Genomic data suggest parallel dental vestigialization within the xenarthran radiation

Christopher A. Emerling*[1,2,3], Gillian C. Gibb[1,4], Marie-Ka Tilak[1], Jonathan J. Hughes[5], Melanie Kuch[5], Ana T. Duggan[5], Hendrik N. Poinar[5], Michael W. Nachman[2], and Frédéric Delsuc*[1]

[1] Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ. Montpellier, CNRS, IRD – Montpellier, France
[2] Museum of Vertebrate Zoology and Department of Integrative Biology, University of California – Berkeley, CA, USA
[3] Biology Department, Reedley College – Reedley, CA, USA
[4] School of Natural Sciences, Massey University – Palmerston North, New Zealand
[5] McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University – Hamilton, ON, Canada

*Corresponding authors
Correspondence: christopher.emerling@reedleycollege.edu, Frederic.Delsuc@umontpellier.fr

ABSTRACT

The recent influx of genomic data has provided greater insights into the molecular basis for regressive evolution, or vestigialization, through gene loss and pseudogenization. As such, the analysis of gene degradation patterns has the potential to provide insights into the evolutionary history of regressed anatomical traits. We specifically applied these principles to the xenarthran radiation (anteaters, sloths, armadillos), which is characterized by taxa with a gradation in regressed dental phenotypes. Whether the pattern among extant xenarthrans is due to an ancient and gradual decay of dental morphology or occurred repeatedly in parallel is unknown. We tested these competing hypotheses by examining 11 core dental genes in most living species of Xenarthra, characterizing shared inactivating mutations and patterns of relaxed selection during their radiation. Here we report evidence of independent and distinct events of dental gene loss in the major xenarthran subclades. First, we found strong evidence of complete enamel loss in the common ancestor of sloths and anteaters, suggested by the inactivation of four enamel-associated genes (*AMELX*, *AMTN*, *MMP20*, *ENAM*). Next, whereas sloths halted their dental regression, presumably a critical event that ultimately permitted adaptation to an herbivorous lifestyle, anteaters continued losing genes on the path towards complete tooth loss. Echoes of this event are recorded in the genomes of all living anteaters, being marked by a 2-bp deletion in a gene critical for dentinogenesis (*DSPP*) and a putative shared 1-bp insertion in a gene linked to tooth retention (*ODAPH*). By contrast, in the two major armadillo clades, genes pertaining to the gingival-tooth junction appear to have been independently inactivated prior to losing all or most enamel. These genomic data provide evidence for a complex evolutionary history with multiple pathways and rates of anatomical regression, and underscore the utility of using pseudogenes to reconstruct evolutionary history when fossils are sparse.

Keywords: Armadillos, Anteaters, Sloths, Dental regression, Gene loss, Molecular evolution, Phylogenetics

Introduction

Regressive evolution involves the vestigialization or loss of formerly adaptive traits over time, with examples ranging from the reduction of wings in flightless birds, the degeneration of limbs in various squamates and aquatic mammals, and the degradation of eyes in species occupying extreme dim-light niches. The reduction of such traits is often thought to result from a lack of adaptive utility after entering a novel niche, resulting in relaxed selection and regression through drift, and/or direct selection against the trait to minimize energetic costs (Fong et al., 1995; Jeffery, 2009; Lahti et al., 2009). The increasing availability of genomic data has permitted the documentation of parallel, and possibly causal, mutational events in genes associated with these and other vestigial phenotypes, providing a greater understanding of the genetic basis of regressive evolution (Albalat and Cañestro, 2016; Burga et al., 2017; Emerling, 2017; Emerling and Springer, 2014; Leal and Cohn, 2016; Sharma et al., 2018).

For instance, while the teeth of gnathostomes (jawed vertebrates) are considered a key adaptive trait in the origin of and radiation of this clade, dentition has subsequently regressed in numerous lineages (Charles et al., 2013; Davit-Béal et al., 2009), such as baleen whales, turtles, birds and pangolins. Notably, there are well-documented corresponding genomic signals behind these numerous instances of dental degeneration. Genetic association studies in humans with dental diseases and mouse knockout models have led to a robust understanding of the genetics underlying tooth development (Meredith et al., 2014; Smith et al., 2017), and comparative genomic analyses of edentulous (toothless) and enamelless vertebrates have revealed that many of these same dental genes were deleted or have eroded into unitary pseudogenes. Indeed, the list of documented dental pseudogenes in such vertebrates includes those encoding (1) enamel matrix proteins (EMPs), which provide a protein scaffold for the seeding of hydroxyapatite crystals during enamel development (ENAM [enamelin], AMELX [amelogenin], AMBN [ameloblastin]) (Choo et al., 2016; Delsuc et al., 2015; Meredith et al., 2014, 2013, 2009; Sire et al., 2008), (2) a metalloproteinase that processes these matrix proteins into their mature forms (MMP20 [enamelysin]) (Meredith et al., 2011a, 2014), (3) other proteins expressed in both enamel-forming ameloblasts and enamel-contacting gingiva (AMTN [amelotin], ODAM [odontogenic ameloblast-associated]) (Gasse et al., 2012; Meredith et al., 2014; Springer et al., 2019), (4) proteins of unknown function but showing clear associations with enamel formation (ACPT [testicular acid phosphatase], ODAPH [odontogenesis-associated phosphoprotein; formerly called *C4orf26*]) (Mu et al. 2021; Sharma et al., 2018; Springer et al., 2016), and (5) a protein that contributes to the dentin matrix (DSPP [dentin sialophosphoprotein]) (Meredith et al., 2014; Sire et al., 2008; McKnight and Fisher 2009).

Given the parallel signals between genes and anatomy during the evolution of degenerated traits, genomic data have the potential to provide insights into the timing, dynamics and consequences of regressive evolution, particularly in cases where the fossil record is limited. Xenarthra, which includes armadillos (Cingulata), sloths (Folivora) and anteaters (Vermilingua), represents a compelling example of the process of regressive evolution in regard to dentition. In contrast to the completely edentulous clades of baleen whales, birds, turtles and pangolins, xenarthrans span a spectrum of regressed dental phenotypes across a continual 68 million year radiation (Charles et al., 2013; Ciancio et al., 2014; Davit-Béal et al., 2009; Gibb et al., 2016; Vizcaíno, 2009). For instance, anteaters are united with sloths in the clade Pilosa, but while anteaters entirely lack teeth, sloths have intermediately-regressed teeth, possessing an edentulous premaxilla and simple, peg-like, single-rooted, enamelless dentition. Indeed their teeth have deviated so far from the ancestral tribosphenic form, that only recent developmental research has provided evidence of dental homologies to other placental mammals (Hautier et al., 2016). Sister to Pilosa are the Cingulata, which are divided into the extant families Dasypodidae and Chlamyphoridae. Most extant armadillos have an edentulous premaxilla, as well as peg-like, single-rooted teeth, which lack enamel in adult animals (Ferigolo, 1985). A notable exception is Dasypodidae, which possess vestigial enamel on deciduous teeth and prismatic or prismless enamel on permanent teeth of juveniles, which wears away with use (Ciancio et al., 2021; Martin, 1916; Spurgin, 1904).

The spectrum of degenerated dental phenotypes across Xenarthra raises a number of distinct questions. The first set concerns the specific history of the xenarthran clade. Anteaters are myrmecophagous mammals, almost exclusively consuming copious amounts of ants and termites by employing an extensive tongue with sticky saliva, and as such have little use for teeth. Sloths, by contrast, are herbivorous and often folivorous, and need to chew fibrous material with their seemingly ill-equipped, enamelless teeth. Among armadillos, some are relatively myrmecophagous (e.g., tolypeutines), while others are more omnivorous (e.g., euphractines). Whereas an extended history of myrmecophagy is universally associated with dental regression (Charles et al., 2013; Davit-Béal et al., 2009; Reiss, 2001), herbivory and omnivory are not. Accordingly, is the dental regression seen in xenarthrans the result of inheriting regressed teeth from their last common ancestor, which possibly had an insectivorous/myrmecophagous diet followed by subsequent dietary shifts to herbivory and omnivory? Or does it represent parallel, independent events within multiple lineages?

A second set of questions concerns the timing and patterns of regressive evolution. First, has the degeneration of teeth in xenarthrans taken place over a short period of time, consistent with selection against their presence, or has it been a gradual process over many millions of years in a manner more in line with relaxed selection and genetic drift? Finally, is there any sort of consistency in the sets of genes that are lost and the timing of those losses, or does the regression of teeth occur via divergent genomic patterns?

To answer these evolutionary questions, we collected genomic data to study patterns of pseudogenization and selection pressure in 11 core dental genes for most living species of Xenarthra. Our results point to xenarthran teeth having repeatedly regressed in parallel, showing distinct patterns of gene loss in different lineages in order to give rise to the variation in dentition observed across the clade today. They further suggest that regressive evolution can take place both gradually, over protracted periods of time, and in relatively rapid, discrete phases for the same trait during the radiation of a single clade.

Methods

To study patterns of dental gene loss in xenarthrans, we assembled a dataset of 11 dental genes: nine genes have well-characterized functions and/or expression patterns tied to tooth development and are frequently pseudogenized in edentulous and enamelless taxa (*ACPT*, *AMBN*, *AMELX*, *AMTN*, *DSPP*, *ENAM*, *MMP20*, *ODAM*, *ODAPH*; Choo et al., 2016; Delsuc et al., 2015; Gasse et al., 2012; McKnight and Fisher, 2009; Meredith et al., 2014, 2013, 2011a, 2009; Mu et al., 2021; Sharma et al., 2018; Sire et al., 2008; Smith et al., 2017; Springer et al., 2019, 2016), and two other genes (*DMP1*, *MEPE*) are expressed in teeth, more specifically in dentin, but do not have the same patterns of inactivation (Sun et al., 2011; Gullard et al., 2016). Our taxonomic coverage included 31 xenarthran species (four anteaters, six sloths, seven dasypodid armadillos, 14 chlamyphorid armadillos) plus 25 outgroup species spanning the remaining three superorders of placental mammals. We used a combination of strategies to reconstruct gene sequences: targeted sequencing of PCR amplified regions, exon-capture, whole-genome sequencing, and retrieval of sequences from publicly available genome assemblies (Supplementary Tables S1, S2).

Biological samples

Xenarthran tissue samples used for DNA extractions and amplifications of dental gene exons came from the Animal Tissue Collection of the Institut des Sciences de Montpellier (Table S1): nine-banded armadillo (*Dasypus novemcinctus* ISEM T-JL556), greater long-nosed armadillo (*Dasypus kappleri* ISEM T-2977), southern naked-tailed armadillo (*Cabassous unicinctus* ISEM T-2291), large hairy armadillo (*Chaetophractus villosus* ISEM NP390), giant armadillo (*Priodontes maximus* ISEM T-2353), southern three-banded armadillo (*Tolypeutes matacus* ISEM T-2348), pichi (*Zaedyus pichiy* ISEM T-6060), pink fairy armadillo (*Chlamyphorus truncatus* ISEM T-CT1), pale-throated three-fingered sloth (*Bradypus tridactylus* ISEM T-1476), brown-throated three-fingered sloth (*Bradypus variegatus* ISEM T-2999), southern two-fingered sloth (*Choloepus didactylus* ISEM T-1722), Hoffmann's two-fingered sloth (*Choloepus hoffmanni* ISEM T-6052), southern tamandua (*Tamandua tetradactyla* ISEM T-6054), giant anteater (*Myrmecophaga tridactyla* ISEM T-2862), and pygmy anteater (*Cyclopes didactylus* ISEM T-1631). Xenarthran specimens and corresponding Illumina genomic libraries used in exon capture experiments were those previously generated in Gibb et al. (2016). The greater fairy armadillo museum sample (*Calyptophractus retusus* ZSM T-Bret) and the pink fairy armadillo (*Chlamyphorus truncatus* ISEM T-CT1) used for whole genome shotgun sequencing were

146 respectively obtained from the Bavarian State Collection of Zoology (Munich, Germany) and provided by Dr.
147 Mariella Superina as previously detailed in Delsuc et al. (2012). The pygmy anteater (*Cyclopes didactylus* JAG
148 M2300) used for whole genome shotgun sequencing came from the JAGUARS animal tissue collection hosted
149 at the Institut Pasteur de la Guyane (Cayenne, French Guiana). Finally, the southern naked-tailed armadillo
150 (*Cabassous unicinctus* MVZ 155190) sample used for whole genome assembly was derived from a frozen
151 tissue sample from the Museum of Vertebrate Zoology (Berkeley, CA, USA).

152 **DNA extractions, PCR amplifications, and Sanger sequencing**

153 Total genomic DNA was extracted from tissue samples (Supplementary Table S1) preserved in 95%
154 ethanol using the QIAampDNA extraction kit (Qiagen). Polymerase chain reaction (PCR) was used to target
155 and amplify *AMBN*, *AMELX*, *DMP1*, *DSPP*, *ENAM*, and *MEPE* exons. To do so, a number of primers pairs were
156 designed from alignments of available sequences (Supplementary Table S3). PCR conditions were as follows:
157 95°C for 5 min, followed by 40 cycles at 95°C for 30 s, 50-55°C for 30 s, 72°C for 45 s, and a final extension
158 step at 72°C for 10 min. All PCR products were then purified from 1% agarose gels using Amicon Ultrafree-DA
159 columns (Millipore Corporation, Bedford, MA, USA) and sequenced on both strands using the polymerase
160 chain reaction primers with the Big Dye Terminator cycle sequencing kit on an Applied ABI Prism 3130XL
161 automated sequencer. Electropherograms were checked by eye and assembled into contigs using Geneious
162 Prime (Kearse et al., 2012).

