Received: 30 July 2018 | Revised: 24 February 2019 | Accepted: 1 April 2019

DOI: 10.1002/jez.b.22857

RESEARCH ARTICLE

WILEY JEZ-B MOLECULAR AND DEVELOPMENTAL EVOLUTION

Reptile enamel matrix proteins: Selection, divergence, and functional constraint

Omar Alazem | John Abramyan

Department of Natural Sciences, University of Michigan-Dearborn, Dearborn, Michigan

Correspondence
John Abramyan, Department of Natural Sciences, University of Michigan-Dearborn, 114 SFC Bldg, 4901 Evergreen Rd., Dearborn, MI 48128.
Email: abramyan@umich.edu

Abstract

The three major enamel matrix proteins (EMPs): amelogenin (AMEL), ameloblastin (AMBN), and enamelin (ENAM), are intrinsically linked to tooth development in tetrapods. However, reptiles and mammals exhibit significant differences in dental patterning and development, potentially affecting how EMPs evolve in each group. In most reptiles, teeth are replaced continuously throughout life, while mammals have reduced replacement to only one or two generations. Reptiles also form structurally simple, aprismatic enamel while mammalian enamel is composed of highly organized hydroxyapatite prisms. These differences, combined with reported low sequence homology in reptiles, led us to predict that reptiles may experience lower selection pressure on their EMPs as compared with mammals. However, we found that like mammals, reptile EMPs are under moderate purifying selection, with some differences evident between *AMEL*, *AMBN*, and *ENAM*. We also demonstrate that sequence homology in reptile EMPs is closely associated with divergence times, with more recently diverged lineages exhibiting high homology, along with strong phylogenetic signal. Lastly, despite sequence divergence, none of the reptile species in our study exhibited mutations consistent with diseases that cause degeneration of enamel (e.g. amelogenesis imperfecta). Despite short tooth retention time and simplicity in enamel structure, reptile EMPs still exhibit purifying selection required to form durable enamel.

KEYWORDS

ameloblastin, amelogenin, enamelin, polyphyodont, reptile, tooth replacement, enamel matrix proteins

1 | INTRODUCTION

The evolution of teeth was a major milestone in vertebrate history. Teeth are unique structures comprised of dentin, cementum, and enamel; tissues found nowhere else in the amniote body (Bluteau, Luder, De Bari, & Mitsiadis, 2008; Jussila & Thesleff, 2012). The layered architecture, plus enamel covering, makes them resistant to wear and damage that can occur while the animal is procuring and processing food (Delgado, Davit-Beal, Allizard, & Sire, 2005). Due to the critical role that teeth play, their durability is crucial for the survival of most dentulous vertebrates, placing strong evolutionary pressure on the structural integrity of dental enamel.

Enamel is unique in its extreme hardness and the almost complete lack of cellular components, providing a durable surface for the tooth (Diekwisch et al., 2002). Epithelial cells called ameloblasts form enamel through the secretion and deposition of an organic extracellular matrix (Fincham, Moradian-Oldak, & Simmer, 1999). This matrix forms through the interaction of several proteins that are encoded by tooth-specific, nonpleiotropic genes (Delgado et al., 2001; Sire, Delgado, & Girondot, 2006; Sire, Delgado, Fromentin, & Girondot, 2005). The three main components of the enamel matrix are amelogenin (AMEL), ameloblastin (AMBN) and enamelin (ENAM), which together are referred to as enamel matrix proteins (EMPs; Kawasaki & Weiss, 2003). EMPs play essential roles

during enamel matrix formation, organization, and biomineralization (Gasse & Sire, 2015). At initial deposition, the enamel matrix is 80–90% protein and 10–20% mineral by volume (Moss-Salentijn, Moss, & Yuan, 1997). As mineralization progresses, the protein matrix is progressively degraded, resulting in nearly protein-free, mature enamel (reviewed by Moradian-Oldak & Goldberg, 2005).

AMEL, *AMBN*, and *ENAM* are phylogenetically related genes. *ENAM* is thought to have arisen first, followed by *AMBN* through tandem duplication, and then *AMEL* through duplication of *AMBN* (Sire et al., 2006). Despite being youngest, AMEL make up ~90% of the organic matrix in developing enamel and is thought to form the transient organic scaffold for mineralization, essential for hydroxyapatite crystal deposition and organization (Moradian-Oldak, Iijima, Bouropoulos, & Wen, 2003). AMBN only makes up ~5% of the organic enamel matrix and it is thought to have a number of roles in enamel development, including formation of enamel prism sheaths (Hu et al., 1997; Nanci et al., 1998), nucleation of calcium crystallite (Ravindranath, Chen, Zeichner-David, Ishima, & Ravindranath, 2004), and as an ameloblast adhesive molecule (Fukumoto et al., 2004; Sonoda et al., 2009). ENAM makes up another ~5% of the enamel matrix and is thought to work as a nucleator during the early phases of enamel mineralization and/or enamel crystal elongation (Al-Hashimi, Sire, & Delgado, 2009; Satchell et al., 2002). In humans, mutations in *AMEL*, *AMBN*, and *ENAM* can result in a condition called amelogenesis imperfecta, which makes enamel structure weak, brittle and more susceptible to wear and damage (reviewed by Smith et al., 2017).

Teeth also exhibit lineage-specific variation in structure and function. Most reptiles and amphibians possess relatively simple teeth, along with polyphyodont dentition, where teeth are continuously replaced throughout life (Edmund, 1960). Mammals, on the other hand, exhibit more complex tooth structure, with either a monophyodont or diphyodont replacement pattern where only one or two generations of teeth develop (Tucker & Fraser, 2014). Reptiles and amphibians also produce structurally simple, "aprismatic" enamel, while mammals evolved "prismatic" enamel, where hydroxyapatite crystallites bundle in an organized pattern (Diekwisch et al., 2009; Line & Novaes, 2005). It is hypothesized that reduction of tooth generations in primitive mammals intensified the need for greater durability of teeth; leading to the evolution of prismatic enamel (Grine, Vrba, & Cruickshank, 1979). This theory is supported by the fact that the only known example of prismatic enamel in reptiles is found in the Uromastyx lizard, which has a limited number of tooth generations due to acrodont teeth that fuse to the jaw and are not replaced (Bertin, Thivichon-Prince, LeBlanc, Caldwell, & Viriot, 2018; Cooper & Poole, 1973; Diekwisch et al., 2009; Throckmorton, 1979).

Along with enamel structure, tooth replacement has also been postulated to affect selection on tooth-associated genes. Assaraf-Weill et al. (2013) have previously hypothesized that amphibian EMPs experience lower constraint and attributed this to two differences between amphibian and mammalian dentition: polyphyodonty and lack of occlusion. Lack of occlusion may reduce wear while polyphyodonty relives the need for long-term resistance to damage. Similarly, Delgado, Ishiyama, and Sire (2007) also postulated that the constraints acting on the enamel structure could be less important in reptiles than in mammals due to their polyphyodont dentition. However, this question remains unexplored, largely due to the limited availability of reptilian sequences.

The authors cited above have studied the available sequences of tetrapod *AMEL*, *AMBN*, and *ENAM* at length, revealing details of their intron-exon boundaries, insertions, deletions, as well as conserved sites (Al-Hashimi et al., 2009; Al-Hashimi, Lafont, Delgado, Kawasaki, & Sire, 2010; Sire, Davit-Beal, Delgado, & Gu, 2007; Sire et al., 2005, 2006). However, the limited amount of nonmammalian sequence data has hindered a more comprehensive understanding of EMP evolution (Davit-Beal, Chisaka, Delgado, & Sire, 2007; Delgado, Couble, Magloire, & Sire, 2006). In this study, we were particularly interested in assessing how the simpler, polyphyodont dentition of reptiles affects selection pressure on reptile EMP orthologs in comparison to those of mammals; furthermore, whether all three EMPs experience the same evolutionary selection pressure and whether or not they are affected in the same manner. To this end, we applied in silico analyses to more than 20 reptile orthologs each of *AMEL*, *AMBN*, and *ENAM*, with the aim of better understanding patterns of homology, selection, and putative functional divergence.

2 | MATERIALS AND METHODS

2.1 | Sequence acquisition and multiple sequence alignment

Reptile genomes were downloaded and searched for *AMBN*, *AMEL*, and *ENAM* on an exon-by-exon basis using stand-alone BLAST v 2.2.18 (Altschul, Gish, Miller, Myers, & Lipman, 1990; Supporting Information File S1). To conserve current nomenclature, we followed previously published exon numbering system of reptile EMPs established by Sire and colleagues: AMEL (Gasse & Sire, 2015; Sire et al., 2005); AMBN (Gasse & Sire, 2015); ENAM (Al-Hashimi et al., 2010). After exons were identified, sequences were visually checked for intron-exon boundary GT-AG splice sites. Exons were concatenated and translated to amino acid sequence using the ExPASy-Translate tool (https://web.expasy.org/translate/) or batch translated at http://www.bioinformatics.org/sms2/translate.html.

