UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Evolution of Tooth Genes in the Degeneration and Loss of Enamel and Teeth in Whales and Pangolins

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Evolution, Ecology, and Organismal Biology

by

Jason Gerard Randall

December 2023

Dissertation Committee:
Dr. Mark S. Springer, Chairperson
Dr. David N. Reznick
Dr. John M. Heraty

The Dissertation of Jason Gerard Randall is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

There are many different people that I would like to thank that have helped me throughout my journey as a student from the initial onset of academia to the final steps of defense and submission of this dissertation. I must first thank my late grandfather Dr. Norman Weinberger for being an inspiration and always encouraging me to pursue further education and an advanced college degree as without him I probably would not be where I am today. I also must thank my parents for the opportunities they have given me in encouraging and allowing me to pursue education from the very beginning all the way to graduation at every level. Along with these pillars of inspiration are all my other family members that have greatly supported me as I have made my way through this academic adventure. I must also thank my future family starting with my fiancée Zoila who has been an incredible source of inspiration to myself, has helped me stay focused and see this degree to the end through hard times, and has had immense patience as I have worked on this degree and truly wish I had more time to spend together through this process (thank you so much for coming to see my defense albeit without my knowledge, and I wanted to thank you here and not last after the non-human animal pets 😊).

I am also very appreciative of my advisor Dr. Mark S. Springer for first believing in me and allowing me to join his lab and giving me such great ideas of research that I would be able to pursue and eventually turn into my dissertation. I must also thank his unbelievable patience throughout this long journey and incredibly helpful advice through all the stages involved with the eventual defense and completion of this dissertation. We have been able to publish several research papers that I would not have been able to do

without his brilliant mind and incredible mentorship. His upbeat demeaner and encouragement through collaboration with the other lab members always made lab meetings and discussions a pleasure to be a part of.

Other faculty members at UCR have also been incredible to work with that have made my time there a joyous experience. My dissertation committee member David Reznick was a pleasure to work with and I feel very lucky to have met him after originally learning about him by reading his interpretation of the Origin of Species as an undergraduate. The other committee member of my dissertation John Heraty was also great to work with and learn from, his enthusiasm for the underappreciated discipline of systematics made studying it much more exciting and interesting (even with his focus on insects when we know mammals are superior, just kidding!) I also want to thank academic coordinators Katie Burnette and Mike Fugate for developing and strengthening my love for teaching science to students, a passion I did not think I had previously. Lastly in conjunction with the amazing people I have had to work with at UCR, I must thank funding sources that made this work possible, with the receipt of the GAANN fellowship from the U.S. Department of Education and in part an NSF grant (DEB-1457735) utilized in the research contributing to several published papers.

In addition, I need to thank my extended support system of all the friends that I have had the pleasure to meet on this journey that I hope to stay in contact with for years to come. My immediate lab members that I have known the longest, Peggy and Sam, it was a joy to have you join the lab when I was first there by myself and have helped encourage me through the struggles of hard times and helped in so many ways in

accomplishing the research needed for this dissertation. Many of the other friends that I have met in my EEOB cohort have also been an incredible source of inspiration. Of all of these friends both Ryan and Robert have been especially helpful in providing support to accomplish and complete this PhD degree and have made my time at UCR an amazing experience. Lastly, the lineage of Trinidadian guppies still going strong provided by Robert and my cat Lillith have been a source of enjoyment and calming presence in my life that helped me stay focused and determined.

ABSTRACT OF THE DISSERTATION

Molecular Evolution of Tooth Genes in the Degeneration and Loss of Enamel and Teeth in Whales and Pangolins

by

Jason Gerard Randall

Doctor of Philosophy, Graduate Program in Evolution, Ecology, and Organismal Biology
University of California, Riverside, December 2023
Dr. Mark S. Springer, Chairperson

Changes in dental morphology have profound consequences on the ecology of species and the niches they are adapted to. Here, changes in teeth, specifically in the main components of enamel and dentin, were investigated. Genes encoding the proteins that produce these structures were examined for mutations and utilized in selection analyses to determine the rates of nonsynonymous versus synonymous substitutions via dN/dS ratios. These values were then used to calculate the timing of inactivation of those genes and by proxy their phenotypes to determine if these phenotypes were linked or decoupled in the evolutionary history of toothless baleen whales (Mysticeti) and pangolins (Pholidota). In addition, the enamel genes for toothed whales (Odontoceti) were investigated for mutations and selection patterns given their degenerative enamel. In chapter 1, mysticetes were discovered to possess numerous inactivating mutations in enamel and dentin/tooth-specific genes, many of which were shared among major clades. The elevated dN/dS values suggest that these genes may have been relaxed in the

common ancestor of Mysticeti. Inactivation times estimated that enamel was lost before dentin/teeth, the latter of which was lost at most two times, independently in the common ancestor of Balaenidae and Plicogulae. The results from Chapter 2 reveal numerous mutations and elevated dN/dS values among enamel genes in many odontocetes. More mutations and increased relaxed/positive selection correlated to taxa with less complex enamel, with a strong correlation of decreasing dN/dS values associated with increasing enamel complexity. Chapter 3 described inactivating mutations among the tooth genes in three living species pangolins, with many of the genes containing shared mutations in all three taxa. Selection intensity was observed to be elevated and relaxed compared to outgroups. Calculations for the inactivation of genes and phenotypes reveal that enamel was lost around 62 million years ago, but unable to retrieve an inactivation date for dentin/teeth. Overall, these chapters have demonstrated the underlying genetic components necessary for enamel and dentin/tooth production have become pseudogenized molecular relics riddled with mutations in edentulous taxa and have started to build up in taxa with degenerative enamel.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Chapter 1

Chapter 2

Chapter 3

Prologue

Macroevolutionary events are major transitions in the evolutionary history of organisms that have spurred the curiosity of biologists and philosophers for centuries. Although not explicitly stated as a macroevolutionary transition, the disparate connection of different groups of organisms was thought about as far back as ancient Greece, with Aristotle's discussion on cetaceans including that they are viviparous, have hair, and breath through blowholes using lungs (*Historia animalium*, *History of Animals*). Cetaceans were eventually classified as mammals by Carl Linnaeus (*Systema Naturae*, 1758), and although even Darwin knew about other major transitions in whales such as the presence of tooth buds in fetal baleen whales, he struggled with the concept of a quadruped terrestrial ancestry of whales. However, Darwin did know about transitions occurring in other groups of organisms, such as the discovery of the "lizard-like" bird *Archaeopteryx*, as well as rudimentary organs such as the pelvis and hind limbs in snakes (Darwin, 1872). Rudimentary and vestigial traits such as these bones in snakes and tooth buds in baleen whales brought up by Darwin has become a tremendously important and useful tool for tracking the evolution of major transformations in organisms and grouping species together that were previously thought to be unrelated. This concept is succinctly captured by Ernst Haeckel in 1866 as "ontogeny recapitulates phylogeny".

In the century and a half since then, research has investigated the adaptive purposes of vestigialization and underlying mechanisms that drive regressive evolution and loss of traits. Only somewhat recently, with the prevalence of DNA sequencing and the acquisition of molecular genomes from abundant numbers of species, has our

understanding of these mechanisms truly come to fruition. However, those mechanisms are complex and not straightforward. There are different forces at work that lead to the degeneration or complete loss of vestigial traits that were once useful to an organism. Only once genetic loci and specific genes that code for these regressive traits were identified could the underlying molecular processes be elucidated. There may be selection acting on these genes such as direct positive selection against a trait to minimize energetic costs in favor of more useful traits or relaxed selection leading to neutral evolution of a trait that is no longer of adaptive use to an organism in their current environment; both types of selection lead to a decrease in these phenotypes due to not positively affecting the fitness of these species (Moran et al., 2023). Other mechanisms involved in producing rudimentary organs that have been suggested includes genetic drift, pleiotropy, and genetic hitchhiking (Espinasa and Espinasa, 2008; Porter and Crandall, 2003). However, one of the most common forces is relaxed selection occurring on genes that have become nonfunctional that mostly originate from gene duplication events (Zhang, 2003). Although gene duplication events can occur and result in many forms, instances where an entire gene is duplicated and remains functional allows for additional copies to evolve new functions or become pseudogenized. This has been the case for the Secretory Calcium binding Phosphoprotein (SCPP) family of genes that arose through iterative gene duplication and produced many of the enamel and dentin/tooth-specific genes investigated throughout this dissertation (Kawasaki and Weiss, 2008).

The origin of this gene family paved the way for the evolution of teeth which was a hallmark macroevolutionary event that originated in stem gnathostomes approximately

450 million years ago (Davit-Béal et al., 2009). Dentition aided these early jawed vertebrates in their survival and eventual extensive speciation and radiation as they adapted to a predatory lifestyle (Davit-Béal et al., 2009). Compared to the simplified aprismatic enamel of amphibians and most reptiles, mammals developed dental occlusion and intricate enamel that allowed for more efficient processing of food and therefore increased intake of nutrients (Line and Novaes 2005). This better use of mastication and energy was a key step in the evolution of mammals that fueled their high metabolism and endothermy which further promoted diversification (Werth et al., 2020). In turn, increased radiation allowed mammals to occupy underutilized niches that involved adaptation to a harder diet which necessarily included increased longevity of teeth with robust and complex enamel microstructure. Interestingly, there have been numerous subsequent instances of convergent regression of dentition including loss of complex enamel as well as enamelless and edentulous phenotypes. Examples of these phenotypes include enamel and tooth reduction in armadillos and platypuses, enamel loss in aardvarks, sloths, and some odontocetes, and complete tooth loss in toads, birds, turtles, echidnas, anteaters, pangolins, and baleen whales (Davit-Béal et al., 2009)

The variation associated with the numerous instances of dental regression in disparate taxa has led researchers to investigate the causes for the degeneration and loss of this significantly adaptive trait that assisted in their successful diversification and radiation. Questions investigated include what level of convergence has occurred to cause these similar phenotypes, such as similar dietary shifts or convergence at the molecular level through substitutions at specific amino acid sites (Mu et al., 2021). Additional

research has examined the amount of time in which tooth regression could occur, taking place over a short period of time related to direct (positive) selection against costly phenotypes or a gradual, longer process associated with relaxed selection and genetic drift (Emerling et al., 2023). In order to more fully understand these nuances, the timing of tooth degeneration and loss has been examined in two of the following chapters (Ch.1 and 3).

Study system and outline

The increase in the number and availability of genomes being released from divergent taxa allows for the detailed investigation into the causes for the regressive evolution of vestigial phenotypes, like the loss of teeth. This is especially the case when there is a sparse fossil record, like for pangolins, where the available genetic data of the underlying genes responsible for these phenotypes can be examined and may provide insights into the circumstances involved in the evolutionary adaptation to these new specialized ecological niches. The following research investigates the degeneration of enamel and teeth in three different clades of placental mammals using recently acquired genome data for new species that have not been examined to date. Taxa explored includes species belonging to two clades in the order Cetartiodactyla (Mysticeti and Odontoceti cetaceans), and the sole members of the order Pholidota (pangolins). Pangolins and baleen whales are the only other members of the superorder Laurasiatheria that have lost their teeth.

The first two chapters investigate tooth loss and enamel degeneration in baleen whales (Mysticeti) and toothed whales (Odontoceti), respectively. Cetaceans are an

appropriately utilized “poster child” for demonstrating one of the best-known cases of a macroevolutionary event in a group of organisms that people know about around the world and have known for centuries. One of the large-scale changes that occurred in the evolution of Mysticeti included the loss of teeth in favor of baleen racks they use to filter feed. Although the transition from a toothed ancestor was known from at least the time Darwin wrote about it, only recently with the ability to obtain DNA sequences for specific genes responsible for enamel and dentin/tooth production has the ability to understand how and when the process of tooth loss occurred in the common ancestor of these animals. Specifically, these genes were investigated for inactivating mutations with an interest on taxa that share such mutations to confirm that the loss of these genes (and encoded proteins and phenotypes) occurred together in the common ancestor and not independently through convergence in separate lineages. Enamel and dentin/tooth-specific genes were examined independently to observe if there was an enamelless condition of teeth that occurred before total tooth loss. This is useful because teeth are not usually found with fossils, which instead are inferred present through different methods and cannot determine if the teeth had enamel or not. This was further corroborated using methods to calculate an approximate timing of enamel and dentin/tooth loss utilizing results from selection intensity analyses (dN/dS). For the toothed whales, unlike baleen whales, they still possess a component of their original dentition. However, many different clades of odontocetes exhibit varying degrees of enamel reduction, ranging from thin, weak enamel that wears away quickly to taxa that are enamelless. Due to these observations, a detailed investigation of their enamel-specific genes was conducted. In

addition to examining these genes for autapomorphic and synapomorphic inactivating mutations, selection analyses were conducted on these genes to determine which taxa contained genes that have been relaxed and released from selective pressures. The results from these selection analyses were compared against enamel complexity that was assigned to taxa via discrete categories that utilized a system of enamel organization that were scored in Werth et al. (2020).

Lastly, pangolins (Pholidota) are another clade of toothless organisms where the history of tooth loss has yet to be fully elucidated. With a sparse fossil record, there are currently no instances of a fossil pangolin with teeth that has been found. This complicates the understanding of tooth loss in the evolution of ancient pangolins of how and when they were lost in their common ancestor. There has so far not been a comprehensive investigation into the molecular structure of tooth genes in pangolins, which this third chapter sets out to accomplish. This research is conducted by utilizing recently released genome sequences for three extant taxa that have not been investigated together in this manner. The coding sequences of enamel and dentin/tooth-specific genes were surveyed for inactivating mutations as well as used for selection analyses. The results of selection intensity were then used in calculations to determine when those genes and by proxy phenotypes were lost.

By developing a better understanding on the genetic mechanisms underlying tooth regression and loss, these studies aim to contribute to the broader understanding of trait reduction or loss among disparate taxonomic groups and how different species may continue to change overtime.

Chapter 1: Molecular Evolutionary Analyses of Tooth Genes Support Sequential Loss of Enamel and Teeth in Baleen Whales (Mysticeti)

Abstract

The loss of teeth and evolution of baleen racks in Mysticeti was a profound transformation that permitted baleen whales to radiate and diversify into a previously underutilized ecological niche of bulk filter-feeding on zooplankton and other small prey. Ancestral state reconstructions suggest that postnatal teeth were lost in the common ancestor of crown Mysticeti. Genomic studies provide some support for this hypothesis and suggest that the genetic toolkit for enamel production was inactivated in the common ancestor of living baleen whales. However, molecular studies to date have not provided direct evidence for the complete loss of teeth, including their dentin component, on the stem mysticete branch. Given these results, several questions remain unanswered: (1) Were teeth lost in a single step or did enamel loss precede dentin loss? (2) Was enamel lost early or late on the stem mysticete branch? (3) If enamel and dentin/tooth loss were decoupled in the ancestry of baleen whales, did dentin loss occur on the stem mysticete branch or independently in different crown mysticete lineages? To address these outstanding questions, we compiled and analyzed complete protein-coding sequences for nine tooth-related genes from cetaceans with available genome data. Seven of these genes are associated with enamel formation (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) whereas two other genes are either dentin-specific (*DSPP*) or tooth-specific (*ODAPH*) but not enamel-specific. Molecular evolutionary analyses indicate that all seven enamel-specific genes have inactivating mutations that are scattered across

branches of the mysticete tree. Three of the enamel genes (*ACP4*, *KLK4*, *MMP20*) have inactivating mutations that are shared by all mysticetes. The two genes that are dentin-specific (*DSPP*) or tooth-specific (*ODAPH*) do not have any inactivating mutations that are shared by all mysticetes, but there are shared mutations in Balaenidae as well as in Plicogulae (Neobalaenidae + Balaenopteroidea). These shared mutations suggest that teeth were lost at most two times. Shared inactivating mutations and dN/dS analyses, in combination with cetacean divergence times, were used to estimate inactivation times of genes and by proxy enamel and tooth phenotypes at ancestral nodes. The results of these analyses are most compatible with a two-step model for the loss of teeth in the ancestry of living baleen whales: enamel was lost very early on the stem Mysticeti branch followed by the independent loss of dentin (and teeth) in the common ancestors of Balaenidae and Plicogulae, respectively. These results imply that some stem mysticetes, and even early crown mysticetes, may have had vestigial teeth comprised of dentin with no enamel. Our results also demonstrate that all odontocete species (in our study) with absent or degenerative enamel have inactivating mutations in one or more of their enamel genes.

1.1. Introduction

Cetaceans are a diverse group of fully aquatic mammals that exhibit a variety of morphological, physiological, and behavioral specializations that evolved in conjunction with their invasion and conquest of aquatic habitats. This clade includes both the largest known vertebrate (*Balaenoptera musculus* [blue whale]) and the mammal with the longest lifespan (*Balaena mysticetus* [bowhead whale]) (Gatesy et al., 2013; Keane et al.,

2015). Aside from their remarkable phenotypic diversity, cetaceans have a remarkably well-documented macroevolutionary history that is illuminated by both fossils and genomes (Bajpai et al., 2009; Berta et al., 2016; Gatesy et al., 2013; McGowen et al., 2014, 2020b; Thewissen et al., 2009). Recent decades of paleontological research have yielded extinct species that document the transition from life on land to the aquatic realm (Gatesy et al., 2013; Gingerich, 2012; Thewissen and Bajpai, 2001; Thewissen et al., 2001, 2009). These fossils record the evolution of key morphological features of the cetacean body plan including development of paddle-shaped forelimbs, 'telescoping' of the skull, and reduction of the hindlimbs (Bejder and Hall, 2002; Gatesy et al., 2013). At the genomic level, molecular evolutionary studies provide evidence for both positive selection and extensive gene loss in association with the transition from land to water. These changes are linked to a variety of anatomical structures and organ systems that were modified in the ancestry of Cetacea including the skin, limbs, lungs, pineal gland, brain, brown fat, eyes, ears, vomeronasal organ, and nose (Emerling et al., 2021; Gatesy et al., 2013; Gaudry et al., 2017; Huelsmann et al., 2019; McGowen et al., 2014, 2020b; Nery et al., 2013; Sharma et al., 2018; Springer and Gatesy, 2018; Springer et al., 2021; Thermudo et al., 2020).

During the late Eocene (~36–37 million years ago [Ma]) there was a cladogenic split in the last common ancestor of Neoceti (crown Cetacea) that resulted in the reciprocally monophyletic clades Odontoceti (toothed whales) and Mysticeti (baleen whales) (Gatesy et al. 2013; McGowen et al., 2009, 2020a). Whereas most odontocetes still possess enamel-capped teeth like their terrestrial ancestors, extant mysticetes have

lost their teeth and instead feed using baleen racks that consist of keratin (Uhen, 2010). The evolution of baleen was associated with a profound dietary transformation that allowed mysticetes to exploit a previously underutilized food resource, zooplankton and other tiny prey, by bulk filter-feeding instead of raptorial or suction feeding on individual prey as in stem cetaceans and odontocetes.

Diverse data support the hypothesis that baleen whales evolved from toothed ancestors: (1) cladistic analyses of phenotypic data and ancestral state reconstructions infer the evolution of edentulous mysticetes from stem mysticete ancestors that possessed teeth (Fitzgerald, 2006, 2009; Gatesy et al., 2013; Meredith et al., 2011; Uhen, 2010), (2) molecular relics of both enamel-related and dentin/tooth-related tooth genes in the genomes of extant mysticetes are remnants from a toothed ancestry (Berta et al., 2016; Deméré et al., 2008; Gatesy et al., 2022; Kawasaki et al., 2014, 2020; Meredith et al., 2009, 2011a; Mu et al., 2021; Springer et al., 2016a, 2019), and (3) the initiation of tooth formation in whale fetuses (i.e., tooth germs that are sometimes mineralized and resorbed before birth) provides ontogenetic evidence for a toothed ancestry (Deméré et al., 2008; Dissel-Scherft and Vervoort, 1954; Lanzetti, 2019; Lanzetti et al., 2019; Thewissen and Williams, 2002; Thewissen et al., 2017). Note that our use of 'teeth' in the remainder of this paper refers to postnatal teeth and does not refer to tooth germs that occur in mysticete fetuses.

However, what has been less clear is how and when postnatal teeth were lost, and if tooth loss occurred before or after the evolution of baleen. The co-occurrence hypothesis suggests that baleen evolved before tooth loss and that teeth and baleen

functioned together before teeth were subsequently lost in the ancestry of extant mysticetes (Boessenecker and Fordyce, 2015; Deméré et al., 2008; Ekdale and Deméré, 2021). A variation of this hypothesis is the dentral filtration hypothesis, which contends that early mysticetes used their teeth to filter feed before developing baleen for this purpose and subsequently losing their teeth (Geisler et al., 2017; also see Hocking et al. [2017] for opposing view). By contrast with these two hypotheses wherein baleen evolved before functional teeth were lost, the toothless suction-feeding hypothesis postulates that stem mysticetes lost their teeth and were edentulous suction feeders prior to the evolution of baleen (Peredo et al., 2017, 2018).

Deméré et al. (2008) advocated for the co-occurrence hypothesis based on the presence of both teeth and inferred baleen (medial to the teeth) in *Aetiocetus weltoni* and two additional Oligocene mysticete species. Baleen was hypothesized based on the presence of lateral palatal foramina and sulci that are used to supply blood and nerves to the ever-growing baleen racks in extant baleen whales (Deméré et al., 2008). If the co-occurrence hypothesis is correct, the lateral palatal foramina in *A. weltoni* should connect internally to the superior alveolar canal, which transmits blood vessels and nerves to the maxillary teeth in odontocetes and to the baleen racks in extant mysticetes (Ekdale and Deméré, 2021). Ekdale and Deméré (2021) used CT scan images of *A. weltoni* to show that this early mysticete exhibits a condition that is intermediate between a representative odontocete (*Tursiops*) and a representative mysticete (*Eschrichtius*). Specifically, the lateral branch of the maxillary canal (superior alveolar canal) has connections with both the dental alveoli and the lateral palatal foramina. Ekdale and Deméré's (2021) results

provide evidence for the co-occurrence hypothesis, but do not bear on the timing of enamel versus dentin loss in the ancestor of modern baleen whales.

To clarify the correct sequence of evolutionary events that culminated in tooth loss, several studies have investigated the functional versus pseudogenic status of tooth-specific genes in extant mysticetes. If teeth were lost on the stem mysticete branch (Fitzgerald, 2006, 2009; Meredith et al., 2011a), we should expect to find inactivating mutations in both enamel and dentin related genes that are shared by all extant mysticetes (Berta et al., 2016; Deméré et al., 2008; Meredith et al., 2009, 2011a; Springer et al., 2016a, 2019). Inactivating mutations include genetic changes that are expected to radically impact or impair a gene's function including frameshift insertions and deletions (frameshift indels), altered start or stop codons, premature stop codons, insertion of long retroelement sequences, modified splice sites at intron/exon boundaries, and deletions of an exon(s) or an entire gene. Meredith et al. (2011a) reported the first evidence of a shared inactivating mutation in mysticetes. Specifically, the enamel-specific gene matrix metallopeptidase 20 (*MMP20*) has a shared retroelement insertion (a CHR-SINE in exon 2 of all living mysticetes that were examined. More recently, Mu et al. (2021) and Gatesy et al. (2022) reported inactivating mutations in acid phosphatase 4 (*ACP4*) and kallikrein related peptidase 4 (*KLK4*), respectively, that are shared by all extant mysticetes that were examined. Like *MMP20*, these genes are related to enamel formation. Other enamel-specific genes (amelogenin X-linked [*AMELX*], ameloblastin [*AMBN*], amelotin [*AMTN*], enamelin [*ENAM*]) have inactivating mutations in multiple mysticetes, but none that are shared by all mysticete species for the exons that were investigated (Deméré et

al., 2008; Gatesy et al., 2022; Meredith et al., 2009, 2011a). The absence of shared inactivating mutations in these enamel-specific genes may reflect a lag time between the initial relaxation of selection pressures for the maintenance of enamel on the stem mysticete branch and the first occurrence of an inactivating mutation in an enamel gene. Given the decelerated rate of molecular evolution in cetaceans relative to most other mammals, such mutational lags are predicted (Meredith et al., 2009).

Two other genes are dentin-specific (dentin sialophosphoprotein [*DSPP*]) (McKnight and Fisher, 2009) or tooth-specific (odontogenesis associated phosphoprotein [*ODAPH*]) (Springer et al., 2016a), but not enamel-specific. *DSPP* plays an essential role in the formation of dentin and the protease-processed products of *DSPP* comprise the largest component of non-collagenous proteins found in dentin (Yamakoshi and Simmer, 2018). Mutations in the human *DSPP* gene are known to cause both dentin dysplasia and dentinogenesis imperfecta (Yamakoshi and Simmer, 2018). *DSPP* exhibits a 1-bp frameshift insertion that is shared by two balaenids (*Balaena mysticetus, Eubalaena japonica*) (Gatesy et al., 2022). However, this exon is deleted in three balaenopteroids that were examined, and it is unclear whether the 1-bp deletion in exon 3 occurred in the ancestor of balaenids or instead is an older mutation that occurred in the common ancestor of extant mysticetes that was subsequently erased in balaenopteroids by the much larger deletion that completely removed exon 3. If the balaenid 1-bp deletion is also present in the neobalaenid *Caperea marginata*, which is the sister group of balaenopteroids, this would provide evidence for inactivation of the genetic toolkit for dentin production on the stem lineage to crown Mysticeti.

In the case of *ODAPH*, this gene is intact in all dentate placental mammals that have been investigated including several species with enamelless teeth (*Orycteropus afer* [aardvark], *Dasypus novemcinctus* [nine-banded armadillo], *Tolypeutes matacus* [southern three-banded armadillo], *Chaetophractus vellerosus* [screaming hairy armadillo], *Choloepus hoffmanni* [Hoffmann's two-toed sloth], *Choloepus didactylus* [Linnaeus's two-toed sloth]) (Gatesy et al., 2022; Springer et al., 2016a). Morever, Springer et al. (2016a) found that dN/dS values on branches leading to species with enamelless teeth are not significantly different from dN/dS values on branches leading to species with enamel-capped teeth. These results suggest that *ODAPH* remains under purifying selection even in species that have lost the enamel caps on their teeth. By contrast, all toothless placental mammals that have been investigated (ten mysticetes, *Manis pentadactyla* [Chinese pangolin], *Tamandua tetradactyla* [collared anteater]) have inactivating mutations in *ODAPH* including complete deletion of this locus in most balaenopteroids (Gatesy et al., 2022; Springer et al., 2016a). The inactivation of *ODAPH* in multiple toothless clades, but not enamelless clades, suggests that *ODAPH* is tooth-specific but not enamel-specific. At the same time, *ODAPH* has a role in enamel formation, and mutations in this gene are known to cause amelogenesis imperfecta (Ji et al., 2021; Liang et al., 2021; Parry et al., 2012). Even though all mysticetes that have been investigated have inactivating mutations in *ODAPH*, there are no inactivating mutations in the protein-coding sequences of this gene that are shared by all mysticetes. However, this gene was lost at most three times based on inactivating mutations that are shared by Balaenidae and by Balaenopteroidea, respectively; the *Caperea* genome has

not been examined to date and could document a third independent inactivation of this gene (Springer et al., 2016a). Importantly, the absence of inactivating mutations shared by all extant mysticetes in *DSPP* and in *ODAPH* does not imply that these genes were necessarily functional in the most recent common ancestor of extant baleen whales. As noted above, mysticetes have very slow rates of molecular evolution, and there may have been a lag time between relaxed selection on the dentition and the first occurrence of an inactivating mutation in one or both of these genes. DN/dS analyses that measure selective constraints and gene inactivation dating have not yet been employed to examine the timing of relaxed selection on *DSPP* and *ODAPH*.

In view of the above, several key questions remain unanswered: (1) Were teeth lost in a single step or did enamel loss precede dentin loss? (2) Was enamel lost early or late on the stem mysticete branch? (3) If enamel and tooth loss were decoupled in the ancestry of baleen whales, did tooth loss occur on the stem mysticete branch or independently in multiple crown mysticete lineages? To address these questions, we compiled and analyzed complete protein-coding sequences for nine tooth-specific genes from cetaceans with assembled genomes or raw Illumina data. Seven of these genes are enamel specific (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) and two are related to dentin/tooth formation (*DSPP*, *ODAPH*) with the caveat that *ODAPH* is pleiotropic and is also important for enamel maturation (Liang et al., 2021). Taxon sampling included representatives of all four crown mysticete families (Balaenidae, Neobalaenidae, Eschrichtiidae, Balaenopteridae); a genomic library for *Caperea marginata* (Neobalaenidae) was used to sequence short reads to recover and assemble

tooth gene sequences from this species. The nine genes listed above were investigated for shared inactivating mutations. We also employed dN/dS values to determine the timing of inactivation for both enamel and dentin genes, and by proxy the inferred loss of both enamel and teeth in Mysticeti.

1.2. Methods

1.2.1. Gene sampling

Nine genes were chosen for study based on prior evidence that these loci are enamel-specific or tooth-specific based on inactivation of these genes in one or more clades of toothless or enamelless vertebrate species (Deméré et al., 2008; Gasse et al., 2012; Gatesy et al., 2022; Kawasaki et al., 2014; McKnight and Fisher, 2009; Meredith et al., 2009, 2011a, 2013, 2014; Mu et al., 2021; Springer et al., 2015, 2016a, 2019). Additional evidence for the enamel or tooth specificity of these genes derives from mutagenesis studies in mice and natural genetic variation in humans that causes nonsyndromic cases of amelogenesis imperfecta, dentinogenesis imperfecta, or dentin dysplasia (Table 1.1). A tenth tooth-related gene (*ODAM*) is also inactivated in enamelless and edentulous mammals, but this gene is also pseudogenized in all toothed whales that were investigated as well as several other clades of mammals with enamel-capped teeth (Springer et al., 2019). For this reason, we omitted *ODAM* from our study. Also, exon 4 of *AMELX* was not included in our analyses because this exon is subject to alternative splicing and is absent in many mammals (Delgado et al., 2005; Sire et al., 2005, 2006, 2007).

1.2.2. Taxon sampling

Taxon sampling for this study included 44 species of which 13 were mysticetes, 14 were odontocetes, and 17 were terrestrial or semiaquatic cetartiodactyl outgroups. Mysticetes included *Balaena mysticetus* (bowhead whale), *Balaenoptera acutorostrata* (common minke whale), *Balaenoptera bonaerensis* (Antarctic minke whale), *Balaenoptera borealis* (sei whale), *Balaenoptera edeni* (Bryde's whale), *Balaenoptera musculus* (blue whale), *Balaenoptera physalus* (fin whale), *Caperea marginata* (pygmy right whale), *Eschrichtus robustus* (gray whale), *Eubalaena glacialis* (North Atlantic right whale), *Eubalaena australis* (southern right whale), *Eubalaena japonica* (North Pacific right whale), and *Megaptera novaeangliae* (humpback whale). Odontocetes included *Delphinapterus leucas* (beluga), *Kogia breviceps* (pygmy sperm whale), *Kogia sima* (dwarf sperm whale), *Lagenorhynchus obliquidens* (Pacific white-sided dolphin), *Lipotes vexillifer* (Chinese river dolphin), *Mesoplodon bidens* (Sowerby's beaked whale), *Monodon monoceros* (narwhal), *Neophocaena asiaeorientalis* (Yangtze finless porpoise), *Orcinus orca* (killer whale), *Physeter macrocephalus* (sperm whale), *Phocoena phocoena* (harbor porpoise), *Sousa chinensis* (Indo-Pacific humpback dolphin), *Tursiops aduncus* (Indo-Pacific bottlenose dolphin), and *Tursiops truncatus* (common bottlenose dolphin). Outgroup taxa included *Bison bison* (American bison), *Bos mutus* (wild yak), *Bubalus bubalis* (water buffalo), *Camelus bactrianus* (Bactrian camel), *Capra hircus* (domestic goat), *Catagonus wagneri* (Chacoan peccary), *Choeropsis liberiensis* (pygmy hippopotamus), *Elaphurus davidianus* (Pere David's deer), *Giraffa camelopardalis* (giraffe), *Hippopotamus amphibius* (river hippopotamus), *Moschus moschiferus* (Siberian

musk deer), *Odocoileus virginianus* (white-tailed deer), *Okapia johnstoni* (okapi), *Ovis aries* (domestic sheep), *Sus scrofa* (wild boar), *Tragulus javanicus* (Java mouse-deer), and *Vicugna pacos* (alpaca).

1.2.3. Data collection

DNA sequences for nine different tooth genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *DSPP*, *ENAM*, *KLK4*, *MMP20*, *ODAPH*) were obtained from (1) assembled genomes at NCBI (https://www.ncbi.nlm.nih.gov/) and the The Bowhead Whale Genome Resource (http://www.bowhead-whale.org/), (2) raw sequence reads at NCBI's Sequence Read Archive (SRA), and (3) newly generated Illumina whole-genome sequence data (John Gatesy and Mark Springer). NCBI's RefSeq and Nucleotide databases were searched using keywords for all nine genes in conjunction with taxon names for four reference species (*Capra hircus*, *Camelus bactrianus*, *Orcinus orca*, *Tursiops truncatus*). Sequences for each reference species were then imported into Geneious Prime (current version 2021.1.1, https://geneious.com) (Kearse et al., 2012), aligned with MAFFT (Katoh and Toh, 2008), and cross-checked against each other for consistent annotations. Sequences for additional species were collected through NCBI's Nucleotide Basic Local Alignment Search Tool (BLAST), which was used to search both assembled and unassembled genomes using the whole-genome shotgun (WGS) and SRA databases, respectively. Each BLAST search employed a query sequence from a closely related species. Megablast was used for highly similar sequences (e.g., taxa in same family), whereas Blastn was used for less similar sequences (e.g., taxa in different families). Top-scoring BLAST results were imported into Geneious Prime. Sequences obtained through

the SRA database were assembled using Geneious Prime's 'Map to Reference' approach, where the reference sequence was a closely related species to the SRA taxon. We allowed for a maximum mismatch of 10% per read and required a minimum of two reads for base calling with a consensus threshold of 65%. Unassembled genome sequences for two additional cetacean species (*Caperea marginata*, *Kogia sima*) were obtained from DNA libraries that were constructed with Illumina's NeoPrep procedure and then sequenced at ~40X coverage at the New York Genome Center with paired-end sequencing (150 bp per read) on a HiSeq 2500 platform. DNA samples for these libraries were provided by Southwest Fisheries Science Center (SWFSC) (*C. marginata* [Lab ID 5989]; *K. sima* [Lab ID 175303]). Sequences for the nine tooth genes were then obtained using a map to reference approach as described above. Accession numbers for these new sequences are OK282856-OK282863 for *C. marginata* and OK391138-OK391144 for *K. sima*.

1.2.4. Alignments and tabulation of inactivating mutations

Complete protein-coding sequences and introns were aligned in Geneious Prime using MAFFT (Katoh and Toh, 2008). Sequences were manually spot-checked for alignment errors using AliView version 1.23 (Larsson, 2014). Alignments were examined for inactivating mutations (frameshift indels, start and stop codon mutations, premature stop codons, splice site mutations), which were annotated in Geneious Prime. Mutations were mapped onto the species tree using delayed transformation (DELTRAN) parsimony optimization.

1.2.5. Phylogenetic analyses

Gene trees were constructed from protein-coding sequences with maximum likelihood using the program RAxML version 8.2.11 in Geneious Prime (raxmlHPC-SSE3-MAC) (Stamatakis, 2014). Rapid bootstrapping (500 pseudoreplicates) and a search for the best tree were performed in the same analysis (Stamatakis et al., 2008). We employed the GTRGAMMA option, which implements the GTR + Γ model of sequence evolution.

1.2.6. Selection analyses

Inferred inactivating mutations, estimates of selection intensity (dN/dS analyses), and divergence times from McGowen et al.'s (2020a) cetacean timetree were used to reconstruct inactivation times of tooth genes and, by proxy, phenotypes. Selection (dN/dS) analyses were conducted with the codeml program of PAML (version 4.9e, Yang, 2007). Analyses were performed with a concatenation of seven enamel-specific genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) that serve as a proxy for enamel and two dentin/tooth-specific genes (*DSPP*, *ODAPH*) that serve as a proxy for the presence of teeth. We used concatenations of enamel and dentin/tooth genes rather than individual genes because larger data sets yield dN/dS that are less impacted by sampling error. For the analysis with *DSPP* + *ODAPH* we excluded exon 4 of *DSPP* because this exon contains short, highly repetitive motifs that are difficult to align. We employed branch-specific codon models with branch categories (background, transitional, pseudogenic) that were based on phenotypes and shared genetic mutations (*sensu* Meredith et al., 2009). The background branch category includes branches that lead to

internal nodes or extant species with functional teeth that are capped with prismatic enamel (seven enamel genes) or teeth irrespective of whether enamel is present (two dentin/tooth genes). These branches are expected to have evolved under purifying selection with dN/dS < 1 in terrestrial cetartiodactyl outgroups and some or all odontocetes (all odontocetes for dentin/tooth genes but only some odontocetes for enamel genes). Transitional branches lead to internal nodes or extant species that lack enamel/prismatic enamel or teeth and contain the first detected occurrence of an inactivating mutation that is shared by all members of a clade. Each transitional branch was given its own branch category. Transitional branches have mixed evolutionary histories that include a period of evolution under purifying selection followed by a period of neutral evolution after selection was relaxed and the phenotype was lost (Meredith et al., 2009). DN/dS values on transitional branches are expected to be intermediate between dN/dS values for background branches and pseudogenic branches. Pseudogenic branches post-date transitional branches and are expected to have neutral evolutionary histories with dN/dS values near 1. DN/dS analyses for the seven enamel genes included nine branch categories (two background [#0, #1], six transitional [#2–#7], one pseudogenic [#8]) as follows: #0 background terrestrial cetartiodactyl outgroups, #1 background Odontoceti (all odontocetes with complex enamel [categories 4 and 5 of Werth et al., 2020]), #2 *Physeter*, #3 stem *Kogia*, #4 *Mesoplodon*, #5 stem Monodontidae, #6 stem Phocoenidae, #7 stem Mysticeti, #8 crown Mysticeti + crown *Kogia* + crown Monodontidae + crown Phocoenidae. Background cetartiodactyl outgroups and background Odontoceti were separated into two categories to determine if

there is relaxed selection on these enamel genes in odontocetes with complex enamel, i.e., enamel with prismatic radial enamel that may also exhibit Hunter-Schreger bands (HSB) or other decussations (Werth et al., 2020). DN/dS analyses with the dentin genes included three branch categories: #0 background terrestrial cetartiodactyl outgroups plus stem and crown Odontoceti, #1 stem Mysticeti + stem Balaenidae, #2 crown Balaenidae. Other mysticete families were excluded from these analyses because the coding sequences for exons 1-3 of *DSPP* and all of *ODAPH* have been deleted (*Balaenoptera musculus* retains *ODAPH* [Springer et al., 2016a] but not exons 1-3 of *DSPP*).

Analyses were performed with two codon frequency models, CF1 and CF2 (Yang, 2007). CF1 estimates codon frequencies from mean nucleotide frequencies across all three codon positions, whereas CF2 estimates frequencies at each of the individual codon positions. Codon positions are absent in pseudogenes so it is important to verify that analyses without base compositional differences at different codon positions (i.e., CF1) yield results that are similar to results that are obtained with a codon frequency model that allows for base compositional differences at 1^{st}, 2^{nd}, and 3^{rd} codon positions (i.e., CF2). Analyses were conducted with both fixed (dN/dS = 1) and estimated values for the fully pseudogenic branch category (dN/dS = 1.0 or dN/dS = estimated); chi-square tests were conducted using the program chi2 (Yang, 2007) to determine whether the analyses with fixed versus estimated dN/dS values for the pseudogenic branch category were significantly different from each other. All frameshift insertions were deleted prior to performing dN/dS analyses. In addition, premature stop codons were recoded as missing data as required for codeml analyses. The species tree used for Cetartiodactyla

relationships was taken from McGowen et al. (2020a) for cetaceans and Hassanin et al. (2012) for outgroups to Cetacea.

1.2.7. Gene inactivation times

Inactivation times for the concatenation of seven enamel genes and for the concatenation of two dentin genes were each estimated using equations from Meredith et al. (2009) that allow for either one or two synonymous substitution rates. The one synonymous substitution rate model assumes that the rate of synonymous substitution is neutral and equal on both functional and pseudogenic branches, whereas the two-rate model assumes that the synonymous substitution rate on functional branches is non-neutral and is 70% of the substitution rate on pseudogenic branches (Bustamante et al., 2002; Meredith et al., 2009). Divergence time estimates were taken from McGowen et al. (2020a).

1.3. Results

1.3.1. Alignments and gene trees

Complete coding sequences with all exons for all genes were recovered for all the non-cetacean cetartiodactyl outgroup taxa that were sampled, as well as for all odontocete species except for the dwarf and pygmy sperm whales (*Kogia* spp.) that lack well-developed enamel (Bianucci and Landini, 2006; Werth et al., 2020). Mysticetes that were sampled had varying degrees of completeness for the coding sequences of the nine tooth genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *DSPP*, *ENAM*, *KLK4*, *MMP20*, *ODAPH*) and in some cases had missing exons or missing genes (see below). Table 1.2 summarizes the presence or absence of 27 well-supported and non-controversial clades in Cetartiodactyla

(Gatesy et al., 2013; Hassanin et al., 2012; McGowen et al., 2014, 2020a; Meredith et al., 2011b) on the individual gene trees. All of the individual gene trees recovered the majority of well-supported clades in Cetartiodactyla. Eight clades were recovered on all gene trees with the constituent taxa and the mean number of recovered clades was 23.78. Coding sequences for *ENAM* are longer than for any of the other enamel-related genes and 27 of 27 clades were recovered on the *ENAM* tree. *ODAPH* contains the shortest coding sequences and recovered the fewest clades (20 of 25).

1.3.2. Inactivating mutations in mysticetes

Table 1.3 provides a summary of inactivating mutations in mysticetes, which are also mapped onto a species tree for Mysticeti in Figure 1.1. Of the seven enamel genes, three (*ACP4, KLK4, MMP20*) were found to have inactivating mutations that are shared by all mysticetes. *ACP4* has a 1-bp deletion in exon 4; *KLK4* has a 1-bp deletion in exon 3; and *MMP20* has a SINE insertion in exon 2. Remnants of this insertion range from 302 to 324 bp in different mysticete species with assembled genomes, and all possible reading frames of this SINE insertion contain premature stop codons that would result in a severely truncated *MMP20* protein (Meredith et al., 2011a). Mu et al. (2021) reported a second 1-bp deletion in exon 5 of *ACP4* that is shared by all mysticetes that were sampled in that study, but expanded taxon sampling revealed that this deletion is not shared by *Caperea marginata* or *Balaenoptera edeni*. Nevertheless, three shared inactivating mutations support the hypothesis that baleen whales lost functional enamel on the stem branch leading to crown Mysticeti (Fig. 1.1).

Among the other enamel-related genes, *ENAM* has inactivating mutations in all mysticetes that were examined, but none that are shared by all taxa (Fig. 1.1). The most inclusive mutations in *ENAM* include a premature stop codon in exon 6 that is shared by *Eschrichtius*, *Megaptera*, and three species of *Balaenoptera*; a premature stop codon in exon 8 that is shared by *Caperea*, *Megaptera*, and four species of *Balaenoptera*; and a premature stop codon in exon 8 that is shared by all four balaenids. *AMBN* has one or more inactivating mutations in all mysticetes except for *Balaenoptera acutorostrata* including a 1-bp insertion in exon 11 that is shared by all four balaenids. *AMELX* has one or more inactivating mutations in all mysticetes except for *Balaenoptera musculus* and *Eubalaena glacialis*. The most inclusive inactivating mutation in *AMELX* is a premature stop codon in exon 6 that is shared by all balaenopteroids except for *Balaenoptera musculus*. Finally, *AMTN* remains intact in most mysticetes, but has inactivating mutations in three species of *Balaenoptera* (*B. acutorostrata*, *B. bonaerensis*, *B. musculus*) and negative BLAST or map to reference results for the entire protein-coding sequence (*B. borealis*) or exons 6–9 of this gene (*B. physalus*). The only definitive shared inactivating mutation in *AMTN* is a 4-bp insertion in exon 4 in the two minke whales (*Balaenoptera acutorostrata*, *B. bonaerensis*).

Mu et al. (2021) reported that seven of nine mysticetes that were examined have exon 1 sequences for *ACP4*, including five species with a mutated start codon (GTG). We recovered exon 1 sequences for *ACP4* in three balaenids, but not in any balaenopteroids. The exon 1 sequences that we found are not orthologous with the putative exon 1 sequences that were reported by Mu et al. (2021) for mysticetes. Instead, our analyses

indicate that Mu et al.'s (2021) exon 1 sequences for mysticetes are more than 4 kb upstream from the correct exon 1 location in cetaceans and are the result of an annotation error by NCBI's automated software (Fig. 1.2A). A ~2.25 kb deletion in balaenopteroids removed approximately 2.1 kb of sequence that is upstream of exon 1, all of exon 1, and 43 bp at the 5' end of intron 1 (Fig. 1.2A). Negative map to reference results suggest that this exon was also deleted in the neobalaenid *Caperea marginata*.

The tooth/dentin genes *DSPP* and *ODAPH* do not have any inactivating mutations that are shared by all mysticetes. However, *DSPP* exhibits two inactivating mutations that are shared by all four balaenids (1-bp insertion in exon 3, donor splice site mutation [AT] in intron 1) (Table 1.3). There is also a large deletion (e.g., 18.7 kb in *Balaenoptera musculus*) that is present in all species of Plicogulae (Fig. 1.2B). This deletion includes the 5' UTR, exon 1, intron 1, exon 2, intron 2, exon 3, intron 3, and most of exon 4. The remnants of exon 4 (including the stop codon) in different species of Plicogulae are only 390 to 585 bp in length. By contrast, the *DSPP* coding sequences in *Sus scrofa* and *Capra hircus* are 3072 and 3528 nucleotides in length (start codon through stop codon), respectively. A caveat regarding the large deletion in Plicogulae is that its 3' boundary is sensitive to alignment settings and may or may not be at a homologous position in balaenopteroids and the neobalaenid *Caperea*. Furthermore, and importantly, it is possible that the 1-bp frameshift insertion in exon 3 of balaenids may have originated in the common ancestor of Mysticeti with subsequent erasure of this evidence for early inactivation on the stem mysticete branch when exon 3 was deleted in balaenopteroids and the neobalaenid. In the case of *ODAPH*, there is a start codon mutation (ATG >

TTG) in all four balaenids, complete deletion of the gene in *Caperea marginata* and all balaenopteroids except for *Balaenoptera musculus*, and two inactivating mutations in *B. musculus ODAPH*. The inactivating mutations in dentin/tooth genes suggest that teeth were lost at most two times in crown Mysticeti, i.e., on the stem branches leading to Balaenidae and Plicogulae, respectively (Fig. 1.1).

1.3.3. Inactivating mutations in odontocetes

All of the enamel genes except for *MMP20* exhibit an inactivating mutation(s) in one or more odontocetes (Table 1.4, Fig. 1.3). With the exception of a donor splice site mutation in intron 4 of *Orcinus orca* (killer whale), all of the inactivating mutations are in odontocetes that lack complex enamel with well-developed prisms (*sensu* Werth et al., 2020) and whose enamel (if present) is often worn away in adults. Non-delphinid odontocetes with inactivating mutations in enamel genes include two monodontids (*Monodon monoceros*, *Delphinapterus leucas*), two phocoenids (*Neophocaena asiaeorientalis*, *Phocoena phocoena*), one ziphiid (*Mesoplodon bidens*), and three physeteroids (*Physeter macrocephalus*, *Kogia breviceps*, *K. sima*). The two monodontids share a donor acceptor splice site mutation (GT → AT) in intron 2 of *AMELX*. One or both of the monodontids also exhibit autapomorphic inactivating mutations in *ACP4*, *AMBN*, *AMTN*, and *KLK4*. The two phocoenids share a premature stop codon in exon 2 of *KLK4* and an acceptor splice site mutation (AG → AT) in intron 2 of *AMTN*. There is also an autapomorphic premature stop codon in *Neophocaena asiaeorientalis AMELX*. Among physeteroids, *Physeter macrocephalus* has three autapomorphic frameshift indels in *ACP4*. The two species of *Kogia*, in turn, exhibit inactivating mutations and/or

negative BLAST/map to reference results for five (*K. sima*) or six (*K. breviceps*) enamel genes. Shared inactivating mutations are present in *ACP4*, *AMELX*, and *ENAM*, and there were negative BLAST results (*K. breviceps*) or no mapped reads (*K. sima*) for *KLK4*. Presumably, *KLK4* is deleted in *Kogia*. Finally, the ziphiid *M. bidens* has three inactivating mutations in *ACP4*.

