ACPT gene is inactivated in mammalian lineages that lack enamel or teeth

Yuan Mu [1], **Xin Huang** [1], **Rui Liu** [1], **Yulin Gai** [1], **Na Liang** [1], **Daiqing Yin** [1], **Lei Shan** [1], **Shixia Xu** [1], **Guang Yang** [Corresp. 1]

[1] Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, China

Corresponding Author: Guang Yang
Email address: gyang@njnu.edu.cn

Loss of tooth or enamel is widespread in multiple mammal lineages. Although several studies have been reported, the evolutionary mechanisms of tooth / enamel loss are still unclear. Most previous studies have found that some tooth-related genes have been inactivated in toothless and / or enamel-less mammals, such as *ENAM*, *ODAM*, *C4orf26*, *AMBN*, *AMTN*, *DSPP*, etc. Here, we conducted evolutionary analyses on *ACPT* playing a key role in amelogenesis, to interrogate the mechanisms. We obtained the *ACPT* sequences from 116 species, including edentulous and enamel-less mammals. The results shows that variant ORF-disrupting mutations were detected in *ACPT* coding region among nine edentulous baleen whales and three enamel-less taxa (pygmy sperm whale, aardvark, nine-banded armadillo). Furtherly, selective pressure uncovered that the selective constraints have been relaxed among all toothless and enamel-less lineages. Moreover, our results support the hypothesis that mineralized teeth were lost or degenerated in the common ancestor of crown Mysticeti through two shared single-base sites deletion in exon 4 and 5 of *ACPT* among all living baleen whales. D_N / d_S values on transitional branches were used to estimate *ACPT* inactivation records. In the case of aardvark, inactivation of *ACPT* was estimated at ~23.60-28.32 Ma, which is earlier than oldest aardvark fossil record (*Orycteropus minutus*, ~19Ma), suggesting that *ACPT* inactivation may result in degeneration or loss of enamel. Conversely, the inactivation time of *ACPT* estimated in armadillo (~10.18-11.30 Ma) is later than oldest fossil record, suggesting that inactivation of *ACPT* may result from degeneration or loss of enamel in these mammals. Our findings suggested that different mechanisms of degeneration of tooth / enamel might exist among toothless and enamel-less lineages during evolution. Our study further considered that *ACPT* is a novel gene for studying tooth evolution.

ACPT gene is inactivated in mammalian lineages that lack enamel or teeth

Yuan Mu[1], Xin Huang[1], Rui Liu[1], Yulin Gai[1], Na Liang[1], Daiqing Yin[1], Lei Shan[1], Shixia Xu[1] and Guang Yang[1]*

[1] Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China

* Corresponding author:
Guang Yang
1# Wenyuan Road, Nanjing, Jiangsu Province, 210023, China
Email address: gyang@njnu.edu.cn

Abstract

Loss of tooth or enamel is widespread in multiple mammal lineages. Although several studies have been reported, the evolutionary mechanisms of tooth / enamel loss are still unclear. Most previous studies have found that some tooth-related genes have been inactivated in toothless and / or enamel-less mammals, such as *ENAM*, *ODAM*, *C4orf26*, *AMBN*, *AMTN*, *DSPP*, etc. Here, we conducted evolutionary analyses on *ACPT* playing a key role in amelogenesis, to interrogate the mechanisms. We obtained the *ACPT* sequences from 116 species, including edentulous and enamel-less mammals. The results shows that variant ORF-disrupting mutations were detected in *ACPT* coding region among nine edentulous baleen whales and three enamel-less taxa (pygmy sperm whale, aardvark, nine-banded armadillo). Furtherly, selective pressure uncovered that the selective constraints have been relaxed among all toothless and enamel-less lineages. Moreover, our results support the hypothesis that mineralized teeth were lost or degenerated in the common ancestor of crown Mysticeti through two shared single-base sites deletion in exon 4 and 5 of *ACPT* among all living baleen whales. D_N / d_S values on transitional branches were used to estimate *ACPT* inactivation records. In the case of aardvark, inactivation of *ACPT* was estimated at ~23.60-28.32 Ma, which is earlier than oldest aardvark fossil record (*Orycteropus minutus*, ~19Ma), suggesting that *ACPT* inactivation may result in degeneration or loss of enamel. Conversely, the inactivation time of *ACPT* estimated in armadillo (~10.18-11.30 Ma) is later than oldest fossil record, suggesting that inactivation of *ACPT* may result from degeneration or loss of enamel in these mammals. Our findings suggested that different mechanisms of degeneration of tooth / enamel might exist among toothless and enamel-less lineages during evolution. Our study further considered that *ACPT* is a novel gene for studying tooth evolution.

Key works: *ACPT*, Tooth evolution, Enamel loss, Mammals, Pseudogene, Inactivation time

Introduction

Dental innovations (such as differentiated dentitions and the evolution of tri-bosphenic molar) have been regarded as the great success of mammalian evolution and adaptation (Ungar, 2010). However, in spite of their importance for animal survival, teeth have been lost independently in multiple mammalian lineages, such as baleen whales and pangolins. In addition, some lineages have lost their outer enamel of teeth, such as pygmy sperm whale and dwarf sperm whale, aardvarks and species from Xenarthra (Davitbeal et al., 2009). Tooth loss and / or enamel loss is one of the most important field for mammalian tooth evolution.

Amelogenesis imperfecta (AI) and tooth loss are the diseases that characterized by genetic defects in the formation of enamel and teeth. Multiple studies have suggested these genetic disorders are mainly caused by mutations of protein-coding genes functioned in formation of enamel and teeth (Stephanopoulos et al., 2005; Smith et al., 2017). Of these genes, three enamel matrix protein genes (EMPs, i.e., *AMELX*, *AMBN* and *ENAM*), two enamel proteases genes (*MMP20* and *KLK4*), and some other related genes (e.g., *C4orf26*, *AMTN*, *ODAM*, *ACPT*, *DSPP*) have been confirmed to be candidate genes responsible for the diseases (Crawford et al., 2007; Smith et al., 2017). The variant inactivating mutations have been detected in these genes among toothless and enamel-less mammalian lineages. However, the mechanisms of tooth loss or enamel loss are still completely unclear.

