Genetic Signatures of Lipid Metabolism Evolution in Cetacea

Yoshinori Endo,[1] Ken-ichiro Kamei,[2]* Miho Inoue-Murayama[1,3]*

[1]*Wildlife Research Center, Kyoto University, 2-24 Tanaka-Sekiden-cho, Sakyo-ku, Kyoto, 606-8203, Japan*

[2]*Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-Ushinomiya-cho, Sakyo-ku, Kyoto, 606-8501, Japan*

[3]*Wildlife Genome Collaborative Research Group, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan*

*Authors for Correspondence:

Ken-ichiro Kamei, Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto, Japan, +81-75-753-9774, +81-75-753-9761, kkamei@icems.kyoto-u.ac.jp

Miho Inoue-Murayama, Wildlife Research Center, Kyoto University, Kyoto, Japan, +81-75-771-4375, +81-75-771-4394, mmurayama@wrc.kyoto-u.ac.jp

Abstract

In mammalian evolutionary history, Cetacea (whales, dolphins, and porpoises) achieved astonishing success by adapting to an aquatic environment. One unique characteristic of Cetacea, contributing to this adaptive success, is efficient lipid utilization. Here we report comparative genetic analysis of aquatic and terrestrial Cetartiodactyla using 144 genes associated with lipid metabolism. We analysed genetic mutation rate, amino acid substitution, and metabolic pathways using genetic data publicly available. Our test detected 18 positively selected genes in Cetacea compared to 13 in Bovidae with little overlap between the lineages. There were lineage specific amino acid substitutions within functional domain regions of these genes. Moreover, a pathway analysis showed that the identified genes in Cetacea were associated with lipid digestion, lipid storage, and energy producing pathways. This study emphasizes the evolutionary context of lipid metabolism modification of Cetacea since the divergence from terrestrial ancestor. Our results provide a foundation for future studies of elucidating the adapted biological mechanisms of Cetacea lipid metabolism and a framework for incorporating ecological context into studies aimed at investigating adaptive evolution.

Key words: Cetacea, mammal, adaptation, positively selected genes, protein domain, lipid metabolism

1. Introduction

The evolution of Cetacea represents one of the most striking adaptations for a habitat transition in mammalian evolutionary history [1, 2]. Although mammals are generally under greater constraints with regard to anatomical, physiological, and behavioural changes than other classes of vertebrates, Cetacea has achieved a remarkable macroevolutionary transition since their divergence from an artiodactyl ancestor approximately 50 million years ago [1]. The past two decades of paleontological and phylogenetic research have thoroughly characterized the evolutionary history of Cetacea [1-3], and there continues to be considerable attention on the molecular evolution behind the dramatic changes in their morphology and ecology upon aquatic adaptation [4]. Understanding the molecular basis underlying the adaptive traits will shed light on this unique evolutionary event from land to water.

Cetacea has developed unique characteristics that are different from those of terrestrial mammals but are common within the lineage [5]. For example, all species of Cetacea are carnivores and have streamline body morphology. Cetacea consumes lipids not only as one of their primary energy resources, but also as an important component of the thick blubber that provides thermo insulation and supports their streamlined morphology [6]. The evolution of lipid usage may be a key factor for the adaptive characteristics enabling Cetacea's transition to an aquatic environment. Recent genomic studies have reported signatures of positive selection on diverse processes, including those of lipid metabolism [7-11]. However, evolutionary analyses of cetaceans have been with distantly related lineages, rather than contrasting them with closely related terrestrial species using the same analytical procedures. Furthermore, given the potential adaptive specificity of each Cetacea species to its own environment and ecology,

analyses at the taxonomic-group level are necessary to interpret the detected genetic signatures as aquatic adaptations.

In this study, we identified the genetic signatures of positive selection on lipid metabolism in Cetacea and Bovidae (figure 1*a*). The goals of our study were to (1) collect genes associated with lipid metabolism from Cetartiodactyla species (even-toed ungulates, including Cetacea), (2) identify the positively selected genes (PSGs) for each lineage, Cetacea and Bovidae, (3) determine the amino acid substitutions that occurred in functional domain regions of the identified PSG , and (4) describe the biological functions and metabolic pathways that have been modified during aquatic adaptation (figure 1*b*). Our comparative analyses revealed unique and a greater degree of signatures for positive selection in the Cetacea lipid metabolism pathways compared with those of Bovidae. This study provides the genetic basis for the evolution of lipid metabolism developed by Cetacea in the context of aquatic adaptation and emphasizes the importance of comparative analyses based on the phylogenetic and ecological perspectives of the study organisms.

2. Materials and Methods

(a) Data sampling and preparation

All the genetic data in this study were downloaded from NCBI (https://www.ncbi.nlm.nih.gov). Genomes of Artiodactyla species that were publicly available and annotated were used. Orthologous cording DNA sequences (CDSs) were identified by blastn search with human genes as references. Genes that were not annotated in any of the study species were excluded from later analyses. CDSs were aligned with PRANK v.140603 [12] and manually edited with Mesquite v.3.2 [13]. To

remove potentially unreliable sequences, the aligned CDSs were filtered with GUIDANCE2 using the default settings [14]. In addition, codon triplets with gaps in more than 50% of the sequences were discarded.

(b) Identification of positively selected genes

PSGs were identified using branch-site models implemented in the CODEML program in the PAML v. 4.8 software package [15]. The branch-site test detects positive selection on *a priori* specified branches. To detect selection in Cetacea and Bovidae separately, we conducted the tests twice for all genes evaluating one lineage at a time. A one-ratio model was employed as a null hypothesis in which all branches were under neutral evolution. A two-ratio model was conducted as an alternative hypothesis in which a specified lineage was under positive selection, and this model estimated the d_N/d_S ratio (ω: nonsynonymous to synonymous substitution ratio). Because the branch-site test estimates the d_N/d_S ratio on a subset of codons, the ω values are not reliable measure of the strength of positive selection [16]. Therefore, the likelihood ratio test was used to determine whether a gene was under positive selection. The likelihood ratio test was performed to compare $2\Delta l$ of the two models to the χ^2 distribution for *p*-value evaluation.

(c) Domain annotation and amino acid substitutions

The identified PSGs were scanned for domains with the pfam_scan utility and HMMER3.1 [17] against the Pfam-A from Pfam v. 31.0 [18]. The domain arrangements were visualized with DoMosaics v. 0.95 [19]. Domain similarity was estimated by using amino acid sequences of *Orcinus orca* and *Bos taurus* as criteria for Cetacea and

Bovidae PSGs, respectively, using DoMosaics, and the dot plots were generated using the R package ggplot2 [20]. The lineage specific amino acid substitutions were identified manually by looking at a conversion of residue that occurred in only one lineage and where all the species in that lineage had identical residue. Those instances were included when one of the outgroup species had a mutation at the same position.