1.3.4. Selection analyses

Selection (dN/dS) analyses were performed on seven enamel genes that were concatenated together and two tooth/dentin genes that were concatenated together. The concatenated alignment of enamel genes included 44 taxa and 9618 bp. The concatenated dentin alignment included 35 taxa and 2127 bp. Plicogulae species were excluded from the concatenated dentin alignment because of their incompleteness. The results of selection analyses on both the concatenated enamel and concatenated tooth/dentin genes are summarized in Table 1.5.

For the enamel genes, dN/dS values for the two background categories, terrestrial cetartiodactyls and odontocetes with complex enamel (Werth et al., 2020), ranged from 0.3939 to 0.4072 for the former and 0.5400 to 0.5577 for the latter. These values are indicative of purifying selection and further suggest that purifying selection has been stronger in terrestrial cetartiodactyls than in odontocetes with complex enamel, perhaps because odontocetes generally do not masticate their food as assiduously as herbivorous/omnivorous terrestrial cetartiodactyls. The dN/dS value for fully pseudogenic branches ranged from 1.168 (CF1) to 1.2028 (CF2) when these values were estimated rather than fixed at 1.0. These values are significantly different from 1.0 based

on log likelihood ratio tests (p = 0.029 [CF1], p = 0.009 [CF2]). Among the transitional branches, the dN/dS value on the stem mysticete branch was intermediate between the functional and pseudogenic dN/dS values and ranged from 0.9507 to 0.9818. DN/dS values for transitional branches in Odontoceti were all higher than dN/dS values for the two functional categories and in three cases (stem *Kogia*, *Physeter*, stem Phocoenidae) were slightly higher than the estimated dN/dS values for the fully pseudogenic branches. However, in each case these elevated values were not significantly higher than the estimated values for the pseudogenic branches when individual transitional branches were constrained to have the same dN/dS value as the fully pseudogenic branches (stem *Kogia*: p = 0.48 [CF1], p = 0.47 [CF2]; *Physeter*: p = 0.74 [CF1], p = 0.76 [CF2]; stem Phocoenidae: p = 0.42 [CF1], p = 0.41 [CF2]).

For selection analyses in which fully pseudogenic dN/dS was estimated, the two dentin genes yielded dN/dS values of 0.5735 (CF1) and 0.5533 (CF2) for the background branches that lead to dentate taxa, 0.8263 (CF1) and 0.7989 (CF2) for the fully pseudogenic branch category of crown mysticetes that are edentulous, and 0.7197 (CF1) and 0.6654 (CF2) for the transitional branch that merges the stem Mysticeti and stem Balaenidae branches (Table 1.5). As expected, the background branches have the lowest dN/dS values, the fully pseudogenic branches have the highest dN/dS values, and the transitional branch has intermediate dN/dS values. The dN/dS value for the fully pseudogenic crown branches is less than the expected value of 1.0, but the difference is not statistically significant based on a likelihood ratio test (p = 0.54 [CF1], p = 0.47 [CF2]).

1.3.5. Gene inactivation times

The timings of enamel and dentin/tooth loss were estimated by proxy using dN/dS values for the concatenations of enamel and dentin/tooth genes, respectively, equations from Meredith et al. (2009), and cetacean divergence dates from McGowen et al. (2020a). The mean of eight different estimates for inactivation of the enamel-specific genes on the stem Mysticeti branch is 34.62 Ma, whereas the mean inactivation time for dentin/tooth-specific genes on the combined stem Mysticeti + stem Balaenidae branch is 19.94 Ma (Table 1.6, Fig. 1.4). Note that species in Plicogulae were not included in the dentin/tooth gene analysis because of exon and gene deletions. These inactivation times suggest that enamel was lost very early on the stem Mysticeti branch, whereas dentin/teeth were lost on the stem Balaenidae branch, which extends from 25.73 Ma to 10.61 Ma based on McGowen et al.'s (2020a) autocorrelated rates timetree (Fig. 1.4). This is consistent with the observation that shared inactivating mutations occur in Balaenidae but not for all mysticete species. There is insufficient information to estimate a reliable date for dentin/tooth loss in the ancestor of Plicogulae based on dN/dS analyses, but exon deletions in *DSPP* and the complete deletion of *ODAPH* in most plicogulans suggest that dentin/teeth were lost on the stem plicogulan branch, which extends from 25.73 to 22.11 million years ago based on McGowen et al.'s (2020a) timetree. Overall, these results suggest a two-step model for the loss of teeth in the ancestry of living baleen whales, where the initial loss of enamel on the mysticete stem lineage was followed by subsequent loss of dentin/teeth on two or more lineages within the crown group.

1.4. Discussion

1.4.1. Tooth loss in mysticetes

Both fossil evidence and ancestral reconstructions of edentulism suggest that postnatal teeth were lost on the stem lineage to crown Mysticeti. Previous studies based on molecular data validate the hypothesis that the genetic toolkit for enamel production was knocked out on the stem mysticete branch (Gatesy et al., 2022; Meredith et al., 2011; Mu et al., 2021). However, shared inactivating mutations in a dentin/tooth-related gene have not been reported in mysticetes. Indeed, it remains unclear if enamel and dentin loss were coupled or if enamel loss preceded edentulism. Gatesy et al. (2022) articulated three hypotheses for the evolution of edentulism in extant baleen whales: (H1) enamel and dentin components of teeth were lost simultaneously on the stem mysticete branch, (H2) enamel and dentin/teeth were lost in a stepwise fashion on the stem mysticete branch (i.e., enamel first, then dentin/teeth), and (H3) enamel loss occurred on the stem mysticete branch followed by the independent loss of dentin (and teeth) in two or more crown mysticete lineages. We addressed these hypotheses through molecular evolutionary analyses of protein-coding sequences for seven enamel-related genes and two tooth/dentin-related genes.

In the case of the seven enamel-related genes we inferred three inactivating mutations that are shared by all mysticetes: a SINE insertion in exon 2 of *MMP20*, a 1-bp deletion in exon 4 of *ACP4*, and a 1-bp deletion in exon 3 of *KLK4*. Each of these inactivating mutations has previously been reported albeit with less comprehensive taxon sampling for mysticetes (*MMP20* [Meredith et al., 2011]; *ACP4* [Mu et al., 2021]; *KLK4*

[Gatesy et al., 2022]). Mutations in *KLK4*, *MMP20*, and *ACP4* are all associated with amelogenesis imperfecta in humans. Defects in the *ACP4* gene cause autosomal recessive hypoplastic amelogenesis imperfecta (Seymen et al., 2016; Smith et al., 2017). Mutations in *KLK4* in are associated with autosomal recessive hypomaturation amelogenesis imperfecta with a phenotype of soft, stained enamel (Hart et al., 2004; Kawasaki et al., 2014; Wang et al., 2013). Finally, mutations in *MMP20* are associated with autosomal recessive hypomaturation amelogenesis imperfecta with enamel that is soft and unusually thin (Wang et al., 2020). In addition, *MMP20* null-mice exhibit hypoplastic and hypomineralized enamel (Caterina et al., 2002; Kawasaki and Susuki, 2012; Wright et al., 2009). Taken together, mutations in these three genes in the common ancestor of Mysticeti would be expected to eliminate the production of normal, prismatic enamel. Moreover, mutations in *MMP20* can also cause reductions in the thickness and mineral density of dentin, possibly because *MMP20* is involved in protease-processing of the precursor *DSPP* protein that is rapidly cleaved in dentin (Wang et al., 2020). Given these findings, the first step in the alteration of normal dentin in living mysticete species may have occurred in their common ancestor as a result of the *MMP20* SINE insertion in exon 2 of this gene.

In addition to these inactivating mutations that are shared by all mysticetes, there are numerous mutations, some of which are shared by two or more species, in all seven of the enamel-related genes (Fig. 1.1; Table 1.3). These results suggest that enamel-related genes have been evolving neutrally since the loss of enamel on the stem mysticete branch. The absence of shared inactivating mutations (for Mysticeti) in four enamel-

related genes (*AMBN*, *AMELX*, *AMTN*, *ENAM*) may reflect a lag time between the onset of neutral evolution and the accumulation of the first inactivating mutation in each of these four genes. This scenario has previously been inferred for cone-specific phototransduction genes in rod monochromatic cetaceans (Springer et al., 2016b). DN/dS analyses validate the hypothesis that the seven-enamel related genes have evolved neutrally in crown Mysticeti after purifying selection was abolished on the stem mysticete branch.

By contrast with the seven enamel-related genes, we did not find inactivating mutations that are shared by all mysticetes in either of the dentin/tooth related genes (*DSPP*, *ODAPH*) that we examined. Springer et al. (2016a) previously reported this result for *ODAPH* (= *C4orf26*) but with more limited taxon sampling. Nevertheless, there are inactivating mutations in *ODAPH* that map to the common ancestors of Balaenidae (start codon mutation) and Plicogulae (gene deletion), respectively, although in the latter case one balaenopterid (*Balaenoptera musculus*) retains a pseudogenic copy of *ODAPH*. One explanation for this pattern is that the deletion of *ODAPH* was polymorphic in the ancestor of Plicogulae with subsequent lineage sorting (Springer et al., 2016a), but alternatively, the presence of *ODAPH* in *B. musculus* might be due to gene flow and introgression with now extinct mysticete lineages that retained functional *ODAPH*. Such hybrid gene flow events are thought to have been extensive in mysticete phylogeny (Arnason et al., 2018; Berube and Aguilar, 1998; Springer et al., 2020). In the case of *DSPP*, a splice site mutation in intron 1 and a 1-bp insertion in exon 3 are shared by all four balaenids. There is also a large deletion that is shared by Plicogulae. This deletion

encompasses exons 1–3 and most of exon 4. The most parsimonious hypothesis is that this deletion occurred in the common ancestor of Plicogulae. However, alignment uncertainties at the 3' boundary of this deletion, which occurs in a highly repetitive region of *DSPP*, allow for the alternate scenario wherein the deletion occurred independently in Neobalaenidae and in Balaenopteroidea. Either way, the deletion of exon 3 in Plicogulae leaves open the possibility that the 1-bp frameshift insertion in Balaenidae originated in the common ancestor of Mysticeti with subsequent loss of evidence for this mutation in Plicogulae when exon 3 was deleted in this clade. In summary, the combined evidence from *ODAPH* and *DSPP* suggests that the toolkit for tooth/dentin formation was inactivated at most two times within crown Mysticeti, once in the common ancestor of Balaenidae and once in the common ancestor of Plicogulae.

Given the very slow rates of molecular evolution that occur in mysticetes (Meredith et al., 2009), selection on the two dentin/tooth genes may have been relaxed in the common ancestor of Mysticeti even though there are no shared inactivating mutations in tooth/dentin genes in this clade. We tested this hypothesis using coding sequences for *ODAPH* and the first three exons of *DSPP* in balaenids (exon 4 is highly repetitive and difficult to align). The results of these analyses suggest that selection on the two tooth/dentin genes was relaxed ~20 Ma in the common ancestor of Balaenidae, which is ~15 million years after selection was relaxed on the seven enamel genes in the common ancestor of crown Mysticeti. Together, shared inactivating mutations and the results of dN/dS analyses provide support for H3 (see above) wherein enamel loss and tooth loss were decoupled in the ancestry of living baleen whales (Fig. 1.4). First, enamel loss

occurred ~35 Ma on the stem mysticete branch. This molecular estimate for enamel loss in mysticetes is older than current paleontological evidence for enamel loss in Mysticeti. The early stem mysticete *Llanocetus denticrenatus* is ~34 million years in age (Fordyce and Marx, 2018) and possessed prismatic enamel with HSB (Loch et al., 2020), but the loss of enamel is recorded much later in the fossil record (Gatesy et al., 2022). Next, independent dentin/tooth loss occurred on the stem balaenid and stem Plicogulae branches, respectively. Tooth/dentin loss occurred ~20 Ma on the stem balaenid branch based on dN/dS analyses and inactivation dating. We were unable to date tooth/dentin loss on the stem Plicogulae branch using our dN/dS methods, but if the loss of *ODAPH* and/or the deletion of most of *DSPP* are shared inactivating mutations in Neobalaenidae + Balaenopteroidea, then dentin/teeth may have been lost on this branch. Estimated divergence dates from McGowen et al.'s (2020a) timetree for Cetacea therefore suggest that dentin/tooth loss on the stem Plicogulae branch occurred 25.73-22.11 Ma.

Our hypothesis for tooth loss (H3) in the ancestry of living whales is incongruent with ancestral reconstructions of edentulism in mysticetes (Deméré et al., 2008; Fitzgerald, 2006, 2010; Meredith et al., 2011a). One explanation for this incongruence is the possibility that small enamelless teeth were set in the gums rather than in bony alveoli in some stem mysticetes and crown mysticetes. This condition is known in some extant ziphiids (Boschma, 1951; Fordyce et al., 1979; Gomerčić et al., 2006). In addition, Lambert et al. (2008) suggested that small teeth may have been embedded in the gums rather than in alveoli in the Miocene ziphiid *Nazcacetus urbinai*. If this condition also occurred in late stem and early crown mysticetes from the Oligocene and Miocene, it is

possible that fossils might not record this anatomical condition. The small, loosely set teeth may have become detached from the jaws of such species by post-mortem taphonomic processes and not recovered in association with the skull (Gatesy et al., 2022). Along these lines, there is suggestive but inconclusive evidence for the presence of very small teeth, set in alveoli, that were located at the tips of the upper and lower jaws in some Oligocene eomysticetid mysticetes including *Yamatocetus canaliculatus* and *Tokarahia* sp. (Boessenecker and Fordyce, 2015; Okazaki, 2012). It remains to be determined if this condition may have occurred in some of the earliest crown mysticetes. More generally, additional paleontological and molecular evolutionary studies will be required to test competing hypotheses pertaining to the loss of teeth in extant mysticetes. Beyond *DSPP* and *ODAPH*, we are unaware of additional genes that are dentin/tooth-specific but not enamel-specific. The discovery of such genes, if they exist, will allow for a more thorough investigation of the timing and pattern of tooth loss in mysticetes.

Molecular evolutionary evidence for asynchronous enamel and tooth loss in mysticetes, with the latter occurring within the crown group, is compatible with the co-occurrence hypothesis for the evolution of baleen wherein transitional forms such as *Aetiocetus weltoni* are interpreted as species that possessed both teeth and baleen (Deméré et al., 2008; Ekdale and Deméré, 2021). More specifically, the evolution of baleen occurred on the stem mysticete branch before teeth were finally lost within crown Mysticeti (Fig. 1.4). By contrast, our results are incompatible with the toothless suction-feeding hypothesis wherein teeth were lost prior to the evolution of baleen (Peredo et al., 2017, 2018). This second hypothesis requires the loss of teeth very deep in the mysticete

tree on the stem lineage, while our analysis of tooth genes implies a much later loss of teeth within crown Mysticeti. Peredo et al. (2018) argued that the Oligocene mysticete *Maiabalaena nesbittae* (~33 Ma) exemplified the toothless, baleenless, suction-feeding intermediate that bridged the transition from feeding with teeth on single prey items to batch filter-feeding with baleen. However, character optimizations based on the mysticete data matrix compiled by Peredo et al. (2018) imply that *Maiabalaena* does not represent the ancestral condition and instead is more likely an aberrant side branch that independently lost the postnatal dentition (Gatesy et al., 2022) millions of years prior to convergent tooth loss in crown Mysticeti (Fig. 1.4).

Thewissen et al. (2017) reported the presence of a mineralized dentin matrix, but not an enamel matrix, in tooth germs of fetal specimens of *Balaena mysticetus* (bowhead) that reach the bell stage of tooth development before they are resorbed. Fetal tooth buds have also been reported for *Megaptera novaeangliae* (humpback whale) (Lanzetti et al., 2020), *Balaenoptera physalus* (fin whale) (Deméré et al., 2008; Dissel-Scherft and Vervoort, 1954), *B. acutorostrata* (common minke whale) (Lanzetti, 2019), and *B. bonaerensis* (Antarctic minke whale) (Ishikawa and Amasaki, 1995; Lanzetti, 2019). [Note: Ishikawa and Amasaki (1995) reported their findings for *B. acutorostrata*, but the fetal specimens they examined were from the 1982/1983 whaling season in Antarctica and correspond to what is now recognized as a full species of minke whale (*B. bonaerensis*) rather than a subspecies (*B. acutorostrata bonaerensis*) (Rice, 1998).] In the case of *B. physalus* and *B. bonaerensis* the tooth germs reach the bell stage, and Ishikawa and Amasaki (1995) reported the presence of predentin that appears in this stage of tooth

development of *B. bonaerensis*. Ishikawa and Amasaki (1995) also noted that both the inner and outer enamel layers form during the bell stage, although the inner enamel layer never differentiates into ameloblasts. Instead, the tooth buds begin to degenerate between dentin formation and ameloblast differentiation. Our results suggest that the predentin/dentin of both Ishikawa and Amasaki (1995) and Thewissen et al. (2017) is a degenerative form of dentin because of inactivating mutations that are present in mysticete *DSPP*. Indeed, type 1 collagen and DSPP are the most abundant protein components of dentin (Yamakoshi, 2009) and of these only DSPP is dentin specific. DSPP is a precursor protein for three other proteins (DSP, DGP, and DPP from amino-terminus to carboxy-terminus) that are found in dentin (Yamakoshi, 2009; Yamakoshi and Simmer, 2018). It will be interesting to learn if any remnants of *DSPP* are expressed in the dentin matrix of fetal mysticetes. In balaenids, the most likely candidate would be a remnant of the amino-terminal DSP protein because there is a frameshift insertion in exon 3 that results in a premature stop codon and a truncated DSPP protein of only 52 amino acids. In Balaenopteridae, a partial dentin phosphoprotein (DPP), which is encoded by the 3' region of exon 4, might be a candidate. For example, the coding sequence remnants of *DSPP* are 387 bp in *B. physalus* and 546 bp in *B. bonaerensis* (stop codons excluded). These would translate truncated protein fragments of the original DPP protein that are just 129 and 182 amino acids, respectively. By contrast, alternative alleles that encode the DPP domain of DSPP in *Sus scrofa* are much longer and have lengths that range from 551 to 594 amino acids (Yamakoshi, 2009). An additional complication is that in toothed mammals, the full length DSPP protein is partially processed by MMP20,

which was inactivated in the common ancestor of living mysticetes. It also remains to be determined if the tooth buds of fetal whales have a role in guiding the formation of the baleen racks (Thewissen et al., 2017) or if they are merely rudiments of the formerly functional dentition that currently serve no purpose.

1.4.2. Enamel loss in odontocetes

Odontocetes exhibit a wide range of enamel phenotypes including highly prismatic enamel with or without HSB, intermediate enamel with less distinct prisms and amorphous crystallite aggregations that run in multiple directions, thin prismless enamel that is easily removed due to wear, and the complete absence of enamel (Bianucci and Landini, 2006; Ishiyama, 1987; Loch et al., 2013a,b; Meredith et al., 2009, 2013; Werth et al., 2020). There is even a recently discovered toothless extinct species, the dwarf dolphin *Inermorostrum xenops*, which belongs to the early diverging odontocete clade Xenorophidae (Boessenecker et al., 2017).

Living odontocetes with highly prismatic enamel (with or without HSB) include delphinids and river dolphins (Werth et al., 2020). These taxa use their teeth to procure, retain, and process their prey more than do other odontocetes (Werth et al., 2020). Our study included five delphinids and one river dolphin (*Lipotes vexillifer*). We are not aware of enamel microstructure data for *L. vexillifer*, but this taxon has highly crenulated enamel (Brownell and Herald, 1972) as in the closely related Amazon River dolphin (*Inia geoffrensis*). In *Inia*, the enamel is highly prismatic with HSB (Werth et al., 2020). Among delphinids and river dolphins in our study, the only inactivating mutation in the seven enamel-related genes is a donor splice site mutation (GT to AT) in intron 4 of

ENAM in *Orcinus orca*. The other eight odontocetes that were included in our study include two phocoenids, two monodontids, one ziphiid, and three physeteroids. Of these, the phocoenids have intermediate enamel and exhibit inactivating mutations in one (*Neophocaena asiaeorientalis*) to three (*Phocoena phocoena*) of the enamel genes. With the exception of *Tasmacetus shepherdi* (Shepherd's beaked whale), which has a full set of teeth, the dentition in extant ziphiids is restricted to tusks and occasionally small vestigial teeth (Loch and van Vuuren, 2016). Four ziphiids that have been investigated (*Berardius bairdii*, *B. arnuxii*, *Mesoplodon densirostris*, *Ziphius cavirostris*) have no enamel or only a thin layer of enamel on their tusks (Ishiyama, 1987; Loch and van Vuuren, 2016; Thewissen, 2018; Werth et al., 2020). The single ziphiid that was included in our study, *Mesoplodon bidens*, exhibits multiple inactivating mutations but all of these are confined to one (*ACP4*) of the seven enamel genes. Among the monodontids, *Monodon monoceros* lacks enamel and *Delphinapterus leucas* has prismless enamel (Ishiyama, 1987). *M. monoceros* has inactivating mutations in four enamel genes and *D. leucas* has inactivating mutations in three enamel genes. The presence of a shared inactivating mutation in *AMELX* in these taxa suggests that enamel degeneration commenced in their common ancestor.

In the case of physeteroids, Bianucci and Landini (2006) reported the presence of enamel in several stem taxa (*Zygophyseter*, *Naganocetus*, *Aulophyseter*) and suggested that enamel was lost in the last common ancestor of crown Physeteroidea, which includes two extant genera (*Kogia*, *Physeter*) as well as several extinct forms (*Orycterocetus*, *Placoziphius*, *Physeterula*, *Scaphokogia*) that were coded by these authors as having no

enamel. Other studies have also reported the absence of enamel in *Physeter* (Flower and Lydekker, 1891) and *Kogia* (Willis and Baird, 1998). By contrast, some authors have noted the presence of a thin layer of prismless enamel in both *Physeter* (Ishiyama, 1987) and *Kogia* (Bloodworth and Odell, 2008; Plön, 2004; Werth et al., 2020). The apparent discrepancy between different studies may be explained by the localized occurrence of very thin enamel at the tips of the teeth and the rapid erosion of this enamel once the teeth have erupted (Bloodworth and Odell, 2008; Ishiyama, 1987; Plön, 2004). At the molecular level, all three extant physeteroid species have inactivating mutations in at least one of their enamel genes. All of the inactivating mutations in *P. macrocephalus* are in *ACP4*, as was also observed for *M. bidens*. For *Kogia*, six enamel genes have inactivating mutations or have been completely deleted in one or both species, but there are no mutations that are shared by *Physeter* and *Kogia*. Nevertheless, the high dN/dS values on the *Physeter* and stem *Kogia* branches (Table 1.5) are consistent with the hypothesis that enamel degeneration from prismatic enamel to prismless enamel occurred very early in the history of crown Physeteroidea.

In summary, nine of 14 odontocete species that were included in our study have one or more inactivating mutations in their battery of seven enamel genes. Odontocete species with absent or degenerative enamel exhibit many more inactivating mutations than odontocetes with highly prismatic enamel such as delphinids, most of which have intact protein-coding sequences and splice sites for all seven enamel genes. Mutations that cause amelogenesis imperfecta in humans and mice are consistent with the lack of prismatic enamel that occurs in various odontocete species that have inactivating

mutations in these same genes. These results also demonstrate that a thin, soft, prismless enamel substance can still be manufactured with a defective genetic toolkit for enamel production. However, this "enamel" is only a shadow of the highly structured and prismatic enamel that is found in terrestrial cetartiodactyls and in some odontocetes (e.g., delphinids).

Tables

Table 1.1. Phenotypes of nine tooth genes in null mice and in humans with various deleterious mutations.

Gene	Previous reports of gene inactivation in mammals with enamelless teeth	Previous reports of gene inactivation in edentulous mammals	Null mice phenotype	Human mutation phenotypes	Selected references
Enamel					
ACP4	Yes (Dasypodidae *Kogia*, Tubulidentata)	Yes (Mysticeti, Pholidota)	Not reported in literature	Autosomal recessive hypoplastic AI	Seymen et al., 2016; Mu et al., 2021
AMBN	Yes (Dasypodidae, Folivora, *Kogia*, Tubulidentata)	Yes (Mysticeti, Pholidota, Vermilingua)	Severe enamel hypoplasia	Autosomal recessive hypoplastic AI	Fukumoto et al., 2004; Meredith et al., 2016; Poulter et al., 2014; Kawasaki and Weiss, 2008
AMELX	Yes (Dasypodidae, Folivora, Tubulidentata)	Yes (Mysticeti, Pholidota, Vermilingua)	Chalky white enamel that is hypoplastic and nonprismatic	X-linked hypoplastic and/or hypomaturation AI	Lagerström et al., 1991; Prakash et al., 2005; Kawasaki et al., 2020
AMTN	Yes (Dasypodidae, Folivora, Tubulidentata)	Yes (Mysticeti, Pholidota)	Mechanically inferior enamel of enamel incisors	Autosomal dominant hypomineralized AI	Nakayama et al., 2015; Smith et al., 2016
ENAM	Yes (Dasypodidae, *Kogia*, Tubulidentata)	Yes (*Hydrodamalis*, Mysticeti, Pholidota)	No true enamel and instead a thick layer of enamel proteins	Autosomal dominant hypoplastic AI	Rajpar et al., 2001; Hu et al., 2008; Meredith et al., 2009; Springer et al., 2015
KLK4	No#	Yes (Mysticeti)	Delayed maturation of enamel crystals; individual enamel crystallites fail to grow together, interlock	Autosomal recessive hypomaturation AI	Hart et al., 2004; Simmer et al., 2009; Kawasaki et al., 2014; Gatesy et al., 2022

*MMP20**	Yes (Folivora, *Kogia*, Tubulidentata)	Yes (Mysticeti)	Severe hypoplastic enamel and accumulated enamel matrix proteins in molars	Autosomal recessive hypomaturation AI with enamel that is soft and unusually thin	Caterina et al., 2002; Kim et al., 2005; Papagerakis et al., 2008; Bartlett et al., 2011; Wang et al., 2020
Dentin/Tooth					
DSPP	No	Yes (Mysticeti, Pholidota)	Hypomineralized predentin that mimics dentinogenesis imperfecta Type III in humans	Dentinogenesis imperfecta and dentin dysplasia	McKnight et al., 2008; Meredith et al., 2014; Zhang et al., 2018
ODAPH	No†	Yes (Mysticeti, Pholidota)	Maturation stage hypomineralization of enamel	Autosomal recessive hypomaturation AI	Springer et al., 2016a; Ji et al., 2021

#There is no evidence that *KLK4* was ever functional outside of Boreoeutheriua. *KLK4* probably originated via a duplication of *KLK5* near the most recent common ancestor of Placentalia and was only maintained in Boreoeutheria where it accelerates the hardening of the enamel, which in turn may have been an exaptation for accelerated eruption of the teeth (Kawasaki et al., 2014).
*Mutations in *MMP20* also impact dentin formation (Wang et al., 2020).
†All mammalian species with enamelless teeth that have been investigated have an intact copy of *ODAPH*, which suggests that this gene is not enamel-specific and instead is tooth-specific (Springer et al., 2016a).
Abbreviations: AI, amelogenesis imperfecta.

Table 1.2. Monophyly of well-supported cetartiodactyl clades on maximum likelihood phylograms for nine tooth related genes.

Clade	Gene								
	ACP4	*AMBN*	*AMELX*	*AMTN*	*DSPP*	*ENAM*	*KLK4*	*MMP20*	*ODAPH*
Camelidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Suiformes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ruminantia + Cetancodonta	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Ruminantia	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pecora	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bovidae	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Bovinae	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Caprinae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cervidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Giraffidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cetancodonta	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Hippopotamidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cetacea	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Odontoceti	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Physeteroidea	No	Yes	Yes	Yes	Yes	Yes	NA	Yes	No
Kogia	Yes	NA	Yes	Yes	Yes	Yes	NA	Yes	Yes
Delphinida + Ziphiidae	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Delphinida	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Delphinoidea	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Monodontidae + Phocoenidae	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Monodontidae	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
Phocoenidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Delphinidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Mysticeti	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Balaenidae	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balaenopteroidea	No	No	Yes	Yes	Yes	Yes	Yes	Yes	NA
Balaenoptera acutorostrata + *B. bonaerensis*	Yes	No	Yes	Yes	No	Yes	Yes	Yes	NA
Total number of clades on gene tree	24/27	22/26	24/27	24/27	25/27	27/27	22/25	26/27	20/25

Table 1.3. Inactivating mutations in enamel- and dentin/tooth-specific genes in Mysticeti.

Taxa with the mutation	**Enamel gene and mutation**	**Dentin gene and mutation**
Mysticeti	*ACP4*: E4: 446D; *KLK4*: E3: 399D; *MMP20*: E2: 214-543I SINE	
All Mysticeti except *Balaena* and *Eubalaena glacialis*†	*KLK4*: E5: 639D	
Balaenidae + *Caperea* + *Balaenoptera borealis* + *B. edeni* + *B. musculus*	*ACP4*: E10: 1135I	
Balaenidae	*ACP4*: E1: 1-3SCM*; *AMBN*: E11: 810I; *ENAM*: In1Ac: AC, E8 2595-2597S; *MMP20*: E1: 1-3 SCM, In7Ac: A-, E8: 1436D, In9Do: AT, E9: 1604-1606S	*DSPP*: E3: 147I, In1Do: AT; *ODAPH*: E1: 1-3SCM
Eubalaena	*ENAM*: E3: 144-146S; *KLK4*: E3: 464D; *MMP20*: E1: 87-91I, In8Ac: AA	
Eubalaena australis + *E. japonica*†	*AMELX*: In3Ac: AA	*DSPP*: E4: 1493-1495S
Eubalaena glacialis + *E. japonica*	*ACP4*: In1AC: AC; *AMBN*: E1: 1-15D; *MMP20*: E4: 914D	*DSPP*: E4: 979-983D
Plicogulae	*ACP4*: E1: 1-111D	
Plicogulae (except *Balaenoptera musculus*)		*ODAPH*: WGD
Balaenopteroidea		*DSPP*: 1-2605D (E1-3, 5' end of E4)
Balaenopteroidea except *Balaenoptera musculus*	*AMELX*: E5: 125-127S	
Balaenopteroidea except *Eschrichtius* and *Balaenoptera edeni*	*MMP20*: E6: 1242-1244S	
Plicogulae (except *Eschrichtius* and *Balaenoptera acutorostrata*)†	*MMP20*: E1: 89I	

Balaenoptera borealis + B. edeni + B. physalus + Megaptera + Eschrichtius	*ACP4*: E4: 425D	
Balaenoptera borealis + B. edeni + B. musculus + Megaptera + Eschrichtius	*ENAM*: E6: 534-536S	
Balaenoptera borealis + B. edeni + B. musculus + B. physalus + Megaptera	*ENAM*: E8: 2562-2564S	
Balaenoptera acutorostrata + B. bonaerensis	*ACP4*: E7: 777-779S; *AMELX*: E6: 277-280D; *AMTN*: E4: 205-208I; *MMP20*: E2: 676-678S	
Balaenoptera acutorostrata + B. bonaerensis + B. physalus	*ENAM*: E8: 1104-1106S; *KLK4*: E3: 206D	
Balaenoptera acutorostrata + B. bonaerensis + B. musculus +B. physalus	*KLK4*: E2: 194D	
Balaenoptera acutorostrata + B. bonaerensis + Eschrichtius†	*ENAM*: E8: 1458-1460S	
Balaenoptera acutorostrata + B. bonaerensis + Eschrichtius + Caperea	*ENAM*: E8: 1845D	
Balaenoptera borealis + B. edeni	*ACP4*: E4: 367-373I, In7Do: CT; *AMBN*: E4: 145-147S; *ENAM*: E8: 2568-2570S; *KLK4* (WGD)	
Balaenoptera musculus + Eschrichtius	*ACP4*: E11: 1303-1304I	
Balaenoptera musculus + Megaptera	*ENAM*: E8: 1778D	
Balaenoptera borealis + B. edeni + B. bonaerensis	*AMBN*: E11: 912I	
Balaenoptera acutorostrata	*ACP4*: E11: 1255D; *AMELX*: E6: 413D; *ENAM*: E8: 3293D	
Balaenoptera bonaerensis	*ACP4*: E10: 1112-1156D; *AMBN*: In2Ac: AA, E11: 1130-1132S, 1250-1251I; *AMTN*: E5: 281-283S;	

	ENAM: E5: 417-419S, E8: 2476-2477D; *KLK4*: E3: 413I, E5: 760-762S; *MMP20*: E2: 685-687S	
Balaenoptera borealis	*ACP4*: E7: 788D; *AMTN*: NRM; *ENAM*: E2: 67I	
Balaenoptera edeni	*ENAM*: E8: 2943-3262D; *MMP20*: E6: NBR	
Balaenoptera musculus	*AMBN*: E2: 120-121D; *AMTN*: 356-643D (3' end of E7, E8, E9)	*ODAPH*: E1: 1-3SCM, In1Do: AT
Balaenoptera physalus	*ACP4*: E11: 1217-1254D, *AMBN*: E9: 669-671S; *AMELX*: In2Do: GK; *AMTN*: E6-9: NBR/NRM; *MMP20*: In1Ac: -G, In2Ac: AT, E8: 1522-1524S	
Megaptera novaeangliae	*ACP4*: In9Ac: AA; *AMBN*: In5Do: AT, In6Do: GG; *KLK4*: E2: 205D; *MMP20*: E7: 1369I	
Eschrichtius robustus	*ACP4*: E11: 1309D; *AMBN*: E11: 929D; *KLK4*: E2: 214-215D; *MMP20*: E1: 1-3SCM, E4: 899-905I	
Caperea marginata	*ACP4*: In9Do: GW, E11: 1302D; *AMBN*: E6: 301-305D (allelic variation); *AMELX*: E2: 1-3SCM; E6: 207-255D; *ENAM*: E5: 486-488S, In7Ac: AR, E8: 1755-1758D (allelic variation); *KLK4*: E1: 1-61 NBR; E2: 688I; E8: 1541I, E9: 1649-1651S	*DSPP*: 1-1875D (E1-E3, 5' end of E4)
Eubalaena australis	*ACP4*: E1-2: 1-220 NBR, E5: 462-560 NBR; *ENAM*: E8: 729-730I	
Eubalaena japonica	*ACP4*: In8Ac: AA	
Eubalaena glacialis	*MMP20*: E7: 1356I	
Balaena mysticetus	ACP4: E11: 1309-1310D; *AMELX*: In2Do: AT:	*DSPP*: In3Do: KT; *ODAPH*: In1Do: GG

	ENAM: E8: 3694D; *MMP20*: In1Do: CT	

Numbers correspond to positions in the protein-coding sequence alignments (CDS) Abbreviations: Ac = acceptor splice site; D = deletion; Do = donor splice site; E = exon; I = insertion; In = intron; NBR = no blast results and possible deletion of exon(s) or gene; NRM = no reads mapped and possible deletion of exon(s) or gene; S = premature stop codon; SCM = start codon mutation; SINE = Short Interspersed Nuclear Element; WGD = whole gene deletion.

*Region containing exon 1 is string of Ns in *Eubalaena australis*.

†Denotes inactivating mutations that map differently with accelerated transformation (ACCTRAN) versus delayed transformation (DELTRAN). Of the five mutations that map differently with these two parsimony optimizations, the only optimization that impacts the timing of gene inactivation is the splice site mutation in *AMELX*. This is because the other four genes all have other inactivating mutations that map equivalently deep or deeper in the tree. For the inactivating mutation in *AMELX*, DELTRAN optimization results in independent inactivation in *Eubalaena australis* and *E. japonica* whereas ACCTRAN optimization results in inactivation of *AMELX* in the common ancestor of *Eubalaena* followed by reversal of this inactivating mutation in *E. glacialis*.

Table 1.4. Inactivating mutations in enamel genes in Odontoceti.

Taxon with the mutation	**Enamel gene and mutation**
Orcinus orca	*ENAM*: In4Do: AT
Phocoenidae	*KLK4*: E2: 73-75S; *AMTN*: In2Ac: AT
Neophocaena asiaeorientalis	*AMELX*: E6: 386-388S
Monodontidae	*AMELX*: In2Do: AT
Monodon monoceros	*ACP4*: In6Do: AT; *AMBN*: E11: 1214-1216S; *KLK4*: E4: 503-505S
Delphinapterus leucas	*ACP4*: E7: 673D; *AMTN*: E8: 571D
Physeter macrocephalus	*ACP4*: E4: 425D, E9: 1015D, E10: 1134D
Kogia	*ACP4*: E1-3 (deletion), E11 (deletion); *AMELX*: E2: 47I, In2Do: GG; *ENAM*: E8: 2042D, 3429-3430D; *KLK4*: NBR/NMR (gene deletion)
Kogia breviceps	*ACP4*: E7: 657-670D, E9: 901D, E10: 1098-1100S, 1109I; *AMBN*: In7Ac: AT, In9AC: AT; *AMELX*: E2: 1-3 SCM; *AMTN*: E8: 373D; *ENAM*: In6Do: CT;
Kogia sima	*AMBN*: NRM (gene deletion); *ENAM*: E1-6: NRM
Mesoplodon bidens	*ACP4*: E2: 146-149I, E5: 536-543D, In8Ac: GG

Numbers correspond to positions in the protein-coding sequence alignments (CDS) Abbreviations: Ac = acceptor splice site; D = deletion; Do = donor splice site; E = exon; I = insertion; In = intron; NBR = no blast results and possible deletion of exon(s) or gene; NRM = no reads mapped and possible deletion of exon(s) or gene; S = premature stop codon; SCM = start codon mutation.

Table 1.5. Selection analyses on seven enamel genes and two tooth/dentin genes.

Branch category	**dN/dS**			
	CF1		**CF2**	
Enamel genes	Fully pseudogenic estimated	Fully pseudogenic fixed	Fully pseudogenic estimated	Fully pseudogenic fixed
Functional terrestrial cetartiodactyls (background)	0.3941	0.3939	0.4072	0.4068
Functional Odontoceti (background)	0.5403	0.5400	0.5577	0.5572
Crown Mysticeti + crown Monodontidae + crown Phocoenidae + crown *Kogia* (fully pseudogenic)	1.1680	1.0000	1.2028	1.0000
Stem Mysticeti (transitional)	0.9507	0.9609	0.9697	0.9818
Stem *Kogia* (transitional)	1.4280	1.4542	1.4780	1.5098
Physeter macrocephalus (transitional)	1.2856	1.2852	1.3134	1.3127
Stem Monodontidae (transitional)	0.6256	0.6254	0.6429	0.6424
Stem Phocoenidae (transitional)	1.4781	1.4771	1.5286	1.5273
Mesoplodon bidens (transitional)	0.9873	0.9863	1.0194	1.0180
Tooth/dentin genes				
Functional Cetartiodactyla (background)	0.5735	0.5737	0.5533	0.5535
Stem Mysticeti + stem Balaenidae (transitional)	0.7197	0.7097	0.6654	0.6549
Crown Balaenidae (fully pseudogenic)	0.8263	1.0000	0.7989	1.0000

Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2.

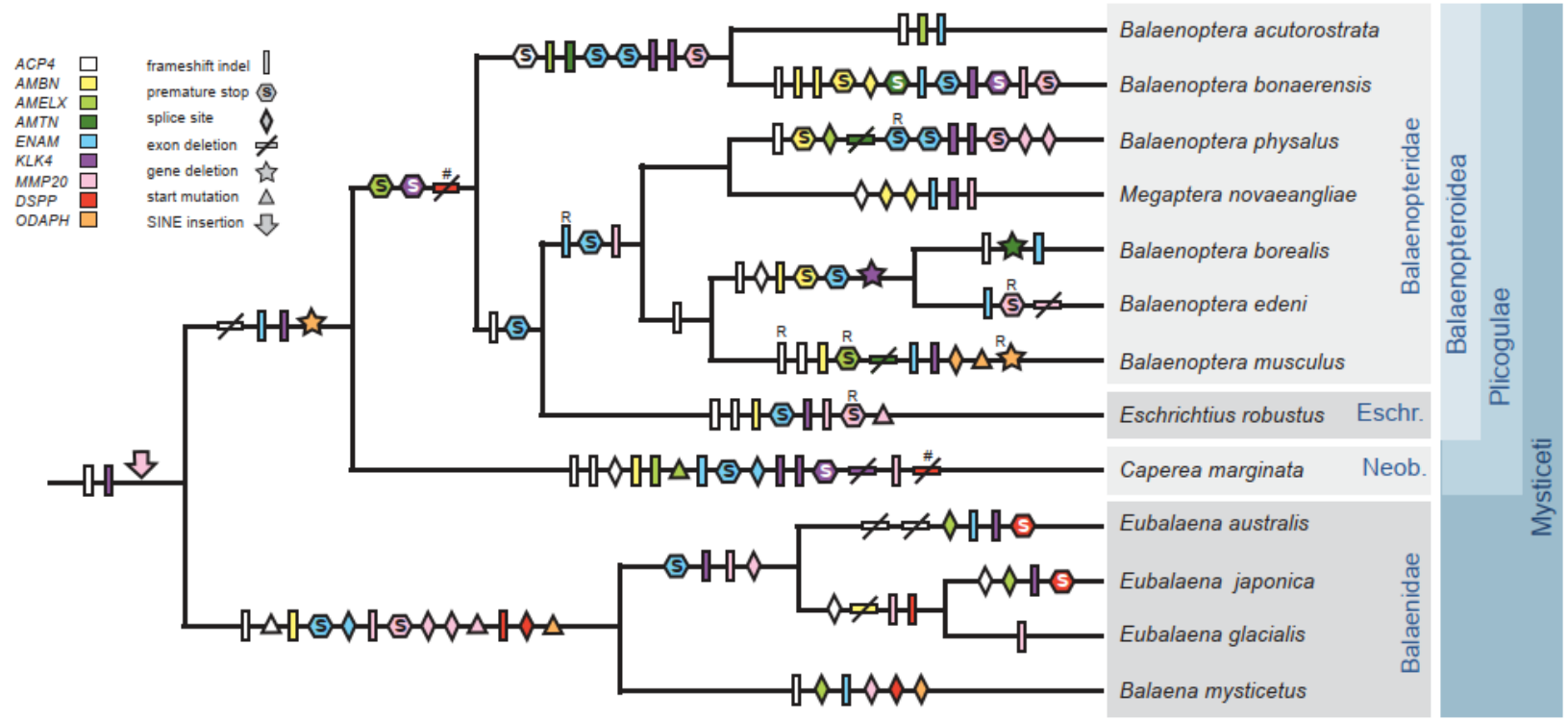
Table 1.6. Inactivation times (Ma) for seven enamel genes on the stem Mysticeti branch and two dentin/tooth genes on the stem Mysticeti + stem Balaenidae branches

Gene concatenation	CF1				CF2				Mean Inactivation Time
	dN/dS estimated		dN/dS = 1		dN/dS estimated		dN/dS = 1		
	1 syn rate	2 syn rates	1 syn rate	2 syn rates	1 syn rate	2 syn rates	1 syn rate	2 syn rates	
Seven enamel genes	33.63	32.79	36.01	35.73	33.50	32.63	36.38	36.24	34.62
Two dentin genes	25.71	23.40	18.94	17.06	22.53	20.28	16.54	15.06	19.94

Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2; dN = nonsynonymous substitutions per nonsynonymous site; dS = synonymous substitutions per synonymous site; Ma = millions of years ago, syn = synonymous.

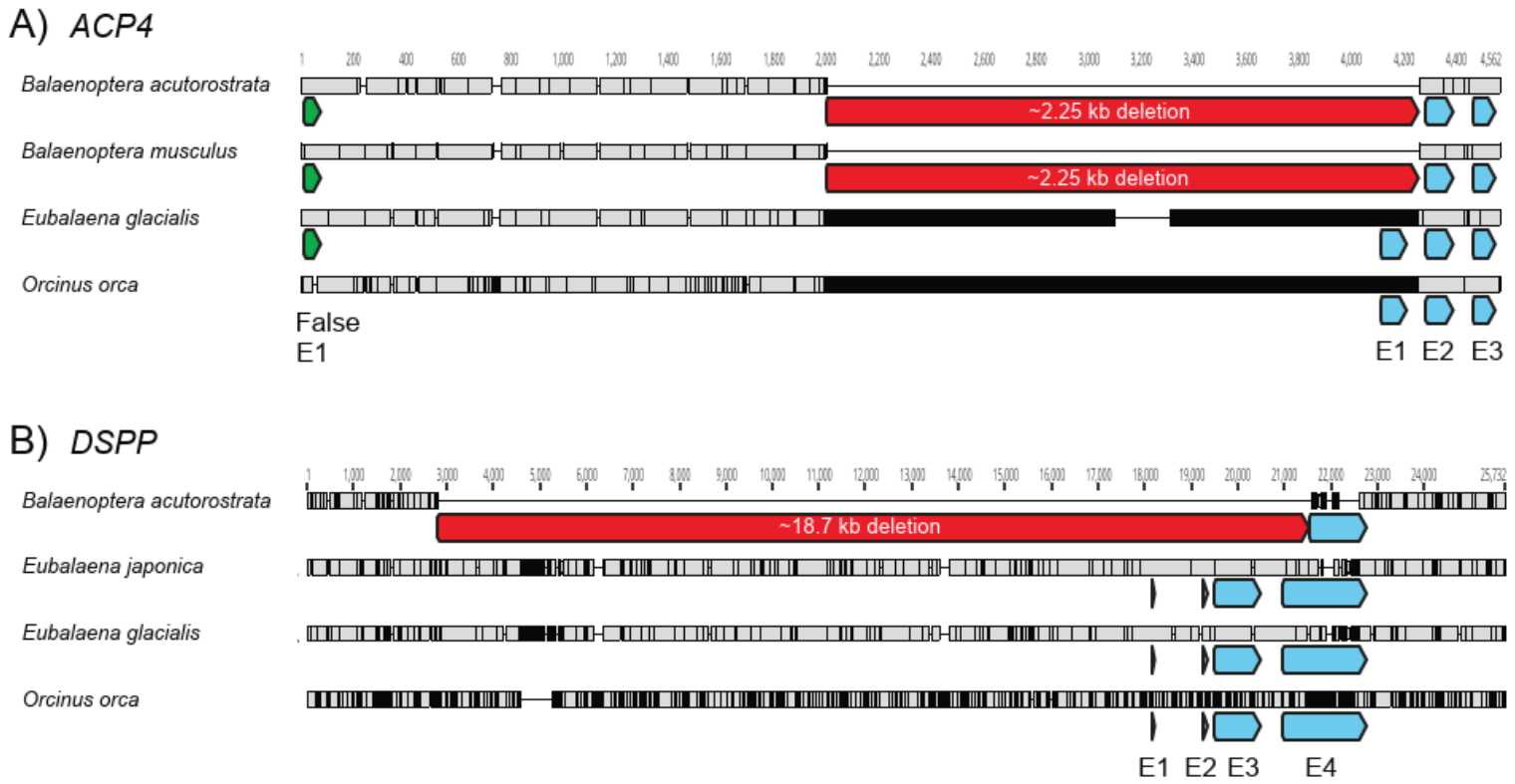
Figures

Figure 1.1. Mapping of inactivating mutations in tooth genes on mysticete tree.