It has been reported that *ACPT* was lower expressed in testicular cancer tissues compared to normal tissues and is regulated by steroid hormones (Yousef et al., 2001). Besides, *ACPT* is also expressed in the brain and acts as a tyrosine phosphatase to modulate signals mediated by ErbB4 (Fleisig et al., 2004). But, it is interesting to note that *ACPT* is expressed in secretory-stage ameloblasts (Seymen et al., 2016), which can induce odontoblasts differentiation, mineralization of dentin, and amelogenesis (Choi et al., 2016). Furthermore, there are some increasing evidences that homozygous missense variants of *ACPT* would lead to AI (e.g., c.226C>T, p.Arg76Cys; c.746C4T, p.P249L) (Seymen et al., 2016; Smith et al., 2017). These evidences suggested that *ACPT* play an important role in amelogenesis.

All extant Mysticeti, descended from toothed ancestors, have no teeth and instead have baleen (Uhen, 2010). Paleontological evidences have shown that mineralized teeth were lost in the common ancestor of crown Mysticeti. Moreover, a transitional stage from tooth to baleen in stem mysticetes have been revealed in some taxa bearing both teeth and baleen (Deméré et al.,

2008). Although many tooth-related genes have been revealed to be inactivated in various living mysticetes (e.g., *AMBN*, *ENAM*, *AMEL*, *AMTN*, *MMP20*, *C4orf26* and *DSPP*) (Deméré et al., 2008; Meredith et al., 2009; Meredith et al., 2011; Gasse et al., 2012; Delsuc et al., 2015; Springer et al., 2016; Springer et al., 2019), only the *MMP20* are commonly inactivated across all the living baleen whales (Meredith et al., 2011). This molecular evidence is consistent with earlier studies of paleontology and anatomy.

Despite its significance in mammalian enamel maturation, very little is known about *ACPT* evolutionary trajectory, relationship and function in mammals. To address this issue, we carried out a series of evolutionary analyses on *ACPT*, aim to uncover the evolutionary pattern of *ACPT* gene among mammalian lineages.

Methods

Sequences mining and BLAST searches

The full-length coding sequences (CDS) of *ACPT* gene were extracted from the OrthoMaM v10b (http://orthomam2.mbb.univ-montp2.fr/OrthoMaM_v10b10/), ENSEMBL (http://www.ensembl.org/index.html?redirect=no) and NCBI (http://www.ensembl.org/index.html?redirect=no) databases (Table S1). *ACPT* of some whales were extracted from their Genome and SRA database of NCBI (Table S2 and S3). To further ensure the sites of inactivating mutation of toothless / enamel-less lineages, we used the CDSs of some representative placental species with well-annotated genomes (*Homo sapiens* [human], *Canis lupus familiaris* [Dog], *Bos taurus* [Cow], *Echinops telfairi* [Lesser hedgehog tenrec]) as queries including ~50bp of flanking sequence on each exon. These sequences were used as queries to BLAST against toothless / enamel-less mammals to confirm the related inactivating mutation among baleen whales.

Identification of inactivating mutations and functional sites and domains

The intact *ACPT* sequences (human, cow, tenrec) were used for identifying inactivating mutations (including mutation of initiation codons, frame-shift insertions and deletions, premature stop codons, splice sites mutation of intron / exon boundary [GT/AG], etc.). The inactivating mutation was identified based on BLAST searches against whole genomes of the relevant taxon from NCBI. The information on gene function, related key amino acid sites/domains was searched from UniProtKB/Swiss-Prot (http://www.uniprot.org/) and some references.

Alignment and phylogenetic analysis of mammalian *ACPT*

The 116 mammalian *ACPT* sequences were aligned based on their amino acid translations using online PRANK (https://www.ebi.ac.uk/goldman-srv/webprank/), and then deleted the gaps and non-homologous regions by using GBLOCK, then we corrected the multiple sequences alignment (MSA) in MEGA 7 (Kumar et al., 2016) by eye.

A gene tree was reconstructed by Mrbayes 3.2 (Ronquist et al., 2012) with a general time reversible (GTR) substitution model and rate heterogeneity modeled with a Gamma distribution, as conducted by MrModeltest version 2 using the Akaike information criterion (AIC) (Nylander, 2004). In bayesian analysis, four simultaneous runs with four chains each were run for two million generations, sampling every 1000 trees. The first 25% of these trees were discarded as burn-in when computing the consensus tree. Tracer v1.5 software was used for checking convergence among chains in Bayesian analysis. When the ESS value is higher than 200, and the average standard deviation of spilt frequencies is lower than 0.01, we think it reach convergence level.

Selection analyses

To evaluate the selective pressure of relevant branches leading to enamel-less and toothless lineages respectively, we implemented *two ratio branch model* to calculate the ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) ($\omega = d_N/d_S$) by running CodeML in PAML 4.8a package (Yang, 2007). We also recoded premature stop codons as missing data. Akaike information criterion (AIC) scores were used to select the most appropriate codon frequency model in CodeML. The *ACPT* gene tree exhibits different topological relationship compared to species tree, which may be unrelated to incomplete lineage sorting. In order to illuminate the detected signal reasonably and accurately, we used a species tree supported by some previous studies (Fig. S1).

Refer to the methods of Springer and Gatesy (Springer and Gatesy, 2018), several different branch categories were considered during selective analyses: (1) One category accounted for 'background' branches, which are lineages with intact teeth and an intact copy of *ACPT*. (2) Nine branch categories to terminal branches with unique inactivating mutations (baleen whales), which lacks teeth. (3) Three branch categories to terminal branches with unique inactivating mutations (pygmy sperm whale, nine-banded armadillo and aardvark), whose enamel has been vestigial. (4) One branch categories were assigned for stem Mysticeti where mineralized teeth

were degraded. (5) One branch categories were assigned for crown Mysticeti.