Multiple sequence alignments of amino acid sequences were generated using Multiple Alignment using Fast Fourier Transform (MAFFT) v.7 (https://mafft.cbrc.jp/alignment/software/; https://mafft.cbrc.jp/alignment/server/; Katoh, Misawa, Kuma, & Miyata, 2002). MAFFT utilizes an iterative, progressive approach and finds homologous segments using Fast Fourier Transform. It is one of the most broadly used sequence alignment programs to date (Bawono et al., 2017). In fact, recent studies consistently rank MAFFT as one of the top multiple sequence alignment methods in terms of accuracy, speed, and consistency, in comparison to other methods such as T-Coffee, ClustalW, ProbCons, and Dialign (Chang, Di Tommaso, & Notredame, 2014; Durand, Hazelhurst, & Coetzer, 2010; Manzoor, Shahid, & Zafar, 2015; Nuin, Wang, &

Tillier, 2006; Pais, Ruy, Oliveira, & Coimbra, 2014; Thompson, Linard, Lecompte, & Poch, 2011; Wang et al., 2011; J. Yang & Warnow, 2011). Coding sequence alignments were subsequently generated by converting amino acid alignments to coding sequence alignments using the PAL2NAL tool (Suyama, Torrents, & Bork, 2006; http://www.bork.embl.de/pal2nal/). See Supporting Information File S2 for alignments.

2.2 | Phylogenetic analyses

Phylogenetic trees were generated in MrBayes v. 3.2 (Ronquist & Huelsenbeck, 2003) using coding sequences. Analysis for AMEL was run with Ngen = 5,000,000, Samplefreq = 500, and burnin = 1,250,000. Analyses for AMBN and ENAM were run with Ngen = 1,000,000, Samplefreq = 500, and burnin = 250,000. All analyses were run until they reached a likelihood score plateau (i.e. stationarity), identified by the standard deviation of split frequencies reaching below 0.01. The resulting trees were used in the generation of a 50%-majority rule consensus tree so that the proportion of trees at each node measured the Bayesian Posterior Probabilities (BPP) of each bipartition. Sequence evolution models for analysis were determined with jModelTest 2.1.10 (Darriba, Taboada, Doallo, & Posada, 2012). jModelTest 2.1.10 selected an HKY substitution model for all three genes (nst = 2), with gamma-distributed rates across sites for AMBN and AMEL (rates = gamma), and a combination of gamma-distributed and a proportion of invariable sites for ENAM (rates = invgamma). Mouse ortholog was used as an outgroup. Consensus tree was viewed using FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) and phylogeny figures traced and modified using Adobe Illustrator CS4 (Adobe Systems).

2.3 | Percent identity calculations

Percent identity values for amino acid sequences were calculated by uploading MAFFT alignments (as described above) into the Geneious software package (Biomatters, Auckland, New Zealand). Amino acid sequences were utilized for sequence identity analyses in order to obtain a more accurate estimate of the changes in the functional units of the proteins, and to avoid inflation of identity differences due to synonymous mutations. *Aspidoscelis* was not used in this analysis due to the low sample size. Gene-wide percent identity graphs were generated in Excel through a sliding window analysis with a window of 10 bases, with an overlap of nine bases between windows. We chose not to use a larger sliding window since this causes a "smoothing" effect, resulting in a loss of information. Pairwise divergence times were obtained from www.timetree.org.

2.4 | ω estimate using phylogenetic analysis using maximum likelihood (PAML)

Rates of synonymous (d*S* – silent) and nonsynonymous (d*N* – amino acid replacement) substitutions were analyzed in the CodeML program in PAML v. 4.4 (Z. Yang, 2007). The d*N*/d*S* ratio (ω) measures selection pressure on amino acids. An $\omega < 1$ estimate indicates purifying selection (d*N* < d*S*), $\omega \approx 1$ indicates neutral selection (d*S* ≈ d*N*), and $\omega > 1$ is considered to infer positive selection (d*N* > d*S*). To estimate ω, a tree-based likelihood approach was implemented as described by Z. Yang (1998). Branch-specific codon model analyses were used to estimate selection along specific branches of the species tree and applied to MAFFT codon alignments. The free-ratio model is the most general, parameter-rich model and allows for different ω values for each branch. The one-ratio model is the simplest and assumes the same ω for all branches. The two-ratio is intermediate and allows for two ω values, allowing an individually labeled branch (or group of branches) to differ (foreground) from the average ω across the unlabeled, "background" branches of the tree. Likelihood estimates assume the codon substitution model of Goldman and Yang (1994). The likelihood estimates for each were compared using a hierarchical likelihood ratio test (LRT) of twice the difference in the log likelihood value of the models being compared ($2\Delta l = 2(l_1 - l_0)$), with the result approximating a chi-square (χ^2) distribution (Z. Yang, 1998). Lizard infraorders were not analyzed separately due to their small number of representative species, since small sample size has been shown to negatively affect PAML analyses (Anisimova, Bielawski, & Yang, 2001; Yang, 2007).

2.5 | Molecular clock test using PAML

Molecular clock tests were conducted for the reptile data set using codon alignments for *AMEL*, *AMBN*, and *ENAM*. Analysis was performed according to Lemey and Posada (2009) with the BaseML program in PAML v. 4.4 (Z. Yang, 2007; Yoder & Yang, 2000), comparing Crocodilia versus Squamata, Serpentes versus Crocodilia + Lacertilia, as well as the individual lizard clades within a squamate-only data set. Both data sets were analyzed with the assumption of no molecular clock (clock = 0), a global molecular clock (clock = 1) and a local clock that tested for a difference between labeled clades in the data set (clock = 2). A likelihood ratio test was used to establish 95% confidence level for rejection of a global molecular clock. The following divergence times (from www.timetree.org) were used as calibration points: Crocodylidae – Alligatoridae – 80 mya; Gekkota - Lacertoidea – 201 mya; Serpentes - Lacertilia – 167 mya; Anguimorpha – Iguania – 165 mya (for clade designations see Reeder et al., 2015).

2.6 | Testing for relaxed selection

To test for relaxed selection, we used the RELAX software (Wertheim, Murrell, Smith, Kosakovsky Pond, & Scheffler, 2015), available on the Datamonkey web server (www.datamonkey.org/RELAX) and as part of the HyPhy software package (Pond, Frost, & Muse, 2005). RELAX estimates ω ratios, similar to PAML, and tests whether the selection is relaxed or intensified on a set of "test branches" compared with "reference branches" in a predefined tree. RELAX distributes sites among three ω classes: those under purifying selection with less nonsynonymous changes than expected ($\omega_1 < 1$),

those under neutral selection with roughly equal synonymous and nonsynonymous changes ($\omega_2 \approx 1$), and those under positive selection with more nonsynonymous changes (amino acid changes) than expected ($\omega_3 > 1$). It then calculates a selection intensity parameter (K), defined as $\omega_{reference} = \omega_{test}^{K}$. In the null model, the selection intensity is constrained to 1 for all branches, whereas in the alternative model, K is allowed to differ between the reference and test groups. Under relaxed selection on test branches, the dN/dS value in the purifying selection class will increase and the dN/dS value in the positive selection class will decrease. Consequently, relaxation of selection will move the sites in both the first ω category ($\omega < 1$) and the third ω category ($\omega > 1$) toward neutral (Wertheim et al., 2015). In other words, test branches should have an ω distribution skewed towards neutrality as compared with the reference branches if they are under relaxed selection. By raising each dN/dS class of the reference branches to the exponent K, the corresponding dN/dS class of the test branches is obtained. Therefore, $K > 1.0$ indicates intensified selection, while $K < 1.0$ indicates relaxed selection.

2.7 | Functional divergence analysis (DIVERGE)

Predicted functional divergence of EMPs was assessed using a maximum likelihood (ML) approach implemented in DIVERGE 2.0 (Gu, 1999, 2001, 2006; Gu & Vander Velden, 2002). DIVERGE uses a phylogenetic tree to assess site-specific changes in evolutionary rates between user-defined, monophyletic subclades after a divergence event (i.e. duplication, speciation) to identify amino acid residues with predicted functional divergence. Type I divergence refers to a shift in evolutionary rate that results in high conservation in one subclade, while the other evolves more freely in that position (Gu, 1999; Gu & Vander Velden, 2002). Type II divergence refers to a radical change in amino acid property resulting in amino acid positions that show clade-specific conservation (complete fixation within each), albeit different amino acids are fixed in each of the two clades, resulting in "conserved-but-different" residues (Gu, 1999, 2006). DIVERGE then calculates Gu's coefficient of evolutionary functional divergence (Θ) which ranges between 0 and 1, and measures changes in site-specific evolutionary rates. $\Theta = 0$ indicates no functional divergence, with an increase in Θ value as functional divergence increases (Gu, 1999, 2001). A position-specific posterior probability (PP) is then calculated, predicting the amino acid sites critical for divergence. Empirical cutoff for significance of PP values are established by sequentially removing the highest scoring residues from the alignment until Theta-ML (for Type I) and Theta-II (for Type II) are no longer significantly different from 0.

2.8 | Test for substitution saturation using DAMBE

When nucleotide substitutions within a site occur repeatedly (usually correlated with time), that position becomes saturated with polymorphisms. This may lead to an underestimation of synonymous substitutions per synonymous site (dS) and inflation of the ω value (Gojobori, 1983). To strengthen the reliability of our ω calculations, we tested saturation by applying the index of substitution saturation approach described by Xia, Zheng, Salemi, Chen, and Wang (2003) and implemented in DAMBE v.6 (Xia & Lemey, 2009; Xia & Xie, 2001). This test calculates an entropy-based index of substitution saturation (I_{SS}) and a critical index of substitution ($I_{SS.C}$). The $I_{SS.C}$ value is calculated from the critical tree length, the sequence length of the alignment, and the number of operational taxonomic units. The $I_{SS.C}$ serves as the cut-off value beyond which sequences will fail to recover the true phylogenetic tree. If I_{SS} is higher than $I_{SS.C}$, the sequences have experienced high level of saturation and have limited use in phylogenetic analyses (Xia & Lemey, 2009). We performed analyses for all three codon positions following Xia and Lemey (2009).