(d) Function and pathway analysis of PSGs

Gene ontology (GO) terms of the identified PSGs were assigned according to DAVID Functional Annotation tool [21, 22] under the biological process domain. The pathways in which the identified PSGs participate were identified using KEGG Mapper [23]. Because a gene can be involved in multiple pathways, which can be categorized in multiple hierarchy, the pathways that was more representative and directly related with lipid metabolism were focused.

3. Results and Discussion

(a) Positively selected genes in Cetacea and Bovidae

To detect signatures of positive selection, we analysed the alignments of lipid metabolism-related genes collected from genomes of Cetartiodactyla. To assess the evolutionary transitions since their divergence from a terrestrial ancestor [1], we categorized the study species by their phylogeny and ecology in the target group (Cetacea), a control group (Bovidae), and an outgroup (Suina and Tylopoda). Among Ruminantia, the family Bovidae was used as a control group. Bovidae is a fully terrestrial and close clade of Cetacea [24]. The ecological and geographical diversity of Bovidae provides an unbiased representation of terrestrial adaptation. We opted to use

only one family, because there are few differences in the adaptive fitness of Bovidae to an aquatic environment. Thus, positive selection detected in Cetacea can suggest the adaptations required to return to an aquatic environment. To gain insight into the evolution of lipid usage, we investigated the genes associated with lipid metabolism. Taken together, the results from these datasets will elucidate plausible molecular adaptations underlying Cetacea macroevolutionary transitions from a terrestrial to an aquatic environment.

A total of 14 Cetartiodactyla species were used in this study: five Cetacea, including the minke whale (*Balaenoptera acutorostrata*), sperm whale (*Physeter catodon*), killer whale (*Orcinus orca*), bottlenose dolphin (*Tursiops truncatus*), and Yangtze River dolphin (*Lipotes vexillifer*); five Bovidae, including cattle (*Bos taurus*), wild yak (*Bos mutus*), American bison (*Bison bison*), domestic goat (*Capra hircus*), and domestic sheep (*Ovis aries*); and Suina and Tylopoda, including pig (*Sus scrofa*), alpaca (*Vicugna pacos*), Bactrian camel (*Camelus bactrianus*), and Arabian camel (*Camelus dromedaries*) (figure 1*c*). For each species we searched orthologous of the 156 genes assigned to functions associated with lipid metabolism in humans. Our filtering strategy retained 144 genes covering the major lipid metabolic pathways (electronic supplementary material, table S1).

To identify genes under positive selection, we performed likelihood ratio tests using branch-site models implemented in PAML [15]. This analysis estimates d_N/d_S ratio (ω) for the lineage of interest, and the PSGs were determined by the likelihood ratio test. Of the 144 lipid metabolism genes, 14 genes were identified as PSG in the lineage of Cetacea, as well as 5 genes in the lineage of Bovidae at $p < 0.05$ (figure 2*a*, table 1, electronic supplementary material, table S2 and S3). Due to the low number of

PSGs for assessing the selection landscape of metabolic pathways, we relaxed the significance level to $0.05 \leq p < 0.1$ and found an additional 4 and 8 PSGs in Cetacea and Bovidae, respectively. Among the identified PSGs, only one gene, *GPAM* (also known as *GPAT1*, glycerol-3-phosphate acyltransferase) was found in common at $p < 0.05$ (figure 2*b*).

To assess whether the detected amino acid substitutions have influence on functions of the identified PSGs, we examined the functional regions of the protein sequences, namely domains, as described in the Method section (figure 3 for *EHHADH*; for other PSGs see electronic supplementary material, figure S1-S31). Whereas the domain arrangements were largely conserved among Cetartiodactyla, the comparisons of the sequences of the domain region showed inter-lineage divergence between Cetacea and Bovidae. Furthermore, we explored amino acid substitutions that occurred only in Cetacea and remain identical within the lineage. Such lineage specific substitutions in the functional domain of genes, evolving under positive selection, would putatively be associated with adaptive phenotype in an aquatic environment. We found lineage specific substitutions in at least one domain of 14 Cetacea PSGs and 11 Bovidae PSGs (table 2 and figure 3*c*). For example, one Cetacea PSG, *EHHADH*, is composed of four domains, except for the domain arrangements of *Tursiops truncates* and *Sus scrofa*, probably because of isoforms (figure 3*a*). The amino acid sequence of each domain differs between Cetacea and Bovidae while they are relatively similar within each lineage (figure 4*b*). *EHHADH* encodes two enzymes, enoyl-CoA hydratase/isomerase (ECH) and 3-hydroxyacyl-CoA dehydrogenase (HCDH), involved in β-oxidation [25]. The amino acid sequences of both of these enzymes contains lineage specific substitutions (figure 4*c*), even though none of the substitutions were

found in active sites [26]. We did not find lineage specific substitutions in four Cetacea PSGs (*AGPAT*, *FBP1*, *GNPAT*, and *GPAM*) nor in two Bovidae PSGs (*GPAM* and *PGAM5*). However, substitutions outside of domains may still have functional influence.

Our comparison of the lipid metabolism genes highlighted greater and unique selective pressures in Cetacea lipid metabolism compared with Bovidae. The results showed that both lineages have modified their lipid metabolism since divergence from a common ancestor, suggesting the adaptive importance of the usage of lipids in both aquatic and terrestrial environments. The relatively higher number of PSGs with $p < 0.05$ Cetacea, however, indicates that evolutionary changes in lipid metabolism were more substantial for the aquatic than the terrestrial lineage. Moreover, the slight overlap between the Cetacea and Bovidae PSGs indicated the differing selection regimes in lipid metabolism of aquatic and terrestrial lineages. Our results (1) provide support for the adaptive evolution of lipid metabolism in the context of aquatic adaptation and (2) highlighted that these adaptations are not species specific for *T. truncatus* [7, 27] or other Cetacea species [8].

(b) Functional analysis of Cetacea PSGs

To gain insights into the functional consequences of the lipid metabolism behind the aquatic adaptation of Cetacea, we assessed the GO terms and metabolic pathways of the identified PSGs (figure 4).

Our functional analyses revealed that six Cetacea PSGs (*APOA2*, *APOB*, *CYP8B1*, *LIPF*, *PNLIP*, and *PNLIPRP2*) were involved in the fat digestion and absorption pathway associated with the lipid catabolic process (GO:0016042) and

lipoprotein biosynthetic process (GO: 0042158). *LIPF*, *PNLIP*, and *PNLIPRP2* encode pancreatic lipase, which hydrolyses dietary lipids in the digestive system [28, 29]. *CYP8B1* encodes a member of the cytochrome P450 superfamily of enzymes which catalyses the synthesis of primary bile acids. Bile acids play important roles not only in absorption of lipids but also in the regulation of lipid metabolism. Additionally, they are believed to function as a therapeutic approach for metabolic syndrome [30]. Lastly, *APOA2* and *APOB* encode apolipoproteins, which aid in transporting hydrophobic lipids through the circulation system and are associated with atherosclerotic cardiovascular diseases [31], implying enhanced tolerance for high lipid concentration in their circulatory systems. Previous studies have also reported some genes in this pathway under selection in Cetacea [7, 8, 10, 32]; moreover, owing to the explicit comparison with their herbivorous relatives, our results provide further support of genetic signatures in Cetacea adaptation in their shift in diet.