To better understand the selective pressure, a series of evolutionary models were compared in the likelihood. We first use the M0 model (Model A), which assumed that all branches in the phylogenetic tree has a common value, and compare it with the null hypothesis (Model B), which assumed that the common value in the phylogenetic tree is 1. To further understand whether the selective pressure on the lineages leading to pseudogenes was relaxed, we constructed Model C, which assumed that the branches with pseudogene had their own selection pressure ω_2, while the background branches without pseudogenization was ω_1, and then compared Model C with Model A. To further confirm whether the selective pressure on the lineages leading to pseudogenes was completely relaxed, we build the Model D, which assumed that the branches with pseudogene had their own selection pressure $\omega_2 = 1$, while the selective pressure of background branches was ω_1, and then compared Model C with Model D.

Estimation of inactivation times

To estimate when *ACPT* was inactivated in different lineages of Placentalia, the method described in Chou et al. (2003) and Zhang et al. (2010) was used. Among the branches along which the gene became pseudogenes, this method presumes that gene evolves under a selective pressure similar to that in other species until it is inactivated. Next, this gene was presumed to accumulate both nonsynonymous and synonymous mutations at an equal rate. The K_a / K_s (K) value was assessed for this entire branch. The average K_a / K_s value was just for a part of the branch, where the gene was under selection (Ks). In addition, the K_a / K_s value for the rest of part of the branch where the gene evolved neutrally ($K_n = 1$). Thus, the evolutionary time was weighted by the proportion, for which the gene was evolving under selection (T_s / T) and neutrally (T_n / T):

$$K = K_s \times T_s / T + K_n \times T_n / T$$

where T is the time since the split from the last common ancestor (LCA). By selecting the lower and upper bound of the confidence interval for the species divergence time T, which was obtained from TimeTree website (http://www.timetree.org/) to estimate a lower and upper bound for T_n as:

$$T_n = T \times (K-K_s) / (1-K_s)$$

which provides an estimate of how long the *ACPT* gene has been evolving neutrally.

Results

Characterization of *ACPT* sequence

120 sequences were obtained in this study. Due to the poor quality and low coverage of sequences among three pangolins (*Manis javanica*, *M. javanica*, *Phataginus tricuspis*) and one sloth (*Choloepus hoffmanni*), they were not used for subsequent analysis. However, some inactivating mutations (most of them are indels) were found in these sequences (Fig. S2). The complete protein-coding sequence of *ACPT* in 116 taxa were used for alignment by PRANK. Interestingly, one or more inactivating mutations (frame-shift mutation, initial codon mutation, premature stop codons, splice site mutations, etc.) were detected in another placental taxa without teeth or without enamel. (Fig. 1, Table S4, Fig. S3). For example, among toothless baleen whales, the initial codon mutation (n. ATG→GTG, p. M→V) was found in *Balaenoptera borealis*, *B. physalus*, *B. musculus*, *Eschrichtius robustus*, *Eubalaena glacialis*. Meanwhile, premature stop codons were found in *B. acutorostrata* and *B. bonaerensis*, frameshift indels were also found in baleen whales. Interestingly, two shared single-base site deletion was found on exon 4 and 5 of *ACPT* among all living baleen whales (Fig. 1, Fig. S3). The splice site mutations were detected in *B. acutorostrata*, *Eubalaena japonica* and *Megaptera novaeangliae* (Table S4). Whilst, the premature stop codons were found in enamel-less *D. novemcinctus* and . *afer*. Besides, frameshift indels were found in enamel-less *Kogia breviceps*.

Except for the species mentioned above, *ACPT* gene in other species whose teeth are intact were found to be activated. Nevertheless, some crucial amino acids mutation was found in toothed species, such as site 76 has been mutated (R76C) in *Neophocaena asiaeorientalis*.

Reconstruction of *ACPT* gene tree

We recovered the *ACPT* gene tree with well-supported values by using Mrbayes method (Fig. 2). In this gene tree, most of orders have been well reconstructed, and have high support rate, e.g., Cetartiodactyla, Perissodactyla, Eulipotyphla, Carnivora, Chiroptera etc. In addition, phylogenetic relationships of higher levels have also been well reconstructed, such as Laurasiatheria, Euarchontoglires, Boreoeutheria and Afrotheria. In this gene tree, bayesian posterior probability (PP) values of nearly 70% nodes are generally greater than 0.70. However, the relationship between some order level were relatively chaotic, such as Lagomorpha didn't cluster with Rodentia, but as the sister group of Primate; Chiroptera and Carnivora clustered together first, and then they became sister group of Perissodactyla.

Evolutionary analyses among toothless and enamel-less mammals

We carried out the PAML analysis to detect the selective pressure of toothless / enamel-less lineages, and found the selective pressure of these toothless / enamel-less lineages (including ancestral nodes, terminal branches and even the whole toothless / enamel-less group) was significantly higher than that of background branches. For example, the terminal branch of *B. physalus*: ω_1=0.116, ω_2=1.883; the terminal branch of *M. novaeangliae*: ω_1=0.116, ω_2=0.641; the terminal branch of *E. robustus*: ω_1=0.116, ω_2=2.688; the terminal branch of *E. glacialis*: ω_1=0.116, ω_2=0.503. A similar tendency was found in the terminal branches of other baleen whales, and further model comparison shows that the selective pressure of these branches had been completely relaxed. Whilst, much higher selective pressure was detected in the ancestral branch of stem mysticeti (ω_1=0.120, ω_2=0.436), even the clade of crown mysticeti (ω_1=0.116, ω_2=0.522). Meanwhile, higher selective pressure was detected among enamel-less lineages, such as the terminal branch of *D. novemcinctus* (ω_1=0.116, ω_2=0.206), the terminal branch of *O. afer* (ω_1=0.116, ω_2=0.414), and the terminal branch of *K. breviceps* (ω_1=0.116, ω_2=0.581). And the selective pressure of these branches had been completely relaxed, except for the terminal branch of *K. breviceps* (Table S5).

