

参赛队员姓名: 杨奕非、龚靖尧、戴川滢

中学: 南京外国语学校

省份: 江苏省

国家/地区: 中国

指导教师姓名: 章熙东、徐士霞

论文题目: **Feeding on animals or plants: evolution of enamel-related genes and its relationship with food habits in mammals**

# **Feeding on animals or plants: evolution of enamel-related genes and its relationship with food habits in mammals**

**Yifei Yang Jingyao Gong Chuanyan Dai**

**(2016 DP I-1, Nanjing Foreign Language School, Nanjing 210000, China)**

## **Abstract**

### **Background**

Mammalian lineages with different diets have different tooth phenotypes, yet the genetic basis of tooth variation remains poorly understood. To determine whether the evolution of tooth-related genes in different mammalian lineages are related to feeding habits, we retrieved genic sequences of seven enamel-related genes (*AMELX*, *AMBN*, *ENAM*, *AMTN*, *ODAM*, *KLK4* and *MMP20*) in mammals with different feeding habits.

### **Results**

We found higher levels of positive selection on enamel-related genes in herbivorous lineages than carnivorous lineages. There was significant relaxation of selective constraints in the terminal and/or ancestral branches of species without enamel or teeth, which may be correlated with their unique feeding habits. In addition, evolutionary rates of enamel-related genes were higher in herbivores than carnivores, and seven parallel substitution sites were identified among different herbivorous lineages. Interestingly, we found a significant association between evolution of enamel-related genes and average enamel thickness in primates.

### **Conclusion**

Genes involved in enamel composition showed stronger positive selection in plant-eating lineages including herbivores and most of omnivores, which is congruent with the better developed enamel in these animals that could enhance their ability to eat plants while protecting teeth from being eroded by fibers. This is further corroborated

by the significant association between evolution of enamel-related genes and average enamel thickness in primates. In summary, the evolution of enamel-related genes may provide insights into the evolution of different feeding habits in mammals.

### **Key Words**

Mammals, Enamel-related gene, Adaptive evolution, Positive selection, Feeding habits

### **Introduction**

Teeth are responsible for cutting, grinding and crushing food, as well as to attack or defend against other animals, playing an important role in the survival and diversity of vertebrates [1]. Teeth originated about 460 million years ago (Ma) [2] and evolved into diverse morphologies in different vertebrates. Mammals have evolved complex tooth patterns, such as the differentiation of incisor, canine, premolar, and molar teeth within an individual, and phenotypic variation among different lineages. The diverse shapes of teeth suggest that they may contribute to different foraging strategies or feeding habits, such as herbivorous, omnivorous and carnivorous diets [3][4].

Most mammalian teeth are composed of two main components, enamel and dentine. Compared with dentine, enamel is a heavily mineralized and hard [5], making teeth strong enough to feed on hard foods, such as plants with high fiber content. Herbivores often have thicker layer of enamel compared to carnivores, and considered to play a role in different dietary adaptations in mammals. The association between tooth morphology and food habits is also observed in species that have lost teeth (baleen whales, pangolins and Monotreme) or enamel (sloth, armadillo and armadillo), likely due to the specialized feeding habits and strategies [6, 7].

Although the association between tooth structure and feeding habits in mammals have been suggested, the molecular evolution of enamel-related traits have not well explored so far. In this study, seven enamel-related genes that are associated with amelogenesis and odontogenesis (*AMELX*, *AMBN*, *ENAM*, *AMTN*, *ODAM*, *KLK4*, *MMP20*) (Table S1) were chosen from representative mammalian species with different

feeding habitats. We used bioinformatic analyses to test whether evolutionary changes in enamel genes are associated with feeding habits in mammals, and provide insights into molecular evolution underlying dietary adaptation in mammals.

## Material and methods

### DNA sequence screen and alignment

The full-length codon DNA sequences (CDS) of seven enamel-related genes were extracted from the OrthoMaM v9 [8], Ensembl [9] and NCBI [10] databases. CDS of bowhead whale (*Balaena mysticetus*) and pangolins [Malayan pangolin (*Manis javanica*) and Chinese pangolin (*Manis pentadactyla*)] were obtained from The Bowhead Whale Genome Resource (<http://pangolin-genome.um.edu.my/index.php/home/main>) and PGD database (<http://www.bowhead-whale.org/>), respectively. For some species, like African wild dog (*Lycaon pictus*) and enamelless/toothless species, we obtained sequences of enamel-related genes through blastn searches of mammalian genomes against the well annotated human (*Homo sapiens*) and cow (*Bos taurus*) enamel-related gene sequences as queries. The CDS were first translated into putative amino acid sequences, and then were aligned using Muscle [11] and MEGA7.0 [12], and alignment was confirmed visually.

### Analysis of selective pressure

Selective pressure were calculated using the ratio of non-synonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitutions ( $d_N/d_S$  or  $\omega$ ) in CODEML of PAML 4.7a [13], where  $\omega < 1$ ,  $\omega = 1$  and  $\omega > 1$  indicates the purifying selection, neutral selection and positive selection, respectively. We performed the evolutionary analyses with two datasets: the complete dataset of all mammalian sequences (dataset I), and a reduced dataset of functional mammalian CDS (dataset II).

Two methods were utilized to detect signatures of positive selection in these genes: 1) *Branch models* allow the  $\omega$  ratio to vary among branches in the phylogeny, and are used to detect selective pressures on particular lineages under different models (Model =

0,1,2) [14, 15]. *One ratio model* (Model A, all branches have one  $\omega$ ; Model B, all branches have one  $\omega$  and  $\omega = 1$ ) assigns the same  $\omega$  ratio for all branches and the *free ratio model* model can accommodate different  $\omega$  in each branch (model = 1). 2) To further test if the positively selected sites are restricted to specific dietary lineages, we used the *branch-sites model* to detect positive selection in lineages with different diets. All the positively selected sites were identified by using Bayes Empirical Bayes (BEB) analysis [13] with posterior probabilities (PP)  $\geq 0.80$ .

### **Identification of convergent sites among distantly related mammals with similar diets**

Parallel/convergent sites among mammals with similar diets were identified according to methods previously described [16]. The ancestral amino acid sequences of each genes were reconstructed using CODEML in PAML [13]. Then, we searched for convergent amino acid substitutions from the ancestor branches along paraphyletic lineages of species that are herbivorous and carnivorous. The software CONVERG 2 [17] was used to test whether the observed convergent amino acid substitutions in focal branches were fixed randomly or by natural selection.

### **Association analysis between root-to-tip $\omega$ and tooth phenotype in primates**

In order to test whether there is a potential relationship between the evolutionary rate of enamel-related genes and tooth with relative enamel thickness of lower M1 (obtained from Shellis et al. [18]), we examined the association in the primate dataset using previously described methods by Montgomery et al. [19]. *Two ratio model* was used to calculate the average  $d_N/d_S$  ratios from the ancestral species to each terminal species (root-to-tip  $\omega$ ), which is more suitable for regressions against phenotypic data from extant species. We then used PGLS (Phylogenetic Generalized Least Squares) analysis in the cape package of R to analyze continuous data that has been applied to estimating adaptive optima [20] and estimating the relationships among traits. A parameter, lambda( $\lambda$ ) value, was estimated by maximum likelihood method to quantitatively

measure the phylogenetic association level: from zero (no phylogenetic signal) to one (significantly phylogenetic signal).