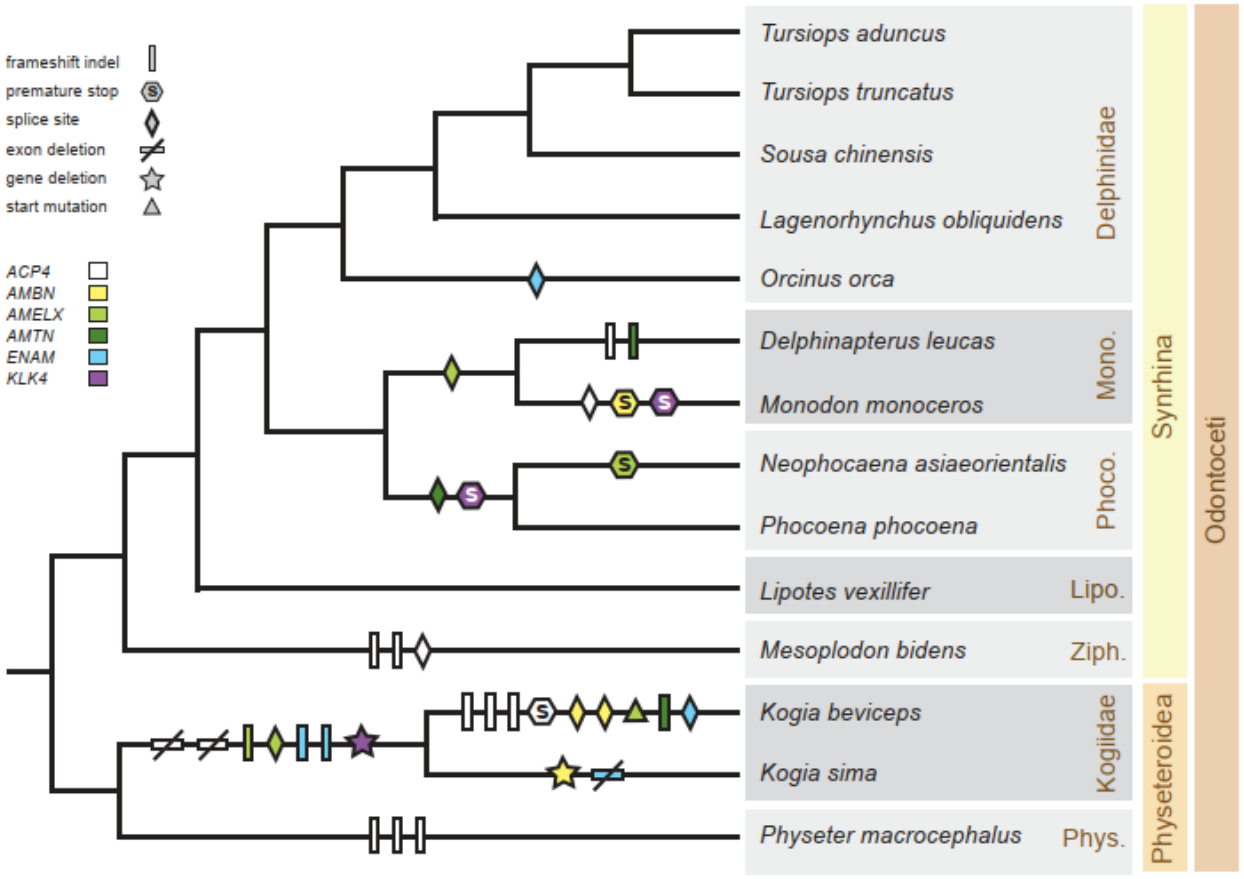
Mapping of inactivating mutations in tooth genes onto a species tree for mysticetes from McGowen et al. (2020a). Inactivating mutations were optimized with DELTRAN. Reversals of inactivating mutations are indicated by an uppercase R above a symbol. In some cases, reversals may be due to lineage sorting of ancestral polymorphism (Springer et al., 2016a) or gene flow among lineages (Árnason et al., 2018). Exons 1-3 and part of exon 4 of *DSPP* are deleted in Plicogulae. This deletion is marked as convergent by pound signs (#) on the stem Balaenopteroidea and *Caperea marginata* branches, but a single deletion that originated in the common ancestor of Plicogulae may instead explain the missing exons (see main text). Abbreviations: Eschr. = Eschrichtiidae, Neob. = Neobalaenidae.

Figure 1.2A-B. Examples of exon deletions in mysticete tooth genes (*ACP4* and *DSPP*).



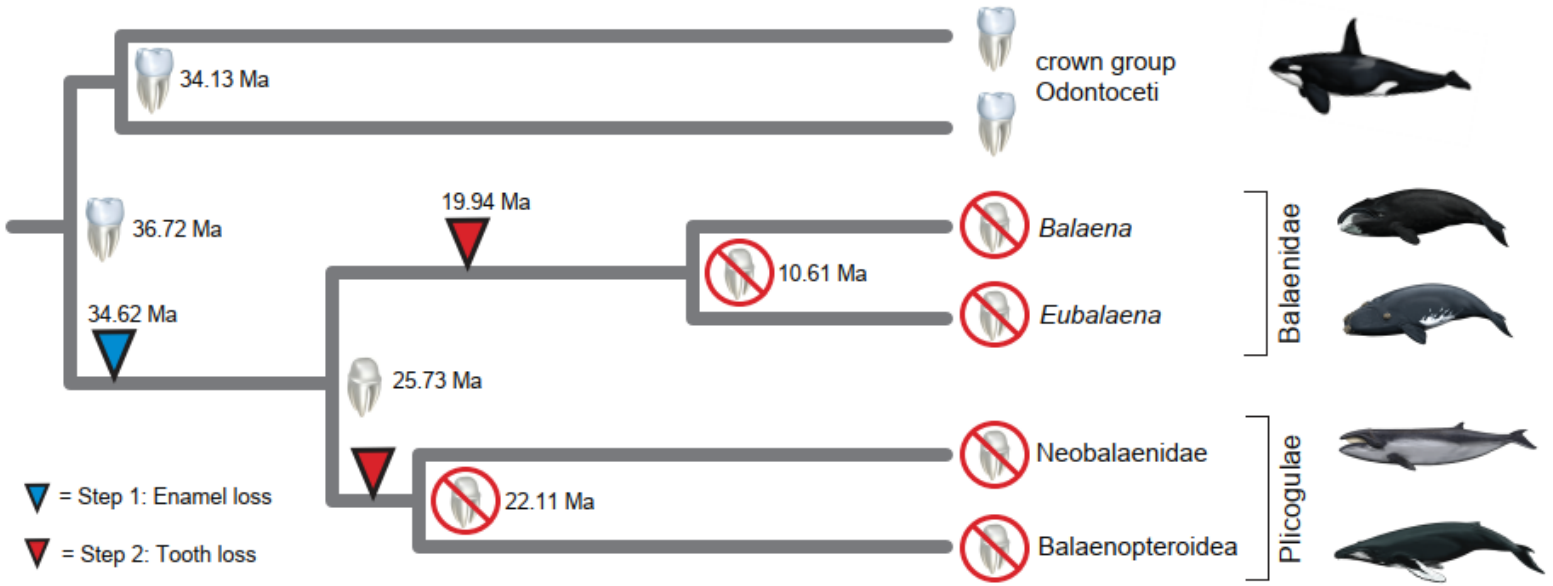
Examples of exon deletions in mysticete tooth genes. **A)** Deletion of exon 1 (E1) of the *ACP4* gene in Balaenopteroidea. The deletion (in red) spans ~2.25 kb and includes ~2.1 kb of sequence that is upstream to exon 1, all of exon 1, and 43-bp at the 5' end of intron 1. Exons are shown in blue. Mu et al. (2021) reported an incorrect (false) version of exon 1 (in green) in several mysticetes based on misannotations in GenBank. **B)** Deletion of exons 1-3 and part of exon 4 of *DSPP* in a representative balaenopteroid (*Balaenoptera acutorostrata* [common minke whale]). Exons 1–3 and part of exon 4 are also deleted in the neobalaenid *Caperea marginata* (not shown). The deletion spans ~ 18.7 kb in *B. acutorostrata*.

Figure 1.3. Mapping of inactivating mutations in tooth genes on odontocete tree.



Mapping of inactivating mutations in enamel-specific genes onto a species tree for odontocetes from McGowen et al. (2020a). Inactivating mutations were optimized with DELTRAN. Abbreviations: Mono. = Monodontidae, Phoco. = Phocoenidae, Lipo. = Lipotidae, Ziph. = Ziphiidae, Phys. = Physeteridae.

Figure 1.4. Two-step model for tooth loss in extant mysticetes.



Two-step model for tooth loss in extant mysticetes based on shared inactivating mutations and gene inactivation dating. Step 1: Enamel was lost from the teeth early on the stem mysticete branch (arrowhead). Shared inactivating mutations in three genes (ACP4, KLK4, MMP20) provide support for enamel loss on the stem mysticete branch, and gene inactivation dating based on seven enamel-related genes places enamel loss at ~35 million years ago. Step 2: Dentin and teeth were lost independently on the stem Balaenidae and stem Plicogulae branches, respectively (red arrowheads). Shared inactivating mutations provide support for dentin/tooth loss on these two branches, and gene inactivation dating places the loss of teeth at ~20 MYA on the stem Balaenidae branch. Gene inactivation dating was not applied to the stem Plicogulae branch because exons 1-3 of DSPP are deleted in this clade. Enamel capped teeth, enamelless teeth, or absence of postnatal teeth are indicated for terminal taxa and at internal nodes of the tree. Divergence dates at nodes are from McGowen et al. (2020).

Chapter 2: Release from Selection Pressure on Enamel Genes in Toothed Whales (Odontoceti) with Degenerative Enamel Phenotypes

Abstract

Different species of toothed whales (Odontoceti) exhibit a variety of tooth forms and enamel types. Some odontocete species have highly prismatic enamel with Hunter-Schreger bands whereas enamel is vestigial or entirely lacking in other species. At the molecular level, previous studies have documented inactivating mutations in as many as seven enamel-specific genes in different odontocete species that lack complex enamel with well-developed prisms. However, it remains to be determined if the complexity of the enamel in odontocetes is correlated with the strength of purifying selection on enamel-specific genes. In this chapter, I examine seven enamel-specific genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) in representatives of all odontocete families and provide a comprehensive catalog of inactivating mutations in species with one or more inactivated enamel genes. I also perform selection (dN/dS) analyses on these genes and test for a correlation between enamel complexity and selection intensity on the previously mentioned enamel-specific genes. Several genera and higher-level clades of odontocetes possess shared inactivating mutations, which suggests shared loss of these genes in their common ancestors. Additionally, selection analyses on enamel genes revealed relaxed or positive selection on many stem branches suggesting that a release from selection pressures or directional selection, respectively, occurred in the ancestry of those clades. Linear regression analyses revealed a strong correlation between selection intensity and enamel complexity, where dN/dS values are negatively correlated with

values for enamel complexity. Stronger purifying selection (low dN/dS) is found in taxa with more complex enamel and weaker purifying selection (higher dN/dS) occurs in taxa with less complex enamel or enamelless teeth.

2.1. Introduction

Cetacea is the well-known taxonomic group that includes whales, dolphins, and porpoises, and is separated into two monophyletic clades, one that includes the massive baleen-bearing filter feeders (Mysticeti), and the other that contains a variety of toothed forms that range from sleek sharp-toothed raptorial hunters to tusk-bearing suction feeders (Odontoceti). Although known as toothed whales, different odontocetes display a variety of tooth phenotypes that are associated with a wide array of feeding strategies and/or behavioral traits. Some of the most striking differences are evident in comparisons of the polydont, enamel-covered teeth of raptorial dolphins (Delphinidae) and porpoises (Phocoenidae) versus the teeth of dwarf and pygmy sperm whales (*Kogia*) that are covered by thin prismless enamel and the toothless condition that is found in suction feeding female beaked whales (Ziphiidae) (Bianucci and Landini, 2006; Heyning and Mead, 1996; Johnston and Berta, 2011; Werth et al. 2020).

There is also extensive variability in the internal enamel microstructure wherein different species exhibit varying degrees of enamel complexity ranging from highly prismatic enamel with Hunter-Schreger bands (HSB) to intermediate or thin prismless enamel (Ishiyama, 1987; Loch et al., 2013a; Werth et al., 2020). Enamel, the outer covering of teeth, is highly mineralized and is composed primarily of hydroxyapatite crystals. The development of these crystals is directed by enamel matrix proteins (EMPs),

which are secreted by ameloblasts that can greatly differ in their morphology in different species (Meredith et al., 2009). Prismatic enamel, which is characteristic of mammals, consists of bundles of crystals (prisms) that are separated from each other by discontinuities (interprisms). Enamel prisms can follow a straight course from the dentin-enamel junction (DEJ) to the tooth surface ("radial enamel") or follow a more complex pattern when the ameloblasts exhibit lateral wavy migration that results in prismatic enamel with HSB (Line and Novaes, 2005). The formation of HSB, with their decussation, increases the strength of enamel and makes it more resistant to fracture, a key feature in the evolution of many mammalian clades (Line and Novaes, 2005). Conversely, the enamel of most other toothed amniotes (amphibians and most reptiles) is prismless ("aprismatic"), and forms from ameloblasts that secrete EMPs from a flat surface. Aprismatic enamel lacks the added strength from decussation that characterizes prismatic enamel (Line and Novaes, 2005).

Aside from delphinids and river dolphins (Iniidae, Lipotidae, Pontoporiidae, Platanistidae), which have the most complex enamel structure among odontocetes, most other families have more simplistic or degenerative enamel with no HSB (Brownell and Herald, 1972; Ishiyama, 1987; Loch et al., 2013a; Werth et al., 2020). Simplified enamel phenotypes may have arisen in conjunction with changes in feeding strategy, changes in behavior and life history, and/or new interactions between teeth and their surrounding aquatic habitat. Among mammals, in addition to prey acquisition and food processing, teeth are also known to be used for social interactions, defense, sexual signaling, and sensing external stimuli (Loch et al., 2013b, 2015; Nweeia et al., 2014; Werth et al.,

2020). However, most of these uses for teeth are unnecessary for cetaceans. Cetaceans do not have many predators due to their large size, are generally not territorial (but see Crespo-Picazo et al., 2021), and do not need to protect breeding ranges (Miller, 2018; Uhen, 2010). Furthermore, odontocetes can communicate with echolocation. For feeding, even though most odontocetes feed by raptorial predation, they tend to swallow their prey whole rather than process their food via mastication (some exceptions are seen in the killer whale *Orcinus orca* and the Amazon River dolphin *Inia geoffrensis*) (Jett et al., 2017; Loch et al., 2015). Other odontocetes, including the narwhal and beaked whales, use facultative suction methods to feed and primarily use their teeth for display or male-male competition (Hooker, 2018; Werth et al., 2020). It has even been suggested that several odontocete taxa may use high intensity sounds to debilitate prey before capturing them, reducing the need for many strong teeth to be able to capture prey (Hooker, 2018, and references within). The reasons for the disuse of teeth in traditional mammalian ways may coincide with the overall restructuring of the odontocete jaw and dental battery, including the transition from the heterodont, diphyodont condition of the dentition in stem taxa to the homodont, monophyodont, polydont condition of the dentition in crownward odontocetes (Armfield et al., 2013; Boessenecker et al., 2017; Johnston and Berta, 2011; Loch et al., 2015; Rodrigues et al., 2019; Uhen, 2010). These changes in how teeth are shaped and replaced may also have led to simplification in the enamel *schmelzmuster* (three-dimensional arrangement of enamel types) in modern odontocetes. (Loch et al., 2015).

The differences in enamel microstructure and how teeth are structured and used in disparate taxa has been investigated for correlations. Specifically, Werth et al. (2020) examined if patterns of enamel complexity are correlated with different ecological parameters, life history variables, and/or the extent to which upper and lower teeth occlude with each other. Do conduct this, Werth et al. (2020) assigned discrete enamel organization scores to each taxon that was included in their study as follows: 1 = no enamel, 2 = prismless enamel, 3 = intermediate or irregular enamel, 4 = prismatic enamel, and 5 = Hunter-Schreger bands or other decussation (table 1 in Werth et al., 2020). These authors concluded that less complex enamel (lower rankings) is associated with feeding on softer prey with less oral processing, possessing less teeth, and having wider jaws (Werth et al., 2020). The reduced amount of mastication and mechanical demand on teeth in various odontocete species therefore may have led to relaxed selection on enamel structure and the underlying enamel-related genes (Loch et al., 2013a).

Genes involved with enamel formation (amelogenesis) include EMPs (amelogenin [*AMELX*], ameloblastin [*AMBN*], and enamelin [*ENAM*]) and acid phosphatase 4 (*ACP4*) that are secreted in the secretory stage of amelogenesis. The EMPs are ultimately degraded by enamel proteases (matrix metallopeptidase 20 [*MMP20*] and kallikrein related peptidase 4 [*KLK4*]) (Ikeda et al., 2018; Mu et al., 2021b; Smith et al, 2017) and replaced by hydroxyapatite mineral (Meredith et al., 2009; Su et al., 2021; Werth et al., 2020). Amelotin [*AMTN*] and odontogenic ameloblast-associated [*ODAM*]), in turn, are maturation stage proteins that have a role in hydroxyapatite nucleation

(Abbarin et al., 2015; Ikeda et al., 2018). The genes for two of the EMPs (*AMBN*, *ENAM*), *AMTN*, and *ODAM* are closely located on the same chromosome in the secretory calcium-binding phosphoprotein (SCPP) gene cluster, which also includes genes for dentin and bone formation; *AMELX* is located on the X-chromosome (Kawasaki, 2011; Kawasaki and Weiss, 2008).

Tooth genes such as these have been investigated for inactivating mutations in cetaceans due to the loss of teeth in mysticetes as well as the peculiarities of enamel in various odontocetes. Springer et al. (2019) reported that the tooth-related gene *ODAM* is inactivated in all the toothed whales that were examined (eight species). The current hypotheses are that *ODAM* loss is either related to the simplified enamel structure in these whales or altered antimicrobial functions of the junctional epithelium necessary for aquatic environments (Springer et al., 2019). *ODAM* is expressed in the junctional epithelium, where the outer enamel surface is in contact with the gingiva. This location matches the observed enamel phenotype of thin enamel on the tips of teeth, which erodes rapidly after tooth eruption in the taxa that have this gene inactivated (Bloodworth and Odell, 2008; Ishiyama, 1987; Plön, 2004; Springer et al., 2019).

Studies investigating enamel-specific genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) have observed a variety of inactivating mutations in some of the odontocete taxa that were examined. Most of these mutations were reported in species that have the least complex enamel phenotypes such as Sowerby's beaked whale (*Mesoplodon bidens*), narwhal (*Monodon monoceros*), sperm whales (*Kogia* spp., *Physeter macrocephalus*), and beluga (*Delphinapterus leucas*) (Meredith et al., 2009,

2011a; Mu et al., 2021a,b; Randall et al., 2021). However, mutations have also been reported in taxa with more complex enamel such as porpoises (Phocoenidae) and even the killer whale (*Orcinus orca*) that has prismatic enamel (Mu et al., 2021a,b; Randall et al., 2021). One interesting observation is that there are shared inactivating mutations in *Kogia*, Monodontidae (narwhal + beluga), and two phocoenids (*Neophocaena asiaeorientalis* + *Phocoena phocoena*). These results suggest that enamel degeneration was initiated in the common ancestors of these clades (Meredith et al., 2009; Randall et al., 2021). At the same time, enamel-specific genes have only been investigated in a small number of odontocete species and it will be illuminating to examine additional taxa.

In addition to reporting inactivating mutations, Randall et al. (2021) performed branch-specific selection analyses on enamel genes in odontocetes and observed elevated dN/dS values on the terminal branches for *Mesoplodon bidens* and *Physeter macrocephalus*, and on the stem and crown branches for *Kogia*, Monodontidae, and Phocoenidae. These dN/dS values were higher than for outgroup taxa with functional enamel-capped teeth, and indicate relaxed selection on the enamel-specific genes investigated in these odontocete lineages and clades. These species either have poor organization of enamel microstructure or no enamel, and for Physeteroidea (*Physeter* + *Kogia*), these results support the previous hypothesis for the loss of enamel complexity in their last common ancestor (Bianucci and Landini, 2006).

Given the fact that odontocete enamel is very thin and that functionally there may be fewer selective constraints on the requirements for enamel-capped teeth, I examined enamel-specific genes from a diverse array of odontocetes for evidence of inactivating

mutations and relaxed selection. To our knowledge, this study is the most extensive investigation of enamel-specific genes in toothed whales, including seven genes for 63 species of odontocetes. The 63 odontocete species comprise ~85% of the total number of living species (74) and including one or more representatives from every family. This taxon sampling also includes at least one species from each genus that Werth et al. (2020) examined in their study of odontocete enamel complexity. This overlapping sampling provides a basis for comparing selection intensity (present study) to enamel complexity as reported by Werth et al. (2020). In our study, Werth et al.'s (2020) enamel scores are used to investigate a possible correlation between enamel complexity and selection intensity (dN/dS). We also assessed whether or not there is a correlation between enamel complexity and the number of inactivating mutations. In addition to enamel complexity data from Werth et al. (2020), we scored eight taxa (four odontocetes, four outgroups) for enamel complexity based on the literature.

The analyses that are performed in this chapter allow us to address three outstanding research questions: (1) Which enamel-specific genes exhibit inactivating mutations in a broad survey of odontocete taxa that includes representatives of all extant families and the great majority (~85%) of living species? (2) Are there more inactivating mutations in odontocete taxa with less complex enamel than in odontocete taxa with more complex enamel? (3) Is there relaxed purifying selection on branches leading to taxa with low enamel complexity? We predict that there will be inactivating mutations in one or more of the enamel-specific genes of odontocete taxa with degenerative enamel, and that more mutations will have accumulated in taxa with lower enamel complexity

scores (Werth et al., 2020). We also predict that dN/dS analyses should reflect relaxed purifying selection on branches leading to species with less complex enamel. This study will provide the first comprehensive assessment of a possible correlation between enamel complexity and selection intensity on enamel-specific genes in a phylogenetic framework.

2.2 Methods

2.2.1. Gene sampling

Genes included in this study (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) were chosen based on (1) prior studies that reported inactivation of these genes in edentulous (pangolins, baleen whales, anteaters, Steller's sea cow) and enamelless taxa (aardvark, armadillo, sloth), (2) mutations in humans that cause amelogenesis imperfecta, and (3) mutagenesis gene knockout studies in mice (Emerling et al., 2023; Gasse et al., 2015; Hu et al., 2008; Kawasaki et al., 2014; Meredith et al., 2009, 2011, 2013; Mu et al., 2021; Smith et al, 2017; Springer et al., 2015; Wright et al., 2009). The tooth-related gene *ODAM* is also pseudogenized in all the toothed whales that were investigated (Springer et al., 2019). However, this gene is also inactivated in several other mammalian clades with enamel-capped teeth, so it was omitted from this study. Exon 4 of *AMELX* was not included in the analyses in this study because this exon is subject to alternative splicing and is absent in many mammals (Delgado et al., 2005; Sire et al., 2005, 2006, 2007).

2.2.2. Taxon sampling

Taxon sampling for this study included 93 species of which 63 are odontocetes, 13 are mysticetes, and 17 are terrestrial or semiaquatic cetartiodactyl outgroups.

Odontocete sampling included *Berardius arnuxii* (Arnoux's beaked whale), *Berardius bairdii* (Baird's Beaked whale), *Cephalorhynchus commersonii* (Commerson's dolphin), *Cephalorhynchus eutropia* (black dolphin/Chilean dolphin), *Cephalorhynchus heavisidii* (Heaviside's dolphin), *Cephalorhynchus hectori* (Hector's dolphin), *Delphinapterus leucas* (beluga), *Delphinus capensis* (long-beaked common dolphin, *Delphinus delphis* (short-beaked dolphin), *Feresa attenuata* (pygmy killer whale), *Globicephala macrorhynchus* (short-finned pilot whale), *Globicephala melas* (long-finned pilot whale), *Grampus griseus* (Risso's dolphin), *Hyperoodon ampullatus* (northern bottlenose whale), *Hyperoodon planifrons* (southern bottlenose whale), *Indopacetus pacificus* (Longman's beaked whale/Tropical bottlenose whale), *Inia geoffrensis* (Amazon River dolphin), *Kogia breviceps* (pygmy sperm whale), *Kogia sima* (dwarf sperm whale), *Lagenodelphis hosei* (Fraser's dolphin), *Lagenorhynchus* ("*Leucopleurus*") *acutus* (Atlantic white-sided dolphin), *Lagenorhynchus albirostris* (white-beaked dolphin), *Lagenorhynchus obliquidens* (Pacific white-sided dolphin), *Lagenorhynchus* ("*Sagmatias*") *obscurus* (dusky dolphin), *Lipotes vexillifer* (Chinese river dolphin/baiji), *Lissodelphis borealis (*northern right whale dolphin*), Lissodelphis peronii* (southern right whale dolphin), *Mesoplodon bidens* (Sowerby's beaked whale), *Mesoplodon bowdoini* (Andrew's beaked whale), *Mesoplodon carlhubbsi* (Hubb's beaked whale), *Mesoplodon densirostris* (Blainville's beaked whale), *Mesoplodon europaeus* (Gervais' beaked whale), *Mesoplodon ginkgodens* (ginkgo-toothed beaked whale), *Mesoplodon grayi* (Gray's beaked whale), *Mesoplodon layardii* (strap-toothed whale), *Mesoplodon mirus* (True's beaked whale), *Mesoplodon perrini* (Perrin's beaked whale),

Mesoplodon peruvianus (pygmy beaked whale/Bandolero/Peruvian beaked whale), *Mesoplodon stejnegeri* (Stejneger's beaked whale), *Monodon monoceros* (narwhal), *Neophocaena asiaeorientalis* (Yangtze finless porpoise), *Orcaella brevirostris* (Irrawaddy dolphin), *Orcaella heinsohni* (Australian snubfin dolphin), *Orcinus orca* (killer whale), *Peponocephala electra* (melon-headed whale), *Phocoena phocoena* (harbor porpoise), *Phocoena sinus* (vaquita), *Physeter macrocephalus* (sperm whale), *Platanista gangetica* (Ganges River dolphin), *Platanista minor* (Indus River dolphin), *Pontoporia blainvillei* (Franciscana dolphin/La Plata dolphin), *Pseudorca crassidens* (False killer whale), *Sousa chinensis* (Indo-Pacific humpback dolphin), *Stenella attenuata* (Pantropical spotted dolphin), *Stenella clymene* (Clymene dolphin), *Stenella coeruleoalba* (striped dolphin), *Stenella frontalis* (Atlantic spotted dolphin), *Stenella longirostris* (Eastern spinner dolphin), *Steno bredanensis* (rough-toothed dolphin), *Tasmacetus shepherdi* (Shepherd's beaked whale), *Tursiops aduncus* (Indo-Pacific bottlenose dolphin), and *Tursiops truncatus* (common bottlenose dolphin), *Ziphius cavirostris* (Cuvier's beaked whale).

Mysticete sampling included *Balaena mysticetus* (bowhead whale), *Balaenoptera acutorostrata* (common minke whale), *Balaenoptera bonaerensis* (Antarctic minke whale), *Balaenoptera borealis* (sei whale), *Balaenoptera edeni* (Bryde's whale), *Balaenoptera musculus* (blue whale), *Balaenoptera physalus* (fin whale), *Caperea marginata* (pygmy right whale), *Eschrichtus robustus* (gray whale), *Eubalaena glacialis* (North Atlantic right whale), *Eubalaena australis* (southern right whale), *Eubalaena japonica* (North Pacific right whale), and *Megaptera novaeangliae* (humpback whale).

Outgroup taxa included *Bison bison* (American bison), *Bos mutus* (wild yak), *Bubalus bubalis* (water buffalo), *Camelus bactrianus* (Bactrian camel), *Capra hircus* (domestic goat), *Catagonus wagneri* (Chacoan peccary), *Choeropsis liberiensis* (pygmy hippopotamus), *Elaphurus davidianus* (Pere David's deer), *Giraffa camelopardalis* (giraffe), *Hippopotamus amphibius* (river hippopotamus), *Moschus moschiferus* (Siberian musk deer), *Odocoileus virginianus* (white-tailed deer), *Okapia johnstoni* (okapi), *Ovis aries* (domestic sheep), *Sus scrofa* (wild boar), *Tragulus javanicus* (Java mouse-deer), and *Vicugna pacos* (alpaca).

In addition to the full data set with 93 species, we performed a subset of analyses on a pruned data set than contained 37 taxa (33 odontocetes and four cetartiodactyl outgroups). The 33 odontocetes in the pruned data set comprise species with molecular data for enamel-specific genes plus scores for Werth et al.'s (2020) index of enamel complexity. This data set includes at least one species from each genus and each family in Werth et al.'s (2020) original dataset (Table 2.1, also see below).

2.2.3. BLAST searches and collection of molecular data

DNA sequences for seven different tooth genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) were obtained from (1) assembled genomes at NCBI (https://www.ncbi.nlm.nih.gov/) and The Bowhead Whale Genome Resource (http://www.bowhead-whale.org/), (2) raw sequence reads at NCBI's Sequence Read Archive (SRA), (3) newly generated Illumina whole-genome sequence data (John Gatesy and Mark Springer, unpublished), and (4) assemblies provided on DNA Zoo (https://www.dnazoo.org/, Dudchenko et al., 2017) (see Supplemental Table S2.1 for

accession numbers and/or sources for each species). NCBI's RefSeq and Nucleotide databases were searched using keywords for all seven genes in conjunction with taxon names for four reference species (*Capra hircus*, *Camelus bactrianus*, *Orcinus orca*, *Tursiops truncatus*). Sequences for each reference species were then imported into Geneious Prime (current version 2022.2.2, https://geneious.com) (Kearse et al., 2012), aligned with MAFFT (Katoh and Toh, 2008), and cross-checked against each other for consistent annotations. Sequences for additional species were collected through NCBI's Nucleotide Basic Local Alignment Search Tool (BLAST), which was used to search both assembled and unassembled genomes using the whole-genome shotgun (WGS) and SRA databases, respectively. Each BLAST search employed a query sequence from a closely related species. Megablast was used for highly similar sequences (e.g., taxa in same family), whereas Blastn was used for less similar sequences (e.g., taxa in different families). Top-scoring BLAST results were imported into Geneious Prime. Sequences obtained through the SRA database were assembled using Geneious Prime's 'Map to Reference' approach, where the reference sequence was a closely related species to the SRA taxon. We allowed for a maximum mismatch of 10% per read and required a minimum of two reads for base calling with a consensus threshold of 65%.

2.2.4. Alignments and inactivating mutations

Complete protein-coding sequences and introns were aligned in Geneious Prime using MAFFT (Katoh and Toh, 2008). Sequences were manually spot-checked for alignment errors using AliView version 1.28 (Larsson, 2014). Alignments were examined for inactivating mutations (frameshift insertions and deletions [indels], start and stop

codon mutations, premature stop codons, splice site mutations, exon deletions, and whole gene deletions), all of which were annotated in Geneious Prime. Synapomorphic inactivating mutations were validated by their occurrence in two or more taxa. Autapomorphic inactivating mutations in odontocete taxa for which we generated Illumina data were validated by their occurrence in multiple Illumina reads. In the case of autapomorphic inactivating mutations that were detected in a genome sequence of a single taxon, we BLASTed (megablast) genomic segments containing these mutations to the associated Illumina reads on NCBI's Sequence Read Archive (Supplemental Table S2.2). Mutations with allelic variation (Table 2.2) were each counted as 0.5 mutations. These were determined by verifying heterozygous mutations in mapped sequence alignments, with a criteria of at least 10X coverage and each allele consisting of at least 40% and at most 60% of the reads (Supplemental Table S2.2). These mutations were removed for codeml selection analyses similar to non-heterozygous mutations (stop codons coded as "NNN", deletions as missing data) (see 2.2.9).

Inactivating mutations were mapped onto species trees (see 2.2.7) with delayed transformation (DELTRAN) character optimization (Figure 2.1). DELTRAN mapping was performed on the full data set for 93 taxa and the pruned data set for 37 taxa. For the latter, the number of inactivating mutations was tabulated for each branch and divided by the length of the branch in millions of years on the corresponding species timetree (see 2.2.7). This resulted in mutations per million years for each branch and these values were used in linear regression analyses to test for correlations between the number of inactivating mutations per million years and enamel complexity (see 2.2.10).

2.2.5. Phylogenetic analyses

Gene trees were constructed from the aligned protein-coding sequences with maximum likelihood using the program RAxML version 8.2.11 in Geneious Prime (raxmlHPC-SSE3-MAC) (Stamatakis, 2014). Rapid bootstrapping (500 pseudoreplicates) and a search for the best tree were performed in the same analysis (Stamatakis et al., 2008). We employed the GTRGAMMA option, which implements the GTR + Γ model of DNA sequence evolution.

2.2.6. Collection of data for enamel complexity

We compiled an enamel complexity data set (Werth Enamel Complexity, Table 2.1) for a subset of species from our full data set. Taxon sampling for this data set included 29 odontocetes with enamel complexity scores from Werth et al. (2020, table 1) plus four additional odontocetes and four cetartiodactyl outgroups for which we scored enamel complexity based on the literature. The additional odontocetes are *Monodon monoceros, Neophocaena asiaeorientalis, Platanista minor*, and *Kogia sima*. The cetartiodactyl outgroups, in turn, are *Hippopotamus amphibius*, *Bos mutus*, *Sus scrofa*, and *Camelus bactrianus.* The enamel complexity categories used in this study were the same as in Werth et al. (2020): 1 = no enamel, 2 = prismless enamel, 3 = intermediate or irregular enamel, 4 = prismatic enamel, and 5 = enamel with Hunter-Schreger bands or other decussation. For taxa that were scored by Werth et al. (2020), we used the same scores, including taxa that were scored with polymorphic enamel categories based on indeterminate results or specimens that showed notable variation. For odontocetes that were not scored by Werth et al. (2020), the narwhal (*M. monoceros*) was scored as

category 1 because this taxon lacks enamel (Ishiyama, 1987). The narrow-ridged finless porpoise (*Neophocaena asiaeorientalis*) was assigned the same score (3) as the finless porpoise (*Neophocaena phocaenoides*) given the very close relationship of these taxa. *N. asiaorientalis* was previously a subspecies of *N. phocaenoides* (Jefferson and Hung, 2004), although it has recently been elevated to a full species based on genomic data that suggests it is an incipient species (Zhou et al., 2018). The Indus River dolphin (*Platanista minor*) was assigned a score of 5, which is the same as its sister-species the Ganges River dolphin (*Platanista gangetica*) (Werth et al., 2020). The Indus River dolphin was described as having well-developed enamel with undulating HSB (Cooper and Maas, 2009). For *Kogia*, we scored the dwarf sperm whale (*K. sima*) as category 2, the same category given to its sister species *Kogia breviceps*. Different studies have reported that this species has lost the enamel on its teeth or has a thin layer of enamel that quickly wears away (Bloodworth and Odell, 2008; Handley, 1966; McAlpine, 2018; Plön, 2004). All of the outgroup taxa were scored as category 5 (Alloing-Séguier et al., 2014; Berkovitz and Shellis, 2018; Radhi, 2017; Sathe, 2000).

2.2.7. Species trees

The 93-taxon species tree that was used for Coevol analyses and mapping inactivating mutations with DELTRAN optimization was constructed from McGowen et al. (2020a) with additions from McGowen et al. (2009: *Berardius arnuxii*, *Cephalorhynchus eutropia*, *Cephalorhynchus hectori*, *Hyperoodon planifrons*, *Phocoena sinus*, *Platanista minor*), Zurano et al. (2019: *Indopacetus pacificus*), and Chehida et al. (2020: *Neophocaena asiaeorientalis*) for odontocete taxa that were not included in

McGowen et al. (2020a). Note that McGowen et al. (2009) and McGowen et al. (2020a) used different names for several species that were included in our analysis: *Delphinus capensis* (McGowen et al., 2009) = *Delphinus delphis bairdii* (McGowen et al., 2020a); *Lagenorhynchus acutus* (McGowen et al., 2009) = *Leucopleurus acutus* (McGowen et al., 2020a). Also, *Lagenorhynchus obliquidens* and *L. obscurus* in the present study = *Sagmatias obliquidens* and *S. obscurus* of McGowen et al. (2009; 2020a). Relationships among cetartiodactyl outgroups are based on Foley et al. (2023). We pruned the 93-taxon tree to 37 taxa for codeml analyses and ancestral reconstructions with the 37-taxon Werth Enamel Complexity data set.

Divergence times for the 37-taxon tree are based on McGowen et al.'s (2020a) table S3 "Full dataset 6-partition AR (Mean)" except for several nodes that were not included in McGowen et al.'s (2020a) timetree. Specifically, divergence times for nodes 7, 42, and 63 were taken from McGowen et al. (2009, mean dates from table 2), and the divergence time for node 52 was taken from Chehida et al. (2020, table S5) (see Figure 2.2 for node numbers). The 93-taxon and 37-taxon species trees, with divergences times for the latter, are provided in Supplemental Files S2.3.

2.2.8. Ancestral reconstructions and enamel complexity values for branches

Ancestral reconstructions for Werth Enamel Complexity were performed in PAUP (version 4.0a, Swofford, 2003) and Mesquite (version 3.70, Maddison and Maddison, 2023). PAUP was employed with DELTRAN, accelerated transformation (ACCTRAN), and most parsimonious reconstruction sets (MPR) settings. PAUP's MPR setting and Mesquite both allow for multiple (equivocal) state assignments for a given

node. Taxa that were scored as polymorphic for Werth Enamel Complexity (e.g., 1/2) were also scored as polymorphic in both PAUP and Mesquite ancestral reconstruction analyses. Enamel complexity was treated as an ordered character for all analyses in both programs. External and internal branches were assigned enamel complexity scores based on the enamel scores for the nodes on the basal and apical sides of each branch (Figure 2.3A–C). If the nodes at both ends of a branch had the same score for Werth Enamel Complexity, then the branch was assigned the same score (e.g., 3 → 3 = 3). If the nodes at the end of a branch had different scores, then the branch was assigned the mean value of these two scores (e.g., 4 → 2 = 3). External nodes with polymorphic character states and internal nodes that were reconstructed as equivocal (i.e., multiple states) were assigned mean values between the two states prior to calculating a value for the relevant branch (e.g., 1 → 1/2 = 1 → 1.5 = 1.25; 1/2 → 1/2 = 1.5 → 1.5 = 1.5).

2.2.9. Selection analyses

Selection analyses (dN/dS) were conducted with two different programs, the codeml program of PAML (version 4.9j, Yang, 2007) and Coevol (version 1.6, Lartillot and Poujol, 2011). Selection analyses were performed with a concatenation of seven enamel-specific genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) that serve as a proxy for enamel. Importantly, analyses with the seven-gene concatenation are less impacted by sampling error than are analyses with individual genes. The analysis with Coevol included 93 taxa and 9987 base pairs (bp) whereas the analyses with codeml employed a reduced data set that included 37 taxa (i.e., Werth Enamel Complexity taxa)

and 9570 bp. We used a rooted species tree for Coevol and an unrooted species tree for codeml (2.2.7. Species trees).

For codeml analyses, we employed a free-ratio model and a branch-specific codon model that used branch categories for Werth Enamel Complexity (Table 2.1). Enamel complexity values for internal branches were based on ancestral reconstructions (Figure 2.3A–C). The analyses that employed ACCTRAN, DELTRAN, and MPR/Mesquite ancestral reconstructions were conducted with 12, 9, and 14 branch categories, respectively, based on the number of possible categories that were non-empty and contained branches.

Selection analyses with codeml were performed with two codon frequency models, CF1 and CF2 (Yang, 2007). CF1 estimates codon frequencies from mean nucleotide frequencies across all three codon positions, whereas CF2 estimates frequencies at each of the individual codon positions. Codon positions are absent in pseudogenes so it is important to verify that analyses without base compositional differences at different codon positions (i.e., CF1) yield results that are similar to results that are obtained with a codon frequency model that allows for base compositional differences at 1st, 2nd, and 3rd codon positions (i.e., CF2) (Randall et al., 2022; Springer et al., 2023). All frameshift insertions were deleted prior to performing dN/dS analyses. In addition, premature stop codons were recoded as missing data as required for codeml analyses.

By contrast with codeml, Coevol utilizes a Bayesian approach and provides a visual representation of dN/dS ratio estimates varying across a phylogeny (Lartillot and

Delsuc, 2012). The Coevol analysis was conducted on the 93-taxon species tree with 10 fossil calibrations from McGowen et al.'s (2020) table 2. The age of the root node for this tree (65.83 Ma, Cetartiodactyla) was taken from McGowen et al.'s (2020) table S3, and was also used as the standard deviation for this analysis. We employed the *dsom* procedure that uses a codon model with two *a priori* independent values (dS and dN/dS) as *a priori* independent variables. The data set was run for at least 200 cycles, and dN/dS values were sampled once every cycle. The burnin was determined with the *tracecomp* command, which was used to check for MCMC convergence by monitoring effective sample size.

2.2.10. Statistical analyses

To test the significance of the relationships between Werth Enamel Complexity values and two different measures of selection at the molecular level (number of inactivating mutations per million years, dN/dS values), we analyzed our dataset using a linear regression model. Analyses were performed with branch categories based on ancestral reconstructions. Regression analyses employed dN/dS values from both free-ratio and branch-specific selection analyses. Analyses using free-ratio dN/dS values omitted branches that were either 0.0001 (no nonsynonymous substitutions) or 999 (synonymous mutations). A confidence interval and R-squared value were calculated for each analysis to examine the strength of the correlation. The statistical analyses were run in RStudio (2022.07.2) using the libraries plyr, dplyr, ggplot2, and sjPlot with the linear regression function lm() (R Core Team, 2022). Molecular variables were selected as the

independent variable and enamel complexity was selected as the dependent variable because enamel genes encode for proteins that produce enamel.

2.3. Results

2.3.1. Alignments and gene trees

Sequence alignments range from 645 bp (*AMTN*) to 3872 bp (*ENAM*). The number of sequences in the individual gene alignments ranges from 88 (*KLK4*) to 93 (*ACP4, AMELX, ENAM, MMP20*). Intact coding sequences were recovered for all of the terrestrial and semiaquatic cetartiodactyl outgroup taxa that were sampled, as well as for most of the odontocete species. Notable exceptions include the absence of *KLK4* in both *Kogia* species and the absence of *AMBN* in *Kogia sima*. Sequences for three mysticetes are also missing for *KLK4* (Table S2.1).

Table 2.3 summarizes the presence or absence of 30 well-supported clades in Cetartiodactyla (Gatesy et al., 2013; Hassanin et al., 2012; McGowen et al., 2009, 2014, 2020a; Meredith et al., 2011b) for all seven gene trees. A majority of the well-supported clades were recovered by all seven gene trees with a mean of 25.43 clades per gene tree (~86%). *ENAM* has the longest coding sequence among enamel-related genes (3522 bp) and the *ENAM* gene tree has the maximum number of recovered clades (30/30). *AMELX*, in turn, has the second shortest coding sequence (693 bp) and the *AMELX* gene tree exhibits the fewest recovered clades (22/30).

2.3.2. Inactivating mutations

Coding sequences for all of the semiaquatic and terrestrial cetartiodactyl taxa that were sampled are intact. By contrast, all seven of the enamel genes exhibit at least one

inactivating mutation in multiple odontocete species. The number of inactivating mutations in different enamel genes ranges from two (*MMP20*) to 14 (*ACP4*). Also, there are fewer mutations in taxa with more complex enamel phenotypes (Werth categories 4 and 5) and more mutations in taxa with less complex enamel phenotypes (Werth categories 1-3) (Figure 2.1, Table 2.2). Werth enamel complexity scores are 4 or 5 for all representatives of *Platanista*, Inioidea, and Delphinidae that have been scored for enamel complexity (Figure 2.1), and there are only two inactivating mutations (one synapomorphic, one autapomorphic) among 36 species that belong to these three clades. Of these, the only synapomorphic mutation is an in-frame deletion of exon 3 in the *AMTN* gene of the Ganges and Indus River dolphins (*Platanista*) that exhibit the most complex enamel phenotype (category 5). By contrast, Werth scores range from 1 to 3 for monodontids, phocoenids, ziphiids, and physeteroids that have been scored for enamel complexity, and there are 46 inactivating mutations (11 synapomorphic, 35 autapomorphic) among 27 taxa that belong to these clades. For these taxa with less complex enamel, the most inclusive synapomorphic inactivating mutations are found in Monodontidae (*Monodon monoceros* + *Delphinapterus leucas*), Phocoenidae, *Kogia*, and *Hyperoodon*. The largest inactivating mutation is a presumed whole gene deletion (*KLK4*) in *Kogia* based on the absence of BLAST hits and map to reference results. The two species of *Kogia* also share presumed deletions of exons 1–3 and 11 in *ACP4* (Figure 2.1, Table 2.2), frameshift indels in *ENAM*, and a frameshift indel and splice site mutation in *AMELX*. Both kogiids have autapomorphic mutations in *AMBN* and *K. breviceps* has a frameshift indel in *AMTN*. Hence, the only enamel gene with intact

coding sequences in both species of *Kogia* is *MMP20*. There are no shared inactivating mutations for Ziphiidae, but 12 of the 19 beaked whales included in this study have at least one gene that contains an inactivating mutation (Figure 2.1, Table 2.2). These mutations are spread across six of the seven enamel genes and the only gene that is intact in all of the ziphiid species is *AMBN*. Shared inactivating mutations are also absent in Physeteroidea, although all three physeteroids exhibit inactivating mutations in at least one gene and *K. breviceps* exhibits mutations in six of the seven enamel genes as noted above.

2.3.3. Selection analyses

Selection (dN/dS) analyses were performed on a concatenation of seven enamel genes with the Coevol and codeml programs. Figure 2.4 shows the results of the Coevol selection analysis and provides a visual portrayal of selection intensity on the concatenation of enamel genes in 93 cetartiodactyls. Relaxed selection intensity (red and orange-red branches) is most evident in toothless mysticetes, but is also apparent in three groups of odontocetes (physeteroids, phocoenids, monodontids) that are characterized by low enamel complexity (Werth scores of 1–3). Further, the gradual increase of relaxed selection on stem and then crown branches is very apparent in both physeteroids and monodontids (Figure 2.4). Ziphiids that have been investigated also have low Werth scores (1–2), but relaxed selection intensity in less evident in this clade than in the aforementioned odontocetes. By contrast, selection intensity remains higher (= lower dN/dS values) in inioids and delphinids that have Werth scores of 4 or 5.

The results of codeml analyses for the 37-taxon dataset are summarized in Tables 2.5 and 2.6. These results are based on analyses with a free-ratio model as well as three different sets of branch categories based on ancestral reconstructions with ACCTRAN, DELTRAN, and MPR/Mesquite (Figure 2.3A–C). The codeml results were incorporated into regression analyses that are reported in the next section.

2.3.4. Regression analyses

We used linear regression to examine the relationships between enamel complexity and two different measures of relaxed selection/neutral evolution – the number of inactivating mutations per million years (Table 2.4) and dN/dS values (Tables 2.5 and 2.6). The results of these analyses are shown in Figures 2.5A–C and 2.6A–L, respectively. The results for the regression of enamel complexity and the number of mutations per million years are sensitive to the method that was used to reconstruct enamel complexity on different branches. Specifically, the results for the ACCTRAN ($p = 0.0428$) and MPR/Mesquite (0.0157) reconstructions are statistically significant whereas the DELTRAN reconstruction (0.0737) is insignificant.

All of the regressions analyses investigating the relationship between enamel complexity and dN/dS values are statistically significant (Figure 2.6A–L). P-values for six analyses with free-ratio dN/dS values range from $p = 0.0003$ (DELTRAN, CF2) to $p = 0.0008$ (ACCTRAN, CF1). P values with binned dN/dS values for different branch categories based on enamel complexity are several orders of magnitude smaller than for analyses with free ratios and range from 6.377e-15 (DELTRAN, CF1) to 2.642e-13 (ACCTRAN, CF1).