163 **Target sequence capture and sequencing**

164 Baits for DNA sequence exon capture were designed using available xenarthran sequences for the
165 complete CDSs of all focal dental genes except *ACPT*. For each gene, 80mer baits were generated with a 4x
166 tiling density and were then BLASTed against the genome assemblies of *Choloepus hoffmanni*
167 (GCA_000164785.2) and *Dasypus novemcinctus* (GCA_000208655.2). Baits with more than one hit and a Tm
168 outside the range 35-40°C were excluded. This resulted in a final set of 5262 baits (Supplementary Dataset
169 S1) that were synthetized as part of a myBaits RNA kit by Arbor Biosciences (Ann Arbor, MI, USA). The
170 xenarthran Illumina libraries previously prepared (Gibb et al., 2016) were subsequently enriched with the
171 designed bait set in order to capture target sequences following previously described methodological
172 procedures (Delsuc et al., 2018). All enriched libraries were pooled together at varying concentrations with
173 the aim of producing one million reads for sequencing. Sequencing of the enrichment set was performed at
174 McMaster Genomics Facility (McMaster University, Hamilton, ON, Canada) on an Illumina MiSeq instrument
175 using 150 bp paired-end reads.

176 Index and adapter sequences were removed from raw reads and overlapping pairs merged with leeHom
177 (Renaud et al., 2014), and then mapped to all xenarthran reference sequences available with a modified
178 version of BWA (Li and Durbin, 2009; Stenzel, 2017). Mapped reads were additionally filtered to those that
179 were either merged or properly paired, had unique 5' and 3' mapping coordinates, and then restricted to
180 reads of at least 24 bp with SAMtools (Li et al., 2009). The bam files were then imported into Geneious to
181 select the best assembly for each sequence depending on the most successful reference sequence. All
182 sequences have been deposited in GenBank at the National Center for Biotechnology Information (NCBI)
183 under accession numbers OP966064 to OP966335.

184 **Whole genome shotgun libraries construction and sequencing**

185 For the greater fairy armadillo museum specimen (*Calyptophractus retusus* ZMS T-Bret), we constructed
186 new whole genome DNA Illumina libraries using a previously extracted genomic DNA and library preparation
187 procedure (Gibb et al., 2016). These degraded DNA libraries were sequenced at the Vincent J. Coates
188 Genomics Sequencing Laboratory (University of California, Berkeley, CA, USA) on an Illumina HiSeq4000
189 instrument using 50 bp single reads. Whole genome DNA Illumina library preparation and sequencing of
190 *Chlamyphorus truncatus* (ISEM T-CT1) and *Cyclopes didactylus* (JAG M2300) were outsourced to Novogene
191 Europe (Cambridge UK). These libraries were sequenced on an Illumina NovaSeq instrument using 150 bp
192 paired-end reads. The resulting short reads were cleaned from sequencing indexes and adaptors and quality
193 filtered using Trimmomatic (Bolger et al., 2014) with default parameters. These were then mapped to their
194 closest available relative xenarthran reference sequence using the Geneious mapping algorithm with default
195 settings. Consensus sequences of mapped reads were called using the 50% majority rule for each targeted
196 gene.

197 **Whole genome sequencing and assemblies**

198 For the southern naked-tailed armadillo (*Cabassous unicinctus* MVZ 155190), a PCR-free library Illumina
199 library with 450 bp inserts was prepared according to the recommendations of the Broad Institute for
200 subsequent genome assembly with DISCOVAR de novo
201 (https://software.broadinstitute.org/software/discovar/blog/?page_id=375) following the strategy of the
202 Zoonomia project (Zoonomia consortium, 2020). The library was sequenced at the Vincent J. Coates
203 Genomics Sequencing Laboratory (University of California, Berkeley, CA, USA) on an Illumina HiSeq4000
204 instrument using 250 bp paired-end reads. A draft genome assembly was produced from these data using
205 the DISCOVAR de novo assembler
206 (https://github.com/broadinstitute/discovar_de_novo/releases/tag/v52488) using the following command:
207 DiscovarDeNovo READS=[ReadsFile] OUT_DIR=[OutputDirectory] NUM_THREADS=24 MAX_MEM_GB=700.

208 We also utilized multiple whole genome assemblies deposited in NCBI's Whole Genome Shotgun (WGS)
209 contig database (Supplementary Table S2). To obtain genes on NCBI, we BLASTed human reference
210 nucleotide sequences, composed of exons, introns and flanking sequence, against the WGS database using
211 discontiguous megablast. We then used the NCBI-derived xenarthran sequences to BLAST using
212 discontiguous megablast against Discovar de novo assemblies imported into Geneious. Sequences derived
213 from assemblies that contained strings of more than 10 unknown nucleotides (Ns) had these stretches of Ns
214 reduced to 10 for ease of alignment. Raw sequencing reads have been deposited in the Short Read Archive
215 (SRA) at the National Center for Biotechnology Information (NCBI) under Bioproject PRJNA907496. The
216 Discovar de novo draft genome assembly for the southern naked-tailed armadillo (Dastaset S24) is available
217 from the associated Zenodo repository (https://doi.org/10.5281/zenodo.7214824).

218 **Dataset assembly**

219 Xenarthran sequences were assembled, and occasionally combined, from the various methodologies
220 employed (Supplementary Dataset S1). Outgroup sequences were retrieved by BLASTing reference
221 sequences against GenBank using discontiguous megablast and obtaining NCBI's annotated gene models
222 (Supplementary Table S2). All orthologs of xenarthrans and outgroup sequences were aligned using MUSCLE
223 (Edgar, 2004) in Geneious, examined by eye and adjusted manually to correct for errors in the automated
224 alignments. After characterizing putative inactivating mutations (see below), we prepared codon alignments
225 suitable for subsequent selection pressure analyses based on dN/dS estimations (Supplementary Dataset S1)
226 by removing insertions, incomplete codons and premature stop codons from individual sequences, and any
227 other portions of sequences with dubious homology, including: exons 7–9 of *AMBN*, given probable exon
228 duplications within the human reference (Toyosawa et al., 2000); *DSPP*'s highly repetitive region at the 3'
229 end of exon 4 (McKnight and Fisher 2009; Fisher, 2011); and exon 2 of *ODAPH* in humans, which is a splice
230 variant that is not common to all placental mammals (Springer et al., 2016). The evolutionary history of each
231 gene was estimated by maximum likelihood phylogenetic reconstruction using RAxML v8.2.11 (Stamatakis,
232 2014) within Geneious (GTR+GAMMA model, Rapid hill-climbing) to detect aberrant sequences (e.g., possible
233 contaminants and annotation errors stemming from poor gene models).

234 **Inactivating mutations and dN/dS ratio analyses**

235 We classified dental genes as (1) enamel-associated (*AMELX*, *AMBN*, *AMTN*, *ENAM*, *MMP20*, *ACPT*,
236 *ODAM*) based on their clear impact on enamel formation (Meredith et al. 2014; Smith et al., 2017), (2)
237 dentin-associated (*DSPP*) due a clear role in dentinogenesis (Sreenath et al., 2003), (3) tooth-associated
238 (*ODAPH*) given an apparent critical link to both enamel and dentin-formation (Parry et al., 2012; Springer at
239 al., 2016; Prasad et al., 2016), and (4) mineralization-associated (*DMP1*, *MEPE*) based on their expression in
240 various mineralized tissues, including teeth (Sun et al., 2011; Gullard et al., 2016). We searched for putative
241 inactivating mutations, including frameshift insertions or deletions, premature stop codons, start codon
242 mutations, and splice site mutations, and noted any mutations shared by multiple species within a clade. We
243 then performed several analyses to estimate the pattern and timing of shifts in selection pressure in tooth
244 genes among our focal xenarthran taxa.

245 Based on the patterns of inactivating mutations, we performed branch model dN/dS ratio (ω) analyses
246 (Yang 1998; Yang and Nielsen, 1998) using *codeml* in PAML v4.8 (Yang, 2007) to estimate the selection
247 pressure experienced by the different tooth genes throughout xenarthran history. We used the same
248 composite placental mammal topology for all analyses (Emerling et al., 2015; Gibb et al., 2016), and

249 employed the following protocol for each gene. First, we ran *codeml* with a one-ratio model, successively
250 employing codon frequency models 0, 1 and 2, then using the Akaike information criterion to determine the
251 best-fitting model (Supplementary Table S4). Next, using the branch model approach, we allowed ω to vary
252 across the phylogeny in a multi-ratio model with branch labels set for (a) each set of branches within a clade
253 that post-date the minimum gene inactivation date inferred from shared inactivation mutations, (b) each
254 branch that coincides with inferred gene inactivation, and (c) certain branches that predate gene
255 inactivation, often grouped together with multiple such branches. Note that this did not correspond to a free
256 ratio model, but rather follows a previously described approach (Meredith et al., 2009). In every case, when
257 taxon representation was sufficient, we set separate branch categories for stem Xenarthra, stem Pilosa, stem
258 Cingulata, stem Chlamyphoridae, and stem Dasypodidae. All non-xenarthran branches were grouped
259 together in a single ω category. Given that all selected outgroup taxa have teeth with enamel, this
260 background dN/dS ratio was assumed to be the average baseline ω estimate for functional dental genes.
261 Finally, subsequent models were run in which focal xenarthran branches were fixed as part of the
262 background ratio and then 1, respectively. For example, the stem xenarthran branch was always freely
263 estimated in the initial multi-ratio model, then a separate model fixed the stem xenarthran branch as part of
264 the background ratio, and then another model fixed the stem xenarthran branch at 1. This was carried out
265 for all non-background branches. Models were then compared statistically using likelihood-ratio tests (Yang,
266 2007).
267 Then, we reconstructed the evolution of the dN/dS ratio for each gene using the Bayesian approach
268 implemented in Coevol 1.4b (Lartillot and Poujol, 2011). Coevol provides a visual representation of the
269 variation in dN/dS ratio estimates across a phylogeny (e.g. Lartillot and Delsuc, 2012; Springer et al., 2019).
270 We used the *dsom* procedure to jointly estimate branch specific dN/dS ratios, divergence times, body sizes,
271 generation times, and ages at sexual maturity modeled as a multivariate Brownian diffusion process across
272 the phylogeny. We used the same topology than in the *codeml* analyses, assumed fossil calibrations from
273 previous studies (Meredith et al., 2011b; Emerling et al, 2015; Foley et al., 2016), and extracted the three life
274 history traits (body size, generation time, and age at sexual maturity) from the PanTHERIA database (Jones et
275 al., 2009). We set the prior on the root node to 97 Ma with a standard deviation of 20 Ma following the
276 molecular dating estimates of Meredith et al. (2011b). For each dataset, we ran two independent MCMC for
277 a total of 1000 cycles sampling parameters every cycle. MCMC convergence was checked by monitoring the
278 effective sample size of the different parameters using the *tracecomp* command of CoEvol. The first 100
279 points of each MCMC were excluded as burnin and posterior inferences were made from the remaining 900
280 sampled points of each chain.
281

282 ## Results

283 **Widespread pseudogenisation of dental genes in xenarthrans**
284 Nine out of the 11 examined tooth genes are inactivated in most xenarthrans, though there are different
285 patterns distinct to certain clades (Figures 1-2; Supplementary Tables S5-S15). The toothless anteaters have
286 the greatest proportion of pseudogenized genes (82%), followed by the enamelless sloths (55%-64%) and
287 chlamyphorid armadillos (45-55%), and the weakly-enamelled dasypodid armadillos (27-45%).
288 Genes encoding the enamel matrix proteins (EMPs) are commonly inactivated in xenarthrans. Among all
289 pilosans (anteaters and sloths), *ENAM*, *AMBN*, and *AMELX* are pseudogenes (Figure 2). Among
290 chlamyphorids, *ENAM* and *AMELX* are inactivated in all species, whereas *AMBN* varies, appearing intact in
291 some euphractines. The sequences for the latter remain incomplete, however, raising the possibility of
292 undetected inactivating mutations. Within dasypodids, EMPs are largely intact: *ENAM* is variable in its
293 inactivation, *AMELX* is never a pseudogene and *AMBN* inactivation is only suggested by a start codon
294 mutation in the nine-banded armdillo (*Dasypus novemcinctus*).
295 The metalloproteinase-encoding *MMP20* gene is clearly pseudogenized in all pilosan species (Figure
296 2). Among armadillos, the data suggest a very different trend: only two species provide evidence of
297 inactivation, with the hairy long-nosed armadillo (*Dasypus pilosus*) having a single stop codon in exon 2, and
298 the northern long-nosed armadillo (*D. sabanicola*) possessing a polymorphic stop codon in exon 8. Missing
299 data is an issue for some exon capture sequences, but complete sequences are provided for species with

genome assemblies, suggesting that *MMP20* is indeed intact in many if not the vast majority of armadillos, a conclusion further supported by selection pressure analyses based on dN/dS ratio (see below).

The genes encoding the ameloblast and gingiva-expressed amelotin (*AMTN*) and odontogenic ameloblast-associated protein (*ODAM*) are almost universally inactivated in xenarthrans (Figure 2). The data are patchier for the exon-capture sequences, but the genome assembly-derived sequences are all inactivated. One prominent exception is *ODAM* in the three-fingered sloths (*Bradypus* spp.), all of which fail to show evidence of inactivation. Although the three-fingered sloth *ODAM* sequences are incomplete, dN/dS ratio analyses (see below) suggest that they are likely intact.

Among the two genes encoding proteins with unknown roles in enamel formation (*ACPT*, *ODAPH*), only *ACPT* shows evidence of widespread pseudogenization, being unambiguously inactivated in every species examined except for the screaming hairy armadillo (*Chaetophractus vellerosus*), for which we were only able to recover two of the 11 exons. By contrast, *ODAPH* is only inactivated in anteaters (Figure 2).

Finally, among the genes encoding dentin-associated proteins (*DSPP*, *DMP1*, *MEPE*), only *DSPP* shows evidence of inactivation, and, again, solely among the anteaters (Figure 2). *DMP1* and *MEPE* have important roles in bone formation (Gullard et al., 2016; Sun et al., 2011), which may explain their universal retention.