***ACPT* inactivation dates**

Estimates of inactivation times for *ACPT* based on d_N / d_S ratios and equations in Sharma et al. (Sharma et al., 2018). The mean estimate for the inactivating time of *ACPT* on the branch of *K. breviceps*, *D. novemcinctus* and *O. afer* is 12.20-15.52Ma, 10.18-11.30Ma and 23.60-28.32Ma, respectively (Fig. 3). The mean estimate for the inactivation of *ACPT* on the Mysticeti clade is 14.05-16.30Ma.

Discussion

***ACPT* is a novel candidate gene for studying mammalian tooth loss and enamel loss**

The well-conserved gene structure in extant species indicates that this organization and arrangement might be present in the last common mammalian ancestor, which represented the vital function for organisms (Madsen, 2009). In our study, the number of *ACPT* exons are 11 in placental mammals, which encode 427 amino acids (human ACPT sequence as the reference sequence). Our study collected that four residues (191N, 269N, 330N and 339N) of the extracellular region were for glycosylation, two residues (41H and 289D) directly involved in catalysis (from UniProt database). In addition, mutation in seven residues were reported that

were responsible for AI (Seymen et al., 2016; Smith et al., 2017) (Fig. S4). Besides, there are three disulfide bond regions, namely, site 159 to 378, site 214 to 312, site 353 to 357. In fact, we detected not only teratogenic mutations but also inactivated mutations in these functional sites and domains. For example, enamel in finless porpoise were degenerated (Ishiyama, 1987), mutation in site 76 (R→C) was found in *N. asiaeorientalis*. Previous research has confirmed that site 76 mutated into Cys (C) in human ACPT would lead to hypoplastic AI (Seymen et al., 2016), from which this result further supported that teeth in finless porpoise were degenerated in molecular level. Of cause, most obvious characteristics of *ACPT* is that different types of inactivating mutations were found in toothless and enamel-less mammals, e.g., baleen whales, pangolins, sloths and so on (Fig. S2, Fig. S3). Therefore, *ACPT* could be a candidate gene for AI and studying mammalian tooth loss and enamel loss.

Degeneration or loss of mineralized teeth in LCA of Mysticeti

Fossil evidence shows that the earliest ancestors of baleen whales possessed complete dentitions without baleen (such as *Janjucetus* and *Mammalodon*), and then evolved the baleen with teeth (such as *Aetiocetus*), until the lineages only baleen existed (e.g., *Eomysticetus* and *Micromysticetus*) (Fitzgerald, 2006; Fitzgerald, 2010; Meredith et al., 2011). However, the fact is all living baleen whales lack teeth and instead baleen (Uhen, 2010). This implied that that mineralized teeth were lost or degenerated gradually in the common ancestors of all modern baleen whales (Boessenecker and Fordyce, 2015). In addition, the successive steps of vestigial tooth development was found in the fetal period of living baleen whales (Davit-Béal et al., 2009; Thewissen, 2018), which was also confirmed by genetic evidence. Molecular sequences of some specific genes, such as *AMBN*, *ENAM*, *AMELX*, *AMTN*, *C4orf26* and *ODAM*, contain different types of inactivating mutations (e.g., stop codons, frameshift mutations, splice site mutations, etc.) in various mysticete species (Deméré et al., 2008; Meredith et al., 2009; Alhashimi et al., 2010; Gasse et al., 2012; Meredith et al., 2013; Delsuc et al., 2015; Springer et al., 2019), which is consistent with loss-of-teeth in this group. But none of the inactivating mutations are shared by all living mysticetes species. Meredith et al. found a common insertion of CHR-2 SINE retroposon in *MMP20* gene among all living baleen whales (Meredith et al., 2011). Previous study has been confirmed that mutations or deletions of *MMP20* gene would result in thin and brittle enamel layer (Caterina et al., 2002). Based on this result, they confirmed the hypothesis that mineralized teeth were lost or degenerated in the common ancestor of crown Mysticeti in the

molecular level.

In this research, we also identified different inactivating mutations was detected among all mysticete species in *ACPT* gene, among which two shared single-base sites deletion were found on exon 4 and 5 of *ACPT* among all living baleen whales, which result in loss of function. Some studies have confirmed that *ACPT* gene is responsible for the development of enamel, and mutations can also lead to *amelogenesis imperfecta* (Choi et al., 2016; Seymen et al., 2016; Smith et al., 2017). Similar to the result of Meredith et al. (2011), our study supported the hypothesis that mineralized teeth were lost or degenerated in the common ancestor of all extant baleen whales.

Is inactivation of *ACPT* neutral or adaptive?

The degeneration and / or loss of some morphological structures (such as limbs, teeth, and eyes, etc.) is a complex process that may result from the relaxation of the negative selection (neutral evolution), adaptive evolution (direct natural / positive selection to conserve energy and / or eliminate the disadvantageous effects of morphological structure), and / or gene pleiotropy (indirect selection on another traits) (Wang et al., 2006; Zhang, 2008; Krishnan and Rohner, 2017). In some conditions, evolutionary change also results from differences in the reproductive success of individuals with different genotypes (Olson,1999). Sharma et al. (2018) revealed that evolutionary gene losses are not only a consequence, but may also be causally involved in phenotypic adaptations. By estimating the inactivation time of pseudogenes, and comparing with oldest fossil records, we might be able to speculate whether gene inactivation is due to the adaptive or neutral selection after the loss of phenotype.

