Evolution of Digestive Enzymes and RNASE1 Provides Insights into Dietary Switch of Cetaceans

Zhengfei Wang,[†,1] Shixia Xu,[†,1] Kexing Du,[1] Fang Huang,[1] Zhuo Chen,[1,2] Kaiya Zhou,[1] Wenhua Ren,[*,1] and Guang Yang[*,1]

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

[2]College of Life Sciences, Henan Normal University, Xinxiang, China

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: 08162@njnu.edu.cn; gyang@njnu.edu.cn.

Associate editor: Bing Su

Abstract

Although cetaceans (whales, porpoises, and dolphins) have multi-chambered stomachs, feeding habits of modern cetaceans have dramatically changed from herbivorous to carnivorous. However, the genetic basis underlying this dietary switch remains unexplored. Here, we present the first systematic investigation of 10 digestive enzymes genes (i.e., *CYP7A1*, *CTRC*, *LIPC*, *LIPF*, *PNLIP*, *PGC*, *PRSS1*, *SI*, *SLC5A1*, and *TMPRSS15*) of representative cetaceans, and the evolutionary trajectory of *RNASE1* in cetartiodactylans. Positive selections were detected with proteinases (i.e., *CTRC*, *PRSS1*, and *TMPRSS15*) and lipases (i.e., *CYP7A1*, *LIPF*, and *PNLIP*) suggesting that cetaceans have evolved an enhanced digestion capacity for proteins and lipids, the major nutritional components of their prey (fishes and invertebrates). In addition, it was found that *RNASE1* gene duplicated after the cetartiodactylan speciation and two independent gene duplication events took place in Camelidae and Ruminantia. Positive selection was detected with *RNASE1* of Camelidae and Bovidae, suggesting enhanced digestive efficiency in the ruminants. Remarkably, even though the ancestors of cetaceans were terrestrial artiodactyls that are herbivorous, modern cetaceans lost the pancreatic *RNASE1* copy with digestive function, which is in accordance with the dietary change from herbivorous to carnivorous. In sum, this is the first study that provides new insights into the evolutionary mechanism of dietary switch in cetaceans.

Key words: **evolution, cetaceans, positive selection, digestive enzymes, RNASE1, dietary switch.**

Article

Introduction

Cetaceans, a highly specialized group of mammals that includes whales, dolphins, and porpoises, are suggested to be evolved from terrestrial artiodactyls by paleontological, morphological, embryological, and molecular processes. Their transition from land to water, and subsequent adaptation to complete aquatic habitats (e.g., loss of limbs, shortening of the skull, thickening of the blubber, and loss of hair) (Reidenberg 2007), make cetaceans evolutionarily significant and remarkable (McGowen et al. 2014). Most notably, in contrast to their terrestrial, herbivorous ancestors, extant cetaceans are carnivorous, with their diet mainly composed of fish, large squid, and zooplankton (Pauly et al. 1998). Although they, like their ruminant relatives, still retain a multi-chambered foregut.

Dietary switch in mammals is intriguing and has attracted the interest of many evolutionary biologists. Genetic basis underlying dietary switches of giant pandas (from carnivorous to herbivorous), dogs (from carnivorous to omnivorous), and colobine monkeys (from omnivorous to herbivorous) have been studied in recent decades (Zhang et al. 2002; Zhao et al. 2010; Axelsson et al. 2013). For example, the duplication of pancreatic ribonuclease (RNASE1, encoded by *RNASE1*) is thought to be in response to a dietary switch from omnivorous to herbivorous required to digest leaves in colobine monkeys (Zhang et al. 2002). However, the molecular basis and consequences of dietary switch of cetacean has not been addressed to date.

There are mainly four kinds of digestive enzymes (i.e., lipases, proteinases, amylases, and cellulases) in the digestive tracts and organs of mammals. Lipases are water-soluble enzymes that hydrolyze ester bonds of water-insoluble substrates such as triglycerides, phospholipids, and cholesteryl esters (Wong and Schotz 2002). Proteases, secreted by the stomach and pancreas, digest proteins and begin with intraluminal hydrolysis. They continue as peptidases in the brush-border membrane of the small intestine, digesting the protein into amino acids and di- and tri-peptides that are then taken up by the enterocytes (Sitrin 2014a). Amylases are present in both the salivary and pancreatic secretions, and are responsible for intraluminal starch digestion (Sitrin 2014b). Cellulases degrade cellulose and release reducing sugars as the end products (Deng and Tabatabai 1995); however, vertebrates seem to lack the ability to produce this enzyme (Stickney and Shumway 1974).

Digestive enzymes of animals with different feeding habits (carnivores, omnivores, and herbivores) possess different characteristics. For example, the activity of amylases in the

herbivores is generally higher than that in the carnivores and omnivores (Agrawal et al. 1974). During transition from land to water, the diet of cetaceans changed from plants to animals, the latter being rich in lipids and proteins (Trumble et al. 2003). In response to such a dramatic change of food habit, digestive enzymes of cetaceans, especially those for digesting lipids and proteins, should have adaptively evolved in order to get enough energy and nutrition from the transitioned foods.

In addition to the digestive enzymes, *RNASE1* is another gene that is associated with food habit. *RNASE1* gene duplication has been found in many mammals with multi-compartmentalized stomachs, such as ruminants (e.g., cows, sheep, and camels) (Kleineidam et al. 1999; Breukelman et al. 2001), and species with ruminant-like or cecal digestions (e.g., leaf-eating colobines and elephants) (Zhang et al. 2002; Yu et al. 2010; Goo and Cho 2013; Zhou et al. 2014). Gene duplication of *RNASE1* is believed to be correlated with the plant-feeding adaptation of foregut-fermenting herbivores (Liu et al. 2014), and ruminant artiodactyls have considerably higher concentrations of pancreatic *RNASE1* than other mammals (Zhang 2006). However, it is still not clear whether cetacean *RNASE1* has experienced a special evolutionary trajectory during the dietary switch from ancestral herbivorous to extant carnivorous.

In the present study, 10 digestive enzyme genes and *RNASE1* were investigated in representative species of major cetacean lineages, and compared with orthologous sequences from their terrestrial relatives, with an aim to uncover the process and pattern of evolution during the transition from herbivorous to carnivorous. Positive selection of digestive enzyme genes and gene loss of *RNASE1* provided some new insights into the molecular mechanism underlying the cetacean dietary switch.

Results

Molecular Evolution of Digestive Enzyme Genes in Cetaceans

Data Set of All Mammals

We used a pair of site models (M8 vs. M8a) to test whether specific codons in the genes of mammalian digestive enzymes have been subjected to positive selection. We compared nested models and found that the M8 model, which incorporated selection, fitted significantly better than the neutral model, M8a, for *CTRC*, *LIPC*, *LIPF*, *PNLIP*, *PRSS1*, *SI*, *SLC5A1*, and *TMPRSS15* genes, whereas no significant evidence of positive selection was found for *CYP7A1* and *PGC* (supplementary table S1, Supplementary Material online). The ω values were calculated as 1.878, 1.549, 2.433, 2.066, 2.377, 2.032, 1.482, and 1.565, respectively, for these eight positive selection genes; whereas 5, 7, 11, 12, 10, 43, 13, and 15 sites were separately identified to be under selection by the BEB approach as having posterior probabilities $\geq$ 0.80 (supplementary table S1, Supplementary Material online).

We then used the branch-site model to test for positive selection in individual codons for the lineage leading to cetaceans and other groups (i.e., cetartiodactyla, perissodactyla, carnivora, chiroptera, primates, lagomorpha, and rodentia) across the mammalian phylogeny (*a*, *ab*–*ag* in supplementary fig. S1, Supplementary Material online). Evidence for positive selection was identified along the lineages leading to cetaceans (branch *a* in supplementary fig. S1, Supplementary Material online) at the *SI* gene. This gene also showed significant signs of positive selection in the lineages of the other groups (i.e., perissodactyla, carnivore, chiroptera, lagomorpha, and rodentia). In addition, *LIPF*, *PNLIP*, *SLC5A1*, *TMPRSS15* in lagomorpha, *CTRC*, *LIPC*, *LIPF* in chiroptera, *CTRC*, *LIPC* in perissodactyla, and *PGC* in carnivora were also found to be positively selected. A summary of these results are shown in supplementary table S2, Supplementary Material online.

Cetaceans-Only Data Set

To determine how selection pattern occurred within the cetacean lineages, we again used the M8 versus M8a pair of site models in the cetaceans-only data set. Six genes (i.e., *CTRC*, *CYP7A1*, *LIPF*, *PNLIP*, *PRSS1*, and *TMPRSS15*) were found to be under positive selection in the cetaceans-only data set, where the likelihood ratio tests (LRTs) of the site model (M8 vs. M8a) were statistically significant (table 1). In the M8 model, average ω values for cetaceans were 3.103, 22.845, 2.856, 4.859, 15.716, and 2.765 for *CTRC*, *CYP7A1*, *LIPF*, *PNLIP*, *PRSS1*, and *TMPRSS15*, respectively, and 14, 1, 22, 9, 7, and 10 codons were identified as being positively selected using the BEB approach with posterior probabilities $\geq$ 0.8 (table 1). Five (*CTRC*, *LIPF*, *PNLIP*, *PRSS1*, and *TMPRSS15*) of these six genes were also found to be under positive selection in the mammalian data set. The fixed-effects likelihood (FEL) and random-effects likelihood (REL) models implemented in the Datamonkey where the ω values calculated based on the d_S whereas d_S was fixed in the PAML analysis. A total of 46 codons (14 in *CTRC*, 1 in *CYP7A1*, 7 in *LIPF*, 10 in *PNLIP*, 4 in *PRSS1*, and 10 in *TMPRSS15*) identified at the six genes in the cetaceans (table 2 and supplementary table S3, Supplementary Material online). Moreover, 60.2% (56/93) of putative positively selected sites from six genes (10 in *CTRC*, 1 in *CYP7A1*, 19 in *LIPF*, 10 in *PNLIP*, 7 in *PRSS1*, and 9 in *TMPRSS15*) were identified to be undergoing radical changes using the protein-level approach implemented in TreeSAAP (table 2 and supplementary table S3, Supplementary Material online), which provided additional evidence for the operation of positive selection on cetaceans.