The storage of fat is an essential part of lipid metabolism, whereby triacylglycerol (TAG) is synthesized to serve as a major component in several biological functions. Synthesis of TAG takes place in glycerophospholipid metabolism via two distinct major pathways, the phosphatidic acid pathway and the monoacylglycerol pathway [33]. We found five Cetacea PSGs (*AGPAT1*, *GPAM*, *LIPF*, *PNLIP*, and *PNLIPRP2*) involved in storage of fat, including the triglyceride metabolic process (GO: 0006641) and phosphatidic acid biosynthetic process (GO: 0006654). Three Bovidae PSGs (*GPAM*, *MOGAT1*, and *MOGAT3*) were also found in this category, including the glycerol metabolic process (GO: 0006071), in which TAG is synthesized from monoacylglycerol. These findings suggest enhanced TAG syntheses in both lineages, but under different molecular mechanisms in aquatic and terrestrial

Cetartiodactyla. Functional evolution of the monoacylglycerol pathway associated with rumen evolution, for digesting C4 grasses that emerged approximately 40 million years ago, has been proposed in Bovidae [34]. Conversely, an evolutionary analysis of Cetacea reported positive selection on genes involved in both pathways of TAG synthesis for blubber thickening [8]. In these studies, one cannot rule out that changes in the monoacylglycerol pathway may have been present in a common ancestor (i.e. contemporaneous evolution) since close relatives were not investigated. Our analyses do not suffer from this shortcoming because both Cetacea and Bovidae were analysed in an equivalent manner. Therefore, our analyses support the evolution of the monoacylglycerol pathway in Bovidae and the phosphatidic acid pathway in Cetacea. The only common PSG between Cetacea and Bovidae, *GPAM*, encodes a mitochondrial enzyme that regulates and synthesizes TAG and other lipids [35]. More research, such as *in vitro* functional experiments, is necessary to determine whether the mutations in this gene produced a different phenotypic outcome between the lineages.

We also inspected evolutionary changes in energy producing pathways, including fatty acid degradation and glycolysis/gluconeogenesis, in which fatty acids and glucose serve as the main energy resource, respectively. We found three Cetacea PSGs (*ACADVL*, *EHHADH*, and *HADHB*) in β-oxidation (GO: 0006635). These three PSGs encode four enzymes including acyl-CoA dehydrogenase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and thiolase which drives the β-oxidation [25]. Our domain analysis revealed that Cetacea specific amino acid substitutions occurred in the functional domains of all the four enzymes (figure 3, table 2, and electronic supplementary material, figure S1, S9, and S13). Given the signatures of selection in the four main steps of this process, the β-oxidation of Cetacea may have been extensively

modified during aquatic adaptation, potentially increasing the efficiency of producing energy from fatty acids. To our knowledge, this is the first report of an implication for evolution of Cetacea β-oxidation, including genomic and physiological studies.

The Cetacea PSGs exhibited three genes (*ACSS2*, *FBP1*, and *PDHB*) that were associated with the glycolysis/gluconeogenesis pathway (GO: 0006096 and GO: 0006094). Interestingly, gene expression of *FBP1* is regulated by bile acids [36], implying a possible functional connection with the other identified PSG, *CYP8B1*. *FBP1* encodes an enzyme at the rate limiting step of gluconeogenesis. Although metabolic studies of carnivores have primarily used domestic cats [37], the selection of the Cetacea gluconeogenesis may provide a new insight into the mechanisms for glucose biosynthesis by carnivorous animals. The remaining two PSGs encodes enzymes involved in Acetyl-CoA synthesis: *PDHB* is a part of the pyruvate dehydrogenase complex that converts pyruvate into acetyl-CoA [38], and *ACSS2* is a part of acetyl-CoA synthetase that catalyses the formation of acetyl-CoA from acetate [39]. Acetyl-CoA is an important molecule that participates in the citrate cycle. It is noteworthy that we identified multiple signatures of positive selection involved in the synthesis of acetyl-CoA and selective pressure on β-oxidation. This emphasizes the importance of energy production for aquatic adaptation.

4. Conclusion.

The comparative analyses between aquatic and terrestrial Cetartiodactyla provided genetic signatures of the evolution of lipid metabolism in Cetacea. This study is unique in that it used a sufficient number of both Cetacea and Bovidae samples in an equal

manner that enabled the identification of molecular changes putatively reflecting their ecologies. The lineage specific amino acid substitutions in domain regions implied functional modifications of multiple genes that are involved in important biological processes for aquatic adaptation. These genes and the associated pathways will be plausible targets for future investigations of Cetacea lipid metabolism.

Data accessibility.

Additional results supporting this article have been uploaded as part of the online electronic supplementary material.

Author's contributions.

Y.E. conceived of the study, designed the study, and carried out the data collection, data analysis, bioinformatics works, and drafted the manuscript; K.K. and M.M. coordinated the study and participated in the design of the study, and helped draft the manuscript. All authors gave final approval for publication.

Competing interests

The authors declare no potential conflict of interests for this study.

Funding.

This work was supported by the Japan Society for the Promotion of Science (JSPS; 16K14660 to K.K. and 25118005 to M.I-M.). This study was also partially supported by Kyoto University Supporting program for interaction-based initiative team studies (SPIRITS) to M.I-M. The WPI-iCeMS was supported by the World Premier

International Research Centre Initiative (WPI), MEXT, Japan.

Acknowledgements.

The authors wish to thank Kishida Takushi and Daisuke Muramatsu for assisting with the data analysis and interpretation, and Maegan Fitzgerald for English correction.

References

1. Thewissen J.G.M., Cooper L.N., George J.C., Bajpai S. 2009 From land to water: the origin of whales, dolphins, and porpoises. *Evo Edu Outreach* **2**, 272-288. (doi:10.1007/s12052-009-0135-2).
2. Uhen M.D. 2010 The origin(s) of whales. *Annu Rev Earth Planet Sci* **38**, 189-219. (doi:10.1146/annurev-earth-040809-152453).
3. Gatesy J., Geisler J.H., Chang J., Buell C., Berta A., Meredith R.W., Springer M.S., McGowen M.R. 2013 A phylogenetic blueprint for a modern whale. *Mol Phylogenet Evol* **66**, 479-506. (doi:10.1016/j.ympev.2012.10.012).
4. McGowen M.R., Gatesy J., Wildman D.E. 2014 Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol Evol* **29**, 336-346. (doi:10.1016/j.tree.2014.04.001).
5. Thewissen J.G.M. 2014 *The Walking Whales: From Land to Water in Eight Million Years*. Oakland (CA), University of California Press.
6. Berta A., Sumich J.L., Kovacs K.M. 2006 *Marine Mammals: Evolutionary Biology*. San Diego (CA), Academic Press.
7. Sun Y.-B., Zhou W.-P., Liu H.-Q., Irwin D.M., Shen Y.-Y., Zhang Y.-P. 2013 Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. *Genome Biol Evol* **5**, 130-139. (doi:10.1093/gbe/evs123).
8. 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* **5**, 14187. (doi:10.1038/srep14187).
9. Tsagkogeorga G., McGowen M.R., Davies K.T.J., Jarman S., Polanowski A., Bertelsen M.F., Rossiter S.J. 2015 A phylogenomic analysis of the role and

timing of molecular adaptation in the aquatic transition of cetartiodactyl mammals. *R Soc Open Sci* **2**, 150156. (doi:10.1098/rsos.150156).

