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

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

[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]Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden

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

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

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

[7]Laboratório de Biologia Evolutiva e Conservação de Vertebrados (Labec), Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, S˜

Article

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

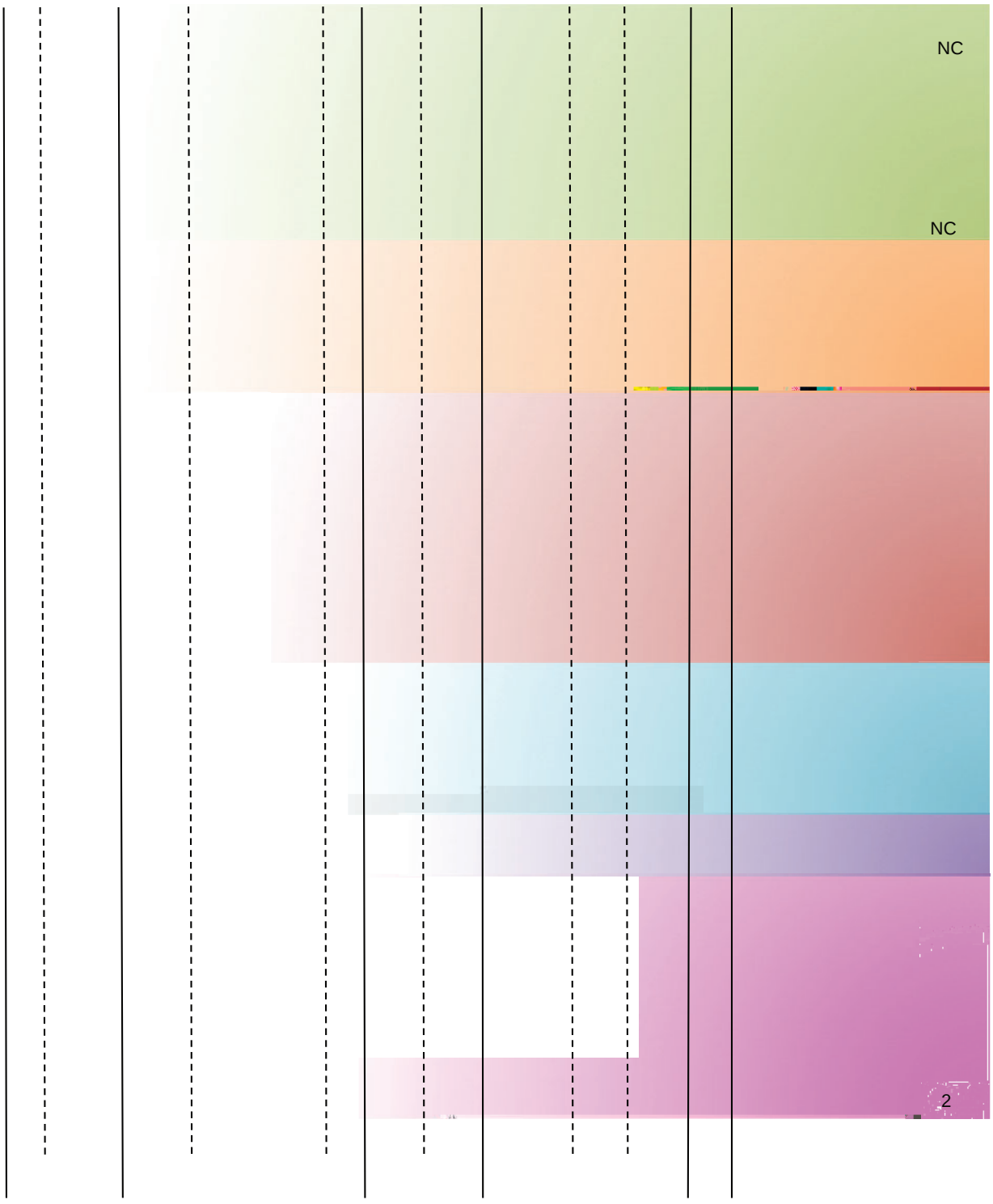
Species	Common Name	Collection Number	Origin	Collection Date	Sample Type	Total Reads	Mito Reads	MtDNA (%)	Mean Coverage Depth	Accession Number
Bradypus pygmaeus	Pygmy sloth	USNM 579179	Panama	1991	Internal organ	31,262,238	383,944	1.23	2,319X	KT818523
Bradypus torquatus	Maned sloth	BA449/11	Brazil	2011	Skin biopsy	39,749,058	61,728	0.16	375X	KT818524
Bradypus tridactylus	Pale-throated three-toed sloth	ISEM T-5013	French Guiana	2006	Ear biopsy	2,016,278	11,916	0.59	42X	KT818525
Bradypus variegatus	Brown-throated three-toed sloth	MVZ 155186	Peru	1978	Internal organ	726,173	30,043	4.14	121X	KT818526
B. variegatus	Brown-throated three-toed sloth									NC_006923
Choloepus didactylus	Southern two-toed sloth	MNHN 1998-1819	French Guiana	1997	Internal organ	670,499	7,289	1.09	26X	KT818537
Cho. didactylus	Southern two-toed sloth									NC_006924
Choloepus hoffmanni	Hoffmann's two-toed sloth	ISEM T-6052	Panama	2001	Kidney	1,258,442	3,174	0.25	11X	KT818538
Cyclopes didactylus	Pygmy anteater	MNHN 1998-234	French Guiana	1995	Internal organ	1,832,592	5,630	0.31	21X	KT818539
Myrmecophaga tridactyla	Giant anteater	ISEM T-2862	French Guiana	2001	Internal organ	1,102,261	7,215	0.65	26X	KT818549
Tamandua mexicana	Northern tamandua	MVZ 192699	Mexico	1977	Internal organ	1,552,407	4,167	0.27	16X	KT818551
Tamandua tetradactyla	Southern tamandua	ISEM T-6054	French Guiana	2001	Ear biopsy	14,370,168	6,373	0.04	30X	KT818552
Ta. tetradactyla	Southern tamandua									NC_004032
Dasypus kappleri	Greater long-nosed armadillo	ISEM T-3365	French Guiana	2001	Internal organ	1,059,009	46,167	4.36	166X	KT818541
Dasypus hybridus	Southern long-nosed armadillo	ZVC M2010	Uruguay	1976	Tail bone	30,663,498	38,967	0.13	123X	KT818540
Dasypus septemcinctus										