## Results

We found that the majority of enamel-associated genes in this study were intact, however some genes had become pseudogenes in species with enamel-capped teeth, such as *ODAM* in galago (*Otolemur garnettii*) and in toothed whales, such as bottlenose dolphin (*Tursiops truncatus*), killer whale (*Orcinus orca*), baiji (*Lipotes vexillifer*) and finless porpoise (*N. phocaenoides*). In particular, edentulous species, such as pangolins, baleen whales and platypus, commonly had premature stop codons, indels and/or splice sites mutations in enamel-related genes (*AMELX*, *AMBN*, *ENAM*, *AMTN*, *ODAM*, *KLK4* and *MMP20*), suggesting pseudogenization.

### Detection of selective pressure

The detection of dataset I (all mammals) with *branch model* showed that *free ratio model* was significantly better than *one ratio model*, suggesting that divergent selective pressure operate on different branches (Table 1). Enamel-related genes in herbivorous lineages often had branches with  $\omega$  ratios greater than one. For example, of the 33, 29, 26 branches for herbivorous, carnivorous and omnivorous lineages within the phylogenetic tree (Fig. S1 a-f), the proportion of positively selected branches in herbivorous lineages were 20%, 8.6%, 22.9% for *AMELX*, *ENAM* and *ODAM* respectively, which was higher than 10.3%, 0, 10.3% in carnivorous lineages and 3.8%, 23.1%, 9.1% in omnivorous lineages, except for the higher positively selected branch percentage of 23.1% in *ENAM* of omnivorous lineage (Fig. 1).

**Table 1** Likelihood and omega values estimated under the *two ratio branch model* of selective pressures on enamel-related genes

Models and special branches	$\omega$	$-\ln L$	np	Models comparison	$2\Delta(\ln L)$	P-value
<i>AMBN</i>						
A. All branches have one $\omega$	0.459	22143.515	124			
B. All branches have one $\omega = 1$	1	22429.326	123	B vs A	571.622	0
C. The terminal branch of <i>Orycteropus afer afer</i> with pseudogenized <i>AMBN</i> has $\omega_2$ , others have $\omega_1$	$\omega_1=0.454$ $\omega_2=0.894$	22140.403	125	A vs C	6.224	0.013
D. The terminal branch of <i>Orycteropus afer afer</i> with pseudogenized <i>AMBN</i> has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.454$ $\omega_2=1$	22140.481	124	D vs C	0.156	0.693
E. The ancestral branch of Pholidota has $\omega_2$ , others have $\omega_1$	$\omega_1=0.450$ , $\omega_2=1.009$	22136.206	125	A vs E	14.618	<0.001
F. The ancestral branch of Pholidota has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.450$ $\omega_2=1$	22136.207	124	F vs E	0.002	0.964
<i>ENAM</i>						
A. All branches have one $\omega$	0.472	64566.977	128			
B. All branches have one $\omega = 1$	1	65277.191	127	B vs A	1420.428	0
C. The terminal branch of <i>Manis javanica</i> with pseudogenized <i>ENAM</i> has $\omega_2$ , others have $\omega_1$	$\omega_1=0.470$ $\omega_2=1.034$	64563.421	129	A vs C	7.112	0.008
D. The terminal branch of <i>Manis javanica</i> with pseudogenized <i>ENAM</i> has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.470$ $\omega_2=1$	64563.426	128	D vs C	0.010	0.920
E. The ancestral branch of Pholidota has $\omega_2$ , others have $\omega_1$	$\omega_1=0.468$ $\omega_2=0.806$	64561.142	129	A vs E	11.670	<0.001
F. The ancestral branch of Pholidota has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.468$ $\omega_2=1$	64561.979	128	F vs E	1.674	0.196
G. The terminal branch of <i>Choloepus hoffmanni</i> with	$\omega_1=0.469$	64562.172	129	A vs G	9.610	0.002

Models and special branches	$\omega$	$-\ln L$	np	Models comparison	$2\Delta(\ln L)$	P-value
pseudogenized <i>ENAM</i> has $\omega_2$ , others have $\omega_1$	$\omega_2=0.914$					
H. The terminal branch of <i>Choloepus hoffmanni</i> with pseudogenized <i>ENAM</i> has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.469$ $\omega_2=1$	64562.249	128	H vs G	0.154	0.695
I. The terminal branch of <i>Dasypus novemcinctus</i> with pseudogenized <i>ENAM</i> has $\omega_2$ , others have $\omega_1$	$\omega_1=0.468$ $\omega_2=0.743$	64562.396	129	A vs I	9.162	0.002
J. The terminal branch of <i>Dasypus novemcinctus</i> with pseudogenized <i>ENAM</i> has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.467$ $\omega_2=1$	64564.069	128	J vs I	3.346	0.067
<i>MMP20</i>						
A. All branches have one $\omega$	0.180	20392.241	128			
B. All branches have one $\omega = 1$	1	21668.035	127	B vs A	2551.588	0
C. The terminal branch of <i>Manis javanica</i> with pseudogenized <i>MMP20</i> has $\omega_2$ , others have $\omega_1$	$\omega_1=0.179$ $\omega_2=0.864$	20388.990	129	A vs C	6.502	0.011
D. The terminal branch of <i>Manis javanica</i> with pseudogenized <i>MMP20</i> has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.179$ $\omega_2=1$	20389.011	128	D vs C	0.043	0.836
E. The ancestral branch of Pholidota has $\omega_2$ , others have $\omega_1$	$\omega_1=0.174$ $\omega_2=0.966$	20372.189	129	A vs E	40.104	<0.001
F. The ancestral branch of Pholidota has $\omega_2 = 1$ , others have $\omega_1$	$\omega_1=0.174$ $\omega_2=1$	20372.195	128	F vs E	0.012	0.913