2.9 | Amelogenesis imperfecta sites

The LOVD database provides a curated list of all published amelogenesis imperfecta mutation sites in human (http://dna2.leeds.ac.uk/LOVD/; Smith et al., 2017). We used the database to evaluate corresponding positions in reptile sequences for *AMEL*, *AMBN*, and *ENAM* for putatively amelogenesis imperfecta-causing mutations.

3 | RESULTS

3.1 | Reptile EMP orthologs

Our sequences largely conformed to those previously described (Al-Hashimi et al., 2010; Gasse & Sire, 2015), yet we did have some novel, noteworthy findings. Gasse and Sire (2015) previously reported the loss of *AMBN* exon 7 in *Anolis*. We show here that this loss is restricted to the members of Iguania (*Anolis* and *Pogona*) only, with all other squamates exhibiting an intact exon 7 in their *AMBN* orthologs. We were also unable to identify *AMBN* exon 4 in all three gecko species, which may indicate a putative loss of this exon in Gekkota.

Alignment of *AMEL* orthologs from 11 mammal species that share a most recent common ancestor (MRCA) ~180 million years ago (mya) revealed 91 conserved amino acid residues (using MAFFT alignments; total length in *Mus* – 219 aa). 23 species of reptiles (MRCA ~280 mya) exhibited 41 conserved amino acid residues (total length in *A. mississippiensis* – 199 aa). Alignment of reptiles and mammals revealed 27 conserved residues across ~310 my divergence. Mammal *AMBN* alignment revealed 85 conserved residues (total length in *Mus* – 422 aa), while alignment of 22 reptile sequences revealed 66 conserved residues (total length in *A. mississippiensis* – 407 aa). Alignment of mammals and reptiles revealed 21 conserved residues. *ENAM* is the longest EMP and mammal alignment revealed 155 conserved residues (total length in *Mus* – 1274 aa), while 22 reptile species shared 93 conserved residues (total length in *A. mississippiensis* – 1092 aa). Mammal + reptile alignment exhibited 42 conserved amino acid residues (Supporting Information File S2).

3.2 | Phylogeny

We used MrBayes to construct phylogenetic trees from the coding sequences of reptile EMPs. Gene phylogenies revealed topologies that were largely similar to known species relationships in squamates (Reeder et al., 2015; Vidal & Hedges, 2009; Wiens et al., 2012; Zheng & Wiens, 2016) and crocodilians (Man, Yishu, Peng, & Xiaobing, 2011; Figure 1). Crocodilians and squamates are reciprocally monophyletic in all three gene trees (*AMEL*, *AMBN* and *ENAM*), with disproportionately long branches leading to crocodilians, squamates, as well as snakes. Species relationships are maintained in crocodilians, except in *AMEL*, where *Paleosuchus* was placed as sister to Crocodylidae (albeit with poor support – BPP = 0.58; Figure 1a). In squamates, all three genes yield monophyletic clades for Gekkota (*Eublepharis*, *Gekko*, *Lepidodactylus*) and Serpentes (snakes), but there are differences in their relationships. *AMEL* reveals a polytomy between *Aspidoscelis*, Gekkota, and Toxicofera (a clade comprised of Anguimorpha – *Ophisaurus*, *Shinisaurus*; Iguania *Pogona*, *Anolis*; and Serpentes), with Anguimorpha as a sister group to Serpentes. *AMBN*, on the other hand, shows Iguania as a sister to Serpentes (Figure 1b). *ENAM* shows Serpentes as sister to a monophyletic clade comprised of Anguimorpha and Iguania (Figure 1c). *Pogona vitticeps* is part of the family Agamidae and is unique in our data set in that it lacks lifelong tooth replacement. However, it did not exhibit significant differences in sequence and its phylogenetic position with *Anolis* was retained for all genes. Therefore, we decided to keep it as part of our data set.

3.3 | Percent identity and molecular clock

To quantify divergence between reptile EMPs, we calculated percent identity values between amino acid sequences of orthologs (see Supporting Information File S3 for a comprehensive table). In crocodilians, AMEL exhibited an average percent identity of 98.79 ± 0.59 (average ± standard deviation), while squamates aver-

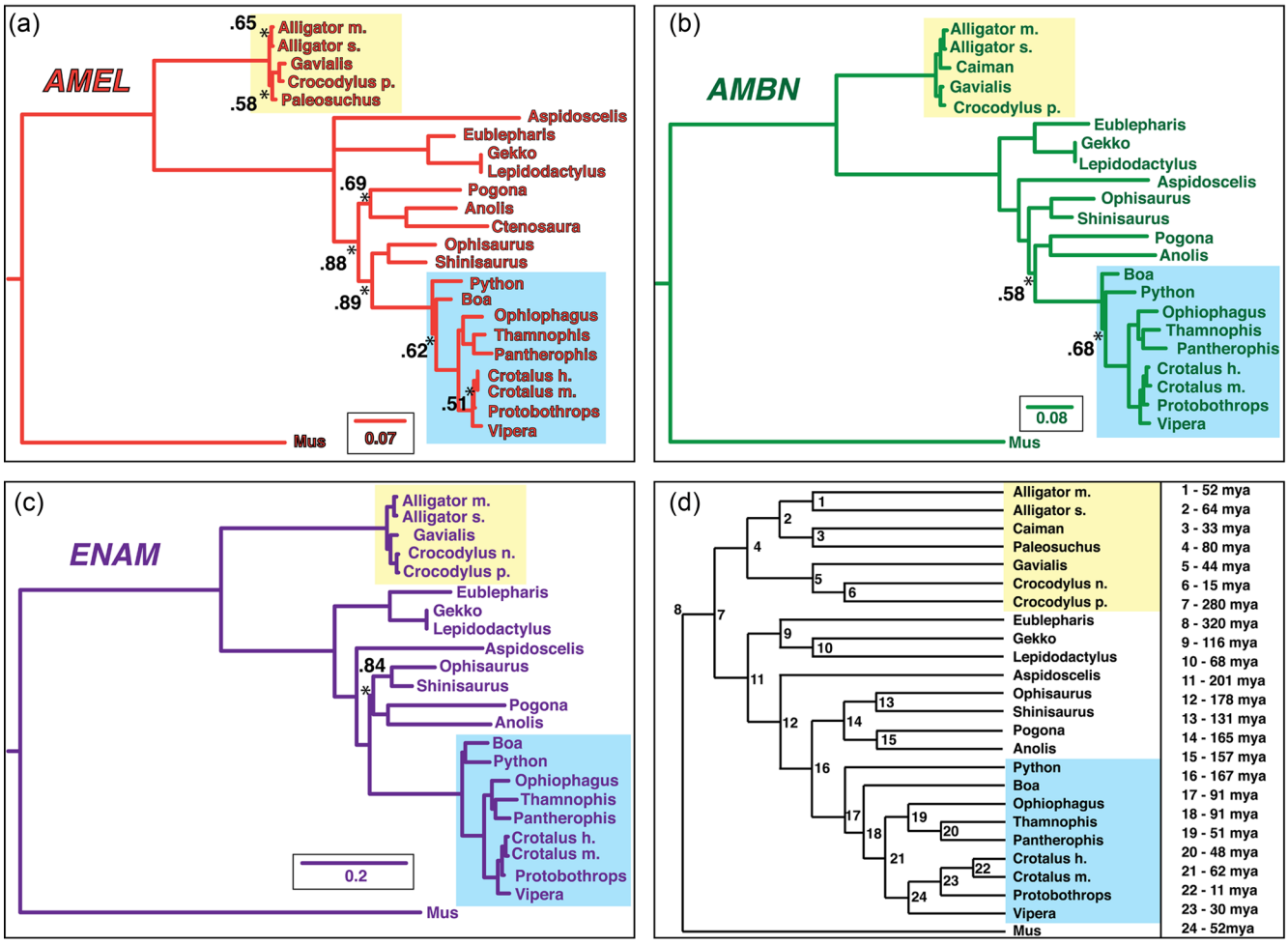

FIGURE 1 Phylogeny of reptile EMPs. (a) Bayesian phylogeny based on MAFFT codon-specific alignments of reptile *AMEL*, *AMBN*, and *ENAM* coding sequences, with mouse orthologs as outgroups (a–c). Multifurcations correspond to branches with BPP < 0.5. Nodes labeled with asterisks indicate BPP 0.50 ≤ 0.90, while unlabeled branches at bifurcations exhibit BPP ≥ 0.90. Monophyletic crocodilian clades are highlighted in yellow boxes, while snakes are highlighted in light blue. Scale bars indicate the number of substitutions per site. (d) Cladogram of known species relationships as estimated by Reeder et al. (2015) for squamates and Man et al. (2011) for crocodilians, including average estimates of divergence times obtained from www.timetree.org. AMBN: ameloblastin; AMEL: amelogenin; BPP: Bayesian posterior probabilities; EMPs: enamel matrix proteins; ENAM: enamelin [Color figure can be viewed at wileyonlinelibrary.com]

aged 68.57 ± 4.35. However within squamate groups, we observe higher values, with snakes exhibiting an average value of 89.24 ± 4.81, while members of the infraorders Gekkota, Anguimorpha, and Iguania averaged 88.09 ± 10.31, 79.79, and 69.16 ± 3.05, respectively (Figure 2a). AMBN was also highly conserved in crocodilians, with an average identity of 94.50 ± 1.74, while squamates were again lower at 68.31 ± 13.89. Within squamates, snakes exhibited an average of 88.05 ± 4.88, while Gekkota, Anguimorpha, and Iguania averaged 81.79 ± 12.42, 84.71, and 61.88, respectively (Figure 2b). ENAM sequences also revealed a similar pattern, albeit with slightly lower values. Crocodilians and squamates averaged 94.71 ± 3.98 and 60.09 ± 17.43, respectively. Within squamates, snakes averaged 82.96 ± 8.54, and members of Gekkota, Anguimorpha and Iguania exhibited percent identity values of 82.61 ± 14.01, 78.71, and 51.33, respectively (Figure 2c). When percent identity was mapped across the entire length of each gene in a sliding window analysis, reptile exons exhibited comparable identity landscapes to mammals for AMEL, AMBN, and ENAM (Figure 2a′–c′).