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

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

The xenarthran mitogenomic tree shows a fair amount of branch length heterogeneity among groups, with fast evolving clades including anteaters and Dasypodinae, and slow evolving clades such as sloths and hairy armadillos (fig. 1). Lineage-specific variation in evolutionary rates in mammalian mitochondrial genomes has been previously characterized (Gissi et al. 2000). Such variation has been linked to differences in mutation rates that correlate well with longevity in mammals (Nabholz et al. 2008). As a result, the mammalian mitochondrial clock is particularly erratic (Nabholz et al. 2009) and substitution rate variations among lineages should be accounted for in dating analyses by using relaxed clock models (Thorne et al. 1998). The selection of the clock

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


NC
NC
2

± 3ode17.[1.6–12.6]

abasso]1 etr

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

Nodes	This study	Delsuc et al. (2012)	Delsuc et al. (2004)
1. Xenarthra[a]	67.7 ± 3.0 [60.4–71.6]	67.8 ± 3.4 [61.3–74.7]	64.7 ± 4.9 [55.3–74.6]
2. Pilosa[a] (anteaters + sloths)	58.4 ± 4.1 [48.6–64.7]	60.1 ± 3.6 [53.1–67.2]	55.2 ± 4.9 [45.8–65.2]
3. Folivora[a] (sloths)	29.9 ± 6.5 [16.5–39.6]	28.3 ± 3.4 [22.0–35.2]	20.8 ± 3.3 [15.0–27.8]
4. Megalonychidae (two-toed sloths)	9.2 ± 3.5 [3.5–16.7]	N.A.	N.A.
5. Bradypodidae (three-toed sloths)	19.0 ± 4.7 [9.6–27.0]	N.A.	N.A.
6. Bradypus pygmaeus/others	7.7 ± 2.4 [3.6–12.6]	N.A.	N.A.
7. Bradypus tridactylus/Bradypus variegatus	5.±	±	±