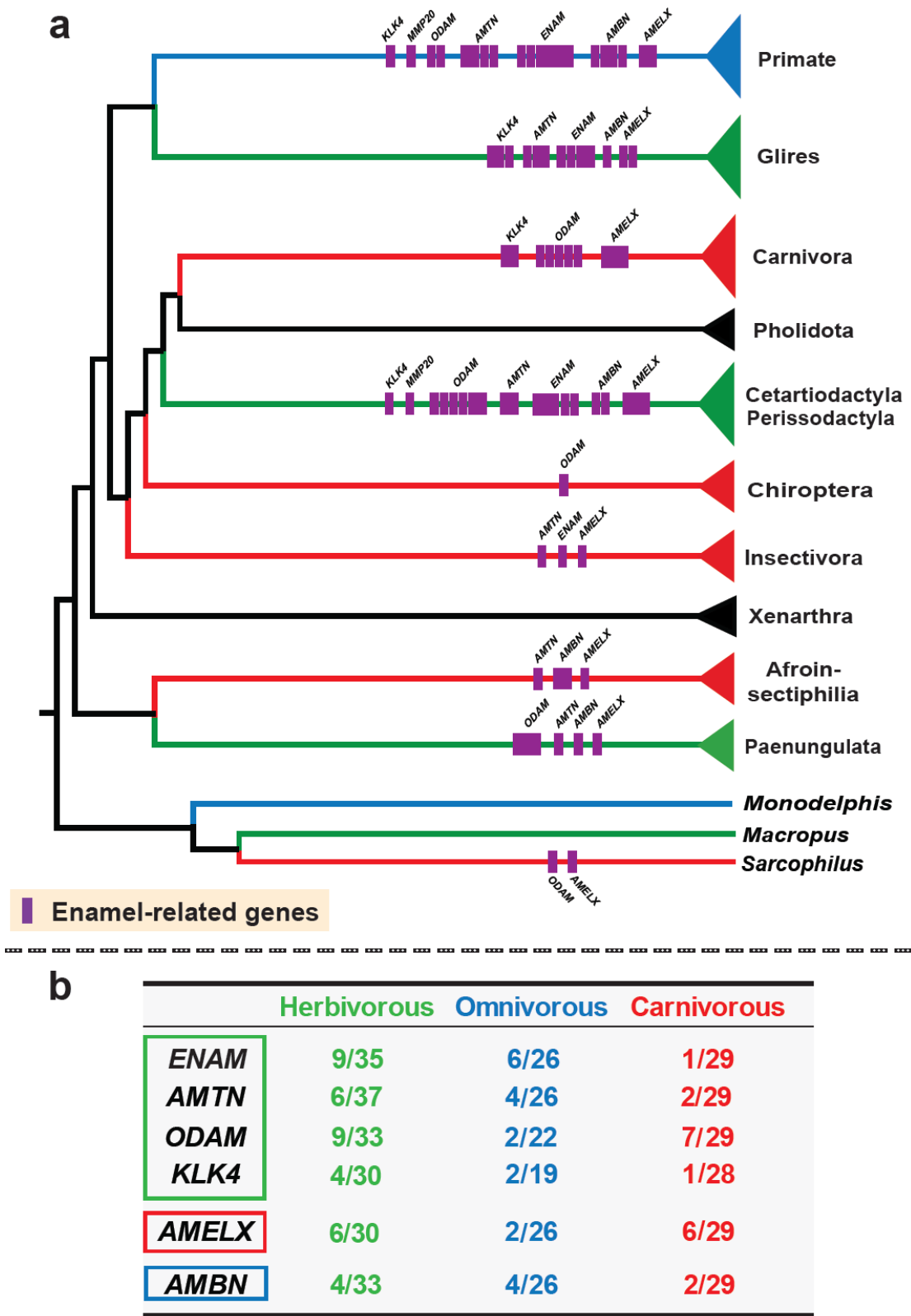


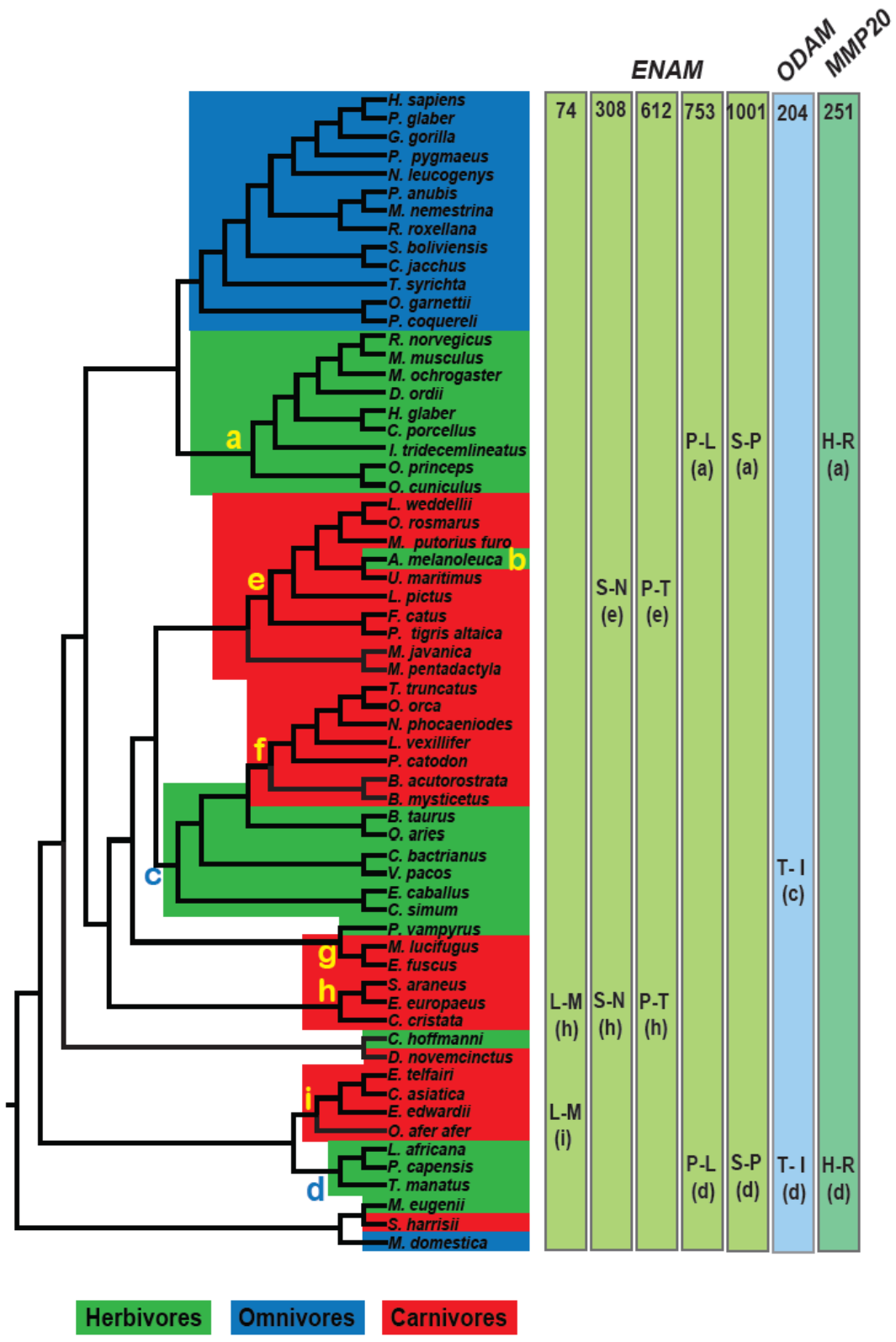
Fig. 1 a. Summary of positive selection of seven tooth-related genes based on the *free ratio model*

and *branch-site model*. The small bar on each clade stands for one positively selected branch detected within the corresponding clade. **b.** The proportion of positively selected branches of each gene in different dietary lineages where the denominator is the number of branches for each different dietary lineage, and numerator is the number of positively selected branches for different dietary lineage respectively.

When we used the more stringent *branch-site model* on dataset II where edentulous/enamelless mammals and some pseudogenized lineages were excluded, positive selection was observed mainly in herbivorous lineages for *AMBN*, *ENAM*, *AMTN* and *KLK4*. The proportions of positively selected branches in herbivorous lineages were generally higher than carnivorous and omnivorous lineages. The percentages were 5.7%, 17.1%, 8.1%, 10%, 43.3%, 28.6%, 18.9%, 26.5% in herbivorous lineages, which was relatively higher than 0, 3.4%, 0, 7.1%, 24.1%, 7.4%, 13.8%, 17.2% in carnivorous lineages and 0, 0, 3.8%, 3.6%, 3.8%, 15.4%, 11.5%, 4.2% in omnivorous lineages.