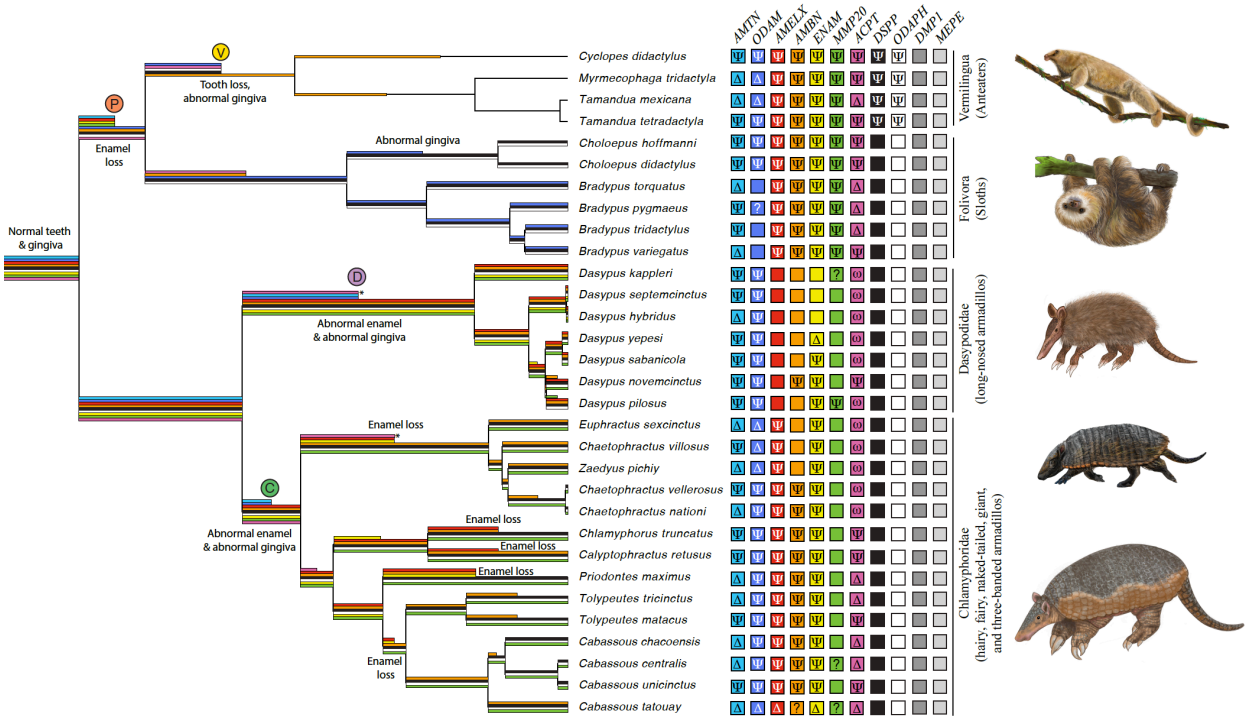


Figure 1: Dental gene inactivations across Xenarthra indicating distribution of gene losses (right) and inferred regressive dental events based on shared inactivating mutations (SIMs) of key genes (left). Abnormal enamel = evidence of inactivation in one or more of the following genes (*AMBN*, *ENAM*, *MMP20*, *ACPT*, *AMTN*); abnormal gingiva = evidence of inactivation in both *AMTN* and *ODAM*; enamelless teeth = inactivation of *AMELX*; edentulous = inactivation of *ODAPH* and possibly *DSPP*. Branches on which gene inactivations occurred are inferred from SIMs, which are summarized in the Results, Figure 2 and Supplementary Tables S5–S15. Gene losses in right columns are indicated by the following: Ψ = direct evidence of pseudogenization; Δ = gene inactivated or deleted, as inferred from phylogenetic bracketing; ω = possible gene inactivation inferred by dN/dS estimates; ? = unknown; empty box = gene intact. Timing of gene inactivations arbitrarily placed midway on the corresponding branch, indicated by termination of colored branch. Colors on branches correspond to genes using the same colors on the right (e.g., *AMELX* = red, *AMBN* = orange, etc.). Colored circles with letters P, V, C and D correspond with letters in Figure 2. * = possible timing of *ACPT* inactivation based on dN/dS estimates, but unclear due to absence of data from sufficient taxa. Paintings by Michelle S. Fabros.

Reconstructing patterns of shared inactivating mutations

The presence of shared inactivating mutations (SIMs) provides strong evidence for the minimum date of pseudogenization in a lineage. If, for example, a frameshift indel or premature stop codon occurs at the same position in all species within a clade, then it is more parsimonious to assume that it was inherited from a common ancestor rather than independently derived. Furthermore, multiple SIMs within a gene strengthen the case for a shared history.

We found no unambiguous inactivating mutations shared among all xenarthran subclades. The genes derived from the whole genome assemblies are sufficiently complete that this inference is unlikely to be a result of missing data (Supplementary Tables S5-S15). However, there are two possible caveats to this conclusion. First, exon 11 of *ACPT* is missing from the assemblies of all the species we examined, except the nine-banded armadillo (*Dasypus novemcinctus*). While *D. novemcinctus* has inactivating mutations in this exon, we cannot determine if they are unique to this species. Second, a 2-bp deletion occurs in exon 7 of the four armadillos and three sloths for which we have data (no data in anteaters), but the alignment for this indel is ambiguous (Supplementary Figure S1) and we believe this deletion is likely the result of independent events. Indeed, dN/dS analyses suggest that inactivation on the stem xenarthran branch is unlikely (see below).

By contrast, we found multiple examples of SIMs among pilosans (Figure 2, Supplementary Tables S7, S8, S11, S13). All ten pilosans we examined shared a premature stop codon in exon 4 of *AMELX*. Exon 4 of *AMTN* is followed by a splice donor mutation (GT → AT) shared by two-fingered sloths (*Choloepus* spp.) and the pygmy (*Cyclopes didactylus*) and giant (*Myrmecophaga tridactyla*) anteaters. Exon 8 of *ENAM* has three SIMs across pilosans, including a 1-bp deletion, premature stop codon and a roughly 520 bp insertion. Among the three sloths and two anteaters for which we assembled all of exon 6 of *MMP20*, they share a 1-bp deletion. Exon 3 of *AMBN* is preceded by a splice acceptor mutation (AG → AT) in sloths, but this exon is missing in anteaters.

Three genes appear to have been inactivated at minimum in a stem or crown sloth lineage. For instance, a three-fingered sloth (*Bradypus variegatus*) and the two-fingered sloths have at least 17 SIMs in *ACPT*, and sloths possess three SIMs in *AMBN*. By contrast, we found no disabling mutations in *ODAM* for three-fingered sloths, and two-fingered sloths only possess a single SIM (exon 8, 10-bp deletion).

Five genes appear to have been lost in the stem or a crown anteater lineage (Figure 1). Anteater SIMs include a shared deletion of both exons 2 and 3 in *ACPT*, a stop codon in exon 9 of *ODAM*, a 1-bp insertion in exon 3 of *ODAPH*, and a 2-bp deletion in exon 3 of *DSPP* (Figure 2). By contrast, *AMBN* only clearly possesses SIMs (four) between the giant (*Myrmecophaga tridactyla*) and lesser anteaters (*Tamandua* spp.).

We found no evidence of SIMs shared among all armadillos, but *AMTN* and *ODAM* both contain SIMs unique to each subclade (Figure 2). Among chlamyphorids, a chlamyphorine (*Calyptophractus retusus*), a euphractine (*Chaetophractus vellerosus*) and two tolypeutine armadillos (*Cabassous unicinctus*, *Tolypeutes matacus*) possess a splice donor mutation in intron 1 of *AMTN* (GT → GG), and the same species share a 15-bp deletion of the splice donor region of intron 2 in *ODAM*. Among dasypodids, a 1-bp deletion is found in exon 7 of *AMTN*, and *ODAM* has premature stop codons in exons 4 and 9. *ACPT* may well have been disabled prior to crown Chlamyphoridae, based on mutations shared by chlamyphorines and tolypeutines (1-bp del, exon 5; splice acceptor GT → GG, intron 5), but the only euphractine sequence (*Chaetophractus vellerosus*) is missing data from the relevant exons.

Among the other genes, the pattern of SIMs in armadillos is uneven. *AMBN*, for instance, only has shared mutations among congeners. By contrast, *AMELX* has SIMs in euphractines and tolypeutines, respectively, but none in chlamyphorines. Finally, *ENAM* possesses SIMs unique to all euphractines, chlamyphorines, and probably tolypeutines, respectively, and within dasypodids, *Dasypus novemcinctus* and *D. pilosus* share a splice donor mutation (AG → AT, intron 4) and *D. novemcinctus* and *D. sabanicola* share a 1-bp insertion (exon 8).

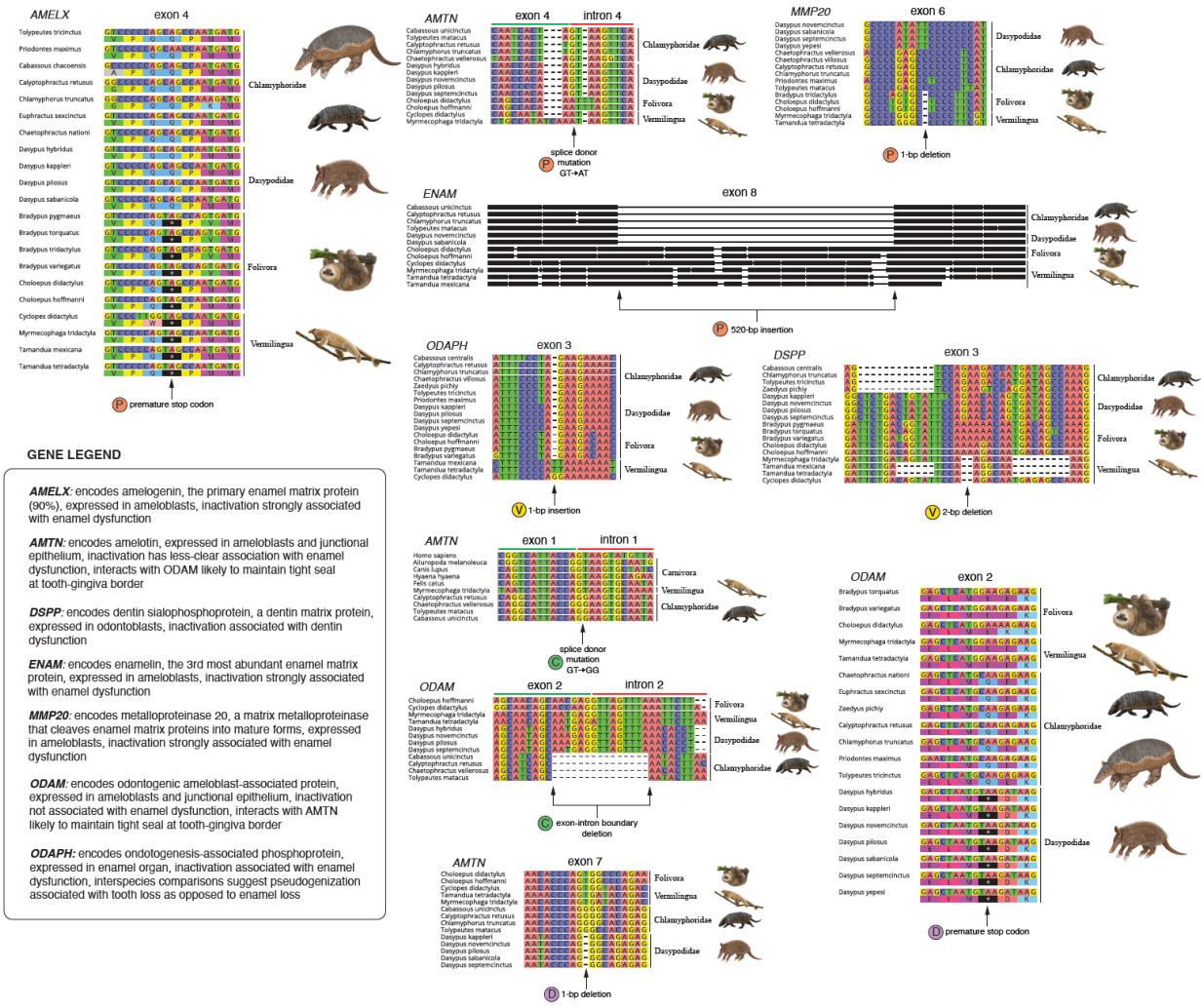


Figure 2: DNA and amino acid alignments provide examples of shared inactivating mutations indicating key dental gene losses during xenarthran evolutionary history. Colored circles with letters P, V, C and D correspond with letters in Figure 1. P = gene inactivation inferred on stem Pilosa branch; V = gene inactivation inferred on stem Vermilingua branch; C = gene inactivation inferred on stem Chlamyphoridae branch; D = gene inactivation inferred on stem Dasypodidae branch. Paintings by Michelle S. Fabros.