To evaluate the diversified selection pressures on various lineages within cetaceans, the optimized branch-site model was used to test for positive selection in individual codon in each lineage across cetacean phylogeny (*b*–*u*, fig. 1). The LRT tests showed evidence of positive selection in three cetacean-specific lineages, that is, the lineage leading to the last common ancestor of delphinoidea and baiji (*Lipotes vexillifer*) (fig. 1) for *PNLIP*, terminal branch of the sperm whale (*Physeter macrocephalus*) for *PRSS1*, and terminal branch of the baiji for *CYP7A1* and *PNLIP*.

Parallel/Convergent Analysis

To test if different carnivorous lineages of mammals, such as cetaceans and carnivorans, showed similar patterns of evolution in their digestive enzyme genes in response to similar food habit, we first reconstructed ancestral nodes and then detected the parallel/convergent amino acids of the six

Table 1. Analysis of M8 and M8a, and Evidence of Positive Selection at 10 Digestive Enzyme Genes of Cetaceans.

Gene	Models	-lnl	*P* value	M8 ω value	Positively selected sites
CTRC	M8	1686.118	<0.01	3.103	5(0.987) 34(0.867) 37(0.967) 54(0.992) 67(0.994) 100(0.950) 102(0.886) 111(0.890) 114(0.976) 119(0.847) 130(0.879) 139(0.977) 152(0.994) 164(0.949)
	M8a	1695.052			
CYP7A1	M8	2794.383	<0.01	22.845	10(0.987)
	M8a	2798.536			
LIPC	M8	2774.549	1.000	1.000	
	M8a	2774.547			
LIPF	M8	2428.593	<0.01	2.856	49(0.893) 150(0.931) 185(0.916) 214(0.923) 229(0.868) 235(0.867) 236(0.904) 238(0.999) 246(0.917) 247(0.980) 249(0.884) 251(0.867) 252(0.896) 253(0.977) 261 (0.982) 266(0.845) 298(0.882) 301(0.979) 304(0.892) 341(0.844) 351(0.918) 382(0.934)
	M8a	2437.265			
PGC	M8	2217.442	0.596	1.412	
	M8a	2217.582			
PNLIP	M8	2783.625	<0.01	4.859	88(0.950) 105(0.838) 112(0.964) 232(0.945) 237(0.938) 306(0.989) 341(0.991) 397(0.882) 451(0.801)
	M8a	2792.623			
PRSS1	M8	1783.570	<0.01	15.716	98(0.945) 102(1.000) 115(0.980) 121(0.855) 125(0.894) 131(0.994) 177(1.000)
	M8a	1806.076			
SI	M8	8501.121	0.283	1.477	
	M8a	8501.698			
SLC5A1	M8	3227.422	0.051	3.403	
	M8a	3229.332			
TMPRSS15	M8	4934.207	<0.01	2.765	12(0.843) 153(0.926) 165(0.859) 182(0.847) 269(0.837) 299(0.854) 408(0.860) 440(0.820) 599(0.853) 789(0.840)
	M8a	4937.784			

Table 2. Positive Selection at 10 Cetacean Digestive Enzyme Genes Based on the Analysis of the All Mammals and Cetaceans-Only Data Sets.

Gene	All mammals (site model)	Cetaceans-only (site model, Datamonkey, TreeSAAP)[a]	Cetaceans-only (branch-site model)
CTRC	Y	Y	N
CYP7A1	N	Y	Y
LIPC	N	N	N
LIPF	Y	Y	N
PGC	N	N	N
PNLIP	Y	Y	Y
PRSS1	Y	Y	Y
SI	Y	N	N
SLC5A1	Y	N	N
TMPRSS15	Y	Y	N

[a]Results are listed in supplementary tables S3, Supplementary Material online.
Y, positive selection was detected; N, no positive selection was detected.

positively selected genes. Four statistically significant ($P < 0.01$) parallel/convergent substitutions between these two lineages were identified in three genes (table 3), that is, one (I4V) at *CTRC*, two (R190K and N208H) at *CYP7A1*, and one (I69L) at *PNLIP*, respectively (branch *a* vs. *ac* in supplementary fig. S1, Supplementary Material online). The position of parallel/convergent nonsynonymous amino acid substitutions that were found in positively selected genes are shown in table 3.

Spatial Distribution of Positively Selected Sites in Protein Structures

Functional domains of each digestive enzyme gene were further examined to determine the significance of the putative positively selected sites. It was observed that the most positively selected sites were located in, or close to the functional regions within the structures of the digestive enzymes (supplementary fig. S2 and table S4, Supplementary Material online). One positively selected sites in *CTRC* (site 5) and *PRSS1* (site 13) were located in the signal peptide, respectively (supplementary table S4, Supplementary Material online). In addition, positively selected sites of the digestive enzymes were located primarily in functional domains that facilitated ligand–receptor interactions, such as glycosylation (*TMPRSS15*: site 440), active site (*LIPF*: site 341), helical transmembrane (*CYP7A1*: site 10), activation peptide(*PRSS1*: site 28), metal binding (*PRSS1*: sites 80 and 85), cytoplasmic topological domains (*TMPRSS15*: site 12), and extracellular

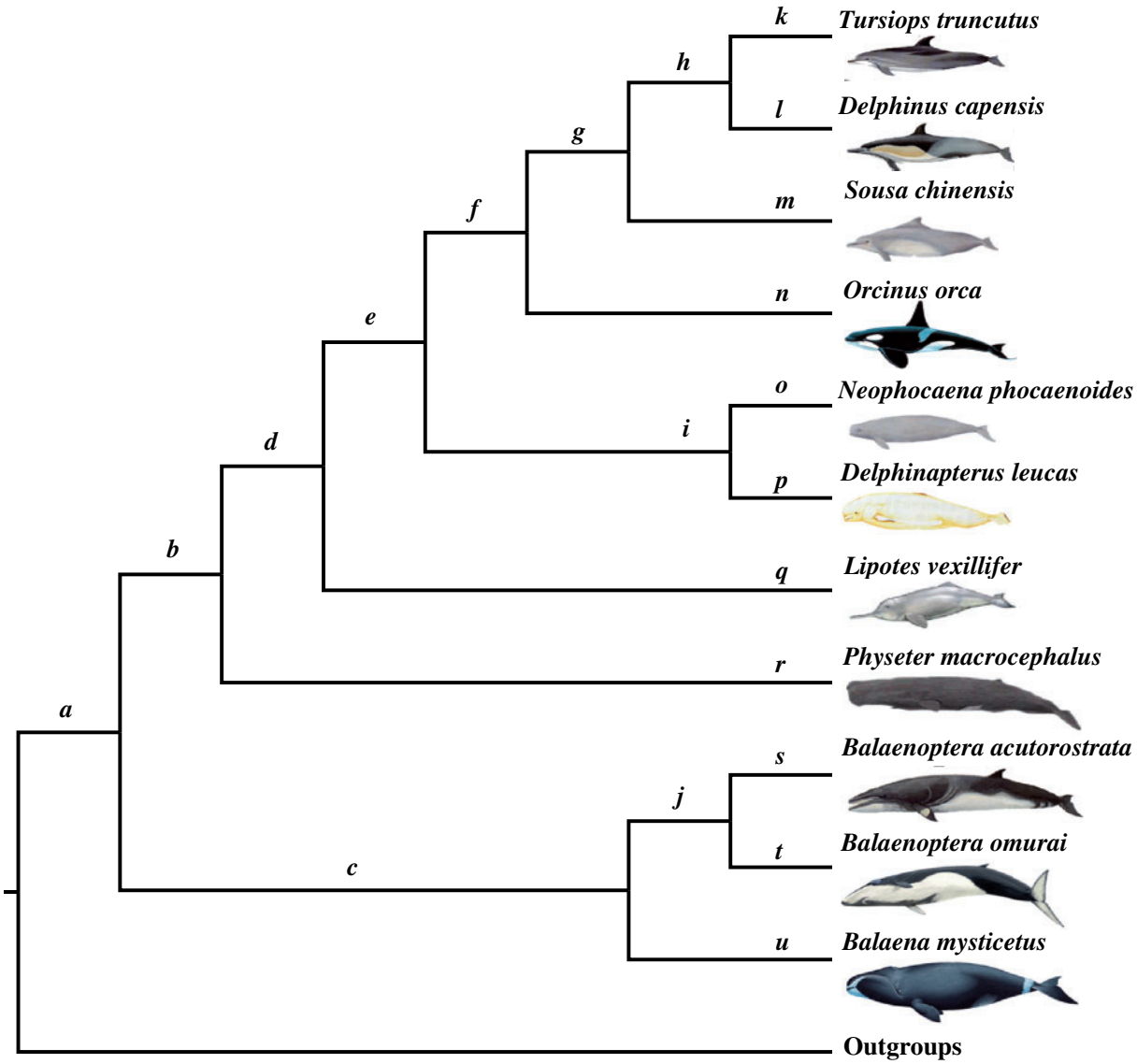

FIG. 1. A well-supported phylogeny of cetaceans used for selective pressure analysis in PAML. Branches *a*–*s* in the tree are used in the branch-site models tests. Pictures of representative cetacean members on the right of the phylogenetic tree are drawn by Professor Kaiya Zhou.