### **Convergent sites among different dietary mammals**

To determine the molecular convergent evolution in distantly related lineages with the same diet, we tested for the convergent amino acid changes in seven enamel-related genes. We identified seven parallel nonsynonymous substitution sites. Four of the substitutions were in herbivorous lineages for three related genes, *ENAM*, *ODAM* and *MMP20*, and three substitutions were identified in *ENAM* in carnivorous lineages (Fig. 2 & Table 2). These parallel substitutions deviated significantly ( $p < 0.05$ ) from the random expectation.



**Fig. 2** Parallel amino acid changes on the phylogenetic tree. Amino acid positions are listed in the top of colored bars (numbers), and parallel changes at each position are listed in the right part of

colored bars corresponding to genes marked in different colors.

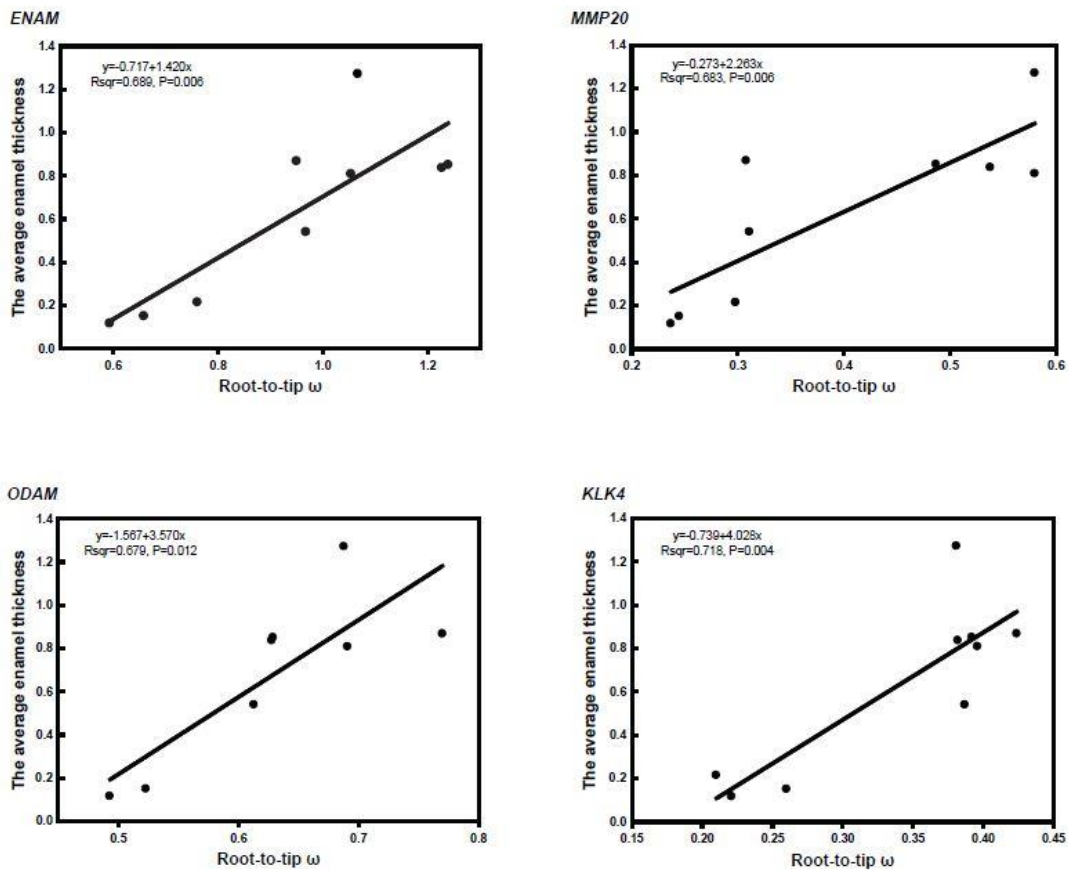
Table 2 Statistical tests for parallel nonsynonymous amino acid substitutions among lineages with different diets.

Genes	Branch pair	Parallel substitution	Observed number	Expected number	P value
<i>ENAM</i>	Gli vs. Pae	753 P-L	5	0	0
		1001 S-P			
	Car vs. Ins	308 S-N			
		621 P-T			
Ins vs. Afr	74 L-M				
<i>ODAM</i>	LCP vs. Pae	204 T-I	1	0	0
<i>MMP20</i>	Gli vs. Pae	251 H-R	1	0	0

Gli, Glires; Pae, Paenungulata; Car, Carnivora; Ins, Insectivora; Afr, Afroinsectiphilia; LCP, LCA of Cetartiodactyla and Perissodactyla; Cet, Cetacean; Pan, Panda

### Association between evolution of genetic sequences and enamel thickness in primates

Lastly, we conducted a PGLS analysis of the primate dataset (dataset III, Fig. S2) to test whether there is a significant correlation with evolution of enamel genes and enamel thickness. The  $\lambda$  value was 0, suggesting that there is no phylogenetic signal in our analysis to estimate the association between root-to-tip  $\omega$  of enamel-related genes and average enamel thickness. We also conducted a OLS (Ordinary Least Squares) analysis and found a significantly positive regression between root-to-tip  $\omega$  and average enamel thickness at *ENAM* ( $R^2=0.689$ ,  $P=0.007$ ), *ODAM* ( $R^2=0.679$ ,  $P=0.012$ ), *MMP20* ( $R^2=0.683$ ,  $P=0.006$ ) and *KLK4* ( $R^2=0.718$ ,  $P=0.004$ ) (Fig. 3).



**Fig. 3** Associations between root-to-tip  $\omega$  and average enamel thickness in primates using OLS.

## Discussion

Understanding the evolution of enamel-related genes can improve our understanding of phenotypic variation in mammalian teeth and further provide insights into the evolutionary mechanism underlying the adaptation to different diets.

In the present study, we conducted evolutionary genetics analyses and detected a series of positively selected sites in seven enamel-related genes in mammals, and found interesting findings with regard to gene evolution and its association with teeth variation and diet included. We found higher percentage of positively selected branches, evidenced by  $\omega$  values greater than one by branch model or more positively selected sites by branch-site models, in enamel-related genes in herbivorous lineages than in carnivorous lineages. There was significant convergence between some enamel-related

genes between distantly related lineages with the same feeding habits. For example, some convergences were found between different herbivorous lineages, such as Glires vs Paenungulata at *ENAM* and *MMP20*, whereas convergence was found between Insectivora vs Carnivora or Afroinsectiphilia. We found significant association between root-to-tip  $\omega$  and average enamel thickness for *ENAM* ( $R^2=0.689$ ,  $P=0.007$ ), *ODAM* ( $R^2=0.679$ ,  $P=0.012$ ), *MMP20* ( $R^2=0.683$ ,  $P=0.006$ ) and *KLK4* ( $R^2=0.718$ ,  $P=0.004$ ) in primates, suggesting that these genes may play an important role in controlling the thickness of enamel in primates. In enamelless or toothless mammals, enamel-related genes often became pseudogenes or non-functional. Taken together, our findings suggest that protein-coding changes in enamel genes are involved in tooth variation in mammals. Specifically, enamel-related genes appear to be under stronger positive selection in herbivores, potentially playing a role in the evolution of feeding on hard plants with fibers.