Selection pressure analyses

Given that shared inactivating mutations provide only a minimum probable date for inactivation, they may underestimate the timing of the onset of relaxed selection on a gene. Gene dysfunction may predate the fixation of a frameshift indel or premature stop codon, e.g., due to disruption of non-coding elements. We thus performed dN/dS branch model analyses to estimate the selective pressure experienced by the 11 dental genes on branches that predate the timing of inactivating mutations in xenarthrans (Supplementary Tables S16-S26; Supplementary Figures S2-S12). These branch model analyses (Yang, 1998; Yang and Nielsen, 1998) estimate how natural selection is acting on a gene by comparing the ratio (ω) of nonsynonymous substitutions (dN) to synonymous substitutions (dS) accumulated in a gene on a given branch or set of branches on a phylogeny. $\omega < 1$ suggests conservation of protein sequence (purifying selection) on average across the gene on that branch, $\omega > 1$ is consistent with change in protein function (positive selection), and $\omega = 1$ is associated with relaxed selection, which is the pattern expected for pseudogenes. However, on a phylogenetic branch that has a mixed history, such as purifying selection followed by relaxed selection, ω should be intermediate between the average strength of selection and one (Meredith et al., 2009). As such,

401 we estimated the average background ω, and tested whether key branches were statistically elevated or
402 lowered compared to this background selection pattern and an ω of 1.
403 We first tested whether ω was ever elevated on branches that pre-dated the earliest SIMs and found
404 several such examples. For instance, *AMELX* has SIMs for *Cabassous* + *Tolypeutes*, with the sister taxon
405 *Priodontes* only possessing unique mutations. However, the branch immediately ancestral to these
406 armadillos has an elevated ω, suggesting relaxed selection on *AMELX* began on this branch. Given such
407 examples, we tested whether elevated ω estimates could be found on the two xenarthran branches that
408 predate any SIMs: stem Xenarthra and stem Cingulata. For the stem xenarthran branch, we did not find
409 elevated ω among any genes in this lineage. By contrast, five of the 11 genes had signals consistent with
410 purifying selection (*AMELX* [ω = 0.0001], *DMP1* [ω = 0.37], *ENAM* [ω = 0.48], *MMP20* [ω = 0.2], *ODAPH* [ω =
411 0.14]), being statistically distinguishable from 1 but not from the background. *ODAPH* had an ω so low that it
412 was statistically distinguishable from both one and the background (ω = 0.48). For the stem Cingulata branch,
413 10 of the 11 genes showed no evidence of elevated ω, with five instead being consistent with purifying
414 selection (*ACPT* [ω = 0.1], *AMBN* [ω = 0.17], *DMP1* [ω = 0.34], *MMP20* [ω = 0.12], *ODAPH* [ω = 0.32]). *AMBN*'s
415 ω was so low that it was distinct from the background (ω = 0.41). The one example of an elevated ω was
416 *DSPP* (ω = 1.46), though we found no evidence of inactivation in this dentin-specific gene among the dentin-
417 retaining armadillos.
418 Additionally, in order to further evaluate changes in selection pressure resulting from dental regression,
419 we reconstructed the global variation in dN/dS across the placental tree (Figure 3; Supplementary Figures
420 S14-S24) using the program Coevol, which implements a form of rate smoothing through the incorporation
421 of a Brownian motion model of continuous trait evolution. These analyses allow for the visualization and
422 localization of shifts in selection pressure that have occurred within xenarthrans, likely corresponding to
423 relaxed selection and pseudogenization events in the 11 focal dental genes. The reconstructed patterns of
424 relaxed selection were strikingly similar among genes involved in the same function during tooth
425 development with enamel matrix-structure (*AMELX*, *AMBN*, *ENAM*), enamel junction-associated (*AMTN*,
426 *ODAM*), and enamel matrix-processing (*ACPT*, *MMP20*) genes presenting clear traces of relaxed selection in
427 specific xenarthran branches (Figure 3).

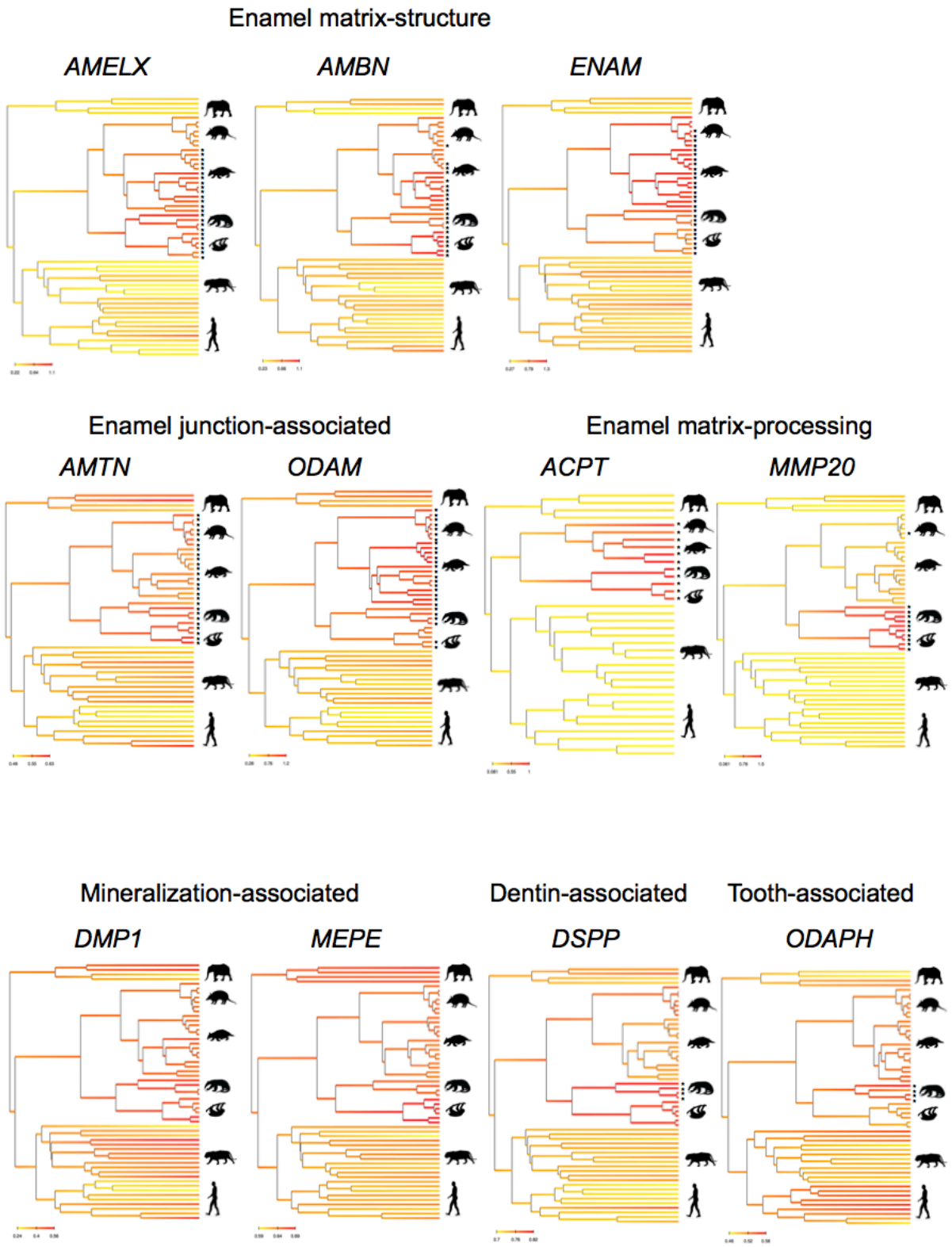


Figure 3: Bayesian reconstruction of dN/dS for 11 dental genes across the placental phylogeny with focus on xenarthrans (armadillos, anteaters, and sloths). The variation of dN/dS was jointly reconstructed with divergence times while controlling the effect of three life-history traits (body mass, longevity, and sexual maturity). Asterisks at the tips of terminal branches indicate non-functional sequences (pseudogenes). The tree is rooted with Afrotheria as the sister-group to all other placentals. Detailed trees are provided as Supplementary Figures S13-S23. Silhouettes were obtained from PhyloPic (http://phylopic.org/).

436 ## Discussion

437 **Reconstructing the dental regression history of xenarthrans**

438 Our results suggest that certain major regressive events in the dental history of xenarthrans took place
439 independently and in a gradual, stepwise fashion. Four lines of evidence suggest that this regression
440 occurred in parallel in the three main xenarthran lineages rather than in their last common ancestor. First,
441 among the nine genes that would ultimately become pseudogenes in at least some xenarthrans, no
442 unambiguous shared inactivating mutations (SIMs) were inferred for the entire xenarthran clade. The single
443 putative SIM, a 2-bp deletion in *ACPT*, has dubious homology between armadillos and sloths, and dN/dS ratio
444 analyses are not consistent with relaxed selection acting on the immediate descendant branches. Second, no
445 SIMs were found between the two major armadillo clades despite the last common ancestor of cingulates
446 dating to roughly 23 million years after the origin of Xenarthra (Gibb et al., 2016). Third, we found no
447 evidence of elevated dN/dS estimates on the stem xenarthran branch. By contrast, we found five genes
448 showing statistically significant purifying selection on this branch, three of which are enamel-specific genes
449 with widespread patterns of pseudogenization in vertebrates with dental regression (*AMELX*, *ENAM*,
450 *MMP20*) (Meredith et al., 2009, 2011a, 2013, 2014; Choo et al., 2016). Fourth, the same can be said for the
451 stem armadillo branch, on which five genes were also inferred to be under purifying selection, with only the
452 dentin-specific *DSPP* having an elevated ω, despite never being inactivated in any armadillos.

453 These four lines of evidence predict that the last common ancestor (LCA) of xenarthrans, the stem-
454 cingulate lineage, and the LCA of Cingulata all had teeth covered with enamel, with descendant lineages
455 subsequently deriving thin enamel (dasypodid armadillos), losing enamel (chlamyphorid armadillos, sloths)
456 or losing teeth entirely (anteaters). However, this does not negate the possibility that some degree of tooth
457 simplification had already occurred by the last common ancestor of xenarthrans and/or armadillos. Notably,
458 the two earliest armadillo fossils that preserve teeth, which are roughly coeval with molecular estimates of
459 the LCA of cingulates (Gibb et al., 2016), have simple, peg-like teeth with thin enamel, although they differ in
460 whether the enamel wore easily (*Utaetus buccatus* [42–39 Mya] or not (*Astegotherium dichotomus* [45
461 mya]). Notably, the former is a putative chlamyphorid relative, and the latter is considered a relative of
462 dasypodids (Ciancio et al., 2014; Simpson, 1932). Their enamel is reminiscent of phenotypes that can arise
463 from amelogenesis imperfecta, a condition that leads to the formation of soft or brittle enamel that wears
464 away with time, which can be caused by the inactivation of any one of numerous enamel-associated genes
465 (Smith et al., 2017). It remains possible that non-coding mutations leading to dental regression could have
466 accumulated prior to the mutations more typically characteristic of pseudogenes, at least by the origin of
467 cingulates.

468 After the origin of cingulates, our data suggest that stem chlamyphorid and stem dasypodid armadillos
469 independently inactivated *AMTN* and *ODAM* by at least 37 Mya and 12 Mya, respectively (Gibb et al., 2016).
470 The proteins expressed by these genes, AMTN and ODAM, are both expressed in enamel-producing
471 ameloblasts (Fouillen et al., 2017), but their link to enamel integrity is tenuous. Though the deletion of exons
472 3-6 in *AMTN* has been linked to amelogenesis imperfecta in humans (Smith et al., 2016), mouse knockouts
473 display a minimally-affected dental phenotype (Nakayama et al., 2015). Specifically, mandibular incisors have
474 chalky enamel, which begins to chip away after 13 weeks of age, but maxillary incisors and molars show no
475 significant alterations. *ODAM* knockout mice do not seem to have affected enamel whatsoever (Wazen et al.,
476 2015), and this gene has not been implicated in enamel malformations in humans (Smith et al., 2017). Both
477 proteins, however, continue to be expressed throughout adulthood in the junctional epithelium (Ganss and
478 Abbarin, 2014). These proteins appear to congregate into an extracellular matrix to maintain a tight seal
479 between the gingiva and teeth (Fouillen et al., 2017), presumably to protect the teeth from microbial
480 exposure (Lee et al., 2015). Notably, *ODAM* knockout mice show a decreased ability to heal the junctional
481 epithelium after damage (Wazen et al., 2015), lending credence to this hypothesis. Accordingly, the losses of
482 both *AMTN* and *ODAM* in stem chlamyphorids and stem dasypodids, respectively, point to a weakening of
483 the gingiva-tooth association and possibly minimal enamel degeneration. The latter is further suggested by
484 evidence of relaxed selection on *ENAM* on both stem branches.

485 After the origin of chlamyphorids (37 Mya), we inferred that further pseudogenizations of *ACPT*, *AMBN*,
486 *AMELX*, and *ENAM* took place, apparently in parallel (Figure 1). Disabling any of these four genes, all of which
487 are expressed in ameloblasts, is associated with AI (Lagerström et al., 1991; Poulter et al., 2014; Rajpar, 2001;
488 Seymen et al., 2016), leaving little doubt that enamel regression and eventually loss occurred during the

diversification of this clade. Curiously, *MMP20* appears universally intact in chlamyphorids but shows evidence of elevated dN/dS estimates, possibly indicating a change in function for this gene that is otherwise strongly linked to enamel development (Meredith et al., 2011a, 2013, 2014; Smith et al., 2017).

By contrast, the comparably late and minimal degree of pseudogenization in dasypodids may explain the retention of vestigial enamel on the milk teeth and thin, easily worn enamel on the permanent teeth of dasypodids (Ciancio et al., 2021; Martin, 1916; Spurgin, 1904). Only *ACPT* shows evidence of early pseudogenization in dasypodids, based on the large number of inactivating mutations and high ω estimate in *Dasypus novemcinctus*, and *ENAM* is the only other gene with clear evidence of pseudogenization in multiple dasypodids. This implies a delay in the historical timing of dental degeneration compared to chlamyphorids. The pattern of gene inactivation in dasypodids is notable in that *ENAM* is inactivated in several species and *Dasypus novemcinctus* has pseudogenic *ENAM*, *AMBN* and *ACPT*, yet thin regressive enamel is still present in these species (Ciancio et al., 2021; Martin, 1916; Spurgin, 1904). This suggests that inactivation of these genes individually or in concert is insufficient for complete enamel loss. Distinctly, *AMELX* appears intact in all dasypodids yet is pseudogenic in all pilosans and chlamyphorids sampled. This implies that *AMELX* may be a strong indicator for the timing of enamel loss in xenarthrans, a possibility made more plausible by its function (see below). If correct, enamel loss appears to have occurred up to five times independently within Chlamyphoridae based on SIMs (Figure 1).