Table 3. Statistical Tests for Parallel Nonsynonymous Amino Acid Substitutions Between Cetaceans and Carnivores.

Gene	Branch pair	Parallel substitution	Observed number	Expected number	*P* value
CTRC	Cet vs. Car	I4V	1	0	<0.01
CYP7A1	Cet vs. Car	R190K N208H	2	0	<0.01
PNLIP	Cet vs. Car	I69L	1	0	<0.01

Cet, Cetacean; Car, Carnivora.

topological domains (*TMPRSS15*: sites 153, 165, 182, 299, 408, and 599) (supplementary table S4, Supplementary Material online). Remarkably, 75% (12/16) positively selected sites located in the functional domain were involved in radical changes in their physicochemical properties (charge, polarity, and volume) (supplementary table S5, Supplementary Material online).

Evolutionary Trajectory of *RNASE1* in Cetartiodactylans

Variation of RNASE1 Gene Number among Mammalian Genomes

A total of 77 *RNASE1* sequences from 43 mammalian species were included in this study (table 4). Of these, 46 sequences were newly identified from 26 mammalian

Table 4. Gene Counts for *RNASE1* Identified in 43 Mammals.

Common name	Latin name	Abbreviation	RNASE1	Total
Bottlenose Dolphin	*Tursiops truncatus*	Tt	1	1
Killer whale	*Orcinus orca*	Oo	1	1
Yangtze finless porpoise	*Neophocaena asiaeorientalis asiaeorientalis*	Naa	1	1
Baiji	*Lipotes vexillifer*	Lv	1	1
Sperm whale	*Physeter macrocephalus*	Pm	1	1
Minke whale	*Balaenoptera acutorostrata*	Ba	1	1
Bowhead whale	*Balaena mysticetus*	Bm	1	1
Goat	*Capra hircus*	Ch	2(1)	3
Sheep	*Ovis aries*	Oar	2(1)	3
Tibetan antelope	*Pantholops hodgsonii*	Ph	2(1)	3
Cow	*Bos taurus*	nBt	3	3
Wild yak	*Bos grunniens mutus*	Bgm	3	3
Pig	*Sus scrofa*	Ss	1	1
Camel	*Camelus ferus*	Cfe	2	2
Alpaca	*Vicugna pacos*	Vpa	2	2
Rhinoceros	*Ceratotherium simum simum*	Css	1	1
Horse	*Equus caballus*	Ec	1	1
Seal	*Leptonychotes weddellii*	Lw	1	1
Elephant seal	*Odobenus rosmarus divergens*	Ord	1	1
Ferret	*Mustela putorius furo*	Mpf	2[1](2)	5
Giant panda	*Ailuropoda melanoleuca*	Am	1	1
Dog	*Canis familaris*	Cf	1	1
Little brown bat	*Myotis lucifugus*	Ml	7	7
Human	*Homo sapiens*	Hs	1	1
Chimpanzee	*Pan troglodytes*	Pt	1	1
Bonobo	*Pan paniscus*	Ppa	1	1
Gorilla	*Gorilla Gorilla*	Gg	1	1
Orangutan	*Pongo pygmaeus*	Pp	1	1
Gibbon	*Nomascus leucogenys*	Nl	1	1
Snub-nosed monkey	*Rhinopithecus roxellana*	Rr	2(1)	3
Rhesus macaque	*Macaca mulatta*	Mmu	1	1
Squirrel monkey	*Saimiri bolivienis*	Sb	1	1
Marmoset	*Callithrix jacchus*	Cj	1	1
Mouse	*Mus musculus*	Mm	1	1
Rat	*Rattus norvegicus*	Rn	3	3
Hamster	*Cricetulus griseus*	Cg	1	1
Guinea pig	*Cavia porcellus*	Cp	3(1)	4
Naked mole rat	*Heterocephalus glaber*	Hg	1(4)	5
Rabbit	*Oryctolagus cuniculus*	Oc	1	1
African elephant	*Loxodonta africana*	La	1(2)	3
Sea cow	*Trichechus manatus latirostris*	Tml	(1)	1
Opossum	*Monodelphis domestica*	Md	1	1
Platypus	*Ornithorhynchus anatinus*	Oa	0	0

NOTE.—Number of pseudogenes are shown in parenthesis, and incomplete genes are listed in square bracket.

genomes, including 23 species first reported and three species (cow, giant panda, and African elephant) corrected from Goo and Cho (2013) (supplementary data set and supplementary table S6, Supplementary Material online), whereas the rest of the 31 sequences were from 17 species, referenced from Goo and Cho (2013). Except for the platypus, other 42 mammalian species have at least one intact *RNASE1* gene copy. Among these 42 species, 28 have a single-copy *RNASE1* gene, and the remaining 14 species ranged from two to seven copies (table 4). Interestingly, compared with their close terrestrial ruminant relatives that have two to three copies of *RNASE1* gene, only a single copy is maintained in the seven cetacean species examined here, suggesting that the *RNASE1* copies have been lost during cetacean evolution.

Phylogenetic Inferences of Mammalian RNASE1 Genes

Similar tree topologies were generated when phylogenetic trees were reconstructed using three different methods (Bayesian inference, BI; neighbor joining, NJ; and maximum parsimony, MP) with all the *RNASE1* genes (supplementary fig. S3, Supplementary Material online) or exception of incomplete and 14 pseudogenes (supplementary fig. S4, Supplementary Material online). The relevant phylogenetic positions of the Chiroptera, Perissodactyla, and Proboscidea showed slightly incongruence in the three phylogenetic trees. Only the BI trees were chosen in the present study (supplementary figs. S3 and S4, Supplementary Material online) since they showed a major congruence with the currently well-recognized mammalian phylogeny. It was found that *RNASE1* copies of different species formed respective monophyletic groups. From the phylogenetic tree, a total of nine

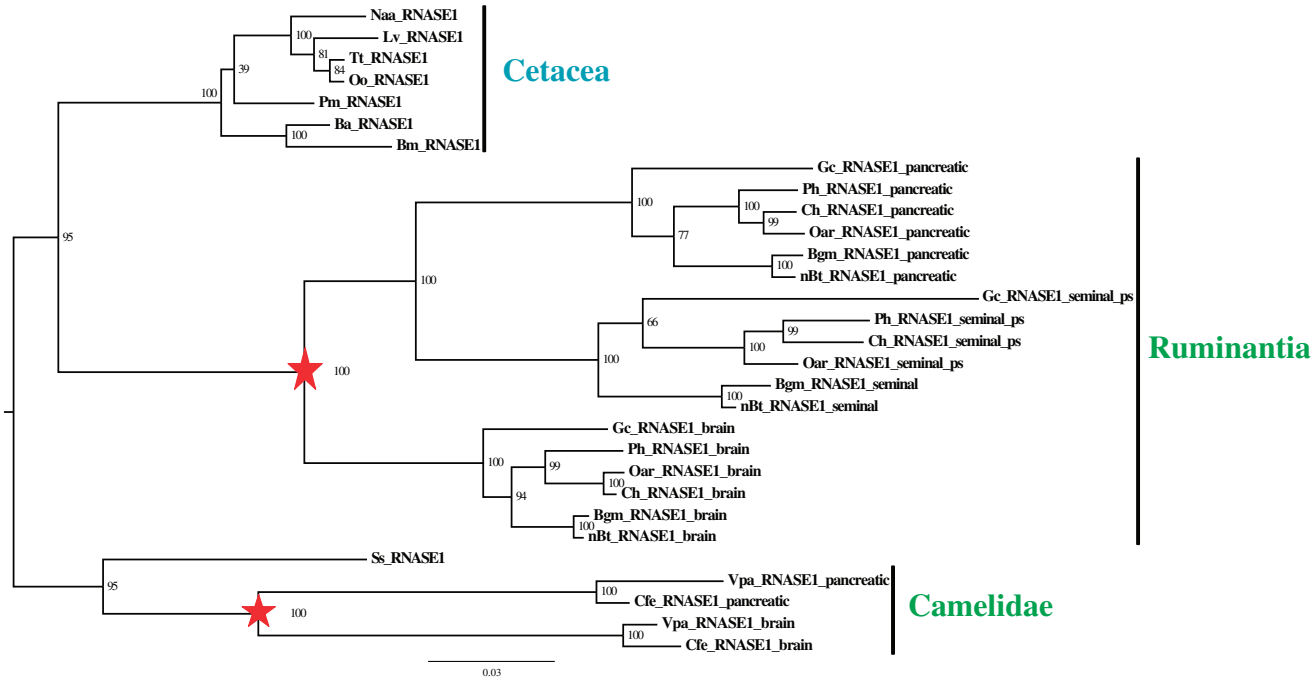

Fig. 2. Phylogenetic tree of cetartiodactylan *RNASE1* genes, inferred by Bayesian inference under the K80 + gamma model. Posterior probabilities values are indicated at the nodes. Gene duplication events are indicated by red five-pointed star.

species-specific duplication of *RNASE1* were identified, including duplications in the rat, the little brown bat, the guinea pig, the naked mole rat, the African elephant (Goo and Cho 2013), the Colobinae (Zhang et al. 2002), the Mustelidae (Liu et al. 2014), the Bovidae and the Camelidae. This suggested strongly that duplication events occurred after the divergence from the last common ancestor of mammals (supplementary figs. S3 and S4, Supplementary Material online), and that multiple events of gene duplication and pseudogenization of *RNASE1* have independently occurred during the mammalian evolutionary history.