10. Foote A.D., Liu Y., Thomas G.W., Vinař T., Alföldi J., Deng J., Dugan S., van Elk C.E., Hunter M.E., Joshi V., et al. 2015 Convergent evolution of the genomes of marine mammals. *Nat Genet* **47**, 272-275. (doi:10.1038/ng.3198).
11. McGowen M.R., Grossman L.I., Wildman D.E. 2012 Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc R Soc Lond B Biol Sci* **279**, 3643-3651. (doi:10.1098/rspb.2012.0869).
12. Löytynoja A., Goldman N. 2005 An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* **102**, 10557-10562. (doi:10.1073/pnas.0409137102).
13. Maddison W.P., Maddison D.R. 2017 Mesquite: A modular system for evolutionary analysis. (3.2 ed).
14. Sela I., Ashkenazy H., Katoh K., Pupko T. 2015 GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res* **43**, W7-W14. (doi:10.1093/nar/gkv318).
15. Yang Z. 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591. (doi:10.1093/molbev/msm088).
16. Yang Z., dos Reis M. 2011 Statistical properties of the branch-site test of positive selection. *Mol Biol Evol* **28**, 1217-1228. (doi:10.1093/molbev/msq303).
17. Finn R.D., Clements J., Arndt W., Miller B.L., Wheeler T.J., Schreiber F., Bateman A., Eddy S.R. 2015 HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30-W38. (doi:10.1093/nar/gkv397).
18. Finn R.D., Coggill P., Eberhardt R.Y., Eddy S.R., Mistry J., Mitchell A.L., Potter S.C., Punta M., Qureshi M., Sangrador-Vegas A., et al. 2016 The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-D285. (doi:10.1093/nar/gkv1344).
19. Moore A.D., Held A., Terrapon N., Weiner r.J., Bornberg-Bauer E. 2014 DoMosaics: software for domain arrangement visualization and domain-centric analysis of proteins. *Bioinformatics* **30**, 282-283. (doi:10.1093/bioinformatics/btt640).
20. Wickham H. 2009 *ggplot2: Elegant Graphics for Data Analysis*. Verlag (NY),

Springer.

21. Huang D.W., Sherman B.T., Lempicki R.A. 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57. (doi:10.1038/nprot.2008.211).
22. Huang D.W., Sherman B.T., Lempicki R.A. 2009 Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13. (doi:10.1093/nar/gkn923).
23. Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. 2017 KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361. (doi:10.1093/nar/gkw1092).
24. Shimamura M., Abe H., Nikaido M., Ohshima K., Okada N. 1999 Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of tRNA(Glu)-derived families of SINEs. *Mol Biol Evol* **16**, 1046-1060. (doi:10.1093/oxfordjournals.molbev.a026194).
25. Bartlett K., Eaton S. 2004 Mitochondrial β-oxidation. *Eur J Biochem* **271**, 462-469. (doi:10.1046/j.1432-1033.2003.03947.x).
26. Agnihotri G., Liu H.W. 2003 Enoyl-CoA hydratase: Reaction, mechanism, and inhibition. *Bioorg Med Chem* **11**, 9-20. (doi:10.1016/S0968-0896(02)00333-4).
27. Nery M.F., González D.J., Opazo J.C. 2013 How to make a dolphin: molecular signature of positive selection in cetacean genome. *PLoS One* **8**, e65491. (doi:10.1371/journal.pone.0065491).
28. Pafumi Y., Lairon D., de la Porte P.L., Juhel C., Storch J., Hamosh M., Armand M. 2002 Mechanisms of inhibition of triacylglycerol hydrolysis by human gastric lipase. *J Biol Chem* **277**, 28070-28079. (doi:10.1074/jbc.M202839200).
29. Mukherjee M. 2003 Human digestive and metabolic lipases—a brief review. *J Mol Catal B Enzym* **22**, 369-376. (doi:10.1016/S1381-1177(03)00052-3).
30. Ma H., Patti M.E. 2014 Bile acids, obesity, and the metabolic syndrome. *Best Pract Res Clin Gastroenterol* **28**, 573-583. (doi:10.1016/j.bpg.2014.07.004).
31. Grundy S.M. 2016 Metabolic syndrome update. *Trends Cardiovasc Med* **26**, 364-373. (doi:10.1016/j.tcm.2015.10.004).
32. Wang Z., Xu S., Du K., Huang F., Chen Z., Zhou K., Ren W., Yang G. 2016 Evolution of digestive enzymes and RNASE1 provides insights into dietary switch of cetaceans. *Mol Biol Evol* **33**, 3144-3157.

(doi:10.1093/molbev/msw191).

33. Takeuchi K., Reue K. 2009 Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. *Am J Physiol Endocrinol Metab* **296**, E1195-E1209. (doi:10.1152/ajpendo.90958.2008).
34. Jiang Y., Xie M., Chen W., Talbot R., Maddox J.F., Faraut T., Wu C., Muzny D.M., Li Y., Zhang W., et al. 2014 The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168-1173. (doi:10.1126/science.1252806).
35. Wendel A.A., Lewin T.M., Coleman R.A. 2009 Glycerol-3-phosphate acyltransferases: Rate limiting enzymes of triacylglycerol biosynthesis. *Biochim Biophys Acta* **1791**, 501-506. (doi:10.1016/j.bbalip.2008.10.010).
36. Yamagata K., Daitoku H., Shimamoto Y., Matsuzaki H., Hirota K., Ishida J., Fukamizu A. 2004 Bile acids regulate gluconeogenic gene expression via small heterodimer partner-mediated repression of hepatocyte nuclear factor 4 and Foxo1. *J Biol Chem* **279**, 23158-23165. (doi:10.1074/jbc.M314322200).
37. Schermerhorn T. 2013 Normal glucose metabolism in carnivores overlaps with diabetes pathology in non-carnivores. *Front Endocrinol (Lausanne)* **4**. (doi:10.3389/fendo.2013.00188).
38. Saunier E., Benelli C., Bortoli S. 2016 The pyruvate dehydrogenase complex in cancer: An old metabolic gatekeeper regulated by new pathways and pharmacological agents. *Int J Cancer* **138**, 809-817. (doi:10.1002/ijc.29564).
39. Fujino T., Kondo J., Ishikawa M., Morikawa K., Yamamoto T.T. 2001 Acetyl-CoA synthetase 2, a mitochondrial matrix enzyme involved in the oxidation of acetate. *J Biol Chem* **276**, 11420-11426. (doi:10.1074/jbc.M008782200).
40. Agnarsson I., May-Collado L.J. 2008 The phylogeny of Cetartiodactyla: The importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. *Mol Phylogenet Evol* **48**, 964-985. (doi:10.1016/j.ympev.2008.05.046).