In Pilosa, our results suggest that the first eight million years of their history (Gibb et al., 2016) resulted in a relatively rapid regression of their dentition. Specifically, we inferred that *AMTN*, *AMELX*, *ENAM*, and *MMP20* all became pseudogenes prior to the last common ancestor of pilosans (Figure 1). Enamel-forming ameloblasts express the enamel matrix proteins (EMPs) AMELX and ENAM, both of which are processed into their mature peptide forms by the metalloproteinase MMP20 (Smith et al., 2017). Building off an earlier study of *ENAM* (Meredith et al., 2009), this strongly suggests that enamel was completely lost by the earliest pilosans, particularly given that AMELX makes up to 90% of the EMP composition and MMP20 is critical for processing the EMPs AMELX, AMBN and ENAM (Smith et al., 2017). The earliest tooth-bearing fossil pilosan is a stem sloth (*Pseudoglyptodon*) dating to the early Oligocene (32 Mya) (McKenna et al., 2006), well after the estimate for the earliest pilosans (58 Mya) (Gibb et al., 2016). As predicted, *Pseudoglyptodon* lacks enamel, but our results predict the future discovery of enamelless stem and/or crown pilosans in the Paleocene.

Our results further suggest that stem vermilinguans inherited this enamelless condition, and then further continued dental regression to the point of complete tooth loss. In addition to all modern anteaters being edentulous, we found a mutation in *ODAPH* that suggests complete tooth loss occurred by the origin of Vermilingua (38 Mya) (Gibb et al., 2016). Although *ODAPH* is considered to be important in enamel formation, a comparative study of this gene in placental mammals found that it is uniquely inactivated in edentulous species (Springer et al., 2016). Perhaps more convincing is the evidence suggesting pseudogenization of the dentin-matrix protein *DSPP* in a stem vermilinguan (Figure 2). Inactivation of *DSPP* leads to dentinogenesis imperfecta in humans (Xiao et al., 2001) and dentin defects in knockout mice (Sreenath et al., 2003), increasing the likelihood that tooth loss occurred prior to the origin of this clade. The earliest known anteater, the edentulous *Protamandua* (Gaudin and Branham, 1998), dates back to the Santacrucian (17.5–16.3 Mya), but our results predict toothless stem vermilinguans likely dating to the Eocene, between 58 and 38 Mya (Gibb et al., 2016).

Broader implications for regressive evolution

Our results reveal a few noteworthy patterns that may inform the study of regressive evolution in other systems. First, these data provide genomic evidence that trait loss can take place in a stepwise manner. Fossil evidence implies that both turtles and birds lost their teeth gradually: turtles first lost marginal dentition, followed by palatal teeth (Li et al., 2008), and birds lost premaxillary teeth prior to becoming completely edentulous (Meredith et al., 2014). Xenarthran dental genes point to a third path towards this phenotype, with discrete events of enamel loss followed by tooth loss in the evolution of anteaters, based on shared and distinct pseudogenization signals in vermilinguans and sloths. This dental regression scenario was also recently inferred in a study on the evolution of baleen whales (Mysticeti) based on a similar comparison of dental genes (Randall et al., 2022).

Second, our results suggest that regressive evolution can vary broadly in timespan and pattern. Although we inferred that enamel loss occurred in stem pilosans within eight million years, dasypodid dentition, by contrast, appears to have taken a much lengthier period of regression. The earliest putative crown cingulate

fossils have simplified teeth with thin enamel and date back to 45 Mya (Ciancio et al., 2014). Assuming that these taxa represent the ancestral condition for crown armadillos, then the loss of *ODAM* and *AMTN* in a stem dasypodid, inactivation of *ENAM* in several crown dasypodids, and individual examples of *MMP20* and *AMBN* pseudogenization, imply a protracted, and perhaps episodic, pattern of regression in this lineage.

Finally, these data provide insights into how the regression of traits, such as the weakening or wholesale loss of enamel, may constrain the evolutionary trajectories lineages can take. The simplification and loss of teeth is a relatively common phenomenon in mammals, with most such species having diets characterized as myrmecophagous (ants and/or termites), vermivorous (soft-bodied worms) or nectarivorous (Charles et al., 2013; Davit-Béal et al., 2009; Freeman, 1995; Rosenberg and Richardson, 1995). Presumably the simplification or loss of teeth in such taxa is due to the softness of their prey (vermivores) or reliance on the tongue for food acquisition and minimized need for mastication (myrmecophagy, nectarivory). Notably, vermilinguan xenarthrans are myrmecophagous, and most armadillos at least partially, and at most extensively, consume social insects (Nowak, 1999), so possessing simplified teeth or being edentulous likely reinforces this diet. Reconstructions of chitinase genes in the earliest pilosans suggest that they were likely highly insectivorous (Emerling et al., 2018), indicating myrmecophagy is a plausible explanation for their early dental regression. However, after deriving enamelless teeth in stem pilosans, sloths predominantly became herbivores (Nowak, 1999; Saarinen and Karme, 2017). In contrast to the simplified, enamelless teeth present in sloths, other extant herbivorous mammals have enamel-capped teeth that tend towards an increased complexity of tooth cusps (Ungar, 2010). Yet, whereas anteaters continued the trend towards dental simplification with complete tooth loss, natural selection likely strongly favored the retention of teeth in sloths, as evidenced by the signal of purifying selection found in *ODAPH* on the stem sloth branch, with elevated dN/dS values on the dentin-associated genes *DMP1* and *MEPE* (Figure 3; Supplementary Figures S6, S9, S18, S21) potentially pointing to positive selection resulting in functional changes to sloth dentin.

In short, our results show how pseudogenes can provide insights into the deep evolutionary history of a clade, and point to the divergent paths that regressive evolution can take. Further study of other systems may provide additional clarity into how traits simplify and degenerate over time.

Acknowledgements

We would like to thank Anaïs Tibi and Delphine Sérol for their contribution to this study. We also thank Sérgio Ferreira-Cardoso and Lionel Hautier for helpful discussions. Jeremy Johnson (Broad Institute, Cambridge, MA, USA) kindly provided early access to xenarthran genome assemblies. This work would not have been possible without the help of the following individuals and institutions in accessing xenarthran tissue samples: Mariella Superina, Jim Loughry, Agustín Jiménez, Rodolfo Rearte, Guido Valverde, and Guillermo Pérez-Jimeno; Philippe Gaucher, Gérard Lievin, Erika Taube, Philippe Cerdan, Michel Blanc, Rodolphe Paowé, Jean-François Mauffrey, Eric Hansen, François Ouhoud-Renoux, and Jean-Christophe Vié (French Guiana); John Trupkiewich (Philadelphia Zoo, USA); 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); Gerhard Haszprunar and Michael Hiermeier (The Bavarian State Collection of Zoology, Munich, Germany); Benoit de Thoisy (Kwata NGO and Institut Pasteur de la Guyane, Cayenne, French Guiana); Géraldine Véron and Violaine Nicolas (Museum National d'Histoire Naturelle, Paris, France); and François Catzeflis (Institut des Sciences de l'Evolution, Montpellier, France). Sanger sequencing data were produced through technical facilities of the SeqGen platform of the Labex CeMEB (Centre Méditerranéen Environnement Biodiversité). Phylogenetic and statistical analyses benefited from the Montpellier Bioinformatics Biodiversity platform (MBB). This is contribution ISEM-2022-XXX of the Institut des Sciences de l'Evolution.

Data, scripts, code, and supplementary information availability

Data are available online: https://doi.org/10.5281/zenodo.7214824
Scripts and code are available online: https://doi.org/10.5281/zenodo.7214824
Supplementary information is available online: https://doi.org/10.5281/zenodo.7214824

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article. The authors declare the following non-financial conflict of interest: Frédéric Delsuc is a recommender of Peer Community In Evolutionary Biology.

Funding

This research was supported by a European Research Council consolidator grant (ConvergeAnt ERC-2015-CoG-683257; FD); the Centre National de la Recherche Scientifique (CNRS; FD); the Scientific Council of the Université de Montpellier (FD); Investissements d'Avenir grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; CEMEB: ANR-10-LABX-0004; FD); a National Science Foundation Postdoctoral Research Fellowship in Biology (award no. 1523943; CAE); a National Science Foundation Postdoctoral Fellow Research Opportunities in Europe award (CAE); the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France (CAE); the France-Berkeley Fund (FD and MWN); and the Natural Sciences and Engineering Research Council of Canada (NSERC, no. RGPIN04184-15) and the Canada Research Chairs program (HNP).

References

Albalat R, Cañestro C (2016) Evolution by gene loss. *Nature Reviews Genetics*, **17**, 379-391. https://doi.org/10.1038/nrg.2016.39

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Burga A, Wang W, Ben-David E, Wolf PC, Ramey AM, Verdugo C, Lyons K, Parker PG, Kruglyak L (2017) A genetic signature of the evolution of loss of flight in the Galapagos cormorant. *Science*, **356**, eaal3345. https://doi.org/10.1126/science.aal3345

Charles C, Solé F, Rodrigues HG, Viriot L (2013) Under pressure? Dental adaptations to termitophagy and vermivory among mammals. *Evolution*, **67**, 1792-1804. https://doi.org/10.1111/evo.12051

Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, Yurchenko AA, Kliver S, Tamazian G, Antunes A, Wilson RK, Warren WC, Koepfli KP, Minx P, Krasheninnikova K, Kotze A, Dalton DL, Vermaak E, Paterson IC, Dobrynin P, Sitam FT, Rovie-Ryan JJ, Johnson WE, Yusoff AM, Luo SJ, Karuppannan KV, Fang G, Zheng D, Gerstein MB, Lipovich L, O'Brien SJ, Wong GJ (2016) Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Research*, **26**, 1312-1322. https://doi.org/10.1101/gr.203521.115

Ciancio MR, Vieytes EC, Carlini AA (2014) When xenarthrans had enamel: Insights on the evolution of their hypsodonty and paleontological support for independent evolution in armadillos. *Naturwissenschaften*, **101**, 715-725. https://doi.org/10.1007/s00114-014-1208-9

Ciancio MR, Vieytes EC, Castro MC, Carlini AA (2021) Dental enamel structure in long-nosed armadillos (Xenarthra: Dasypus) and its evolutionary implications. *Zoological Journal of the Linnean Society*, **192**, 1237–1252. https://doi.org/10.1093/zoolinnean/zlaa119

Davit-Béal T, Tucker AS, Sire JY (2009) Loss of teeth and enamel in tetrapods: Fossil record, genetic data and morphological adaptations. *Journal of Anatomy*, **214**, 477-501. https://doi.org/10.1111%2Fj.1469-7580.2009.01060.x

Delsuc F, Superina M, Tilak M-K, Douzery EJP, Hassanin A (2012) Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Molecular Phylogenetics and Evolution*, **62**, 673-680. https://doi.org/10.1016/j.ympev.2011.11.008

Delsuc F, Gasse B, Sire JY (2015) Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. *BMC Evolutionary Biology*, **15**, 148. https://doi.org/10.1186/s12862-015-0431-0

Delsuc F, Kuch M, Gibb GC, Hughes J, Szpak P, Southon J, Enk J, Duggan AT, Poinar HN (2018) Resolving the phylogenetic position of Darwin's extinct ground sloth (*Mylodon darwinii*) using mitogenomic and nuclear exon data. *Proceedings of the Royal Society B*, **285**, 20180214. https://doi.org/10.1098/rspb.2018.0214

DISCOVAR de novo: Large genome assembler. https://www.broadinstitute.org/software/discovar/blog

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797. https://doi.org/10.1093/nar/gkh340

Emerling CA, Springer MS (2014) Eyes underground: Regression of visual protein networks in subterranean mammals. *Molecular Phylogenetics and Evolution*, **78**, 260-270. https://doi.org/10.1016/j.ympev.2014.05.016

Emerling CA, Huynh HT, Nguyen MA, Meredith RW, Springer MS (2015) Spectral shifts of mammalian ultraviolet-sensitive pigments (short wavelength-sensitive opsin 1) are associated with eye length and photic niche evolution. *Proceedings of the Royal Society B*, **282**, 20151817. https://doi.org/10.1098/rspb.2015.1817

Emerling CA (2017) Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. *Molecular Phylogenetics and Evolution*, **115**, 40-49. https://doi.org/10.1016/j.ympev.2017.07.014

Emerling CA, Delsuc F, Nachman MW (2018) Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Science Advances*, **4**, eaar6478. https://doi.org/10.1126/sciadv.aar6478

Ferigolo J (1985) Evolutionary trends of the histological pattern in the teeth of Edentata (Xenarthra). *Archives of Oral Biology*, **30**, 71-82. https://doi.org/10.1016/0003-9969(85)90027-5

Fisher LW (2011) DMP1 and DSPP: Evidence for duplication and convergent evolution of two SIBLING proteins. *Cells Tissues Organs*, **194**, 113-118. https://doi.org/10.1159/000324254

Foley NM, Springer MS, Teeling EC (2016) Mammal madness: is the mammal tree of life not yet resolved? Philosophical Transactions of the Royal Society B, **371**, 20150140. https://doi.org/10.1098/rstb.2015.0140

Fong D, Kane T, Culver D (1995) Vestigialization and loss of nonfunctional characters. *Annual Review of Ecology and Systematics*, **26**, 249-268. https://doi.org/10.1146/annurev.es.26.110195.001341

Fouillen A, Dos Santos Neves J, Mary C, Castonguay JD, Moffatt P, Baron C, Nanci A (2017) Interactions of AMTN, ODAM and SCPPPQ1 proteins of a specialized basal lamina that attaches epithelial cells to tooth mineral. *Scientific Reports*, **7**, 46683. https://doi.org/10.1038/srep46683

Freeman PW (1995) Nectarivorous feeding mechanisms in bats. *Biological Journal of the Linnean Society*, **56**, 439-463. https://doi.org/10.1111/j.1095-8312.1995.tb01104.x