Phylogenetic trees for the multiple *RNASE1* sequences from Cetartiodactyla showed that they clustered within family-specific clusters with well nodal supports (Bovidae BS = 100%, Camelidae BS = 98% in supplementary fig. S3, Supplementary Material online). Furthermore, to identify whether a gene duplication event took place in the ancestor of ruminantia, we added *RNASE1* sequences from the giraffe (*Giraffa camelopardalis*) to reconstruct the ancestral nodes of the Cetartiodactylan phylogeny. The resultant phylogenetic tree favored independent duplication patterns in ruminantia and camelidae (fig. 2), supporting the hypothesis that they arose from a common ruminantia ancestor through ancient duplication events which is consistent with a previously reported study (Confalone et al. 1995).

Selective Pattern of RNASE1 Genes in Mammals

To examine whether some branches leading to gene duplication events in *RNASE1* have been positively selected, we performed LRTs based on the branch-site models. Of the nine branches leading to the gene duplication events (branches *a*, *d*, *h*, *i*, *k*, *m*, *n*, *o*, *q* in fig. 3), signals of positive selection were detected only in branches *a* (Bovidae, $P < 0.001$) and *d* (Camelidae, $P < 0.001$), the former having two positively selected residues (37 and 160) and the latter having six residues (6, 8, 14, 46, 105, and 134) under positive selection (table 5 and fig. 3). Furthermore, significant evidence of positive selection was also detected along the ancestral branch of cetartiodactyls (branch *e* in fig. 3, $P < 0.05$), with one positively selected site (160, PP = 0.916) identified using BEB analysis (table 5 and fig. 3).

Isoelectric Points (pI) and Functional Assays

Twelve amino acid changes were uniquely detected in cetacean *RNASE1* when compared with ruminantia pancreatic *RNASE1*, 10 of which had radical effects on their physicochemical properties (charge, polarity, and volume) (supplementary table S7, Supplementary Material online). RNA is negatively charged, thus the net charge of *RNASE1* could influence its catalytic performance and also its interaction with substrates. The pI of a protein can be determined by its amino acid composition and is involved with the charge of protein. The computed pI for herbivorous pancreatic *RNASE1* (Camelidae and Bovidae, mean pI = 7.771) is significantly lower than the *RNASE1* of the cetaceans (mean pI = 8.553) ($t = 2.7458$, df = 6.861, $P = 0.029$, supplementary table S8, Supplementary Material online). To test the optimal pH of *RNASE1* and its relationship with catalytic activity and function, we prepared recombinant RNASE1 protein from baiji, common minke whale (*Balaenoptera acutorostrata*), common bottlenose dolphin (*Tursiops truncatus*), as well as protein from the pancreatic *RNASE1* gene of cow. Their ribonucleolytic activity was then examined at different

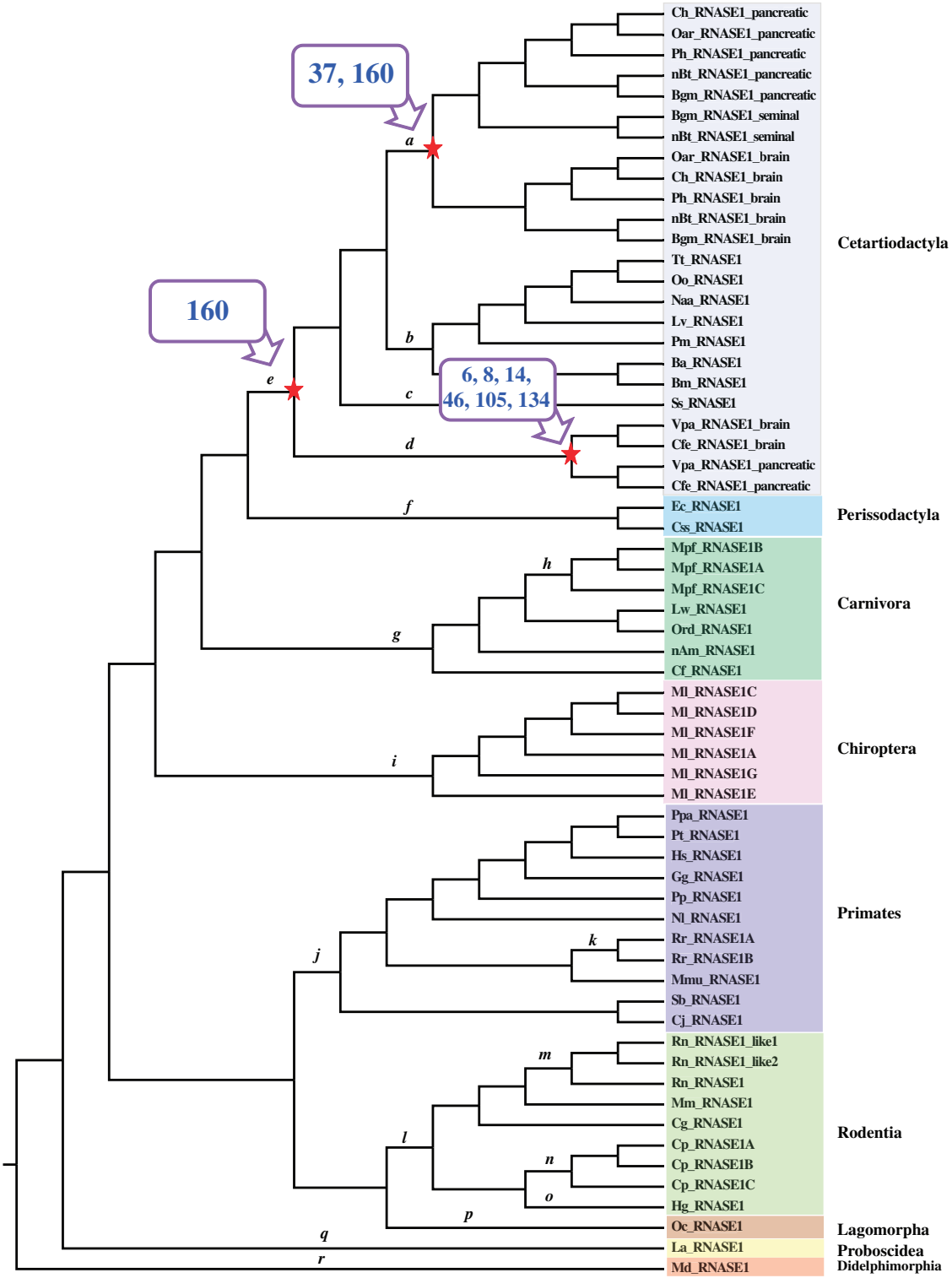

Fig. 3. Phylogenetic tree of putative functional *RNASE1* sequences. Tree topologies of Laurasiatheria and Primate were from Zhou *et al.* (2012) and Perelman *et al.* (2011), respectively. Branches *a*–*u* in the tree are used in the branch-site model tests. Positively selected branches and sites identified in branch-site model are indicated.

pH levels in a standard *RNASE1* assay against yeast tRNA. It was found that the optimal pH of cow pancreatic *RNASE1* ranged from 6.8 to 8.4 (fig. 4A), whereas those of bottlenose dolphin *RNASE1* ranged from 7.2 to 10.0, baiji *RNASE1* ranged from 6.8 to 10.0, and minke whale *RNASE1* ranged from 6.4 to 10.0 (fig. 4B).

Discussion

Enhanced Digestion of Lipids and Proteins in Cetaceans

Although cetaceans have multi-chambered stomachs which is similar to the ruminant mammals (Herwig et al. 1984), they

Table 5. Analysis of Selective Pressure Acting on *RNASE1*.