Figure Legends

Figure 1. Genetic signatures of aquatic adaptation in an evolutionary context. (a) Diagram describing the evolutionary relationship between Cetacea and Bovidae. Cetacea and Bovidae diverged from a terrestrial ancestor and adapted to aquatic and terrestrial habitats. The adaptive features for the aquatic environment indicate rearrangement of biological processes associated with lipid metabolism. (b) Schematic illustration of methods used to investigate the adaptive signatures of Cetacea lipid metabolism. (c) Cladogram showing the evolutionary relationships among Cetartiodactyla species. Branch colour represents the lineages: five Cetacea (*blue*: target aquatic group), five Bovidae (*orange*: control terrestrial group), and one Suina and three Tylopoda (*black*: outgroup). The phylogenetic positions of the species were reconstructed according to previous works [3, 40]. The lineages tested for positively selected genes are indicated with (i) and (ii).

Figure 2. Number of identified positively selected genes (PSGs). (a) Bar graph showing the number of PSGs in Cetacea and Bovidae (*blue*: $p < 0.05$; *light blue*: $0.05 \leq p < 0.1$). (b) Venn diagram showing unique and overlapping PSGs in Cetacea and Bovidae (*blue*: $p < 0.05$; *light blue*: $0.05 \leq p < 0.1$).

Figure 3. Amino acid substitutions that were found in functional domains of positively selected genes (PSGs), *EHHADH* as an example. (a) Domain arrangements showing the name and position of domains. (b) Dot plots for the amino acid sequence similarity of each domain. *Blue* dots represent cetacea species, and *orange* dots represent bovidae species. (c) Lineage specific amino acid substitution that were unique

to and identical within Cetacea.

Figure 4. Unique and overlapping metabolic pathways based on gene ontological (GO) terms of the positively selected genes (PSGs) in Cetacea (*blue*), Bovidae (*orange*), and both (*green*). Solid arrow represents transformation of the substrates in one-step, and dashed arrow represents multiple steps. Each asterisk represents the rate limiting step. Each metabolic pathway is indicated in a *pink* rounded-square.

Table 1
List of positively selected genes (PSGs) identified in Cetacea and Bovidae.

Lineage	Gene symbol	Estimates of ω	κ	lnL one-ratio	lnL two-ratio	*p* value	
Cetacean	*CYP8B1*	11.92734	5.85792	-4291.819841	-4285.67461	4.55.E-04	**
	PDHB	5.98748	4.36259	-2494.18574	-2491.079352	1.27.E-02	*
	AGPAT1	7.05413	4.86528	-1855.189217	-1852.248895	1.53.E-02	*
	LIPF	4.44878	2.25004	-6287.045085	-6284.171286	1.65.E-02	*
	PNLIPRP2	4.54093	3.64754	-4524.349375	-4521.629714	1.97.E-02	*
	APOB	4.05216	4.19478	-41746.17128	-41743.46198	1.99.E-02	*
	CYP2J2	13.63547	3.19121	-5139.202241	-5136.697766	2.52.E-02	*
	EHHADH	14.52325	3.44028	-5810.991244	-5808.631236	2.98.E-02	*
	ACADVL	3.40418	4.11549	-5098.363465	-5096.023469	3.05.E-02	*
	GPAM	4.35956	4.01882	-6046.804276	-6044.470464	3.07.E-02	*
	CYP2E1	2.37475	4.51029	-4844.274535	-4842.07278	3.59.E-02	*
	HADHB	6.58348	4.73191	-3491.684049	-3489.483724	3.59.E-02	*
	APOA2	3.60699	4.43456	-961.661901	-959.657031	4.52.E-02	*
	PGD	3.30916	3.05453	-3813.94106	-3812.008851	4.93.E-02	*
	GNPAT	2.10957	4.04115	-5941.958899	-5940.047374	5.06.E-02	
	ACSS2	8.09817	4.63572	-4902.585196	-4900.799437	5.88.E-02	
	PNLIP	3.79609	3.17642	-4448.612671	-4446.922816	6.60.E-02	
	FBP1	2.87282	3.22927	-2728.680035	-2727.07769	7.34.E-02	
Bovidae	*MOGAT3*	6.44231	3.04369	-4562.007537	-4552.517873	1.32.E-05	**
	ABHD12	10.698	4.10908	-2503.418086	-2499.826239	7.36.E-03	**
	APOE	32.57059	4.04548	-3013.622556	-3010.088442	7.85.E-03	**
	MLYCD	12.46396	4.31351	-4458.618387	-4456.297488	3.12.E-02	*
	GPAM	8.83467	4.03293	-6050.187322	-6047.92294	3.33.E-02	*
	IDH2	45.30329	6.95175	-3056.751425	-3054.899266	5.43.E-02	
	HSD17B4	16.02856	4.15483	-5827.966765	-5826.198171	6.00.E-02	
	SOAT1	45.93865	3.48586	-4577.415397	-4575.845606	7.64.E-02	
	PGAM5	22.60281	5.77771	-1980.355128	-1978.812842	7.90.E-02	
	ACAD9	2.64777	4.91697	-5713.112028	-5711.57532	7.96.E-02	
	MOGAT1	5.222	3.44578	-2989.32785	-2987.877786	8.86.E-02	
	ACADM	21.5195	2.95373	-2911.486719	-2910.067882	9.21.E-02	
	SOAT2	12.78966	4.68574	-4299.927829	-4298.536813	9.53.E-02	