Ganss B, Abbarin N (2014) Maturation and beyond: Proteins in the developmental continuum from enamel epithelium to junctional epithelium. *Frontiers in Physiology*, **5**, 371. https://doi.org/10.3389/fphys.2014.00371

Gasse B, Silvent JF, Sire JY (2012) Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI. *Journal of Dental Research*, **91**, 1085-1089. https://doi.org/10.1177/0022034512460551

Gaudin TJ, Branham DG (1998) The phylogeny of the Myrmecophagidae (Mammalia, Xenarthra, Vermilingua) and the relationship of *Eurotamandua* to the Vermilingua. *Journal of Mammalian Evolution*, **5**, 237-265. https://doi.org/10.1023/A:1020512529767

Gibb GC, Condamine FL, Kuch M, Enk J, Moraes-Barros N, Superina M, Poinar HN, Delsuc F (2016) Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. *Molecaular Biology and Evolution*, **33**, 621-642. https://doi.org/10.1093/molbev/msv250

Gullard A, Gluhak-Heinrich J, Papagerakis S, Sohn P, Unterbrink A, Chen S, MacDougall M (2016) MEPE localization in the craniofacial complex and function in tooth dentin formation. *Journal of Histochemistry and Cytochemistry*, **64**, 224-236. https://doi.org/10.1369/0022155416635569

687 Hautier L, Gomes Rodrigues H, Billet G, Asher RJ (2016) The hidden teeth of sloths: Evolutionary vestiges and
688 the development of a simplified dentition. *Scientific Reports*, **6**, 27763.
689 https://doi.org/10.1038/srep27763
690 Jeffery WR (2009) Regressive evolution in *Astyanax* cavefish. *Annual Review of Genetics*, **43**, 25-47.
691 https://doi.org/10.1146/annurev-genet-102108-134216
692 Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C, Connolly
693 C, Cutts MJ, Foster JK, Grenyer R, Habib M, Plaster CA, Price SA, Rigby EA, Rist J, Teacher A, Bininda-
694 Emonds ORP, Gittleman JL, Mace GM, Purvis A (2009) PanTHERIA: a species-level database of life history,
695 ecology, and geography of extant and recently extinct mammals. *Ecology*, **90**, 2648.
696 https://doi.org/10.1890/08-1494.1
697 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran
698 C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable
699 desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-
700 1649. https://doi.org/10.1093/bioinformatics/bts199
701 Lagerström M, Dahl N, Nakahori Y, Nakagome Y, Bäckman B, Landegren U, Pettersson U (1991) A deletion in
702 the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). *Genomics*, **10**, 971-975.
703 https://doi.org/10.1016/0888-7543(91)90187-J
704 Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA (2009)
705 Relaxed selection in the wild. *Trends in Ecology and Evolution*, **24**, 487-496.
706 https://doi.org/10.1016/j.tree.2009.03.010
707 Lartillot N, Poujol R (2011) A phylogenetic model for investigating correlated evolution of substitution rates
708 and continuous phenotypic characters. *Molecular Biology and Evolution*, **28**, 729-744.
709 https://doi.org/doi:10.1093/molbev/msq244
710 Lartillot N, Delsuc F (2012) Joint reconstruction of divergence times and life-history evolution in placental
711 mammals using a phylogenetic covariance model. *Evolution*, **66**, 1773-1787.
712 https://doi.org/10.1111/j.1558-5646.2011.01558.x
713 Leal F, Cohn MJ (2016) Loss and re-emergence of legs in snakes by modular evolution of Sonic hedgehog and
714 HOXD enhancers. *Current Biology*, **26**, 2966-2973. https://doi.org/10.1016/j.cub.2016.09.020
715 Lee HK, Ji S, Park SJ, Choung HW, Choi Y, Lee HJ, Park SY, Park JC (2015) Odontogenic ameloblast-associated
716 protein (ODAM) mediates junctional epithelium attachment to teeth via integrin-odam-rho guanine
717 nucleotide exchange factor 5 (ARHGEF5)-RhoA signaling. *Journal of Biological Chemistry*, **290**, 14740-
718 14753. https://doi.org/10.1074/jbc.M115.648022
719 Li C, Wu XC, Rieppel O, Wang LT, Zhao LJ (2008) An ancestral turtle from the Late Triassic of southwestern
720 China. *Nature*, **456**, 497-501. https://doi.org/10.1038/nature07533
721 Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
722 *Bioinformatics*, **25**, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324
723 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The
724 Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
725 https://doi.org/10.1093/bioinformatics/btp352
726 Martin BE (1916) Tooth development in *Dasypus novemcinctus*. *Journal of Morphology*, **27**, 647-691.
727 McKenna MC, Wyss AR, Flynn JJ (2006) Paleogene pseudoglyptodont xenarthrans from central Chile and
728 Argentine patagonia. *American Museum Novitates*, **3536**, 1-18. https://doi.org/10.1206/0003-
729 0082(2006)3536[1:PPXFCC]2.0.CO;2
730 McKnight DA, Fisher LW (2009) Molecular evolution of dentin phosphoprotein among toothed and toothless
731 animals. *BMC Evolutionary Biology*, **9**, 299. https://doi.org/10.1186/1471-2148-9-299
732 Meredith RW, Gatesy J, Murphy WJ, Ryder OA, Springer MS (2009) Molecular decay of the tooth gene
733 enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genetics*, **5**,
734 e1000634. https://doi.org/10.1371/journal.pgen.1000634
735 Meredith RW, Gatesy J, Cheng J, Springer MS (2011a) Pseudogenization of the tooth gene enamelysin
736 (MMP20) in the common ancestor of extant baleen whales. *Proceedings of the Royal Society B*, **278**, 993-
737 1002. https://doi.org/10.1098/rspb.2010.1280
738 Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T,
739 Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A,
740 Westerman M, Ayoub NA, Springer MS, Murphy WJ (2011b) Impacts of the Cretaceous Terrestrial

Revolution and KPg extinction on mammal diversification. *Science*, **334**, 521-524. https://doi.org/10.1126/science.1211028

Meredith RW, Gatesy J, Springer MS (2013) Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evolutionary Biology*, **13**, 20. https://doi.org/10.1186/1471-2148-13-20

Meredith RW, Zhang G, Gilbert MTP, Jarvis ED, Springer MS (2014) Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science*, **346**, 1254390. https://doi.org/10.1126/science.1254390

Mu Y, Huang X, Liu R, Gai Y, Liang N, Yin D, Shan L, Xu S, Yang G (2021) ACPT gene is inactivated in mammalian lineages that lack enamel or teeth. *PeerJ*, **9**, e10219. https://doi.org/10.7717/peerj.10219

Nakayama Y, Holcroft J, Ganss B (2015) Enamel hypomineralization and structural defects in amelotin-deficient mice. *Journal of Dental Research*, **94**, 697-705. https://doi.org/10.1177/0022034514566214

Nowak RM (1999) Walker's Mammals of the World, Vol. 1, 6th ed. John Hopkins University Press, Baltimore, Maryland.

Parry DA, Brookes SJ, Logan CV, Poulter JA, El-Sayed W, Al-Bahlani S, Al Harasi S, Sayed J, Raïf EF, Shore RC, Dashash M, Barron M, Morgan JE, Carr IM, Taylor GR, Johnson CA, Aldred MJ, Dixon MJ, Wright JT, Kirkham J, Inglehearn CF, Mighell AJ (2012) Mutations in C4orf26, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. *American Journal of Human Genetics*, **91**, 565-571. https://doi.org/10.1016/j.ajhg.2012.07.020

Prasad MK, Laouina S, El Alloussi M, Dollfus H, Bloch-Zupan A (2016) Amelogenesis Imperfecta: 1 Family, 2 Phenotypes, and 2 Mutated Genes. *Journal of Dental Research*, **95**, 1457-1463. https://doi.org/10.1177/0022034516663200

Poulter JA, Murillo G, Brookes SJ, Smith CEL, Parry DA, Silva S, Kirkham J, Inglehearn CF, Mighell AJ (2014) Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta. *Human Molecular Genetics*, **23**, 5317-5324. https://doi.org/10.1093/hmg/ddu247

Rajpar MH (2001) Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. *Human Molecular Genetics*, **10**, 1673-1677. https://doi.org/10.1093/hmg/10.16.1673

Randall JG, Gatesy J, Springer MS (2022) Molecular evolutionary analyses of tooth genes support sequential loss of enamel and teeth in baleen whales (Mysticeti). *Molecular Phylogenetics and Evolution*, **171**, 107463. https://doi.org/10.1016/j.ympev.2022.107463

Reiss KZ (2001) Using phylogenies to study convergence: the case of the ant-eating mammals. *American Zoologist*, **41**, 507-525. https://doi.org/10.1093/icb/41.3.507

Renaud G, Stenzel U, Kelso J (2014). LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research*, **42**, e141. https://doi.org/10.1093/nar/gku699

Rosenberg HI, Richardson KC (1995) Cephalic morphology of the honey possum, *Tarsipes rostratus* (Marsupialia: Tarsipedidae); an obligate nectarivore. *Journal of Morphology*, **223**, 303-323. https://doi.org/10.1002/jmor.1052230307

Saarinen J, Karme A (2017) Tooth wear and diets of extant and fossil xenarthrans (Mammalia, Xenarthra) – Applying a new mesowear approach. *Palaeogeography Palaeoclimatology Palaeoecology*, **476**, 42-54. https://doi.org/10.1016/j.palaeo.2017.03.027

Seymen F, Kim YJ, Lee YJ, Kang J, Kim TH, Choi H, Koruyucu M, Kasimoglu Y, Tuna EB, Gencay K, Shin TJ, Hyun HK, Kim YJ, Lee SH, Lee ZH, Zhang H, Hu JCC, Simmer JP, Cho ES, Kim JW (2016) Recessive mutations in ACPT, Encoding Testicular Acid Phosphatase, cause hypoplastic amelogenesis imperfecta. *American Journal of Human Genetics*, **99**, 1199-1205. https://doi.org/10.1016/j.ajhg.2016.09.018

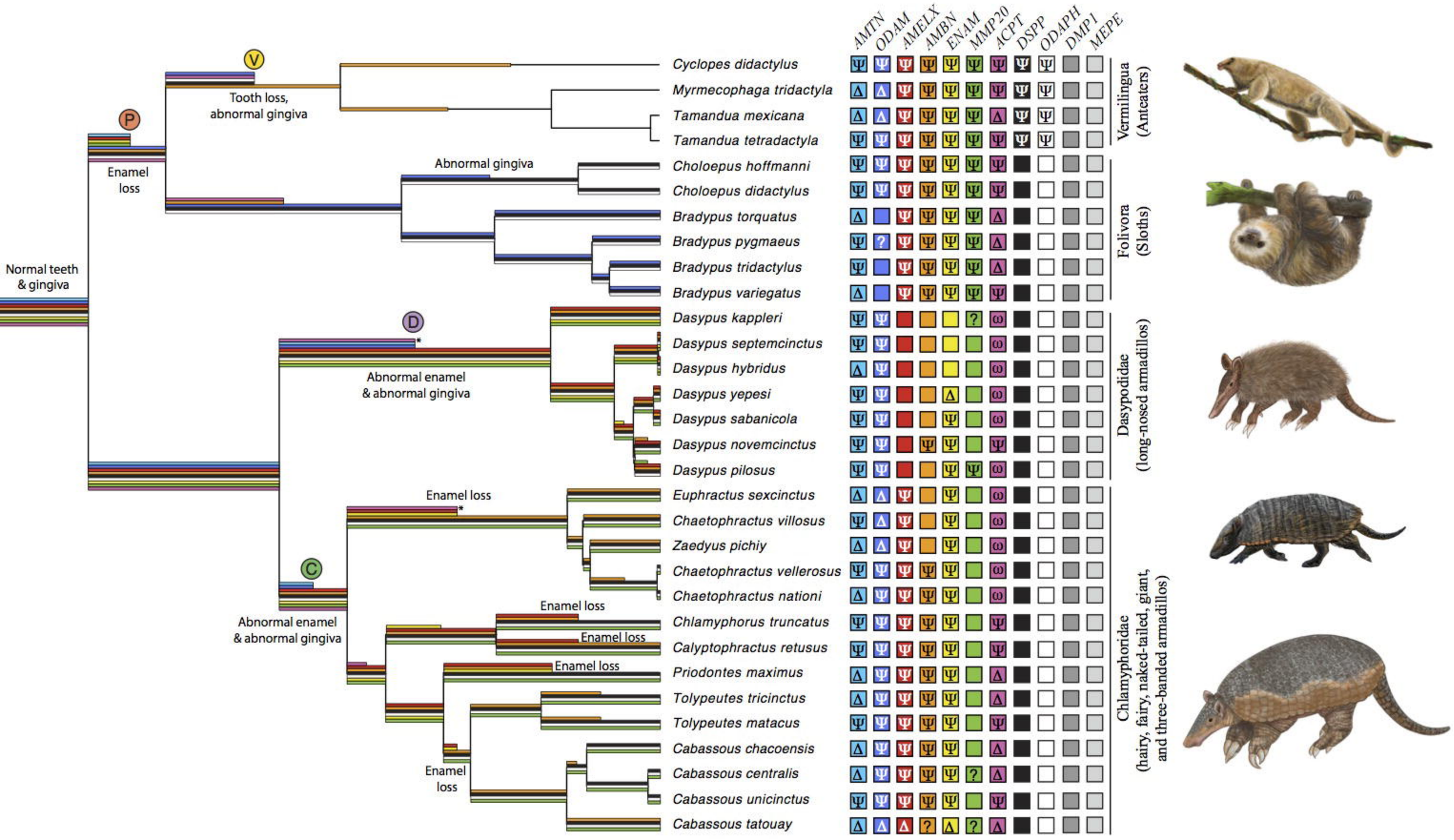
Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M (2018) A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications*, **9**, 1-9. https://doi.org/10.1038/s41467-018-03667-1