2.4. Discussion

2.4.1. Inactivating mutations in enamel genes

Odontocetes display a variety of tooth morphologies and enamel phenotypes. Some river dolphins (i.e., *Platanista* spp., *Inia geoffrensis*) exhibit polydont teeth with highly prismatic enamel whereas the narwal (*Monodon monoceros*) possesses only a single enamelless tooth ("tusk") that is usually found only in males. Other taxa exhibit intermediate conditions in enamel complexity, e.g., porpoises (Phocoenidae) exhibit intermediate enamel with amorphous crystallite aggregations that run in multiple directions without a clear arrangement (Werth et al., 2020). However, the connection between the various enamel phenotypes in odontocetes and the integrity of the underlying genes responsible for proper enamel formation has only been investigated in a limited number of odontocete species. Meredith et al. (2009, 2011a) and Mu et al. (2021a) documented inactivating mutations in two enamel genes (*ENAM, MMP20*) in one or both species of *Kogia*, and Mu et al. (2021a) reported inactivating mutations in *ACP4* in *K. breviceps*. Mu et al. (2021b), in turn, documented premature stop codons in the *AMELX* and *KLK4* genes of the Yangtze finless porpoise (*Neophocaena asiaeorientalis*). Most recently, Randall et al. (2022) examined seven enamel genes in 13 mysticetes and 14 odontocetes and expanded the catalog of inactivating mutations in both groups. In the case of odontocetes, Randall et al. (2022) provided evidence for inactivating mutations in four additional genes (*AMBN, AMELX, AMTN, KLK4*) that are found in one or both species of *Kogia*. These authors also documented a heterozygous splice site mutation in the *ENAM* gene of the killer whale (*Orcinus orca*), a shared splice site mutation in the

AMELX gene of both monodontids (*Delphinapterus leucas, Monodon monoceros*), autapomorphic inactivating mutations in assorted tooth genes in both monodontids, shared inactivating mutations in the *AMTN* and *KLK4* genes of two phocoenids, and multiple inactivating mutations in the *ACP4* gene of both Sowerby's beaked whale (*Mesoplodon bidens*) and the sperm whale (*Physeter macrocephalus*). To further investigate the association between the molecular components of enamel production and the morphology of odontocete enamel, taxon sampling was thoroughly expanded in the present study to include 63 odontocetes. DELTRAN mapping documented a total of 51 inactivating mutations in these taxa. The majority of inactivating mutations that were reported in previous studies (Meredith et al., 2009, 2011a; Mu et al., 2021a,b; Randall et al., 2022) were confirmed here, but there are a few exceptions. First, Mu et al. (2021b) and Randall et al. (2022) both reported a stop codon in exon 6 of *AMELX* in *Neophocaena asiaeorientalis*. However, this stop codon is a CAG instead of a TAG in the individual of *N. asiaeorientalis* that we included in our study. Second, we examined a more complete genome of *Kogia breviceps* than Randall et al. (2022) to determine if we could find any remnants of *ACP4* exons 1-3 and 11 that were reported as missing by Randall et al. (2022). We did not find any remnants of these exons in the second individual of *K. breviceps*, but did find that a frameshift deletion in exon 7 is present in the individual examined by Randall et al. (2022) but not in the individual examined in this study. Finally, differences in Randall et al.'s (2022) 44-taxon alignment for *ACP4* and our 93-taxon alignment for this gene resulted in a slightly upstream location for a 1-

bp insertion in exon 10 of *K. breviceps* that also nullified a putative stop codon in the original reading frame of this gene.

Statistical analyses also support the conclusion that inactivating mutations are more plentiful in lineages with less complex enamel. Specifically, regression analyses demonstrate that the number of inactivating mutations per million years is negatively correlated with enamel complexity as measured by Werth Enamel Complexity scores. Odontocete taxa with lower enamel complexity scores having accumulated more mutations than taxa with higher enamel complexity scores. Selection intensity is also negatively correlated with enamel complexity. Specifically, higher dN/dS values are associated with lower Werth Enamel Complexity scores and there is a trend of decreasing dN/dS values as enamel complexity increases.

2.4.2. Dolphins: categories 4–5

Odontocetes with the most complex enamel examined for this study include river and oceanic dolphins (delphinids, *Inia geoffrensis*, *Pontoporia blainvillei*, and *Platanista*: categories 4–5; *Lipotes vexillifer* included but not categorized). These species overall exhibit a lower number of inactivating mutations and less mutations per million years compared to the less complex enamel categories, possessing just two inactivating mutations in 35 species. One of these mutations is a heterozygous splice site mutation in intron 4 of the *ENAM* gene in the killer whale (*Orcinus orca*, category 4) and the second is an in-frame deletion of *AMTN* exon 3 in *AMTN* in both *Platanista* species (Ganges River dolphin [*Platanista gangetica*], Indus River dolphin [*Platanista minor*], both category 5). This latter mutation is the more surprising result as these species are reported

to have among the most complex enamel among species examined. However there have been reports of a thin prismless layer in the outer enamel surface in this species which may explain the mutation that is observed (Loch et al., 2015). The free-ratio model using the Werth only concatenated dataset showed slightly elevated dN/dS values on the stem *Platanista* branch when compared to the outgroups (Table 2.5). These results suggest slight relaxed selection occurring on these genes, most likely somewhat recently over their long history. Both *Platanista* have relatively short terminal branches only diverging around 0.52 million years ago, so there may have not been many changes that occurred on these branches which is the reason the free-ratio dN/dS analyses resulted in undefined for both *Platanista* species (999, CF1 and CF2). Indeed, these values are based on only a few substitutions, with a total of 4.2 (CF1 and CF2) nonsynonymous mutations and 0.0 (CF1 and CF2) synonymous mutations between both species. The Chinese river dolphin (also known as the baiji or Yangtze River dolphin, *Lipotes vexillifer*) wasn't included in the Werth category analyses since the enamel microstructure is not known from low specimen counts due to its probable recent extinction. However, *L. vexillifer* has been reported as having a wrinkled rugose enamel phenotype, similar to the closely related Amazon River dolphin (*Inia geoffrensis*) (Brownell and Herald, 1972). Both these species have similar dN/dS results from the Coevol analysis.

2.4.3. Porpoises: category 3

Porpoises (Phocoenidae) are known to use both ram and suction feeding techniques and have teeth that are flatter and more spade-shaped than dolphins (Berkovitz and Shellis, 2018). Porpoises that have been examined generally have prismless enamel

with a low degree of mineralization (Ishiyama, 1987; Loch et al., 2013b) although the narrow-ridged finless porpoise (*Neophocaena phocaenoides*) has been observed to have enamel prisms near the enamel-dentin junction that gradually transition to prismless enamel further away from this junction (Ishiyama, 1987). Even so, the enamel prisms are simple and interprismatic regions cannot be discerned (Ishiyama, 1987). The three porpoises included in our study (*N. asiaeorientalis, Phocoena phocoena,* and *Phocoena sinus*) share one inactivating mutation in *AMTN* and a second inactivating mutation in *KLK4*. Interestingly, Núñez et al. (2016) examined enamel maturation in *AMTN* and *KLK4* null mice. These authors concluded that AMTN and KLK4 are both essential for proper enamel maturation in mice, and that the absence of both proteins had a more severe effect than the absence of a single protein. Importantly, enamel mineral density was significantly reduced in *AMTN*$^{-/-}$ *KLK4*$^{-/-}$ mice (Núñez et al., 2016). This finding mirrors the empirical observation that a representative porpoise (*Phocoena spinipinnis* [Burmeister's porpoise]) has reduced enamel mineral density relative to dolphins and river dolphins that have intact copies of all seven enamel genes (Loch et al., 2014).

2.4.4. Monodontids: categories 1–2/3

Closely related to porpoises, Monodontidae has only two living species, the enigmatic narwhal (*Monodon monoceros*) and the "white whale" more commonly known as the beluga (*Delphinapterus leucas*). The beluga has slightly less complex enamel compared to porpoises, with thin prismless enamel on unerupted teeth and has exposed dentin at the tip of erupted teeth (Ishiyama, 1987). Werth et al. (2020) scored beluga as indeterminate from irregular and prismless enamel and notes that the enamel is extremely

thin and soft, suggesting it readily wears away throughout the animal's lifespan. In the free-ratio model the beluga had elevated dN/dS results suggesting relaxed selection (Table 2.5). The internal branch leading to Monodontidae from the free-ratio model had dN/dS values that were elevated over the outgroups suggesting recent relaxed selection occurring on this branch (Table 2.5). For the Coevol analysis, Monodontidae is slightly higher than the dN/dS values for the outgroups (Fig. 2.4). The narwhal is unique for its long enamelless tusk in males which it uses to sense external stimuli in its environment and in intrasexual selection, while females also have enamelless tusks that are instead left unerupted in the maxilla (Ishiyama, 1987; Nweeia et al., 2014). So, it is not surprising then that four of the narwhal's enamel genes have been inactivated. The relaxed selection occurring in the enamel genes for these taxa is plausible since the main diet of monodontids consists of invertebrates such as squid and shrimp that they capture with suction which is similar to taxa with even more degenerated enamel phenotypes such as sperm and beaked whales (Berkovitz and Shellis, 2018; Heyning and Mead, 1996)

2.4.5. Physeteroids: categories 1/2–2

The description of enamel in living physeteroids (Physeteroidea: *Physeter* and *Kogia*) has been debated by several authors with some suggesting a total enamelless condition with others reporting the presence of a thin layer of prismless enamel (Bloodworth and Odell, 2008; Flower and Lydekker, 1891; Ishiyama, 1987; Werth et al., 2020; Willis and Baird, 1998). An explanation of this inconsistency could be due to notable variation among specimens examined from the occurrence of localized thin enamel at the tips of teeth that wears away rapidly after tooth eruption (Bloodworth and

Odell, 2008; Ishiyama, 1987; Werth et al., 2020). This was the case for Werth et al. (2020) which assigned the enamel complexity categories of 1/2 for the sperm whale (*Physeter macrocephalus*) although gave the definitive category of 2 to the pygmy sperm whale (*Kogia breviceps*) where they observed prismless enamel for their specimens of this species. Selection analysis results for those categories in the branch-specific model as well as for these specific taxa from the free-ratio model suggest that selection has been relaxed on these genes. Also, both *Kogia* species and *Physeter* showed elevated dN/dS values from the Coevol analyses. Evolutionarily, Bianucci and Landini (2006) suggested that enamel was lost in the last common ancestor of crown Physeteroidea as they reported several extinct stem taxa that had no enamel. There are no shared inactivating mutations between *Kogia* and *Physter*, but the shared loss of enamel complexity on the stem branch to physeteroids is consistent with the dN/dS value on this branch that is slightly elevated over the outgroups and the high dN/dS values on the *Physeter* and stem *Kogia* branches suggesting early relaxation of enamel genes in this clade (Table 2.5). In addition, Kogiidae share multiple inactivating mutations across four of the enamel genes, with at least one species having an inactivating mutation in six of the seven genes. Of these genes with shared mutations in *Kogia*, *ACP4* was also observed to contain multiple autapomorphic frameshift mutations in *Physeter* which ostensibly suggests functional loss of at least this gene in crown physeteroids and is the underlying cause of enamel degeneration in these taxa. Although there are no shared mutations in *ACP4* or the enamel genes for both genera, the case for convergent loss is unlikely with the interpretation of enamel evolution mentioned above from Bianucci and Landini (2006).

ACP4 is expressed in secretory stage ameloblasts and plays a key role in amelogenesis and differentiation of odontoblasts and mutations in humans causes autosomal recessive hypoplastic amelogenesis imperfecta in humans (Mu et al., 2021a; Seymen et al., 2016; Smith et al., 2017). This disease in humans is related to different enamel phenotypes but specifically when *ACP4* is nonfunctional there is insufficient or absence of the enamel matrix which causes thin, variably mineralized enamel (Smith et al., 2017). This degenerative enamel phenotype from mutations in humans is similar to what was observed in physeteroids and could relate to the variation in specimens and discrepancies among the different studies listed. Interestingly, this phenotype is also like that of some beaked whales where *ACP4* is inactivated in 3 species. In addition, *Kogia* share inactivating mutations in *AMELX*, which could relate to their prismless enamel phenotype similar to its inactivation mentioned in Phocoenidae (Werth et al., 2020). Finally, *Kogia* also share frameshift deletion mutations in *ENAM* which also could be a factor in producing prismless enamel where the decay of enamelin has similar effects as defective amelogenin where there is shortened activity of ameloblasts secreting the extracellular/enamel matrix proteins so the ameloblasts lose their Tome's processes relatively early (Loch et al., 2013a).

2.4.6. Ziphiids: categories 1–1/2

Extinct stem ziphiids (Ziphiidae) are characterized by having complete functional upper and lower teeth used to hunt and capture prey as raptorial feeders (Lambert et al., 2017). However, aside from Shepherd's beaked whale (*Tasmacetus shepherdi*) which has a complete set of functional teeth and tusks, most extant beaked whales only possess one

or two pairs of tusk-like teeth which usually only erupt and are functional in males albeit sometimes erupt in females that are smaller and usually do not develop and remain unerupted in the gums like seen in Cuvier's beaked whale (*Ziphius cavirostris*), and the two species of *Berardius* (Arnoux's beaked whale [*B. arnuxii*] and Baird's beaked whale [*B. bairdii*]); this latter case of unerupted, embedded teeth also occurs in males in the two species of *Hyperoodon* (northern [*H. ampullatus*] and southern [H. *planifrons*] bottlenose whales) (Dalebout et al., 2008; Loch and van Vuuren, 2016; MacLeod, 2017). These tusks can have very thin enamel or no enamel and exposed dentin (Ishiyama, 1987; Werth et al., 2020). Because of this sexual dimorphism and lack of complex enamel, it is not surprising that like other odontocetes ziphiids mainly use suction to capture their prey, feeding primarily on cephalopods, and instead most species of beaked whales use their tusks in male-male competition (Dalebout et al., 2008, MacLeod, 1998). This use for their tusks could be selected for as suggested by MacLeod (1998) in conjunction with the selection for scars to remain permanently unpigmented which would lead to a decrease in combat and therefore less injuries since scars are used as an indicator of male 'quality' as a signal to other males of a stronger individual with higher social status. Conversely, Dalebout et al. (2008) suggests selection on tusks as weapons may have been relaxed in some species such as the ginko-toothed beaked whale (*Mesoplodon ginkgodens*) since male intrasexual selection no longer occurs, as well as the fact that tusks can be costly to grow and, in some cases, hinder feeding ability. It would seem likely then that taxa that no longer require teeth for feeding are freed from selective pressure for numerous enamel-capped teeth, but instead of total degeneration and loss of

teeth, some are retained and altered via selection for this other function of combat in species with aggressive social interactions (MacLeod, 1998). Selection analysis results for the beaked whales from the free-ratio model dataset are all above the outgroup taxa and indicative of weaker purifying selection occurring on these branches. The stem branch that leads to these three species in this model had dN/dS values near suggesting that enamel genes have evolved neutrally on this branch and were relaxed near the base of this branch. The results of the Coevol analysis for all beaked whales have elevated dN/dS values. These results corroborate the free-ratio model results that enamel genes have been relaxed and are evolving neutrally in ziphiids. Along with dN/dS results of being elevated over the higher enamel complexity categories, there were also more mutations observed in beaked whales and lower category taxa. Including all the beaked whales from the odontocete dataset in this study, 12 of the total 19 beaked whales have at least one gene that contains an inactivating mutation (Table 2.2). The only taxa that have a shared inactivating mutation in Ziphiidae are the two species of *Hyperoodon* which have a shared frameshift deletion in *AMTN*. Although there is no specific data on enamel complexity in these species, it has been noted that along with the teeth that are left concealed in the gums, teeth are also rudimentary and functionless, and when found to have tips protruding out of the gums are loosely attached and easily fall out of their follicles; this has also been observed for *Z. cavirostris* and *Mesoplodon bidens* (Sowerby's beaked whale), both of which also have inactivating mutations (Boschma 1951).

Although there are no inactivating mutations shared by all ziphiids examined this could be due to the disparate types of enamel and tusk phenotypes seen in this family. Dalebout et al. (2008) suggested that tusks in this family evolved via parallel evolution, arising independently several times and may be the result of retention or re-emergence of ancestral character states. The interesting differences in tusk morphology and enamel microstructure phenotypes in beaked whales follows the results reported in our study, with a reduction of selection pressure on enamel genes resulting in relaxed selection on ziphiid branches from complex enamel not being needed for prey capture in most species where they instead use suction feeding, and that the underlying dentin/cementum remains strong enough to enable tusks to be used as weapons in some species (Davit-Béal et al., 2009; Loch and van Vuuren, 2016; Nweeia et al., 2014). Perhaps one reason for not observing more inactivating mutations in this clade could be that selection has been maintained for male-male competition albeit elevated above taxa that are known to use their teeth for prey acquisition and mastication.

2.4.7. Conclusions

Our study builds upon previous reports of inactivating mutations present in a variety of odontocetes with differing enamel microstructures including enamelless species. Perhaps the thin enamel phenotype observed in many odontocete species can lead to an evolutionary advantageous adaptation in taxa that no longer require the hard protective outer layer of enamel on teeth that was once required to aid in capture and mastication of prey or defense against predators. In narwhal, the lack of enamel on the elongated tusk in males exposes porous cementum which allows water to enter dentine

tubules that lets the animal sense characteristics in its surrounding environment like differences in temperature, salinity, and pressure among other stimuli; females may also possess this ability in unerupted tusks albeit in a much more limited capacity (Nweeia et al., 2014).

There is also research that has reported that in some non-odontocete taxa investigated where there is the pre-adaption of a secondary tool or method of food uptake which enhances the ability of its owner to occupy a specific ecological niche is selected for and leads to relaxed functional constraints on teeth and leads to enamel reduction and perhaps eventual tooth loss; examples of taxa and their tools leading to enamel simplification or tooth loss include beaks in birds, turtles, echidnas and platypus, elongated sticky tongues in toads, pangolins, and anteaters, baleen in mysticetes, and hyposodonty/hypselodonty in armadillos, sloths and aardvarks (Davit-Béal et al., 2009). Perhaps this could also apply to some odontocetes where suction feeding is the pre-adaptive tool letting species have access to a food source in an uncontested niche shifting total jaw function and morphology to this new phenotypic space which involves further degeneration of dentition involving enamel and tooth loss. Indeed, this is not totally farfetched as there have been instances of edentulous taxa that have evolved suction feeding from raptorial ancestors that possessed prismatic enamel-capped teeth such as in a recently discovered extinct toothless dwarf dolphin (*Inermorostrum xenops*) and several living baleen whale species (Boessenecker et al., 2017). This process of evolution has also occurred in several other taxa that still possess teeth such as in extant beaked whales, narwhal, and some seals indicating that the adaptation to suction feeding has happened

several times previously (Hocking et al., 2017). These observations along with the release from selection pressure on enamel genes reported in our study potentially provide evidence for further changes in the dentition of toothed whales that already do not utilize their teeth for traditional mammalian functions. These taxa may very well adapt to obligate suction feeding involving teeth to become rudimentary, vestigial, and functionless or become lost altogether.

Tables

Table 2.1. Odontocete taxa with a score for Werth Enamel Complexity.

Odontoceti Species	Werth Enamel Complexity
Berardius bairdii	2
Cephalorhynchus hectori	4
Delphinapterus leucas	2/3
Delphinus capensis	4
Delphinus delphis	4
Globicephala macrorhynchus	4
Globicephala melas	4
Grampus griseus	4
Inia geoffrensis	5
Kogia breviceps	2
Kogia sima	2*
Lagenodelphis hosei	4
Lagenorhynchus acutus	4
Lagenorhynchus albirostris	4
Lagenorhynchus obscurus	4
Mesoplodon densirostris	1/2
Monodon monoceros	1*
Neophocaena asiaeorientalis	3*
Orcaella brevirostris	4
Orcinus orca	4
Phocoena phocoena	3
Physeter macrocephalus	1/2
Platanista gangetica	5
Platanista minor	5*
Pontoporia blainvillei	4
Pseudorca crassidens	4
Stenella attenuata	4
Stenella clymene	4
Stenella coeruleoalba	4
Stenella frontalis	4
Steno bredanensis	4
Tursiops truncatus	4
Ziphius cavirostris	1

Outgroups	
Bos mutus	5*
Camelus bactrianus	5*
Hippopotamus amphibius	5*
Sus scrofa	5*

Taxa with assigned enamel organization from Werth et al. (2020). *Denotes species that were assigned a Werth enamel complexity score in this paper. Categories with a decimal are the mean of two categories in parentheses (see main text). 1 = no enamel, 2 = prismless enamel, 3 = intermediate or irregular enamel, 4 = prismatic enamel, and 5 = Hunter-Schreger bands or other decussation.

Table 2.2. Inactivating mutations in enamel genes in Odontoceti.

Odontoceti taxon	**Enamel gene and mutation**
Kogiidae (*K. breviceps* + *K. sima*)	*ACP4*: E1-3, 11: NBR/NRM; *AMELX*: E2: 47I, In2Do:GG; *ENAM*: E8: 2403D, 3751-3752D; *KLK4*: WGD (NBR/NRM)
Phocoenidae (*N. asiaeorientalis* + *P. phocoena* + *P. sinus*)	*AMTN*: In2Ac:AT; *KLK4*: E2: 73-75S
Plastanista (*P. gangetica* + *P. minor*)	*AMTN*: E3: NBR
Hyperoodon (*H. ampullatus* + *H. planifrons)*	*AMTN*: E8: 566D
Monodontidae (*D. leucas* + *M. monoceros*)	*AMELX*: In2Do: AT
Berardius arnuxii	*AMELX*: E7: NBR/NRM; *ENAM*: E8: 648I
Berardius bairdii	*MMP20*: E5:1095-1097S
Delphinapterus leucas	*ACP4*: E7: 674D; *AMTN*: E8: 576D
Hyperoodon ampullatus	*ENAM*: In6Do: TT
Hyperoodon planifrons	*AMELX*: E7: NBR/NRM
Kogia breviceps	*ACP4*: E9: 900D; E10: 1100I; *AMBN*: In7Ac: AT; In9Ac: AT; *AMELX*: E2: SCM; *AMTN*: E8: 377D; *ENAM*: In6Do: CT
Kogia sima	*AMBN:* WGD (NBR/NRM); *ENAM*: E1-6: NBR/NRM
Mesoplodon bidens	*ACP4*: E2: 145-148I, E5: 537-544D, In8Ac: GG
Mesoplodon densirostris	*ACP4*: E5: 537-539S (allelic variation); *AMTN*: E5: 249-252I (allelic variation)
Mesoplodon grayi	*ACP4*: E8: 802I (allelic variation); *AMELX*: In2Ac: GG/AG (allelic variation)
Mesoplodon layardii	*KLK4*: In2Do: CG/GT (allelic variation)
Mesoplodon perrini	*AMTN*: E8-9: NRM (no stop)
Mesoplodon peruvianus	*AMELX*: E7: NBR/NRM; *AMTN*: E8: 376I
Monodon monoceros	*ACP4*: In6Do: AT; *AMBN*: E11: 1214-1216S; *KLK4*: E4: 503-505S
Orcinus orca	*ENAM*: In4Do: AT/GT (allelic variation)
Physeter macrocephalus	*ACP4*: E4: 427D; E9: 1015D; E10: 1132D (allelic variation)
Tasmacetus shepherdi	*KLK4*: E5: 767-769
Ziphius cavirostris	*KLK4*: In3Ac: GG

Numbers correspond to positions in the protein-coding sequence alignments (CDS). Abbreviations: Ac = acceptor splice site; D = deletion; Do = donor splice site; E = exon; I = insertion; In = intron; NBR = no blast results and possible deletion of exon(s) or gene; NRM = no reads mapped and possible deletion of exon(s) or gene; S = premature stop codon; SCM = start codon mutation.

Table 2.3. Monophyly of well-supported cetartiodactyl clades on maximum likelihood phylograms for seven enamel related genes.

Clade	Gene						
	ACP4	***AMBN***	***AMELX***	***AMTN***	***ENAM***	***KLK4***	***MMP20***
Delphinidae	yes	no	yes	yes	yes	yes	polytomy
Phocoenidae	yes	yes	yes	yes	yes	yes	yes
Monodontidae	yes	yes	no	yes	yes	yes	yes
Phocoenidae + Monodontidae	yes	yes	no	yes	yes	no	no
Delphinoidea	yes	yes	yes	yes	yes	yes	yes
Delphinida	yes	yes	yes	yes	yes	yes	polytomy
Inioidea	yes	yes	yes	yes	yes	yes	yes
Iniidae + Pontoporiidae + Lipotidae	no	no	yes	no	yes	no	yes
Ziphiidae	yes	yes	yes	yes	yes	yes	yes
Delphinida + Ziphiidae	yes	yes	no	polytomy	yes	yes	yes
Platanistidae	yes	yes	yes	yes	yes	yes	yes
Kogiidae	yes	**n/a**	yes	yes	yes	**n/a**	yes
Physeteroidea	no	yes	yes	yes	yes	**n/a**	yes
Odontoceti	no	yes	no	no	yes	yes	yes
Balaenopteroidea	no	no	yes	no	yes	yes	yes
Balaenidae	yes	yes	yes	yes	yes	yes	yes
Mysticeti	yes	yes	yes	no	yes	yes	yes
Cetacea	yes	yes	yes	yes	yes	yes	yes
Hippopotamidae	yes	yes	yes	yes	yes	yes	yes
Cetancodonta*	yes	no	no	yes	yes	no	yes
Cetancodonta + Ruminantia	yes	yes	no	yes	yes	yes	yes
Giraffidae	yes	yes	no	yes	yes	yes	yes
Cervidae	yes	yes	yes	yes	yes	yes	yes
Caprinae	yes	yes	yes	yes	yes	yes	yes
Bovinae	yes	yes	yes	no	yes	yes	yes
Bovidae	yes	no	no	yes	yes	yes	yes
Pecora	yes	yes	yes	yes	yes	yes	yes
Ruminantia	yes	yes	yes	yes	yes	yes	yes

Suina**	yes	yes	yes	yes	yes	yes	yes
Camelidae	yes	yes	yes	yes	yes	yes	yes
Total number of clades recovered	26/30	24/29	22/30	24/30	30/30	25/28	27/30
Percentage of clades recovered	86.67 %	82.8%	73.3%	80%	100%	89.29 %	90%

*Same as Whippomorpha (Waddell et al., 1999), **sensu Spaulding et al. (1999) and Hassanin et al. (2012).

Delphinoidea = Delphinidae + Phocoenidae + Monodontidae; Delphinida = Inioidea + Lipotidae + Delphinoidea; Inioidea = Iniidae + Pontoporiidae; Physeteroidea = Kogiidae + Physeteridae.

Table 2.4. Mutation counts mutations per million years.

Werth Categories	Number of Branches	Number of Mutations	Branch length (MY)	Mutations/MY
ACCTRAN				
1	1	1	11.2	0.0893
1.25	1	1	11.2	0.0893
1.5	2	3	12.56	0.2389
1.75	1	2.5	22.42	0.1115
2	4	17	45.78	0.3713
2.25	1	2	8.15	0.2454
2.5	1	1	7.17	0.1395
3	5	2	48.11	0.0416
3.5	1	0	4.46	0.0000
4	43	0.5	154.03	0.0032
4.5	3	1	71.32	0.0140
5	9	0	251.87	0.0000
DELTRAN				
1.5	1	1	11.2	0.0893
1.75	2	3.5	33.62	0.1041
2	6	20	58.34	0.3428
2.75	1	2	8.15	0.2454
3	5	3	43.57	0.0689
3.5	2	0	16.17	0.0000
4	41	0.5	151.17	0.0033
4.5	2	0	20.78	0.0000
5	12	1	305.27	0.0033
MPR				
1.25	1	1	11.2	0.0893
1.5	1	1	11.2	0.0893
1.75	3	5.5	34.98	0.1572
2	4	17	49	0.3469
2.5	1	2	8.15	0.2454
2.75	1	1	7.17	0.1395
3	4	2	36.4	0.0549
3.25	1	0	11.71	0.0000
3.5	1	0	4.46	0.0000

4	41	0.5	150.13	0.0033
4.25	1	0	1.16	0.0000
4.5	2	0	21.32	0.0000
4.75	2	1	51.7	0.0193
5	9	0	251.87	0.0000

Abbreviations: ACCTRAN = accelerated transformation; DELTRAN = delayed transformation; MPR = most parsimonious reconstruction sets in PAUP; MY = millions of years.

Table 2.5. Selection analyses (dN/dS) for the free-ratio model.

Species	dN/dS	dN/dS	
	FreeRatio-CF1	FreeRatio-CF2	Werth Category
Berardius bairdii	0.5836	0.6035	2
Cephalorhynchus hectori	1.8138	1.8686	4
Delphinapterus leucas	1.9918	2.0672	2.5
Delphinus capensis	1.2811	1.3193	4
Delphinus delphis	0.0001	0.0001	4
Globicephala macrorhynchus	1.7106	1.7603	4
Globicephala melas	1.6476	1.6972	4
Grampus griseus	0.4269	0.4397	4
Inia geoffrensis	0.4405	0.4519	5
Kogia breviceps	1.3944	1.4303	2
Kogia sima	1.4179	1.4541	2*
Lagenodelphis hosei	0.8511	0.8774	4
Lagenorhynchus acutus	0.7598	0.7792	4
Lagenorhynchus albirostris	0.8469	0.8671	4
Lagenorhynchus obscurus	0.6420	0.6614	4
Mesoplodon densirostris	0.7021	0.7264	1.5
Monodon monoceros	1.6126	1.6527	1*
Neophocaena asiaeorientalis	0.5375	0.5536	3*
Orcaella brevirostris	0.4010	0.4089	4
Orcinus orca	0.6224	0.6454	4
Phocoena phocoena	1.3526	1.3934	3
Physeter macrocephalus	1.1977	1.2290	1.5
Platanista gangetica	999	999	5
Platanista minor	999	999	5*
Pontoporia blainvillei	0.5166	0.5374	4
Pseudorca crassidens	1.4938	1.5388	4
Stenella attenuata	1.2886	1.3247	4
Stenella clymene	0.3655	0.3767	4
Stenella coeruleoalba	999	999	4
Stenella frontalis	0.3734	0.3847	4
Steno bredanensis	0.8806	0.9145	4
Tursiops truncatus	0.6062	0.6260	4
Ziphius cavirostris	0.5840	0.6026	1

Outgroups	dN/dS	dN/dS	Werth Category
Bos mutus	0.4542	0.4675	5
Camelus bactrianus	0.3754	0.3880	5
Hippopotamus amphibius	0.4940	0.5074	5
Sus scrofa	0.4622	0.4767	5
Internal Branch Grouping Listed by Nodes	**dN/dS**	**dN/dS**	**Reconstruction Method and Category**
3 → 4	0.2568	0.2637	All: 5
4 → 5	0.4013	0.4221	All: 5
5 → 6	0.3714	0.3851	ACC: 4.5, DEL: 5, MPR: 4.75
6 → 66	0.4968	0.5104	ACC: 3, DEL: 3.5, MPR: 3.25
66 → 67	1.4677	1.5189	All: 2
6 → 7	0.7828	0.8488	ACC: 4, DEL: 5, MPR: 4.5
7 → 63	0.5799	0.5949	ACC: 4.5, DEL: 5, MPR: 4.75
7 → 8	0.6543	0.6937	ACC: 4, DEL: 4.5, MPR: 4.25
8 → 58	1.0176	1.0427	All: 3
58 → 59	0.8524	0.8791	ACC: 1.5, DEL: 2, MPR: 1.75
8 → 9	0.3957	0.4155	All: 4
9 → 55	0.3161	0.3245	All: 4
9 → 10	0.6273	0.6430	All: 4
10 → 48	1.0727	1.0916	All: 3.5
48 → 49	0.6211	0.6394	ACC: 2.5, DEL: 3, MPR: 2.75
48 → 52	1.4551	1.5016	All: 3
10 → 11	0.6878	0.7088	All: 4
11 → 12	0.3215	0.3288	All: 4
12 → 13	0.5682	0.2669	All: 4
13 → 14	0.1815	0.1878	All: 4
14 → 42	0.2653	0.2744	All: 4
14 → 15	0.6410	0.6605	All: 4
15 → 31	999	999	All: 4
31 → 32	999	999	All: 4
32 → 33	0.9620	0.9620	All: 4
33 → 34	1.3256	0.5460	All: 4
34 → 35	0.4572	0.4711	All: 4
15 → 16	0.0001	0.0001	All: 4
16 → 22	999	999	All: 4

22 → 23	0.5996	0.6170	All: 4
23 → 24	0.0001	0.0001	All: 4
23 → 27	0.6405	0.6599	All: 4
16 → 17	0.5894	0.3239	All: 4
17 → 19	0.7200	0.7392	All: 4

Abbreviations: ACC, accelerated transformation; All, all reconstruction methods; CF1, codon frequency model 1; CF2, codon frequency model 2; DEL, delayed transformation optimization; MPR, most parsimonious reconstruction sets in PAUP (equivalent to reconstructions in Mesquite).
Node numbers from Figure 2.2.

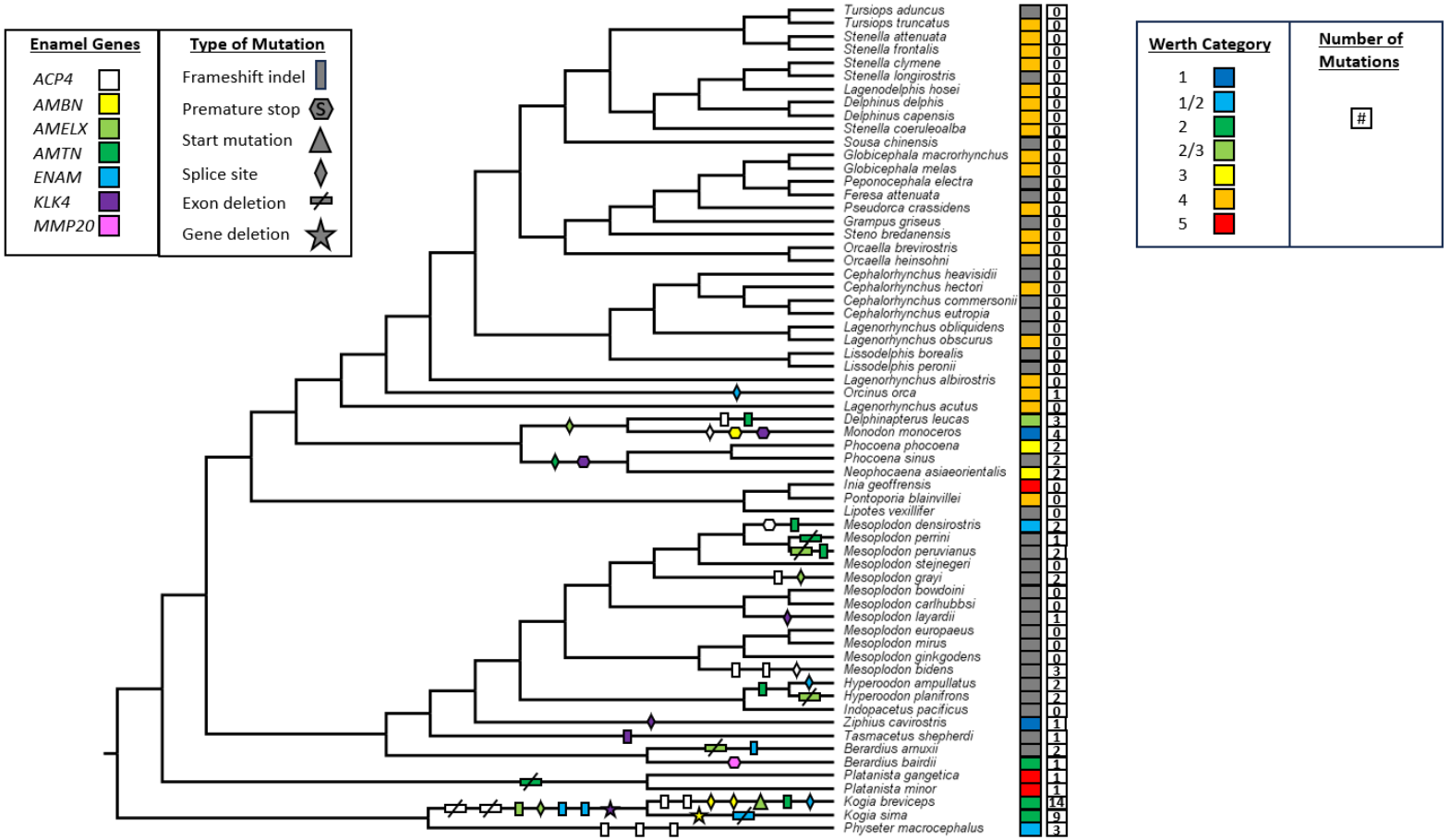
Table 2.6. Selection analyses (dN/dS) for the branch-specific codon models from all three ancestral reconstructions.

Branch End Points (basal, apical)			Enamel Complexity Branch Category for Regression Analyses	dN/dS Values					
				ACCTRAN		DELTRAN		MPR/Mesquite	
ACCTRAN	DELTRAN	MPR/Mesquite		CF1	CF2	CF1	CF2	CF1	CF2
1, 1	X	X	1	0.5748	0.5934	X	X	X	X
1, 1/2	X	1/2, 1	1.25	0.707	0.7312	X	X	0.5765	0.595
2, 1	2, 1	1/2, 1/2	1.5	1.2968	1.3335	0.5695	0.588	0.7058	0.73
2, 1/2	2, 1/2	2, 1/2; 2/3, 1	1.75	1.1115	1.1418	0.8962	0.9244	1.2283	1.2614
2, 2	2, 2; 3, 1	2, 2	2	1.0931	1.1268	1.117	1.1511	1.0924	1.126
2, 2/3	X	X	2.25	1.9852	2.0594	X	X	X	X
3, 2	X	2/3, 2/3	2.5	0.6236	0.6418	X	X	1.9973	2.0737
X	3, 2/3	3, 2/3	2.75	X	X	2.0209	2.0979	0.6229	0.6411
3, 3; 4, 2	3, 3; 4, 2	3, 3; 4, 2	3	1.0063	1.036	1.027	1.0563	1.1315	1.1632
X	X	4/5, 2	3.25	X	X	X	X	0.5134	0.5289
4, 3	4, 3; 5, 2	4, 3	3.5	1.0642	1.0849	0.7276	0.748	1.058	1.0781
4, 4	4, 4	4, 4	4	0.5918	0.6119	0.5893	0.6085	0.5887	0.6079
X	X	4/5, 4	4.25	X	X	X	X	0.7023	0.7438
5, 4; 4, 5	5, 4; 4, 5	4, 5; 4/5, 4/5	4.5	0.4427	0.4568	0.4426	0.4549	0.4824	0.4993
X	X	5, 4/5	4.75	X	X	X	X	0.4451	0.4601
5, 5	5, 5	5, 5	5	0.4248	0.4378	0.4289	0.4424	0.4248	0.4378

Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2. Gray cells with 'X' represent enamel branch categories that have zero reconstructed branches with this enamel complexity value.

Figures

Figure 2.1. Mapping of inactivating mutations.



Mapping of inactivating mutations in enamel genes onto species tree with delayed transformation (DELTRAN) character optimization. Enamel gene and type of mutation in legend on left. First column of boxes to right of taxa represents Werth Category shown in legend, gray color represents taxa with no assigned category. Second column of boxes shows number of mutations in each taxon, including shared mutations.

Figure 2.2. Species tree with node numbers for 37 taxa with Werth Enamel Complexity scores.

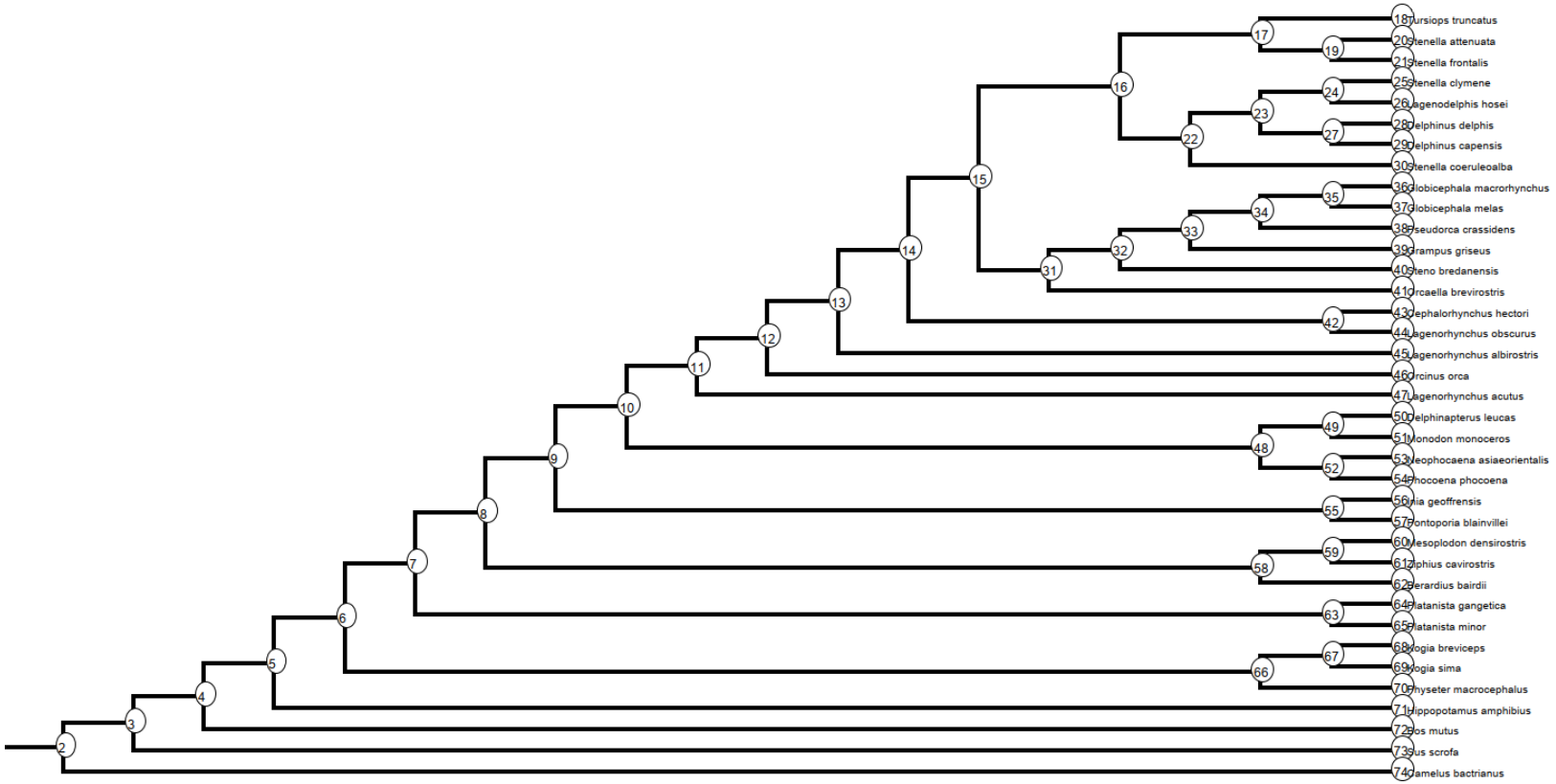
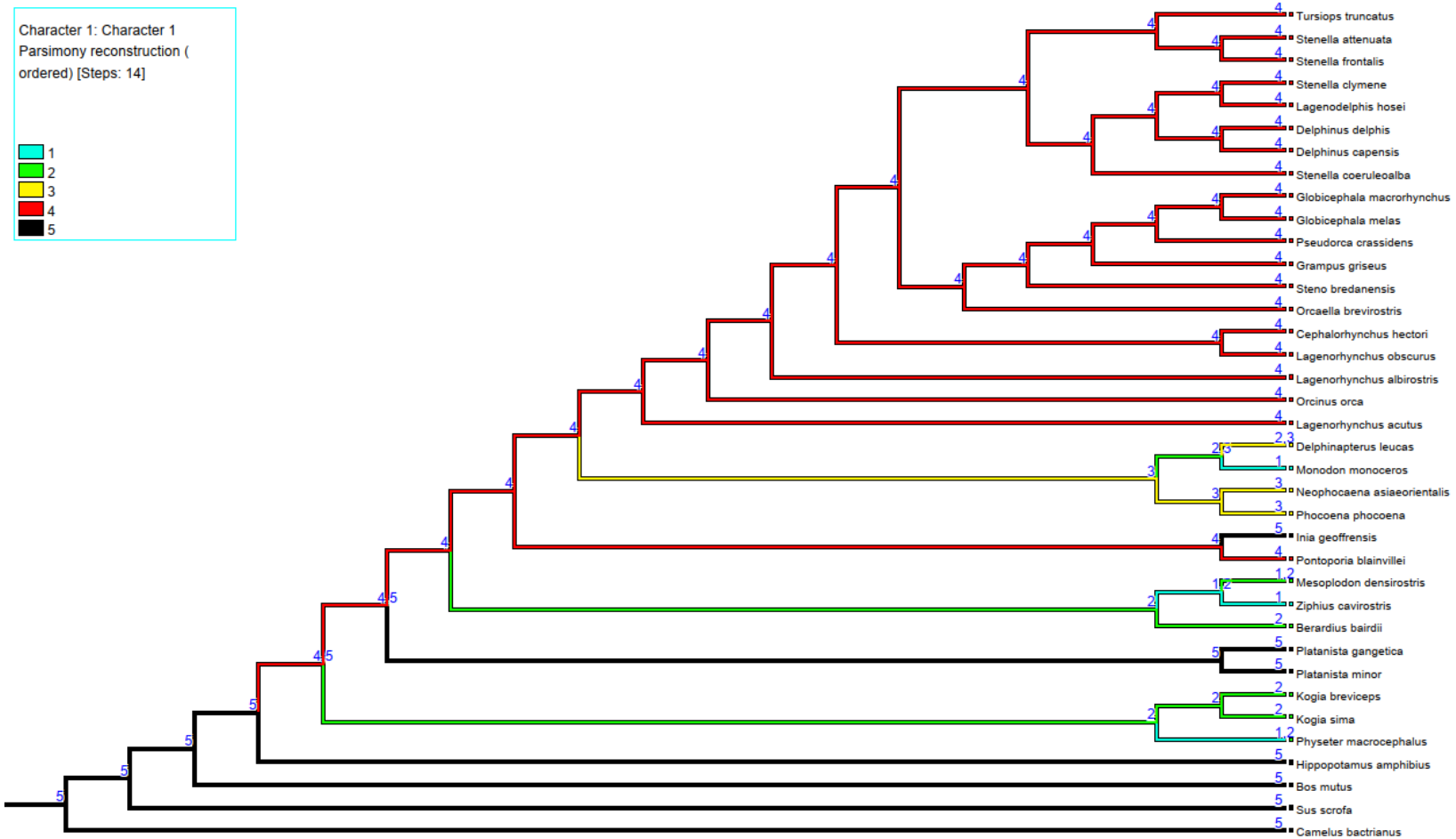
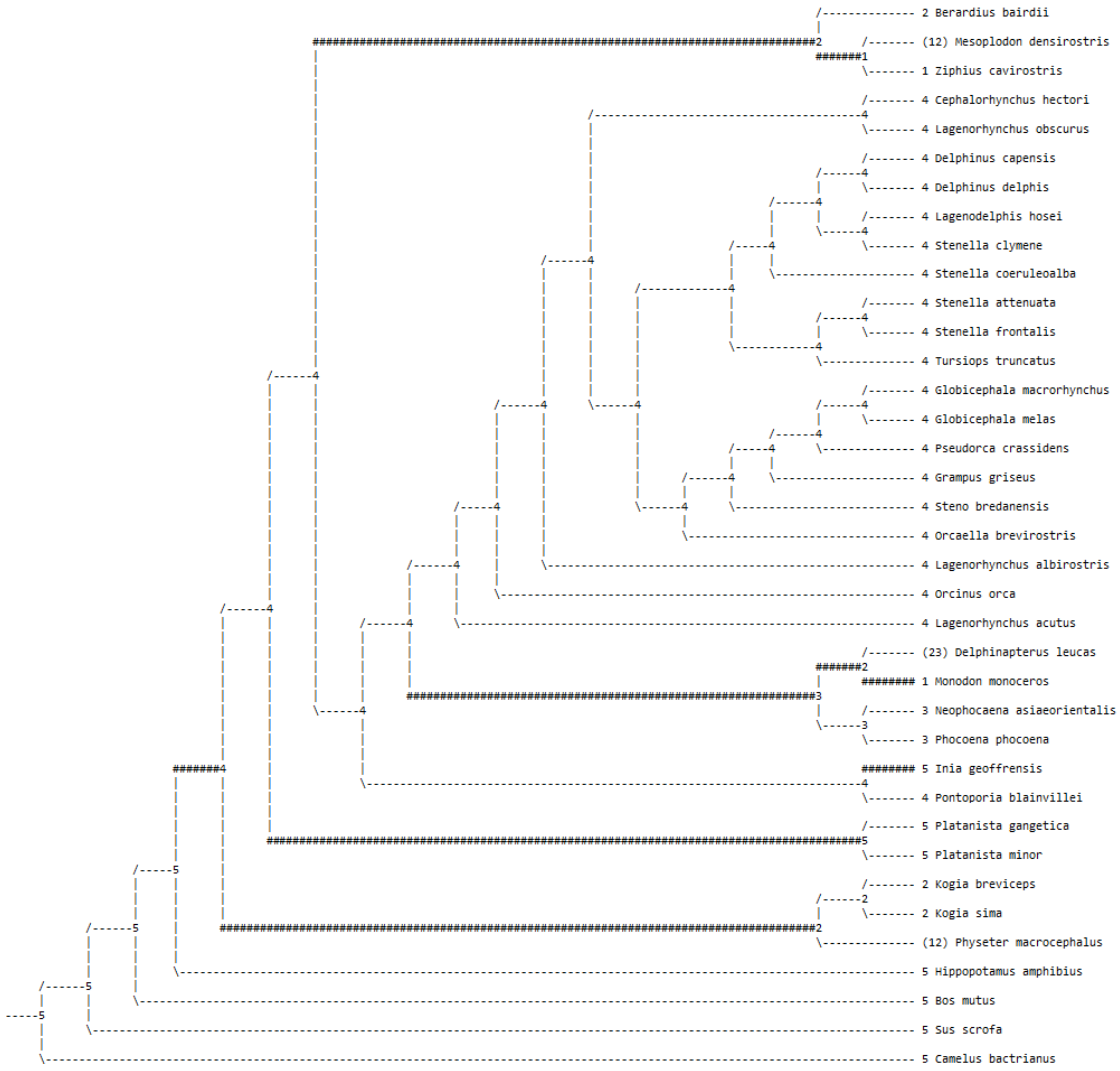


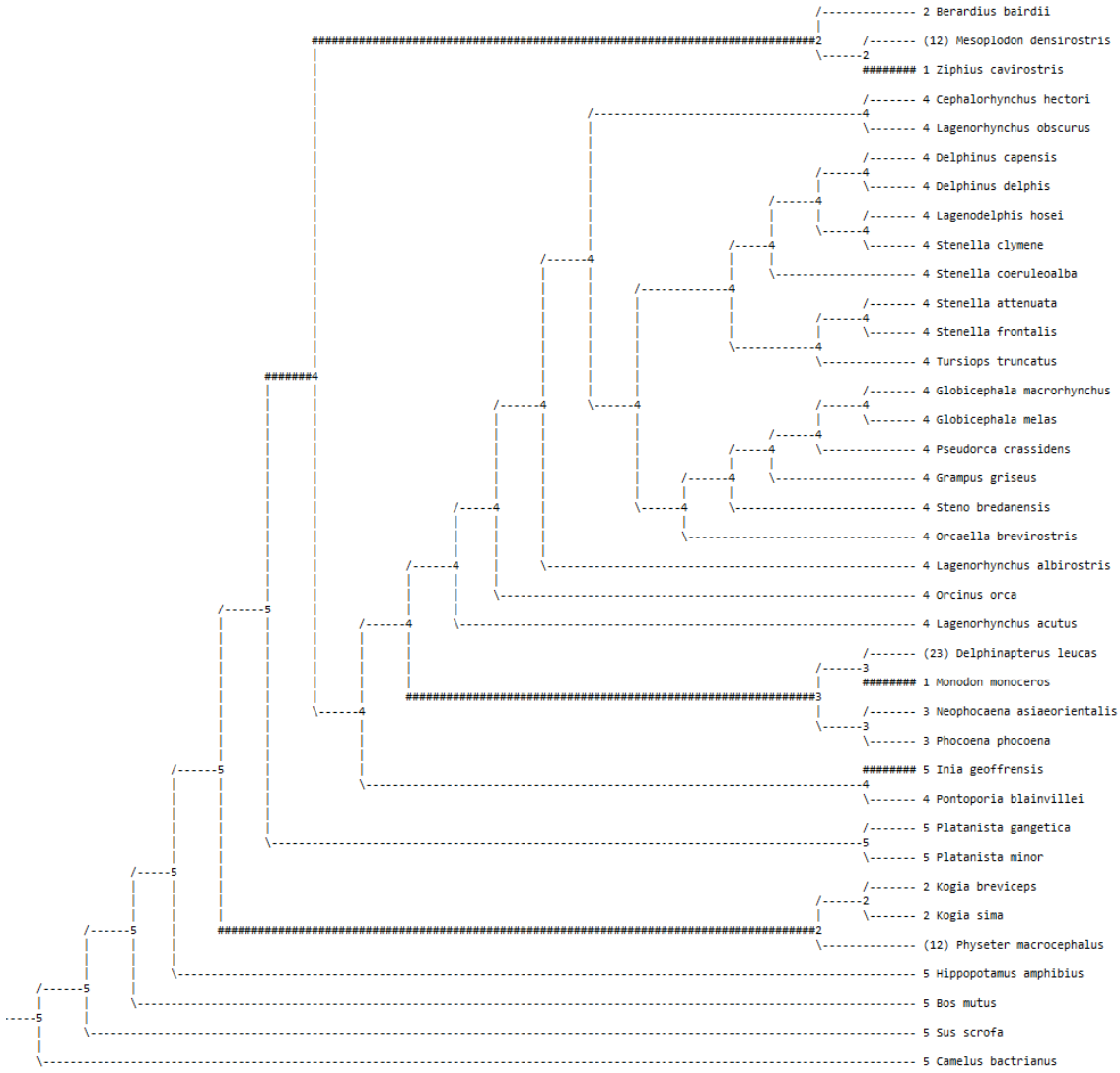
Figure 2.3A–C. Ancestral reconstructions.



A) Reconstruction from Mesquite and PAUP with the MPR model. Werth Enamel Complexity represented by numbers on nodes and colors shown in the legend. Nodes with multiple numbers are polymorphic complexity categories.



B) Reconstruction from PAUP with the ACCTRAN model. Branches drawn with "#" indicate reconstructed state changes. Taxa with two states are polymorphic.



C) Reconstruction from PAUP with the DELTRAN model. Branches drawn with "#" indicate reconstructed state changes. Taxa with two states are polymorphic.

Figure 2.4. Coevol selection analysis (dN/dS).

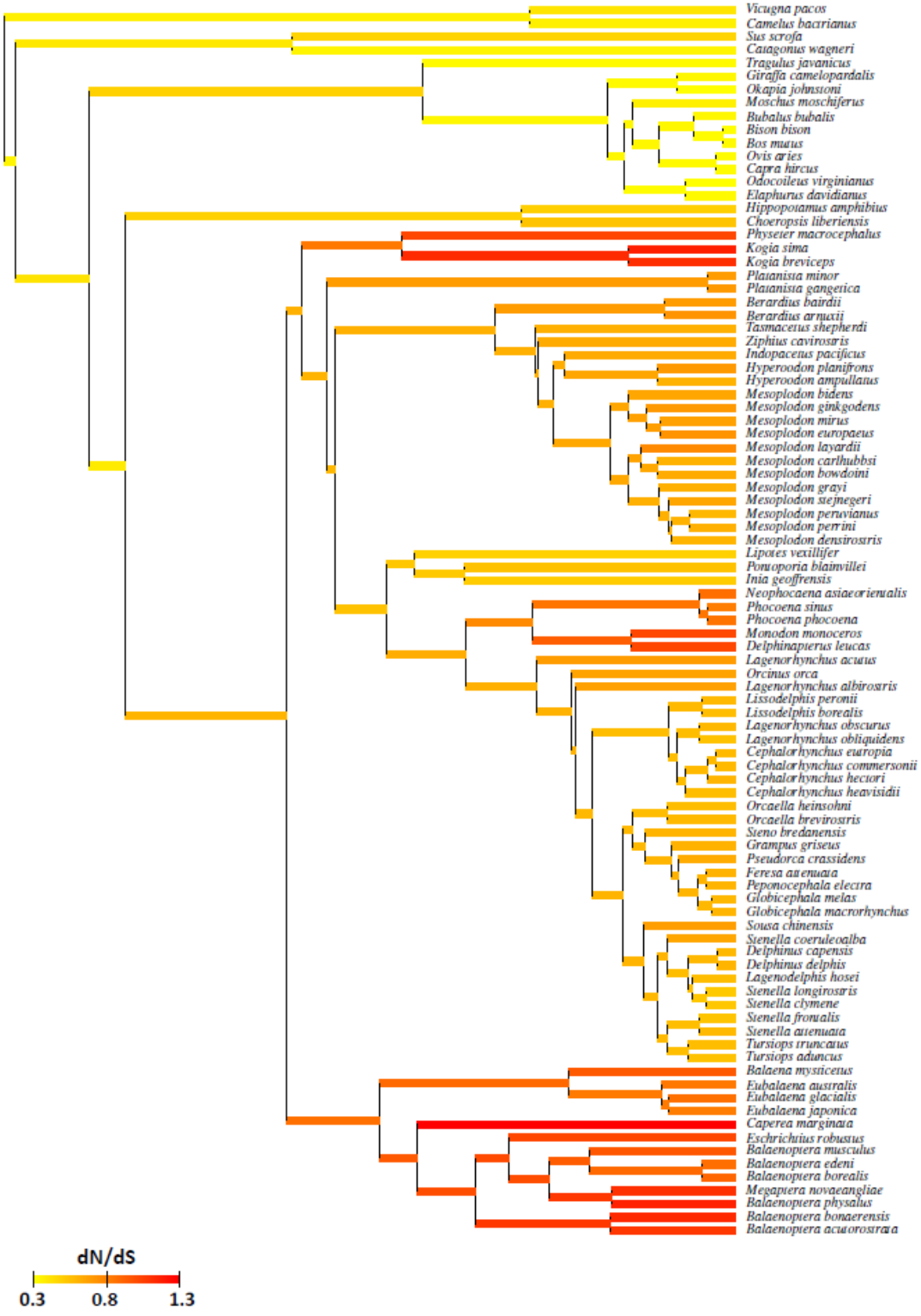
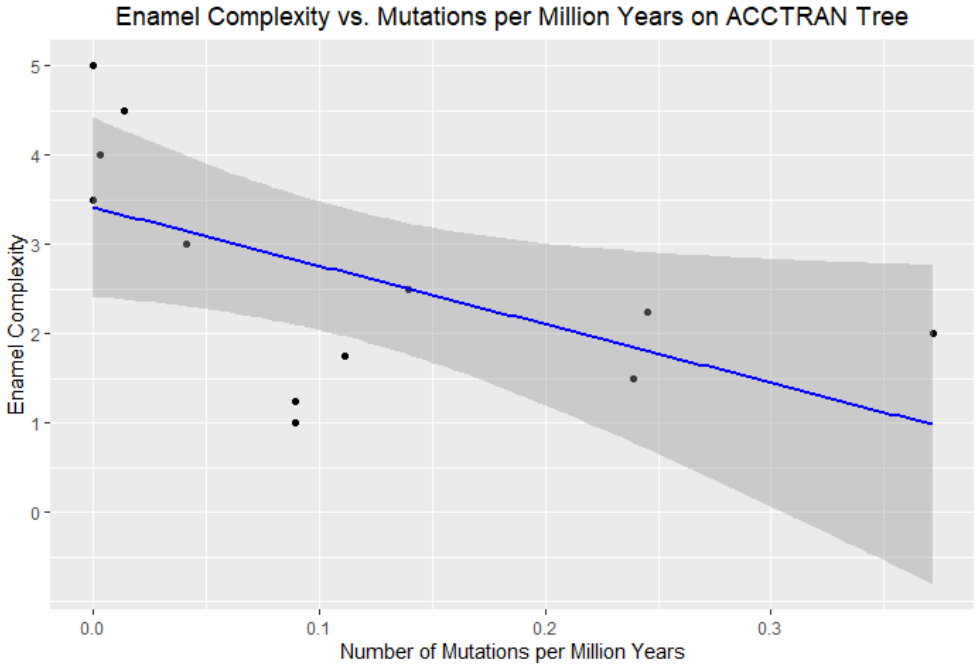
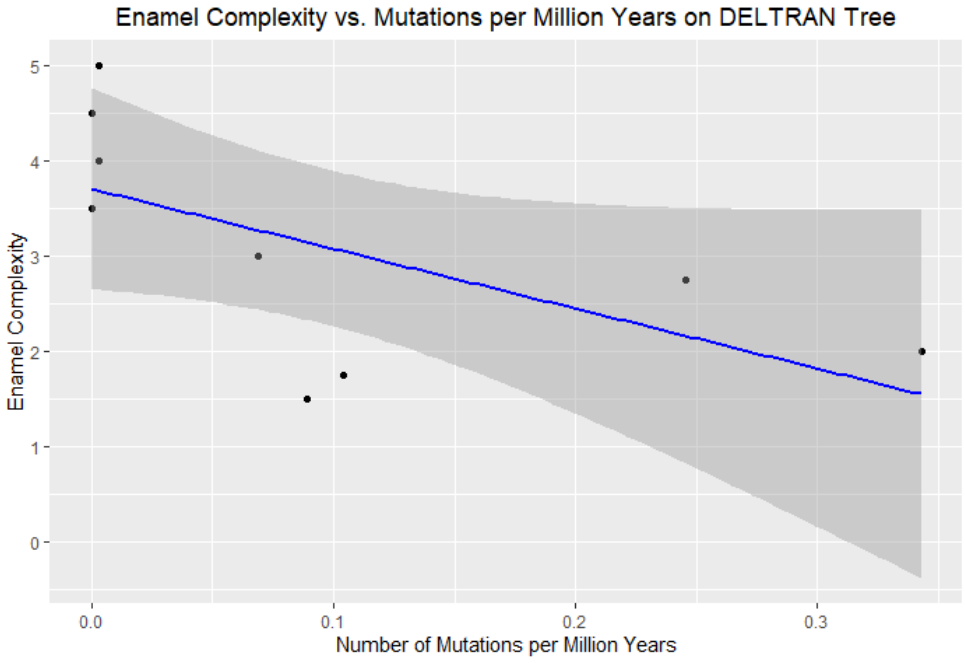


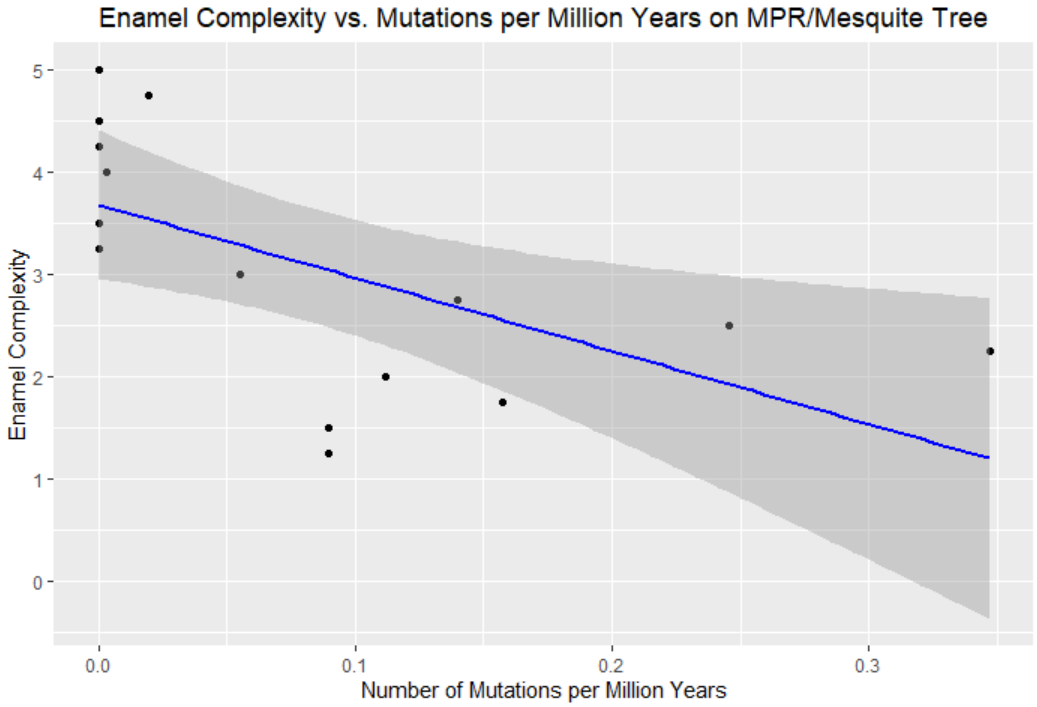
Figure 2.5A–C. Regression analyses of enamel organization categories versus mutations per million years. Shaded areas represent 95% confidence intervals on the fitted values.



A) Branch lengths from branches with ACCTRAN reconstructed Werth Categories. Slope: -6.5638. Adjusted R-squared: 02848. p-value: 0.0428*.



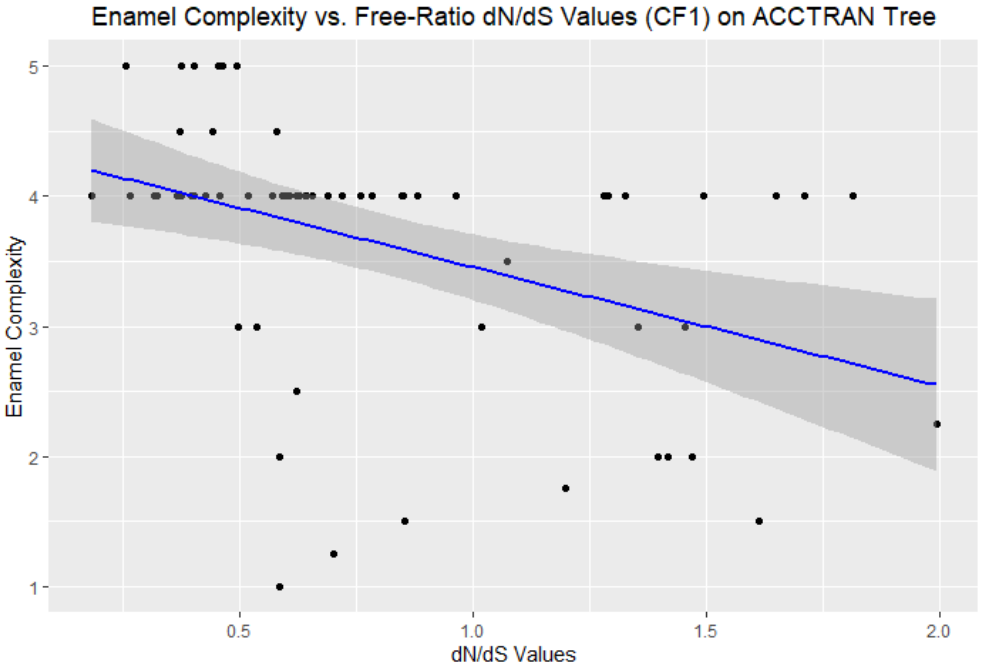
B) Branch lengths from branches with DELTRAN reconstructed Werth Categories. Slope: -6.2960. Adjusted R-squared: 0.2993. p-value: 0.0737.



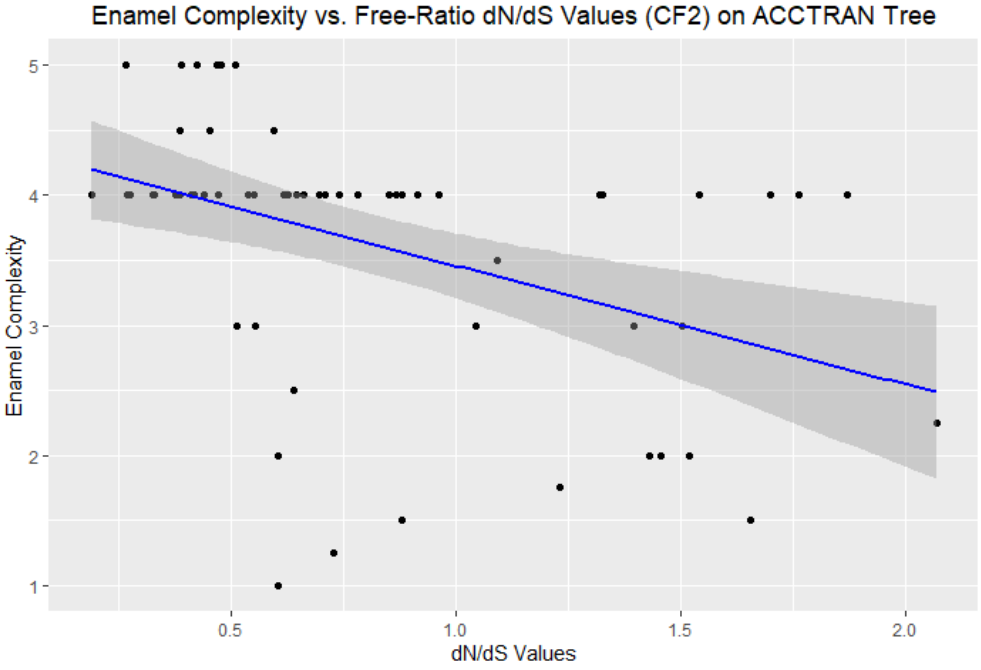
C) Branch lengths from branches with MPR/Mesquite reconstructed Werth Categories. Slope: -7.1648. Adjusted R-squared: 0.3244. p-value: 0.0157*.

* = significant p-values.

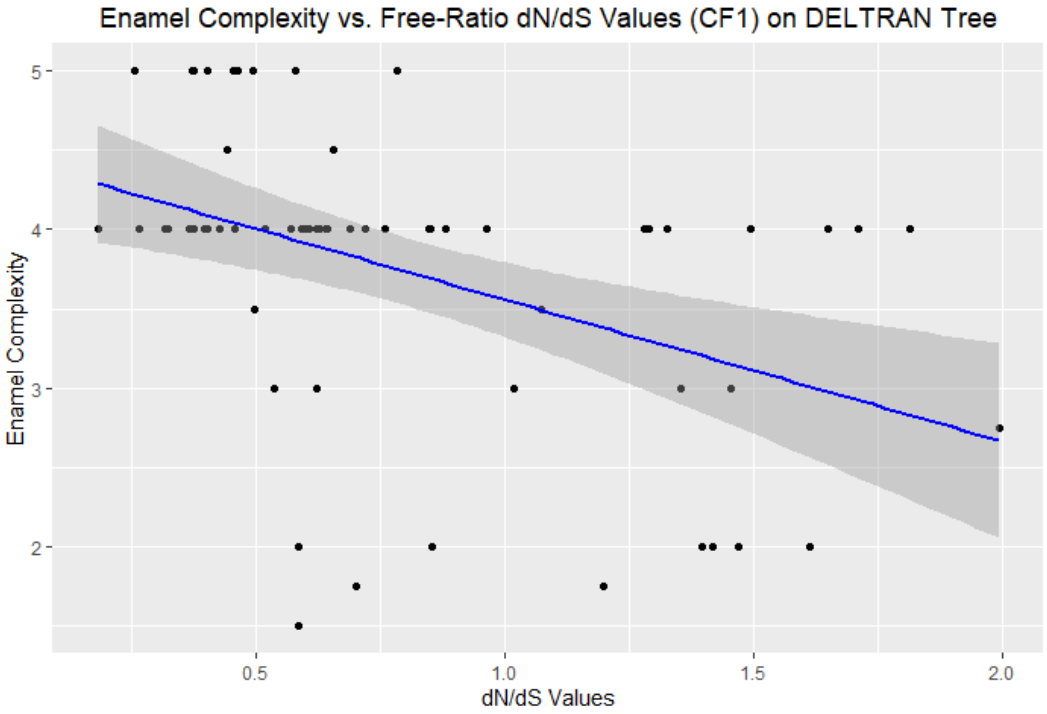
Figure 2.6A–L. Regression analyses of enamel organization categories versus dN/dS values. Shaded areas represent 95% confidence intervals on the fitted values.



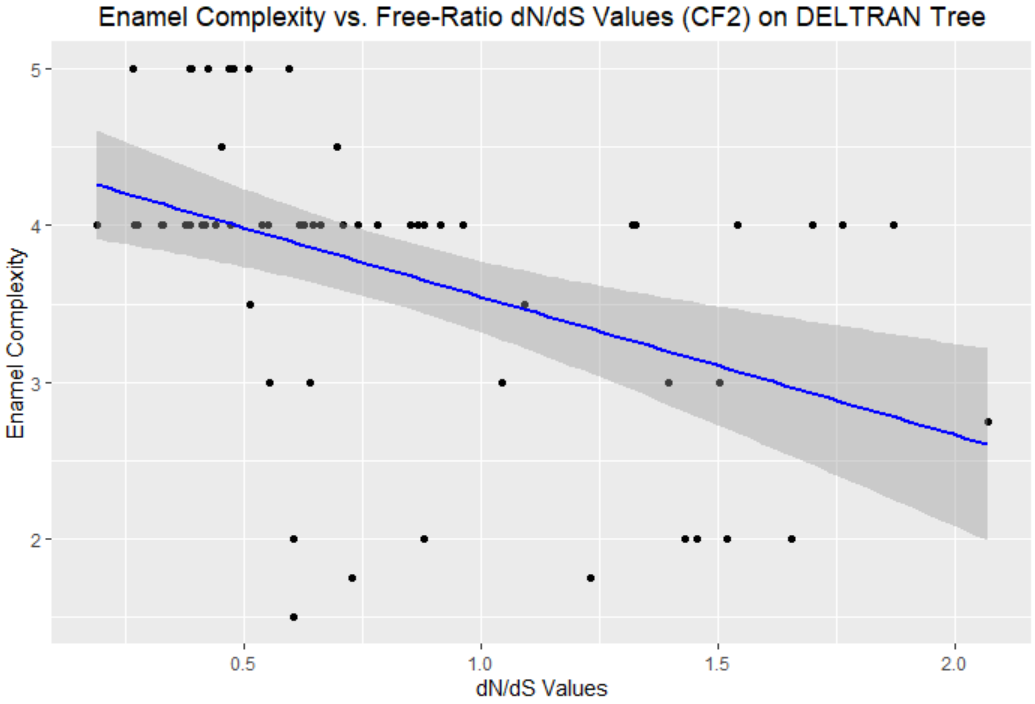
A) Free-ratio dN/dS model (CF1) with ACCTRAN reconstructed Werth Categories. Slope: -0.9133. Adjusted R-squared: 0.1581. p-value: 0.0008*.



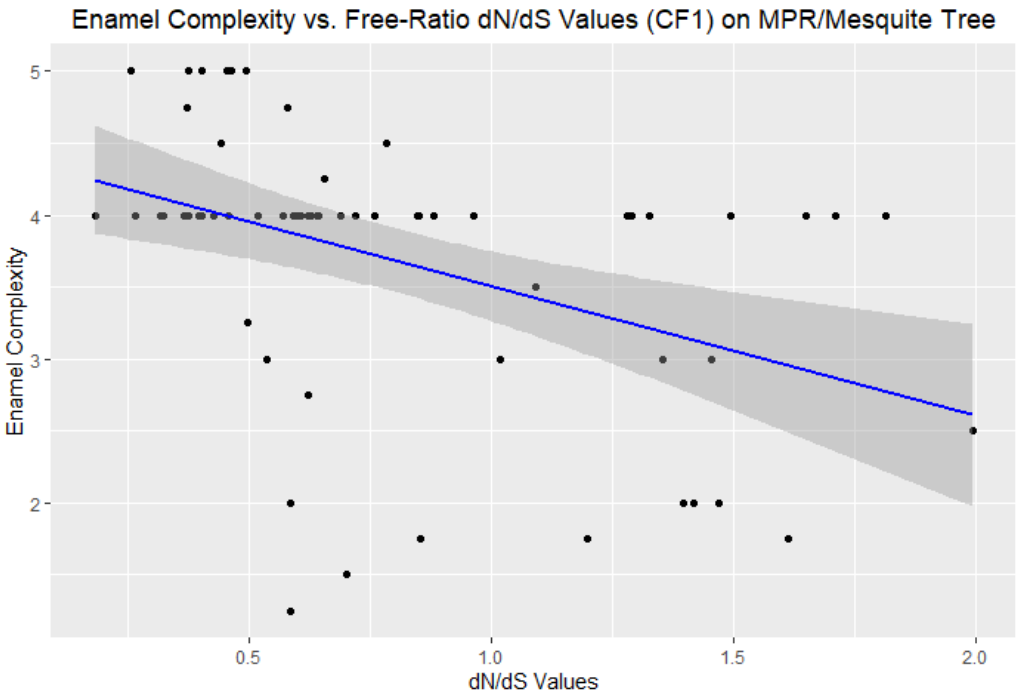
B) Free-ratio dN/dS model (CF2) with ACCTRAN reconstructed Werth Categories. Slope: -0.9113. Adjusted R-squared: 0.1705. p-value: 0.0005*.



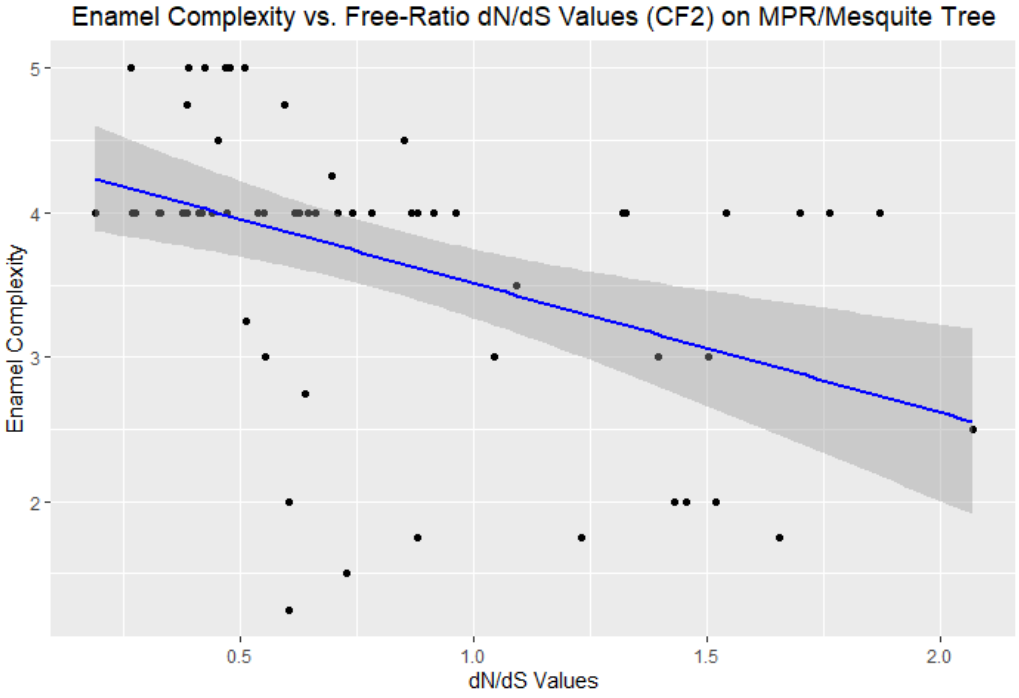
C) Free-ratio dN/dS model (CF1) with DELTRAN reconstructed Werth Categories. Slope: -0.8945. Adjusted R-squared: 0.1726. p-value: 0.0005*.



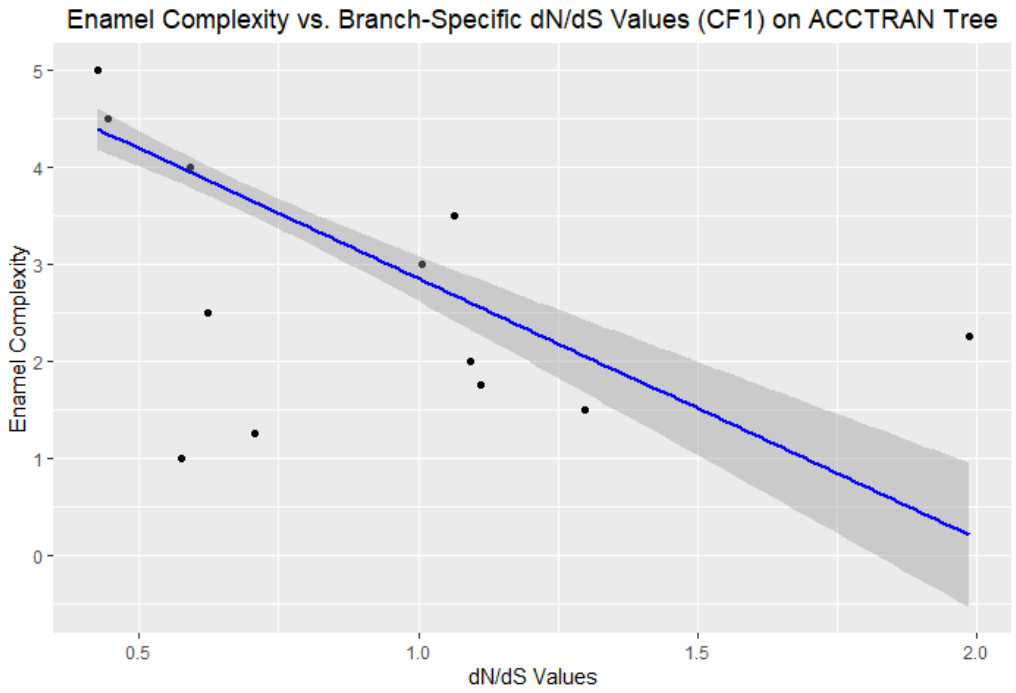
D) Free-ratio dN/dS model (CF2) with DELTRAN reconstructed Werth Categories. Slope: -0.8825. Adjusted R-squared: 0.1874. p-value: 0.0003*.



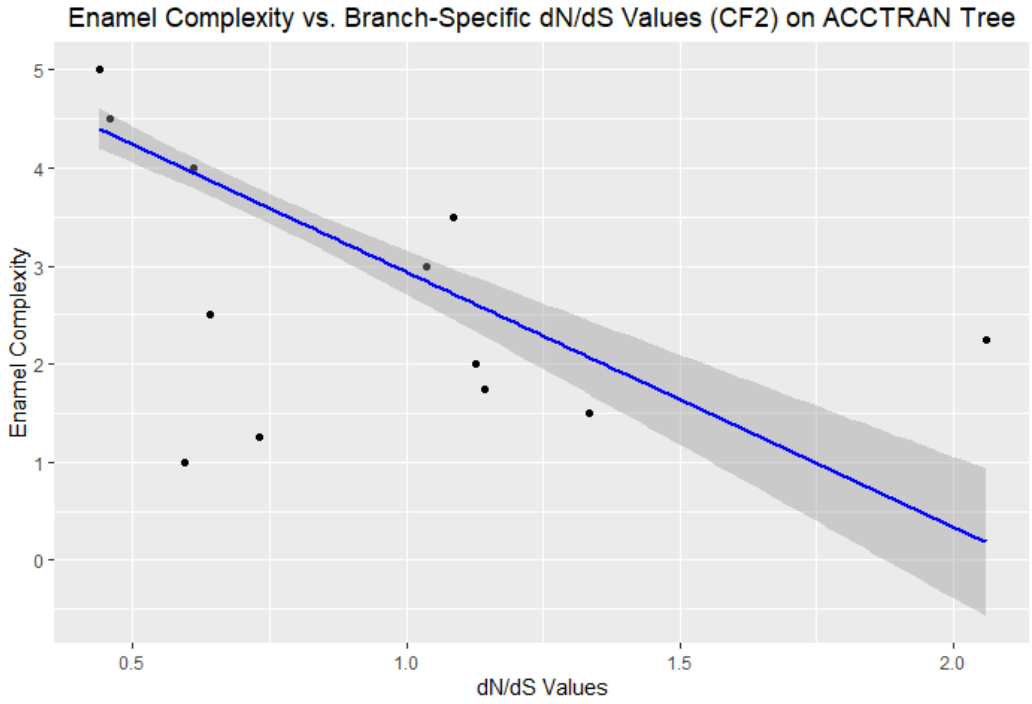
E) Free-ratio dN/dS model (CF1) with MPR/Mesquite reconstructed Werth Categories. Slope: -0.9034. Adjusted R-squared: 0.1672. p-value: 0.0006*.



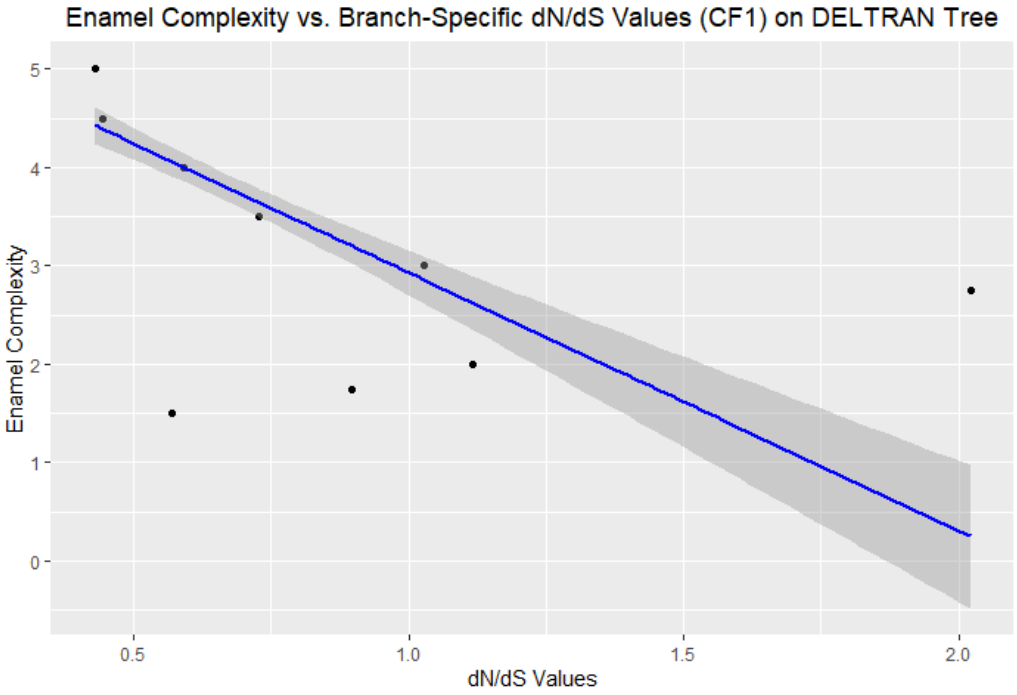
F) Free-ratio dN/dS model (CF2) with MPR/Mesquite reconstructed Werth Categories. Slope: -0.8951 Adjusted R-squared: 0.1773. p-value: 0.0004*.



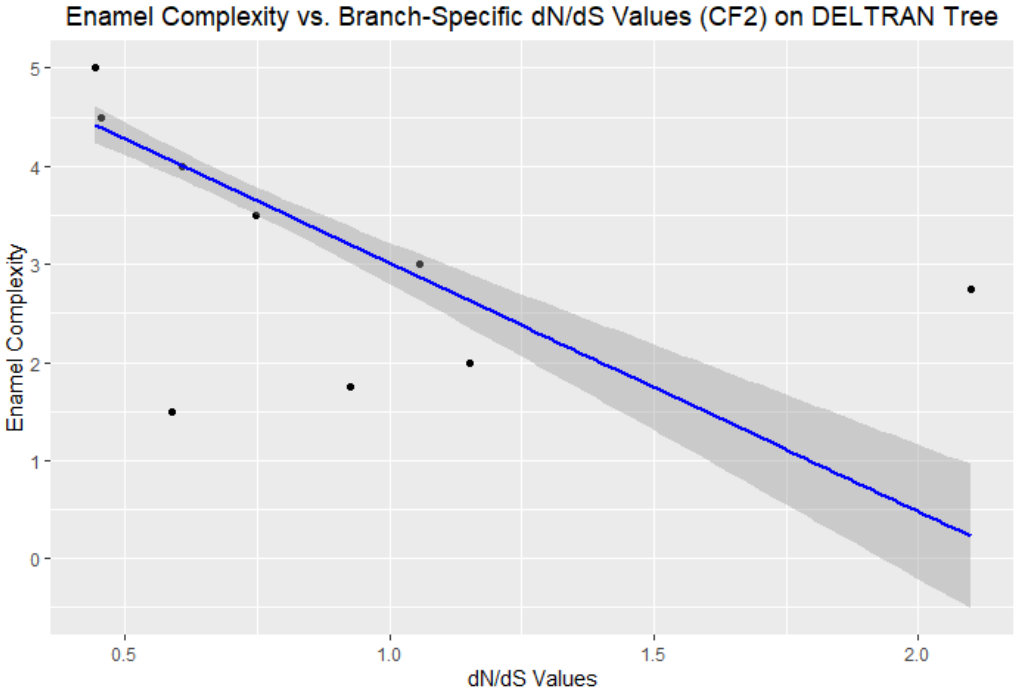
G) Branch-specific dN/dS model (CF1) with ACCTRAN reconstructed Werth Categories. Slope: -2.6820. Adjusted R-squared: 0.5647. p-value: 2.642e-13*.



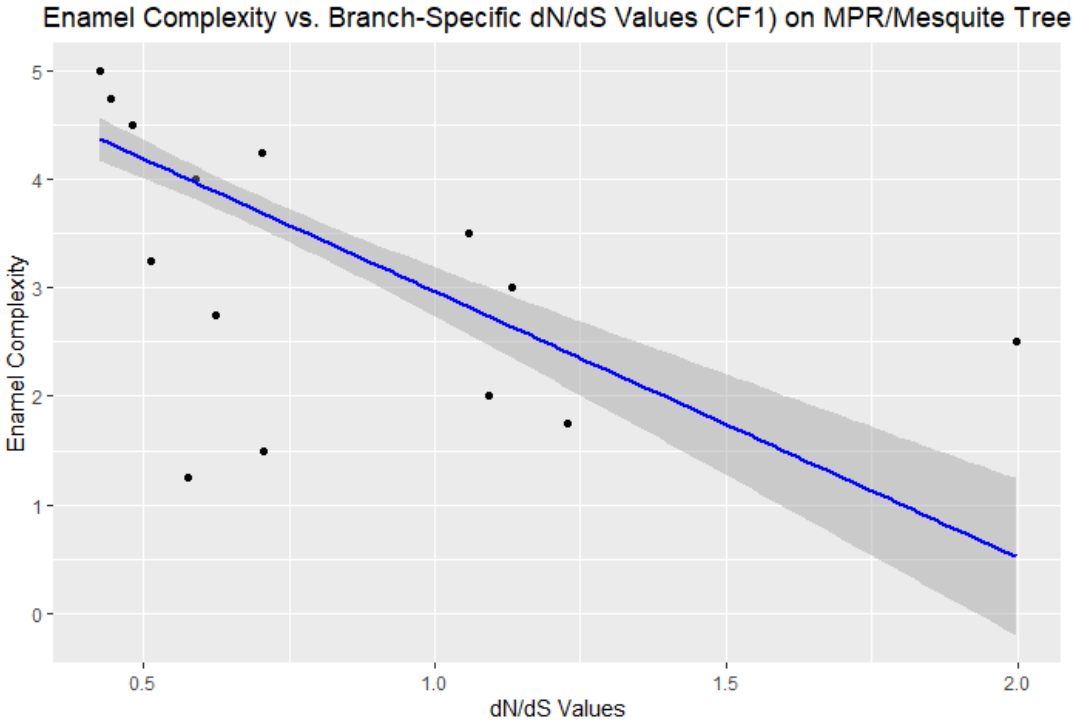
H) Branch-specific dN/dS model (CF2) with ACCTRAN reconstructed Werth Categories. Slope: -2.5993. Adjusted R-squared: 0.5630. p-value: 3.011e-14*.



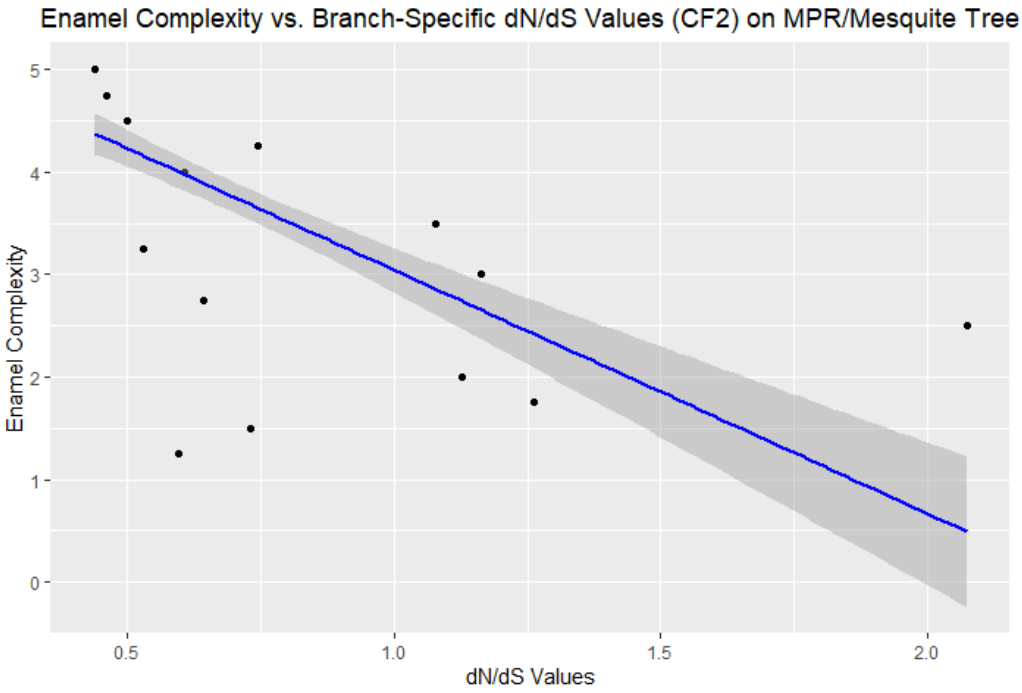
I) Branch-specific dN/dS model (CF1) with DELTRAN reconstructed Werth Categories. Slope: -2.6239. Adjusted R-squared: 0.5821. p-value: 6.377e-15*.



J) Branch-specific dN/dS model (CF2) with DELTRAN reconstructed Werth Categories. Slope: -2.5329. Adjusted R-squared: 0.5789. p-value: 8.284e-15*.



K) Branch-specific dN/dS model (CF1) with MPR/Mesquite reconstructed Werth Categories. Slope: -2.4533. Adjusted R-squared: 0.5400. p-value: 1.8e-13*.



L) Branch-specific dN/dS model (CF2) with MPR/Mesquite reconstructed Werth Categories. Slope: -2.3751. Adjusted R-squared: 0.5372. p-value: 2.231e-13*.

* = significant p-values.

Chapter 3: An Updated Inactivation Date of Enamel and Selection Intensity on Tooth Genes in the Ancestry of Pangolins (Pholidota)

Abstract

Pangolins (Order Pholidota) are toothless mammals with numerous specialized features for myrmecophagy. Cladistic studies suggest that pangolins evolved from toothed ancestors. Extinct palaeanodonts, which had reduced dentitions, are the closest relatives to pangolins. The extant order Carnivora, in turn, is the sister-taxon to Pholidota plus Palaeanodonta. The oldest fossil pangolins (~47 million years) are known from the Messel Pit in Germany and were already edentulous at this time. However, no fossil pangolins possessing teeth have been discovered, complicating a detailed understanding of their evolutionary history of when and how they lost their teeth. Previous efforts to elucidate this transition have begun to investigate the underlying genes necessary for enamel and dentin/tooth production and used results of selection intensity on these genes to determine when these genes and by proxy the associated phenotypes were lost in the common ancestor of pholidotans. To build on and contribute to the story of enamel and tooth loss in pangolins, we targeted seven enamel (*ACP4*, *AMBN*, *AMELX*, *AMTN, ENAM*, *KLK4*, *MMP20*) and two dentin/tooth-specific (*DSPP*, *ODAPH*) genes in the published genomes of three living species of Pholidota (*Manis javanica, Manis pentadactyla,* and *Phataginus tricuspis*) and representative outgroups from Carnivora, Perissodactyla, and Cetartiodactyla. We tabulated inactivating mutations, performed selection analyses, and estimated dates for the commencement of relaxed selection (neutral evolution) on these genes in Pholidota. All nine of the enamel and tooth-specific

genes that we investigated contain numerous inactivating mutations or are completely deleted in each of the three pangolin species. With the exceptions of *KLK4* and *DSPP*, which were not recovered for the two *Manis* species and are presumed to have been deleted, each gene exhibits one or more inactivating mutations that are shared by all three taxa. In addition, dN/dS values suggest an early release from selection pressure on enamel and dentin/tooth genes. We estimate that selection was relaxed on enamel genes, and by proxy the enamel phenotype, ~62–53 million years ago. This estimate accords well with previous reports. By contrast, we were unable to retrieve a reliable estimate for the loss of dentin/teeth because of the small sample size for remnant nucleotides in the associated genes. The timing of enamel loss in the ancestry of pangolins suggests that there may be yet to be discovered fossil specimens with enamelless teeth or teeth with degenerative enamel.