Models	lnL	2lnL	p level	Parameters	Positive sites
Heterocephalus glaber					
ma	−8776.623104			$\omega0$=0.135 $\omega1$=1.0 $\omega2$=4.954	
ma0	−8777.074367	0.902526	0.342105	$\omega0$=0.135 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Cetaceans					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Rhinopithecus roxellana					
ma	−8775.637815			$\omega0$=0.134 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8775.637815	0	1	$\omega0$=0.134 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Camelidae					
ma	−8770.492166			$\omega0$=0.143 $\omega1$=1.0 $\omega2$=999.0	6(0.954) 8(0.583) 14(0.502) 46(0.652) 105(0.533) 134(0.545)
ma0	−8777.588463	14.19259	0.000165	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Hystricomorpha					
ma	−8776.673059			$\omega0$=0.142 $\omega1$=1.0 $\omega2$=13.188	
ma0	−8777.587417	1.828716	0.176279	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Bovidae					
ma	−8770.8902			$\omega0$=0.133 $\omega1$=1.0 $\omega2$=999.0	37(0.729) 160(0.937)
ma0	−8777.588463	13.39653	0.000252	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Rattus norvegicus					
ma	−8775.606673			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=14.467	
ma0	−8777.356705	3.500064	0.061366	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Perissodactyla					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Cetartiodactyla					
ma	−8772.969839			$\omega0$=0.134 $\omega1$=1.0 $\omega2$=999.0	160(0.916)
ma0	−8777.588463	9.237248	0.002371	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Monodelphis domestica					
ma	−8767.799709			$\omega0$=0.125 $\omega1$=1.0 $\omega2$=242.342	
ma0	−8774.366106	13.13279	0.00029	$\omega0$=0.127 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Primates					
ma	−8774.919114			$\omega0$=0.135 $\omega1$=1.0 $\omega2$=39.317	
ma0	−8776.727856	3.617484	0.057175	$\omega0$=0.137 $\omega1$=1.0 $\omega2$=1.0	
Myotis lucifugus					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Sus scrofa					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Loxodonta africana					
ma	−8774.450622			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=35.0	
ma0	−8775.910746	2.920248	0.087474	$\omega0$=0.14 $\omega1$=1.0 $\omega2$=1.0	
Oryctolagus cuniculus					
ma	−8773.98303			$\omega0$=0.13 $\omega1$=1.0 $\omega2$=4.549	
ma0	−8775.006927	2.047794	0.152427	$\omega0$=0.126 $\omega1$=1.0 $\omega2$=1.0	
Cavia porcellus					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Carnivora					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Mustela putorius furo					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.588463	0	1	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
Ancestral Rodentia					
ma	−8777.588463			$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	
ma0	−8777.58853	0.000134	0.990764	$\omega0$=0.139 $\omega1$=1.0 $\omega2$=1.0	

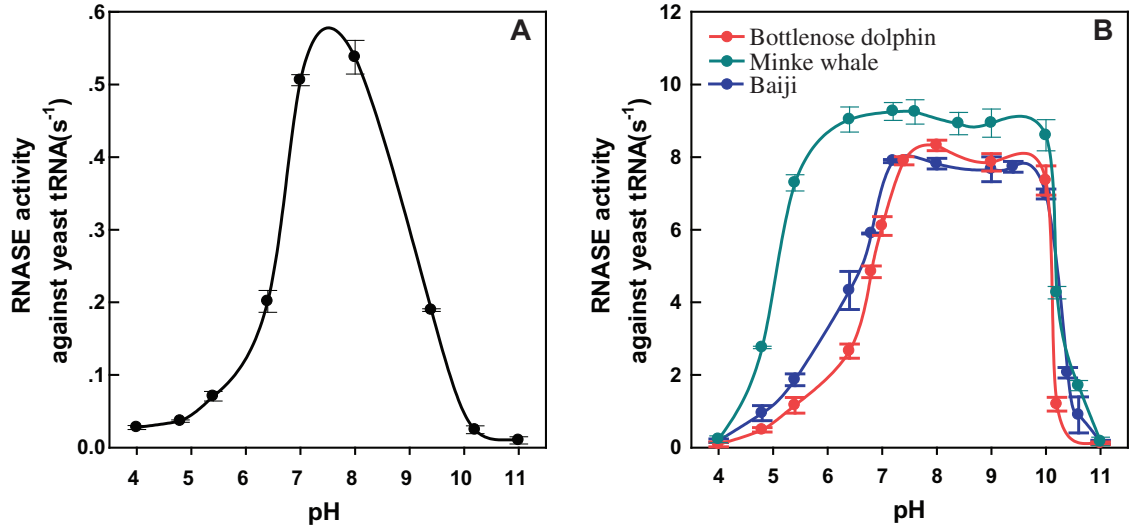

FIG. 4. RNASE1 activities against yeast tRNA at different pHs. (*A*) Recombinant proteins of cow pancreatic RNASE1. (*B*) Recombinant proteins of bottlenose dolphin, baiji and minke whale RNASEs. The mean enzyme activities from three replicates and the associated 95% confidence intervals (bars) are given.

have dramatically changed their feeding habits from a predominantly herbivorous to a carnivorous diet, the latter of which is mainly composed of fishes, large squids, and zooplanktons (Pauly et al. 1998). However, the genetic basis underlying such dietary switch remains poorly explored.

Digestive enzymes play a predominant role in digestion and they must adapt to the diet when substrate (food) intake is altered (Corring 1980). Along with the transition from land to sea, cetaceans had to change their diet from feeding on a vegetarian diet to preying mainly on fishes and aquatic invertebrates. Consequently, not only did lipids and proteins become the major nutrition composition of cetacean foods, but digestion of animal foods (as well as absorption of nutrition from lipids and proteins) became an important challenge for the survival of cetaceans in sea. The present analysis of positive selection of proteinase and lipase genes could potentially permit some predictions to be made whether cetaceans adapted to be able to digest lipids and proteins during the dietary switch.

Proteases, having an enhanced level of specificity towards proteins or polypeptide chains, are very important enzymes that are mainly located at three different sites in the digestive tract; stomach (PGC), pancreas (CTRC and PRSS1) and intestine (TMPRSS15) (Infante and Cahu 2007). The pancreas is the major source of proteases in the digestive system for breaking down ingested proteins and the proteases play a central role in the digestive process (Whitcomb and Lowe 2007). Of the three positively selected protease genes that were identified, CTRC and PRSS1 enzymes are secreted from the pancreas: CTRC is a proteolytic regulator of trypsinogen auto activation which cleaves the trypsinogen activation peptide to stimulate auto activation (Németh et al. 2013), and PRSS1 enzyme cleaves internal bonds involving the carboxyl group of lysine or arginine and activates other pancreatic proenzymes (Whitcomb and Lowe 2007). The third positively selected protease gene from the intestine is the serine proteinase, *TMPRSS15* (also known as *PRSS7*) encoding enterokinase, that activates trypsinogen to trypsin which in turn activates other proenzymes including chymotrypsinogen, procarboxypeptidases, and proelastases (Imamura and Kitamoto 2003). Functional analysis of the TMPRSS15 variants showed enterokinase deficiency, and cause a malabsorption disorder characterized by diarrhea and failure to thrive (Szczałuba et al. 2015). When we mapped these positively selected residues onto the three-dimensional crystal structure of these genes, it was found *CTRC* (T5S) and *PRSS1* (A13V) each had a positively selected site located on the signal peptide (supplementary table S4, Supplementary Material online), which directs the protein to the endoplasmic reticulum and associated with the secretion of the mature peptide (Király et al. 2007). In particular, codon 28 (Y28H) of *PRSS1* was situated in activation peptide, which is hydrolyzed as the first step in the activation process. Moreover, most of positively selected amino acids in proteases were located primarily in functional domains that facilitated ligand-receptor interactions (supplementary table S4, Supplementary Material online). That is to say, these proteases should have adaptively modified to recognize and bind novel substrate, and increased the digestibility of protein. Thus, these positively selected amino acids in the key residues of proteinase may have positive impact on the cetaceans protein absorbed. The detection of positive selection with these proteinase genes suggested that cetaceans may have acquired an enhanced capacity for proteins digestion and in turn, provide clues to our understanding of the dietary switch in cetacean.

Lipases (triacylglycerol acyl hydrolase) perform a key function in dietary fat absorption by hydrolyzing triglycerides into diglycerides and subsequently into monoglycerides and free fatty acids (Winkler et al. 1990). Three lipase genes (*LIPF*, *PNLIP*, and *CYP7A1*) were found to be under positive selection in cetaceans (table 2). Among these three genes, LIPF, agastric lipase, is able to hydrolyze both short- and long-chain

triacylglycerols at comparable rates and initiates the hydrolysis of alimentary triacylglycerols in the stomach (Gargouri et al. 1988), whereas the pancreatic lipase, PNLIP is an exocrine secreted enzyme that hydrolyzes dietary triglycerides in the small intestine (Hegele et al. 2001). In contrast to *LIPF* and *PNLIP*, *CYP7A1*, encoding for the enzyme cholesterol 7α-hydroxylase and promoting the digestion of lipids, is the first and rate-limiting enzyme of the classical pathway that accounts for the majority of total bile-acid synthesis. Bile acids are important for lipid digestion and essential for the formation of mixed micelles in the small intestine that facilitate solubilization, digestion, and absorption of dietary lipids and fat-soluble vitamins (De Aguiar Vallim et al. 2013). Human *CYP7A1* deficiency would cause decrease in bile acid production and accumulation of cholesterol in the liver, leading to down regulation of LDL receptors and consequent hypercholesterolemia (Pullinger et al. 2002). The findings of positive selection at these three genes suggested that cetaceans have evolved a strengthened ability not only to hydrolyze triacylglycerol chain but to improve the solubilization of lipids. Notably, codon 10 (I10A) of *CYP7A1* localized within predicted helical transmembrane segments, and one positively selected codon (N341H) in *LIPF* is active site, which performs the 2-fold function of binding the substrate and catalyzing the reaction (Schechter and Berger 1967). Those replacements contribute functional divergence by affecting enzyme activities or substrate binding in an indirect or complementary way, and these radical change sites may promote the enzymatic reactions of lipolytic. The evidence presented here provides support that these lipase genes have important roles in enhancing digestion of lipids, and may in part be responsible for the shift in diet during cetacean evolution.