The simplest explanation for the sequence identity values described above would be the correlation with the ages of each clade, which would imply a uniform rate of evolution (i.e. a molecular clock). Therefore we decided to test whether a global molecular clock exists among reptiles. For all analyses described in Table 1, we first compared a parameter-rich, no molecular clock model (clock = 0) with a global molecular clock model (clock = 1), and always found clock = 0 fit the data significantly better than clock = 1 (data not shown). We then compared clock = 1 (global clock) model with a local clock model (clock = 2), which allow for separate rate estimates for two predefined groups. When a two-rate model with crocodilians and

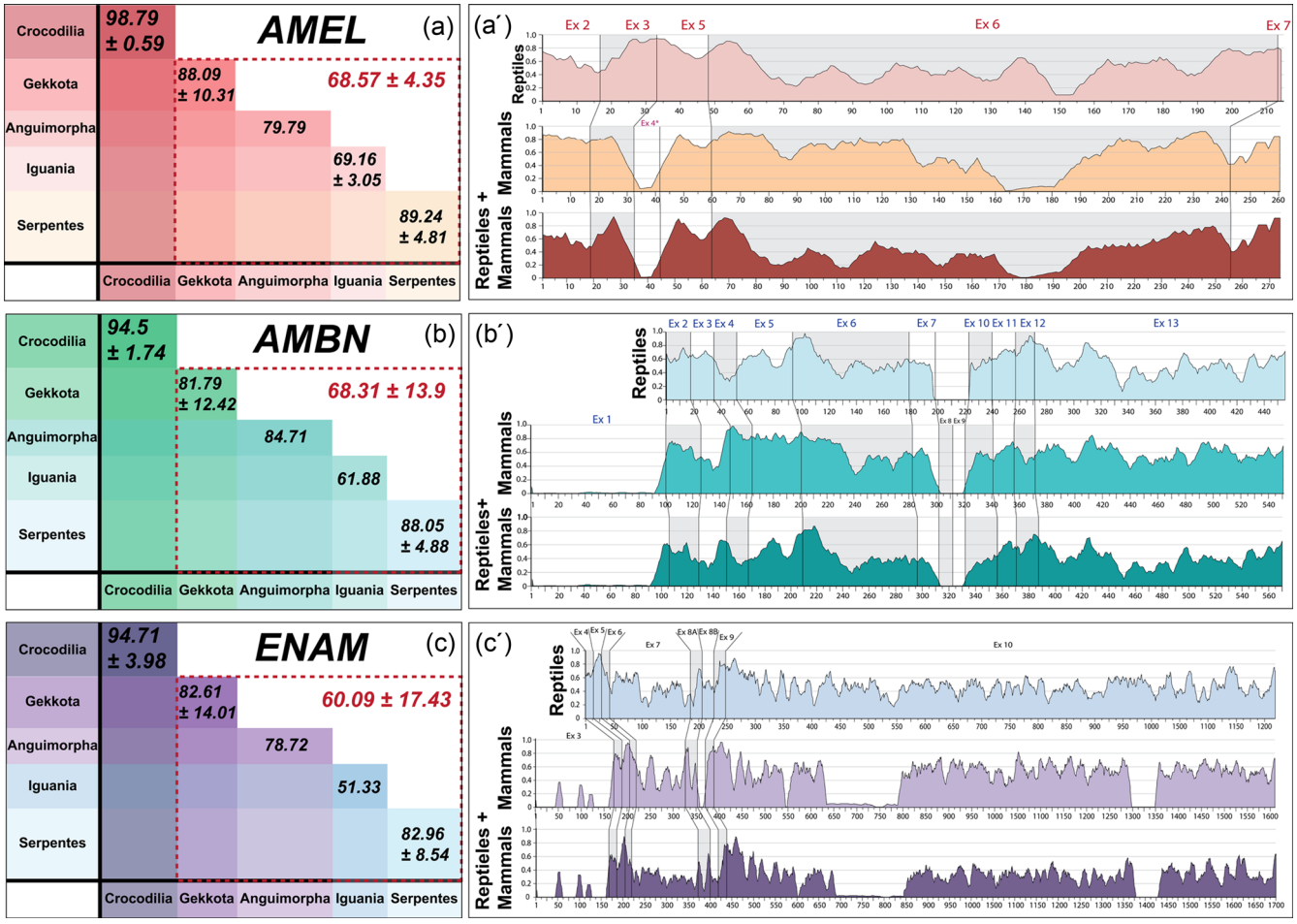

FIGURE 2 Amino acid sequence identity. (a-c) Average percent identity calculation utilizing a MAFFT alignment of full-length reptile EMP amino acid sequences. Values represent average percentage identity ± standard deviation within each group. For all three EMPs: Gekkota = *Eublepharis*, *Gekko*, *Lepidodactylus*; Anguimorpha = *Ophisaurus*, *Shinisaurus*; Serpentes = *Python*, *Boa*, *Vipera*, *Protobothrops*, *Crotalus m.*, *Crotalus h.*, *Thamnophis*, *Ophiophagus*, and *Pantherophis*. AMEL: Crocodilia = *Alligator m.*, *Alligator s.*, *Paleosuchus*, *Gavialis*, and *Crocodylus p.*; Iguania = *Pogona*, *Ctenosaura*, *Anolis*. AMBN: Crocodilia = *Alligator m.*, *Alligator s.*, *Caiman*, *Gavialis* and *Crocodylus p.*; Iguania = *Pogona*, *Anolis*; ENAM: Crocodilia = *Alligator m.*, *Alligator s.*, *Gavialis*, *Crocodylus p.*, and *Crocodylus n.*; Iguania = *Pogona*, *Anolis*. Panels a'–c' depict gene-wide amino acid percent identity graphs with sliding window averages of 10 bp, with a 9 bp overlap within the entire reptile data set "Reptiles," the entire mammalian data set "Mammals," and percent identity across both data sets aligned together "Reptiles + Mammals." AMBN: ameloblastin; AMEL: amelogenin; EMPs: enamel matrix proteins; ENAM: enamelin; MAFFT: Multiple Alignment using Fast Fourier Transform [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Molecular clock analysis for reptile data set

	Gene	Model	nP	Parameters	ln *L*		2Δ*l*	DF	*p* Value
Reptiles only data set									
A	AMEL	Clock 1	21	Global clock	−4550.02				
B		Clock 2	22	Crocodilia [f] Squamata [b]	−4502.95	**A vs. B**	94.15	1	1.6e−22
C		Clock 2	22	Serpentes [f] Crocodilia + Lacertilia [b]	−4549.95	A vs. C	0.15	1	0.974
A	AMBN	Clock 1	20	Global clock	−9993.64				
B		Clock 2	21	Crocodilia [f] Squamata [b]	−9905.45	**A vs. B**	176.38	1	1.5e−40
C		Clock 2	21	Serpentes [f] Crocodilia + Lacertilia [b]	−9985.10	**A vs. C**	17.07	1	2e−05
A	ENAM	Clock 1	20	Global clock	−29977.64				
B		Clock 2	21	Crocodilia [f] Squamata [b]	−29704.64	**A vs. B**	546.01	**1**	4.7e−12
C		Clock 2	21	Serpentes [f] Crocodilia + Lacertilia [b]	−29929.84	**A vs. C**	95.61	**1**	9.6e−23
Squamate only data set									
A	AMBN	Clock 1	17	Global clock	−8253.68				
B		Clock 2	18	Serpentes [f]	−8252.88	A vs. B	1.60	1	0.142
C		Clock 2	18	Anguimorpha [f]	−8229.14	**A vs. C**	49.08	1	1.3e−12
D		Clock 2	18	Gekkota [f]	−8252.98	A vs. D	1.40	1	0.167
E		Clock 2	18	Iguania [f]	−8246.74	A vs. E	13.88	1	0.001
A	AMEL	Clock 1	18	Global clock	−3835.68				
B		Clock 2	19	Serpentes [f]	−3835.64	A vs. B	0.08	1	1.39
C		Clock 2	19	Anguimorpha [f]	−3829.02	**A vs. C**	13.32	1	0.001
D		Clock 2	19	Gekkota [f]	−3835.68	A vs. D	0	1	–
E		Clock 2	19	Iguania [f]	−3834.51	A vs. E	2.34	1	0.081
A	ENAM	Clock 1	17	Global clock	−25064.85				
B		Clock 2	18	Serpentes [f]	−25047.59	**A vs. B**	34.52	1	1.7e−09
C		Clock 2	18	Anguimorpha [f]	−24940.81	**A vs. C**	248.08	1	3.4e−56
D		Clock 2	18	Gekkota [f]	−25064.85	A vs. D	0	1	–
E		Clock 2	18	Iguania [f]	−25044.63	**A vs. E**	40.44	1	1.0e−10

Note. AMBN: ameloblastin; AMEL: amelogenin; ENAM: enamelin; [f]: foreground; [b]: background.

squamates was tested against a global model across both, the two-rate model was a significantly better fit for *AMEL*, *AMBN*, and *ENAM* (Table 1). A similar two-rate model comparison with snakes as foreground and the rest of the reptiles (crocodilians + Lacertilia) as background fit *AMBN* and *ENAM*, but not *AMEL*. We also tested a squamate-only data set with each of the squamate subclades individually set as foreground (Serpents, Anguimorpha, Gekkota, and Iguania) and uncovered more variation in rates. A two-rate model with snakes as foreground only fit *ENAM*, while *AMEL* and *AMBN* did not exhibit rate heterogeneity between snakes and lizards. Anguimorpha in the foreground exhibited rate heterogeneity in *AMEL*, *AMBN*, and *ENAM*, while Iguania as foreground exhibited rate heterogeneity for *AMBN* and *ENAM*, but not *AMEL* (Table 1). Two-rate models with Gekkota as foreground did not allow rejection of the global molecular clock model for any of the genes analyzed.

