CellPress

Correspondence

The phylogenetic affinities of the extinct glyptodonts

Frédéric Delsuc[1,*], Gillian C. Gibb[1,2], Melanie Kuch[3], Guillaume Billet[4], Lionel Hautier[1], John Southon[5], Jean-Marie Rouillard[6], Juan Carlos Fernicola[7], Sergio F. Vizcaíno[8], Ross D.E. MacPhee[9], and Hendrik N. Poinar[3,*]

Among the fossils of hitherto unknown mammals that Darwin collected in South America between 1832 and 1833 during the Beagle expedition [1] were examples of the large, heavily armored herbivores later known as glyptodonts. Ever since, glyptodonts have fascinated evolutionary biologists because of their remarkable skeletal adaptations and seemingly isolated phylogenetic position even within their natural group, the cingulate xenarthrans (armadillos and their allies [2]). In possessing a carapace comprised of fused osteoderms, the glyptodonts were clearly related to other cingulates, but their precise phylogenetic position as suggested by morphology remains unresolved [3,4]. To provide a molecular perspective on this issue, we designed sequence-capture baits using *in silico* reconstructed ancestral sequences and successfully assembled the complete mitochondrial genome of *Doedicurus* sp., one of the largest glyptodonts. Our phylogenetic reconstructions establish that glyptodonts are in fact deeply nested within the armadillo crown-group, representing a distinct subfamily (Glyptodontinae) within family Chlamyphoridae [5]. Molecular dating suggests that glyptodonts diverged no earlier than around 35 million years ago, in good agreement with their fossil record. Our results highlight the derived nature of the glyptodont morphotype, one aspect of which is a spectacular increase in body size until their extinction at the end of the last ice age.

Although the phylogenetic unity of order Cingulata has never been seriously questioned, how its three constituent groups (armadillos, glyptodonts, and pampatheres) are related to one another has been difficult to resolve in fine detail. Of special interest in this regard is the recent proposal that, despite numerous differences in body size and carapace structure, glyptodonts do not constitute a sister-group to armadillos, as traditionally assumed [2], but are instead nested within them [4,6]. This hypothesis is, however, based on a restricted set of cranio-dental characters. Here, we put this proposition to the test by analyzing the mitochondrial genome of a specimen of the late surviving glyptodont *Doedicurus*. One of the largest members of its clade, with an estimated body mass of ~1.5 tons [7], *Doedicurus* exhibited numerous distinctive characters, famously including a club-shaped, armored tail adorned with spikes, presumably used in intraspecific combat.

Using ancient DNA (aDNA) extraction techniques, we recovered endogenous DNA from a carapace fragment (MACN Pv 6744) dated to 12,015 ± 50 ^{14}C radiocarbon years before present (Supplemental information). Utilizing a recently assembled dataset encompassing all modern xenarthran species [5], we reconstructed, *in silico*, a set of ancestral mitogenomic sequences, which permitted the synthesis of a set of target capture RNA baits. Baits constructed in this way may allow for a more specific sequence capture of phylogenetically distant ancient specimens than baits based solely on available modern sequences. This permitted the reconstruction of a nearly complete mitochondrial genome of *Doedicurus* at 76x coverage. Illumina reads mapping to the newly assembled *Doedicurus* mitogenome were 45 base pairs on average and displayed C-to-T damage patterns at both 3' and 5' ends, characteristic of authentic aDNA. We have ruled out the possibility of the inadvertent enrichment of nuclear copies of mitochondrial origin (NUMTs) by performing additional phylogenetic controls (Supplemental information).