3.1. Introduction

Pangolins (Order Pholidota) are enigmatic mammals known for their overlapping scales and most notably for being the most trafficked animals in the world due to the consumption of their meat as a delicacy and the use of their scales in traditional medicines by people in certain regions (Aisher, 2016; Choo et al., 2016). All species have a conservation status listed as vulnerable, endangered, or critically endangered due to their interactions with humans, low fecundity rates, and habitat loss (Aisher, 2016; Zhang et al., 2015, 2016). The destruction of their natural habitat is due to humans and climate change and has had a negative impact on pangolins as they cannot thrive outside of their narrow habitat range of tropical and inter-tropical zones in Africa and Asia (Gaubert et

al., 2018; Xian et al., 2022). Indeed, the loss of UCP1-mediated non-shivering thermogenesis in pangolins may restrict them to tropical/subtropical environments (Gaudry et al., 2017). In addition, they are most recently known as potential intermediate hosts for COVID-19 (Han, 2020). These unique animals join baleen whales as the only other edentulous clade in the superorder Laurasiatheria, with the only other toothless mammals being anteaters in the order Xenarthra), echidnas in the order Monotremata, and the recently extinct Steller's sea cow in the order Sirenia (Davit-Béal et al., 2009; Meredith et al., 2009; Springer et al., 2016a).

Pangolins are also known as 'scaley anteaters' due to their resemblance to New World anteaters. Both pangolins and anteaters are similarly adapted to myrmecophagy (diet of ants and termites) and have evolved long sticky tongues to compensate for their loss of teeth. However, it is unclear if early pholidotans already evolved this tongue prior to or after relaxed selection pressure on tooth genes (Davit-Béal et al., 2009). Other adaptations to this specialized diet include fossoriality with large claws to retrieve prey, a gizzard-like stomach region filled with keratinized spines, sand, and stones that aid in digestion via grinding that makes up for the lack of mastication by teeth in the jaw, and overlapping scales and strong muscles to close their nostrils and ears for protection against biting (Davit-Béal et al., 2009; Gaubert et al., 2018). Pangolins are also known for having poor vision, but at the same time have an enhanced sense of smell that is made possible by an expanded repertoire of olfactory receptor genes (Choo et al., 2016).

Although pangolins resemble anteaters and were thought to be closely related to them based on morphology (Novacek, 1994), those morphological resemblances are

convergent. Instead, molecular phylogenies have revealed that pangolins are the sister-group of Carnivora in the clade Ferae (Foley et al., 2023; Gaubert et al., 2018; Meredith et al., 2011b; Murphy et al. 2001). Ferae also includes the extinct taxon Palaeanodonta, which is the stem group to Pholidota (Gaudin et al., 2009). Together, Pholidota and Palaeanodonta comprise the clade Pholidotamorpha. Whereas the monophyly of Ferae is well supported (Foley et al., 2023; Meredith et al., 2011), the relationship of Ferae to other laurasiatherian mammals has been highly contested. Doronina et al. (2022) discussed the difficulty in discerning laurasiatherian phylogenetic relationships with obtaining contradictory results in different datasets that suggested that Ferae is either sister to Cetartiodactyla or to a clade of Cetartiodactyla + Perissodactyla. However, with the increase in whole-genome sequence availability, some recent publications have utilized genome-wide sequence alignments in their phylogenetic analyses of placental mammals and have obtained results with Ferae sister to Perissodactyla (Foley et al., 2023; Murphy et al., 2021).

The most recent common ancestor of Ferae occurred ~79.8–66 million years ago (Ma), with crown Manidae appearing ~41.34–25.26 Ma which subsequently split into African and Asian lineages, with the African group bifurcating into larger terrestrial (*Smutsia*) and smaller arboreal (*Phataginus*) clades, with *Manis* being the sole genus of Asia (Gaubert et al., 2018). Based on molecular clock dating techniques, there is a consensus on an approximate age for the Ferae divergence, with calculations ranging from 79.8–66 Ma, corresponding to the Upper Cretaceous before the K-Pg boundary (Heighton et al., 2023: 79.47 Ma [95% highest probability density {HPD} 87.24–67.66

Ma]; Foley et al., 2023: Benton averaged IRM and ARM model 73.62 Ma [95% Confidence Interval 79.33–67.89]; Foley et al., 2016: 73.35 Ma; Gaubert et al., 2018: 78.9 Ma [95% HPD 87.0–69.6 Ma]; references within Murphy and Eizirik, 2009: 79.8 [95% Credibility Interval 85–75 Ma], 74.0 Ma [95% Credibility Interval 81–67 Ma]; two methods in Zhou et al., 2012: 76 Ma [95% HPD 87–58 Ma], 66 Ma [95% HPD 82–55 Ma]; Murphy et al., 2021; Springer et al., 2011).

Pholidota is currently classified into three families, two of which are extinct (Eomanidae [middle Eocene *Eomanis*] and Patriomanidae [late Eocene *Patriomanis* and *Cryptomanis*]) and one of which is extant (Manidae). Two other genera from the middle Eocene (*Euromanis* and *Eurotamandua*) have not yet been assigned to a family (Gaudin et al., 2009). Manidae are currently recognized as having eight extant species among three genera (*Manis*, *Smutsia*, and *Phataginus*). However, Gaubert et al. (2018) suggested there may be as many as 13 species based on species delimitation analyses with mitogenomes. The possible additions to the number of extant species are the result of splitting *Phataginus tricuspis* into as many as six species.

3.1.1. Evolutionary history

All members of Pholidota, living and extinct, are edentulous, and the finding that they are most closely related to palaeanodonts and carnivorans and not other toothless taxa has led researchers to investigate when, where, and how they lost their teeth. However, the lack of fossils for this clade makes it difficult to elucidate an accurate account of when they lost their teeth, which will be addressed here in this study. This sparse fossil record is partly due to the facts that pangolins (1) do not have teeth

traditionally used to assist in fossil reconstructions, (2) occur in low population numbers, and (3) inhabit tropical forest biomes that tend to not preserve fossils as readily as other environments (Gaudin et al., 2019a). There is currently only an estimated upper and lower bound of tooth loss on the stem Pholidota branch, with the earliest known edentulous pangolin found in the fossil record and the last known relatives that possessed teeth, members of the strongly supported sister-group Palaeanodonta (discussed below) (Gaudin et al., 2009, 2019b). Knowledge of these estimated time boundaries will help clarify the timing and circumstances that occurred in the evolution of tooth loss in extinct pangolins.

Pholidota presumably had its origin in Laurasia, with the oldest fossils being discovered in Europe in the Messel fauna deposits of Germany in the middle Eocene around ~45 Ma. This fauna includes the taxa *Eomanis*, *Eurotamandua*, and *Euromanis* (now on referred to as the Messel taxa for brevity and clarity) (Gaudin et al. 2009, 2019a). It has been debated which of these three genera is the oldest pholidotan as all three were found in the Messel deposits and have had their phylogenetic placements reconstructed differently by separate authors (Gaudin et al., 2009; Rose 2012). Rose (2012) mentions that *Eomanis* has been considered to claim the basal node for this clade based on its possession of several known pholidotan synapomorphies. Storch (2003, as cited in Gaudin et al., 2009) concluded that *Eomanis* is the oldest pangolin based on its combination of pangolin synapomorphies and plesiomorphic features and that are more similar to Palaeanodonta than Pholidota. However, other attempts to parse together these phylogenetic relationships have instead placed *Euromanis* as the basal taxa of Pholidota

(Gaudin et al., 2009). Gaudin et al. (2009) made this distinction as *Eomanis* had been comprised of two species, *Eomanis waldi* and "*Eomanis" krebsi*, the latter of which was placed into the new genus *Euromanis* as it was determined to be older than *Eurotamandua joresi* that was even older than *Eomanis waldi*, which made the original *Eomanis* no longer monophyletic. This study also linked *Eurotamandua* with more inclusive pangolins with the exclusion of *Euromanis krebsi* with the observation of several synapomorphies that were not possessed by *E. krebsi* (Gaudin et al., 2009). In addition, Gaudin et al. (2009) singled out *E. waldi* as being associated with more inclusive taxa with its possession of epidermal scales and then further suggested this species could justifiably be the basal node of Pholidota proper due to this recognizable feature of modern "scaly" pangolins. However, several recent studies have formed a consensus that suggest the oldest fossil species is *Eomanis,* with radiometric dating increasing the age of the deposits this fossil was discovered in to around 47 Ma (Franzen, 2005; Rose, 2012; Springer et al., 2019). This species therefore provides helpful information on setting the upper time boundary of the appearance of edentulous pangolins.

For the lower time boundary of tooth loss, knowledge of the most recent toothed ancestor is invaluable. The closest known extinct relatives of pangolins are currently presumed to be the order Palaeanodonta, although this hypothesis was only established recently. It was long debated whether this group was closer to xenarthrans or pangolins and had been grouped with both in the now invalid order "Edentata" which was determined to be polyphyletic (Davit-Béal 2009, Gunnell and Gingerich, 1993). Through

extensive morphological analyses, palaeanodonts are now undoubtedly the ancestral relatives of Pholidota. This was determined through larger datasets of scored osteological characters and a detailed examination of the anatomy of the inner ear which discovered and described novel derived auditory features (Gaudin et al., 2009, 2019b; O'Leary et al., 2013). Given the known affinities between these two orders, it is beneficial to examine the morphology of teeth in palaeanodonts to potentially aid in resolving the state of teeth that were in their most recent common ancestor as well as early stem pholidotans to reveal more about the evolution of tooth loss in pangolins.

These extinct toothed anteater-like mammals were small burrowing mammals that ranged from the early Paleocene to early Oligocene and had reduced dentition with a majority of palaeanodonts possessing no incisors, one large canine, and various numbers of postcanine teeth that were not strongly rooted in the jaw (Davit-Béal et al., 2009). The earliest forms had teeth with thin enamel with a trend of even more enamel reduction and the loss of postcanine teeth (Schoch, 1984). Palaeanodonta has four proposed families, including Escavadodontidae, Epoicotheriidae, Ernanodontidae, and Metacheiromyidae, each having its first represented fossil species discovered in the Paleocene (Kondrashov and Agadjanian, 2012, Rose and Lucas 2000, Secord et al., 2002). However, their interfamilial relationship to each other and to Pholidota is not totally clear. For taxa listed below, genera listed without a species and their description of dental anatomy is assumed to be from the type species for brevity.

The oldest palaeanodonts come from the Paleocene in North America, with the earliest fossil discovered in the early Paleocene (Torrejonian, New Mexico) named

Escavadodon zygus and became the type species for the newly created family Escavadodontidae (Rose & Lucas, 2000). This species has dental features akin to other mammals, with having the general placental postcanine tooth count of four premolars and three molars (4:3), with all cheekteeth having two roots except for premolar 1. The authors that described the species, Rose and Lucas (2000), mentioned that their phylogenetic analysis revealed *Escavadodon* as the sister taxon to the palaeanodonts included in their study and considered it to be the most primitive palaeanodont based on derived similarities. The earliest known Epoicotheriidae species *Amelotabes simpsoni* (Late Paleocene, Wyoming) depicts the beginning of progressive dental simplification in this family, with thin enamel on the occlusal surfaces of cheekteeth and one single-rooted premolar. Although *Amelotabes* still has four premolars and three molars (4:3), later genera of epoicotheriids have lost premolars, with *Pentapassalus* (Early Eocene) having five upper (2:3) and six lower (3:3) postcanines, *Tubulodon* (Early Eocene) believed to have four or five lower postcanines (1:3 or 2:3), and *Xenocranium* (Early Oligocene) having either four or five upper postcanines (Rose 2008; Schoch, 1984). In addition to tooth loss, some taxa are also found to have postcanines that have regressed to small, simple peg-like teeth with single-roots (Gunnell and Gingerich 1993; Secord et al., 2002). The most derived palaeanodonts in terms of dental anatomy is the Metacheiromyidae family. Secord et al. (2002) listed several characteristics that have been used to diagnose metacheiromyids from epoicotheriids, including simplifications to postcanine teeth such as the loss of enamel, a decrease in the number of teeth, reduction to simple peg-like structures, and larger alveoli than the teeth in them. The earliest metacheiromyid known

is *Propalaeanodon* from the late Paleocene (Wyoming), which has similar dentition to the other two Paleocene species from Escavadodontidae and Epoicotheriidae (*Escavadodon* and *Amelotabes* respectively) with each having a postcanine tooth count of four premolars and three molars (4:3) (Schoch 1984; Secord et al., 2002). However, unlike epoicotheriids, subsequent species drop to postcanine tooth counts of 4:2 in *Mylanodon*, to 4:1 in *Palaeanodon* and *Brachianodon* and then most reduced in *Metacheiromys* with two cheekteeth (1:1) (Gunnell and Gingerich, 1993; Schoch, 1984; Secord et al., 2002). The most recent of these metacheiromyids were found in the middle Eocene in North America (Wyoming) and had the most regressed postcanine teeth among palaeanodonts, where they either lacked enamel or had a thin layer of enamel that would wear rapidly and were small vestigial single-rooted peg-like teeth (Gunnel and Gingerich, 1993; Schoch 1984).

It is curious that it appears that the trend of postcanine loss in metacheiromyids occurred from back to front, whereas in epoicotheriids it occurred from front to back, or at least started in the front since the molars are still found in the later taxa. Knowledge of the order of tooth loss could provide some insight into the order that they developed as the first tooth lost in a series (premolars versus molars) is the last tooth to develop; although carnivorans are known to have variable eruption patterns, we can see some similarities to the possible eruption patterns in the nearby related palaeanodonts and pholidotans (Berkovitz and Shellis, 2018; Slaughter et al., 1974). In addition to the gradual loss of postcanines and enamel, there is also the observation of the regression of roots progressing through time in metacheiromyids and epoicotheriids, with a gradual

reduction of the number of teeth with double-roots to single-roots. This possibly can provide some information about when palaeanodonts transitioned from a possible diet of small invertebrates to that of myrmecophagy in living pangolins as the larger surface area of roots in postcanines correlates with a stronger bite force and a harder diet (Berkovitz and Shellis, 2018; Rose, 2008).

Given the observations on the progressive dental simplification in Palaeanodonta, and assuming that Palaeanodonta is a paraphyletic stem group and not a monophyletic sister-group to Pholidota, it would be beneficial to clarify which palaeanodont family Pholidota diverged from as well as when this divergence occurred. To aid in achieving this, the interfamilial relationships among palaeanodonts needs to be resolved. It has been proposed that *Escavadodon* is plausibly the ancestor of all later palaeanodonts since it is the oldest known fossil discovered coming from the early Paleocene and has morphological similarities to the oldest epoicotheriid, *Amelotabes*, and oldest metacheiromyid, *Propalaeanodon* (Rose, 2008). It has also been suggested that *Amelotabes* may represent the primitive morphological condition for Metacheiromyidae (Secord et al., 2002). Of the purported morphological links between these groups, there exists similarities in dental and jaw morphology, with *Amelotabes* already showing dental regression and metacheiromyids possessing the most specialized jaw morphology and tooth regression among Palaeanodonta (Schoch, 1984; Secord et al., 2002). If indeed Epoicotheriidae and Metacheiromyidae are closely related, an ancient metacheiromyid would have needed to diverge from an epoicotheriid near *Amelotabes*, as a majority of Epoicotheriidae species have postcanine teeth that extend to the back of the jaw with a

thin layer of enamel, while species of Metacheiromyidae begin a trend of tooth simplification and loss (Rose, 1979, as cited in Secord et al., 2002). Regardless, it is possible that each of these families ostensibly evolved independently from a nearby related common ancestor of *Escavadodon* at the basal node of Palaeanodonta due to the disparity in morphology and dental anatomy in later taxa as previously mentioned (Rose 2008). Indeed, some of these characteristics may have been lost independently via convergence in separate Eocene palaeanodonts, with some leading to an evolutionary dead-end and the other representing the true stem species; this hypothesis is also plausible with the understanding of Dollo's law of irreversibility (Marshall et al., 1994; Secord et al., 2002). This seems to be the case for at least the metacheiromyid *Brachianodon westorum*, which was observed as having possessed the most robust enamel among its family albeit being from the middle Eocene (Gunnell and Gingerich, 1993).

As for the relationship and divergence between Palaeanodonta and Pholidota, it had been suggested in some older studies that metacheiromyids could be the primitive condition for pangolins due to the noticeable tooth reduction in that clade (Emry, 1970 as cited in Schoch 1984). However, with oldest known pangolins being discovered from the middle Eocene (the Messel taxa), they were contemporaneous with middle-late Eocene to early Oligocene palaeanodont species (like *Epoicotherium* and *Xenocranium*), so a phylogenetic relationship between the two seems similar to that of the divergence among palaeanodont families, with a divergence time that would have most likely occurred in the early Paleocene or even earlier (Rose, 2008). This suggests that the immense tooth

regression in these later palaeanodonts is convergent with stem pholidotans and that there may be fossil pangolins with teeth that have yet to be discovered (this is further supported from molecular clock studies on the origin of Pholidota) (Rose, 2008). If indeed the loss of teeth in Palaeanodonta and Pholidota are convergent, the progress of tooth regression in palaeanodonts that was partially outlined here could still provide evidence on how tooth loss occurred in pangolins as both inhabited similar ecological niches.

Recent molecular based studies have contributed to understanding the relationship between Palaeanodonta and Pholidota by investigating the divergence time between pangolins and their closest living relatives Carnivora. This sets a lower time boundary on the origin of palaeanodonts (and supports the lower time boundary of tooth loss in pangolins) as they are located early on the stem Pholidotamorpha branch (Palaeanodonta + Pholidota) that diverged from the Ferae node. These range of dates for the divergence of Ferae fits with the fossil record of palaeanodonts as the average dates are older than the oldest palaeanodont currently known, *Escavadodon*, found at the end of the early Paleocene around 61 Ma (Rose & Lucas, 2000). Given that *Escavadodon* and later palaeanodonts that were contemporaneous with the first pangolins are found in North America along with palaeanodonts found in Asian and Europe, this may suggest an early divergence and dispersal across Laurasia, with Pholidota evolving from Palaeanodonta in Europe during the early Eocene or Paleocene (Gheerbrant et al., 2005; Kondrasov and Agadjanian, 2012; Rose, 2012). This would have had to occur when there was land still connecting Europe and North America in the early Eocene (Gheerbrant et al., 2005; Holynska et al., 2016). However, the land connecting these subsequent continents in

Laurasia, known as the Thule Land Bridge, that connected North America to Eurasia through Greenland may have actually existed up until the Oligocene (Seton et al., 2012). This may explain the discovery of *Patriomanis*, an extinct pangolin species more derived than the Messel taxa, in North America and Asia in the late Eocene (Rose, 2012). It is plausible that after the divergence of the Messel taxa from a palaeanodont in Europe there was a migration of *Patriomanis* related pangolins back to North America from Europe before the breakup of Laurasia (Rose, 2008). If this hypothesis is true, there may be older fossil pangolins that have not yet been discovered in North America. This is possible with the recent discoveries of fossil pangolins that are now filling in parts of the fossil record that have had missing taxa. These include the newly described genus *Necromanis* from Oligocene-Miocene deposits in the Iberian Peninsula in Europe that place it at an age of around 16 Ma (Alba et al., 2018). There was also the more recent discovery of a fossil pangolin species placed in the genus *Smutsia* found in Europe in sediment aging this specimen to ~2.2–1.9 Ma. This suggests that this genus had a much larger biogeographic range and were in Europe during the Pleistocene. This also implies that they inhabited a more open grasslands and woodlands environment, different from previous inferred studies and the environments of living pangolins (Terhune et al., 2021).

3.1.2. Previous reports on tooth gene evolution and inactivation in pangolins

The calculations for the age of divergence of Ferae and the hypothetical convergence of tooth loss between pangolins and palaeanodonts is further corroborated with molecular studies investigating the timing of tooth loss in pangolins. This was conducted by examining the genes responsible for tooth production to see if they are still

functional. Genes that were determined to have become pseudogenic via inactivating mutations were utilized in dN/dS selection analyses to determine selection intensity and subsequently the timing of relaxed selection on those genes. This in turn was used to reconstruct inactivation times of those tooth genes and by proxy the inferred loss of the encoded phenotypes to determine when they were lost during the evolution of ancient stem pholidotans. These analyses are expanded on in the research presented here in this study.

Previous studies have determined that several genes responsible for proper tooth development including enamel and dentin production have inactivating mutations in the pangolins investigated (*Manis javanica* for *ACP4*, *AMBN*, *AMELX*, *ENAM*, and *ODAM*, *Phataginus* ["*Manis*"] *tricuspis* for *ACP4*, *AMBN*, *AMELX*, and *ENAM* and *Manis pentadactyla* for all the genes listed below). The genes these studies investigated include the enamel-specific genes acid phosphatase 4 (*ACP4* [previously = *ACPT*]), ameloblastin (*AMBN*), amelogenin X-linked (*AMELX*), amelotin (*AMTN*), enamelin (*ENAM*), and matrix metallopeptidase 20 (*MMP20*), the dentin/tooth-related genes dentin sialophosphoprotein (*DSPP*) and odontogenesis associated phosphoprotein (*ODAPH* [previously = *C4orf26*]), and a tooth development-related gene odontogenic ameloblast-associated (*ODAM*) (Choo et al., 2016; Meredith et al., 2014; Springer et al., 2016a, 2019). These genes (along with *KLK4* mentioned below) have been classified as being enamel, dentin, or tooth related based on inactivation of these genes in enamelless or edentulous vertebrates as well as diseases in humans such as amelogenesis imperfecta and dentinogenesis imperfecta, respectively. *ODAM* is implicated as being a tooth

development-related gene and not necessarily as enamel or dentin-specific because of its inactivation in mammals with enamel-capped teeth as well as enamelless and toothless species (Springer et al., 2019). In addition to the reported inactivating mutations in the genes mentioned above, the enamel-specific gene kallikrein related peptidase 4 (*KLK4*), was also investigated for mutations. *KLK4* was believed to be lost in the pangolins examined (*M. pentadactyla* and *M. javanica*) based on its absence in the genomes examined via BLASTN searches (Mu et al., 2021b). Emerling et al. (2017) included *KLK4* in a list of known genes that have been inactivated in enamelless taxa but this gene was not indicated as being pseudogenic or deleted in the pangolins that were investigated.

The only two genes that were investigated for the timing of relaxed selection and inactivation time so far include *ENAM* and *ODAM*. The loss of *ENAM* was estimated to occur 59.4–54.9 Ma, with *ODAM* inactivated ~65 Ma (average of different models used ranging from 73.04–57.34 Ma) (Meredith et al., 2009; Springer et al. 2019). These estimates for *ENAM* and *ODAM* loss are both older than the oldest fossil pangolin *Eomanis* dated at ~47 Ma. This suggests that tooth loss or at least enamel loss occurred well before the earliest known edentulous pangolin. Also, this estimated timing of enamel loss is younger than the calculated age of divergence of Pholidota from Carnivora (79.8–66 Ma) and further suggests there may be a significant ghost lineage of older fossil pholidotans that may reveal the progression of tooth simplification and tooth loss akin to palaeanodonts. Furthermore, the presence of palaeanodonts with enamel in the Eocene and the calculated loss of enamel in stem pangolins in the Paleocene suggests that the loss of enamel and tooth regression is convergent. It is possible that there may have been

an intermediate period of enamelless teeth between the transition from enamel-covered teeth to edentulism in pangolins, as there was a drastic reduction in enamel thickness noted in most palaeanodont species (Davit-Béal et al., 2009). This is plausible with the living taxonomic groups that lack enamel on their teeth, such as sloths, armadillos, aardvarks, and some odontocetes.

3.1.3. This study

A more comprehensive examination into the genes responsible for tooth production is explored here to tabulate inactivating mutations in taxa that were not previously obtained. This study includes *M. pentadactyla*, *M. javanica*, and *P. tricuspis*, adding at least one species not used in previous studies. This study also investigates several additional genes for both *P. tricuspis* and *M. javanica* not previously retrieved. Additionally, complete coding sequences were recovered for all three pangolin species for *ACP4*, and several exons of *KLK4* from *P. tricuspis*. *ENAM* was examined for shared inactivating mutations with the inclusion of all three taxa in this research. Along with these observations, there are additional outstanding research questions being addressed in this study: (1) How many inactivating mutations have accumulated in the seven enamel and two dentin/tooth specific genes obtained for all three pangolin species included and are there mutations that are shared among taxa, and (2) were teeth lost in a single step or did enamel loss precede dentin/tooth loss? It is expected that there will be inactivating mutations observed in all the tooth genes obtained here in each of the taxa included as well as shared mutations between *Manis* and among all three Manidae species given the inactivating mutations discovered in the previous studies as well as the amount of time

since their divergence from their toothed relatives and the earliest known edentulous fossil. Similar to the previous research methods mentioned above, shared inactivating mutations observed in these genes that confirm inactivation on the stem transition branch are used for dN/dS selection analyses and then employed to calculate the timing of pseudogenization of those tooth genes and the inferred loss of the respective phenotypes. In addition to the expected inactivating mutations present in these genes, the rates of evolution for nonsynonymous compared to synonymous mutations (dN/dS values) should be elevated compared to outgroups with enamel-covered teeth. The timing of inactivation of these genes and phenotypes should be like that of the previous studies (~65–54.9 Ma) and at least older than the age of *Eomanis* (~47 Ma). Furthermore, with separating out enamel from dentin/tooth genes, we hypothesize that enamel may be inactivated before dentin/tooth genes, with an intermediate enamelless condition in extinct pangolins akin to what is observed in their palaeanodont relatives and many other taxonomic groups such as the discovery of the decoupling of enamel and dentin/tooth loss in baleen whales (Randall et al., 2022). This is also plausible given the observed presence of vestigial teeth that begin to form in fetal pangolins but are subsequently resorbed prior to birth which is similar to what is observed in baleen whales (Mu et al., 2021b; Randall et al., 2022). All of the tooth genes investigated in pangolins mentioned above are inactivated among most baleen whale species examined.

By incorporating these additional species and genes obtained for this study, further progress can be made in resolving the story of tooth loss in pangolins, ultimately aiding in elucidating the timing of enamel and dentin/tooth loss. Moreover, these answers

may further aid in helping to explain and unravel the evolutionary history and adaptation to their specialized ecological niche in these enigmatic mammals.

3.2. Methods

3.2.1. Gene sampling

Nine genes were included in for this study, including seven enamel-specific (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*) and two dentin/tooth-specific genes (*DSPP*, *ODAPH*) that were chosen due to evidence that they were inactivated in either enamelless or toothless clades of vertebrates respectively (Deméré et al., 2008; Gasse et al., 2012; Kawasaki et al., 2014; McKnight and Fisher, 2009; Meredith et al., 2009, 2011, 2013, 2014; Mu et al., 2021; Springer et al., 2016a, 2019). Additional evidence for the enamel or tooth specificity of these genes derives from mutagenesis studies in mice and natural genetic variation in humans that causes nonsyndromic cases of amelogenesis imperfecta, dentinogenesis imperfecta, or dentin dysplasia (Gasse et al., 2012; Hu et al., 2008; Kawasaki et al., 2014; Meredith et al., 2009, 2011, 2013, 2014; Mu et al., 2021; Parry et al., 2012; Smith et al, 2016, 2017; Springer et al., 2016a; Wright et al., 2009). A tenth tooth-related gene (*ODAM*) is also inactivated in enamelless and edentulous mammals, but this gene is also pseudogenized in all toothed whales that were investigated as well as several other clades of mammals with enamel-capped teeth (Springer et al., 2019). To help clarify the difference in activation between enamel and dentin/tooth related genes, we omitted *ODAM* from our study. Also, exon 4 of *AMELX* was not included in our analyses because this exon is subject to alternative splicing and is absent in many mammals (Delgado et al., 2005; Sire et al., 2005, 2006, 2007). Similarly,

exon 2 of *ODAPH* is not included due to this short exon being involved in an alternative transcript variant with a functional copy of this exon being deleted from many of the placental mammals investigated in Springer et al. (2016a), and further using exon 2 with an alternative splice site changes the reading frame of exon 3 introducing premature stop codons even further complicating dN/dS analyses.

3.2.2. Taxon sampling

A total of 20 taxa were used including three pangolins (Pholidota) and 17 outgroups from three different orders: five carnivores (Carnivora), seven odd-toed ungulates (Perissodactyla), and four even-toed ungulates (Cetartiodactyla). Pangolins included *Manis javanica* (Sunda pangolin), *Manis pentadactyla* (Chinese pangolin), and *Phataginus tricuspis* (tree pangolin). Carnivorans included *Acinonyx jubatus* (cheetah), *Canis lupus* (dog [domestic]), *Cryptoprocta ferox* (fossa), *Lutra lutra* (Eurasian otter), *Odobenus rosmarus* (walrus), and *Ursus arctos* (brown bear). Perissodactyls used include *Ceratotherium simum* (southern white rhino), *Diceros bicornis* (African black rhino), *Equus caballus* (horse [domestic]), *Rhinoceros unicornis* (Indian rhino), *Tapirella bairdii* (Baird's tapir), *Tapirus indicus* (Malayan tapir), and *Tapirus terrestris* (South American tapir). Lastly the cetartiodactyls are *Bos mutus* (wild yak), *Camelus bactrianus* (Bactrian camel), *Capra hircus* (goat [domestic]), and *Sus scrofa* (wild boar).

3.2.3. Data collection

DNA sequences for all nine genes were collected either from assembled genomes at NCBI (https://www.ncbi.nlm.nih.gov/), or through assemblies provided on DNAZOO (https://www.dnazoo.org/, Dudchenko et al., 2017) (see Supplemental Table S3.1 for

sources and accession numbers for each species). NCBI's Nucleotide databases were searched using keywords for all nine genes in conjunction with taxon names for four reference species, two included in this study, and two from a previous study (*Camelus bactrianus*/*Capra hircus,* and *Orcinus orca/Tursiops truncatus* respectively). Sequences for each reference species were then imported into Geneious Prime (current version 2023.2.1, https://geneious.com) (Kearse et al., 2012), aligned with MAFFT (Katoh and Toh, 2008), and cross-checked against each other for consistent annotations. These sequences were then utilized to obtain sequences for additional species through NCBI's Nucleotide Basic Local Alignment Search Tool (BLAST), which was used to search assembled genomes using the whole-genome shotgun (WGS). Each BLAST search employed a query sequence from a closely related species. Megablast was used for highly similar sequences (e.g., taxa in same family), whereas Blastn was used for less similar sequences (e.g., taxa in different families). Top-scoring BLAST results were imported into Geneious Prime. Sequences obtained from DNAZOO were Hi-C genome assemblies that were downloaded and imported into Geneious Prime to be used as a BLAST database to be searched against using query-sequence alignment.

3.2.4. Alignments and inactivating mutation annotation

Complete protein-coding sequences and introns were aligned in Geneious Prime using MAFFT (Katoh and Toh, 2008). Sequences were manually spot-checked for alignment errors using AliView version 1.28 (Larsson, 2014). Alignments were examined for inactivating mutations (frameshift indels [insertions and deletions], start and stop

codon mutations, premature stop codons, and splice site mutations), which were annotated in Geneious Prime.

3.2.5. Phylogenetic analyses

Protein coding sequences were used to construct gene trees with maximum likelihood using the program RAxML in the CIPRES Science Gateway (RAxML-HPC2 on XSEDE tool, bootstrapping phase: GTRGAMMA; https://www.phylo.org; Miller et al., 2010; Stamatakis, 2006). The GTRGAMMA option was used which implements the GTR + Γ model of sequence evolution. Rapid bootstrap analyses were performed with 500 pseudoreplicates to search for best trees, all other parameters were set to default (Stamatakis et al., 2008).

3.2.6. Selection analyses

Selection analyses (dN/dS) were conducted in the codeml program of PAML (version 4.9j, Yang 2007). Analyses were conducted on each of the seven enamel genes (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *KLK4*, *MMP20*), a concatenation of these seven genes, each of the two dentin/tooth genes (*DSPP*, *ODAPH*), and a concatenation of these two genes. For the dentin/tooth gene analysis, it should be noted that although the length of the stem branch was the same as used for the enamel gene analyses, most of the sequence data comes from *P. tricuspis* as both *Manis DSPP* sequences are presumed to be deleted from negative BLAST/map to reference results, but all three species having mostly intact coding sequences for *ODAPH* (*DSPP*: 1659 base pairs [bp], *ODAPH*: 537 bp). For the concatenated datasets, a free-ratio model where all the branches are estimated independently and a branch-specific codon model based on the phenotypes of

the taxa were utilized (background, transitional, pseudogenic). The branch-specific model was also used for analyses of individual genes. The background category of branches lead to internal nodes or extant taxa with functional dentition with enamel-capped teeth, transitional branches lead to internal nodes that are edentulous, and pseudogenic branches post-date the occurrence of transitional branches. Background category branches are expected to have evolved under purifying selection ($dN/dS < 1$) and pseudogenic branches expected to have evolved neutrally ($dN/dS \approx 1$) after the loss of enamel or teeth on the stem transitional branch. The transitional stem pangolin branch has a mixed evolutionary history with a portion of the branch evolving under purifying selection followed by a period of neutral evolution when teeth were lost and selection was relaxed (Meredith et al., 2009). The results of the selection analysis (dN/dS) on this branch are expected to be intermediate between the dN/dS values of the background and pseudogenic branches. Analyses were conducted with a fixed and estimated value for the pseudogenic branch ($dN/dS = 1.0$ or dN/dS = estimated). The branch categories were labeled as follows: background (#0), transitional (#1), and pseudogenic branches (#2).

These selection analyses were employed with two codon frequency models, CF1 and CF2 (Yang, 2007). CF1 estimates codon frequencies from mean nucleotide frequencies across all three codon positions, whereas CF2 estimates frequencies at each of the individual codon positions. Codon positions are absent in pseudogenes, so it is important to verify that analyses without base compositional differences at different codon positions (CF1) yield results that are similar to results that are obtained with a codon frequency model that allows for base compositional differences at 1st, 2nd, and 3rd

codon positions (CF2). Each of the CF1 and CF2 analyses were conducted with both fixed (dN/dS = 1) and estimated values for the fully pseudogenic branch category (dN/dS = 1.0 or dN/dS = estimated); chi-square tests were conducted using the program chi2 (Yang, 2007) to determine whether the analyses with fixed versus estimated dN/dS values for the pseudogenic branch category were significantly different from each other. All frameshift insertions were deleted prior to performing dN/dS analyses. In addition, premature stop codons were recoded as missing data as required for codeml analyses. The species tree utilized for these analyses were delineated and combined from Gaubert et al. (2018) for pangolins and Foley et al. (2023) and Westbury et al. (2017) for outgroups.

Selection analyses were also constructed in the program Coevol (version 1.6, Lartillot and Poujol, 2011) for the enamel and dentin/tooth gene concatenated datasets. Coevol utilizes a Bayesian approach and provides a visual representation of dN/dS ratio estimates varying across a phylogeny (Lartillot and Delsuc, 2012). The *dsom* procedure was used and employs a codon model which uses dS and dN/dS as a priori independent variables. The tree topology used for this analysis was the same species tree used in the codeml analyses. Each data set ran as two independent MCMC chains for at least 1000 cycles, sampling parameters every cycle. The *tracecomp* command was used to check for MCMC convergence by monitoring effective sample size. For the first analyses, the enamel data set had a burn-in of the first 201 cycles, leaving 809 sampled cycles and dentin/tooth data had a burn-in of 269 cycles, leaving 1080 cycles sampled. The second analysis for enamel had a burn-in of 230, leaving 922 sampled cycles, and the second dentin/tooth analysis had 335 burn-in leaving 1341 sampled cycles.

3.2.7. Gene inactivation times

Inactivation times for the concatenation of seven enamel genes and for the concatenation of two dentin/tooth genes were each estimated using equations from Meredith et al. (2009) that allow for either one or two synonymous substitution rates. The one synonymous substitution rate model assumes that the rate of synonymous substitution is neutral and equal on both functional and pseudogenic branches, whereas the two-rate model assumes that the synonymous substitution rate on functional branches is non-neutral and is 70% of the substitution rate on pseudogenic branches (Bustamante et al., 2002; Meredith et al., 2009). Divergence time estimates were taken from Gaubert et al. (2018).

3.3. Results

3.3.1. Alignments and gene trees

Complete coding sequences with all exons were recovered for all outgroups except for two exons among the tapirs that were sampled, (*Tapirella bairdii*: *AMELX* E7 and *Tapirus terrestris*: *KLK4* E2, no reads mapped [NRM] to the reference sequence and negative BLAST results [NBR]). The coding sequences for the three pangolins sampled were mostly recovered aside from the absence of all the exons from *KLK4* and *DSPP* for both *Manis* (NRM/NBR), and then varying degrees of completeness among the other genes for all three species (Table 3.1). The presence or absence of 12 well-supported Laurasiatheria clades (Foley et al., 2023; Gaubert et al., 2018) is summarized in Table 3.2. All the individual gene trees recovered the monophyly groupings of all four of the orders used in this study. All 12 clades were recovered for all gene trees except for

AMELX which did not recover Pholidota sister to Carnivora; *AMELX* is the third shortest gene sequence used with 642 bp.

3.3.2. Inactivating mutations

Numerous inactivating mutations have accumulated in all the genes recovered for all three pangolin species. A summary list of all these mutations is available in Table 3.1. In every enamel gene, there is at least one inactivating mutation or missing exon that is shared by all three pangolin taxa. These inactivating mutations include either a mutation within an exon(s) (e.g., premature stop codon, frameshift mutation) and/or an intron splice site mutation. Aside from the total absence of *KLK4* in both *Manis* spp. and the deletion of exons 3–5 in *P. tricuspis*, the next gene with the largest mutation is *AMTN* with exons 2–5 deleted in all three species and additional exons 6–7 deleted in *Manis*. Genes *ACP4*, *AMBN*, *AMELX*, *ENAM*, *KLK4*, and *MMP20* have numerous mutations that are shared, including exon deletions, frameshift indels, premature stop codons, and splice site mutations. Among manids, *ACP4* has eight shared mutations (six indels, one premature stop, one splice), *AMBN* three (three indels), *AMELX* one (splice), *ENAM* five (one range of exons deleted, four indels), *KLK4* one (range of exons deleted), and *MMP20* two (one premature stop, one splice), while both *Manis* share even more mutations between the two species for all these genes. In addition to the observed shared inactivating mutations listed here, all three pangolins have unique inactivating mutations in each gene that was recovered, consisting of exon deletions, frameshift indels, premature stop codons, splice site mutations, or a combination of all types (Table 3.1).

For the two dentin genes, all of the exons for *DSPP* are missing for both *Manis* species. By contrast, all exons were recovered for *Phataginus* albeit with numerous inactivating mutations including frameshift indels, premature stop codons and a splice site mutation (Table 3.1). *ODAPH*, in turn, has three inactivating mutations shared among all three manids (two deletions [including deletion of the start codon] and one splice), while both *Manis* species share one mutation (deletion). There are additionally several unique inactivating mutations in *ODAPH* for each species, with *M. pentadactyla* having three mutations (two indels, one premature stop codon) and *M. javanica* and *P. tricuspis* each having one deletion (Table 3.1). This deletion in both *M. javanica* and *P. tricuspis* is 1-bp and located at the same site in the protein-coding sequence.

3.3.3. Selection analyses

Selection analyses (dN/dS ratios) were conducted on each of the seven enamel genes and on the two dentin/tooth genes as well as on a concatenated alignment of enamel genes and a separate concatenation for the dentin/tooth genes. The concatenated alignment of enamel genes comprised 20 taxa and 9639 bp, and the dentin/tooth gene alignment had the same 20 taxa with 2196 bp. A summary of the results of these selection analyses can be found in Table 3.3A–C. The results of the codeml selection analyses for individual genes which used the branch-specific model typically had the expected values of low dN/dS ratios for outgroups in the background category, elevated dN/dS values near 1 for the pseudogenic crown Pholidota group category, and transitional stem Pholidota dN/dS values that were between these prior two categories (Table 3.3A). There were some outliers observed for *AMELX* and *ODAPH* which had stem dN/dS values that

were higher than the estimated values and fixed value of 1.0 for the pseudogenic branch categories (Table 3.3A). For *KLK4* and *DSPP*, since there were no sequences obtained, there was no difference between the transitional and pseudogenic branch categories. For these genes, the branch that was estimated or fixed at 1.0 was for *P. tricuspis* and was elevated compared to the background category with the pseudogenic branch. Results for these branches for *KLK4* had dN/dS values of 0.7485 (CF1) and 0.5120 (CF2), and *DSPP* had values of 0.9857 (CF1) and 0.9738 (CF2).

The concatenated datasets using the free-ratio model yielded dN/dS values for pangolin branches that were consistently elevated compared to outgroup taxa. For enamel genes, pangolin values ranged from 0.7650 to 0.9915 and dentin/tooth genes ranged from 0.5522 to 0.8556 (Table 3.3B). The analyses with the branch-specific model that had the concatenated alignment of enamel genes with the estimated values for the pseudogenic branch category generated dN/dS results of 0.9300 for CF1, and 0.9565 for CF2. Based on log likelihood ratio tests, these were not significantly different from expected values when the pseudogenic branch was fixed at 1 (CF1: $p = 0.39$, CF2: $p = 0.60$). For the background outgroup branches, dN/dS values were 0.3556 (CF1) and 0.3664 (CF2). These low values are indicative of purifying selection in outgroup taxa as expected. For the transitional stem pangolin branch, dN/dS values were intermediate between the functional background and pseudogenic branch categories, with values of 0.7328 (CF1) and 0.7577 (CF2). The selection analyses that utilized a fixed pseudogenic branch dN/dS value of 1.0 yielded results that were very similar to the previous analyses. The

background branch category values are 0.3557 (CF1) and 0.3665 (CF2), with the transitional branch category of 0.7287 (CF1) and 0.7551 (CF2).

For the concatenated dataset for dentin/tooth genes, the estimated pseudogenic branch category had dN/dS values of 1.0276 (CF1) and 0.9477 (CF2), background branches of 0.5609 (CF1) and 0.5313 (CF2), and transitional branches of 1.0528 (CF1) and 1.0182 (CF2). Fixed pseudogenic branch analyses had similar results, 0.5609 (CF1) and 0.5313 (CF2) for background branches, and 1.0566 (CF1) and 1.0113 (CF2) for transitional branch categories. The pseudogenic branch category value when estimated was not significantly different than the expected value of 1.0 when this category was fixed based on log likelihood ratio tests (CF1: $p = 0.94$, CF2: $p = 0.89$). Although the transitional branches are slightly elevated compared to the pseudogenic branches, these values are indicative of relaxed selection occurring early on this branch and suggest dentin/tooth genes were evolving neutrally near the most recent common ancestor of Pholidota. The low values for the background category indicate purifying selection in these genes, like the results for the enamel genes background category.

The results from the Coevol dN/dS analyses are depicted in Figure 3.1A–D. These results are similar to the codeml selection analyses, showing elevated dN/dS values on the stem and crown pangolin branches near a value of ~1 with outgroups consistently around ~0.5 and lower for enamel genes, and ~0.7 and lower for dentin/tooth genes.

3.3.4. Inactivation times

The timings of enamel and dentin/tooth loss were estimated by proxy using dN/dS values from the concatenations of enamel and dentin/tooth genes, respectively. Equations

used come from Meredith et al. (2009), and pangolin divergence times from Gaubert et al. (2018), with the stem Pholidota branch used extending from 78.9 Ma to 37.9 Ma. From these calculations, the mean of the eight different estimates for inactivation of the enamel-specific genes on the stem Pholidota branch is 61.89 Ma (ranging from 65.09–58.01 Ma) (Table 3.4). Unfortunately, the inactivation time for dentin/tooth-specific genes could not be accurately calculated due to transitional branch category dN/dS ratios coming out higher than the fully pseudogenic branch category (Table 3.3C). However, given that the transitional branch dN/dS values (~1.0113–1.0566) for the dentin/tooth genes were slightly higher than the pseudogenic branch category (0.9477–1.0276, and fixed at 1.0) suggests these genes evolved neutrally for most of the stem branch and was possibly relaxed very near the most recent common ancestor of Pholidota and in tandem with enamel genes and enamel loss.

3.4. Discussion and Conclusion

3.4.1. Inactivating mutations

Previous studies that have reported inactivating mutations in pholidotan tooth genes include Meredith et al.'s (2009) examination of exon 10 of *ENAM* in *Manis pentadactyla* and *Phataginus tricuspis*, Meredith et al.'s (2014) examination of six tooth genes (*DSPP, AMBN, AMELX, AMTN, ENAM, MMP20*) in *Manis pentadactyla*, Choo et al.'s (2016) screen of 107 tooth development-related genes in the genomes of two Asian pholidotans (*M. pentadactyla, M. javanica*) with additional PCR and sequencing for three genes in four African pangolins, Springer et al.'s (2016a) assessment of *ODAPH* in *M. pentadactyla*, Mu et al.'s (2021a) reporting of inactivating mutations in the *ACP4* gene in

three species of pangolins (*M. pentadactyla, M. javanica, P. tricuspis*), and Mu et al.'s (2021b) overview of some of the inactivating mutations in eight tooth-specific genes in enamelless/edentulous mammals including the pholidotans *M. pentadactyla, M. javanica,* and *P. tricuspis*

Our observations confirm many of the previously reported inactivating mutations in *ACP4, AMTN, AMBN, AMELX, ENAM, MMP20*, and *ODAPH*. We also extend the list of inactivating mutations for these genes (e.g., Meredith et al. [2014] only reported frameshift indels for *Manis pentadactyla*) and provide the first evidence for the recovery of portions of *DSPP* and *KLK4* in the pangolin *Phataginus tricuspis* (these genes appear to be deleted in both *Manis* species that were investigated). Finally, some of our observations contradict previous reports of inactivating mutations.