3.4 | Branch-specific selection analysis

After observing rate heterogeneity in *AMEL*, *AMBN*, and *ENAM* between the various reptile lineages, we decided to explore whether they are under different selection regimes despite their close functional relationship. To test for selection intensity, we applied a

TABLE 2 In silico assessment of ω using branch models in PAML, selection intensity parameter (K) in RELAX, and coefficient of functional divergence (Θ) in DIVERGE between orthologs of AMEL, AMBN, and ENAM

	PAML		RELAX		DIVERGE	
Mammals vs. Reptiles	**Reptile-ω**	**Mammal-ω**	**Reptiles [T]**	***K***	**Type I – Θ_1**	**Type II – Θ_{1I}**
AMEL	0.5798	0.3486	Relaxation	0.43	ns	ns
AMBN	0.3655	0.4371	Relaxation	0.66	0.63	ns
ENAM	0.4832	0.4268	ns	0.95	0.56	−0.88
Reptiles only	**PAML**		**RELAX**		**DIVERGE**	
Crocodilians vs. squamates	**Crocodilian-ω**	**Squamate-ω**	**Crocodilian [T]**	***K***	**Type I – Θ_1**	**Type II – Θ_{1I}**
AMEL	0.1673	0.5454	Relaxation	0	ns	ns
AMBN	0.7953	0.3471	ns	0.77	ns	ns
ENAM	0.4784 ns		Relaxation	0.51	ns	ns
Reptiles only	**PAML**		**RELAX**		**DIVERGE**	
Snakes vs. background reptiles (crocodilians and lizards)	**Snake-ω**	**Croc + Lizard-ω**	**Snake [T]**	***K***	**Type I – Θ_1**	**Type II – Θ_{1I}**
AMEL	0.4919 ns		ns		ns	ns
AMBN	0.3609 ns		ns		0.55	ns
ENAM	0.4784 ns		ns		0.50	ns
Squamates only	**PAML**		**RELAX**			
Snakes vs. background squamates (lizards)	**Snake-ω**	**Lizard-ω**	**Snakes [T]**	***K***		
AMEL	0.3988	0.5869	Relaxation	0.55		
AMBN	0.3367 ns		ns			
ENAM	0.4960	0.4533	ns			

Note. $p \geq 0.05$ (LRT) = ns (detailed calculations are shown in Supporting Information Files S4–S7).
AMBN: ameloblastin; AMEL: amelogenin; ENAM: enamelin; ns: not significant; [T]: test branches (rest of the data set is left as reference [R] branch set.

series of ML branch-based models of selection in the CodeML program in PAML v.4.7 (Z. Yang, 2007). For each gene, we first compared a free-ratio model, which assumes different ω for each branch, to a one-ratio model, which assumes the same ω for all branches (Supporting Information File S4). In all cases, LRT found that the difference between the models is highly significant ($p < 0.01$), rejecting the simpler, one-ratio model and confirming that branches are indeed evolving at different rates. Subsequently, the one-ratio model was compared with a two-ratio model with reptiles labeled as foreground (ω_f) and mammals retained as background (ω_b). LRT again found a significant difference ($p \leq 0.01$) for all three EMPs, rejecting the one-ratio model, and showing that reptile and mammal EMPs are evolving under different selective constraints. Results for *AMEL* and *ENAM* yielded higher ω estimates for reptiles than mammals (*AMEL*: $\omega_f = 0.5798$ vs. $\omega_b = 0.3486$ and *ENAM*: $\omega_f = 0.4839$ vs. $\omega_b = 0.4268$), while *AMBN* revealed an opposite pattern $\omega_f = 0.3655$ versus $\omega_b = 0.4371$ (Table 2, Supporting Information File S4).

We also decided to investigate potential differences in ω between major reptile groups as part of a reptile-only data set. When crocodilians (ω_f) were compared with squamates (ω_b), the estimate for *AMEL* found $\omega_f = 0.1673$ and $\omega_b = 0.5454$. For *AMBN*, crocodilians exhibited $\omega_f = 0.7953$ compared with squamates $\omega_b = 0.3471$. Analysis of *ENAM* found that a two-ratio model did not fit the data significantly better than a one-ratio model, which estimated $\omega = 0.4784$ for all reptiles (Table 2, Supporting Information File S5). When snakes were labeled as foreground (ω_f) and the rest of the reptile branches (lizards and crocodilians) as background (ω_b), the two-ratio model did not fit better than the one-ratio model for any of the EMPs. When only squamates were analyzed, there was no difference in ω value between snake *AMBN* and lizards, while snakes exhibited a lower ω for *AMEL* and slightly higher ω for *ENAM* (Table 2, Supporting Information File S5). Individual lizard clades were not analyzed due to small sample sizes.

3.5 | Testing for relaxed selection

It is often difficult to assess whether a difference in ω is due to intensification or relaxation of selection since both positive selection as well as relaxed selection may result in elevation of ω (Wertheim et al., 2015). We, therefore, implemented RELAX to identify cases of truly relaxed selection. When reptiles were compared with mammals, RELAX identified significant signatures of relaxed selection in the analysis of *AMEL* and *AMBN* (AMEL – $K = 0.43$ and AMBN – $K = 0.66$), but not *ENAM* (Table 2; Supporting Information File S6). Within reptiles, crocodilian *AMEL* and *ENAM* exhibited significant signatures of relaxed selection ($K = 0$ and $K = 0.51$, respectively) in comparison

to squamates, while analysis of *AMBN* failed to find a significant selection difference (Supporting Information File S6). When snakes were tested against the rest of the reptile data set (crocodilians + lizards), they did not reveal statistically significant relaxation, while snakes compared with lizards, exhibited relaxation in *AMEL* ($K = 0.55$), but not in *AMBN* or *ENAM* (Table 2; Supporting Information File S6).

3.6 | Functional divergence analysis

Since we observed significant differences in selection between mammals and reptiles, as well as reptile subclades, we implemented DIVERGE to assess whether these differences could translate to functional divergence between orthologs. When mammals and reptiles were compared, sites with significant putative Type I divergence were identified in AMBN ($\Theta_1 = 0.63 \pm 0.08$; $p < 0.01$) and ENAM ($\Theta_1 = 0.56 \pm 0.05$; $p < 0.01$) (Table 2; sites highlighted in Supporting Information File S7). ENAM also exhibited putative Type II divergent sites ($\Theta_2 = -0.88 \pm 0.28$; $p < 0.01$) between reptiles and mammals, while AMEL did not exhibit sites with either divergence type (Table 2; Supporting Information File S7).

When crocodilians and squamates were compared, none of the EMPs exhibited sites with predicted Type I or Type II functional divergence. However, when snakes were compared with the rest of the reptiles (crocodilians + lizards), sites with putative Type I divergence were predicted for AMBN ($\Theta_1 = 0.55 \pm 0.14$; $p < 0.01$) and ENAM ($\Theta_1 = 0.50 \pm 0.10$; $p < 0.01$) (Table 2; Supporting Information File S7). When snakes and crocodilians were compared, all three EMPS exhibited sites with predicted Type II divergence (Supporting Information File S7). Interestingly, none of the analyses predicted functional divergence of either type for AMEL.

3.7 | Saturation analysis

The substantial length of time since the divergence of mammals and reptiles, as well as crocodilians and squamates (~310 and ~280 mya, respectively), increases the risk of substitution saturation that could affect d*N*/d*S* calculations by underestimating d*S*. To strengthen the reliability of the d*N*/d*S* estimates performed by PAML and RELAX, saturation levels were measured using DAMBE (Xia & Xie, 2001). The test was applied to all three EMPs for positions 1 + 2 and position 3 separately, for mammals-only, reptiles-only, and mammal + reptile alignments. All alignments exhibited I_{SS} indices significantly lower than the corresponding $I_{SS.C}$ index for symmetric trees, indicating little to no saturation (Xia & Lemey, 2009) (Supporting Information File S8).

3.8 | Amelogenesis imperfecta sites

We searched the reptile sequences for potential amelogenesis imperfecta-causing mutations identified in the LOVD database and found none of the mutations previously identified. That said, several of the disease-associated positions did exhibit a different amino acid in reptile AMEL and ENAM than their human orthologs, warranting further investigation into these sites.