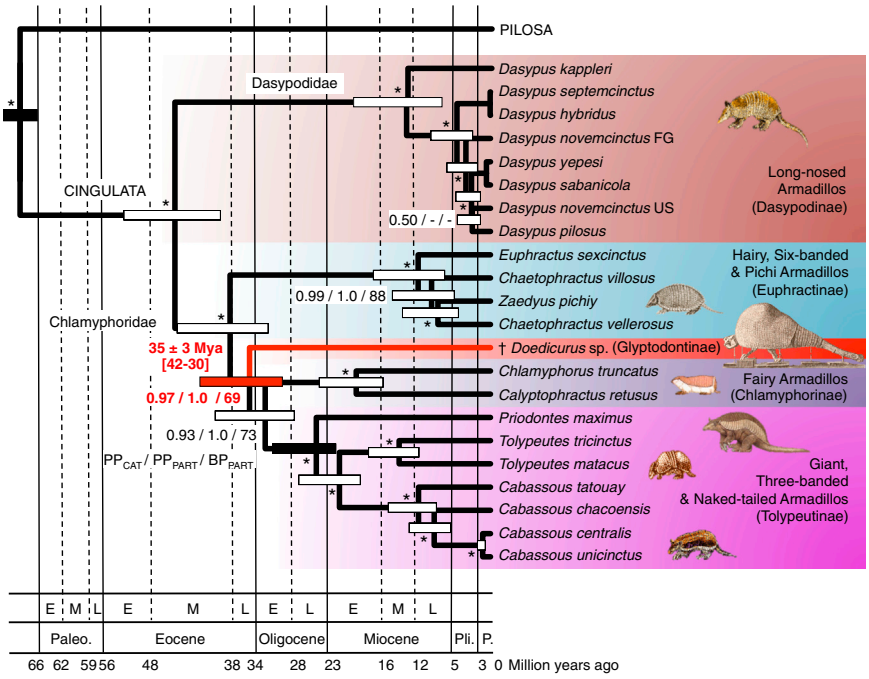

Figure 1. Phylogenetic position of glyptodonts.
Phylogeny and molecular timescale of extant armadillos including the extinct glyptodont *Doedicurus* sp. (in red). Bayesian chronogram was obtained using a rate-autocorrelated log-normal relaxed molecular clock model using PhyloBayes under the CAT-GTR-G mixture model with a birth death prior on the diversification process, and six soft calibration constraints. Mean divergence dates and associated 95% credibility intervals are represented as node bars. Plain black node bars indicated calibration constraints. The main geological periods follow Geological Time Scale of the Geological Society of America (E = Early, M = Middle, L = Late; Paleo. = Paleocene, Pli. = Pliocene, P. = Pleistocene). Statistical support values obtained from three different phylogenetic reconstruction methods (PP_{CAT}: Bayesian Posterior Probability under the CAT-GTR+G mixture model; PP_{PART}: Bayesian PP under the best partition model; BP_{PART}: Maximum likelihood Bootstrap Percentage under the best partition model) are indicated with stars corresponding to nodes with PP > 0.95 and BP > 90. The full chronogram and phylogram are provided in Figure S2.

By comparing our ancient mitogenome to those of living xenarthrans (Figure 1), we were able to confidently place *Doedicurus* within armadillos as the sister-group of a clade composed of Chlamyphorinae (fairy armadillos) and Tolypeutinae (three-banded, naked-tailed and giant armadillos; Supplemental information). This clearly contradicts the old view that glyptodonts must have diverged from other cingulates at a very early point in their phylogenetic history, on the grounds that, for example, they possessed such features as a completely fused carapace lacking movable bands [3]. Our results are more compatible but still incongruent with recent morphological cladistic analyses [4,6] that position glyptodonts within a more inclusive but nevertheless paraphyletic Euphractinae.

To examine the consequences of this novel phylogenetic placement, we incorporated *Doedicurus* into the morphological character matrix of Billet *et al*. [6], but were unable to identify any exclusive synapomorphies justifying grouping of the former with Chlamyphorinae + Tolypeutinae. In our study, only two characters, pertaining to the shape and position of the mandibular coronoid process, might qualify as potential synapomorphies, but only under the assumption that both have reverted to ancestral states in three-banded armadillos. This analysis nevertheless revealed other morphological similarities between glyptodonts and fairy armadillos (Supplemental information).

We estimate that glyptodonts diverged from Chlamyphorinae + Tolypeutinae 35 ± 3 million years ago, close to the Eocene–Oligocene transition (Figure 1). This molecular estimate is compatible with the age of the oldest and widely accepted glyptodont remains (Mustersan *Glyptatelus* osteoderms [8], ca. 36–38 million years old [9]). Tarsal bones from the Early Eocene locality of Itaboraí (Brazil), currently dated to more than 50 Myr [9]), have been interpreted as glyptodont, but the elements in question are better interpreted as belonging to indeterminate dasypodoids [10]. According to our results, they might belong to stem cingulates that evolved before basal divergences occurred within the armadillo crown group, an event we date to ca. 45 million years ago (Figure 1).

While our results are based strictly on the comparison of mitogenomes, the global congruence observed with previous nuclear-based phylogenies as well as molecular dating analyses provides convincing evidence for the proposed xenarthran evolutionary history [5]. On this evidence, glyptodonts (Glyptodontinae) comprised a distinct, Late Paleogene lineage of chlamyphorid armadillos [5]. Such a radical repositioning of glyptodonts within the armadillo crown group has major consequences for interpreting aspects of cingulate evolution. For example, the dome-shaped, tightly-fused carapace of glyptodonts has long been thought to be fundamentally different from that of armadillos and pampatheres, in which the carapace consists of articulated sections. Our results imply that the unarticulated carapace is in fact a derived feature, which in turn provides an explanation for the apparent presence of movable bands in some Miocene glyptodonts [3].

Glyptodonts were a group of ambulatory specialized herbivores that reached giant size bracketed between two extant clades of armadillos that do not share either of these characteristics. Based on our new phylogenetic framework, we performed a statistical reconstruction of ancestral body masses. According to our analysis, the mean ancestral body mass estimate of the last common ancestor of Glytodontinae + Chlamyphorinae + Tolypeutinae was a mere 6 kg (95% credibility interval: 1–19 kg), implying a spectacular increase in glyptodont body mass during the Neogene (Supplemental Information). This inference is in line with the fossil record, which indicates that glyptodonts evolved from medium-sized forms in the Miocene (e.g., *Propalaehoplophorus*, ~80 kg) to become true megafauna in the Pleistocene (e.g., *Glyptodon clavipes*, ~2,000 kg) before disappearing with most other South American large mammals some 10,000 years ago [7].

SUPPLEMENTAL INFORMATION

Supplemental Information including results, acknowledgements, experimental procedures and two figures can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.01.039.