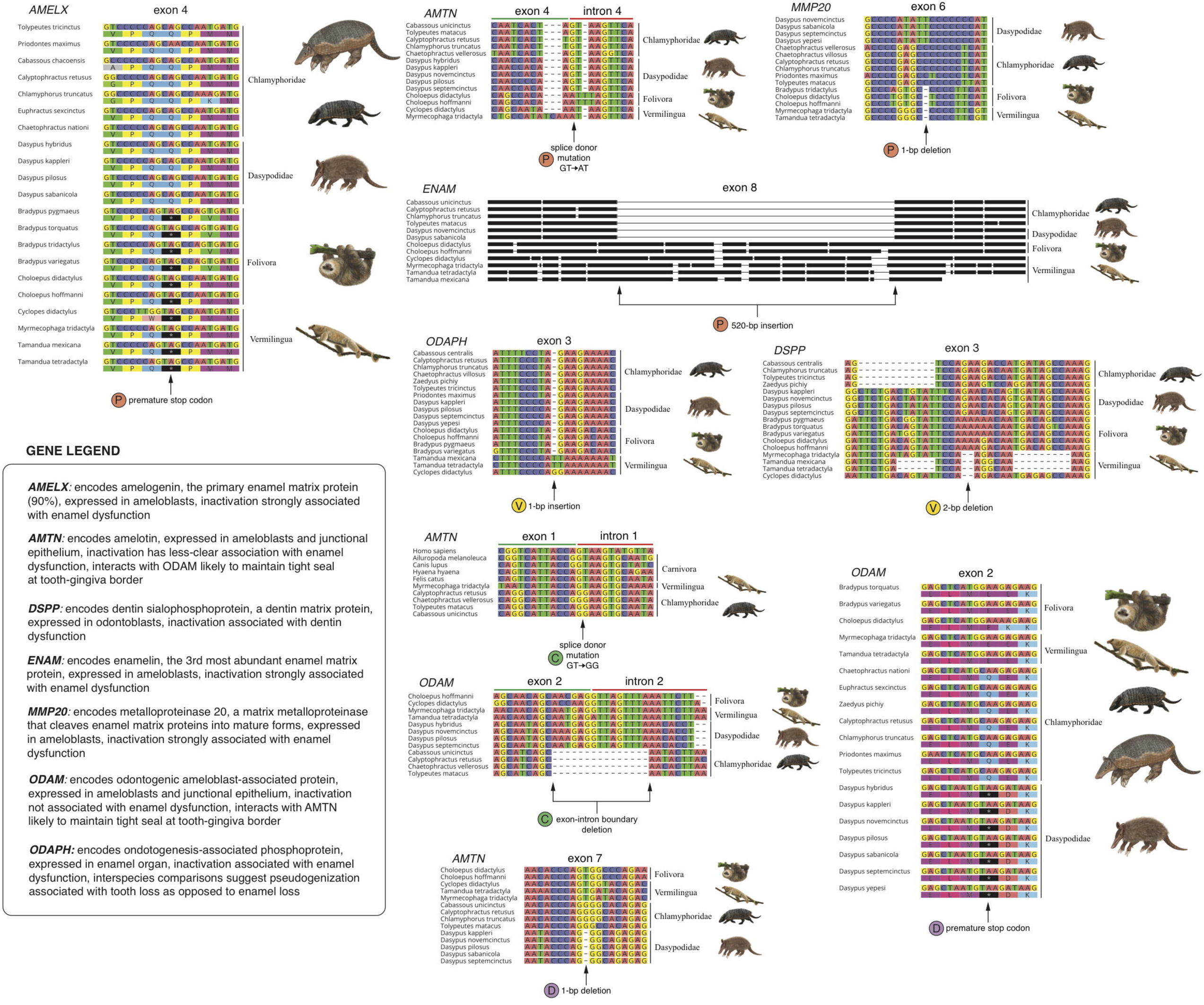
Simpson GG (1932) Enamel on the teeth of an Eocene edentate. *American Museum Novitates*, **4**, 1-4.

Sire JY, Delgado SC, Girondot M (2008) Hen's teeth with enamel cap: From dream to impossibility. *BMC Evolutionary Biology*, **8**, 246. https://doi.org/10.1186/1471-2148-8-246

Smith CEL, Murillo G, Brookes SJ, Poulter JA, Silva S, Kirkham J, Inglehearn CF, Mighell AJ (2016) Deletion of amelotin exons 3-6 is associated with amelogenesis imperfecta. *Human Molecular Genetics*, **25**, 3578-3587. https://doi.org/10.1093/hmg/ddw203

793 Smith CEL, Poulter JA, Antanaviciute A, Kirkham J, Brookes SJ, Inglehearn CF, Mighell AJ (2017) Amelogenesis
794 imperfecta; genes, proteins, and pathways. *Frontiers in Physiology*, **8**, 435.
795 https://doi.org/10.3389/fphys.2017.00435
796 Springer MS, Starrett J, Morin PA, Lanzetti A, Hayashi C, Gatesy J (2016) Inactivation of C4orf26 in toothless
797 placental mammals. *Molecular Phylogenetics and Evolution*, **95**, 34-45.
798 https://doi.org/10.1016/j.ympev.2015.11.002
799 Springer MS, Emerling CA, Gatesy J, Randall J, Collin MA, Hecker N, Hiller M, Delsuc F (2019) Odontogenic
800 ameloblast-associated (ODAM) is inactivated in toothless/enamelless placental mammals and toothed
801 whales. *BMC Evolutionary Biology*, **19**, 31. https://doi.org/10.1186/s12862-019-1359-6
802 Spurgin AM (1904) Enamel in the teeth of an embryo edentate (*Dasypus novemcinctus* linn). *American*
803 *Journal of Anatomy*, **3**, 75-84. https://doi.org/10.1002/aja.1000030105
804 Sreenath T, Thyagarajan T, Hall B, Longenecker G, D'Souza R, Hong S., Wright JT, MacDougall M, Sauk J,
805 Kulkarni AB (2003) Dentin Sialophosphoprotein knockout mouse teeth display widened predentin zone
806 and develop defective dentin mineralization similar to human dentinogenesis imperfecta type III. *Journal*
807 *of Biological Chemistry*, **278**, 24874-24880. https://doi.org/10.1074/jbc.M303908200
808 Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.
809 *Bioinformatics*, **30**, 1312-1313. https://doi.org/10.1093/bioinformatics/btu033
810 Stenzel U (2017) Network-aware-bwa (accessed 5.9.17). https://github.com/mpieva/network-aware-bwa
811 Sun Y, Chen L, Ma S, Zhou J, Zhang H, Feng JQ, Qin C (2011) Roles of DMP1 processing in osteogenesis,
812 dentinogenesis and chondrogenesis. *Cells Tissues Organs*, **194**, 199-204.
813 https://doi.org/10.1159/000324672
814 Toyosawa S, Fujiwara T, Ooshima T, Shintani S, Sato A, Ogawa Y, Sobue S, Ijuhin N (2000) Cloning and
815 characterization of the human ameloblastin gene. *Gene*, **256**, 1-11. https://doi.org/10.1016/S0378-
816 1119(00)00379-6
817 Ungar PS (2010) Mammal teeth: Origin, evolution and diversity. John Hopkins University Press, Baltimore,
818 Maryland.
819 Vizcaíno SF (2009) The teeth of the "toothless": novelties and key innovations in the evolution of xenarthrans
820 (Mammalia, Xenarthra). *Paleobiology*, **35**, 343-366. https://doi.org/10.1666/0094-8373-35.3.343
821 Wazen RM, Moffatt P, Ponce KJ, Kuroda S, Nishio C, Nanci A (2015) Inactivation of the odontogenic
822 ameloblast-associated gene affects the integrity of the junctional epithelium and gingival healing.
823 *European Cells and Materials*, **30**, 187-199. https://doi.org/10.22203/eCM.v030a13
824 Xiao S, Yu C, Chou X, Yuan W, Wang Y, Bu L, Fu G, Qian M, Yang J, Shi Y, Hu L, Han B, Wang Z, Huang W, Liu J,
825 Chen Z, Zhao G, Kong X (2001) Dentinogenesis imperfecta 1 with or without progressive hearing loss is
826 associated with distinct mutations in DSPP. *Nature Genetics*, **27**, 201-204. https://doi.org/10.1038/84848
827 Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme
828 evolution. *Molecular Biology and Evolution*, **15**, 568-573.
829 https://doi.org/10.1093/oxfordjournals.molbev.a025957
830 Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals.
831 *Journal of Molecular Evolution*, **46**, 409-418. https://doi.org/10.1007/PL00006320
832 Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**,
833 1586–1591. https://doi.org/10.1093/molbev/msm088
834 Zoonomia Consortium (2020) A comparative genomics multitool for scientific discovery and conservation.
835 *Nature*, **587**, 240-245. https://doi.org/10.1038/s41586-020-2876-6


AMTN
ODAM
AMELX
AMBN
ENAM
MMP20
ACPT
DSPP
ODAPH
DMP1
MEPE
V
Tooth loss, abnormal gingiva
P
Enamel loss
Normal teeth & gingiva
Abnormal gingiva
D
Abnormal enamel & abnormal gingiva
C
Abnormal enamel & abnormal gingiva
Enamel loss
Enamel loss
Enamel loss
Enamel loss
Enamel loss
Cyclopes didactylus
Myrmecophaga tridactyla
Tamandua mexicana
Tamandua tetradactyla
Choloepus hoffmanni
Choloepus didactylus
Bradypus torquatus
Bradypus pygmaeus
Bradypus tridactylus
Bradypus variegatus
Dasypus kappleri
Dasypus septemcinctus
Dasypus hybridus
Dasypus yepesi
Dasypus sabanicola
Dasypus novemcinctus
Dasypus pilosus
Euphractus sexcinctus
Chaetophractus villosus
Zaedyus pichiy
Chaetophractus vellerosus
Chaetophractus nationi
Chlamyphorus truncatus
Calyptophractus retusus
Priodontes maximus
Tolypeutes tricinctus
Tolypeutes matacus
Cabassous chacoensis
Cabassous centralis
Cabassous unicinctus
Cabassous tatouay
Vermilingua (Anteaters)
Folivora (Sloths)
Dasypodidae (long-nosed armadillos)
Chlamyphoridae (hairy, fairy, naked-tailed, giant, and three-banded armadillos)



GENE LEGEND

***AMELX**:* encodes amelogenin, the primary enamel matrix protein (90%), expressed in ameloblasts, inactivation strongly associated with enamel dysfunction

***AMTN**:* encodes amelotin, expressed in ameloblasts and junctional epithelium, inactivation has less-clear association with enamel dysfunction, interacts with ODAM likely to maintain tight seal at tooth-gingiva border

***DSPP**:* encodes dentin sialophosphoprotein, a dentin matrix protein, expressed in odontoblasts, inactivation associated with dentin dysfunction

***ENAM**:* encodes enamelin, the 3rd most abundant enamel matrix protein, expressed in ameloblasts, inactivation strongly associated with enamel dysfunction

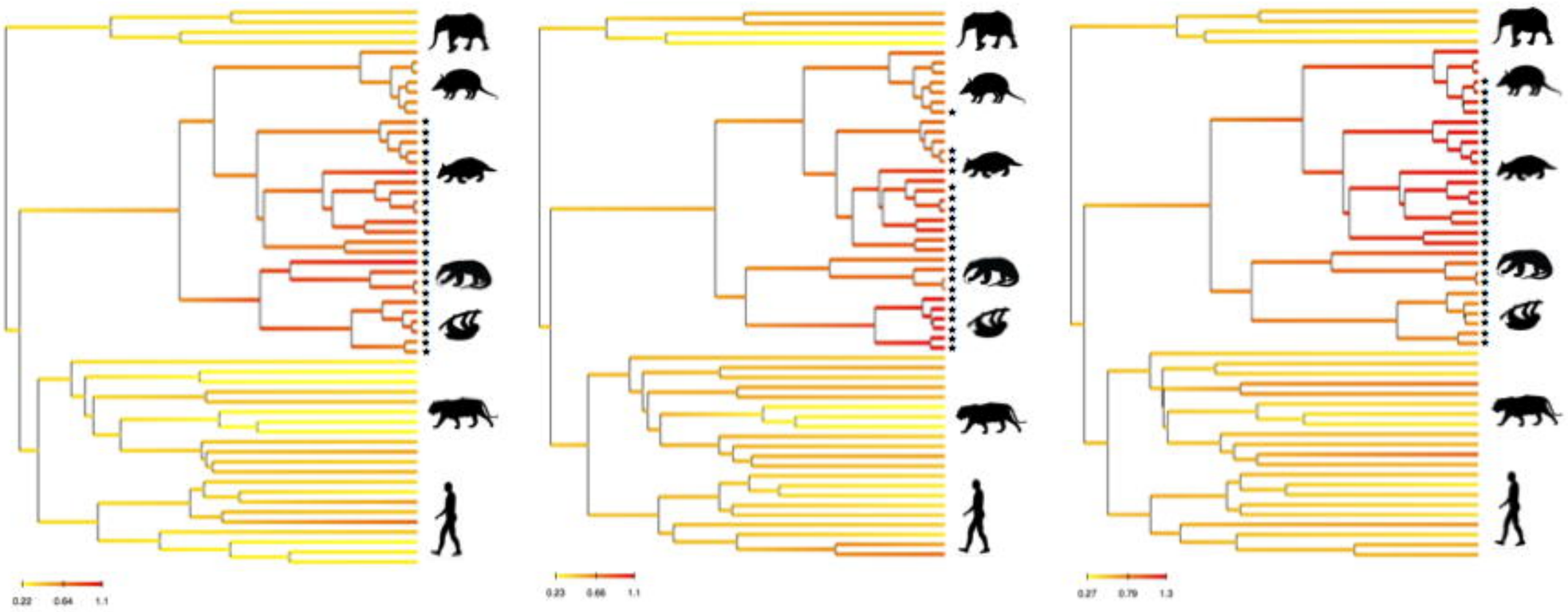
***MMP20**:* encodes metalloproteinase 20, a matrix metalloproteinase that cleaves enamel matrix proteins into mature forms, expressed in ameloblasts, inactivation strongly associated with enamel dysfunction