4 | DISCUSSION

It is well documented that major changes in amniote dentition such as edentulation or enamel loss have a demonstrable effect on the underlying tooth-specific genes. Moreover, the degree of effect is gradual, with a weak case of relaxed selection seen in the platypus (Al-Hashimi et al., 2009), to enamel loss in Xenarthrans (Delsuc, Gasse, & Sire, 2015; Meredith, Gatesy, Murphy, Ryder, & Springer, 2009), and finally, complete edentulation (e.g. birds, turtles), which results in pseudogenization of EMPs as well as other tooth-associated genes (Meredith, Gatesy, & Springer, 2013; Shaffer et al., 2013). Here we show that AMEL, AMBN and ENAM differ in both selection pressure and the resulting changes in coding sequence that they experience.

4.1 | Sequence homology in reptile EMPs

Percent identity between sequences is commonly used as a universal metric to describe the degree of homology (Jones, Taylor, & Thornton, 1992), and studies have previously noted low sequence identity between reptile EMPs (Delgado et al., 2006; Sire et al., 2006). We show here that this phenomenon is correlated with the age of the clade, potentially in conjunction with substitution rate heterogeneity. Crocodilians and snakes represent relatively younger divergences than the various lizard groups (80 and 91 my, respectively), and exhibit substantially higher percent identity values. They also exhibit short terminal branches (which correlate with the high identity values) and the prediction of Type II functional divergence, which indicates fixation of a different amino acid residue in each clade. That said, both crocodilians and snakes exhibit a disproportionately long branch leading to the group, indicating substantial accumulation of variation which has since slowed, presumably due to fixation. Even within lizards, where we see relatively low sequence identity, values correlated with divergence time. For example, *Anolis* and *Pogona* diverged 157 mya and exhibited the lowest sequence identity values in our analysis (68.75, 61.88, and 51.33 for *AMEL*, *AMBN*, and *ENAM*, respectively), while *Shinisaurus* and *Ophisaurus* diverged 131mya and exhibit slightly higher identity values of 79.79, 84.71, and 78.72 for *AMEL*, *AMBN*, and *ENAM*, respectively. Still, divergence time may not sufficiently explain the aforementioned identity values in their entirety.

Crocodilians particularly stood out for their high sequence identity. These values are high even for the relatively recent divergence between *Alligator* and *Crocodylus* (~80 my). In fact, they are comparable to primates of approximately similar divergence time (~75 my between human and lemur – Richard, Delgado, Gorry, & Sire, 2007). On the other hand, bears (*Ursus* sp.) and pigs (*Sus* sp.) diverged ~78 my ago and exhibit much more divergence in amino acid sequence (AMEL – 88.35%; AMBN – 79.42%; ENAM – 79.77%). Thus, even in comparison to mammals, crocodilians are on the conservative end of the spectrum. A possible reason for the high sequence identity in crocodilian EMPs may be associated with an

exceptionally low substitution rate found in the group (Eo & DeWoody, 2010; Green et al., 2014). Our molecular clock analysis also identified a significant difference in substitution rate between Crocodilia and Squamata for all three EMPs.

Squamates exhibit much lower sequence identity, but we show here that most of the sequence variation arises from differences between subclades, while identity was much higher within snakes and lizard infraorders. With only two to three species in each of the lizard groups, it is premature to draw collusions until more sequences are available. However, we did have a significant number of snake species in the analysis. Other studies have found that snakes have an accelerated substitution rate (Eo & DeWoody, 2010; Green et al., 2014), but our analyses did not reflect this for EMPs. In fact, we did not find a difference in substitution rate between snakes and other squamates for *AMEL* and *AMBN*, and furthermore identified high percent identity values in amino acid residues between snake species, which indicates lower number of nonsynonymous substitutions. A possible explanation for this disparity could be that most calculations utilize a molecular clock that assumes a uniform rate throughout time. However, our phylogenetic reconstructions reveal relatively short terminal branches with a longer branch leading to all snakes (e.g. ENAM), indicating a large accumulation of changes early in snake evolution (after divergence from the rest of Toxicofera), with relatively less change in the extant assemblage, as may be observed in Figure 1, and highlighted in Supporting Information File S9. Furthermore, this pattern is recapitulated across many studies. Jiang et al. (2007) even describe a similar pattern of evolution for the mitochondrial genomes of snakes. In their study, analysis of mitochondrial protein-coding sequences revealed a disproportionately long branch leading to all snakes, with complete elimination of gene-specific relative rate differences in terminal lineages. Terminal branches ultimately revealed mitochondrial genome evolution to be similar to other vertebrates, despite the initial accelerated mutation rate along early branches. Phylograms depicting squamate diversification also reveal a similar pattern for Serpentes, both for those derived from a mixture of nuclear and mitochondrial genes (Reeder et al., 2015; Zheng & Wiens, 2016), as well as additional morphological traits, as described by Reeder et al. (2015). This pattern may be due to the relatively recent diversification of the snake assemblage in our study (91 my). Alternatively, this may highlight a potential recent slowdown in substitution rates in snake EMPs.

4.2 | General patterns of EMP evolution

All three EMPs were revealed to be under moderate purifying selection in both mammals and reptiles. AMEL forms the majority of the protein in the developing enamel matrix (Termine, Belcourt, Christner, Conn, & Nylen, 1980), and stood out when selection and divergence were analyzed amongst the various clades. Between reptiles and mammals, both PAML and RELAX identified reptiles *AMEL* orthologs as exhibiting signals of lower selective constraint, which supports our initial prediction of a more lenient selection regime in reptile EMPs than mammals. AMEL evolution is known to slow as enamel complexity increases (Mathur & Polly, 2000). Therefore, a likely interpretation of our results is that the evolution of prismatic enamel in mammals has intensified selective pressure on the associated genes (viz., *AMEL*), while modern reptiles have retained simpler enamel and comparatively less stringent selection. However, this oversimplifies the matter, as we show here, when selection within reptiles is considered.

PAML analysis of *AMEL* revealed a significant signal of strong purifying selection in crocodilians when compared with squamates. RELAX, on the other hand, detected "relaxation" of selection in crocodilians, which at first seems contradictory. However, a closer examination of the RELAX output revealed that the crocodilians exhibit $\omega = 1.00$ in the third category (ω_3), indicating little to no positive selection in the group. The shift of the third category to 1.00 (in comparison to the 37.01 for squamates) fit the designation of "relaxation", despite an overall lower ω estimate for crocodilians than squamates. Indeed, the Partitioned Descriptive Model from RELAX confirms that 99.46% of sites in crocodilian AMEL are under purifying selection (Supporting Information File S6). Therefore, squamates may be overwhelmingly responsible for the higher overall ω value for *AMEL* in reptiles when compared with mammals. Interestingly, the differences in selection pressure did not lead to the generation of residues with the predicted functional divergence between mammals and reptiles, or between reptile subclades for that matter, which may reflect the conserved and important role of AMEL in amelogenesis.

AMBN and ENAM together represent the other ~9–10% of the enamel matrix (Termine et al., 1980) and we found them to be evolving under moderate purifying selection in both mammals and reptile. In fact, our ω estimate for mammal *AMBN* was very close to that of Delsuc et al. (2015), with 0.44 and 0.46, respectively. ENAM has been noted for its high degree of variation between orthologs, such as the presence/absence of exons 3 and 8b (Gasse & Sire, 2015), as well as an elevated variation in exon 10, including several insertions and deletions which seem to have no negative consequence on protein function and enamel structure (Al-Hashimi et al., 2009). Despite purifying selection, AMBN and ENAM did reveal some sites with signatures of predicted functional divergence. It is noteworthy, however, that the majority of the analyses to detect functionally divergent sites across all three EMPs yielded nonsignificant results. This likely reflects the essential role that these proteins play in reptile amelogenesis, despite polyphyodonty and nonprismatic enamel.

4.3 | Amelogenesis imperfecta

Amelogenesis imperfecta is a clinically and genetically diverse group of disorders affecting the development of enamel (Witkop, 1988). Mutations in all three EMPs have been identified as underlying causes (reviewed by Smith et al., 2017). The condition generally results in poor enamel quality, leading to enamel that is brittle and prone to wear and breakage. Polyphyodont dentition could theoretically offer some relief from its symptoms, since teeth are continuously replaced. Thus, one might expect to find putative disease-causing mutations retained in the EMPs of polyphyodont species. However, we found no such mutations in our data set. Therefore, even if selection were less stringent in reptile EMPs,

amelogenesis imperfecta-causing sites are likely still under strong purifying selection to facilitate proper enamel development.

5 | CONCLUSIONS

In this study, we have uncovered an intricate landscape of gene evolutions, selection, and functional constraint. Our results show that reptile EMPs still operate under moderate purifying selection, similar selection pressure to mammals, despite polyphyodonty and simpler enamel structure. Furthermore, while reptile EMPs seem to exhibit low sequence identity, we show here that this is limited to differences between the major reptile clades. Within crocodilians, snakes and individual lizard infraorders, we observe relatively high sequence identity that appears to be correlated to divergence times. Interestingly, we also found that reptile EMPs are not equal in their evolutionary backdrop, with *AMEL* existing under a unique selection regime compared with *AMBN* and *ENAM*. Additionally, while estimates of selection pressure on *AMEL* differed between reptiles and mammals, as well as within reptiles, we did not detect any signatures of sites exhibiting predicted functional diversification. In contrast, *AMBN* and *ENAM* evolve under more moderate selection regimes and do exhibit sites with predicted divergence.