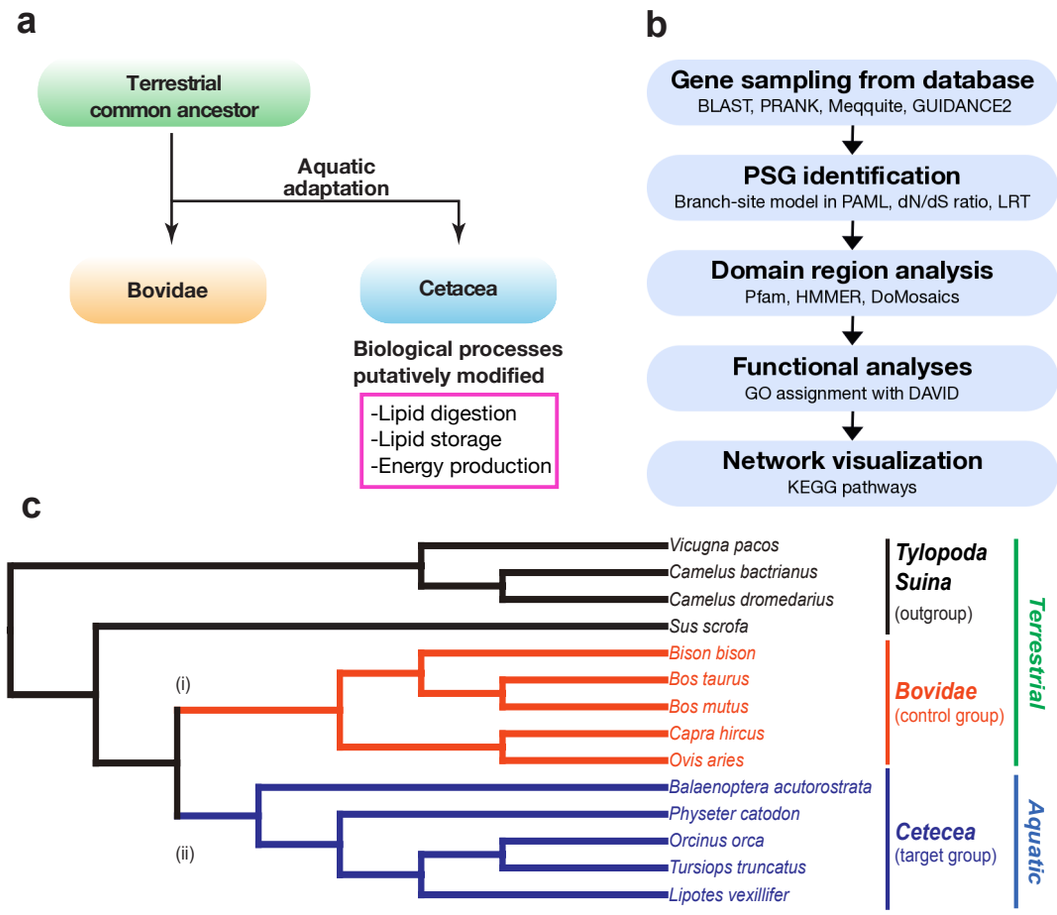
Notes.—κ, transition/transversion rate ratio; ln L one-ratio and ln L two-ratio, log likelihood value for null and alternative hypothesis, respectively; *, significant (P < 0.05); **, highly significant (P < 0.01).

Table 2
Lineage specific amino acid substitutions that occurred in domain regions of positively selected genes (PSGs)

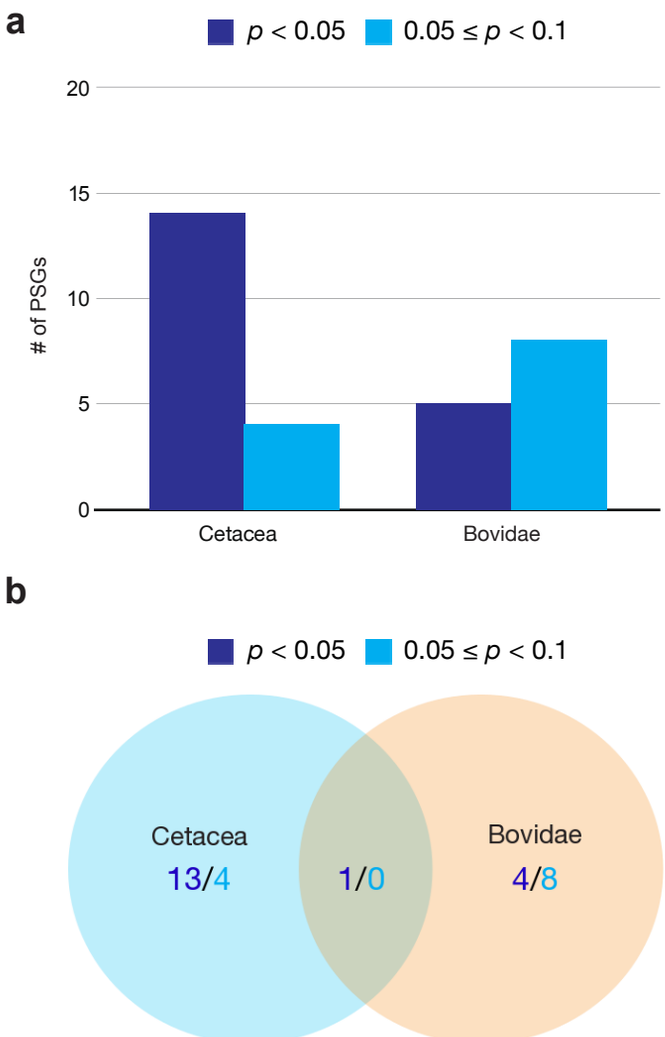
Cetacea PSGs	Domain ID	Domain name	Domain position	Lineage specific amino acid substitution							
ACADVL	PF02771	Acyl-CoA dehydrogenase, N-terminal domain	97 - 209	V174L	G175A						
	PF02770	Acyl-CoA dehydrogenase, middle domain	213 - 315	T268K	Y311H						
	PF00441	Acyl-CoA dehydrogenase, C-terminal domain	327 - 473	L378M							
	PF00441	Acyl-CoA dehydrogenase, C-terminal domain	537 - 614	T545A	I608T						
ACSS2	PF16177	Acetyl-coenzyme A synthetase N-terminus	47 - 107	NA							
	PF00501	AMP-binding enzyme	115 - 574	L287M	N396S	K426R	I500V				
	PF13193	AMP-binding enzyme C-terminal domain	583 - 661	P625S	G653S						
AGPAT1	PF01553	Acyltransferase	79 - 207	NA							
APOA2	PF04711	Apolipoprotein A-II	24 - 97	T68A							
APOB	PF01347	Lipoprotein amino terminal region	44 - 596	K130R	R227K	G355A	S358N				
	PF09172	Domain of unknown function	630 - 935	A651G	E691Q	V756A	H788Q	I807V	M867I	I924V	
	PF06448	Domain of unknown function	958 - 1069	G977V	E1038K	L1047M	R1050T				
	PF12491	Apolipoprotein B100 C terminal	4499 - 4554	S4508A	S4549L						
CYP2E1	PF00067	Cytochrome P450	33 - 489	I42V L199M	L45I M200L	D49E L225I	K94R R249G	K110Q V396L	N111D A453S	R123Q	F153L
CTP2J2	PF00067	Cytochrome P450	44 - 497	N130I	R146K	N179D	D268G	Q283K	L378M	D475N	G486S
CYP8B1	PF00067	Cytochrome P450	34 - 491	R121Q	I365S	P398S					
EHHADH	PF00378	Enoyl-CoA hydratase/isomerase	8 - 201	I24L	C58S	C199Y					
	PF02737	3-hydroxyacyl-CoA dehydrogenase, NAD binding domain	300 - 475	V302I	V313A	Q320K	A348V	V459I			
	PF00725	3-hydroxyacyl-CoA dehydrogenase, C-terminal domain	478 - 582	S487P	I529V	K577E					
	PF00725	3-hydroxyacyl-CoA dehydrogenase, C-terminal domain	619 - 710	I649V	L676V						
FBP1	PF00316	Fructose-1-6-bisphosphatase, N-terminal domain	12 - 198	NA							
GNPAT	PF01553	Acyltransferase	146 - 283	NA							
GPAM	PF01553	Acyltransferase	213 - 355	NA							
HADHB	PF00108	Thiolase, N-terminal domain	55 - 325	S221N	M322L						
	PF02803	Thiolase, C-terminal domain	332 - 471	R447Q							
LIPF	PF00561	alpha/beta hydrolase fold	77 - 376	L85F							
PDHB	PF02779	Transketolase, pyrimidine binding domain	33 - 206	NA							
	PF02780	Transketolase, C-terminal domain	226 - 349	G247S	A322V	I337V					
PGD	PF03446	NAD binding domain of 6-phosphogluconate dehydrogenase	5 - 165	NA							
	PF00393	6-phosphogluconate dehydrogenase, C-terminal domain	180 - 469	K241R	K302Q						
PNLIP	PF00151	Lipase	18 - 353	V19I	I69L	L160P	A178S	M226I			
	PF01477	PLAT/LH2 domain	358 - 458	T369R	K380R	K430R					
PNLIPRP2	PF00151	Lipase	18 - 354	E50K	P105I	H127R	E157Q				
	PF01477	PLAT/LH2 domain	359 - 460	T448A	N455S	T457M					
Bovidae PSGs											
ABHD12	PF12146	Serine aminopeptidase, S33	170 - 286	A232V							
ACAD9	PF02771	Acyl-CoA dehydrogenase, N-terminal domain	67 - 172	S132G	E158K						
	PF02770	Acyl-CoA dehydrogenase, middle domain	176 - 276	E199A	L262M						
	PF00441	Acyl-CoA dehydrogenase, C-terminal domain	289 - 436	G364A	Q368A						
	PF00441	Acyl-CoA dehydrogenase, C-terminal domain	517 - 577	NA							