The record of enamel-degenerated armadillo fossil is significantly earlier than the estimated time of *ACPT* inactivation (10.18-11.30Ma) (Ciancio et al., 2014), which suggested gene loss as a consequence of adaptation is likely the result of the relaxation of the negative selection. The results further supported the previous study (Sharma et al., 2018). Besides, during the tooth evolution, some enamel-related genes (e.g., *ODAM*, *ENAM*, *AMBN*) also have gone through the similar evolutionary trajectory. By integrating different results from different methods, we may better understand the evolution of teeth and enamel. The inactivation time of *ENAM* (~45.5Ma) and *ODAM* (~40.43 Ma, range 36.38-45.45Ma) is much earlier than inactivation date for *ACPT* in armadillo (Springer et al., 2019). *ACPT* inactivation is later than the fossil record, conversely, the inactivation time of *ENAM* is relatively earlier than the fossil record, which implied the

various mechanisms of enamel loss in armadillo. Here, the inactivation of *ENAM* gene might be the causes of degeneration / loss of tooth enamel in armadillos, *ACPT* inactivation might be the consequence of enamel loss.

For *O. afer*, even the inactivation date for *ACPT* (23.60-28.32Ma) is relatively younger than inactivation dates for *ENAM* (28.8-35.3Ma) and *ODAM* (~30.7Ma) in *O. afer* (Meredith et al., 2009; Springer et al., 2019). However, the estimated inactivation times by *ACPT*, *ODAM* and *ENAM* gene markers are all earlier than the oldest fossil record of aardvark (*O. minutus*, ~19Ma) (Patterson, 1975). It should be suggested that gene loss may be the reason, not the consequence, for degeneration and / or loss of enamel. Moreover, due to the difference of species number, sequences quality and topological structure of species tree, the result of *ACPT* inactivation time is different from the result of Sharma et al. (Sharma et al., 2018).

Cetacean includes both toothless Mysticeti and enamel-less *Kogia*. Relaxation of selective pressure was detected in both crown and stem Mysticeti (Table S5), which is consistent with the archaic toothless mysticete, namely, all stem Mysticeti were toothless. For example, *Eomysticetus whitmorei*, an edentulous species, was the geologically oldest mysticete (Deméré et al., 2008). Molecular evidence shows *ACPT* has been lost its function in LCA of Mysticeti. However, the inactivation time of *ACPT* in Mysticeti is 14.05-16.30Ma, which is much younger than the toothless mysticete (~30Ma) and the split of Mysticeti (~25.9Ma). Obviously, this is not consistent with the facts. It might be associated with relatively lower rates of frameshift accumulation during evolution of mysticete pseudogenes and long lifespan of mysticete (Meredith et al., 2009; Meredith et al., 2011). Whether adaptive or neutral, the shared single-base site deletion in *ACPT* fills an important gap in our understanding of the macroevolutionary transition leading from the LCA of crown Cetacean to the LCA of crown Mysticeti. Stem physeteroids (sperm whales) are known from the Miocene and had teeth with enamel (Bianucci and Landini, 2010). Our results provide support for loss of the intact *ACPT* in *K. breviceps*. ACPT was reported that play key roles in amelogenesis and differentiation of odontoblasts (Choi et al., 2016; Seymen et al., 2016; Smith et al., 2017). Our result is in line with the enamel-less morphological structure in *K. breviceps*.

CONCLUSIONS

We detected the different types of inactivated mutation in *ACPT*. Furthermore, selective pressure uncovered that the selective constraints have been relaxed among all toothless and enamel-less

lineages. In addition, our results supported the hypothesis that mineralized teeth were lost or degenerated in the common ancestor of crown Mysticeti through two shared single-base sites deletion in exon 4 and 5 of *ACPT* among all living baleen whales. Together with our evidence, *ACPT* might be a good marker to research the mechanism of tooth loss. By comparing the molecular time with the fossil time, we found there might be different mechanisms of degeneration of tooth / among toothless and enamel-less lineages during evolution, which is needed further researches.

Acknowledgements

We thank members of the Jiangsu Key Laboratory for Biodiversity and Biotechnology, Nanjing Normal University, for their contributions to this paper. The authors thank Mr. Xinrong Xu, Dr. Di Sun and Dr. Ran Tian, Dr. Zepeng Zhang, Dr. Simin Chai and Dr. Zhenpeng Yu for some helpful discussion. Special thanks to Dr. Zhengfei Wang for technical supports.

References

Alhashimi N, Lafont AG, Delgado S, Kawasaki K, Sire JY. 2010. The enamelin genes in lizard, crocodile, and frog and the pseudogene in the chicken provide new insights on enamelin evolution in tetrapods. *Molecular Biology and Evolution* **27**:2078-2094.

Bianucci G, Landini W. 2010. Killer sperm whale: a new basal physeteroid (Mammalia, Cetacea) from the Late Miocene of Italy. *Zoological Journal of the Linnean Society* **148**:103-131.

Boessenecker RW, Fordyce RE. 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.

Caterina JJ, Skobe Z, Shi J, Ding Y, Simmer JP, Birkedalhansen H, Bartlett JD. 2002. Enamelysin (Matrix Metalloproteinase 20) deficient mice display an *amelogenesis imperfecta* phenotype. *Journal of Biological Chemistry* **277**:49598-49604.

Choi H, Kim TH, Yun CY, Kim JW, Cho ES. 2016. Testicular acid phosphatase induces odontoblast differentiation and mineralization. *Cell and Tissue Research* **364**:95-103.

Chou HH, Hayakawa T, Diaz S, Krings M, Indriati E, Leakey M, Paabo S, Satta Y, Takahata N, Varki A. 2002. Inactivation of CMP-*N*-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **99**:11736–11741.

Ciancio MR, Vieytes EC, Carlini AA. 2014. When xenarthrans had enamel: insights on the evolution of their hypsodonty and paleontological support for independent evolution in armadillos. *Naturwissenschaften* **101**:715-725.

Crawford PJ, Aldred M, Bloch-Zupan A. 2007. *Amelogenesis imperfecta*. *Orphanet Journal of Rare Diseases* **2**:17.