However, our study includes a limited selection of enamel genes, and a comprehensive understanding of the tooth evolution of mammals with different diets would require further investigation of more tooth-related genes, including additional enamel or dentine-related genes.

## References

1. Bergqvist LP: **The role of teeth in mammal history.** *Brazilian Journal of Oral Sciences* 2003, **2**(6):249-257.
2. Smith MM, Coates MI: **Evolutionary origins of the vertebrate dentition: phylogenetic patterns and developmental evolution.** *European Journal of Oral Sciences* 1998, **106**(s1):482-500.
3. Koussoulakou DS, Margaritis LH, Koussoulakos SL: **A curriculum vitae of teeth: evolution, generation, regeneration.** *International Journal of Biological Sciences* 2008, **5**(3):226-243.
4. Walker A, Hoeck HN, Perez L: **Mecrowear of mammalian teeth as an indicator of diet.** *Science* 1978, **201**(4359):908-910.
5. Craig RG, Peyton FA: **The micro-hardness of enamel and dentin.** *Journal of Dental Research* 1958, **37**:661-668.
6. Davit-Béal T, Tucker AS, Sire JY: **Loss of teeth and enamel in tetrapods: fossil record,**

- genetic data and morphological adaptations.** *Journal of Anatomy* 2009, **214**(214):477-501.
7. Werth A: **Feeding in Marine Mammals:** Academic Press; 2000.
  8. Ranwez V, Delsuc F, Ranwez S, Belkhir K, Tilak MK, Douzery EJ: **OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics.** *BMC Evolutionary Biology* 2007, **7**(1):241.
  9. <http://www.ensembl.org/index.html>
  10. <http://www.ncbi.nlm.nih.gov>
  11. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004, **32**(5):1792-1797.
  12. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular Biology & Evolution* 2016, **33**(7):1870.
  13. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Molecular Biology & Evolution* 2007, **24**(24):1586-1591.
  14. Yang Z, Nielsen R: **Synonymous and nonsynonymous rate variation in nuclear genes of mammals.** *Journal of Molecular Evolution* 1998, **46**(4):409-418.
  15. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Molecular Biology & Evolution* 1998, **15**(15):568-573.
  16. Foote AD, Liu Y, Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V: **Convergent evolution of the genomes of marine mammals.** *Nature Genetics* 2015, **47**(3):272-275.
  17. Zhang J, Kumar S: **Detection of convergent and parallel evolution at the amino acid sequence level.** *Molecular Biology & Evolution* 1997, **14**(5):527-536.
  18. Shellis RP, Beynon AD, Reid DJ, Hiiemae KM: **Variations in molar enamel thickness among primates.** *Journal of Human Evolution* 1998, **35**(4-5):507-522.
  19. Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI: **Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates.** *Molecular Biology & Evolution* 2010, **28**(1):625-638.
  20. Butler MA, King AA: **Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution.** *American Naturalist* 2004, **164**(6):683-695.