Table 2
Continue

Bovidae PSGs	Domain ID	Domain name	Domain position	Lineage specific amino acid substitution							
ACADM	PF02771	Acyl-CoA dehydrogenase, N-terminal domain	42 - 150	T109S	G125A						
	PF02770	Acyl-CoA dehydrogenase, middle domain	157 - 255	G164V							
	PF00441	Acyl-CoA dehydrogenase, C-terminal domain	267 - 414	G287A	V314I	V368L	T370S				
APOE	PF01442	Apolipoprotein A1/A4/E domain	80 - 291	T146S E251Q	L154M K279R	E170D	Q173K	R180Q	R206Q	R221L	R232H
GPAM	PF01553	Acyltransferase	213 - 355	NA							
HSD17B4	PF00106	Short chain dehydrogenase	11 - 196	A78E	V82I	V95I	G109S				
	PF13452	N-terminal half of MaoC dehydratase	347 - 446	L361M	G388N	V404I	K424R	V432I			
	PF01575	MaoC like domain	484 - 599	NA							
	PF02036	SCP-2 sterol transfer family	629 - 731	N670T	M721L						
IDH2	PF00180	Isocitrate/isopropylmalate dehydrogenase	45 - 441	A204P	E268A	N281H					
MLYCD	PF17408	Malonyl-CoA decarboxylase N-terminal domain	100 - 195	G105S	L152M	D153E	Q161R				
	PF05292	Malonyl-CoA decarboxylase C-terminal domain	198 - 462	K216R V391A	T330A N427H	E359G M436L	C366S M442L	K367Q S453A	I372L	V377A	K382Q
MOGAT1	PF03982	Diacylglycerol acyltransferase	42 - 335	D73E	T79A	F145Y	M275I	H296N	P299S		
MOGAT3	PF03982	Diacylglycerol acyltransferase	34 - 328	F51L H183N	E66Q S185A	R78S N264K	H141Y F271L	V144G L289P	F164Y V320L	I165V	Q173R
PGAM5	PF00300	Histidine phosphatase superfamily (branch 1)	100 - 272	NA							
SOAT1	PF03062	MBOAT, membrane-bound O-acyltransferase family	172 - 521	L223I	A325T	V364I	L444R	F445R	S447A		
SOAT2	PF03062	MBOAT, membrane-bound O-acyltransferase family	163 - 507	I166T	A222V	R341Q	M476T	M493L			