Davit-Béal T, Tucker AS, Sire JY. 2009. Loss of teeth and enamel in tetrapods: fossil record, genetic data and morphological adaptations. *Journal of Anatomy* **214**:477-501.

Delsuc F, Gasse B, Sire JY. 2015. Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for *amelogenesis imperfecta*. *BMC Evolutionary Biology* **15**:148.

Deméré TA, Mcgowen MR, Berta A, Gatesy J. 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Systematic*

Biology **57**:15-37.

Fitzgerald EM. 2006. A bizarre new toothed mysticete (Cetacea) from Australia and the early evolution of baleen whales. *Proceedings of the Royal Society B: Biological Sciences* **273**:2955-2963.

Fitzgerald EMG. 2010. The morphology and systematics of Mammalodon colliveri (Cetacea: Mysticeti), a toothed mysticete from the Oligocene of Australia. *Zoological Journal of the Linnean Society* **158**:367-476.

Fleisig H, El-Husseini ED, Vincent SR. 2004. Regulation of ErbB4 phosphorylation and cleavage by a novel histidine acid phosphatase. *Neuroscience* **127**:91-100.

Gasse B, Silvent J, Sire JY. 2012. Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI. *Journal of Dental Research* **91**:1085.

Ishiyama M. 1987. Enamel structure in odontocete whales. *Scanning microscopy* **1**:1071-1079.

Krishnan J, Rohner N. 2017. Cavefish and the basis for eye loss. *Philosophical Transactions of The Royal Society B Biological Sciences* **372**:20150487.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870-1874.

Madsen O. 2009. *Mammals. In: The time tree of life*. New York, NY: Oxford University Press.

Meredith RW, Gatesy J, Cheng J, Springer MS. 2011. Pseudogenization of the tooth gene enamelysin (*MMP20*) in the common ancestor of extant baleen whales. *Proceedings of the Royal Society B: Biological Sciences* **278**:993-1002.

Meredith RW, Gatesy J, Murphy WJ, Ryder OA, Springer MS. 2009. Molecular decay of the tooth gene enamelin (*ENAM*) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genetics* **5**:e1000634.

Meredith RW, Gatesy J, Springer MS. 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evolutionary Biology* **13**:20.

Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Olson MV. 1999. When Less Is More: Gene Loss as an Engine of Evolutionary Change[J]. *American Journal of Human Genetics* **64**:18-23.

Patterson B. 1975. The fossil aardvarks (Mammalia: Tubulidentata). *Bulletin of the Museum of Comparative Zoology* **147**:185-237.

Ronquist F, Teslenko M, Der Mark PV, Ayres DL, Darling AE, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology* **61**:539-542.

Seymen F, Kim YJ, Lee YJ, Kang J, Kim TH, Choi H, Koruyucu M, Kasimoglu Y, Tuna EB, Gencay K. 2016. Recessive mutations in *ACPT*, encoding testicular acid phosphatase, cause hypoplastic *amelogenesis imperfecta*. *The American Journal of Human Genetics* **99**:1199-1205.

Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications* **9**:1215.

Smith CE, Whitehouse LL, Poulter JA, Brookes SJ, Day PF, Soldani F, Kirkham J, Inglehearn CF, Mighell AJ. 2017. Defects in the acid phosphatase ACPT cause recessive hypoplastic *amelogenesis imperfecta*. *European Journal of Human Genetics* **25**:1015-1019.

Smith CEL, Poulter JA, Antanaviciute A, Kirkham J, Brookes SJ, Inglehearn CF, Mighell AJ. 2017. *Amelogenesis imperfecta*: genes, proteins, and pathways. *Frontiers in Physiology* **8**:435.

Springer M, Emerling C, Gatesy J, Randall J, Collin M, Hecker N, Hiller M, Delsuc F. 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. *BMC Evolutionary Biology* **19**:31.

Springer MS, Gatesy J. 2018. Evolution of the *MC5R* gene in placental mammals with evidence for its inactivation in multiple lineages that lack sebaceous glands. *Molecular Phylogenetics and Evolution* **120**:364-374.

Springer MS, Starrett J, Morin PA, Lanzetti A, Hayashi C, Gatesy J. 2016. Inactivation of *C4orf26* in toothless placental mammals. *Molecular Phylogenetics and Evolution* **95**:34-45.

Stephanopoulos G, Garefalaki ME, Lyroudia K. 2005. Genes and related proteins involved in *amelogenesis imperfecta*. *Journal of Dental Research* **84**:1117.

Thewissen JGM. 2018. Highlights of cetacean embryology. *Aquatic Mammals* **44**:591-602.

Uhen MD. 2010. The Origin(s) of Whales. *Annual Review of Earth and Planetary Sciences* **38**:189-219.

Ungar PS. 2010. *Mammal teeth: origin, evolution, and diversity*: Johns Hopkins University Press, Baltimore, MD.

Wang X, Grus WE, Zhang J. 2006. Gene losses during human origins. *PLoS Biology* **4**:e52.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**:1586-1591.

Yousef GM, Diamandis M, Jung K, Diamandis EP. 2001. Molecular cloning of a novel human acid phosphatase gene (*ACPT*) that is highly expressed in the testis. *Genomics* **74**:385-395.

Zhang JZ. 2008. Positive selection, not negative selection, in the pseudogenization of rcsA in Yersinia pestis. *Proceedings of the National Academy of Sciences* **105**:E69.

Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. 2010. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biology* **11**:R26.

Figure 1

The inactivating mutation of *ACPT* gene in toothless/enamel-less mammals.

ICM, initiation codon mutation; Del, deletion; Ins, insertion; PSC, premature stop codon. The images are from the PHYLOPIC database: http://phylopic.org/.