Table S1 Summary of the seven candidate enamel-related genes used in this study

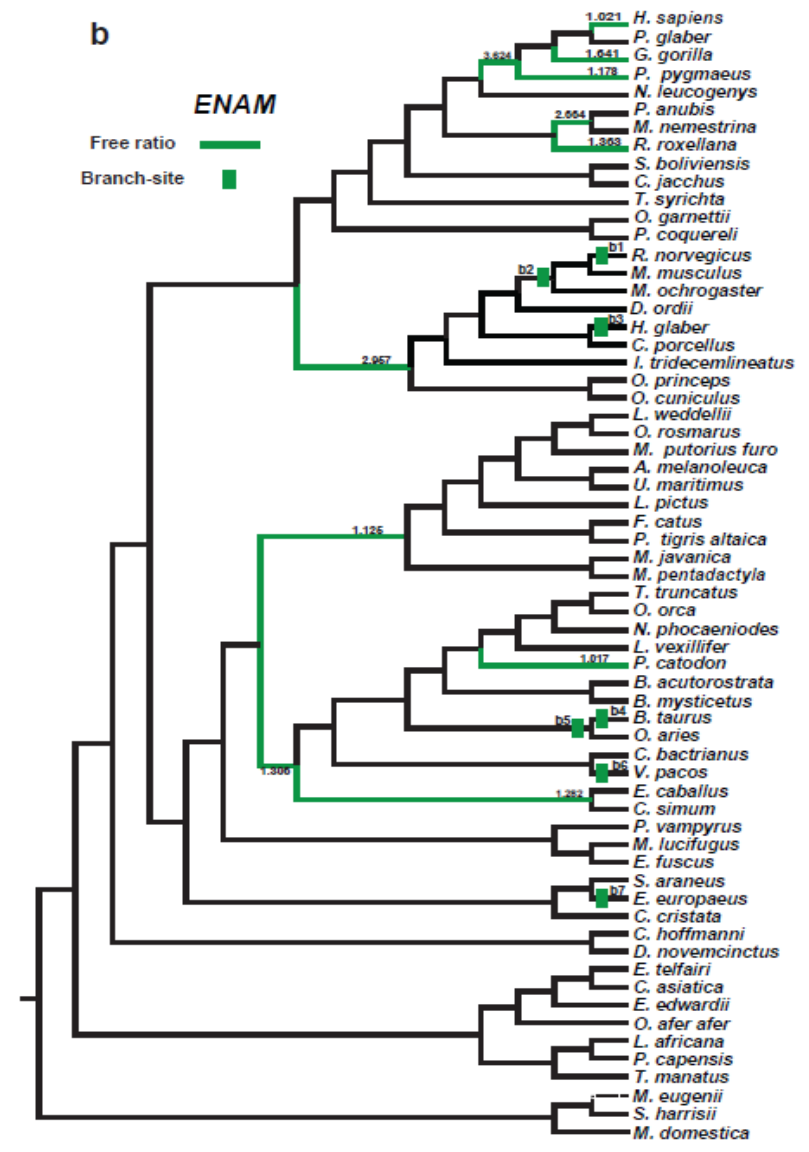
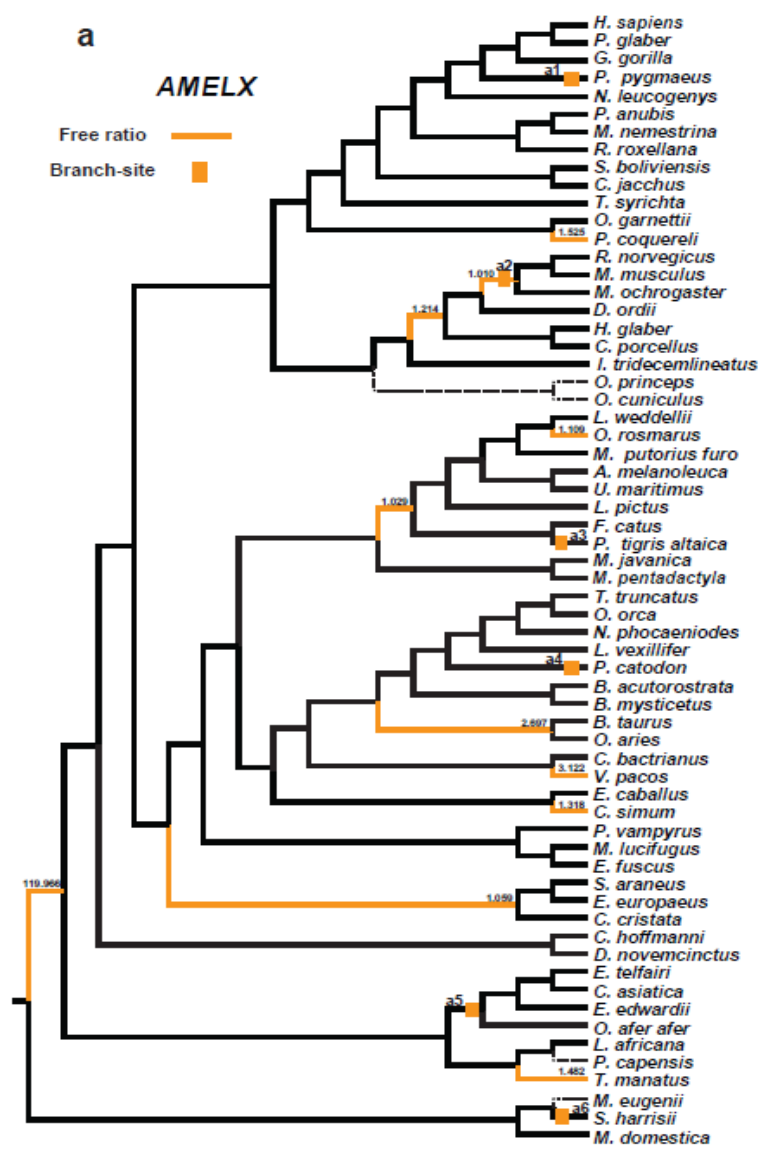
Genes <sup>A</sup>	Full name	Chromosome location	Function and disorders/diseases <sup>B</sup>	Formation stage (and distribution)
<i>AMELX</i>	X-link Amelogenin	Xp22.2	Amelogenins are involved in biomineralization during tooth enamel development. Comprises approximately 80-90% of total enamel protein, Mutations in this gene cause X-linked amelogenesis imperfecta. Indeed, enamel is present in <i>AMEL</i> <sup>-/-</sup> mice, but it turns out to severe AI.	Secretory stage (Enamel)
<i>AMBN</i>	Ameloblastin	4q13.3	The encoded protein may be important in enamel matrix formation and mineralization. Comprises roughly 5% of enamel protein matrix. A potential function in the adhesion of the ameloblasts. Mutations in this gene may be associated with dentinogenesis imperfect and autosomal dominant amelogenesis imperfect.	Express at the secretory stage, diminishes at the maturation stage (Enamel)
<i>ENAM</i>	Enamelin	4q13.3	ENAM, the largest protein in the enamel matrix, comprises only 5% of the total EMPs. And it is a glycosylated, phosphorylated protein. The intact protein is only observed at the mineralization front, suggesting it plays a role in crystal elongation. Diseases associated with ENAM include Amelogenesis Imperfecta, Type Ib and Amelogenesis Imperfecta, Type Ic.	Secretory stage (Enamel)
<i>AMTN</i>	Amelotin	4q13.3	AMTN is specifically expressed in maturation-stage ameloblasts. It may function at basal lamina during tooth formation, which is a promoter of calcium phosphate mineralization, playing a critical role in the formation of the compact, mineralized, aprismatic enamel surface layer during the maturation stage of amelogenesis.	Maturation stage (Enamel)
<i>ODAM</i>	Odontogenic, ameloblast-associated	4q13.3	ODAM probably plays a role in odontogenesis, the complex process that results in the initiation and generation of the tooth. May be incorporated in the enamel matrix at the end of mineralization process. It also plays a role in attachment of the junctional epithelium to the tooth surface.	Maturation stage (Enamel, Milk, Saliva)
<i>MMP20</i>	Matrix Metalloproteinase 20	11q22.2	MMP20 can degrade AMEL, the major protein component of dental enamel matrix, and thus thought to function in tooth enamel formation. Mutations of <i>MMP20</i> and <i>KLK4</i> both cause autosomal recessive AI, a soft characteristics, become a condition that porous enamel	Throughout the secretory stage and into early maturation stage