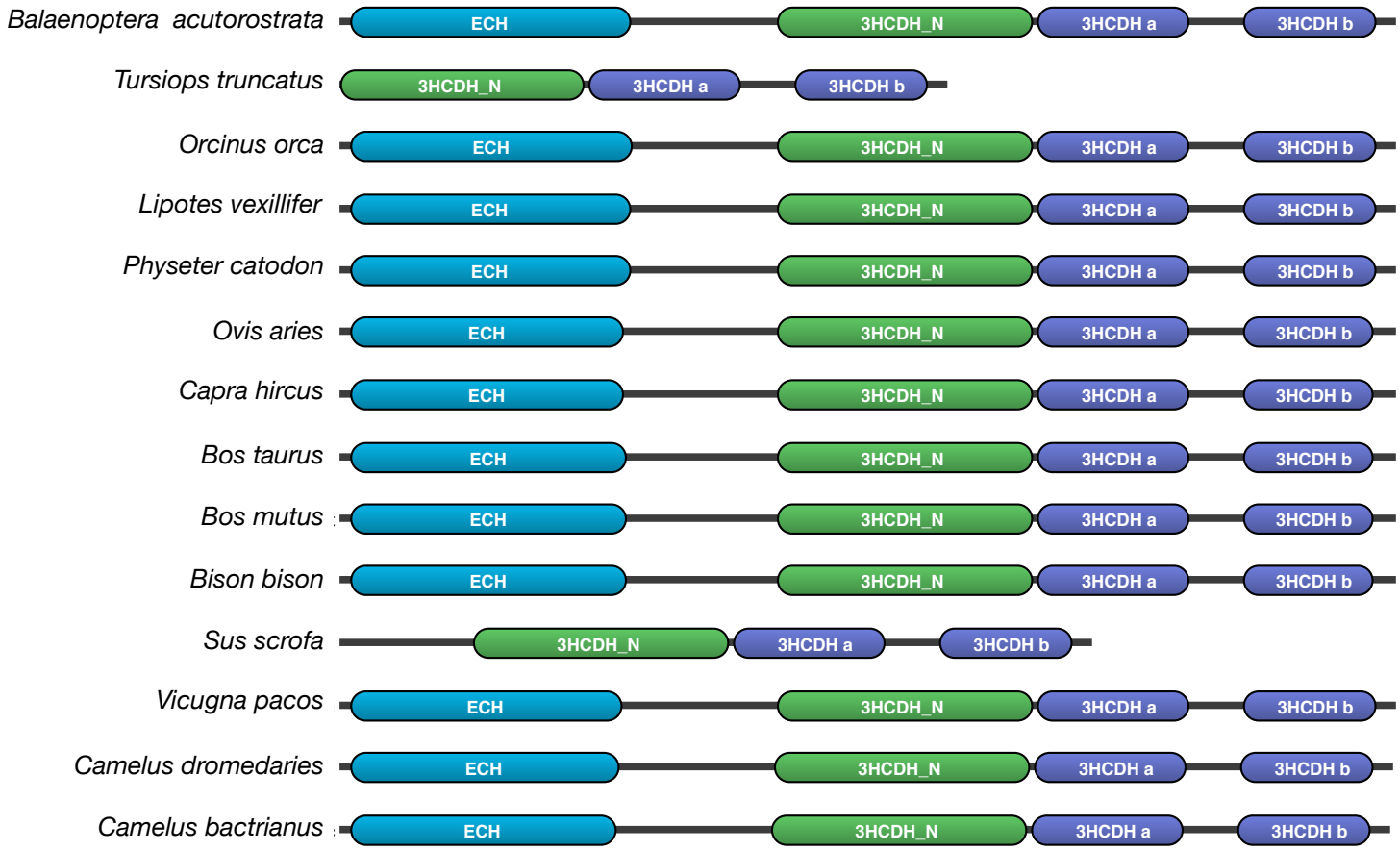


Endo et al., Figure 1

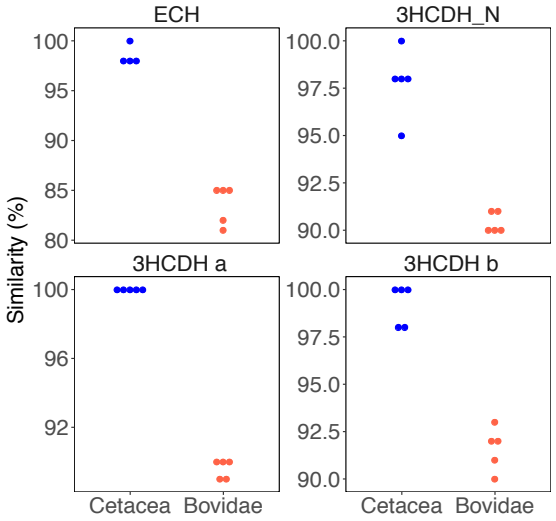


Endo et al., Figure 2

a



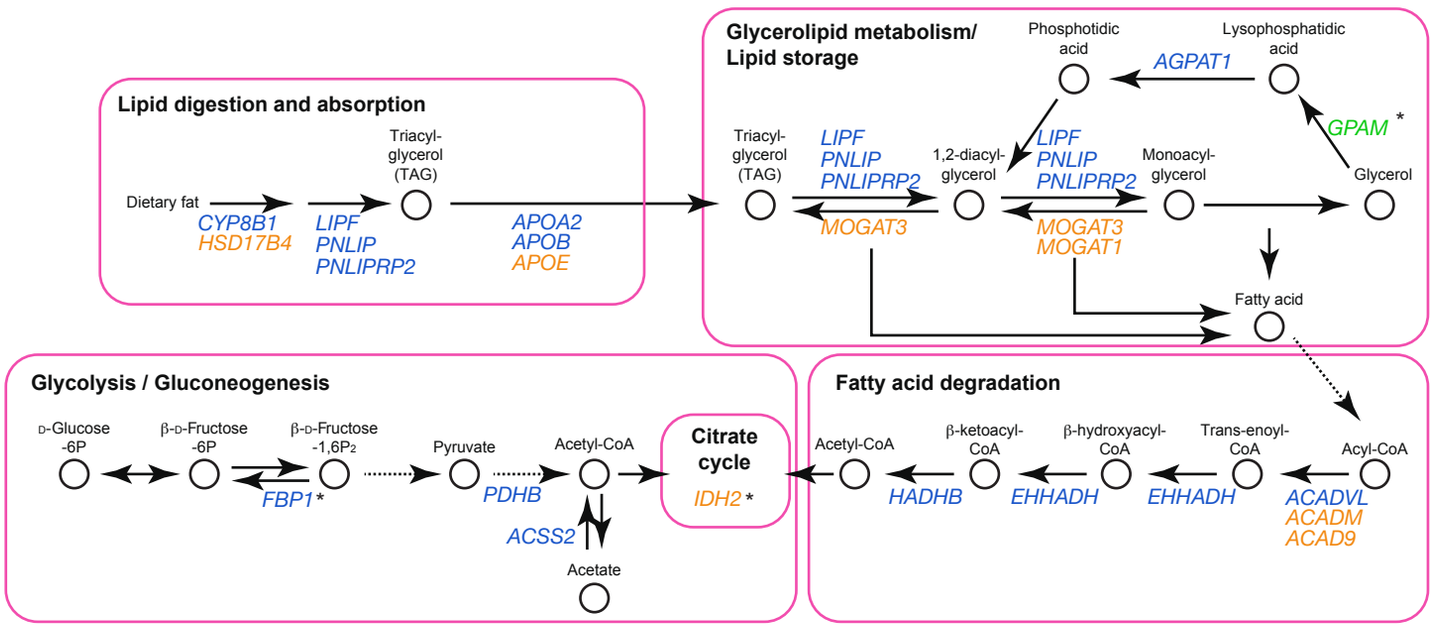
b



c

	ECH domain, C58S										3HCDH domain, K577E									
Balaenoptera acutorostrata	D	G	I	F	**S**	A	G	A	D	I	Q	K	T	G	**E**	G	W	Y	L	Y
Tursiops truncatus	-	-	-	-	-	-	-	-	-	-	Q	K	T	G	**E**	G	W	Y	L	Y
Orcinus orca	D	G	I	F	**S**	A	G	A	D	I	Q	K	T	G	**E**	G	W	Y	L	Y
Lipotes vexillifer	D	G	I	F	**S**	A	G	A	D	I	Q	K	T	G	**E**	G	W	Y	L	Y
Physeter catodon	D	G	I	F	**S**	A	G	A	D	I	Q	K	T	G	**E**	G	W	Y	L	Y
Ovis aries	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Capra hircus	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Bos taurus	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Bos mutus	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Bison bison	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Sus scrofa	-	-	-	-	-	-	-	-	-	-	Q	K	T	G	K	G	W	Y	L	Y
Vicugna pacos	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Camelus dromedarius	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y
Camelus bactrianus	D	G	I	F	C	A	G	A	D	I	Q	K	T	G	K	G	W	Y	L	Y

Endo et al., Figure 3



Endo et al., Figure 4