuco, Brazil.

Distribution: *Dasypus novemcinctus* is distributed from the eastern slopes of the Andes from Ecuador, Colombia, and Venezuela to Brazil, Peru, Bolivia, Paraguay, Uruguay, and northern Argentina. This species is absent in the Guiana Shield, except in the peripheral zone of eastern Venezuela, where it may occur in sympatry with *D. guianensis*.

Diagnosis: *Dasypus novemcinctus* is the smallest species of the complex (Tables S5, S6). It is characterized by a smaller and flatter skull with a narrow snout with short premaxillary bones, a narrow interorbital width, a long and slender jugal part of the zygomatic arch, long pterygoid processes, and a basicranium aligned with the palate in lateral view. Its paranasal sinuses are easily differentiated by an anteroposteriorly reduced posterior chain of caudal frontal sinuses (CFS1 to 5) and an elongated rostral frontal recess 1 (RFR1). The CFS1 are always in contact with each other medially or just at their posteromedial corner, if the CFS0 are present. The anteroposterior length of the CFS never exceeds that of the RFR1 and the CFS5 is generally shorter than the other CFS (Billet et al. 2017). In comparison with the other species of the *D. novemcinctus* complex, *D. novemcinctus* has a markedly short nasal bone and tail.

Remarks: In light of the new taxonomic arrangement of the *novemcinctus* complex proposed in this study, we consider *Dasypus octocinctus* Schreber, 1774, *Dasypus longicaudatus* kerr, 1792, and *Loricatus niger* Desmarest, 1804 as *nomen dubia* as they all represent composite species, lack specific type locality, their type material are not known to exist, and their description cannot be applied with certainty to only one species of the *novemcinctus* complex. Therefore, we removed them from the synonym list of *D. novemcinctus* Linnaeus, 1758.

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*? Revista del Museo Argentino de Ciencias Naturales. 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.

Anderson G., Macdonald J.I., Potts J., Feutry P., Grewe P.M., Boutigny M., Davies C.R., Muir J.A., Roupsard F., Sanchez C., Nicol S.J. 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. Cham, Switzerland: 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 2.3. https://CRAN.R-project.org/package=gridExtra.

Bailey V. 1905. Biological survey of Texas. N. 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., Pääbo S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. U.S.A. 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–s13742-015-0048.

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. 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., Farias I.P., Hrbek T. 2019. The Munduruku marmoset: a new monkey species from southern Amazonia. PeerJ 7:e7019.

Costa-Araújo R., Silva, J.S. Jr, Boubli J.P., Rossi R.V., Canale G.R., Melo F.R., Bertuol F., Silva F.E., Silva D.A., Nash S.D., Sampaio I., Farias I.P., Hrbek T. 2021. An integrative analysis uncovers a new, pseudo-cryptic species of Amazonian marmoset (Primates:

Callitrichidae: *Mico*) from the arc of deforestation. Sci. Rep. 11:15665.
Coyne J.A., Orr H.A. 2004. Speciation. MA: Sinauer Associates Sunderland.
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., McVean G., Durbin R.; 1000 Genomes Project Analysis Group. 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., Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10:giab008.
de Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.
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. New York, NY, USA: 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.
Ewart K.M., Johnson R.N., Ogden R., Joseph L., Frankham G.J., Lo N. 2019. Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species. Mol. Ecol. Resour. 19:1578–1592.
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., 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. Zoologia 39:e22007.
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., Golmard L., Houdayer C. 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., Bierne N., Vekemans X., Roux C. 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 7:134.
Garrison E., Marth G. 2012. Haplotype-based variant detection from short-read sequencing. GitHub Repository: https://github.com/freebayes/freebayes
Ge D., Wen Z., Feijó A., Lissovsky A., Zhang W., Cheng J., Yan C., She H., Zhang D., Cheng Y., Lu L., Wu X., Mu D., Zhang Y., Xia L., Qu Y., Vogler A.P., Yang Q. 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., Hansen N.F., Durand E.Y., Malaspinas A.-S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gušic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. 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., Prüfer K., Siebauer M., Burbano H.A., Ronan M., Rothberg J.M., Egholm M., Rudan P., Brajković D., Kućan Z., Gusić I., Wikström M., Laakkonen L., Kelso J., Slatkin M., Pääbo S. 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.
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.
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. ZooKeys 324:1–83.
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 Bioinf. 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. 6:59–59.
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. Fourth Edition. London, UK. International Trust for Zoological Nomenclature. https://www.iczn.org/the-code/the-code-online/
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. 2024. Pervasive cranial allometry at different anatomical scales and variational levels in extant armadillos. Evolution 28:423–441.
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. GitHub repository. https://github.com/lh3/bwa
Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.; 1000 Genome Project Data Processing Subgroup. 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. 7:28–16.
Marcgrave G. 1648. Historiae Naturalis Brasiliae. Haack, e Elzevier, Leiden e Amsterdam.
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. Boston, MS, USA: Harvard 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., 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. 17:2103–2106.
Padial J.M., Miralles A., De la Riva I., Vences M. 2010. The integrative future of taxonomy. Front. Zool. 7:16–14.
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. https://github.com/broadinstitute/picard
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. Boston, MA, USA: Posit Software, PBC.
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., Sham P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.
R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org
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 P.M. Jr. 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., de Carvalho Clarissa F., Safran Rebecca J., Sandoval Cristina P., Feder J., Gries R., Crespi Bernard J., Gries G., Gompert Z., Nosil P. 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., Taylor B.L. 2017. Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. Mar. Mamm. 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. New York, NY, USA: Chapman and Hall/CRC Press.
Stankowski S., Ravinet M. 2021. Defining the speciation continuum. Evolution 75:1256–1273.
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. U.S.A. 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.P. 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 (2nd ed.). Measurement: Interdisciplinary Research and Perspectives 17:160–167.
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. Mammals of South America 1:128–156.
Wetzel R.M., Mondolfi E. 1979. The subgenera and species of long-nosed armadillos, genus *Dasypus* L. In: JF Eisenberg, editor. Vertebrate ecology in the northern neotropics. WA, 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., Kuhn M., Pedersen T., Miller E., Bache S., Müller K., Ooms J., Robinson D., Seidel D., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H. 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. https://dplyr.tidyverse.org/
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 1.1.3. https://wilkelab.org/cowplot/
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. Boston, MS, USA: 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.