The evidence of adaptive evolution of digestive enzymes in cetaceans can also be supported by the parallel or convergent evolution with carnivores, another group of specialized mammalian predators preying on other animals. This suggestion was further corroborated by the fact that four parallel/convergent nonsynonymous amino acid substitutions were identified between both carnivorous clades in proteinase (*CTRC*) and lipases (*CYP7A1* and *PNLIP*) (table 3). In addition, previous functional assays indicated that parallel/convergent amino acid changes are responsible for parallel/convergent functional changes (Yokoyama and Radlwimmer 1998; Zhang 2006). As well known, both carnivorous lineages (cetaceans and carnivores) are mainly dependent on their digestive enzymes to digest their meat based food, although cetaceans have multi-chambered stomachs whereas carnivores have single stomach. Accordingly, parallel/convergent amino acid substitutions identified in digestive enzymes of both carnivorous clades suggested that both groups might have evolved a similar mechanism in terms of enhanced digestion of proteins and lipids, and to some degree, in response to the pressure from similar carnivorous diet. Of course, further functional assays are needed to affirm this association in the future.

In the present study, digestive enzymes were determined to have undergone positive selection not only in cetaceans and carnivores, but also in chiropterans, lagomorphs and perissodactylans. Chiropterans, as omnivorous, has a huge diversity in their diets, including insects, other arthropods, fish, reptiles, amphibians, mammals, birds, fruits, flowers, nectar, pollen, foliage, and blood (Altringham and Altringham 1999). Around 70% of chiropteran species are insectivorous, and some of these are also carnivorous or piscivorous (Zhao et al. 2010). Thus, lipases and proteases subject to positive selection in chiropterans may be related with their complex dietary and the capabilities of digesting proteins and lipids. By contrast, positive selection identified in herbivorous (lagomorphans and perissodactylans) seems to be difficult to understand. Actually, herbivorous mammals can be divided into two groups, ruminants and nonruminants, according to their different digestive process. It is well known that ruminants (such as cattle, sheep, goats, and buffalo), usually with a four-compartment stomach, can ferment plant-based food in a specialized stomach prior to digestion principally through microbial actions. However, nonruminant herbivorous mammals (such as horses and rabbits), also called "monogastrics", acquire nutrients only by digestion through endogenous enzymes (De Caro et al. 2008). In addition, it has been reported that the capacity of digesting fats is well developed in rabbits from birth, as lipids in the milk (10–25% on fresh basis) are their main source of energy supply (Gidenne and Fortun-Lamothe 2002), and the digestibility of vegetable oil in rabbits is as much as 95% (Lanari 1975). Thus, lipases and proteases in monogastric herbivorous mammals actually act as an important role in the digestion of plant protein and lipid. For this reason, it is not unreasonable to identify some positively selected lipases and proteases genes in these monogastric herbivores. Accordingly, this may also reflect the considerable difference in the digestive process between monogastric and ruminant animals. This was also evidenced by the fact that similar with carnivorous animals, nonruminant herbivores such as rabbits and horses have only a single-copy *RNASE1* gene whereas gene duplications were identified in ruminants. More positively selected genes and radical specific amino acid substitutions (supplementary table S5, Supplementary Material online) were also detected in digestive enzyme genes in cetaceans when compared with plant-feeding herbivores. The stronger selective intensity identified in cetaceans than that of other plant-feeding animals might be in response to greater pressure for digestion of protein and lipid during dietary switch in cetaceans. Of course, further studies should be conducted to focus on this interesting phenomenon in the future. In summary, positive selection at proteinase and lipase genes, and the convergent evolution with other carnivorous mammals suggest that cetaceans have evolved an enhanced ability for digestion of proteins and lipids, which is critical for their dietary switch from herbivorous to carnivorous.

Loss of Pancreatic *RNASE1* Gene Associated with Cetacean Dietary Switch?

The association between food habit and *RNASE1* gene evolution has been discussed in some other mammals (Zhang et al. 2002; Zhang 2006; Yu et al. 2010; Zhou et al. 2014). Gene duplications of this gene have been found in many mammals,

and gene duplications and diversified selection after duplication were believed to be correlated with the plant-feeding adaptation of foregut-fermenting herbivores; for example, sheep, camel, leaf-eating colobines (Zhang et al. 2002) and elephants (Dubois et al. 2003).

In the present study, two to three copies of *RNASE1* gene were identified in two lineages of the order Cetartiodactyla, Ruminantia, and Camelidae, which supported the gene duplication events that occurred during the evolution of ancestral ruminantia (Confalone et al. 1995). This result is congruent with the hypothesis that gene duplication of *RNASE1* gene is associated with plant-feeding adaptation as revealed in colobines (Zhang et al. 2002; Zhang 2006). However, compared with their terrestrial relatives such as Camelidae and Bovidae, extant cetaceans reserved one *RNASE1* gene, whereas another copies of *RNASE1* disappeared during evolution, which was suggested to be related with their dietary change from herbivorous to carnivorous. This suggestion was further corroborated by functional assays that revealed cetacean *RNASE1* had higher isoelectric point values and optimal pH than those of their terrestrial relatives, that is, plant-feeding cows.

It was now well accepted that pancreatic *RNASE1* in the ruminants and other ruminant-like plant-eating mammals is in response to digesting high concentrations of bacterial RNA derived from the symbiotic microflora in the stomach required to recover nutrients (Zhou et al. 2014). Previous studies also found the lower pI and optimal pH of pancreatic *RNASE1* were necessary for its catalytic activity in degrading dietary RNA (Zhang et al. 2002; Zhang 2006; Zhou et al. 2014). However, it was noted that cetaceans had an intact *RNASE1* although lost some copies during evolution, which may be related to a lot of RNAs in their meat diet.

Conclusions

The present study represents the first systematic survey of the molecular genetic basis underlying the dietary switch from herbivorous to carnivorous in cetaceans. Strong signals of positive selection were detected in genes encoding proteinases (i.e., *CTRC*, *PRSS1*, and *TMPRSS15*) and lipases (i.e., *CYP7A1*, *LIPF*, and *PNLIP*), suggesting that cetaceans have evolved an enhanced ability of digesting proteins and lipids in response to their new feeding habits of preying mainly on animal foods. Additionally, it is interesting to note that the pancreatic *RNASE1* copy with digestive functions in plant-eating mammals has been evolutionarily lost in cetaceans, which is congruent with the dietary change from herbivorous to carnivorous.

Methods

Candidate Digestive Enzyme Genes

Many signal pathways and genes are involved in lipid and protein metabolism. For example, three major lipases, LIPC, LIPF, and PNLIP and the rate limiting enzyme of bile acid synthesis, CYP7A1, have been shown to play an essential role in digestion, solubilization and absorption of the lipids (De Aguiar Vallim et al. 2013; Sitrin 2014b; Wang et al. 2015). Four representative proteinases, CTRC, PGC, PRSS1, and TMPRSS15 have been suggested to play important roles in protein digestion (Giovanni and Fairlie 2005). The glucose transporters, SLC5A1 help in the absorption of monosaccharides, SI is thought to function during the final stages of carbohydrate digestion. These 10 genes (i.e., *CYP7A1*, *CTRC*, *LIPC*, *LIPF*, *PNLIP*, *PGC*, *PRSS1*, *SI*, *SLC5A1*, and *TMPRSS15*), with structure and function having been well studied and being considered for the key enzymes in digesting lipid, protein and monosaccharide, were chosen as are therefore important candidates for exploring the molecular basis of cetacean dietary switch.

Amplification and Sequencing of Cetacean Digestive Enzyme Genes

The genomes of five cetacean species (one mysticetes and four odontocetes); Omura's baleen whale (*Balaenoptera omurai*), finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*), long-beaked common dolphin (*Delphinus capensis*), beluga (*Delphinapterus leucas*), and Chinese white dolphins (*Sousa chinensis*) were sequenced during this study. All cetacean samples used in the study were collected from dead individuals in the wild and sampling was conducted systematically in accordance with all the ethical guidelines and legal requirements in China. The protocol of this study was approved by the Institutional Review Board of Nanjing Normal University (NNU). Voucher specimens were preserved at Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University (NNU), China.

Total genomic DNA was extracted from muscle with a standard phenol/chloroform procedure followed by ethanol precipitation (Green and Sambrook 2012). The DNA integrity was checked by 1% agarose gel electrophoresis. Primers were designed for the conserved regions based on an alignment of genomic data from the cow (*Bos taurus*) (http://asia.ensembl.org/Bos_taurus/Info/Index) and bottlenose dolphin (http://asia.ensembl.org/Tursiops_truncatus/Info/Index). All PCR amplifications were conducted using a Bio RAD PTC-200 with 2× Easy Taq PCR Super Mix (Trans Gen Biotech) and the thermo cycling profile were 34 cycles at 94 °C for 5 min, 94 °C for 30 s, 53–59 °C for 30 s, and 72 °C for 30 s, followed by a 10-min extension at 72 °C. The amplified PCR products were purified and sequenced in both directions using an ABI 3730 automated genetic analyzer. The specificity of these newly generated sequences was examined by comparing with the published nucleotide database at GenBank by BLAST (NCBI).