(an997.9(Me5(Myrmecophaga/Tamandu7.6(–13.3)6.1–5.(F1)-14(f)0.7(6)...
3.6 53.1-.24(–79.)6.5(60)-4036(6)6.4(0)-.3(.50]TJ/F51 1 Tf11.6221 0 TD0 Tc()Tj/F1 1 Tf.9373 0 TD-.0204 Tc[(3)6.7(.).1(3)-395([21.2(15)6.7(.041(9

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

Sloths (Pilosa; Folivora)

The six living species of sloths belong to two genera, with two-toed sloths (genus Choloepus) and three-toed sloths (genus Bradypus) having been placed in two distinct families (Megalonychidae and Bradypodidae, respectively) to reflect their numerous morphological differences and a probably diphyletic origin from two different fossil groups (Webb 1985). Their independent adaptation to the arboreal lifestyle also led to a number of anatomical convergences related to their peculiar suspensory locomotion (Nyakatura 2012). Our results confirm this deep dichotomy with a divergence date between the two genera around 30 Ma (fig. 2 and table 2), which appears more ancient than previously estimated with nuclear data (Delsuc et al. 2004). This difference might stem from our increased taxon sampling, because only a single representative species of each genus was previously considered. Their considerable molecular divergence nevertheless supports the classification of the two modern sloth genera into distinct families.

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

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

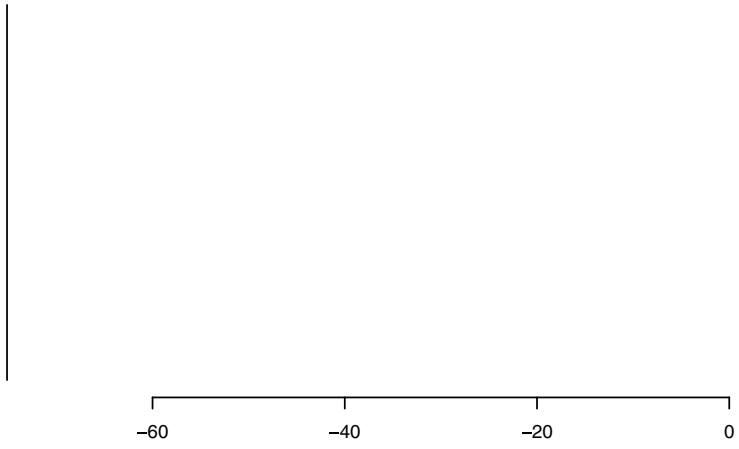
Dasypodinae. This subfamily includes seven species of long-nosed armadillos belonging to the single genus Dasypus. Species in this genus are characterized by an elongated nose that can be functionally related to the use of the tongue to gather ants, termites, and a diversity of soil invertebrates (Loughry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

results of Castro et al. (2015) who found

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

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

Our mitogenomic study is the first to include all four recognized species of the conspicuous and fossorial naked-tailed armadillos. The greater naked-tailed armadillo (Cabassous tatouay) is the first to diverge, followed by the Chacoan naked-tailed armadillo (Cabassous chacoensis) and the two closely related northern (Cabassous centralis) and southern (Cabassous unicinctus) naked-tailed armadillos (fig. 1). The emergence of


−60
−40
−20
0

extinction of the largest forms at the terminus of the Pleistocene (Lyons et al. 2004), a very recent event that is hardly detectable with current methods of macroevolutionary analyses. This effect is especially resonant for Pilosa in which the successful northern emigrants such as giant ground sloths were drawn to extinction (McDonald 2005).