REFERENCES

1. Fernicola, J.C., Vizcaíno, S.F., and De Iuliis, G. (2009). The fossil mammals collected by Charles Darwin in South America during his travels on board the HMS Beagle. Revist. Asoc. Geol. Argent. *64*, 147–159.
2. Hoffstetter, R. (1958). Xenarthra. In Traité de paléontologie, Vol. 2, no. 6., P. Piveteau, ed. (Masson et Cie: Paris), pp. 535–636.
3. Engelmann, G.F. (1985). The phylogeny of the Xenarthra. In The evolution and ecology of armadillos, sloths, and vermilinguas. G.G. Montgomery, ed. (Smithsonian Institution Press: Washington DC), pp. 51–64.
4. Gaudin, T.J., and Wible, J.R. (2006). The phylogeny of living and extinct armadillos (Mammalia, Xenarthra, Cingulata): a craniodental analysis. In Amniote paleobiology: perspectives on the evolution of mammals, birds and reptiles, M.T. Carrano, T.J. Gaudin, R.W. Blob, and J.R. Wible, eds. (University of Chicago Press: Chicago), pp. 153–198.
5. Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H.N., and Delsuc, F. (in press). Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. Mol. Biol. Evol. http://dx.doi.org/10.1093/molbev/msv250.
6. Billet, G., Hautier, L., De Muizon, C., and Valentin, X. (2011). Oldest cingulate skulls provide congruence between morphological and molecular scenarios of armadillo evolution. Proc. Biol. Sci. *278*, 2791–2797.
7. Vizcaíno, S.F., Cassini, G.H., Toledo, N., and Bargo, M.S. (2012). On the evolution of large size in mammalian herbivores of Cenozoic faunas of southern South America. In Bones, clones and biomes: an 80-million year history of recent Neotropical mammals, B. Patterson and L. Costa, eds. (University of Chicago Press: Chicago), pp. 76–101.
8. Simpson, G.G. (1948). The beginning of the age of mammals in South America. Part 1. Introduction. Systematics: Marsupialia, Edentata, Condylarthra, Litopterna, and Notioprogonia. Bull. Am. Mus. Nat. Hist. *91*, 1–227.
9. Woodburne, M.O., Goin, F.J., Bond, M., Carlini, A.A., Gelfo, J.N., López, G.M., Iglesias, A., and Zimicz, A.N. (2014). Paleogene land mammal faunas of South America; a response to global climatic changes and indigenous floral diversity. J. Mammal. Evol. *21*, 1–73.
10. Bergqvist, L.P., Abrantes, E.A., and Dos Santos Avilla, L. (2004). The Xenarthra (Mammalia) of São José de Itaboraí basin (Upper Paleocene, Itaboraian), Rio de Janeiro, Brazil. Geodiversitas *26*, 323–337.

[1]Institut des Sciences de l'Evolution, UMR 5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France. [2]Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand. [3]McMaster Ancient DNA Centre, Department of Anthropology, Biology and Biochemistry, McMaster University, Hamilton, Canada. [4]Sorbonne Universités, CR2P, UMR 7207, CNRS, Muséum National d'Histoire naturelle, Université Paris 06, Paris, France. [5]Keck-Carbon Cycle AMS facility, Department of Earth System Science, University of California, Irvine, CA, USA. [6]MYcroarray, Ann Arbor, MI 48105, United States / Chemical Engineering Department, University of Michigan, Ann Arbor, MI, USA. [7]Consejo Nacional de Investigaciones Científicas y Técnicas, Sección Paleontología de Vertebrados, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires, Argentina. [8]División Paleontología Vertebrados, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, CONICET, La Plata, Argentina. [9]Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY, USA.
*E-mail: Frederic.Delsuc@umontpellier.fr (F.D.), poinarh@mcmaster.ca (H.N.P.)

Supplemental Information

The phylogenetic affinities of the extinct glyptodonts

Frédéric Delsuc, Gillian C. Gibb, Melanie Kuch, Guillaume Billet, Lionel Hautier, John Southon, Jean-Marie Rouillard, Juan Carlos Fernicola, Sergio F. Vizcaíno, Ross D. E. MacPhee, Hendrik N. Poinar

Supplemental Results

Ancestral sequence reconstruction and capture specificity

Our study is based on a fragment of *Doedicurus* sp. carapace that contains a small quantity of endogenous DNA (Figure S1a). Given the low percent on-target endogenous DNA in this sample, shotgun Illumina sequencing would have been cost prohibitive. Targeted enrichment, using baits designed on the basis of sequence data from modern species, has been shown to provide an efficient way of retrieving endogeneous genetic material from the DNA extracts of extinct remains [S1]. However, in cases in which the extinct taxon has uncertain phylogenetic affinities, it may be advantageous to use baits designed from ancestral sequences reconstructed *in silico* from a phylogenetic analysis of modern sequences. Results using these ancestrally reconstructed baits show that our procedure allows for the capture of endogenous mitochondrial reads of extinct *Doedicurus* with increased specificity when compared to baits designed strictly on modern species genomes (Fig. S2b). As can be seen, the percentage of baits mapping to the assembled *Doedicurus* mitogenome at different thresholds (allowing from 5% to 40% mismatches) was consistently higher for ancestrally designed baits than it was for baits derived from extant species alone. This is particularly evident for the more stringent mapping thresholds (5% to 20%) in which higher mapping percentages were obtained with baits designed from the ancestral xenarthran mitogenome sequence. For example, when 15% mismatches are allowed between the reads and the *Doedicurus* mitogenome, 61% of the baits designed from the ancestral xenarthran mitogenome mapped, versus only 6% of the baits designed from the extant *Bradypus tridactylus* mitogenome. Even though low specificity baits (up to 40% sequence divergence) have been shown to be relatively efficient at capturing target regions [S2], our procedure based on ancestral sequence reconstruction appears to improve the specificity of capture in the case of *Doedicurus*, which represents a distinct armadillo lineage without close living relatives. As specificity is important in aDNA studies in which target sequences typically represent a minority of sequence data within an extract, the method presented here could potentially prove useful for future aDNA studies.

Doedicurus mitogenome validation

In order to verify the authenticity of our reconstructed *Doedicurus* mitogenome, we examined all our reads for the presence of an excess of C-T and G-A transitions caused by post-mortem mutations, commonly found in ancient DNA, using mapDamage 2.0 [S3]. The reads mapping to the *Doedicurus* assembled mitogenome exhibit substantial levels of DNA

damage with up to 25% cytosine deamination on the 5' strand and 20% on the 3' strand (Figure S1c). Such a pattern is typical of ancient DNA molecules and strongly argues in favor of the endogenous origin of the captured reads. We also analyzed our sequences in a systematic fashion to ensure that they did not represent nuclear copies of mitochondrial DNA (NUMTs). The final version of the *Doedicurus* mitogenome was aligned with all available xenarthran mitogenomes and the protein-coding regions were checked to confirm that no indels or stop codons originating from potential NUMTs integration in the assembly were present. To further rule out the presence of NUMTs, we cut our final alignment into thirty-one 500bp fragments and performed a variety of tests. First, we verified that all 31 *Doedicurus* 500bp fragments have sequence similarity with other xenarthran mitogenomes using BLASTN searches against the NCBI non-redundant (nr) database. Each fragment had its best hit to available modern xenarthran mitogenomes but without any perfect match, which rules out contaminants represented in the database. Second, we reconstructed maximum likelihood (ML) trees using RAxML [S4] under the GTRGAMMA model for each of the 31 partitions. None of the inferred maximum likelihood trees showed a spurious relationship or a particularly long or short branch for *Doedicurus*, which was always recovered nested within armadillos. We feel these results argue in favor of a genuine, mitochondrial genome from this specimen of an extinct taxon.

Phylogenetic results

Analyses of the mitogenomic dataset using Bayesian and ML methods implemented in both partitioned and mixture models, resulted in the same strongly supported topology (Fig. S2a,b). The statistical support for the positioning of *Doedicurus* within Chlamyphoridae, as a sister-group to Chlamyphorinae + Tolypeutinae, was high with the Bayesian mixture model (PP_{CAT} = 0.97), maximal with the Bayesian mixed model based on the best fitting partitioned scheme (PP_{PART} = 1), and more moderate with the partitioned ML model (BP_{PART} = 69). The monophyly of Chlamyphorinae + Tolypeutinae showed similar support (PP_{CAT} = 0.93; PP_{PART} = 1; BP_{PART} = 73). With the exception of one particular node within the genus *Dasypus*, all nodes were strongly supported and congruent with the recently published complete xenarthran mitogenomic study of Gibb et al. [S5]. This mitogenomic phylogeny of xenarthrans is fully congruent with previous studies conducted at the genus level, and using nuclear exons [S6,S7] as well as non-coding retroposon flanking sequences [S8]. The inferred molecular timescale (Fig. 2c,d) is also consistent with the ones previously obtained with nuclear data [S9]. As discussed in Gibb et al. [S5], the mitochondrial/nuclear congruence strongly argues for the adequacy of mitogenomic data for reconstructing xenarthran evolutionary history in terms of both phylogenetic relationships and divergence times.

The analysis of morphological data onto the molecular topology, via character mapping using maximum parsimony, failed to identify any unambiguous synapomorphies for the grouping of *Doedicurus* with the Chlamyphorinae + Tolypeutinae clade. Only two characters (Character 14, state 1, and Character 15, state 1) pertaining to the shape and position of the

mandibular coronoid process could represent potential synapomorphies under the assumption that both have reverted to ancestral states in three-banded armadillos (genus *Tolypeutes*). Nevertheless, some intriguing morphological resemblances between glyptodonts and chlamyphorines have been revealed by these analyses. For example, both chlamyphorines and glyptodonts are characterized by a ventral surface of the auditory region that is located well dorsal to the palate (Character 78, state 1). This derived character is also present in pampatheres, which are conventionally allied to glyptodonts based on morphological data [S10]. The dorsal shift of the auditory region may represent an important character for disentangling the origins of glyptodonts, but it is absent in tolypeutines and therefore appears inconsistent with our molecular hypothesis. In any case, such morphological resemblances, combined with our newly proposed position of glyptodonts within crown cingulates, raises the distinct possibility that morphological features relevant to armadillo systematics may have been overlooked.

Supplemental Experimental Procedures

Fossil specimen and dating

A fragment of the carapace of *Doedicurus* sp. (Figure S1a) was sampled from the collection of the Museo Argentino de Ciencias Naturales “Bernardino Rivadavia” (MACN) in Buenos Aires (Argentina). The specimen, MAC2010.18 in our voucher list, bears a museum label reading: “MACN Pv 6744, *Doedicurus*, carapacio”. Although we cannot be certain, it was most likely collected in the late 19th century by Carlos Ameghino somewhere along the Río Salado in Buenos Aires province (Argentina). An aliquot of freeze-dried ultrafiltered gelatin prepared from the sample was dated at 12,015 ± 50 ^{14}C yrpb (radiocarbon years before present) by the Keck Carbon Cycle AMS facility of the University of California Irvine (USA).

DNA extraction and library preparation

The specimen was initially subsampled by one of the authors (R.D.E.M.) at MACN in 2010. A 100mg subsample was processed in a dedicated ancient DNA laboratory facility at the McMaster Ancient DNA Centre. Using a hammer and chisel we further reduced the carapace subsample to small particle sizes of 1-5mm. The subsample was then demineralized with 0.5M EDTA (pH 8.0) for 24h at room temperature, and the supernatant removed following centrifugation. The pellet was digested using a Tris-HCl-based proteinase K digestion solution with 0.5% sodium lauryl sarcosine (Fisher Scientific), 1% polyvinylpolypyrrolidone (PVP, Fisher scientific), 50mM dithiothreitol, 2.5mM N-phenacyl thiazolium bromide (PTB, Prime Organics), and 5mM calcium chloride (CaCl2). Proteinase digestions were performed for 24h at room temperature with agitation. Following centrifugation, the digestion supernatants were removed and pooled with the demineralization supernatants. We repeated this process three times, pooling supernatants with the original rounds. Organics were then extracted from the pooled supernatants using phenol:chloroform:isoamyl alcohol

(PCI, 25:24:1), and the resulting post-centrifugation aqueous solution was again extracted with chloroform. The final aqueous solution was concentrated using 10kDA Amicon centrifuge filters (Millipore) 14k × g, with up to four washes of 0.1× TE buffer (pH 8) to provide a final desalted concentrate of 50μL.

We purified the extract with a MinElute column (QIAGEN) to 50μL EBT and converted it to a double-stranded, Illumina sequencing library according to the protocol of Meyer and Kircher [S11]. The resulting library was then double-indexed with P5 and P7 indexing primers [S12] and purified again with MinElute to 15μL EBT. It has been previously shown that a short single-locus quantitative PCR assay can be used with some accuracy to predict on-target ancient DNA high-throughput sequencing read counts both before and after targeted enrichment [S13]. Therefore, to ensure that we had endogenous signal after library preparation, we screened our library for a 47bp portion of the xenarthran mitochondrial 16S rRNA gene using the following quantitative PCR protocol employing 1μL of the library in a total reaction volume of 10ul: 1X PCR Buffer II, 2.5 mM MgCl2, 250 μM dNTP mix, 1 mg/ml BSA, 250 nM each primer (Xen_16S_F2, Xen_16S_R2), 0.167X SYBRgreen, 0.5 U AmpliTaq Gold. The library of the extract was positive and the library of the extraction blank was clean.

Ancestral sequence reconstruction and bait design

In order to maximize the capture of potentially divergent sequences from extinct xenarthran taxa, we designed a xenarthran bait set which was composed of 100bp baits tiled every 5 bases across a representative sample of living xenarthran mitochondrial genome sequences [S12] as well as 13 ancestrally inferred mitochondrial genomes including the ancestral xenarthran mitogenome. Ancestral sequences were inferred for each node of the modern xenarthran tree under a single GTR+G model using the program baseml of the PAML 4 package [S14]. The variable tandem repeat section (VNTR) of the D-loop was masked with 10 Ns prior to bait design, as it is too long to resolve with short read sequencing. In light of evidence that bait coverage across targets can result in coverage biases in target read coverage [S13,S15], we chose not to collapse baits of identical sequence prior to manufacture. The 5,207 final baits were then synthesized at MYcroarray (http://www.mycroarray.com/) as part of several MYbaits targeted enrichment kits.

Target enrichment and Illumina sequencing

We performed a first round of enrichment at 50°C followed by a second round at 55°C using 7.47μL indexed library for 36-38 hours, following the manufacturer's protocol. Phosphate-group end-blocked oligonucleotides matching one strand of the regions flanking the 7bp indexes of the library adapters were included. We used 25ng of baits per reaction, which is what we have found to be sufficient for very sensitive capture of a small target region [S13]. Following hybridization, the reaction was cleaned according to the suggested protocol except that we used 200μL rather than 500μL volumes of wash buffers for each wash step, to accommodate a 96-well plate-format. Hot washes were performed at 50/55°C. The enriched

library was eluted and then purified with MinElute to 15μL EBT, which we then re-amplified according to the protocol above and again purified this time to 10μL EBT.

The enriched library was pooled and sequencing was performed on the Illumina HiSeq system using the TruSeq Rapid (v1) chemistry with initial hybridization on the cBot. Each lane included a 1% spike-in of Illumina's PhiX v3 control library. Paired-end reads of either 111 or 85 bp were performed, along with dual 7 bp indexing on both runs. Raw data were processed with either HCS v2.0.10.0 or v2.0.12.0 and RTA v1.17.21.3. File conversion and demultiplexing using each 7bp reverse index (requiring a 100% match) was performed with the Illumina CASAVA software version 1.8.2.

Mitogenome assembly and annotation

Raw sequence reads were first trimmed to remove adapter and index tag sequences using CutAdapt [S16]. Duplicate paired reads were removed using FastUniq [S17] and the remaining reads were imported into Geneious R9 [S18]. Any potential human contamination was removed by mapping the reads to the human mitogenome using the 'low sensitivity' setting in Geneious, and any reads that mapped with a high similarity were discarded. The remaining trimmed reads were then mapped to the *Euphractus sexcinctus* reference mitogenome (Accession Number NC_028571) using the 'medium sensitivity' setting in Geneious. Iterative mapping to successive "hybrid" consensus genomes was repeated until no more reads were included. Any remaining regions that lacked any coverage were filled in with question marks (?) in order to create a draft mitogenome. Iterative mapping to successive draft genomes was repeated until there were no further improvements in extending coverage into the gap regions.

The completed draft mitogenome assembly was scanned by eye to check for the inclusion of any conflicting reads. These were examined using BLAST and any obvious contaminant reads were removed. To estimate the depth of coverage, the trimmed reads were merged using FLASH [S19] and re-mapped to the draft mitogenome. The mapping of 26,307 reads of mean length 45 bp led to a 76x coverage. For the final consensus mitogenome, all regions with less than 3x coverage were excluded, and the consensus called using 75% read agreement. The completed genome was aligned to the *E. sexcinctus* reference mitogenome and checked to confirm there was no accidental inclusion of *E. sexcinctus* sequence in the final version. The final *Doedicurus* mitogenome was annotated by alignment with published xenarthran mitogenomes and deposited in GenBank with Accession Number KU517659.

Mitogenomic dataset construction

We selected 31 representative living xenarthran species from the complete mitochondrial genome data set assembled by Gibb et al. [S5] for modern species together with three afrotherian outgroup taxa. We then added *Doedicurus* sequences (excluding the control region) and aligned each gene data set individually using MAFFT G-INSI [S20] within Geneious for the 22 tRNA and the two rRNA genes, and using the amino acid translation for

the 13 protein-coding genes. Selection of unambiguously aligned sites was performed on each individual data set with Gblocks [S21] using default relaxed settings and the codon option for protein-coding genes. The final concatenation contained 15,214 unambiguously aligned nucleotide sites for 35 taxa.

Molecular phylogenetics and dating analyses

The best-fitting partition schemes and associated optimal models of sequence evolution were determined using PartitionFinder v1.1.1 [S22] for subsequent use in phylogenetic analyses. The greedy algorithm was used starting from 41 a priori defined partitions corresponding to the three codon positions of the 13 protein-coding genes (3 x 13 = 39 partitions), the two rRNAs (1), and all 24 tRNAs (1). Branch lengths have been unlinked among partitions and the Bayesian Information Criterion (BIC) was used for selecting the best-fitting partition scheme. The best partitioning scheme consisted of three separate GTR+I+G models for the following partitions: 1) ATP6_p1, ATP8_p1, ATP8_p2, ATP8_p3, CYTB_p1, ND1_p1, ND2_p1, ND3_p1, ND4L_p1, ND4_p1, ND5_p1, ND6_p2, ND6_p3, rRNAs, tRNAs; 2) ATP6_p2, COX1_p1, COX1_p2, COX2_p1, COX2_p2, COX3_p1, COX3_p2, CYTB_p2, ND1_p2, ND2_p2, ND3_p2, ND4L_p2, ND4_p2, ND5_p2; 3) ATP6_p3, COX1_p3, COX2_p3, COX3_p3, CYTB_p3, ND1_p3, ND2_p3, ND3_p3, ND4L_p3, ND4_p3, ND5_p3, ND6_p1. ML reconstruction was conducted with RAxML using separated GTRGAMMAI models with branches unlinked for each of the three best-fitting partitions determined by PartitionFinder. Maximum Likelihood bootstrap values (BP_{PART}) were computed by repeating the same ML heuristic search using 100 pseudo-replicates.

Bayesian phylogenetic inference under a mixed model was conducted using the MPI version of MrBayes 3.2.3 [S23]. Following PartitionFinder results, we used separate GTR+G8+I models for each of the three selected partitions with parameters unlinked across partitions. Two independent runs of four incrementally-heated MCMCMC starting from a random tree were performed. MCMCMC were run for 1,000,000 generations with trees and associated model parameters being sampled every 1,000 generations. The initial 250 trees in each run were discarded as burn-in samples after convergence checking. The 50% majority-rule Bayesian consensus tree and the associated posterior probabilities (PP_{PART}) were then computed from the 1,500 combined trees sampled in the two independent runs. Bayesian phylogenetic reconstruction was also conducted under the CAT-GTR-G_4 mixture model using PhyloBayes MPI 1.5a [S24]. Two independent Markov Chain Monte Carlo (MCMC) starting from a random tree were run for 50,000 cycles with trees and associated model parameters sampled every 10 cycles during 2,750,000 tree generations. The initial 500 trees (10%) sampled in each MCMC run were discarded as the burn-in after convergence checking using PhyloBayes diagnostic tools bpcomp and tracecomp. The 50% majority-rule Bayesian consensus tree and the associated posterior probabilities (PP_{CAT}) were then computed from the remaining combined 9,000 (2 x 4,500) trees.

Molecular dating analyses were conducted using PhyloBayes 3.3f [S25] under the CAT-GTR+G mixture model and a log-normal autocorrelated relaxed clock with a birth–death prior

on divergence times combined with soft fossil calibrations. We used the same six fossil calibrations and root prior (100 Myr) as in Gibb et al. [S5]. Calculations were conducted by running two independent MCMC chains for a total 50,000 cycles sampling parameters every 10 cycles. The first 500 samples (10%) of each MCMC were excluded as the burn-in after convergence diagnostics. Posterior estimates of divergence dates were then computed from the remaining 4,500 samples of each MCMC.

Bayesian reconstruction of body mass evolution was performed with CoEvol 1.2 [S26] using the concatenation of the codons of the 12 protein-coding genes encoded on the light strand (excluding ND6). This approach allows joint reconstruction of variations in molecular evolutionary rates, divergence times, and continuous variables by modeling these parameters as a multivariate Brownian diffusion process along the branches of the phylogenetic tree while taking their covariance into account. The Bayesian chronogram was used as a fixed topology to reconstruct ancestral body mass under the dsom procedure, which allows accounting for variation in non-synonymous to synonymous substitution ratio. Body mass values for extant xenarthrans were obtained from the PanTheria database [S27]. We used an estimated body mass of 1,468 kg for *Doedicurus* [S28]. CoEvol was run for 5,000 cycles sampling parameters at each step. The first 500 samples were discarded as the burnin and posterior averages were estimated on the remaining 4,500 points.