Among enamel-specific genes, Mu et al. (2021a) documented many inactivating mutations in the *ACP4* gene of the pangolins that were investigated (*M. pentadactyla*, *M. javanica*, and *P. tricuspis*) in their supplemental materials. Most of these are confirmed in our results. However, there are several mutations that were not recovered in our sequence alignment. Some examples include a 20-bp deletion that spans exons 4 and 5 in *M. javanica* (553–572 in our alignment), a 1-bp deletion at the end of exon 8 in *P. tricuspis* (868 in our alignment), and a 43-bp deletion in exon 9 for all three species that was instead recovered in our alignment to be a 24-bp deletion that is in-frame and the other 16-bp being recovered (950–989 in our alignment). In addition, there were long regions of deleted exons in the alignment in Mu et al. (2021a) that were found to be present and are summarized in Figure 3.2. Long deletions that we recovered include exons 1, 4, 11,

and part of exon 2 for *P. tricuspis*, exon 1, and 4 for *M. javanica*, and exon 3, 4, 11 and part of exon 2 for *M. pentadactyla*. Exon 3 for *M. pentadactyla* in Mu et al.'s (2021a) alignment is instead exon 1 in our alignments which also removes a 4-bp insertion they had for this species in this portion of sequence. Also, for *M. pentadactyla*, in Mu et al.'s (2021a) alignment they had a mutation in the stop codon while our alignment revealed this portion of sequence is not accurate and not due to alignment errors. For *P. tricuspis*, portions of exon 1 in Mu et al.'s (2021a) alignment aligns to exon 3. Differences between Mu et al. (2021a) and the results reported here derive from different qualities of sequences that were obtained from different sources and from different alignments. Mu et al. (2021a) provided a supplementary table with accession numbers for their sequences, but accession numbers for all three pangolin species were missing from this table. However, the authors did mention that their pangolin sequences were of poor quality and had low coverage. By contrast, our sequences were obtained from high-quality Hi-C genome assemblies downloaded from DNA Zoo. Alignment differences for the portions of DNA sequences that are similar are due to the use of different alignment programs (Mu et al. [2021a] used PRANK and we used MAFFT in Geneious) and not taxon sampling issues as they included a total of 116 mammals. Aside from these differences, our results do confirm several shared inactivating mutations in Mu et al. (2021a). These include mutations in both *Manis* species (819–820D [D = deletion], 880–881D, and 1129D in our results) and two shared mutations in all three species (655–658D [although they had 7-bp more deleted in *Manis*] and 903D in our results). Mu et al. (2021a) also reported mutations for one species that were found to be shared with the two other pangolins in

their deleted regions (93D, 322–323D, and 1217I [I = insertion] in our results). Our results also documented several other shared inactivating mutations in all three species that were not noted in Mu et al. (2021a) including a deletion, a premature stop codon, and a splice site mutation (Table 3.1). In addition to these shared mutations, there is possibly another in exon 2 where there is a shared 83-bp deletion (137–219D in our results) in *M. pentadactyla* and *P. tricuspis* along with the entire deletion of exon 2 in *M. javanica*. This may suggest the 83-bp deletion was shared for all three species prior to the entire loss of this exon in *M. javanica* especially given the species relationships; additional taxon sampling may help resolve this.

For *AMTN*, Meredith et al. (2014) reported that protein-coding exons 2–7 are deleted in *M. pentadactyla*. Our results demonstrate that this deletion is shared with *M. javanica* and that exons 2–5 are deleted in *P. tricuspis*. However, the two indels in exon 8 reported by Meredith et al. (2014) for *M. pentadactyla* (498D and 508I in their supplementary materials) were not recovered in our sequence alignments even though the nucleotides for exon 8 are identical in both studies. The differences in our alignment versus Meredith et al.'s (2014) alignment could be due to a smaller sampling of mammals (8 species) in the latter and/or the use of different alignment programs (Geneious with Muscle versus MAFFT).

For *MMP20*, several of the inactivating indels in *M. pentadactyla* noted in Meredith et al. (2014) are confirmed for this species in our alignments. However, two other mutations reported by Meredith et al. (2014) were not recovered in our alignments. Specifically, a 4-bp insertion in exon 6 (1505–1508I in Meredith et al.'s [2014]

supplementary materials) and a 1-bp deletion in exon 7 (1661D in Meredith et al.'s [2014] supplementary materials). Similar to *AMTN*, the differences in our alignment versus Meredith et al.'s (2014) alignment could be the result of different taxon sampling or different alignment algorithms. Meredith et al. (2014) showed a scaffold alignment that included exons 6 and 7 in their figure S10 but these exons were not included in the nexus file for exons. We also recovered an exon 1 sequence for *M. pentadactyla* that was not reported in Meredith et al. (2014). This exon includes an autapomorphic 1-bp deletion (47D in our results) and a second 1-bp deletion that is shared with *M. javanica* (93D in our results).

For *ENAM*, we confirm the shared 10-bp frameshift insertion that was previously reported by Meredith et al. (2009) for *M. pentadactyla* and *P. tricuspis* and Choo et al. (2016) for *M. pentadactyla* and *M. javanica*. Meredith et al. (2009) also reported a 5-bp deletion (1344–1348 [D] in Meredith et al.'s [2009] supplementary materials) and a 1-bp deletion (1362 [D] in Meredith et al.'s [2009] supplementary materials) that are shared by *M. pentadactyla* and *P. tricuspis*. These deletions are only separated by 4-bp and were not recovered in our alignment. Instead, these two deletions were recovered as an in-frame 6-bp deletion in our alignment (1972–1977D in our alignments, Figure 3.3). These differences may be the result of using different alignment programs (CLUSTAL [Meredith] versus MAFFT [present study]) with datasets that have different species composition and taxonomic breadth (49 taxa including placentals and marsupials [Meredith] versus 20 fereuungulate placentals [present study]). Even so, our results provide evidence for three additional frameshift indels in *ENAM* in the common ancestor

of Manidae (Table 3.1). These mutations were not apparent in *P. tricuspis* with the incomplete sequence coverage of Meredith et al.'s (2009) PCR fragments for this gene, although Choo et al. (2016) reported the presence of two of these mutations (7-bp and 2-bp deletions: 574–580 and 604–605, respectively, in Choo et al. [2016], and 1287–1293D and 1319-1320D in our results) in both *Manis* species. The two other mutations reported in Choo et al. (2016) for *ENAM* (8-bp deletion in *M. javanica* and 1-bp deletion in both *Manis* species: 559–566 and 710, respectively, in Choo et al. [2016], and 1274–1281D and 1425D in our results) are confirmed as well as the mutations in Meredith et al. (2009) with some of the reported mutations in *M. pentadactyla* being shared with *M. javanica* (two 1-bp deletions: 1527 [D] and 2842 [D] in Meredith et al. [2009], and 2141D and 2916D, respectively, in our results) that were not reported in Choo et al. (2016). There are also several inconsistencies in mutations listed for *P. tricuspis* in Meredith et al. (2009). Two 1-bp indels were not observed in our results (785 [I] and 826 [D] in Meredith et al.'s [2009] supplementary materials) while our sequence has a 2-bp deletion in between these two indels not in their sequence (1459–1460D in our results), these differences are a result from different sequences and not a difference in alignments (Figure 3.4). There is also a 1-bp deletion reported (917 [D] in Meredith et al.'s [2009] supplementary materials) that was instead aligned as a transition mutation in our alignments as this occurs after a 16-bp deletion in our results (1552–1567D) which without the insertion in Meredith et al. (2009) would be the same albeit they listed this mutation as too long by including an insertion in another species (898–915 [D] in Meredith et al.'s [2009] supplementary materials). In addition, we documented several

premature stop codons and splice site mutations that were not reported in these studies (Table 3.1).

The three frameshift indels in *AMBN* that were reported by Meredith et al. (2014) for *Manis pentadactyla* and Choo et al. (2016) for *M. pentadactyla*, *M. javanica*, and *P. tricuspis* are confirmed in our results. We also documented several other inactivating mutations among these taxa, with three premature stop codons and two splice mutations shared by both *Manis* species, and additional autapomorphic mutations in *M. javanica* and *P. tricuspis* (Table 3.1).

For *AMELX*, we are unaware of previous reports that list the shared splice site mutations that we recovered for *Manis* and Manidae. The shared splice site mutation for Manidae is the only shared inactivating mutation in *AMELX* that maps to the stem Pholidota branch (Intron 5 acceptor: AG → GG). Other mutations recovered include a 4-bp frameshift deletion in exon 6 in *M. pentadactyla* that was also reported in Meredith et al. (2014) (563–566D in our results and 810–813D in Meredith et al.'s [2014] supplementary materials), and a 1-bp deletion and premature stop codon in exon 6 (288D and 601–603S in our results) in *P. tricuspis* that was not reported in the text or supplementary files in Choo et al.'s (2016) investigation of *AMELX*.

We confirm the absence of *KLK4* in both *Manis* species as reported in Mu et al. (2021b). Our results also document the deletion of exons 3–5 in *P. tricuspis*. It remains unclear if the complete deletion of *KLK4* in *Manis* overprinted an earlier deletion of exons 3–5 in the common ancestor of Manidae or if the deletions in *Manis* and *P.*

tricuspis occurred on the stem *Manis* and *P. tricuspis* branches. Additional taxon sampling may resolve this issue.

For dentin/tooth-specific genes, the missing *DSPP* exons in *M. pentadactyla* were also reported by Meredith et al. (2014). Our finding that *DSPP* also appears to be deleted in *M. javanica* suggests that *DSPP* loss is a shared feature of these two species. With the inclusion of *Phataginus tricuspis* in our dataset, we observed some molecular remnants of *DSPP* in Manidae as well, albeit with numerous inactivating mutations (Table 3.1). For *ODAPH*, Springer et al. (2016a) previously reported several mutations in *M. pentadactyla* that were confirmed here, including a 34-bp deletion at the start of exon 1 and a 4-bp deletion that also occurs in exon 1. These two mutations are shared by all three manid species. Springer et al. (2016a) also reported a 5-bp deletion in *M. pentadactyla* that includes the last 3-bp of exon 1 and the canonical donor splice site (GT) in intron 1. This mutation is also present in *M. javanica* and *P. tricuspis*, although there is alignment ambiguity in this region and it is unclear if this deletion spans the exon-intron boundary (Fig. 3.5A) or is located entirely in the intron where it would include the donor splice site or (Fig. 3.5B) just the "T" of the donor splice site (Fig. 3.5C). There were also two other frameshift indels observed in exon 2 of *M. pentadactyla* that are autapomorphic for this species (Springer et al. 2016a). In our results for *ODAPH*, *M. javanica* and *P. tricuspis* were observed to have a 1-bp deletion at the same site in the protein-coding sequence of exon 2 that is not seen in *M. pentadactyla*. However, this mutation occurs in a poly T region where unequal recombination may have overwritten the deletion in *M. pentadactyla* if the deletion originated in the common ancestor of Manidae. Indeed, the

hotspot nature of the poly T region for inactivating mutations in Manidae is evidenced by a 2-bp insertion in *M. pentadactyla* that is several bp upstream of the 1-bp deletion in the other two manids. The observed accumulation of inactivating mutations in all the genes investigated here as well as the shared mutations among enamel and dentin/tooth genes confirm that these genes and by proxy the inferred phenotypes were lost in the common ancestor of extant pangolins, as opposed to convergent loss of enamel and dentin/teeth. Indeed, the occurrence of inactivating mutations in enamel and dentin/tooth genes that are shared by all pangolins is expected given that the edentulous pangolin fossil *Eomanis* is older (~47 MY) (Franzen, 2005) than are relaxed molecular clock estimates for the most recent common ancestor of crown Pholidota (~25.3 MY [Meredith et al., 2011]; 37.9 MY [Gaubert et al., 2018]).

3.4.2. Selection analyses

Results from the selection analyses indicate an early loss of selection pressures on enamel and dentin/tooth genes in the common ancestor of pangolins as dN/dS values were all elevated above those in the outgroup taxa with enamel-capped teeth. This suggests that most of the stem branch leading to extant species was evolving under neutral evolution after the release from selection pressures on these genes and the necessity of possessing teeth in their feeding method. Comparing our results with previous studies, Meredith et al. (2009) examined *ENAM* and obtained a dN/dS value of 0.82 for their stem pangolin branch albeit for *M. pentadactyla* and "*Manis*" (*Phataginus*) *tricuspis* (although their believed shared mutations were confirmed as being shared in our Manidae taxa as mentioned earlier). Our results for *ENAM* with the inclusion of *M.*

javanica corroborates this finding, obtaining stem transitional branch category values ranging from 0.8196–0.8621. Although this was the only gene investigated in Meredith et al. (2009) this is the longest enamel gene, having more than double the number of base-pairs as the next longest gene. These results are similar to the concatenated dataset of enamel genes, which had transitional branch categories ranging from 0.7219 to 0.7506, and the pseudogenic branch category having values of 0.9353 (CF1) and 0.9631 (CF2) when this category was estimated and not fixed at 1.0.

For the dentin/tooth gene *ODAPH*, Springer et al. (2016a) reported a high dN/dS value of 2.0012 (CF1) and 1.9501 (CF2) for their pangolin branch, which only included *M. pentadactyla*, suggesting an early inactivation of this gene that was under positive selection after divergence from Carnivora, however the authors also cited possible alignment errors or model misspecification in PAML. With the inclusion of *M. javanica* and *P. tricuspis* for this gene, our results yielded dN/dS values ranging from 1.2951 to 1.3073 for the transitional stem branch category and from 1.1605 to 1.1790 for the pseudogenic branch category when it was estimated instead of fixed at 1.0. These dN/dS results on our stem branch are closer to the expected value of near 1 than the dN/dS values of ~1.95–2.00 obtained in Springer et al. (2016a), although the elevated values compared to our crown pangolin dN/dS values may still suggest positive selection on this and an early inactivation of this gene on the stem Pholidota branch. The dN/dS values for *ODAPH* were similar to *DSPP* dN/dS values obtained for *P. tricuspis* of 0.9857 (CF1) and 0.9738 (CF2). For the concatenated dataset for both *DSPP* and *ODAPH* genes, the stem and pseudogenic branches were similarly elevated and near the expected values of

1. The transitional category for these genes ranged from 1.0113 to 1.0566, and pseudogenic category was 1.0276 (CF1) and 0.9477 (CF2) when estimated.

3.4.3 Inactivation times and conclusion

Furthermore, our results support that at least enamel was lost early on the stem Pholidota branch, with our results ranging from 63.08 to 53.46 Ma, which is near the calculated loss of the tooth genes *ENAM* and *ODAM* from previous studies, which were reported as 59.4–54.9 Ma and ~65 Ma respectively (Meredith et al., 2009; Springer et al. 2019). The projected timing for the loss of enamel reported in our study is around ~16 to 6 Ma before the first known edentulous pangolin (47 Ma), and around ~17 to 12.5 Ma after the split from Carnivora (79.8–66 Ma). This amount of time on either side of this projected loss suggests that there may be additional fossil pangolin species to be discovered that will fill in the gap of enamel and tooth loss in ancient taxa. Given that many early palaeanodont species already showed signs of thinning enamel and the calculation of enamel loss is very early on the stem Pholidota branch, it is probable that the earliest forms of diverging pangolins already had thin enamel due to mutations in several enamel production-related genes that they most likely inherited from their palaeanodont relatives. Although the timing of dentin/tooth genes and phenotype loss could not be inferred from inactivation calculations, the elevated dN/dS values recovered from the codeml and Coevol analyses suggest loss near the base of the Pholidota branch, and perhaps in the most recent common ancestor along with the loss of enamel disproving the hypothesis of the possibility that enamel and dentin/teeth were decoupled and lost at separate times. Even further, obtaining and examining the enamel structure of

fossil pangolin teeth is perhaps unlikely with the reported difficulty in locating some teeth and peculiar jaw morphology of many Paleocene and Eocene palaeanodonts. Such species have been reported to contain alveoli that are significantly larger than their teeth, poorly held single-rooted teeth, and some alveoli that are empty or infilled with bone indicating that these teeth were lost prematurely during the life of the animal which could similarly resemble the early evolution of pholidotans (Gheerbrant et al., 2005; Secord et al., 2002).

With the addition of the results of this study adding to our understanding of extinct pangolin biology, we now have a more robust estimate of the timing of enamel and tooth loss in early stem pangolins which may aid in elucidating the evolutionary adaptations that occurred that led to their specialized diet of and unique ecological niche. Specifically, this understanding may clarify when there was a loss of selection pressure on tooth genes in their common ancestor compared to when they developed their more derived specializations for myrmecophagy that coincides with being edentulous.

Further research into the pseudogenized tooth genes in pangolins and other toothless mammals can help us better understand the evolutionary processes that have produced their unique morphologies and feeding habits. Additionally, this research can contribute to our understanding of tooth development and tooth loss in mammals more broadly, shedding light on the genetic and environmental factors that influence the evolution of these traits.

Tables

Table 3.1. Inactivating mutations in enamel- and dentin/tooth-specific genes in Pholidota.

Taxa with mutation	**Enamel gene with mutation**	**Dentin gene with mutation**
Manidae	*ACP4*: E1: 70D, 93D, E4: 322-323D, E7: 655-658D, E8: 826-829S*, In8Do: --, E9: 903D, E11: 1217I; *AMBN*: E11: 869-882D, 1039D, 1181-1184I; *AMELX*: In5Ac: GG; *AMTN*: E2-5: NRM/NBR; *ENAM*: E1-5: NRM/NBR, E8: 1074I, 1287-1293D, 1319-1320D, 2768-2777I; *KLK4*: E3-5: NRM/NBR; *MMP20*: E7: 1031-1033S, In7Do: CT	*ODAPH*: E1: 1-34D, 46-49D, In1Do: -- or G- or GG
Manis	*ACP4*: E1: 25-56D, E4: 379-398D, E8: 819-820D, E9: 880-881D, In9Do: G-, E10: 1129D; *AMBN*: E3: 125-127S, In6Do: AT, In7Do: AT, E11: 1215-1217S; *AMELX*: In5Do: GA; *AMTN*: E6-7: NRM/NBR; *ENAM*: E6: 571D, E7: 612D, E8: 1425D, 1560-1562S, 1941-1943S, 2141D, 2916D, 3104-3105I; *KLK4*: E1-2: NRM/NBR; *MMP20*: E1: 93D, E2: NRM/NBR, E4: 559-561S, E6: 924I, E7: 1070-1080D, 1089-1090I, 1096-1098S, E8: 1170D, In8Ac:AT	*DSPP*: NRM/NBR; *ODAPH*: E2: 70D
Manis javanica	*ACP4*: E2: NRM/NBR, In5Ac: GA, E6: 565-567S, E7: 776-777D, In7Ac: --, E8: 785-788D; *AMBN*: E2: 80I, E4: 158-160S, E11: 736D, 1269-1271S; *ENAM*: E6: 555D, E8: 938I, 1274-1281D, 1443-1445S, 1768D, 2332I, 3132-3134S, 3283D; *MMP20*: E3: 385-387S, E4: 634-636S, In8Do: GA	*ODAPH*: E2: 315D

Manis pentadactyla	*ACP4*: E2: 137-219D, E5: 454-493D, In5Ac: AC, In6Ac: A-, E7: 781-783S, E8: 803D, 827I, E9: 932-934S; *AMELX*: E6: 563-566D, In6Do: CT; *ENAM*: In6Ac: GG, E8: 712-714S, 1232D, 2485-2487S, 3434D; *MMP20*: E1: 47D, E4: 642D, E5: 726-730D, In8Do: AA, E9: 1291I, 1361-1363S, In9Ac: -G (allelic variation)	*ODAPH*: E2: 184-186S, 228-229D, 313-314I
Phataginus tricuspis	*ACP4*: E1: 10-26D, 46D, E2: 137-219D, In7Ac: AA, E8: 785-822D, E9: 924D, In10Ac: GG; *AMBN*: E1: NRM/NBR, E3: 104D, In3Do: TT, E7: 572-574S (allelic variation), In9Ac: GG (allelic variation); *AMELX*: In2Ac: AC, E6: 288D, 601-603S; *AMTN*: E7: 337-346D, E8: 430-439D (allelic variation); *ENAM*: E6: 553-555S, In6Do: AT, E8: 633-660D, 787-789S, 904-906S, 1122-1124S, 1459-1460D (allelic variation), 1552-1567D, 1660-1663D, 2247-2249S, 2734-2736S, 2810-2813I, 3043D, 3052D, 3087-3090I, 3386I; *KLK4*: E2: 112-114S; *MMP20*: E1, 3: NRM/NBR, E5-6: NRM/NBR, E8: 1174-1176S, 1199D, 1234-1235D, In8Do: GA, In8Ac: --, E9: 1260-1325D, E10:1442-1444S	*DSPP*: In1Do: AT, E2: 98-99D, In2Do: CT, E3: 190-192S, 530D, 544-546S, 866D, 964-967D, 1071D, E4: 1394I, 1409-1411S; *ODAPH*: E2: 315D

Numbers correspond to positions in the protein-coding sequence alignments (CDS). Abbreviations: Ac = acceptor splice site; D = deletion; Do = donor splice site; E = exon; I = insertion; In = intron; NRM = no reads mapped and possible deletion of exon(s) or gene; NBR = no blast results and possible deletion of exon(s) or gene; S = premature stop codon; SCM = start codon mutation.

*, with subsequent 1-bp insertion in *Manis pentadactyla*.

Table 3.2. Monophyly of well-supported laurasiatherian clades on maximum likelihood phylograms for seven enamel and two dentin/tooth related genes.

Clade	Gene								
	ACP4	***AMBN***	***AMELX***	***AMTN***	***DSPP***	***ENAM***	***KLK4***	***MMP20***	***ODAPH***
Pholidota	yes	yes	yes	yes	**n/a**	yes	**n/a**	yes	yes
Manis	yes	yes	yes	yes	**n/a**	yes	**n/a**	yes	yes
Carnivora	yes	yes	yes	yes	yes	yes	yes	yes	yes
Feliformia	yes	yes	yes	yes	yes	yes	yes	yes	yes
Caniformia	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ferae	yes	yes	no	yes	yes	yes	yes	yes	yes
Perissodactyla	yes	yes	yes	yes	yes	yes	yes	yes	yes
Tapiridae	yes	yes	yes	yes	yes	yes	yes	yes	yes
Rhinocerotidae	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ferae + Perissodactyla	yes	yes	yes	yes	yes	yes	yes	yes	yes
Cetartiodactyla	yes	yes	yes	yes	yes	yes	yes	yes	yes
Bovidae	yes	yes	yes	yes	yes	yes	yes	yes	yes
Total number of clades recovered	12/12	12/12	11/12	12/12	10/10	12/12	10/10	12/12	12/12
Percentage of clades recovered	100%	100%	91.67%	100%	100%	100%	100%	100%	100%

Table 3.3A–C. Selection analyses for seven enamel genes and two dentin/tooth genes with branch-specific models, and concatenations of enamel and dentin/tooth genes for free-ratio and branch-specific models.

Branch Category	CF1								
	Fully pseudogenic estimated								
	ACP4	***AMBN***	***AMELX***	***AMTN***	***ENAM***	***KLK4***	***MMP20***	***DSPP***	***ODAPH***
Crown Pholidota	0.8691	1.0411	0.7575	0.9449	0.8915	0.7485	0.9092	0.9857	1.1605
Stem Pholidota	0.4128	0.7873	0.8298	0.7654	0.8621		0.4713		1.2951
Outgroups	0.1532	0.4029	0.5463	0.4902	0.5397	0.3137	0.1208	0.5558	0.6075
	Full pseudogenic fixed (=1.0)								
	ACP4	***AMBN***	***AMELX***	***AMTN***	***ENAM***	***KLK4***	***MMP20***	***DSPP***	***ODAPH***
Crown Pholidota	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Stem Pholidota	0.4098	0.7877	0.8087	0.7641	0.8601		0.4438		1.2952
Outgroups	0.1532	0.4029	0.5469	0.4903	0.5400	0.3137	0.1208	0.5559	0.6075
Branch Category	**CF2**								
	Fully pseudogenic estimated								
	ACP4	***AMBN***	***AMELX***	***AMTN***	***ENAM***	***KLK4***	***MMP20***	***DSPP***	***ODAPH***
Crown Pholidota	0.7593	1.1156	0.8215	1.0939	0.8213	0.5120	0.9211	0.9738	1.1790
Stem Pholidota	0.3546	0.8419	0.8758	0.8769	0.8225		0.4837		1.3038
Outgroups	0.1233	0.4325	0.6051	0.5317	0.5081	0.2203	0.1166	0.5251	0.6170
	Full pseudogenic fixed (=1.0)								
	ACP4	***AMBN***	***AMELX***	***AMTN***	***ENAM***	***KLK4***	***MMP20***	***DSPP***	***ODAPH***
Crown Pholidota	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Stem Pholidota	0.3501	0.8430	0.8597	0.8792	0.8196		0.4591		1.3073
Outgroups	0.1233	0.4323	0.6059	0.5317	0.5085	0.2202	0.1166	0.5251	0.6169

A) Branch-specific analyses for individual genes for estimated and fixed values for the pseudogenic branch category. Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2.

	7 Enamel		2 Dentin	
Taxa	**CF1**	**CF2**	**CF1**	**CF2**
Acinonyx jubatus	0.2166	0.2243	0.4736	0.4394
Bos mutus	0.4018	0.4133	0.4388	0.4198
Camelus bactrianus	0.3706	0.3831	0.7447	0.6902
Canis lupus	0.2711	0.2784	0.5371	0.5087
Capra hircus	0.3669	0.3767	0.3875	0.3770
Ceratotherium simum	0.5127	0.5201	0.2363	0.2271
Cryptoprocta ferox	0.2072	0.2131	0.5753	0.5291
Diceros bicornis	0.5331	0.5476	1.0579	1.0188
Equus caballus	0.4107	0.4237	0.6247	0.6022
Lutra lutra	0.2244	0.2311	0.5590	0.5071
Manis javanica	**0.9735**	**0.9987**	**0.8064**	**0.7343**
Manis pentadactyla	**0.8801**	**0.9050**	**0.8556**	**0.7661**
Odobenus rosmarus	0.2918	0.3019	0.7194	0.6713
Phataginus tricuspis	**0.7513**	**0.7840**	**0.5741**	**0.5522**
Rhinoceros unicornis	0.4137	0.4262	999.0000	588.9900
Sus scrofa	0.4638	0.4764	0.5800	0.5755
Tapirella bairdii	0.3639	0.3822	0.1426	0.1484
Tapirus indicus	0.2454	0.2539	0.6573	0.6436
Tapirus terrestris	0.3240	0.3329	0.0001	0.0001
Ursus arctos	0.2813	0.2899	0.3951	0.4046
Stem Pholidota	**0.7123**	**0.7375**	**1.1150**	**1.0648**

B) Free-ratio analyses for concatenated genes, CF1 and CF2. Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2.

	CF1			
	Fully pseudogenic estimated	Full pseudogenic fixed (=1.0)	Fully pseudogenic estimated	Full pseudogenic fixed (=1.0)
Branch Category	**7 Enamel**		**2 Dentin**	
Crown Pholidota	0.9353	1.0000	1.0276	1.0000
Stem Pholidota	0.7256	0.7219	1.0528	1.0566
Outgroups	0.3554	0.3555	0.5609	0.5609
lnL	-49943.5219	-49943.8497	-12682.2871	-12682.2896
p-value	**0.42**		**0.94**	
	CF2			
	Fully pseudogenic estimated	Full pseudogenic fixed (=1.0)	Fully pseudogenic estimated	Full pseudogenic fixed (=1.0)
Branch Category	**7 Enamel**		**2 Dentin**	
Crown Pholidota	0.9631	1.0000	0.9477	1.0000
Stem Pholidota	0.7506	0.7484	1.0182	1.0113
Outgroups	0.3662	0.3662	0.5313	0.5313
lnL	-49952.9091	-49953.0132	-12749.4923	-12749.5020
p-value	**0.65**		**0.89**	

C) Branch-specific analyses for concatenated genes, CF1 and CF2. Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2.

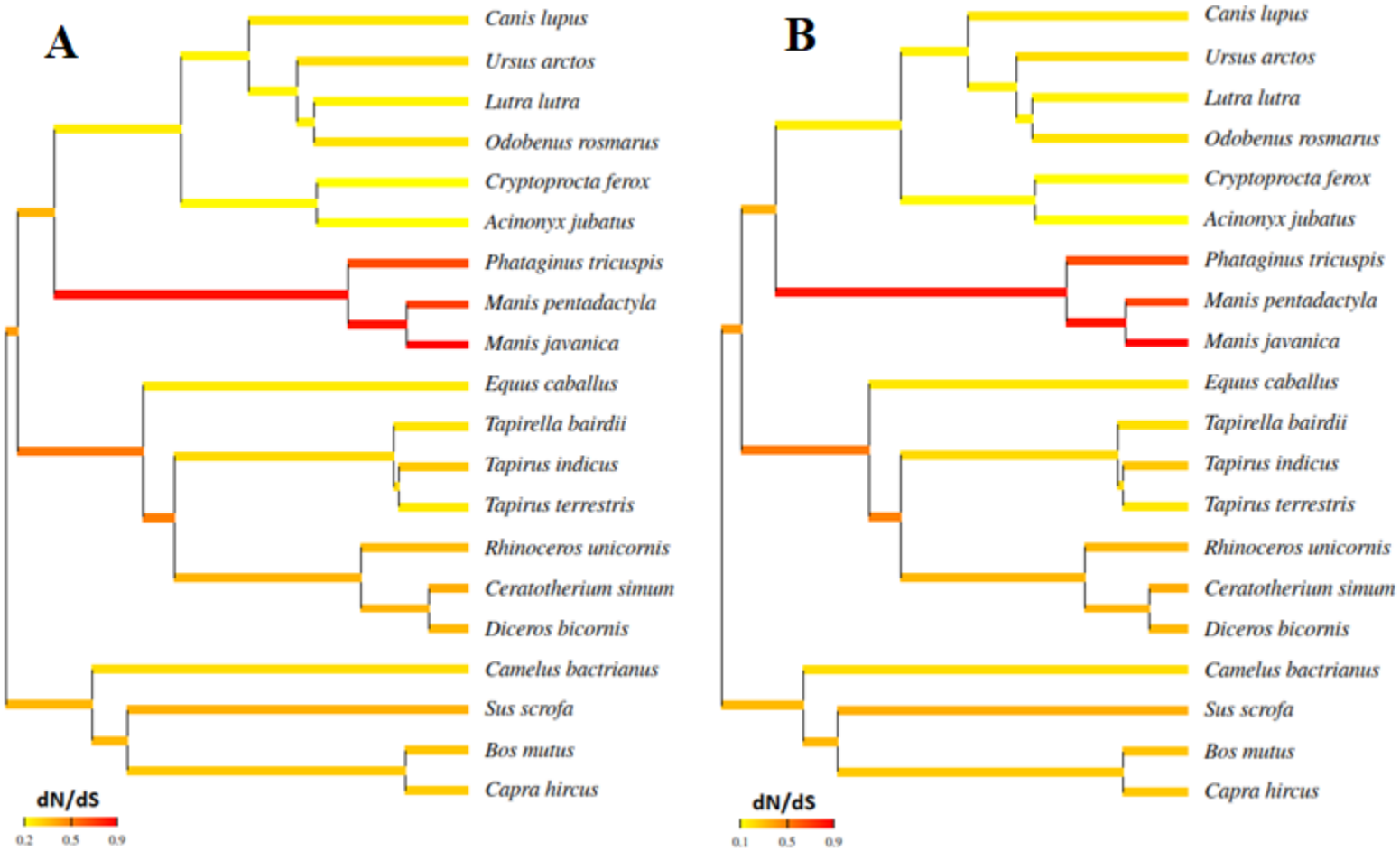
Table 3.4. Inactivation times (Ma) for seven enamel genes on the stem Pholidota branch.

Gene Concatenation: Enamel Genes	CF1				CF2				Mean Inactivation Time
	dN/dS estimated		dN/dS = 1		dN/dS estimated		dN/dS = 1		
	1 syn rate	2 syn rates	1 syn rate	2 syn rates	1 syn rate	2 syn rates	1 syn rate	2 syn rates	
Gaubert et al., 2018	64.07	60.56	61.21	57.57	64.3	60.81	62.62	59.03	61.27
Meredith et al., 2009	71.03	66.58	57.4	52.8	71.32	66.9	59.2	54.65	62.48
Meredith et al., 2011	59.96	55.3	56.16	51.34	60.27	55.63	58.04	53.27	56.25
Foley et al., 2016	56.62	52.65	53.38	49.28	56.88	62.93	56.62	52.65	53.46
Heighton et al., 2023	65.68	62.41	63.01	59.63	65.9	62.64	64.33	60.99	63.08

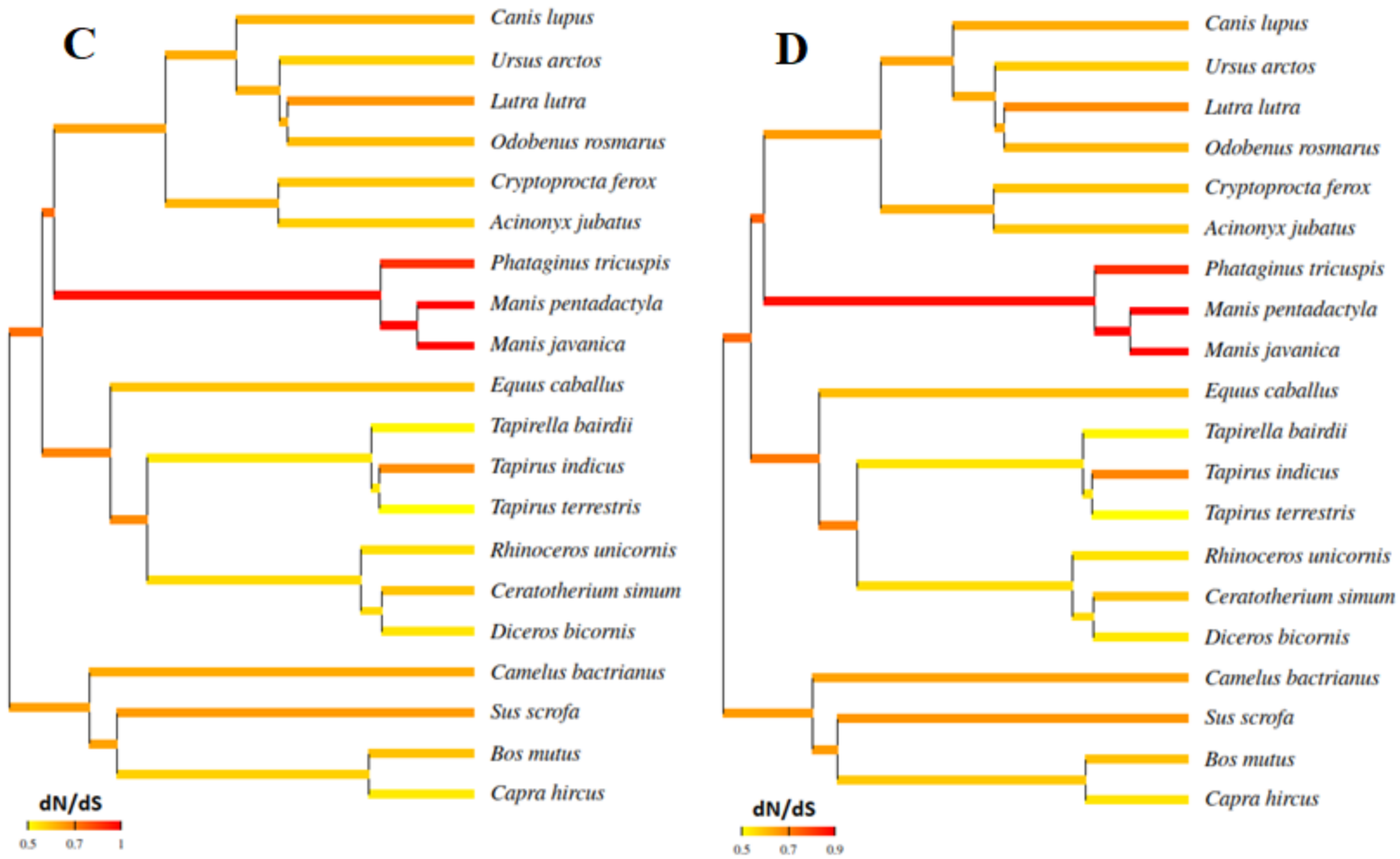
Abbreviations: CF1, codon frequency model 1; CF2, codon frequency model 2; dN = nonsynonymous substitutions per nonsynonymous site; dS = synonymous substitutions per synonymous site; Ma = millions of years ago, syn = synonymous.

Figures

Figure 3.1A–D. Coevol selection analyses (dN/dS) for the concatenation of seven enamel and two dentin/tooth genes.



A (left), B (right): Concatenation of seven enamel genes, two Coevol chains.



C (left), D (right): Concatenation of two dentin/tooth genes, two Coevol chains.

Figure 3.2. Differences in *ACP4* sequences. Shows additional sequences in our sequences compared to Mu et al. (2021a).

Figure 3.3. Differences in *ENAM* exon 8 alignments. Meredith et al. (2009) has two frameshift deletions compared to the in-frame deletion in our sequences.

Figure 3.4. Differences in *ENAM* exon 8 sequences of *Phataginus tricuspis*. Meredith et al. (2009) has two 1-bp frameshift deletions compared to the one 2-bp deletion in our sequence.

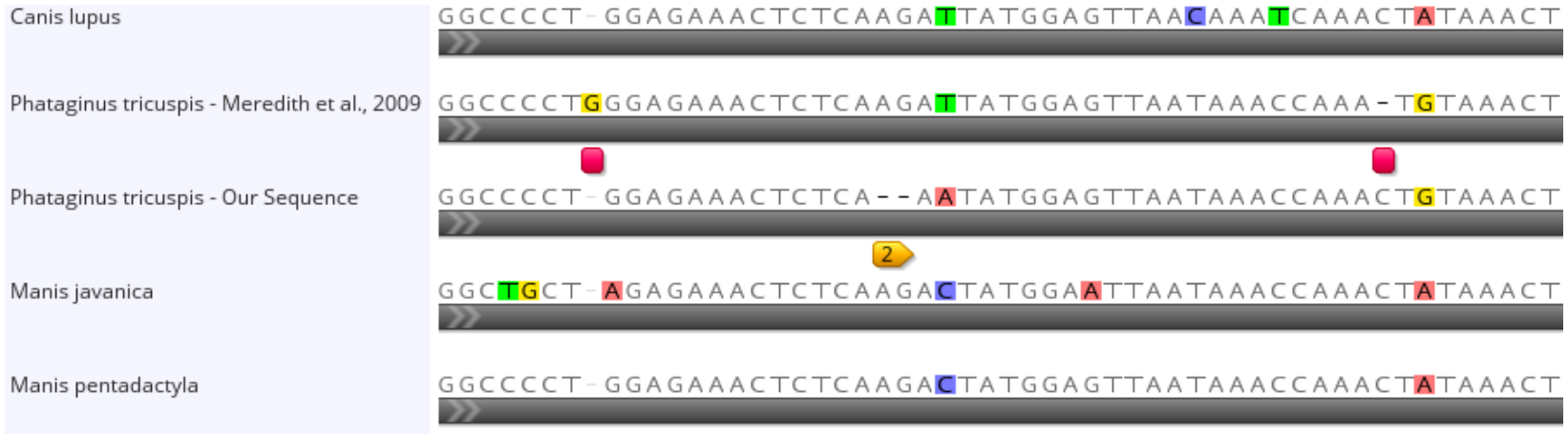
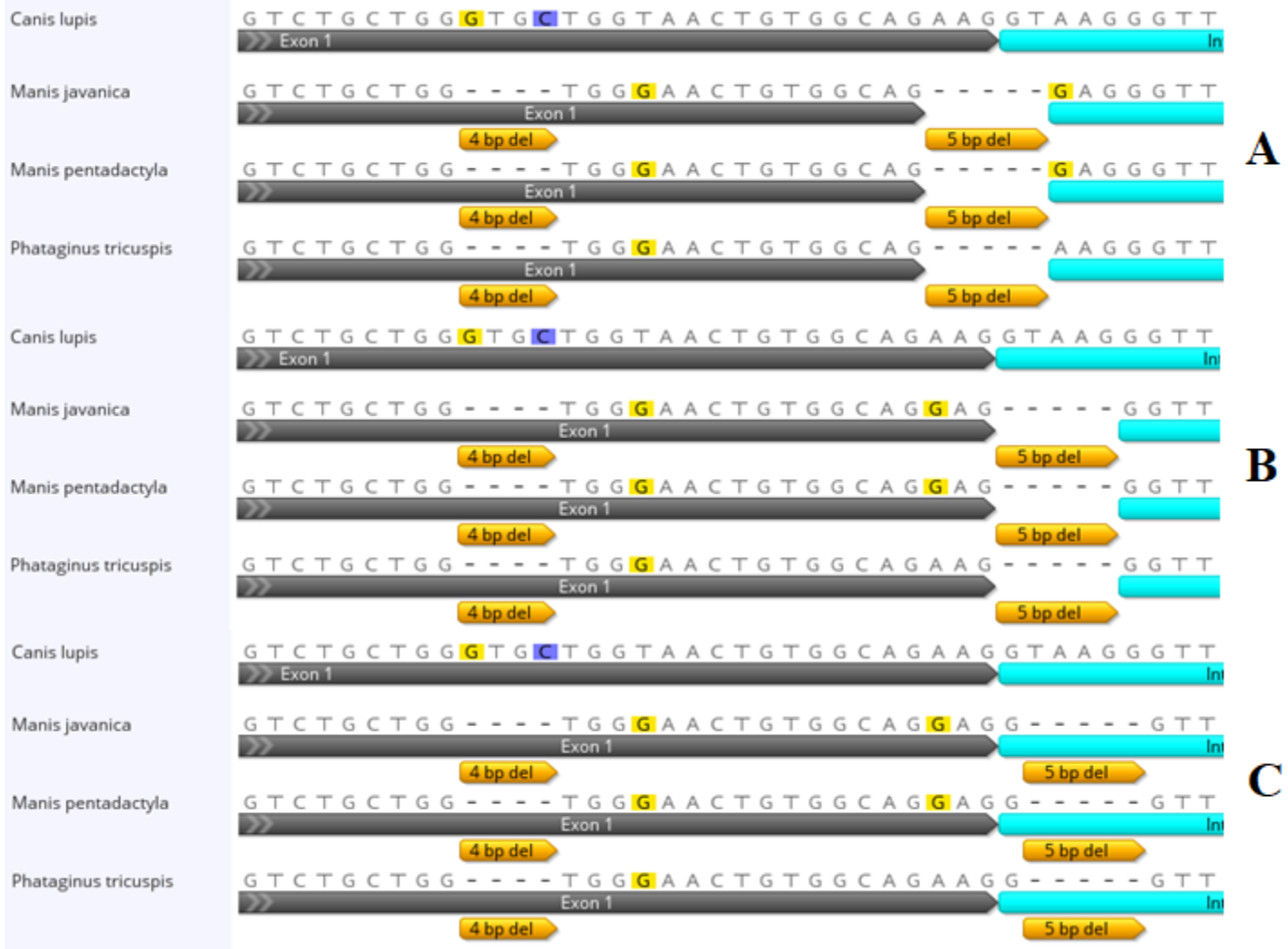


Figure 3.5A–C. Possible different alignments for *ODAPH* which changes interpretation of the mutation at the exon-intron boundary for exon 1 and intron 1.



Epilogue

Macroevolutionary changes are profound occurrences in the history of the tree of life. These events have let us understand more about the evolution of organisms and the transitions they have made in acquiring new adaptations that allowed for the exploitation of previously underutilized ecological niches. Genetic research is now able to help answer these questions with the advent of DNA sequencing technology including the recent progress made in obtaining full genomes from more species in each coming year. Throughout these chapters I have contributed to the understanding of the underlying causes of tooth and enamel loss by examining the molecular precursors to those phenotypes via the gene sequences involved in their developmental pathways. I demonstrated additional sources and causes for enamel and dentin/teeth degeneration in three taxonomic groups, both clades of cetaceans, Mysticeti and Odontoceti, and pholidotans.

In the first chapter, the large, long living baleen whales were investigated for underlying mutations in their enamel and dentin/tooth genes to elucidate the timing of the loss of those phenotypes to answer the long outstanding question of when and how tooth loss occurred in these species. Known as a "poster child" for macroevolution, cetaceans are one of the best-known cases of a large scale ecological evolutionary event, transforming from terrestrial quadrupeds to limbless obligate aquatic animals. This transition involved wholesale changes in the anatomy and behavior of these animals which were discovered through an extensive fossil record and the possession of vestigial and rudimentary traits in living species. Although the transition from land to water and

from teeth to baleen has been known from at least Darwin's time, the underlying genetic components that have altered these structures is a recent and emerging field of study. Building upon work in this discipline, this chapter has added to our understanding of the change from toothed raptorial hunters to gigantic baleen filter-feeders. This study has also help elucidate the debate of current hypotheses of if ancient mysticetes possessed teeth at the same time as baleen, and if so, what state those teeth were in. Examining several genes responsible for enamel and dentin/tooth production, we reported numerous inactivating mutations, with some being shared among major clades of mysticetes. Calculating the timing of the loss of enamel and dentin/tooth genes, it was determined that these phenotypes were decoupled, and enamel was lost before dentin/teeth which was lost at most two times, leaving an enamelless condition of teeth with baleen plates in the jaws very likely before total tooth loss in an extinct common ancestor.

Odontocetes are the other clade of cetaceans and are extremely similar to their sister taxa but still possess teeth. Although they are named the toothed whales, they display a wide array of enamel phenotypes including degenerative enamel which was examined in chapter two. DNA coding sequences for enamel-specific genes were surveyed for mutations and utilized in selection analyses to determine rates of molecular change occurring in these genes on branches leading to different clades of species with similar enamel microstructure phenotypes. Groupings of taxa were determined by similar enamel complexity as reported in previous studies, assigned with categories utilized in Werth et al. (2020). Overall, taxa with less complex enamel phenotypes possess enamel genes with more inactivating mutations and higher dN/dS values implying relaxed

selection compared to less mutations and lower dN/dS values indicative of stronger purifying selection in odontocetes possessing more complex enamel microstructure as well as terrestrial non-cetacean Cetartiodactyla outgroups. Also, linear regression analyses results indicated strong correlations between both mutation counts and selection pressure with enamel complexity. All these results taken together provide evidence in confirming the stated hypotheses of this study.

Observing similar questions as the first chapter, chapter three investigated the timing and method of tooth loss in the enigmatic pangolins. Enamel and dentin/tooth genes were gathered and examined for inactivating mutations as well as used in selection analyses to determine the rate of nonsynonymous versus synonymous substitutions occurring on these genes. These ratios were then used to perform calculations for the inactivation of those genes and by proxy their encoded phenotypes. Although we were not able to determine a date of inactivation for dentin/teeth, the loss of enamel was determined to be similar to previous investigations on the timing of inactivation of certain enamel-specific genes. The timing in the loss of enamel was compared to the earliest known edentulous fossil pangolin and latest known relative that still possessed teeth to provide additional information on where, how, and when dental regression occurred in the early origin of Pholidota. Although not completely elucidated, this study has provided more evidence into this macroevolutionary occurrence and their biogeographical origins of enamel loss relative to their stem relatives Palaeanodonta and their sister group Carnivora.

Hopefully the research in these three previous chapters will increase our knowledge and understanding of macroevolutionary changes in disparate taxa and help lead to possible conservation efforts of species in an ever-changing world. It is my hope that we can continue research into all animals and living creatures in our surrounding environments and help teach non-biologists the importance of these species and the necessity to protect our biodiversity and open people's minds to the interconnectedness of all living things.

References

Abbarin, N., San Miguel, S., Holcroft, J., Iwasaki, K., Ganss, B., 2015. The enamel protein amelotin is a promoter of hydroxyapatite mineralization. J. Bone Miner. Res. 30, 775–785. https://doi.org/10.1002/jbmr.2411

Aisher, A., 2016. Scarcity, Alterity and Value: Decline of the Pangolin, the World's Most Trafficked Mammal. Conserv. Soc. 14, 317–329. https://doi.org/10.4103/0972-4923.197610

Alba, D.M., Hammond, A.S., Vinuesa, V., Casanovas-Vilar, I., 2018. First record of a Miocene Pangolin (Pholidota, Manoidea) from the Iberian Peninsula. J. Vertebr. Paleontol. 38. https://doi.org/10.1080/02724634.2017.1424716

Alloing-Séguier, L., Lihoreau, F., Boisserie, J.R., Charruault, A.L., Orliac, M., Tabuce, R., 2014. Enamel microstructure evolution in anthracotheres (Mammalia, Cetartiodactyla) and new insights on hippopotamoid phylogeny. Zool. J. Linn. Soc. 171, 668–695. https://doi.org/10.1111/zoj.12143

Armfield, B.A., Zheng, Z., Bajpai, S., Vinyard, C.J., Thewissen, J., 2013. Development and evolution of the unique cetacean dentition. PeerJ 1, e24. https://doi.org/10.7717/peerj.24

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., Janke, A., 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci. Adv. 4. https://doi.org/10.1126/sciadv.aap9873

Bajpai, S., Thewissen, J.G.M., Sahni, A., 2009. The origin and early evolution of whales: macroevolution documented on the Indian Subcontinent. J. Biosci. 34, 673–686.