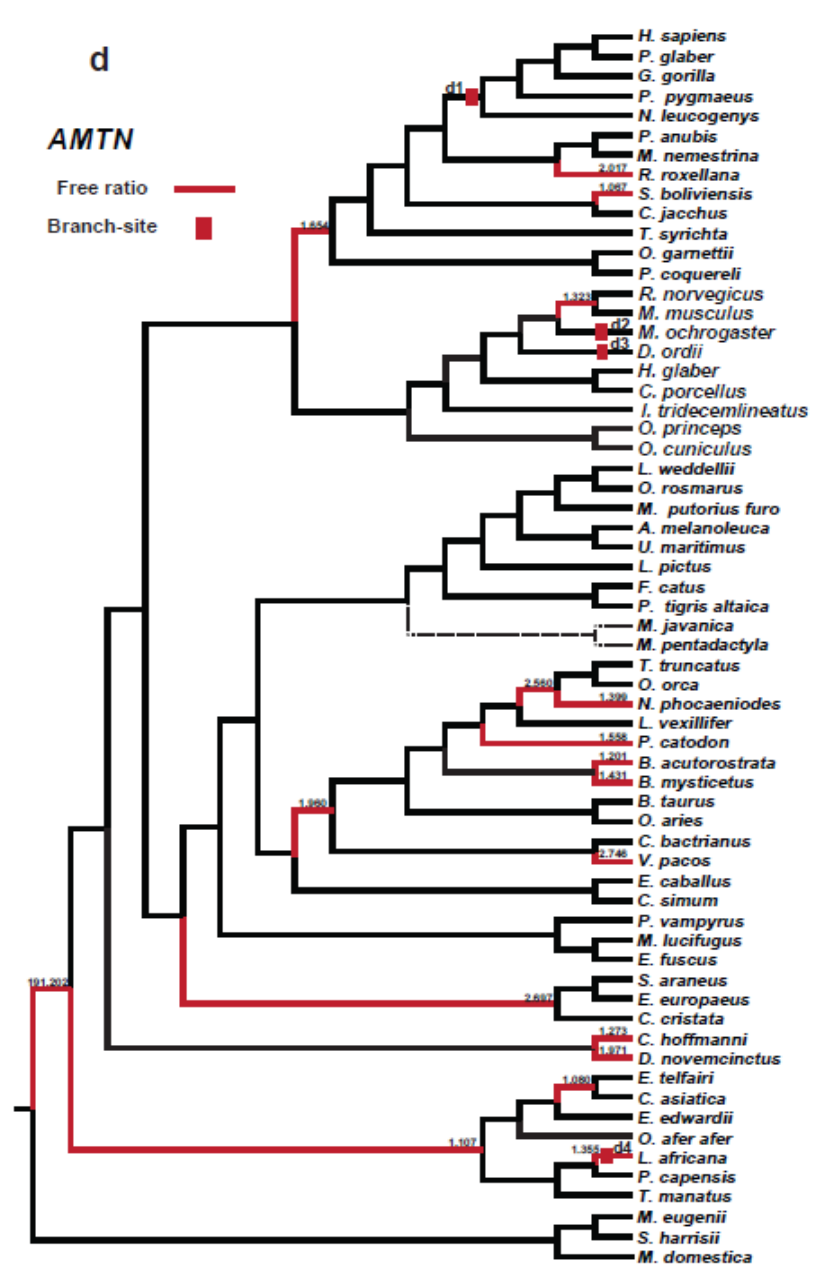
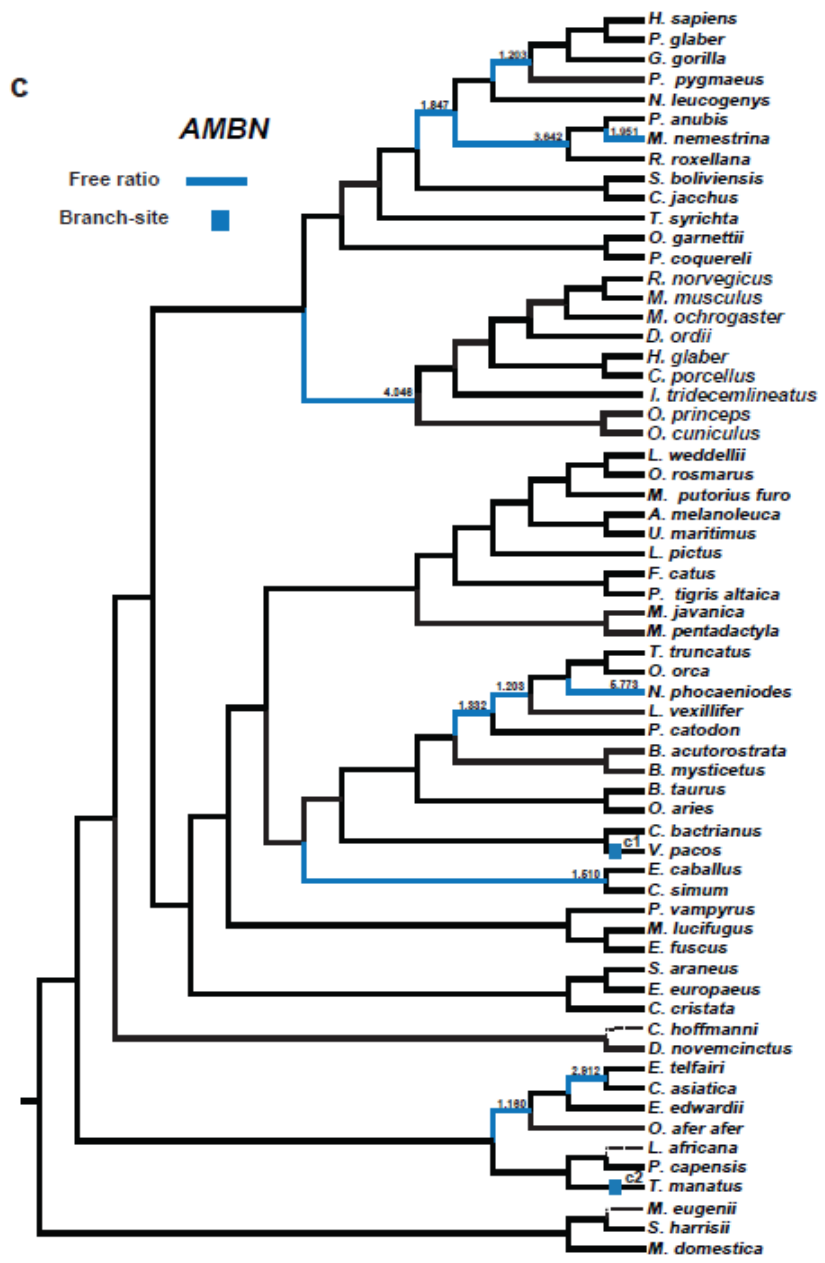


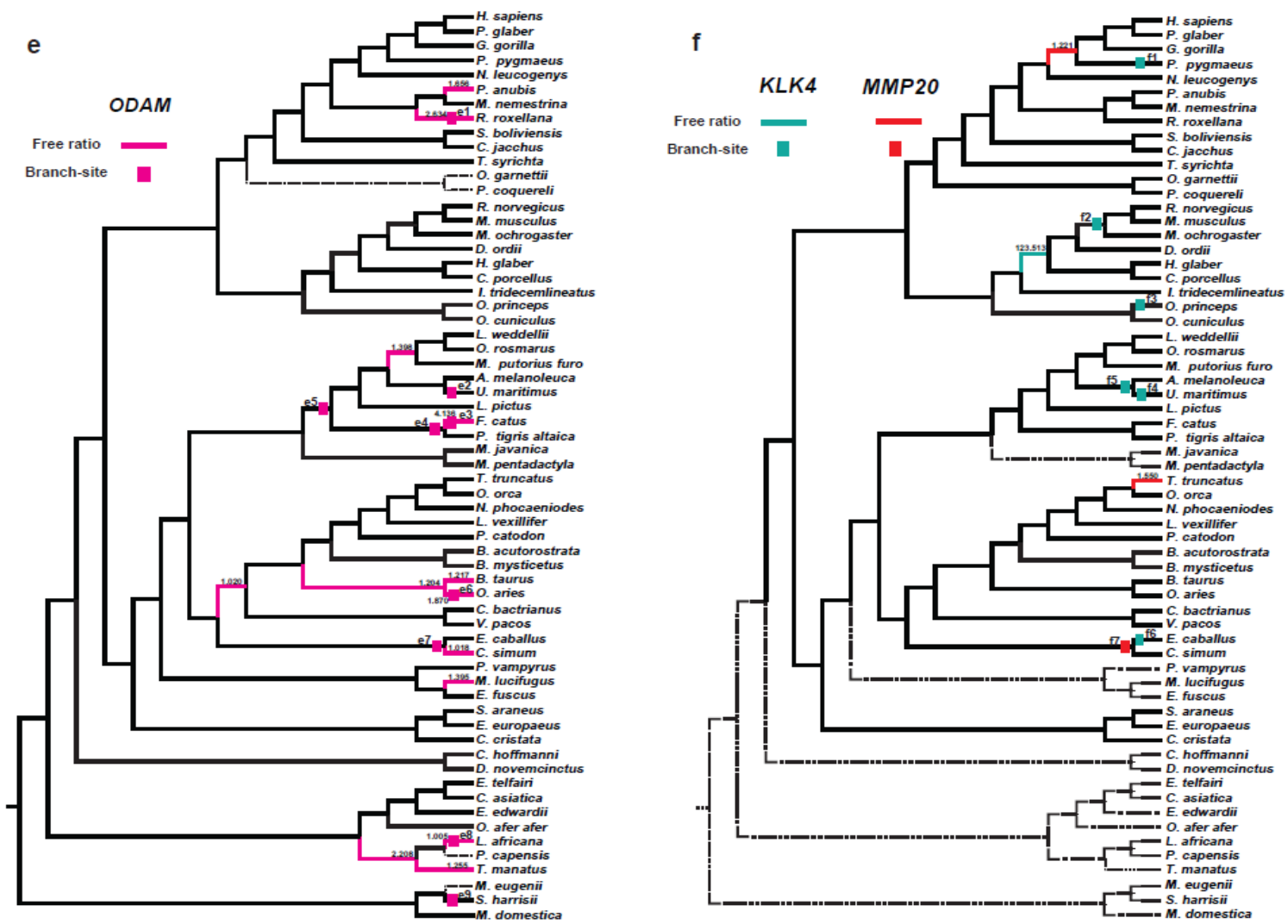
Genes <sup>A</sup>	Full name	Chromosome location	Function and disorders/diseases <sup>B</sup>	Formation stage (and distribution)
<i>KLK4</i>	Kallikrein Related Peptidase 4	19q13.41	containing residual proteins In some tissues its expression is hormonally regulated. After the EMPs formation and incassation, soon hydrolyzed by two main metalloprotease, MMP20 and KLK4, which increase mineralization and harden the enamel. Diseases associated with KLK4 include Amelogenesis Imperfecta, Type Iia1 and Amelogenesis Imperfecta Hypomaturation Type.	(Enamel) Starting in transition/early maturation and continuing through tooth eruption (Enamel)