Source of Data and Primary Treatments

Five digestive enzyme genes, *CTRC*, *LIPC*, *LIPF*, *PGC*, and *PNLIP* were successfully sequenced from five cetacean species mentioned above and these new sequences were deposited in GenBank under accession numbers KX057380–KX057403. Preliminary alignment of the sequences showed no frameshift mutations or premature stop codons in cetaceans. The exons of each gene were sequenced and concatenated before being analyzed together. Only high-quality and high-integrity

sequences were used in the subsequent analyses. In addition, gene sequence from bottlenose dolphin were searched and primarily downloaded from the OrthoMaM (Ranwez et al. 2007). Sequences derived from the genomes of baiji, common minke whale, killer whale (*Orcinus orca*), sperm whale, seal (*Leptonychotes weddellii*), walrus (*Odobenus rosmarus divergens*), manatee (*Trichechus manatus latirostris*), and bowhead whale (*Balaena mysticetus*) (http://www.bowhead-whale.org/) were also used to test for positive Darwinian selection. A total of 10 digestive enzyme genes; *CYP7A1*, *CTRC*, *LIPC*, *LIPF*, *PNLIP*, *PGC*, *PRSS1*, *SI*, *SLC5A1*, and *TMPRSS15*, as well as 34–44 species from representatives mammalian lineages (i.e., cetaceans, cetartiodactyla, perissodactyla, carnivora, chiroptera, primates, lagomorpha, and rodentia) were analyzed in the present study (supplementary tables S1 and S2, Supplementary Material online).

Identification of *RNASE1* Genes

All currently available human and cow *RNASE1* sequences were used as queries to run TBLASTN and BLASTN searches in the 26 completed mammalian genome sequence databases available at the National Center for Biotechnology Information (NCBI) and the Ensembl genome database. The assembly versions and the coverage number of the genomes used in this study are listed in supplementary table S6, Supplementary Material online. We used 10^{-10} as the *E* value cutoff in all TBLASTN and BLASTN searches. As previously reported (Confaloneet al. 1995), three *RNASE1* genes isolated from different tissues of bovine were named pancreatic, brain, and seminal *RNASE1*s, respectively. Here, we report *RNASE1* sequences that demonstrate the existence of the orthologues of bovine pancreatic, seminal, and cerebral ribonucleases coding sequences in the genomes of other ruminant (e.g., sheep, goat, tibetan antelope, and wild yak). So, orthologous *RNASE1* copies in ruminantia were marked with tissue suffix "pancreatic", "brain" and "seminal", respectively. The nucleotide and amino acid sequences of all the *RNASE1* genes are provided in supplementary data set, Supplementary Material online, and the genome location information of *RNASE1* genes list in supplementary table S9, Supplementary Material online.

Sequence Alignment and Evolutionary Analysis

Nucleotide sequences of each gene and their deduced amino acid sequences were aligned separately using MUSCLE 3.8 (Edgar 2004) and MEGA 5.0 (Tamura et al. 2011) and then manually adjusted with GeneDoc. The RNASE1 phylogenetic trees were reconstructed using MEGA 5.0 with neighbor-joining (NJ) and maximum parsimony (MP) methods. In the NJ analysis, the Kimura 2-parameter model (K2P) with complete deletion option for gaps was used. In the MP analysis, a heuristic search strategy was employed with the tree-bisection-reconnection branch-swapping algorithm, random addition of taxa and 1,000 replicates per search. The reliability of the tree topologies was evaluated using bootstrap support (BS) (Felsenstein 1985) with 2,000 replicates for NJ and MP analyses.

Bayesian inference (BI) with Markov-chain Monte Carlo (MCMC) sampling was performed using MrBayes 3.2.6 (Huelsenbeck and Ronquist 2001) run for one million generations. We made two simultaneous runs, sampling trees every 1,000 generations, with three heated and one cold chain to encourage swapping among the MCMC chains and to avoid the analysis remaining in local rather than global optima. ModelTest (Darriba et al. 2012) was used to select the optimal models based on the Akaike Information Criterion (AIC). Convergence of sampled parameters and potential autocorrelation (effective sampling size/ESS for all parameters >200) was investigated in Tracer 1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Additionally, the average SD of split frequencies between both runs was checked (<0.01). The Bayesian posterior probabilities were obtained from the 50% majority rule consensus of the post-burn-in trees sampled at stationarity, after removing the first 25% of trees as a "burn-in" stage.

The codon-based maximum likelihood models implemented in CODEML program in PAML 4.7 (Yang 2007) were applied to estimate the rates of synonymous (d_S) and nonsynonymous substitutions (d_N), as well as d_N/d_S ratio (omega, ω). The nonsynonymous to synonymous rate ratio ω indicates changes in selective pressures, where $\omega = 1$, $\omega < 1$, and $\omega > 1$ correspond to neutral evolution, purifying, and positive selection, respectively. The well-supported phylogeny of Laurasiatheria (Zhou et al. 2012) and Primates (Perelman et al. 2011) was used as the input tree in all analyses (Tree file: supplementary fig. S1, Supplementary Material online).

To identify the probabilities of sites under positive selection in each gene for the mammals and cetacean species being examined, site models were implemented where ω could vary among sites. All the positively selected sites in site models were identified by using Bayes Empirical Bayes (BEB) analysis (Yang 2007) with posterior probabilities of ≥ 0.80. Positive selection was also detected using branch-site model A, in which ω can vary among sites along specific lineages (Zhang et al. 2005). Modified branch-site model A (test 2) was performed for every gene in each foreground lineage, which facilitated the analysis of data sets including cetaceans only (digestive enzyme genes, branches *b*–*u* in fig. 1) and all mammals (*RNASE1* genes, branches *b*–*u* in fig. 3). The likelihood ratio test (LRT) statistic ($2\Delta L$) approximates to a chi-square distribution and was used to compare nested likelihood models. In addition, the improved statistical methods in Datamonkey web server (Pond and Frost 2005), which computed nonsynonymous and synonymous substitutions at each codon position, was used to further evaluate the selection. Sequences within each gene were analyzed in cetaceans-only data set by using two distinct models: namely, fixed-effect likelihood (FEL) and random effect likelihood (REL). Finally, in support for the PAML results, a complementary protein-level approach was used in the TreeSAAP (Woolley et al. 2003).

Identification of Parallel/Convergent Sites among Carnivorous Mammals

The parallel/convergent sites among carnivorous mammals were identified according to the methods previously described (Foote et al. 2015). Briefly, we reconstructed the

ancestral amino acids sequences for six positively selected genes using the Bayesian approach (Rate Ancestor = 1) implemented in the BASEML program from the PAML package (Yang 2007). For each of the two groups (cetaceans and carnivores), the extant sequences at each position were compared with the ancestral sequence at the node corresponding to the most recent ancestor. We used the software CONVERG 2 (Zhang and Kumar 1997) to test whether the number of observed parallel/convergent amino acid substitutions was significantly higher than that expected by chance, given the total number of amino acid replacements in the two evolutionary lineages under investigation.

Mapping of Positively Selected Sites onto Protein Structures

To gain insights into the functional significance of the putatively selected sites, we mapped the sites under positive selection to crystal structures (Zhang 2008). The protein sequences of positively selected genes were derived from the common bottlenose dolphin genome, which were obtained from the Ensembl genome database (http://www.ensembl.org/index.html). In addition, the functional information of genes identified as being under positive selection was derived from uniprot (http://www.uniprot.org/).

Recombinant Protein Preparation and Functional Assays

Change of physicochemical properties of amino acids were detected according to the methods previously described (Zhang 2000). Isoelectric points (pI) of mature peptides were computed by the web program Protein Calculator (http://protcalc.sourceforge.net/), and we selected the independent-samples t test for the groups differences. We subcloned the baiji, common minke whale, and common bottlenose dolphin *RNASE1* as well as the cow pancreatic RNASE1 into the bacterial expression vector, pET28a, and verified them by sequencing. We measured the RNASE1 activity of the recombinant proteins against yeast tRNA at different pHs (40 mM acetate buffer with pH 4.0–6.2 and 40 mM sodium phosphate buffer with pH 6.4–12) at 25°C for 10 min. Purified RNASE1 (4.5×10^{-2} nmol) was added to 0.8 ml of the aforementioned buffer with 1.6 nmol tRNA. The reaction was stopped by addition of 0.5 ml of an ice-cold fresh solution of 20 mM lanthanum nitrate with 3% perchloric acid. The time zero control was prepared by addition of stop solution to the phosphate and protein-containing reaction mixture prior to the addition of yeast tRNA. After being held on ice for 15 min, the solution was centrifuged for 10 min at 12,000×g to remove the insoluble tRNA. We determined the absorbance at 260 nm of the supernatant fraction as a measure of the amount of solubilized RNA. The catalytic activity of RNASE1 was computed as nmol of RNA digested per second per nmol of RNASE1.

Supplementary Material

Supplementary figures S1–S4, tables S1–S9 and data set are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Science Fund for Distinguished Young Scholars (grant no. 31325025 to G.Y.), the National Key Programme of Research and Development, Ministry of Science and Technology (grant no. 2016YFC0503200 to G.Y. and S.X.), the National Natural Science Foundation of China (NSFC) (grant no. 31370401 to W.R. and 31570379 to S. X.), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) to G.Y. and S.X., the Natural Science Foundation of Jiangsu Province (grant no. BK20141449 to S.X.), and the Cultivation Plan for Excellent Doctorial Dissertations of Nanjing Normal University to Z.W., and the Jiangsu Provincial Innovation Project for Scientific Research of Graduate Students in Universities (grant no. KYLX_0708 to Z.W.). We are also grateful to Mr Xinrong Xu, Dr Anli Gao, and some students who have ever studied or are studying at NNU for their assistance in sample collection. We are indebted to Lin Zhang, Xu Niu and Guiting Li for support and discussions.