ACKNOWLEDGMENTS

We would like to thank Prof. Peter Baumann of Johannes Gutenberg University (Mainz, Germany) for allowing us early access to *Aspidoscelis marmorata* genome. We would also like to thank Dr. Matthew Fujita of the University of Texas at Arlington and Dr. Sonal Singhal of California State University, Dominguez Hills for early access to *Lepidodactylus lugubris* genome. We would also like to thank Dr. Iva Vesela and Dr. Nicholas Borotto for critical reading of the manuscript. Lastly, we would like to thank the two anonymous reviewers for their invaluable input which substantially improved our manuscript.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

J. A. wrote the manuscript, annotated reptile EMP sequences, and carried out PAML and RELAX analyses. O. A. carried out analyses for PAML, DIVERGE and Amelogenesis imperfecta. Funding was provided by startup funds from the University of Michigan-Dearborn to J. A.

DATA AVAILABILITY

The data that support the findings of this study are included in Supporting Information File S2.

ORCID

John Abramyan http://orcid.org/0000-0001-7594-8566

REFERENCES

Al-Hashimi, N., Lafont, A. G., Delgado, S., Kawasaki, K., & Sire, J. Y. (2010). The enamelin genes in lizard, crocodile, and frog and the pseudogene in the chicken provide new insights on enamelin evolution in tetrapods. *Molecular Biology and Evolution*, *27*(9), 2078–2094. https://doi.org/10.1093/molbev/msq098

Al-Hashimi, N., Sire, J. Y., & Delgado, S. (2009). Evolutionary analysis of mammalian enamelin, the largest enamel protein, supports a crucial role for the 32-kDa peptide and reveals selective adaptation in rodents and primates. *Journal of Molecular Evolution*, *69*(6), 635–656. https://doi.org/10.1007/s00239-009-9302-x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Evolution*, *215*(3), 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2

Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution*, *18*(8), 1585–1592. https://doi.org/10.1093/oxfordjournals.molbev.a003945

Assaraf-Weill, N., Gasse, B., Al-Hashimi, N., Delgado, S., Sire, J. Y., & Davit-Beal, T. (2013). Conservation of amelogenin gene expression during tetrapod evolution. *Journal of Experimental Biology Part B: Molecular and Developmental Evolution*, *320*(4), 200–209. https://doi.org/10.1002/jez.b.22494

Bawono, P., Dijkstra, M., Pirovano, W., Feenstra, A., Abeln, S., & Heringa, J. (2017). Multiple sequence alignment. *Methods in Molecular Biology*, *1525*, 167–189. doi:10.1007/978-1-4939-6622-6_8

Bertin, T. J. C., Thivichon-Prince, B., LeBlanc, A. R. H., Caldwell, M. W., & Viriot, L. (2018). Current perspectives on tooth implantation, attachment, and replacement in Amniota. *Frontiers in Physiology*, *9*, 1630. https://doi.org/10.3389/fphys.2018.01630

Bluteau, G., Luder, H. U., De Bari, C., & Mitsiadis, T. A. (2008). Stem cells for tooth engineering. *European Cells & Materials Journal*, *16*, 1–9.

Chang, J. M., Di Tommaso, P., & Notredame, C. (2014). TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Molecular Biology and Evolution*, *31*(6), 1625–1637. https://doi.org/10.1093/molbev/msu117

Cooper, J. S., & Poole, D. F. G. (1973). The dentition and dental tissues of the agamid lizard, Uromastyx. *Journal of Zoology*, *169*(1), 85–100.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Davit-Beal, T., Chisaka, H., Delgado, S., & Sire, J. Y. (2007). Amphibian teeth: Current knowledge, unanswered questions, and some directions for future research. *Biological Reviews of the Cambridge Philosophical Society*, *82*(1), 49–81. https://doi.org/10.1111/j.1469-185X.2006.00003.x

Delgado, S., Casane, D., Bonnaud, L., Laurin, M., Sire, J. Y., & Girondot, M. (2001). Molecular evidence for precambrian origin of amelogenin, the major protein of vertebrate enamel. *Molecular Biology and Evolution*, *18*(12), 2146–2153. https://doi.org/10.1093/oxfordjournals.molbev.a003760

Delgado, S., Couble, M. L., Magloire, H., & Sire, J. Y. (2006). Cloning, sequencing, and expression of the amelogenin gene in two scincid lizards. *Journal of Dental Research*, *85*(2), 138–143. https://doi.org/10.1177/154405910608500205

Delgado, S., Davit-Beal, T., Allizard, F., & Sire, J. Y. (2005). Tooth development in a scincid lizard, Chalcides viridanus (Squamata), with particular attention to enamel formation. *Cell and Tissue Research*, *319*(1), 71–89. https://doi.org/10.1007/s00441-004-0950-2

Delgado, S., Ishiyama, M., & Sire, J. Y. (2007). Validation of amelogenesis imperfecta inferred from amelogenin evolution. *Journal of Dental Research*, *86*(4), 326–330. https://doi.org/10.1177/154405910708600405

Delsuc, F., Gasse, B., & Sire, J. Y. (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

Diekwisch, T. G., Berman, B. J., Anderton, X., Gurinsky, B., Ortega, A. J., Satchell, P. G., & Shuler, C. F. (2002). Membranes, minerals, and proteins of developing vertebrate enamel. *Microscopy Research and Technique*, *59*(5), 373–395. https://doi.org/10.1002/jemt.10218

Diekwisch, T. G., Jin, T., Wang, X., Ito, Y., Schmidt, M., Druzinsky, R., & Luan, X. (2009). Amelogenin evolution and tetrapod enamel structure. *Frontiers of Oral Biology*, *13*, 74–79. https://doi.org/10.1159/000242395

Durand, P. M., Hazelhurst, S., & Coetzer, T. L. (2010). Evolutionary rates at codon sites may be used to align sequences and infer protein domain function. *BMC Bioinformatics*, *11*, 151. https://doi.org/10.1186/1471-2105-11-151

Edmund, A. G. (1960). *Tooth replacement phenomena in the lower vertebrates* Vol.52). Toronto, Ontario, Canada: University of Toronto Press.

Eo, S. H., & DeWoody, J. A. (2010). Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society B: Biological Sciences*, *1700*, 3587–3592. https://doi.org/10.1098/rspb.2010.0965

Fincham, A. G., Moradian-Oldak, J., & Simmer, J. P. (1999). The structural biology of the developing dental enamel matrix. *Journal of Structural Biology*, *126*(3), 270–299. https://doi.org/10.1006/jsbi.1999.4130

Fukumoto, S., Kiba, T., Hall, B., Iehara, N., Nakamura, T., Longenecker, G., & Yamada, Y. (2004). Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts. *Journal of Cell Biology*, *167*(5), 973–983. https://doi.org/10.1083/jcb.200409077

Gasse, B., & Sire, J. Y. (2015). Comparative expression of the four enamel matrix protein genes, amelogenin, ameloblastin, enamelin and amelotin during amelogenesis in the lizard *Anolis carolinensis*. *EvoDevo*, *6*, 29. https://doi.org/10.1186/s13227-015-0024-4

Gojobori, T. (1983). Codon substitution in evolution and the "saturation" of synonymous changes. *Genetics*, *105*(4), 1011–1027.

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *5*, 725–736.

Green, R. E., Braun, E. L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., & Ray, D. A. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*, *346*(6215), 1254449. https://doi.org/10.1126/science.1254449

Grine, F. E., Vrba, E. S., & Cruickshank, A. R. I. (1979). Enamel prisms and diphyodonty: Linked apomorphies of Mammalia. *South African Journal of Science*, *75*, 114–120.

Gu, X. (1999). Statistical methods for testing functional divergence after gene duplication. *Molecular Biology and Evolution*, *16*(12), 1664–1674. https://doi.org/10.1093/oxfordjournals.molbev.a026080

Gu, X. (2001). Maximum-likelihood approach for gene family evolution under functional divergence. *Molecular Biology and Evolution*, *18*(4), 453–464. https://doi.org/10.1093/oxfordjournals.molbev.a003824

Gu, X. (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Molecular Biology and Evolution*, *23*(10), 1937–1945. https://doi.org/10.1093/molbev/msl056

Gu, X., & Vander Velden, K. (2002). DIVERGE: Phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics*, *18*(3), 500–501.

Hu, C. C., Fukae, M., Uchida, T., Qian, Q., Zhang, C. H., Ryu, O. H., & Simmer, J. P. (1997). Sheathlin: Cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel sheath proteins. *Journal of Dental Research*, *76*(2), 648–657. https://doi.org/10.1177/00220345970760020501

Jiang, Z. J., Castoe, T. A., Austin, C. C., Burbrink, F. T., Herron, M. D., McGuire, J. A., & Pollock, D. D. (2007). Comparative mitochondrial genomics of snakes: Extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evolutionary Biology*, *7*, 123. https://doi.org/10.1186/1471-2148-7-123

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275–282.

Jussila, M., & Thesleff, I. (2012). Signaling networks regulating tooth organogenesis and regeneration, and the specification of dental mesenchymal and epithelial cell lineages. *Cold Spring Harbor Perspectives in Biology*, *4*(4), a008425. https://doi.org/10.1101/cshperspect.a008425

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066.