A Human genes that were chosen as the reference for BLAST search.

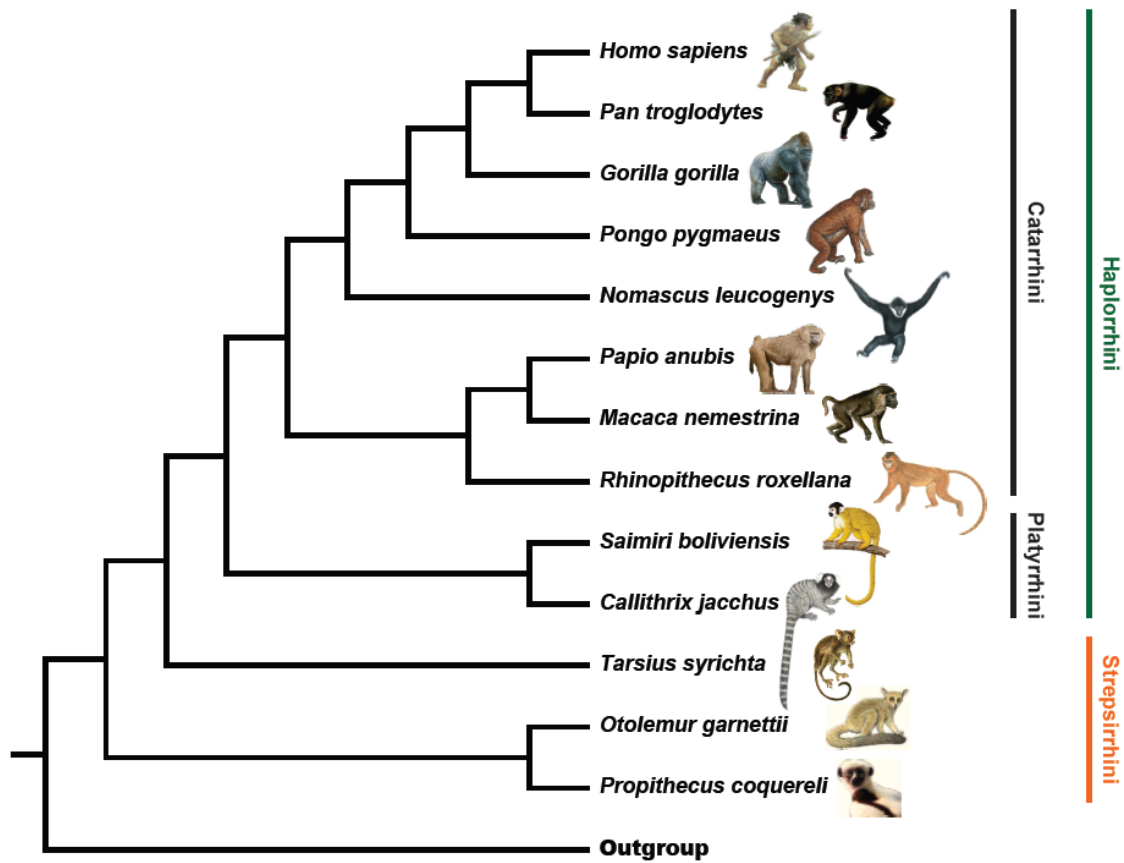
B Summary of gene functions and related disorders/diseases from NCBI [<https://www.ncbi.nlm.nih.gov/gene/>] and GeneCard [<http://www.genecards.org/>].







**Fig. S1a-f** Distribution of positive selection of each genes on the species tree. Dashed line indicates that data was unavailable for the species.



**Fig. S2** The primate dataset and phylogeny used for association analysis between root-to-tip  $\omega$  and average enamel thickness.

(Images are derived from WiKi website: <https://en.wikipedia.org/wiki/Primate>; AWD website: <http://animaldiversity.org/>, respectively)

## 项目分工及致谢:

杨奕非在本项研究中承担了部分序列的搜索和比对、选择压力分析, 及论文初稿的撰写。

龚靖尧承担了不同哺乳类哺乳动物趋同进化位点的鉴定、灵长类 root-to-tip  $\omega$  值和牙齿表型之间的关系分析, 参与论文的修改。

戴川滢承担了部分序列的搜索、比对, 以及序列校对工作, 构建系统发育树。

感谢南京师范大学生命科学学院博士研究生沐远对基因 CDS 序列的校对, 及在分析过程中提供的指导和帮助!

感谢两位老师对本项目实施过程及论文撰写和修改的指导!

章熙东: 男, 硕士, 1979 年 12 月出生, 2006 年毕业于南京师范大学生命科学学院, 现为南京外国语学校高中生物教师, 生物教研组组长, 执教 14 年, 指导的学生多人获得全国中学生生物学联赛一等奖。并多次参与市级、省级和国家级的课题研究, 发表论文数篇。2017 指导学生参加 IYNT 国际比赛获得铜牌。教学中从传授学生知识为主的学习方式, 到为学生构建开放的学习环境, 提供多渠道获取知识并将学到的知识加以综合应用于实践的机会, 促进他们形成积极的学习态度和良好的学习策略, 培养创新精神和实践能力。

徐士霞: 女, 博士, 1978 年生, 2007 年毕业于南京师范大学, 2007 年至 2009 年中国科学院动物研究所做博士后。现为南京师范大学生命科学学院教授, 动物学专业硕士生导师。曾获 2008 年度江苏省优秀博士论文, 2012 年度江苏省高校和南京师范大学“青蓝工程”优秀青年骨干教师。主要研究方向: 运用比较基因组学探讨哺乳动物适应性进化机制。

此页为学术诚信声明

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知,除了文中特别加以标注和致谢中所罗列的内容以外,论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处,本人愿意承担一切相关责任。

参赛队员: 杨奕非、龚靖尧、戴川滢 指导老师: 章熙东、徐士霞

2017年9月10日