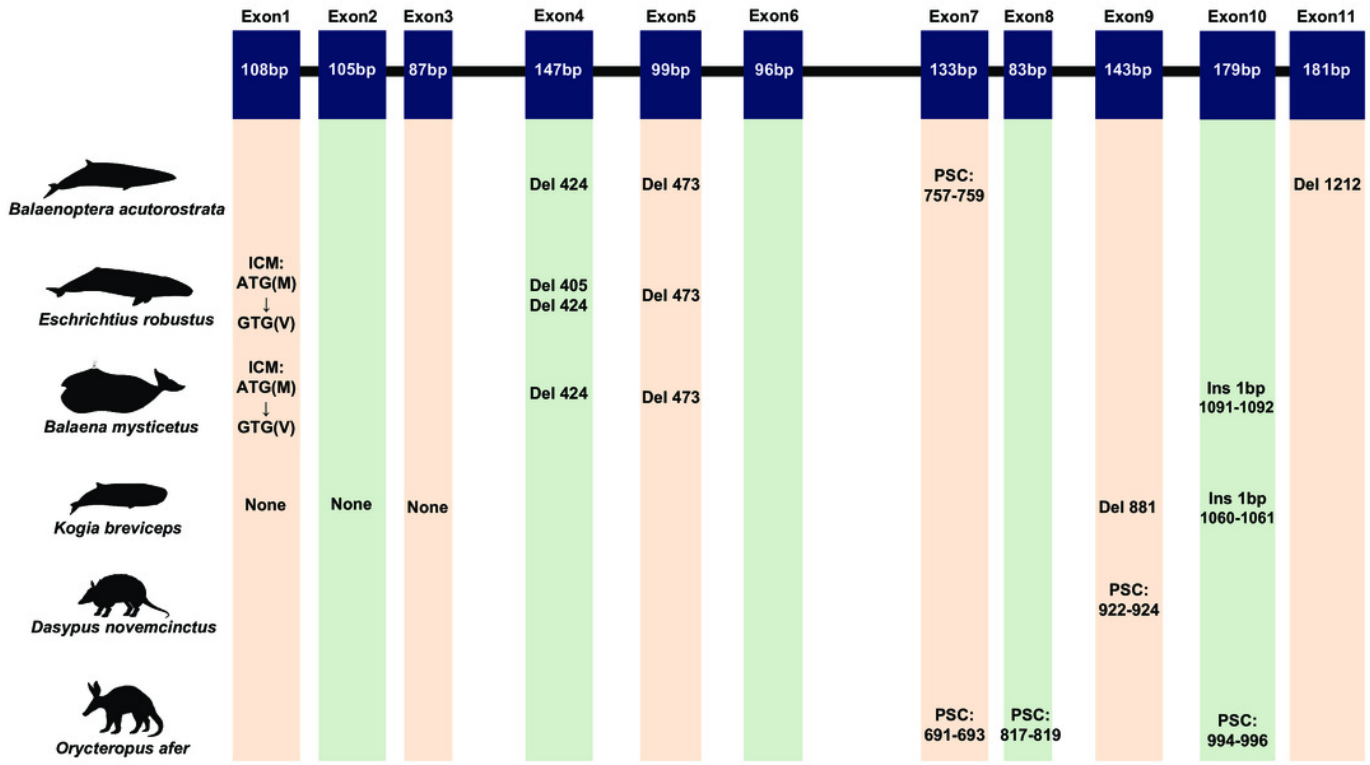


Figure 2

The BI phylogenetic relationship of mammalian *ACPT* gene used in this study.

Nucleotide optimal substitution model: GTR+GAMMA; green box indicates toothless taxa, red boxes indicate enamel-less taxa. The images are from the PHYLOPIC database: http://phylopic.org/.