Morphological character matrix construction and analysis

In order to test if the pattern we found by the molecular analysis could be supported by the morphological data currently available, we added *Doedicurus* to the cranio-dental matrix of Billet et al. [S29], which is in turn primarily based on the matrix assembled by Gaudin and Wible [S10]. *Doedicurus* was scored from the cranio-dental anatomy of specimen MACN PV 2762 (*Doedicurus clavicaudatus*) conserved at the Museo Argentino de Ciencias Naturales “Bernardino Rivadavia” in Buenos Aires (Argentina). This matrix of 125 morphological characters was reduced to 11 terminal taxa representing extant cingulate genera and *Doedicurus*, plus two pilosan outgroups. The previously inferred molecular scaffold was used as follows: ((*Bradypus*, *Tamandua*), (*Dasypus*, ((*Euphractus*, (*Chaetophractus*, *Zaedyus*)), (*Doedicurus*, (*Chlamyphorus*, (*Priodontes*, (*Cabassous*, *Tolypeutes*)))))))) to retrace morphological character history using maximum parsimony in Mesquite [S30].

All datasets are available upon request.

Author contributions

Conceptualization, F.D., R.D.E.M. and H.N.P.; Formal Analysis, F.D., G.C.G., M.K., G.B., L.H. and H.N.P; Investigation, F.D., G.C.G., M.K., G.B., L.H., J.S., R.D.E.M. and H.N.P.; Resources, F.D., J.S., J.-M.R., R.D.E.M. and H.N.P.; Writing – Original Draft, F.D., G.B., L.H., J.C.F., S.F.V., R.D.E.M. and H.N.P.; Writing – Review & Editing, F.D., G.C.G., M.K., G.B., L.H., R.D.E.M. and H.N.P.; Supervision, F.D. and H.N.P.; Funding Acquisition, F.D. and H.N.P.

Acknowledgments

This paper is dedicated to the memory of Audrey Illovies. Alejandro Kramarz (Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires, Argentina) kindly provided access to the Doedicurus specimen. This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS), the Scientific Council of Université Montpellier 2 (UM2), and Investissement d'Avenir grants of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01) to FD, the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs program to HNP. HNP thanks the McMaster Ancient DNA Centre members and D. Poinar for constructive feedback. J.-M.R. has financial interest in MYcroarray (Ann Arbor, MI, USA) who provided in-solution capture baits used in this work. All other authors declare that they have no conflicts of interest. This is contribution ISEM 2015-XXX-S of the Institut des Sciences de l'Evolution de Montpellier.

Supplemental References

S1. Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Fortea, J., Rosas, A., Pääbo, S. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science *325*, 318-321.

S2. Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G. J. (2013). Capturing protein-coding genes across highly divergent species. Biotechniques *54*, 321-326.

S3. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013). mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682-1684.

S4. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312-1313.

S5. Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H.N., and Delsuc, F. (in press). Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. Mol. Biol. Evol. doi:10.1093/molbev/msv250.

S6. Delsuc, F., Scally, M., Madsen, O., Stanhope, M. J., De Jong, W. W., Catzeflis, F. M., Springer, M. S., Douzery, E. J. P. (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol. Biol. Evol. *19*, 1656-1671.

S7. Delsuc, F., Superina, M., Tilak, M. K., Douzery, E. J. P., Hassanin, A. (2012). Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. Mol. Phylogenet. Evol. *62*, 673-680.

S8. Möller-Krull, M., Delsuc, F., Churakov, G., Marker, C., Superina, M., Brosius, J., Douzery, E. J. P., Schmitz, J. (2007). Retroposed elements and their flanking regions resolve the

evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths). Mol. Biol. Evol. *24*, 2573-2582.

S9. Delsuc, F., Vizcaíno, S. F., Douzery, E. J. P. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. BMC Evol. Biol. *4*, 11.

S10. Gaudin, T.J., and Wible, J.R. (2006). The phylogeny of living and extinct armadillos (Mammalia, Xenarthra, Cingulata): a craniodental analysis. In Amniote paleobiology: perspectives on the evolution of mammals, birds and reptiles, M.T. Carrano, T.J. Gaudin, R.W. Blob, J.R. Wible, eds. (University of Chicago Press: Chicago), pp 153-98.

S11. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, t5448.

S12. Kircher, M., Sawyer, S., and Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

S13. Enk, J., Rouillard, J. M., and Poinar, H. (2013). Quantitative PCR as a predictor of aligned ancient DNA read counts following targeted enrichment. BioTechniques *55*, 300-309.

S14. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586-1591.

S15. Mokry, M., Feitsma, H., Nijman, I. J., de Bruijn, E., van der Zaag, P. J., Guryev, V., and Cuppen, E. (2010). Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. Nucleic Acids Res. *38*, e116-e116.

S16. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10-12.

S17. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., and Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One *7*, e52249.

S18. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

S19. Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957-2963.

S20. Katoh, K., Kuma, K. I., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. *33*, 511-518.

S21. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540-552.

S22. Lanfear, R., Calcott, B., Ho, S. Y., and Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. *29*, 1695-1701.

S23. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539-542.

S24. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI. Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611-615.

S25. Lartillot, N., Lepage, T., & Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286-2288.

S26. Lartillot, N., & Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. *28*, 729-744.

S27. Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A., Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M., and Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. Ecology *90*, 2648-2648.

S28. Vizcaíno, S.F., Cassini, G.H., Toledo, N., and Bargo, M.S. (2012). On the evolution of large size in mammalian herbivores of Cenozoic faunas of southern South America. In Bones, clones and biomes: an 80-million year history of recent Neotropical mammals, B. Patterson and L. Costa, eds. (University of Chicago Press: Chicago), pp. 76-101.

S29. Billet, G., Hautier, L., De Muizon, C., and Valentin, X. (2011). Oldest cingulate skulls provide congruence between morphological and molecular scenarios of armadillo evolution. Proc. Biol. Sci. *278*, 2791-2797.