Kawasaki, K., & Weiss, K. M. (2003). Mineralized tissue and vertebrate evolution: The secretory calcium-binding phosphoprotein gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4060–4065. https://doi.org/10.1073/pnas.0638023100

Lemey, P., & Posada, D. (2009). Molecular clock analysis. In M. Salemi, A. M. Vandamme, & P. Lemey (Eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing* (pp. 362–380). Cambridge University Press.

Line, S. R. P., & Novaes, P. D. (2005). The development and evolution of mammalian enamel: Structural and functional aspects. *Journal of Morphological Science*, *22*(1), 67–72.

Man, Z., Yishu, W., Peng, Y., & Xiaobing, W. (2011). Crocodilian phylogeny inferred from twelve mitochondrial protein-coding genes, with new complete mitochondrial genomic sequences for *Crocodylus acutus* and *Crocodylus novaeguineae*. *Molecular Phylogenetics and Evolution*, *60*(1), 62–67. https://doi.org/10.1016/j.ympev.2011.03.029

Manzoor, U., Shahid, S., & Zafar, B. (2015). A comparative analysis of multiple sequence alignments for biological data. *Biomedical Materials and Engineering*, *26*(Suppl 1), S1781–S1789. https://doi.org/10.3233/bme-151479

Mathur, A. K., & Polly, P. D. (2000). The evolution of enamel microstructure: How important is amelogenin? *Journal of Mammalian Evolution*, *7*(1), 23–42.

Meredith, R. W., Gatesy, J., Murphy, W. J., Ryder, O. A., & Springer, M. S. (2009). Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genetics*, *5*(9), e1000634. https://doi.org/10.1371/journal.pgen.1000634

Meredith, R. W., Gatesy, J., & Springer, M. S. (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

Moradian-Oldak, J., & Goldberg, M. (2005). Amelogenin supra-molecular assembly in vitro compared with the architecture of the forming enamel matrix. *Cells Tissues Organs*, *181*(3-4), 202–218. https://doi.org/10.1159/000091382

Moradian-Oldak, J., Iijima, M., Bouropoulos, N., & Wen, H. B. (2003). Assembly of amelogenin proteolytic products and control of octacalcium phosphate crystal morphology. *Connective Tissue Research*, *44*(Suppl 1), 58–64.

Moss-Salentijn, L., Moss, M. L., & Yuan, M. S. T. (1997). The ontogeny of mammalian enamel. In W. V. Koenigswald, & P. M. Sander (Eds.), *Tooth Enamel Microstructure* (pp. 5–30). Rotterdam: Balkema & Brookfield.

Nanci, A., Zalzal, S., Lavoie, P., Kunikata, M., Chen, W., Krebsbach, P. H., & Smith, C. E. (1998). Comparative immunochemical analyses of the developmental expression and distribution of ameloblastin and amelogenin in rat incisors. *Journal of Histochemistry & Cytochemistry*, *46*(8), 911–934. https://doi.org/10.1177/002215549804600806

Nuin, P. A., Wang, Z., & Tillier, E. R. (2006). The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics*, *7*, 471. https://doi.org/10.1186/1471-2105-7-471

Pais, F. S., Ruy, P. C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology: AMB*, *9*(1), 4. https://doi.org/10.1186/1748-7188-9-4

Pond, S. L., Frost, S. D., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*(5), 676–679. https://doi.org/10.1093/bioinformatics/bti079

Ravindranath, H. H., Chen, L. S., Zeichner-David, M., Ishima, R., & Ravindranath, R. M. (2004). Interaction between the enamel matrix proteins amelogenin and ameloblastin. *Biochemical and Biophysical Research Communications*, *323*(3), 1075–1083. https://doi.org/10.1016/j.bbrc.2004.08.207

Reeder, T. W., Townsend, T. M., Mulcahy, D. G., Noonan, B. P., Wood, P. L., Jr., Sites, J. W., Jr., & Wiens, J. J. (2015). Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLOS One*, *10*(3), e0118199. https://doi.org/10.1371/journal.pone.0118199

Richard, B., Delgado, S., Gorry, P., & Sire, J. Y. (2007). A study of polymorphism in human AMELX. *Archives of Oral Biology*, *52*(11), 1026–1031. https://doi.org/10.1016/j.archoralbio.2007.06.001

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574.

Satchell, P. G., Anderton, X., Ryu, O. H., Luan, X., Ortega, A. J., Opamen, R., & Diekwisch, T. G. (2002). Conservation and variation in enamel protein distribution during vertebrate tooth development. *Journal of Experimental Zoology*, *294*(2), 91–106. https://doi.org/10.1002/jez.10148

Shaffer, H. B., Minx, P., Warren, D. E., Shedlock, A. M., Thomson, R. C., Valenzuela, N., & Wilson, R. K. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, *14*(3), R28. https://doi.org/10.1186/gb-2013-14-3-r28

Sire, J. Y., Davit-Beal, T., Delgado, S., & Gu, X. (2007). The origin and evolution of enamel mineralization genes. *Cells Tissues Organs*, *186*(1), 25–48. https://doi.org/10.1159/000102679

Sire, J. Y., Delgado, S., Fromentin, D., & Girondot, M. (2005). Amelogenin: Lessons from evolution. *Archives of Oral Biology*, *50*(2), 205–212. https://doi.org/10.1016/j.archoralbio.2004.09.004

Sire, J. Y., Delgado, S., & Girondot, M. (2006). The amelogenin story: Origin and evolution. *European Journal of Oral Sciences*, *114*(Suppl 1), 64–77. discussion 93-65, 379-380. https://doi.org/10.1111/j.1600-0722.2006.00297.x

Smith, C. E. L., Poulter, J. A., Antanaviciute, A., Kirkham, J., Brookes, S. J., Inglehearn, C. F., & Mighell, A. J. (2017). Amelogenesis imperfecta; Genes, proteins, and pathways. *Frontiers in Physiology*, *8*, 435. https://doi.org/10.3389/fphys.2017.00435

Sonoda, A., Iwamoto, T., Nakamura, T., Fukumoto, E., Yoshizaki, K., Yamada, A., & Fukumoto, S. (2009). Critical role of heparin binding domains of ameloblastin for dental epithelium cell adhesion and ameloblastoma proliferation. *Journal of Biological Chemistry*, *284*(40), 27176–27184. https://doi.org/10.1074/jbc.M109.033464

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(Web Server issue), W609–W612. https://doi.org/10.1093/nar/gkl315

Termine, J. D., Belcourt, A. B., Christner, P. J., Conn, K. M., & Nylen, M. U. (1980). Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel. *Journal of Biological Chemistry*, *255*(20), 9760–9768.

Thompson, J. D., Linard, B., Lecompte, O., & Poch, O. (2011). A comprehensive benchmark study of multiple sequence alignment methods: Current challenges and future perspectives. *PLOS One*, *6*(3), e18093. https://doi.org/10.1371/journal.pone.0018093

Throckmorton, G. S. (1979). The effect of wear on the cheek teeth and associated dental tissues of the lizard Uromastix aegyptius (Agamidae). *Journal of Morphology*, *160*(2), 195–208. https://doi.org/10.1002/jmor.1051600206

Tucker, A. S., & Fraser, G. J. (2014). Evolution and developmental diversity of tooth regeneration. *Seminars in Cell and Developmental Biology*, *25-26*, 71–80. https://doi.org/10.1016/j.semcdb.2013.12.013

Vidal, N., & Hedges, S. B. (2009). The molecular evolutionary tree of lizards, snakes, and amphisbaenians. *Comptes Rendus Biologies*, *332*(2-3), 129–139. https://doi.org/10.1016/j.crvi.2008.07.010

Wang, L. S., Leebens-Mack, J., Kerr Wall, P., Beckmann, K., dePamphilis, C. W., & Warnow, T. (2011). The impact of multiple protein sequence alignment on phylogenetic estimation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(4), 1108–1119. https://doi.org/10.1109/tcbb.2009.68

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, *32*(3), 820–832. https://doi.org/10.1093/molbev/msu400

Wiens, J. J., Hutter, C. R., Mulcahy, D. G., Noonan, B. P., Townsend, T. M., Sites, J. W., Jr., & Reeder, T. W. (2012). Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biology Letters*, *8*(6), 1043–1046. https://doi.org/10.1098/rsbl.2012.0703

Witkop, C. J., Jr. (1988). Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: Problems in classification. *Journal of Oral Pathology & Medicine*, *17*(9-10), 547–553.

Xia, X., & Lemey, P. (2009). Assessing substitution saturation with DAMBE. In S. Marco, & A. Vandamme (Eds.), *The phylogenetic handbook: A practical approach to DNA and protein phylogeny*. Cambridge: Cambridge University Press.

Xia, X., & Xie, Z. (2001). DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity*, *92*(4), 371–373.

Xia, X., Zheng, X., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, *26*(1), 1–7.

Yang, J., & Warnow, T. (2011). Fast and accurate methods for phylogenomic analyses. *BMC Bioinformatics*, *12*(Suppl 9), S4. https://doi.org/10.1186/1471-2105-12-s9-s4

Yang, Z. (1998). On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, *47*(1), 125–133.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yoder, A. D., & Yang, Z. (2000). Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, *17*(7), 1081–1090. https://doi.org/10.1093/oxfordjournals.molbev.a026389

Zheng, Y., & Wiens, J. J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, *94*, 537–547.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Alazem O, Abramyan J. Reptile enamel matrix proteins: Selection, divergence, and functional constraint. *J Exp Zool (Mol Dev Evol)*. 2019;332:136–148. https://doi.org/10.1002/jez.b.22857