Bartlett, J.D., Skobe, Z., Nanci, A., Smith, C.E., 2011. Matrix metalloproteinase 20 promotes a smooth enamel surface, a strong dentino–enamel junction, and a decussating enamel rod pattern. Eur. J. Oral Sci. 119, 199–205.

Bejder, L., Hall, B.K., 2002. Limbs in whales and limblessness in other vertebrates: mechanisms of evolutionary and developmental transformation and loss. Evol. Dev. 4, 445–458.

Berkovitz, B., Shellis, P., 2018. *The Teeth of Mammalian Vertebrates*. https://doi.org/10.1016/B978-0-12-802818-6.01001-2

Berta, A., Lanzetti, A., Ekdale, E.G., Deméré, T.A., 2016. From Teeth to Baleen and Raptorial to Bulk Filter Feeding in Mysticete Cetaceans: The Role of

Paleontological, Genetic, and Geochemical Data in Feeding Evolution and Ecology. Integr. Comp. Biol. 56, 1271–1284. https://doi.org/10.1093/icb/icw128

Bérubé, M., Aguilar, A., 1998. A new hybrid between a blue whale, *Balaenoptera musculus*, and a fin whale, *B. physalus*: frequency and implications of hybridization. Mar. Mamm. Sci. 14, 82–98.

Bianucci, G., Landini, W., 2006. Killer sperm whale: A new basal physeteroid (Mammalia, Cetacea) from the Late Miocene of Italy. Zool. J. Linn. Soc. 148, 103–131. https://doi.org/10.1111/j.1096-3642.2006.00228.x

Bloodworth, B.E., Odell, D.K. 2008. *Kogia breviceps* (Cetacea: Kogiidae). Mamma. Species 819, 1–12.

Boessenecker, R.W., Fordyce, R.E., 2015. Anatomy, feeding ecology, and ontogeny of a transitional baleen whale: a new genus and species of Eomysticetidae (Mammalia: Cetacea) from the Oligocene of New Zealand. PeerJ 3, e1129. https://doi.org/10.7717/peerj.1129

Boessenecker, R.W., Fraser, D., Churchill, M., Geisler, J.H., 2017. A toothless dwarf dolphin (Odontoceti: Xenorophidae) points to explosive feeding diversification of modern whales (Neoceti). Proc. Roy. Soc. B 284, 20170531.

Boschma, H. 1951. Rows of small teeth in ziphioid whales. Zoologische Mededelingen 31, 139–148.

Brownell, R.L., Herald, E.S., 1972. *Lipotes vexillifer*. Mammalian Species 10, 1–4.

Bustamante, C.D., Nielsen, R., Hartl, D.L., 2002. A maximum likelihood method for analyzing pseudogene evolution: Implications for silent site evolution in humans and rodents. Mol. Biol. Evol. 19, 110–117. https://doi.org/10.1093/oxfordjournals.molbev.a003975

Caterina, J.J., Skobe, Z., Shi, J., Ding, Y., Simmer, J.P., Birkedal-Hansen, H. and Bartlett, J.D., 2002. Enamelysin (matrix metalloproteinase 20)-deficient mice display an amelogenesis imperfecta phenotype. J. Biol. Chem. 277, 49598–49604.

Chehida, B., Y., Thumloup, J., Schumacher, C., Harkins, T., Aguilar, A., Borrell, A., Ferreira, M., Rojas-Bracho, L., Robertson, K.M., Taylor, B.L., Víkingsson, G.A., Weyna, A., Romiguier, J., Morin, P.A., Fontaine, M.C., 2020. Mitochondrial genomics reveals the evolutionary history of the porpoises (Phocoenidae) across the speciation continuum. Sci. Rep. 10, 1–18. https://doi.org/10.1038/s41598-020-71603-9

Choo, S.W., Rayko, M., Tan, T.K., Hari, R., Komissarov, A., Wee, W.Y., Yurchenko, A.A., Kliver, S., Tamazian, G., Antunes, A., Wilson, R.K., Warren, W.C., Koepfli, K.P., Minx, P., Krasheninnikova, K., Kotze, A., Dalton, D.L., Vermaak, E., Paterson, I.C., Dobrynin, P., Sitam, F.T., Rovie-Ryan, J.J., Johnson, W.E., Yusoff, A.M., Luo, S.J., Karuppannan, K.V., Fang, G., Zheng, D., Gerstein, M.B., Lipovich, L., O'Brien, S.J., Wong, G.J., 2016. Pangolin genomes and the evolution of mammalian scales and immunity. Genome Res. 26, 1312–1322. https://doi.org/10.1101/gr.203521.1153

Cooper, L.N., Maas, M.C., 2009. Bones and teeth, histology of. Encycl. Mar. Mamm. 124–129. https://doi.org/10.1016/B978-0-12-373553-9.00034-1

Crespo-Picazo, J.L., Rubio-Guerri, C., Jiménez, M.A., Aznar, F.J., Marco-Cabedo, V., Melero, M., Sánchez-Vizcaíno, J.M., Gozalbes, P., García-Párraga, D., 2021. Bottlenose dolphins (Tursiops truncatus) aggressive behavior towards other cetacean species in the western Mediterranean. Sci. Rep. 11, 1–9. https://doi.org/10.1038/s41598-021-00867-6

Dalebout, M.L., Steel, D., Baker, C.S., 2008. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: Implications for the evolution of male tusks. Syst. Biol. 57, 857–875. https://doi.org/10.1080/10635150802559257

Darwin, C., 1872. *Origin of Species 6th Ed*.

Davit-Béal, T., Tucker, A.S., Sire, J.-Y., 2009. Journal of Anatomy - 2009 - Davit-B al - Loss of teeth and enamel in tetrapods fossil record genetic data and.pdf. J. Anat. 214, 477–501. https://doi.org/10.1111/j.1469-7580.2009.01060.x

Delgado, S., Girondot, M., Sire, J.Y., 2005. Molecular evolution of amelogenin in mammals. J. Mol. Evol. 60, 12–30. https://doi.org/10.1007/s00239-003-0070-8

Deméré, T.A., McGowen, M.R., Berta, A., Gatesy, J., 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. Syst. Biol. 57, 15–37. https://doi.org/10.1080/10635150701884632

Dissel-Scherft, M.C.V., Vervoort, W. 1954. Development of the teeth in fetal *Balaeonoptera physalus* (L.) (Cetacea, Mystacoceti). Proc. Nederl. Akad. Wetensch Amsterdam 57, 196–210.

Doronina, L., Hughes, G.M., Moreno-Santillan, D., Lawless, C., Lonergan, T., Ryan, L., Jebb, D., Kirilenko, B.M., Korstian, J.M., Dávalos, L.M., Vernes, S.C., Myers, E.W., Teeling, E.C., Hiller, M., Jermiin, L.S., Schmitz, J., Springer, M.S., Ray,

D.A., 2022. Contradictory Phylogenetic Signals in the Laurasiatheria Anomaly Zone. Genes (Basel). 13. https://doi.org/10.3390/genes13050766

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L., 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, *356*(6333), 92–95. https://doi.org/10.1126/science.aal3327

Ekdale, E.G., Deméré, T.A., 2021. Neurovascular evidence for a co-occurrence of teeth and baleen in an Oligocene mysticete and the transition to filter-feeding in baleen whales. Zool. J. Linn. Soc. 1–21. https://doi.org/10.1093/zoolinnean/zlab017

Emerling, C.A., Widjaja, A.D., Nguyen, N.N., Springer, M.S., 2017. Their loss is our gain: regressive evolution in vertebrates provides genomic models for uncovering human disease loci. J. Med. Genet. 54, 787–794. https://doi.org/10.1136/jmedgenet-2017-104837

Emerling, C.A., Springer, M.S., Gatesy, J., Jones, Z., Hamilton, D., Xia-Zhu, D., Collin, M., Delsuc, F., 2021. Genomic evidence for the parallel regression of melatonin synthesis and signaling pathways in placental mammals. Open Res. Eur. 1, 75. https://doi.org/10.12688/openreseurope.13795.1

Emerling, C.A., Tilak, M., Jonathan, J., Kuch, M., Ana, T., Gibb, G.C., Tilak, M., Hughes, J.J., Duggan, A.T., Poinar, H.N., Nachman, M.W., Delsuc, F., 2023. Genomic data suggest parallel dental vestigialization within the xenarthran radiation. Peer Community J. 3. https://doi.org/https://doi.org/10.24072/pcjournal.303

Espinasa, M., Espinasa, L., 2008. Losing Sight of Regressive Evolution. Evol. Educ. Outreach 1, 509–516. https://doi.org/10.1007/s12052-008-0094-z

Fitzgerald, E.M.G., 2006. A bizarre new toothed mysticete (Cetacea) from Australia and the early evolution of baleen whales. Proc. Roy. Soc. B 273, 2955–2963.

Fitzgerald, E.M.G., 2010. The morphology and systematics of *Mammalodon colliveri* (Cetacea: Mysticeti), a toothed mysticete from the Oligocene of Australia. Zool. J. Linn. Soc.-Lond. 158, 367–476.

Flower, W.H., Lydekker, R. 1891. An Introduction to the Study of Mammals Living and Extinct. Adam and Charles Black, London.

Foley, N.M., Mason, V.C., Harris, A.J., Bredemeyer, K.R., Damas, J., Lewin, H.A., Eizirik, E., Gatesy, J., Karlsson, E.K., Lindblad-Toh, K., Springer, M.S., Murphy, W.J., Andrews, G., Armstrong, J.C., Bianchi, M., Birren, B.W., Bredemeyer, K.R., Breit, A.M., Christmas, M.J., Clawson, H., Damas, J., Di Palma, F., Diekhans, M.,

Dong, M.X., Eizirik, E., Fan, K., Fanter, C., Foley, N.M., Forsberg-Nilsson, K., Garcia, C.J., Gatesy, J., Gazal, S., Genereux, D.P., Goodman, L., Grimshaw, J., Halsey, M.K., Harris, A.J., Hickey, G., Hiller, M., Hindle, A.G., Hubley, R.M., Hughes, G.M., Johnson, J., Juan, D., Kaplow, I.M., Karlsson, E.K., Keough, K.C., Kirilenko, B., Koepfli, K.-P., Korstian, J.M., Kowalczyk, A., Kozyrev, S. V., Lawler, A.J., Lawless, C., Lehmann, T., Levesque, D.L., Lewin, H.A., Li, X., Lind, A., Lindblad-Toh, K., Mackay-Smith, A., Marinescu, V.D., Marques-Bonet, T., Mason, V.C., Meadows, J.R.S., Meyer, W.K., Moore, J.E., Moreira, L.R., Moreno-Santillan, D.D., Morrill, K.M., Muntané, G., Murphy, W.J., Navarro, A., Nweeia, M., Ortmann, S., Osmanski, A., Paten, B., Paulat, N.S., Pfenning, A.R., Phan, B.N., Pollard, K.S., Pratt, H.E., Ray, D.A., Reilly, S.K., Rosen, J.R., Ruf, I., Ryan, L., Ryder, O.A., Sabeti, P.C., Schäffer, D.E., Serres, A., Shapiro, B., Smit, A.F.A., Springer, M., Srinivasan, C., Steiner, C., Storer, J.M., Sullivan, K.A.M., Sullivan, P.F., Sundström, E., Supple, M.A., Swofford, R., Talbot, J.-E., Teeling, E., Turner-Maier, J., Valenzuela, A., Wagner, F., Wallerman, O., Wang, C., Wang, J., Weng, Z., Wilder, A.P., Wirthlin, M.E., Xue, J.R., Zhang, X., 2023. A genomic timescale for placental mammal evolution. Science (80-.). 380. https://doi.org/10.1126/science.abl8189

Fordyce, R.E., Marx, F.G., 2018. Gigantism precedes filter feeding in baleen whale evolution. Curr. Biol. 28, 1670–1676.

Fordyce, R.E., Mattlin, R.H., Wilson, G.J., 1979. Stranding of a Cuvier's beaked whale, *Ziphius cavirostris* Cuvier, 1823, at New Brighton, New Zealand. Mauri Ora 7, 73-82.

Franzen, J.L., 2005. The implications of the numerical dating of the Messel fossil deposit (Eocene, Germany) for mammalian biochronology. Ann. Paleontol. 91, 329–335. https://doi.org/10.1016/j.annpal.2005.04.002

Fukumoto, S., Kiba, T., Hall, B., Iehara, N., Nakamura, T., Longenecker, G., Krebsbach, P.H., Nanci, A., Kulkarni, A.B., Yamada, Y., 2004. Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts. J. Cell Biol 167, 973-983.

Gasse, B., Silvent, J., Sire, J.-Y., 2012. Evolutionary analysis suggests that *AMTN* is enamel-specific and a candidate for AI. J. Dent. Res. 91 (11), 1085–1089.

Gatesy, J., Geisler, J.H., Chang, J., Buell, C., Berta, A., Meredith, R.W., Springer, M.S., McGowen, M.R., (2013). A phylogenetic blueprint for a modern whale. Mol. Phylogenet. Evol. 66, 479–506. https://doi.org/10.1016/j.ympev.2012.10.012

Gatesy, J., Ekdale, E.G., Deméré, T.A., Lanzetti, A., Randall, J., Berta, A., El Adli, J.J., Springer, M.S., McGowen, M.R., 2022. Anatomical, ontogenetic, and genomic

homologies constrain reconstructions of the teeth-to-baleen transition in mysticete whales. J. Mamm. Evol.

Gaubert, P., Antunes, A., Meng, H., Miao, L., Peigné, S., Justy, F., Njiokou, F., Dufour, S., Danquah, E., Alahakoon, J., Verheyen, E., Stanley, W.T., O'Brien, S.J., Johnson, W.E., Luo, S.J., 2018. The Complete Phylogeny of Pangolins: Scaling Up Resources for the Molecular Tracing of the Most Trafficked Mammals on Earth. J. Hered. 109, 347–359. https://doi.org/10.1093/jhered/esx097

Gaudin, T.J., Emry, R.J., Wible, J.R., 2009. The phylogeny of living and extinct pangolins (mammalia, pholidota) and associated taxa: A morphology based analysis. J. Mamm. Evol. 16, 235–305. https://doi.org/10.1007/s10914-009-9119-9

Gaudin, T.J., Gaubert, P., Billet, G., Hautier, L., Ferreira-Cardoso, S., Wible, J.R., 2019a. Evolution and morphology. Pangolins Sci. Soc. Conserv. 5–23. https://doi.org/10.1016/B978-0-12-815507-3.00001-0

Gaudin, T. J., Wible, J. R., Rose, K. D., Emry, R. J., Spaulding, M., 2019b. Skeletal Anatomy of the Basicranium and Auditory Region in the Metacheiromyid Palaeanodont *Metacheiromys* (Mammalia, Pholidotamorpha) Based on High-Resolution CT Scans. Fossil Imprint, 75(3-4): 484–503, Praha. ISSN 2533–4050 (print), ISSN 2533–4069

Gaudry, M.J., Jastroch, M., Treberg, J.R., Hofreiter, M., Paijmans, J.L.A., Starrett, J., Wales, N., Signore, A.V., Springer, M.S., and Campbell, K.L., 2017. Inactivation of thermogenic UCP1 as a historical contingency in multiple placental mammal clades. Sci. Adv. 3, e1602878.

Gheerbrant, E., Rose, K.D., Godinot, M., 2005. First palaeanodont (?pholidotan) mammal from the Eocene of Europe. Acta Palaeontol. Pol. 50, 209–218.

Geisler, J.H., Boessenecker, R.W., Brown, M., Beatty, B.L., 2017. The origin of filter feeding in whales. Curr. Biol. 27, 2036-2042. https://doi.org/10.1016/j.cub.2017.06.003

Gingerich, P.D., 2012. Evolution of whales from land to sea. Proc. Am. Philos. Soc. 156, 309–323.

Gomerčić, H., Gomerčić, M.D., Gomerčić, T., Lucić, H., Dalebout, M., Galov, A., Škrtić, D., Ćurković, S., Vuković, S., Huber, D., 2006. Biological aspects of Cuvier's beaked whale (*Ziphius cavirostris*) recorded in the Croatian part of the Adriatic Sea. Eur. J. Wildl. Res. 52, 182–187.

Gunnell, G.F., Gingerich, P.D., 1993. Skeleton of *Brachianodon westorum*, a new middle Eocene metacheiromyid (Mammalia, Palaeanodonta) from the early Bridgerian (Bridger A) of the southern Green River Basin, Wyoming. Contrib. from Museum Paleontol. Univ. Michigan 28, 365–392.

Han, G.Z., 2020. Pangolins Harbor SARS-CoV-2-Related Coronaviruses. Trends Microbiol. 28, 515–517. https://doi.org/10.1016/j.tim.2020.04.001

Handley Jr, C. O., 1966. A synopsis of the genus *Kogia* (pygmy sperm whales). Whales, dolphins and porpoises, KS Norris (ed.). University of California Press, Berkeley, 62-69.

Hart, P.S., Hart, T.C., Michalec, M.D., Ryu, O.H., Simmons, D., Hong, S., Wright, J.T., 2004. Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta. J. Med. Genet. 41, 545–549.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B.J., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C. R. Biol. 335, 32–50. https://doi.org/10.1016/j.crvi.2011.11.002

Heyning, J.E., Mead, J.G., 1996. Suction feeding in beaked whales: morphological and observational evidence. Contrib. Sci. https://doi.org/10.5962/p.226802

Hocking, D.P., Marx, F.G., Fitzgerald, E.M.G., Evans, A.R., 2017. Ancient whales did not filter feed with their teeth. Biol. Lett. 13, 20170348. https://doi.org/10.1098/rsbl.2017.0348

Hołyńska, M., Leggitt, L., Kotov, A.A., 2016. Miocene Cyclopid Copepod from a Saline Paleolake in Mojave, California. Acta Palaeontol. Pol. 61, 345–361. https://doi.org/10.4202/app.00137.2014

Hooker, S.K., 2018. Toothed Whales (Odontoceti), in: Encyclopedia of Marine Mammals. Elsevier, pp. 1004–1010.

Hu, J.-C., Hu, Y., Smith, C.E., McKee, M.D., Wright, J.T., Yamakoshi, Y., Papagerakis, P., Hunter, G.K., Feng, J.Q., Yamakoshi, F., Simmer, J.P., 2008. Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice. J. Biol. Chem. 283 (16), 10858–10871.

Huelsmann, M., Hecker, N., Springer, M.S., Gatesy, J., Sharma, V., and Hiller, M., 2019. Genes lost during the transition from land to water in cetaceans highlight genomic changes involved in aquatic adaptations. Sci. Adv. 5, eaaw6671.

Ikeda, Y., Neshatian, M., Holcroft, J., Ganss, B., 2018. The enamel protein ODAM promotes mineralization in a collagen matrix. Connect. Tissue Res. 59, 62–66. https://doi.org/10.1080/03008207.2017.1408603

Ishikawa, H., Amasaki, H. 1995. Development and physiological degradation of tooth buds and development of rudiment of baleen plate in southern minke whale, *Balaenoptera acutorostrata*. J. Vet. Med. Sci. 57, 665–670.

Ishiyama, M., 1987. Enamel structure in odontocete whales. Scanning Microsc. 1, 1071–1079.

Jefferson, T.A., Hung, S.K., 2004. *Neophocaena phocaenoides*. Mammalian Species, (746), pp.1-12. https://doi.org/10.1644/746

Jett, J., Visser, I.N., Ventre, J., Waltz, J., Loch, C., 2017. Tooth damage in captive orcas (*Orcinus orca*). Arch. Oral Biol. 84, 151–160. https://doi.org/10.1016/j.archoralbio.2017.09.031

Ji, Y., Li, C., Tian, Y., Gao, Y., Dong, Z., Xiang, L., Xu, Z., Gao, Y., Zhang, L., 2021. Maturation stage enamel defects in Odontogenesis-associated phosphoprotein (Odaph) deficient mice. Dev. Dyn. 250, 1505–1517. https://doi.org/10.1002/dvdy.336

Johnston, C., Berta, A., 2011. Comparative anatomy and evolutionary history of suction feeding in cetaceans. Mar. Mammal Sci. 27, 493–513. https://doi.org/10.1111/j.1748-7692.2010.00420.x

Kawasaki, K., 2011. The SCPP gene family and the complexity of hard tissues in vertebrates. Cells Tissues Organs 194, 108–112. https://doi.org/10.1159/000324225

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9 (4), 286–298.

Kawasaki, K., Weiss, K.M., 2008. SCPP gene evolution and the dental mineralization continuum. J. Dent. Res. 87, 520–531.

Kawasaki, K., Hu, J.-C.-C., Simmer, J.P., 2014. Evolution of *Klk4* and enamel maturation in eutherians. Biol. Chem. 395, 1003–1013. https://doi.org/10.1515/hsz-2014- 0122

Kawasaki K., Mikami M., Goto M, Shindo J., Amano M., Ishiyama M., 2020. The evolution of unusually small amelogenin genes in cetaceans; pseudogenization, X–Y gene conversion, and feeding strategy. J. Mol. Evol. 88,122–135. doi.org/10.1007/s00239-019-09917-0

Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., Michalak, P., Kang, L., Bhak, J., Yim, Y.-S., Grishin, N.V., Hielsen, N.H., Heide-Jørgensen, Oziolor, E.M., Matson, C.W., Church, G.M., Stuart, G.W., Patton, J.C., George, J.C., Suydam, R., Larsen, K., López-Otín, C., O'Connell, J.O., Bickham, J.W., Thomsen, B., de Magalhães, J.P., 2015. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 10, 112–122. https://doi.org/10.1016/j.celrep.2014.12.008

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

Kim, J.W., Simmer, J.P., Hart, T.C., Hart, P.S., Ramaswami, M.D., Bartlett, J.D., Hu, J.C., 2005. MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta. J. Med. Genet. 42, 271–275.

Kondrashov, P., Agadjanian, A.K., 2012. A nearly complete skeleton of Ernanodon (Mammalia, Palaeanodonta) from Mongolia: Morphofunctional analysis. J. Vertebr. Paleontol. 32, 983–1001. https://doi.org/10.1080/02724634.2012.694319

Lagerström, M., Dahl, N., Nakahori, Y., Nakagome, Y., Bäckman, B., Landegren, U., Pettersson, U., 1991. A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). Genomics 10, 971–975.

Lambert, O., Bianucci, G., Post, K., 2008. A new beaked whale (Odontoceti, Ziphiidae) from the Middle Miocene of Peru. J. Vertebr. Paleontol. 29, 910–922.

Lambert, O., Bianucci, G., de Muizon, C., 2017. Sperm and Beaked Whales, Evolution. Encycl. Mar. Mamm. 916–918. https://doi.org/10.1016/b978-0-12-804327-1.00241-7

Lanzetti, A., 2019. Prenatal developmental sequence of the skull of minke whales and its implications for the evolution of mysticetes and the teeth-to-baleen transition. J. Anat. 235, 725–748. https://doi.org/10.1111/joa.13029

Lanzetti, A., Berta, A., Ekdale, E.G., 2020. Prenatal development of the humpback whale: growth rate, tooth loss and skull shape changes in an evolutionary framework. Anat. Rec. 303, 180–204.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30(22): 3276-3278. http://dx.doi.org/10.1093/bioinformatics/btu531

Lartillot, N., Delsuc, F., 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. Evolution (N. Y). 66, 1773–1787. https://doi.org/10.1111/j.1558-5646.2011.01558.x

Lartillot, N., Poujol, R., 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28, 729–744. https://doi.org/10.1093/molbev/msq244

Leary, M. a O., Bloch, J.I., Flynn, J.J., Gaudin, T.J., Giallombardo, A., Giannini, N.P., Goldberg, S.L., Kraatz, B.P., Luo, Z., Meng, J., Ni, X., Novacek, M.J., Perini, F. a, Randall, Z.S., Rougier, G.W., Sargis, E.J., Silcox, M.T., Simmons, N.B., Spaulding, M., Velazco, P.M., Weksler, M., Wible, J.R., Cirranello, A.L., 2013. of Placentals. Science (80-.). 339, 662–667.

Line, S.R.P., Novaes, P.D., 2005. The Development and Evolution of Mammalian Enamel: Structural and Functional Aspects. Braz. J. morphol. Sci 22, 67–72.

Loch, C., van Vuuren, L.J., 2016. Ultrastructure, biomechanical and chemical properties of the vestigial dentition of a Cuvier's beaked whale. New Zeal. J. Zool. 43, 171–178.

Loch, C., Duncan, W., Simões-Lopes, P.C., Kieser, J.A., Fordyce, R.E., 2013a. Ultrastructure of enamel and dentine in extant dolphins (Cetacea: Delphinoidea and Inioidea). Zoomorphology 132, 215–225.

Loch, C., Swain, M.V., van Vuuren, L.J., Kieser, J.A., Fordyce, R.E., 2013b. Mechanical properties of dental tissues in dolphins (Cetacea: Delphinoidea and Inioidea). Arch. Oral Biol. 58, 773–779.

Loch, C., Swain, M. V., Fraser, S.J., Gordon, K.C., Kieser, J.A., Fordyce, R.E., 2014. Elemental and chemical characterization of dolphin enamel and dentine using X-ray and Raman microanalyzes (Cetacea: Delphinoidea and Inioidea). J. Struct. Biol. 185, 58–68. https://doi.org/10.1016/j.jsb.2013.11.006

Loch, C., Kieser, J.A., Fordyce, R.E., 2015. Enamel ultrastructure in fossil cetaceans (Cetacea: Archaeoceti and Odontoceti). PLoS One 10, 1–14. https://doi.org/10.1371/journal.pone.0116557

Loch, C., Buono, M.R., Kalthoff, D.C., Mörs, T., Fernández, M.S., 2020. Enamel microstructure in Eocene cetaceans from Antarctica (Archaeoceti and Mysticeti). J. Mamm. Evol. 27, 289–298.

MacLeod, C.D., 1998. Intraspecific scarring in odontocete cetaceans: An indicator of male "quality" in aggressive social interactions? J. Zool. 244, 71–77. https://doi.org/10.1017/S0952836998001083

MacLeod, C.D., 2017. Beaked Whales, Overview. Encycl. Mar. Mamm. 80–83. https://doi.org/10.1016/b978-0-12-804327-1.00062-5

Maddison, W. P., Maddison, D.R., 2023. Mesquite: a modular system for evolutionary analysis. Version 3.70 http://www.mesquiteproject.org

McAlpine, D. F., 2018. "Pygmy and dwarf sperm whales: *Kogia breviceps* and *K. sima*." Encyclopedia of marine mammals. Academic Press. 786-788.

McGowen, M.R., Spaulding, M., Gatesy, J., 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol. Phylogenet. Evol. 53, 891–906. https://doi.org/10.1016/j.ympev.2009.08.018

McGowen, M.R., Gatesy, J., Wildman, D.E., 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. Trends Ecol. Evol. 29, 336–346.

McGowen, M.R., Tsagkogeorga, G., Álvarez-Carretero, S., dos Reis, M., Struebig, M., Deaville, R., Jepson, P.D., Jarman, S., Polanowski, A., Morin, P.A., Rossiter, S.J., 2020a. Phylogenomic resolution of the cetacean tree of life using target sequence capture. Syst. Biol. 69, 79–501.

McGowen, M.R., Tsagkogeorga, G., Williamson, J., Morin, P.A., Rossiter, S.J., 2020b. Positive selection and inactivation in the vision and hearing genes of cetaceans. Mol. Biol. Evol. 37, 2069–2083.

McKnight, D.A., Fisher, L.W., 2009. Molecular evolution of dentin phosphoprotein among toothed and toothless animals. BMC Evol. Biol. 9, 299. https://doi.org/10.1186/1471-2148-9-299

McKnight, D.A., Simmer, J.P., Hart, P.S., Hart, T.C., Fisher, L.W., 2008. Overlapping *DSPP* mutations cause dentin dysplasia and dentinogenesis imperfecta. J. Dent. Res. 87, 1108–1111.

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 Genet. 5, e1000634. https://doi.org/10.1371/journal.pgen.1000634

Meredith, R.W., Gatesy, J., Cheng, J., Springer, M.S., 2011a. Pseudogenization of the tooth gene enamelysin (*MMP20*) in the common ancestor of extant baleen whales. Proc. R. Soc. B Biol. Sci. 278, 993–1002. https://doi.org/10.1098/rspb.2010.1280

Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011b. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334, 521–524.

Meredith, R.W., Gatesy, J., Springer, M.S., 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. BMC Evol. Biol. 13, 20. https://doi.org/10.1186/1471-2148-13-20

Meredith, R.W., Zhang, G., Gilbert, M.T.P., Jarvis, E.D., Springer, M.S., 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. Science (80-.). 346. https://doi.org/10.1126/science.1254390

Miller, E.H., 2018. Territorial Behavior. Encycl. Mar. Mamm. 1156–1166. https://doi.org/10.1016/B978-0-12-373553-9.00266-2

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010. https://doi.org/10.1109/GCE.2010.5676129

Moran, R.L., Richards, E.J., Ornelas-García, C.P., Gross, J.B., Donny, A., Wiese, J., Keene, A.C., Kowalko, J.E., Rohner, N., McGaugh, S.E., 2023. Selection-driven trait loss in independently evolved cavefish populations. Nat. Commun. 14, 1–19. https://doi.org/10.1038/s41467-023-37909-8

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

Mu, Y., Tian, R., Xiao, L., Sun, D., Zhang, Z., Xu, S., Yang, G., 2021b. Molecular Evolution of Tooth-Related Genes Provides New Insights into Dietary Adaptations of Mammals. J. Mol. Evol. 89, 458–471. https://doi.org/10.1007/s00239-021-10017-1

Murphy, W.J., Eizirik, E., 2016. 4. Placental Mammals (Eutheria). Princet. F. Guid. to Prehist. Mamm. 47–50. https://doi.org/10.1515/9781400884452-006

Murphy, W.J., Foley, N.M., Bredemeyer, K.R., Gatesy, J., Springer, M.S., 2021. Phylogenomics and the Genetic Architecture of the Placental Mammal Radiation. Annu. Rev. Anim. Biosci. 9, 29–53. https://doi.org/10.1146/annurev-animal-061220-023149

Nakayama, Y., Holcroft, J., Ganss, B., 2015. Enamel hypomineralization and structural defects in amelotin deficient mice. J. Dent. Res. 94, 697–705.

Nery, M.F., González, D.J., Opazo, J.C., 2013. How to make a dolphin: molecular signature of positive selection in cetacean genome. PLoS ONE, 8, e65491.

Novacek, M.J., 1994. The Radiation of Placental Mammals. Short Courses Paleontol. 7, 220–237. https://doi.org/10.1017/s2475263000001331

Núñez, S.M., Chun, Y.H.P., Ganss, B., Hu, Y., Richardson, A.S., Schmitz, J.E., Fajardo, R., Yang, J., Hu, J.C.C., Simmer, J.P., 2016. Maturation stage enamel malformations in *Amtn* and *Klk4* null mice. Matrix Biol. 52–54, 219–233. https://doi.org/10.1016/j.matbio.2015.11.007

Nweeia, M.T., Eichmiller, F.C., Hauschka, P. V., Donahue, G.A., Orr, J.R., Ferguson, S.H., Watt, C.A., Mead, J.G., Potter, C.W., Dietz, R., Giuseppetti, A.A., Black, S.R., Trachtenberg, A.J., Kuo, W.P., 2014. Sensory ability in the narwhal tooth organ system. Anat. Rec. 297, 599–617. https://doi.org/10.1002/ar.22886

Okazaki, Y., 2012. A new mysticete from the upper Oligocene Ashiya Group, Kyushu, Japan, and its significance to mysticete evolution. Bull. Kitakyushu Mus. Nat. Hist. Hum. Hist. Series A (Nat. Hist.) 10, 129–152.

Papagerakis, P., Lin, H.K., Lee, K.Y., Hu, Y., Simmer, J.P., Bartlett, J.D., Hu, J.C., 2008. Premature stop codon in *MMP20* causing amelogenesis imperfecta. J. Dent. Res. 87, 56–59.

Parry, D.A., Brookes, S.J., Logan, C.V., Poulter, J.A., El-Sayed, W., Al-Bahlani, S., Al Harasi, S., Sayed, J., Raif, E.M., Shore, R.C., Dashash, M., Barron, M., Morgan, J.E., Carr, I.M., Taylor, G.R., Johnson, C.A., Aldred, M.J., Dixon, M.J., Wright, J.T., Kirkham, J., Inglehearn, C.F., Mighell, A.J., 2012. Mutations in *C4orf26*, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. Am. J. Hum. Gen. 91, 565–571.

Peredo, C.M., Pyenson, N.D., Boersma, A.T., 2017. Decoupling tooth loss from the evolution of baleen in whales. Front. Mar. Sci. 4, 67. https://doi.org/10.3389/fmars.2017.00067

Peredo, C.M., Pyenson, N.D., Marshall, C.D., Uhen, M.D., 2018. Tooth loss precedes the origin of baleen in whales. Curr. Biol. 28, 3992-4000. https://doi.org/10.1016/j.cub.2018.10.047

Plön, S. 2004. The status and natural history of pygmy (*Kogia breviceps*) and dwarf (*K. sima*) sperm whales off Southern Africa. Unpublished Ph.D. dissertation, Rhodes University.

Porter, M.L., Crandall, K.A., 2003. Lost along the way: the significance of evolution in reverse. Trends Ecol. Evol. 18, 541–547. https://doi.org/10.1016/S0169-5347(03)00244-1

Poulter, J.A., Murillo, G., Brookes, S.J., Smith, C.E., Parry, D.A., Silva, S., Kirkham, J., Inglehearn, C.F., Mighell, A.J., 2014. Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta. Hum. Mol. Genet. 23, 5317–5324.

Prakash, S.K., Gibson, C.W., Wright, J.T., Boyd, C., Cormier, T., Sierra, R., Li, Y., Abrams, W.R., Aragon, M.A., Yuan, Z.A., Van den Veyver, I.B., 2005. Tooth enamel defects in mice with a deletion at the Arhgap6/AmelX locus. Calcif. Tissue Int. 77, 23–29.

R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radhi, A., 2017. A quantitative study of hunter- schreger bands in the tooth enamel of *camelus dromedarius*. Ph.D. dissertation, Royal College of Surgeons in Ireland.

Rajpar, M.H., Harley, K., Laing, C., Davies, R.M., Dixon, M.J., 2001. Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. Hum. Mol. Genet. 10, 1673–1677.

Randall, J.G., Gatesy, J., Springer, M.S., 2022. Molecular evolutionary analyses of tooth genes support sequential loss of enamel and teeth in baleen whales (Mysticeti). Mol. Phylogenet. Evol. 171, 107463. https://doi.org/10.1016/j.ympev.2022.107463

Rice, D.W., 1998. Marine mammals of the world: systematics and distribution. Soc. Mar. Mamm. Spec. Publ. 4, 1–231.

Rodrigues, H.G., Lihoreau, F., Orliac, M., Thewissen, J.G.M., Boisserie, J.R., 2019. Unexpected evolutionary patterns of dental ontogenetic traits in cetartiodactyl mammals. Proc. R. Soc. B Biol. Sci. 286. https://doi.org/10.1098/rspb.2018.2417

Rose, K.D., 2012. The importance of Messel for interpreting Eocene Holarctic mammalian faunas. Palaeobiodiversity and Palaeoenvironments 92, 631–647. https://doi.org/10.1007/s12549-012-0090-8

Rose, K.D., 2008. Palaeanodonta and pholidota. Evol. Tert. Mamm. North Am. Vol. 2 Small Mammals, Xenarthrans, Mar. Mamm. 2, 135–146. https://doi.org/10.1017/CBO9780511541438.010

Rose, K.D., Lucas, S.G., 2000. An early paleocene palaeanodont (mammalia, ?pholidota) from new mexico, and the origin of palaeanodonta. J. Vertebr. Paleontol. 20, 139–156. https://doi.org/10.1671/0272-4634(2000)020[0139:AEPPMP]2.0.CO;2

Sathe, V., 2000. Enamel Ultrastructure of Cattle from the Quaternary Period in India. Environ. Archaeol. 5, 107–115. https://doi.org/10.1179/env.2000.5.1.107

Schoch, R.M., 1984. Revision of *Metacheiromys* Wortman, 1903 and a Review of the Palaeanodonta. Postilla.

Secord, R., Gingerich, P.D., Bloch, J.I., 2002. *Mylandon rosei*, a new Metacheiromyid (Mammalia, Palaeanodonta) from the Late Tiffanian (Late Paleocene) of Northwestern Wyoming. Contrib. from Museum Paleontol. Univ. Michigan 30, 385–399.

Seton, M., Müller, R.D., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M., Maus, S., Chandler, M., 2012. Global continental and ocean basin reconstructions since 200Ma. Earth-Science Rev. 113, 212–270. https://doi.org/10.1016/j.earscirev.2012.03.002

Seymen, F., Kim, Youn Jung, Lee, Y.J., Kang, J., Kim, T.H., Choi, H., Koruyucu, M., Kasimoglu, Y., Tuna, E.B., Gencay, K., Shin, T.J., Hyun, H.K., Kim, Young Jae, Lee, S.H., Lee, Z.H., Zhang, H., Hu, J.C.C., Simmer, J.P., Cho, E.S., Kim, J.W., 2016. Recessive Mutations in *ACPT*, Encoding Testicular Acid Phosphatase, Cause Hypoplastic Amelogenesis Imperfecta. Am. J. Hum. Genet. 99, 1199–1205. https://doi.org/10.1016/j.ajhg.2016.09.018

Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., Langer, B.E., Hiller, M., 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat. Comm. 9, 1215.

Simmer, J.P., Hu, Y., Lertlam, R., Yamakoshi, Y., Hu, J.C.C., 2009. Hypomaturation enamel defects in *Klk4* knockout/LacZ knockin mice. J. Biol. Chem. 284, 19110–19121.

Sire, J.-Y., Delgado, S., Fromentin, D., Girondot, M., 2005. Amelogenin: lessons from evolution. Arch. Oral Biol. 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. Eur. J. Oral Sci. 114 (s1), 64–77. https://doi.org/10.1111/j.1600-0722.2006.00297

Sire, J.-Y., Davit-B´eal, T., Delgado, S., Gu, X., 2007. The origin and evolution of enamel mineralization genes. Cells Tissues Organs 186, 25–48. https://doi.org/10.1159/000102679

Slaughter, B.H., Pine, R.H., Pine, N.E., 1974. Eruption of cheek teeth in insectivora and carnivora. J. Mammal. 55, 115–125. https://doi.org/10.2307/1379261

Smith, C.E.L., Murillo, G., Brookes, S.J., Poulter, J.A., Silva, S., Kirkham, J., Inglehearn, C.F., Mighell, A.J., 2016. Deletion of amelotin exons 3–6 is associated with amelogenesis imperfecta. Hum. Mol. Genet. 25, 3578–3587.

Smith, C.E., Whitehouse, L.L., Poulter, J.A., Brookes, S.J., Day, P.F., Soldani, F., Kirkham, J., Inglehearn, C.F., Mighell, A.J., 2017. Defects in the acid phosphatase *ACPT* cause recessive hypoplastic amelogenesis imperfecta. Eur. J. Hum. Genet. 25, 1015–1019.

Spaulding, M., O'Leary, M.A., Gatesy, J., 2009. Relationships of Cetacea (Artiodactyla) among mammals: Increased taxon sampling alters interpretations of key fossils and character evolution. PLoS One 4. https://doi.org/10.1371/journal.pone.0007062

Springer, M.S., Gatesy, J., 2018. Evolution of the MC5R gene in placental mammals with evidence for its inactivation in multiple lineages that lack sebaceous glands. Mol. Phylogenet. Evol. 120, 364–374. https://doi.org/10.1016/j.ympev.2017.12.010

Springer, M.S., Signore, A.V., Paijmans, J.L.A., Vélez-Juarbe, J., Domning, D.P., Bauer, C.E., He, K., Crerar, L., Campos, P.F., Murphy, W.J., Meredith, R.W., Gatesy, J., Willerslev, E., MacPhee, R.D.E., Hofreiter, M., Campbell, K.L., 2015. Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. Mol. Phylogenet. Evol. 91, 178–193.

Springer, M.S., Starrett, J., Morin, P.A., Lanzetti, A., Hayashi, C., Gatesy, J., 2016a. Inactivation of *C4orf26* in toothless placental mammals. Mol. Phylogenet. Evol. 95, 34–45. https://doi.org/10.1016/j.ympev.2015.11.002

Springer, M.S., Emerling, C.A., Fugate, N., Patel, R., Starrett, J., Morin, P.A., Hayashi, C., Gatesy, J., 2016b. Inactivation of cone-specific phototransduction genes in rod monochromatic cetaceans. Front. Ecol. Evol. 4, 61.

Springer, M.S., Emerling, C.A., Gatesy, J., Randall, J., Collin, M.A., Hecker, N., Hiller, M., Delsuc, F., 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. BMC Evol. Biol. 19. https://doi.org/10.1186/s12862-019-1359-6

Springer, M.S., Molloy, E.K., Sloan, D.B., Simmons, M.P., Gatesy, J., 2020. ILS-aware analysis of low-homoplasy retroelement insertions: inference of species trees and introgression using quartets. J. Hered. 111, 147–168.

Springer, M.S., Guerrero-Juarez, C.F., Huelsmann, M., Collin, M.A., Danil, K., McGowen, M.R., Oh, J.W., Ramos, R., Hiller, M., Plikus, M.V., Gatesy, J., 2021. Genomic and anatomical comparisons of skin support independent adaptation to life in water by cetaceans and hippos. Curr. Biol. 31, 1-16. https://doi.org/10.1016/j.cub.2021.02.057

Springer, M.S., Emerling, C.A., Gatesy, J., 2023. Three Blind Moles: Molecular Evolutionary Insights on the Tempo and Mode of Convergent Eye Degeneration in *Notoryctes typhlops* (Southern Marsupial Mole) and Two Chrysochlorids (Golden Moles). Genes (Basel). https://doi.org/https://doi.org/10.3390/genes14112018

Stamatakis, A., 2006. Phylogenetic models of rate heterogeneity: A high performance computing perspective. 20th Int. Parallel Distrib. Process. Symp. IPDPS 2006 2006. https://doi.org/10.1109/IPDPS.2006.1639535

Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 20, 1312–1313.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771. https://doi.org/10.1080/10635150802429642

Swofford, D.L., 2003. PAUP* Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0 http://paup.csit.fsu.edu/

Terhune, C.E., Gaudin, T., Curran, S., Petculescu, A., 2021. The youngest pangolin (Mammalia, Pholidota) from Europe. J. Vertebr. Paleontol. 41, 1–10. https://doi.org/10.1080/02724634.2021.1990075

Thermudo, G.E., Alves, L.Q., Machado, A.M., Lopes-Marques, M., de Fonseca, R.R., Fonseca, M., Ruivo, R., Castro, L.F.C., 2020. Losing genes: the evolutionary remodeling of Cetacea skin. Front. Mar. Sci. 7, 592375.

Thewissen, J.G.M. 2018. Berardius beaked whales: *Berardius bairdii* and *B. arnuxii*. In: Würsig, B., Thewissen, J.G.M., Kovacs, K.M. (Eds.), Encyclopedia of Marine Mammals. Academic Press, Cambridge, MA, pp. 97–99.

Thewissen, J.G.M., Bajpai, S., 2001. Whale origins as a poster child for macroevolution. BioScience 51, 1037–1049.

Thewissen, J.G.M., Williams, E.M., Roe, L.J., Hussain, S.T., 2001. Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. Nature 413, 277–281.

Thewissen, J.G.M., Williams, E.M., 2002. The early radiations of Cetacea (Mammalia): evolutionary pattern and developmental correlations. Annu. Rev. Ecol. Syst. 33, 73–90. https://doi.org/10.1146/annurev.ecolsys.33.020602.095426

Thewissen, J.G.M., Cooper, L.N., George, J.C., Bajpai, S., 2009. From land to water: the origin of whales, dolphins, and porpoises. Evo. Edu. Outreach 2, 272–288.

Thewissen, J.G.M., Hieronymus, T.L., George, J.C., Suydam, R., Stimmelmayr, R., McBurney, D. 2017. Evolutionary aspects of the development of teeth and baleen in the bowhead whale. J. Anat. 230, 549–566.

Uhen, M.D., 2010. The origin(s) of whales. Annu. Rev. Earth Planet Sci. 38, 189–219. https://doi.org/10.1146/annurev-earth-040809-152453

Waddell, P.J., Okada, N., Hasegawa, M., 1999. Towards Resolving the Interordinal Relationships of Placental Mammals. Syst. Biol. 48, 1–5.

Werth, A.J., Loch, C., Fordyce, R.E., 2020. Enamel microstructure in Cetacea: a case study in evolutionary loss of complexity. J. Mamm. Evol. 27, 789–805. https://doi.org/10.1007/s10914-019-09484-7

Westbury, M., Baleka, S., Barlow, A., Hartmann, S., Paijmans, J.L.A., Kramarz, A., Forasiepi, A.M., Bond, M., Gelfo, J.N., Reguero, M.A., López-Mendoza, P., Taglioretti, M., Scaglia, F., Rinderknecht, A., Jones, W., Mena, F., Billet, G., De Muizon, C., Aguilar, J.L., MacPhee, R.D.E., Hofreiter, M., 2017. A mitogenomic timetree for Darwin's enigmatic South American mammal *Macrauchenia patachonica*. Nat. Commun. 8, 1–8. https://doi.org/10.1038/ncomms15951

Willis, P.M., Baird, R.W. 1998. Status of the dwarf sperm whale, *Kogia simus*, with special reference to Canada. Can. Field-Nat. 112, 114–125.

Wright, J.T., Hart, T.C., Hart, P.S., Simmons, D., Suggs, C., Daley, B., Simmer, J., Hu, J., Bartlett, J.D., Li, Y., Yuan, Z.A., 2009. Human and mouse enamel phenotypes resulting from mutation or altered expression of *AMEL*, *ENAM*, *MMP20* and *KLK4*. Cells Tissues Organs 189, 224–229.

Xian, Y., Lu, Y., Liu, G., 2022. Is climate change threatening or beneficial to the habitat distribution of global pangolin species? Evidence from species distribution modeling. Sci. Total Environ. 811, 151385. https://doi.org/10.1016/j.scitotenv.2021.151385

Yamakoshi, Y. 2009. Dentinogenesis and dentin sialophosphoprotein (*DSPP*). J. Oral Biosci. 51, 134–142.

Yamakoshi, Y., Simmer, J.P. 2018. Structural features, processing mechanism and gene splice variants of dentin sialophosphoprotein. Jap. Dent. Sci. Rev. 54, 183–196.

Yang, Z., 1997. User Guide PAML 4: Phylogenetic Analysis by Maximum Likelihood. User Guide. 6, 1–70. https://doi.org/10.1093/molbev/msm088

Zhang, J., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8

Zhang, F., Wu, S., Yang, L., Zhang, L., Sun, R., Li, S., 2015. Reproductive parameters of the Sunda pangolin *Manis javanica*. Folia Zool. 64, 129–135. https://doi.org/10.25225/fozo.v64.i2.a6.2015

Zhang, F., Wu, S., Zou, C., Wang, Q., Li, S., Sun, R., 2016. A note on captive breeding and reproductive parameters of the Chinese pangolin, Manis pentadactyla Linnaeus, 1758. Zookeys 2016, 129–144. https://doi.org/10.3897/zookeys.618.8886

Zhang, H., Xie, X., Liu, P., Liang, T., Lu, Y., Qin, C., 2018. Transgenic expression of dentin phosphoprotein (DPP) partially rescued the dentin defects of *DSPP*-null mice. PLoS ONE 13, e0195854.

Zhou, X., Guang, X., Sun, D., Xu, S., Li, M., Seim, I., Jie, W., Yang, L., Zhu, Q., Xu, J., Gao, Q., Kaya, A., Dou, Q., Chen, B., Ren, W., Li, S., Zhou, K., Gladyshev, V.N., Nielsen, R., Fang, X., Yang, G., 2018. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat. Commun. 9, 1–8. https://doi.org/10.1038/s41467-018-03722-x

Zurano, J.P., Magalhães, F.M., Asato, A.E., Silva, G., Bidau, C.J., Mesquita, D.O., Costa, G.C., 2019. Cetartiodactyla: Updating a time-calibrated molecular phylogeny. Mol. Phylogenet. Evol. 133, 256–262. https://doi.org/10.1016/j.ympev.2018.12.015