S30. Maddison, W. P., & Maddison, D. R. (2015). Mesquite: a modular system for evolutionary analysis. Version 2.75. *URL http://mesquiteproject. org*.

Supplemental Figures

Figure S1. (a) Carapace fragment of *Doedicurus* sp. used for DNA analyses. (b) Sequence capture specificity of different bait sets (ASR: ancestral sequence reconstruction) expressed as the percentage of baits mapping to the assembled *Doedicurus* mitogenome at different thresholds (from 5% to 40% mismatches allowed). (c) Length distribution of Illumina reads used to reconstruct the *Doedicurus* mitogenome. (d) Cytosine deamination as indicated via fragment misincorporation plots with 5' and 3' frequency for both strands of sequenced reads.

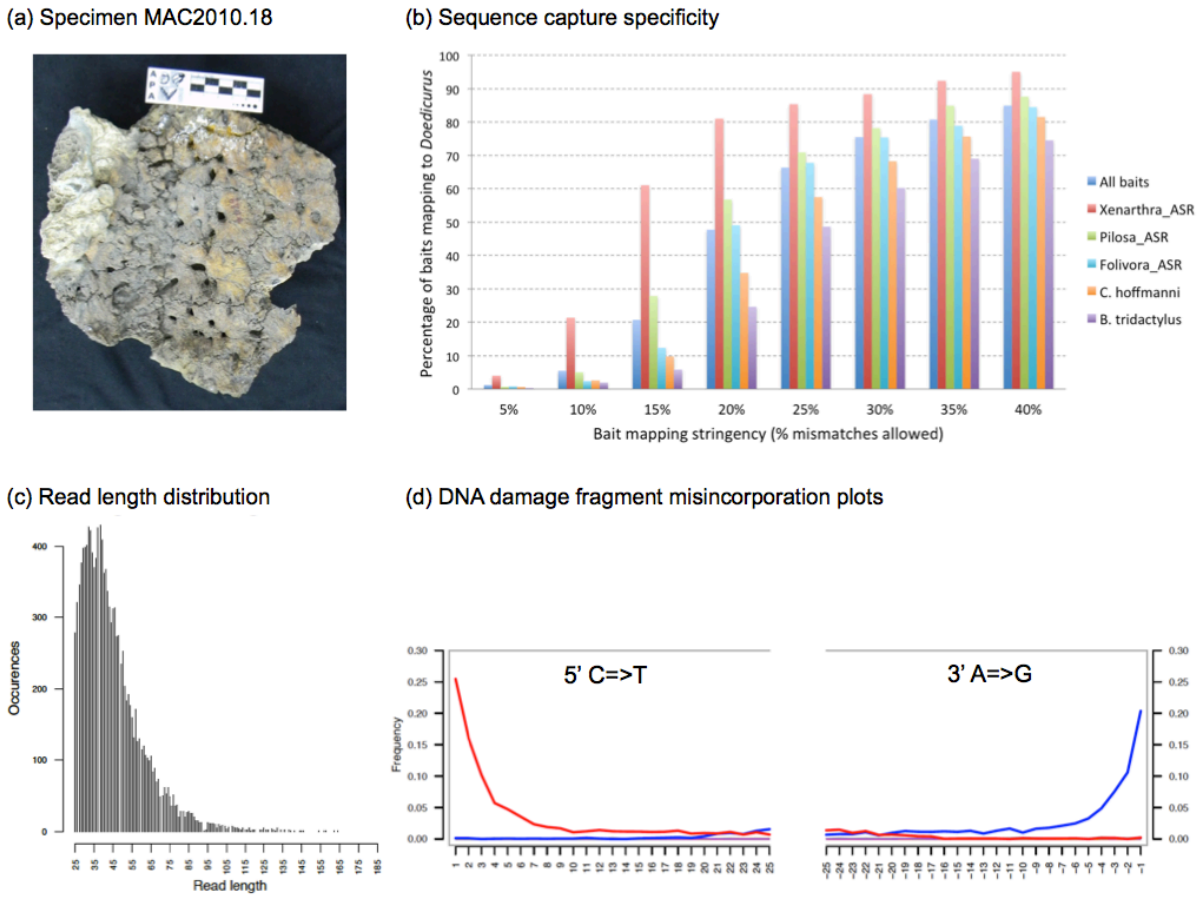

Figure S2. (a) Illustrated excerpt of the Bayesian consensus phylogram obtained using PhyloBayes under the CAT-GTR+G mixture model focused on armadillos. Statistical support values obtained from three different phylogenetic reconstruction methods (PP_{CAT}: Bayesian Posterior Probability under the CAT-GTR+G mixture model; PP_{PART}: Bayesian PP under the best partition model; BP_{PART}: Maximum likelihood Bootstrap Percentage under the best partition model) are indicated with stars corresponding to nodes with PP > 0.95 and BP > 90. (b) Full Bayesian consensus phylogram obtained using PhyloBayes under the CAT-GTR+G mixture model. (c) Molecular timescale for representative extant xenarthran species including the extinct glyptodont *Doedicurus*. The Bayesian chronogram was obtained using a rate-autocorrelated log-normal relaxed molecular clock model using PhyloBayes under the CAT-GTR+G mixture model with a birth-death prior on the diversification process, and six soft calibration constraints. Node bars indicate the uncertainty around mean age estimates based on 95% credibility intervals. (d) Bayesian reconstruction of body mass evolution in xenarthrans. The circle for *Doedicurus* has been scaled down by a factor of 10 for graphical purpose.