***ODAM**:* encodes odontogenic ameloblast-associated protein, expressed in ameloblasts and junctional epithelium, inactivation not associated with enamel dysfunction, interacts with AMTN likely to maintain tight seal at tooth-gingiva border

ODAPH: encodes ondotogenesis-associated phosphoprotein, expressed in enamel organ, inactivation associated with enamel dysfunction, interspecies comparisons suggest pseudogenization associated with tooth loss as opposed to enamel loss

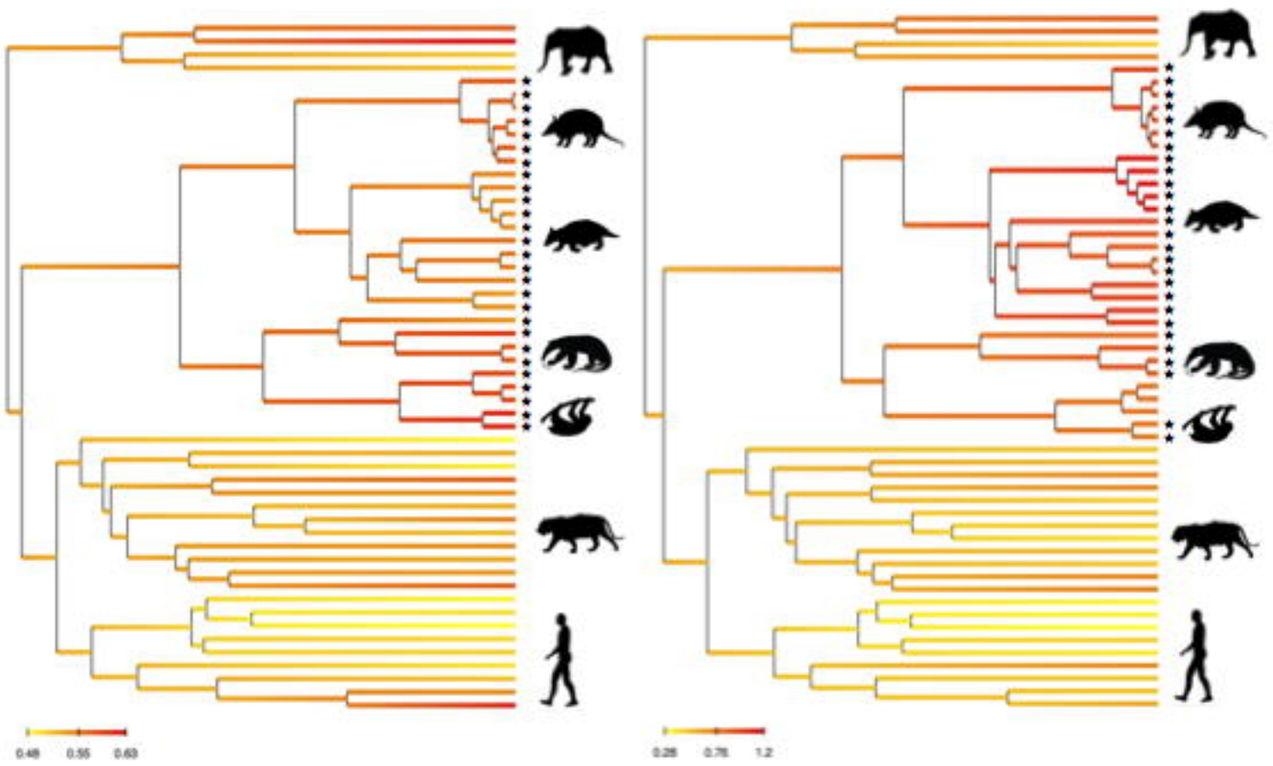
Enamel matrix-structure

AMELX *AMBN* *ENAM*



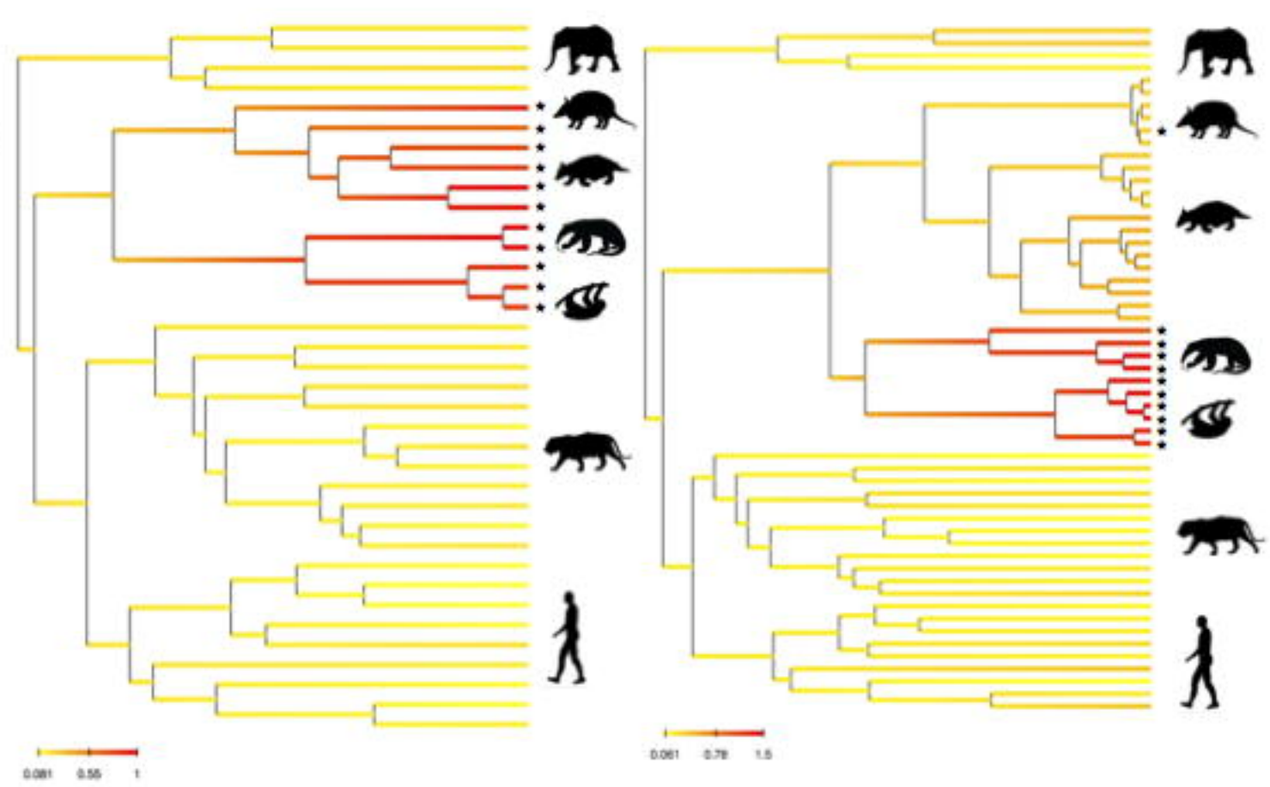
Enamel junction-associated

AMTN *ODAM*



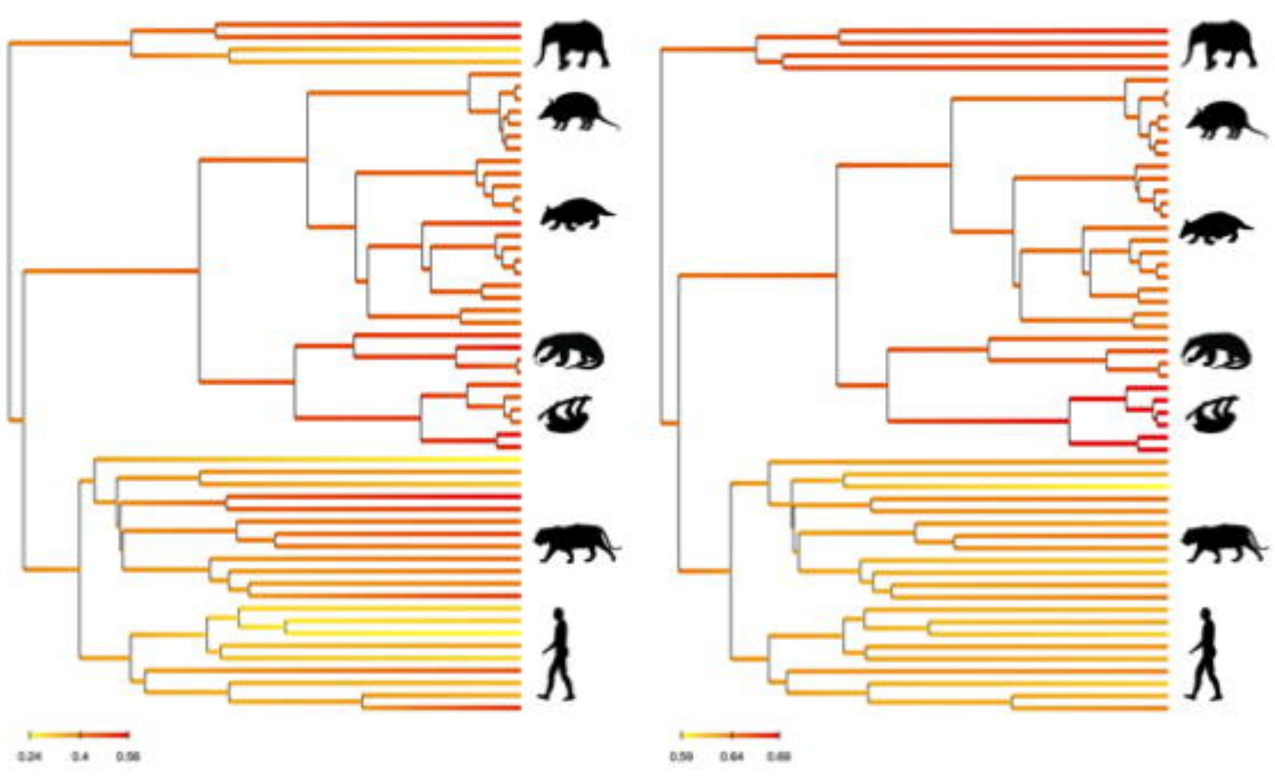
Enamel matrix-processing

ACPT *MMP20*



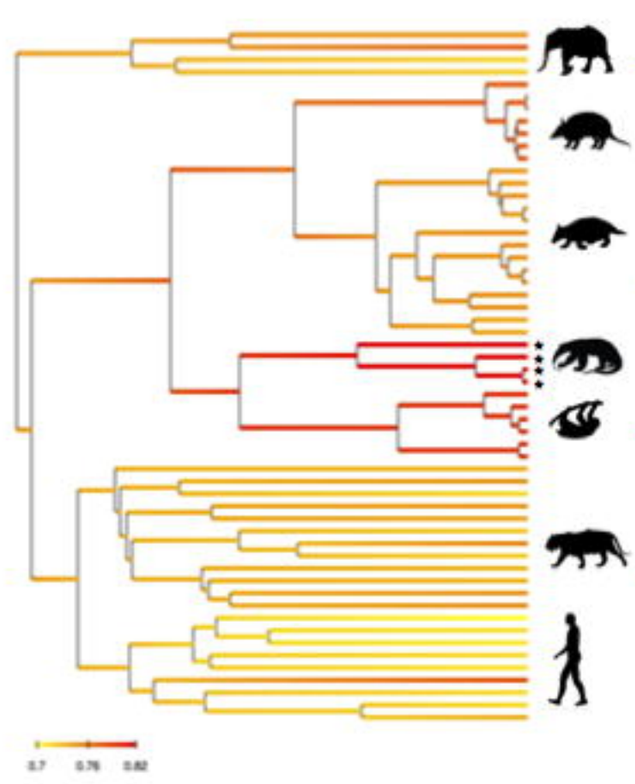
Mineralization-associated

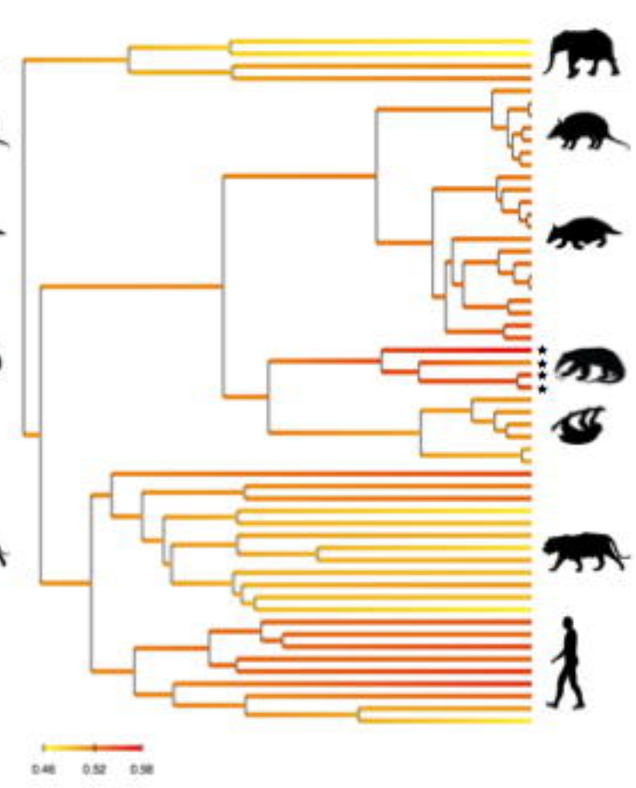
DMP1 *MEPE*



Dentin-associated

DSPP



Tooth-associated

ODAPH