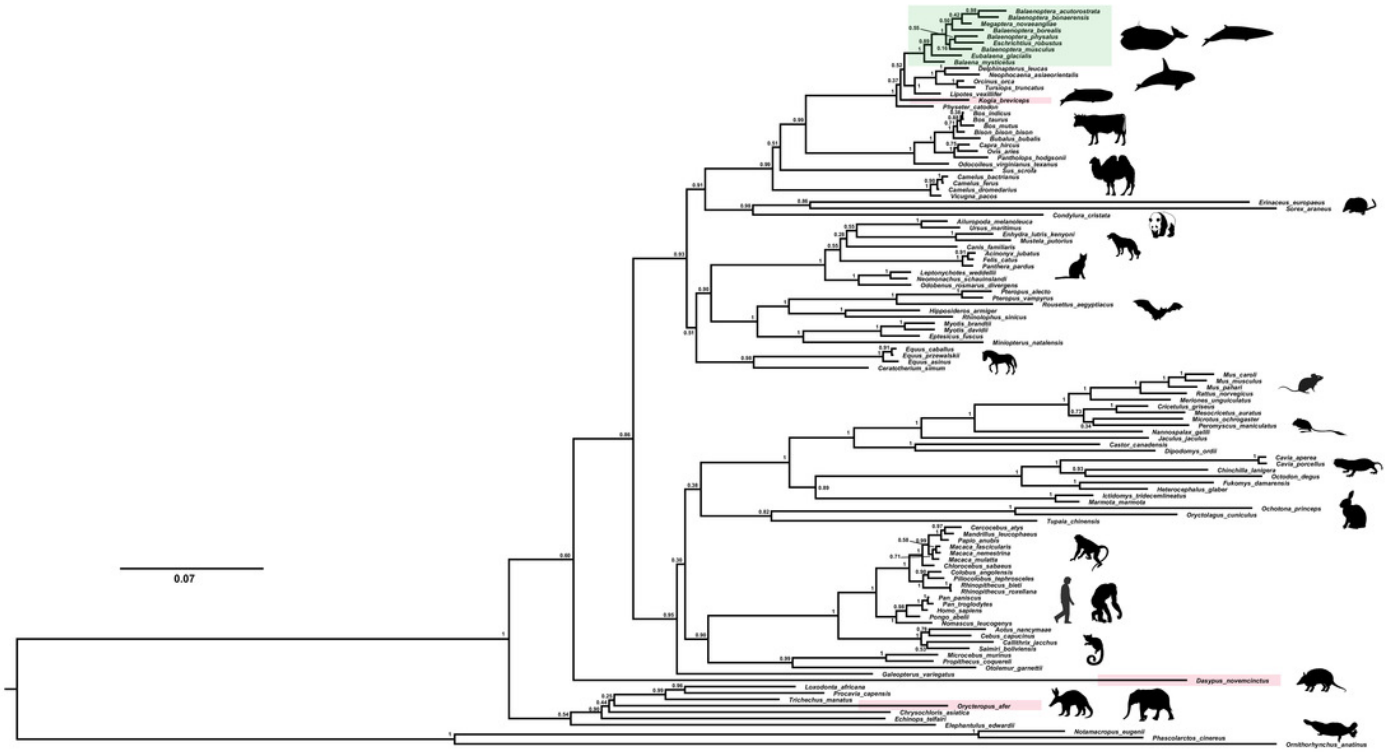
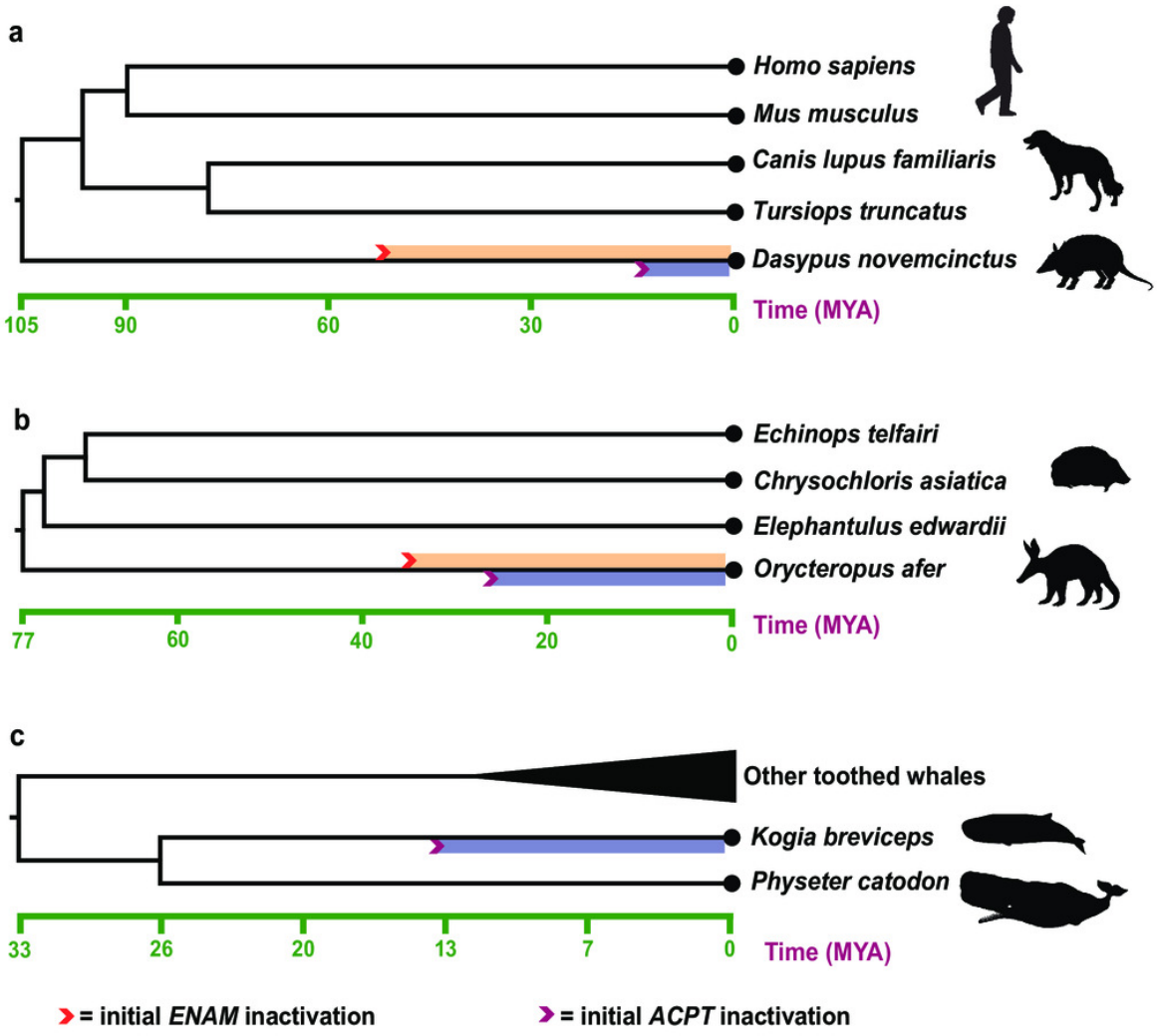


Figure 3

Estimated inactivation times of *ACPT* versus *ENAM*.

(a) *Dasypus novemcinctus* (nine-banded armadillo), (b) *Orycteropus afer* (aardvark), (c) *Kogia breviceps* (pygmy spermwhale). The inactivation times of *ENAM* is from (Meredith et al., 2009; Springer et al., 2019). The images are from the PHYLOPIC database: http://phylopic.org/.


a
Homo sapiens
Mus musculus
Canis lupus familiaris
Tursiops truncatus
Dasypus novemcinctus
105
90
60
30
0
Time (MYA)
b
Echinops telfairi
Chrysochloris asiatica
Elephantulus edwardii
Orycteropus afer
77
60
40
20
0
Time (MYA)
c
Other toothed whales
Kogia breviceps
Physeter catodon
33
26
20
13
7
0
Time (MYA)
= initial ENAM inactivation
= initial ACPT inactivation