References

Agrawal VP, Sastry KV, Kaushab SK. 1974. Digestive enzymes of three teleost fishes. *Acta Physiol Acad Sci Hung* 46:93–98.

Altringham JD. 1999. Bats: biology and behavior. Oxford: Oxford University Press.

Axelsson E, Ratnakumar A, Arendt M, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar Å, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495:360–364.

Breukelman HJ, Jekel PA, Dubois JYF, Mulder PP, Warmels HW, Beintema JJ. 2001. Secretory ribonucleases in the primitive ruminant chevrotain (*Tragulus javanicus*). *Eur J Biochem*. 268:3890–3897.

Confalone E, Beintema JJ, Sasso MP, Carsana A, Palmieri M, Vento MT, Furial A. 1995. Molecular evolution of genes encoding ribonucleases in ruminant species. *J Mol Evol*. 41:850–858.

Corring T. 1980. The adaptation of digestive enzymes to the diet: its physiological significance. *Reprod Nutr Dev*. 20:1217–1235.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772.

De Aguiar Vallim TQ, Tarling EJ, Edwards PA. 2013. Pleiotropic roles of bile acids in metabolism. *Cell Metab*. 17:657–669.

De Caro J, Eydoux C, Chérif S, Lebrun R, Gargouri Y, Carrière F, De Caro A. 2008. Occurrence of pancreatic lipase-related protein-2 in various species and its relationship with herbivore diet. *Comp Biochem Physiol B Biochem Mol Biol*. 150:1–9.

Deng SP, Tabatabai MA. 1995. Cellulase activity of soils: effect of trace elements. *Soil Biol Biochem*. 27:977–979.

Dubois JF, Ursing BM, Kolkman JA, Beintema JJ. 2003. Molecular evolution of mammalian ribonucleases 1. *Mol Phylogenet Evol*. 27:453–463.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.

Foote AD, Liu Y, Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47:272–275.

Gargouri Y, Moreau H, Pieroni G, Verger R. 1988. Human gastric lipase: a sulfhydryl enzyme. *J Biol Chem*. 263:2159–2162.

Gidenne T, Fortun-Lamothe L. 2002. Feeding strategy for young rabbits around weaning: a review of digestive capacity and nutritional needs. *Anim Sci*. 75:169–184.

Giovanni A, Fairlie DP. 2005. Protease inhibitors in the clinic. *Med Chem*. 1:71–104.

Goo SM, Cho S. 2013. The expansion and functional diversification of the mammalian ribonuclease a superfamily epitomizes the efficiency of multigene families at generating biological novelty. *Genome Biol Evol*. 5:2124–2140.

Green MR, Sambrook J. 2012. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Hegele RA, Ramdath DD, Ban MR, Carruthers MN, Carrington CV, Cao H. 2001. Polymorphisms in PNLIP, encoding pancreatic lipase, and associations with metabolic traits. *J Hum Genet*. 46:320–324.

Herwig RP, Staley JT, Nerini MK, Braham HW. 1984. Baleen whales: preliminary evidence for fore stomach microbial fermentation. *J Appl Environ Microbiol*. 47:421–423.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Imamura T, Kitamoto Y. 2003. Expression of enteropeptidase in differentiated enterocytes, goblet cells, and the tumor cells in human duodenum. *Am J Physiol Gastrointest Liver Physiol*. 285: G1235–G1241.

Infante JZ, Cahu CL. 2007. Dietary modulation of some digestive enzymes and metabolic processes in developing marine fish: applications to diet formulation. *Aquaculture* 268:98–105.

Király O, Boulling A, Witt H, Le Maréchal C, Chen JM, Rosendahl J, Battaggia C, Wartmann T, SahinTóth M, Férec C. 2007. Signal peptide variants that impair secretion of pancreatic secretory trypsin inhibitor (SPINK1) cause autosomal dominant hereditary pancreatitis. *Hum Mutat*. 28:469–476.

Kleineidam RG, Pesole G, Breukelman HJ, Beintema JJ, Kastelein RA. 1999. Inclusion of cetaceans within the order Artiodactyla based on phylogenetic analysis of pancreatic ribonuclease genes. *J Mol Evol*. 48:360–368.

Lanari D. 1975. Nutrient requirements of rabbits. *Spun* 18:114–116.

Liu J, Wang XP, Cho S, Lim BK, Irwin DM, Ryder OA, Zhang YP, Yu L. 2014. Evolutionary and functional novelty of pancreatic ribonuclease: a study of Musteloidea (order Carnivora). *Sci Rep UK* 4:3520–3527.

McGowen MR, Gatesy J, Wildman DE. 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol Evol*. 29:336–346.

Németh BC, Wartmann T, Halangk W, Sahin-Tóth M. 2013. Autoactivation of mouse trypsinogens is regulated by chymotrypsin C via cleavage of the autolysis loop. *J Biol Chem*. 288:24049–24062.

Pauly D, Trites AW, Capuli E, Christensen V. 1998. Diet composition and trophic levels of marine mammals. *ICES J Mar Sci*. 55:467–481.

Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M, Rumpler Y. 2011. A molecular phylogeny of living primates. *PLoS Genet*. 7:e1001342.

Pond SLK, Frost SD. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

Pullinger CR, Eng C, Salen G, Shefer S, Batta AK, Erickson SK, Verhagen A, Rivera CR, Mulvihill SJ, Malloy MJ, et al. 2002. Human cholesterol 7α-hydroxylase (CYP7A1) deficiency has a hyper cholesterolemic phenotype. *J Clin Invest*. 110:109–117.

Ranwez V, Delsuc F, Ranwez S, Belkhir K, Tilak M, Douzery EJ. 2007. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol Biol*. 7:241.

Reidenberg JS. 2007. Anatomical adaptations of aquatic mammals. *Anat Rec* 290:507–513.

Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun*. 27:157–162.

Sitrin MD. 2014a. Digestion and absorption of carbohydrates and proteins. The gastrointestinal system. Netherlands: Springer. p. 137–158.

Sitrin MD. 2014b. Digestion and absorption of dietary triglycerides. The gastrointestinal system. Netherlands: Springer. p. 159–178.

Stickney RR, Shumway SE. 1974. Occurrence of cellulase activity in the stomachs of fishes. *J Fish Biol*. 6:779–790.

Szczałuba K, Nowakowska BA, Sobecka K, Smyk M, Castaneda J, Dudkiewicz Z, Kutkowska-Kaźmierczak A, Sąsiadek MM, Śmigiel R, Bocian E. 2015. High-Resolution Array Comparative Genomic Hybridization Utility in Polish Newborns with Isolated Cleft Lip and Palate. *Neonatology* 107:173–178.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28:2731–2739.

Trumble SJ, Barboza PS, Castellini MA. 2003. Digestive constraints on an aquatic carnivore: effects of feeding frequency and prey composition on harbor seals. *J Comp Physiol B*. 173:501–509.

Wang Z, Chen Z, Xu S, Ren W, Zhou K, Yang G. 2015. 'Obesity' is healthy for cetaceans? Evidence from pervasive positive selection in genes related to triacylglycerol metabolism. *Sci Rep UK*. 5:14187.

Whitcomb DC, Lowe ME. 2007. Human pancreatic digestive enzymes. *Digest Dis Sci*. 52:1–17.

Winkler FK, D'Arcy A, Hunziker W. 1990. Structure of human pancreatic lipase. *Nature* 343:771–774.

Wong H, Schotz MC. 2002. The lipase gene family. *J Lipid Res*. 43:993–999.

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671–672.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Yokoyama S, Radlwimmer FB. 1998. The "five-sites" rule and the evolution of red and green color vision in mammals. *Mol Biol Evol*. 15:560–567.

Yu L, Wang X, Jin W, Luan P, Ting N, Zhang Y. 2010. Adaptive evolution of digestive RNASE1 genes in leaf-eating monkeys revisited: new insights from ten additional colobines. *Mol Biol Evol*. 27:121–131.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol*. 50:56–68.

Zhang J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet*. 38:819–823.

Zhang J, Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol Biol Evol*. 14:527–536.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.

Zhang J, Zhang Y, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet*. 30:411–415.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.

Zhao H, Yang J, Xu H, Zhang J. 2010. Pseudogenization of the umami taste receptor gene Tas1r1 in the giant panda coincided with its dietary switch to bamboo. *Mol Biol Evol*. 27:2669–2673.

Zhao H, Zhou Y, Pinto CM, Charles-Dominique P, Galindo-González J, Zhang S, Zhang J. 2010. Evolution of the sweet taste receptor gene Tas1r2 in bats. *Mol Biol Evol*. 27:2642–2650.

Zhou X, Wang B, Pan Q, Zhang J, Kumar S, Sun X, Liu Z, Pan H, Lin Y, Liu G. 2014. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat Genet*. 46:1303–1310.

Zhou X, Xu S, Xu J, Chen B, Zhou K, Yang G. 2012. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the Laurasiatherian mammals. *Syst Biol*. 61:150–164.