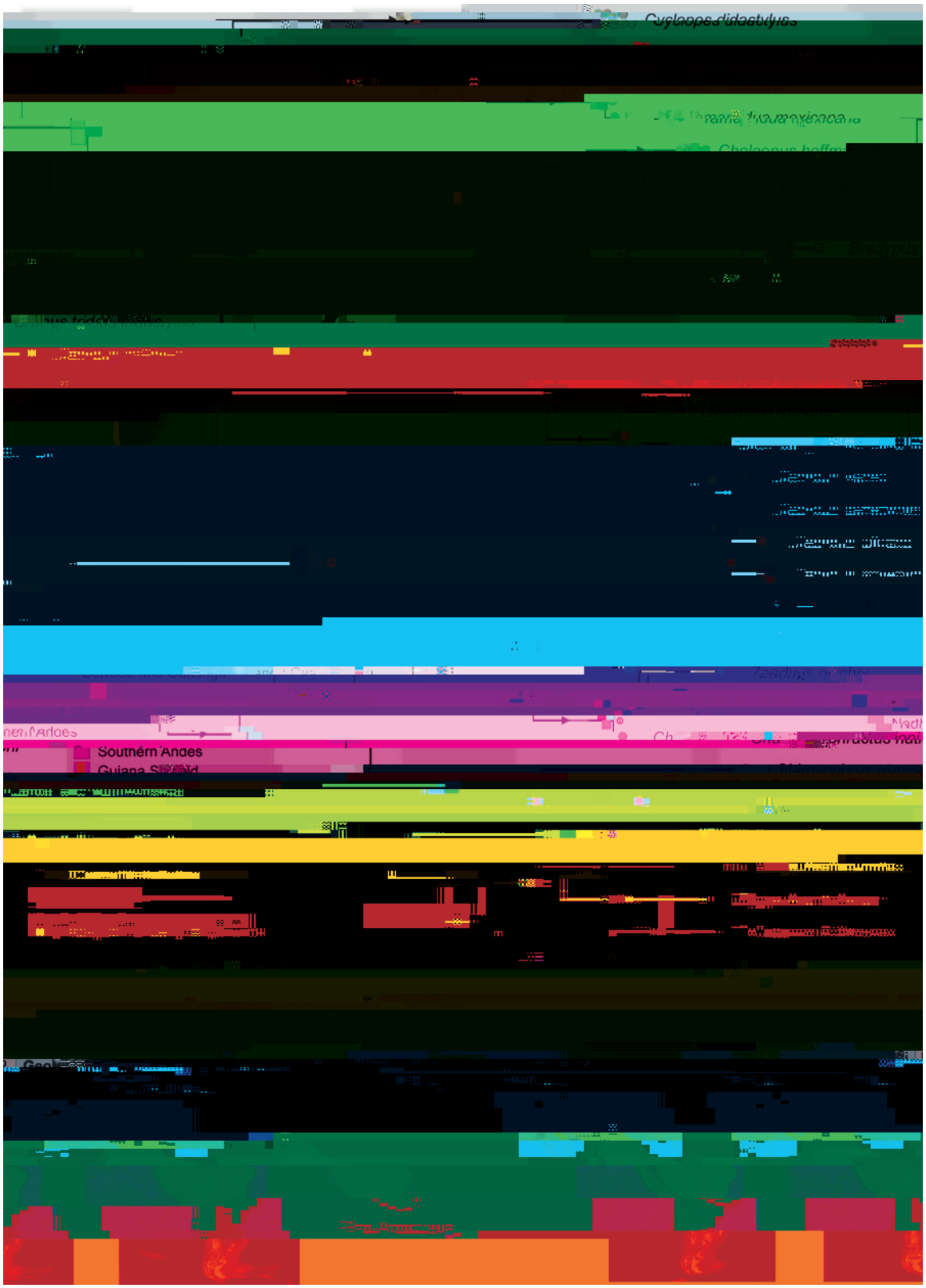


FIG. 5. Historical biogeography of living xenarthrans. The biogeographical range estimation was inferred under the Dispersal-Extinction-Cladogenesis model taking into account the change of connectivity and dispersal ability between areas defined as the main biomes of the American continent.

extend to Trish McLenachan and David Penny who sequenced and made freely available the first sloth mitogenomes, Lionel Hautier and Philippe Gaubert for taxonomical advice, and two anonymous referees for helpful comments. This work was supported by grants from the EU's Seventh Framework Programme (No 286431) to N.M.-B., the Centre National de la Recherche Scientifique (CNRS), the Scientific Council of Université Montpellier 2 (UM2), and Investissement d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01) to F.D., and the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs program to H.N.P. This is contribution ISEM 2015-227-S of the Institut des Sciences de l'Evolution de Montpellier.

References

Feijó A, Garbino GST, Campos BATP, Rocha PA, Ferrari SF, Langguth A. 2015. Distribution of Tolypeutes Illiger, 1811 (Xenarthra: Cingulata) with comments on its biogeography and conservation. Zool Sci. 32:77-87.

Feng X, Papes M. 2015. Ecological niche modelling confirms potential north-east range expansion of the nine-banded armadillo (Dasypus novemcinctus

flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths).