# Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals

Maria Chikina,*[,1] Joseph D. Robinson,[2] and Nathan L. Clark*[,1]
[1]Department of Computational and Systems Biology, University of Pittsburgh
[2]Department of Molecular and Cell Biology, University of California Berkeley
*Corresponding author: E-mail: mchikina@pitt.edu; nclark@pitt.edu.
Associate editor: David Irwin

## Abstract

**Mammal species have made the transition to the marine environment several times, and their lineages represent one of the classical examples of convergent evolution in morphological and physiological traits. Nevertheless, the genetic mechanisms of their phenotypic transition are poorly understood, and investigations into convergence at the molecular level have been inconclusive. While past studies have searched for convergent changes at specific amino acid sites, we propose an alternative strategy to identify those genes that experienced convergent changes in their selective pressures, visible as changes in evolutionary rate specifically in the marine lineages. We present evidence of widespread convergence at the gene level by identifying parallel shifts in evolutionary rate during three independent episodes of mammalian adaptation to the marine environment. Hundreds of genes accelerated their evolutionary rates in all three marine mammal lineages during their transition to aquatic life. These marine-accelerated genes are highly enriched for pathways that control recognized functional adaptations in marine mammals, including muscle physiology, lipid-metabolism, sensory systems, and skin and connective tissue. The accelerations resulted from both adaptive evolution as seen in skin and lung genes, and loss of function as in gustatory and olfactory genes. In regard to sensory systems, this finding provides further evidence that reduced senses of taste and smell are ubiquitous in marine mammals. Our analysis demonstrates the feasibility of identifying genes underlying convergent organism-level characteristics on a genome-wide scale and without prior knowledge of adaptations, and provides a powerful approach for investigating the physiological functions of mammalian genes.**

***Key words:*** **convergent evolution, convergence, marine mammals, adaptive evolution, relaxation of constraint, functional constraint.**

## Introduction

Drastic environmental changes such as the transition from a terrestrial to marine habitat should select for numerous evolutionary adaptations. Multiple mammalian lineages have made this transition, and there has been concerted effort to study convergence of their morphological and physiological adaptations. At a macroscopic level, many convergent changes are observed: the morphologies of marine mammals have become streamlined through limb reduction (Fish and Hui 1991; Fish et al. 2008), and their skins have evolved a thick, compact epidermis, presumably as a result of selection to reduce friction, permeability, and heat loss (Whittow 1987; Rosen and Renouf 1997; Meyer et al. 2011). Sensory systems have also evolved in to forms that process light and sound in ways suited to life underwater (Wartzok and Ketten 1999). Respiratory physiology was remodeled for diving under great pressure, namely through structural changes allowing the lung and airways to remain functional after extreme compression (Kooyman 2006). Marine mammal lineages also have increased blood gas capacities and altered gas exchange mechanisms to help them during periods of hypoxia, paired with restricted blood circulation to prolong essential functions (Andersen 1966). Attesting to the strength of selection in the marine environment, these adaptive traits have appeared independently in each mammalian lineage that made the transition, and they are absent from their terrestrial cousins. Morphological and physiological convergence is therefore strong in marine mammals and provides a powerful criterion to identify specific adaptations.

Convergence at the molecular level has been more difficult to identify. Recent studies have described genes with signatures of positive selection within single marine lineages (McGowen et al. 2012; Sun et al. 2013; Yim et al. 2014), but while these selective events may underlie marine adaptations, they may also have resulted from other lineage-specific evolutionary pressures unrelated to the marine environment. To overcome this limitation, a strategy identifying convergent changes across independent marine lineages should provide the needed leverage to reveal marine-specific molecular adaptations, including those that are not apparent at an organismal scale. A recent study by Foote et al. (2015) identified genes with convergent substitutions at specific amino acid sites across marine lineages (manatee, walrus, bottlenose dolphin, killer whale). However, they found the number of

convergent amino acid changes in their control species group (cow, dog, elephant) was similar, thus raising the possibility that the convergent changes on marine branches resulted from neutral processes and were not driven by specific adaptations to the marine environment.

One possible explanation for this result is that adaptively driven convergence at specific amino acid sites may be exceedingly rare (Zhang and Kumar 1997); indeed, amino acids sometimes exhibit drastically different selective effects depending on the host genome (Kulathinal et al. 2004), and many distinct beneficial mutations appear in the context of simple and reproducible selective pressures (Gresham et al. 2008; Karasov et al. 2010; Spor et al. 2014; Bailey et al. 2015). A series of studies have used evolutionary models to demonstrate that the power to identify convergent amino acid changes is low, and that reported cases of convergence are not found in excess of background (Zhang and Kumar 1997; Zou and Zhang 2015a). It seems that the candidates emanating from genome-wide searches for amino acid convergence can be sufficiently explained by neutral models of evolution (Thomas and Hahn 2015; Zou and Zhang 2015b). On the other hand, there are compelling reports of convergence at specific amino acid sites, such as in response to repeated adaptation to a specific biochemical substrate (Dobler et al. 2012) or parallel changes in the oxidative environment during respiration or photosynthesis (Castoe et al. 2008; Christin et al. 2008; Castoe et al. 2009). Multiple studies also report convergence at specific amino acids in hearing-related genes in echolocating species (Li et al. 2010; Liu et al. 2010; Davies et al. 2012; Shen et al. 2012). These studies present evidence that there are episodes of selection leading to convergence at amino acid sites, likely when the systems are highly constrained (Christin et al. 2010; Stern 2013). However, convergent amino acid changes may only represent a fraction of the total adaptive changes that occurred. We also expect many convergent adaptations in complex, multigenic traits to have evolved via nonidentical genetic changes. Detecting these changes will require a different analytical strategy.

We propose that many convergent phenotypes evolve through nonidentical changes in common sets of genes. We expect that these genes experienced parallel changes in selective pressure in the convergent species, and thus exhibit convergent shifts in whole-gene evolutionary rate. Therefore, our strategy is to use convergent changes in evolutionary rate to identify genes underlying convergent phenotypic evolution. Specifically, we use mammalian lineages making the land-to-water transition, a high-level environmental change, to test this strategy. Throughout this report, we will use the term "convergent" not only in reference to traits, but also to refer to shifts in evolutionary rate that are shared between the genes of marine mammals. Convergent changes in selective pressure could result when the marine environment selects for a particular phenotype, and that phenotype is realized by different adaptive changes in the same gene in each marine lineage. In addition to adaptive evolution, environmental change may also release certain functions from constraint (Leys et al. 2005; Clark et al. 2013; Feng et al. 2014). Both scenarios—adaptation and relaxation—are predicted to result in a more rapid rate of sequence evolution. Alternatively, environmental change may also apply additional constraint on a gene if its function becomes more important, resulting in a slower evolutionary rate. Here, we test these hypotheses and present evidence for pervasive convergence at the gene level during mammalian shifts to the marine environment. We identified hundreds of genes that shifted toward these three evolutionary regimes—adaptation, relaxation, and additional constraint—in response to the marine environment by scanning for convergent changes in evolutionary rate in three independent groups of marine mammals: cetaceans (dolphins and whales), pinnipeds (seals, walruses, sea lions), and sirens (manatees and dugongs).

## Results

### Hundreds of Genes Experienced Rate Accelerations or Decelerations Associated with the Marine Environment

Using 59 placental mammal genomes, we calculated relative rates of evolution for all branches in amino acid alignments of 18,049 protein-coding gene trees (see Materials and Methods). Briefly, relative rates indicate whether that gene evolved at a faster or slower rate than the expectation after normalization for both the genome-wide average rate for that branch, and the gene-specific rate across all species. In the next step, we employed these relative rates to identify those genes that convergently shifted to higher (or lower) than average rates, specifically in the marine environment as represented by the branches leading to five marine species—bottlenose dolphin, killer whale, walrus, Weddell seal, and West Indian manatee—plus the branch leading to the common ancestor of the bottlenose dolphin and killer whale (fig. 1A). To illustrate this approach, the relative rates of the branches of the *GNAT3* gene tree are plotted in figure 2A. *GNAT3* encodes a G protein specific to taste and gustatory signaling. The relative rate of the cetacean ancestral branch (the branch leading to dolphins and whales) is positive and large, indicating that this gene accumulated more changes than we would have expected based on the average amount of divergence on that branch for all genes. In representing rates in this way, it is clear that the other marine mammal branches (red) also show higher rates compared to nonmarine branches (blue), which are centered at a relative rate of zero. The higher relative rates on marine branches in *GNAT3* compared with all terrestrial branches is highly statistically significant, as measured using a nonparametric test ($P = 8.23 \times 10^{-5}$, Wilcoxon-rank sum test). The same analytical steps were repeated for all 18,049 genes to identify those with convergently higher or lower relative rates in marine lineages (table 1, supplementary tables S1 and S3, Supplementary Material online). We hereafter refer to genes with higher rates on marine branches as "marine-accelerated" and those with lower rates as "marine-decelerated." This test does not require all marine branches to have positive relative rates, but because we employ a rank-based test, low *P* values indicate that a gene has consistently low or high rates in a

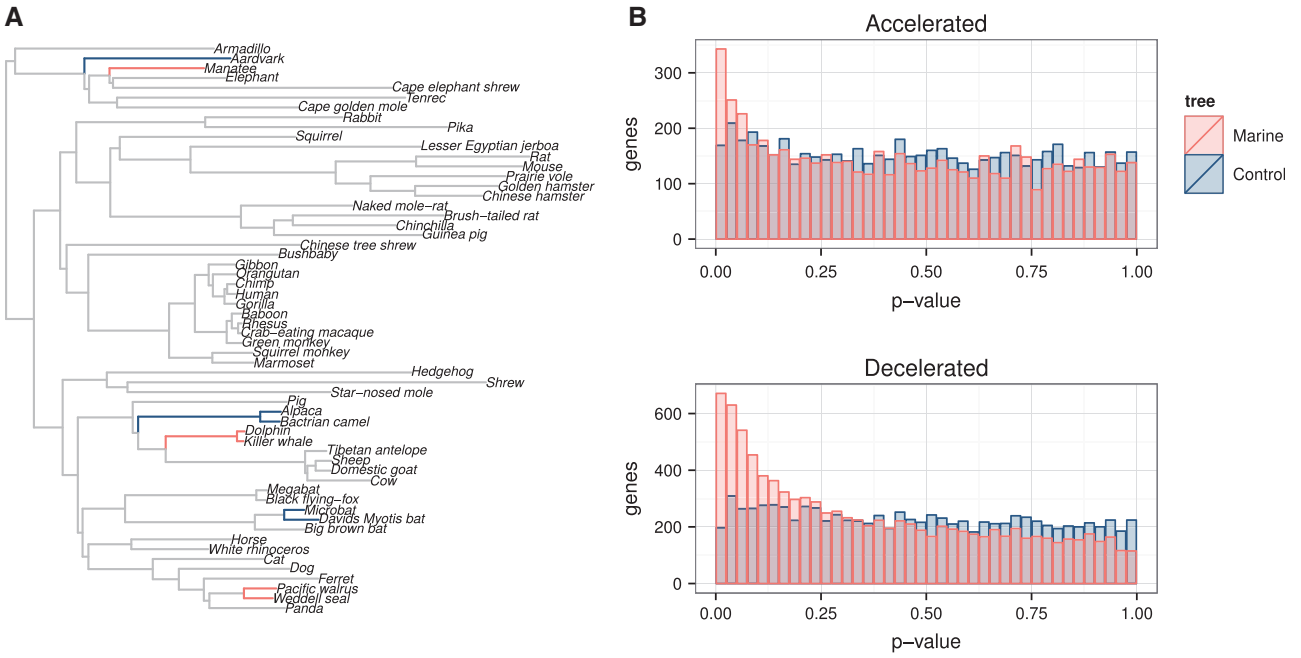


**Fig. 1.** Evolutionary rate shifted in a convergent manner for hundreds of genes in marine mammals. (*A*) Mammals moved from terrestrial to marine environments in 3 independent lineages in this phylogenetic tree of 59 placental mammals. Branches used to infer rate changes associated with the marine phenotype are those leading to the manatee, dolphin, killer whale, Pacific walrus, Weddell seal, and the ancestral branch of the dolphin and killer whale. Other branches represent a set of control branches not expected to show convergence, and which were selected to approximate the topology and branch lengths of the marine set. (*B*) Gene-by-gene evidence for shifts to faster rates (accelerated) and slower rates (decelerated) associated with the highlighted branches was assessed using a Wilcoxon rank-sum test. The resulting genome-wide distributions for the control branch group show no enrichment of low *P* values for either acceleration or deceleration. In contrast, the marine branch group exhibited a dramatic shift towards low *P* values for both faster and slower rates suggesting that the marine environment consistently altered the evolutionary pressures on a large number of genes.

majority of marine branches when compared with terrestrial branches.

We see in figure 1*B* that for genes accelerated or decelerated on marine branches, the genome-wide distribution of *P* values is markedly skewed downward, supporting the hypothesis of convergent shifts in selective pressure on hundreds of genes. In contrast, a control set of mammalian species that was matched for branch length and topology but lacked obvious phenotypic convergence—aardvark, alpaca, camel, microbat, and David's myotis bat—was only modestly enriched for low *P* values. The marine and control *P* value distributions were significantly different for accelerated and decelerated genes ($P = 2.46 \times 10^{-8}$ and $< 2.2 \times 10^{-16}$, respectively, Kolmogorov–Smirnov test). Using the distribution of *P* values in our control species set we apply the q-value method for estimating the fraction of true null hypotheses and infer that genome-wide approximately 560 genes are convergently accelerated and 2,700 genes are decelerated in the marine species. In this analysis we did not designate the pinniped ancestral branch (that leading to walrus and seal) as marine, because the extent to which their last common ancestor was marine is unclear (Arnason et al. 2006; Rybczynski et al. 2009). This was a conservative decision, because if their ancestor were indeed marine, omitting it would only decrease the signal of convergence in our analytical framework. Nevertheless, when the analysis is done with the pinniped ancestral branch as marine, it did not greatly change the results of our study. The gene ranks from our original and the pinniped-ancestor analyses were highly correlated ($R = 0.88$, *P* value $<$ 2.2e-16).

## Marine-Accelerated Genes Are Enriched for Functions Consistent with Marine Adaptations

To ask which biological functions were enriched among marine-accelerated genes, we searched databases of gene pathways and mutant phenotypes (Smith and Eppig 2009; Liberzon et al. 2011). The convergently marine-accelerated genes contained statistically significant enrichment for functional categories that align well with current knowledge of marine mammal adaptations: sensory systems, muscle function, skin and connective tissue, lung function, and lipid metabolism (fig. 3 and table 1; supplementary tables S1 and S2, Supplementary Material online). The same functional enrichments were observed in the additional analysis with the pinniped ancestor as a marine branch. Marine-decelerated genes on the other hand contained a very different set of enriched functions, broadly described as DNA repair, chromosomal maintenance, immune response and apoptosis (supplementary tables S4 and S5, Supplementary Material online). The slower rate of change in these functions is consistent with increased constraint on somatic cell maintenance as would be required in these relatively long-lived and large-bodied mammals. In fact, this trend is illustrated by the additional large and long-lived species with slower rates in these genes (e.g., double-strand break repair gene *XRCC4* is also highly constrained in elephant; figure 2*B* "*XRCC4*," supplementary

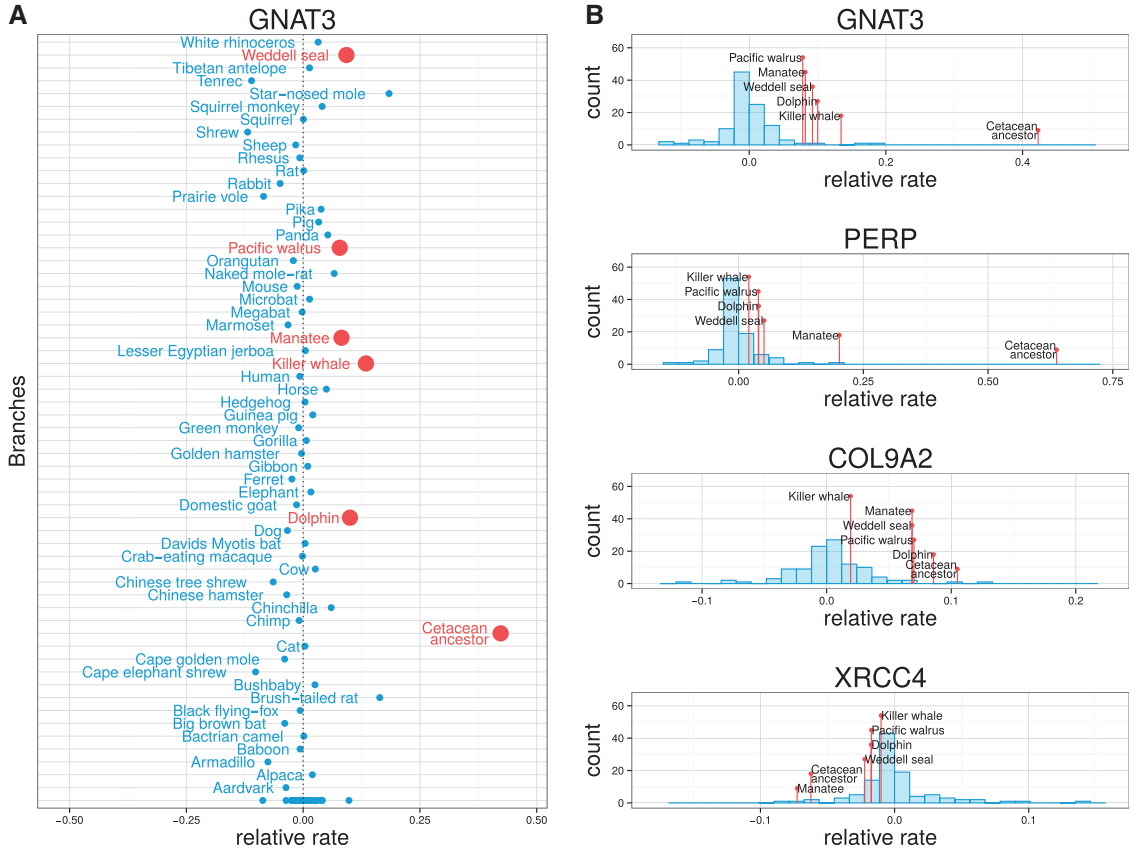


**Fig. 2.** Representative cases of marine-accelerated and decelerated genes. (*A*) Relative evolutionary rates of GNAT3, a guanine nucleotide binding protein involved in taste transduction, for 59 placental mammals and their ancestral branches (points in bottom row). Marine branches (large points) had consistently higher rates, resulting in a low probability that the marine rates are the same as the terrestrial species (small points) ($P = 8.23 \times 10^{-5}$). GNAT3 showed the strongest rate acceleration. This and other taste-related genes provide strong evidence that gustatory senses are reduced or absent in all marine mammals. (*B*) A compacted view of GNAT3 and other genes with significant marine associations. The background distribution over all non-marine branches is shown as a histogram and the rate values of marine branches are flagged with vertical bars. Other depicted examples are PERP ($P = 4.92 \times 10^{-4}$), a desmosome component (an epithelial cell-cell junction structure); COL9A2 ($P = 2.32 \times 10^{-4}$), a cartilage-specific fibril-associated collagen; and XRCC4 ($P = 1.53 \times 10^{-3}$), a DNA repair protein with convergently slower rates of evolution in marine mammals.

fig. S2, Supplementary Material online). Generally, the marine-decelerated set encompasses more genes, but their associations are less specific to the marine environment, which is also reflected in their less extreme *P* values (supplementary table S4 vs. S1, Supplementary Material online). We then asked whether the marine-accelerated genes were associated with large body size as well. We identified the enriched functional categories for genes accelerated in a control set of six large-bodied, terrestrial species (alpaca, camel, elephant, horse, and rhino). Only one enriched functional category, "muscle contractility," was shared between this terrestrial, large-bodied control, and the marine species (supplementary table S7, Supplementary Material online). None of the other significantly enriched functions for marine-accelerated genes were shared with this control.

## Positive Selection Drove Marine-Acceleration in Skin and Lung Proteins

Marine-accelerated genes could have resulted from either bursts of adaptive evolution or relaxed constraint in response to the environment. We assigned each gene to these alternative selective regimes using tests of lineage-specific positive selection in CODEML. Specifically, we compared a set of three nested likelihood models: a neutral sites model (M1), the neutral branch-site model (BS1) and the branch-site model allowing positive selection (BS2) (Yang 2007). Rejection of M1 in favor of BS1 indicated relaxed constraint in marine branches, and further rejection of BS1 to prefer BS2 indicated positive selection in the marine species. A majority of the affected functions in marine-accelerated genes were under relaxed constraint, indicating a decline in their importance in the marine environment (table 2; supplementary table S1, Supplementary Material online). However, a high proportion of genes encoding structural components of cartilage and skin showed evidence of positive selection. Prominent examples of positively selected marine-accelerated genes included components of the cell junction-forming desmosome (e.g., *PERP*), cartilage collagen (*COL9A2*), and the epidermal cross-linking transglutaminase (*TGM3*). All of the marine-accelerated genes under positive selection also showed significant evidence of positive selection across the mammalian tree (supplementary table S1, Supplementary Material online), so their positive selection is not specific to marine branches. However, they did nevertheless show increased

**Table 1.** Top Marine-Accelerated Genes.

| Gene | *P* Value* | Description | Function | Evolutionary Mode |
|---|---|---|---|---|
| GNAT3 | 0.00008 | G protein subunit in bitter, sweet, and umami taste transduction | Taste | Relaxed |
| OR6B1 | 0.00014 | Olfactory receptor | Olfaction | Relaxed |
| CALHM1 | 0.00014 | Ion channel required for sweet, bitter and umami tastes | Taste | Relaxed |
| SSTR4 | 0.00016 | Somatostatin receptor 4 | | Relaxed |
| COL9A2 | 0.00023 | Collagen, type IX, alpha 2; hyaline joint cartilage protein | Cartilage | Pos. selection |
| FGF11 | 0.00033 | Fibroblast growth factor. Nervous system development | | Relaxed |
| HMGCS2 | 0.00042 | Catalyzes ketogenesis, which provides lipid-derived energy | Lipid metabolism | Relaxed |
| CLSTN2 | 0.00046 | Calsyntenin 2 | Nervous system | Relaxed |
| PERP | 0.00049 | Component of intercellular desmosome junctions in epithelia | Skin | Pos. selection |
| S100A5 | 0.00059 | S100 calcium binding protein A5 | | Relaxed |
| OR51M1 | 0.00076 | Olfactory receptor, family 51, subfamily M, member 1 | Olfaction | Pos. selection |
| OR51I2 | 0.00083 | Olfactory receptor, family 51, subfamily I, member 2 | Olfaction | Relaxed |
| KRTAP10-3 | 0.00094 | Keratin associated protein 10-3 | Hair | Relaxed |
| OR51I1 | 0.00096 | Olfactory receptor, family 51, subfamily I, member 1 | Olfaction | Relaxed |
| TRIM29 | 0.00099 | Tripartite motif containing 29 | | Relaxed |
| CNGA4 | 0.00102 | Cyclic nucleotide gated channel alpha 4, subunit of channels that transduce signals in olfactory sensory neurons | Olfactory neurons | Relaxed |
| PAPL | 0.00102 | Iron/zinc purple acid phosphatase-like protein | | Relaxed |
| GRIN2C | 0.00107 | Subunit of the NMDA receptor in central nervous system | Nervous system | Relaxed |
| GRIN3B | 0.00112 | Subunit of the NMDA receptor in central nervous system | Nervous system | Relaxed |
| FER1L4 | 0.00114 | Fer-1-like protein 4 | | Relaxed |
| KRT80 | 0.00114 | Keratin 80, type II epithelial keratin | Skin | Relaxed |
| GLRA4 | 0.00122 | Glycine receptor, alpha 4, neurotransmitter-gated ion channel | Nervous system | Relaxed |
| RAG1 | 0.00122 | VDJ-recombination activating gene 1 | Immune system | Relaxed |
| DSP | 0.00134 | Desmoplakin; obligate component of functional desmosomes | Skin | Relaxed |
| CAB39L | 0.00140 | Calcium binding protein 39-like | | Relaxed |
| OR52D1 | 0.00146 | Olfactory receptor, family 52, subfamily D, member 1 | Olfaction | Relaxed |
| STOML3 | 0.00150 | Stomatin (EPB72)-like 3 | | Relaxed |
| PIK3R5 | 0.00153 | Phosphoinositide-3-kinase, regulatory subunit 5 | | Relaxed |
| RNF222 | 0.00154 | Ring finger protein 222 | | Relaxed |
| OR10Z1 | 0.00164 | Olfactory receptor, family 10, subfamily Z, member 1 | Olfaction | Relaxed |
| TGM3 | 0.00166 | Transglutaminase 3, cell envelope formation in the epidermis and hair follicle | Skin, Hair | Pos. selection |
| SFTPB | 0.00168 | Lung-specific surfactant protein B | Lung | Relaxed |

*P values refer to the Wilcoxon-rank sum test for rate acceleration on marine branches when compared with terrestrial branches.

evolutionary rates on marine branches, suggesting these genes experienced a greater pressure to adapt in the marine environment (supplementary table S3, Supplementary Material online).

## Marine-Accelerated Muscle and Sensory Proteins Were under Relaxed Constraint

The remaining marine-accelerated genes represent apparent cases of relaxed constraint, which include relaxation with continued functionality and complete loss of function. It is notable however that some of these genes may also have experienced short episodes of positive selection that are not detected in statistical models but which nonetheless led to rate acceleration (Anisimova et al. 2001). The four muscle-related genes seem to remain functional in most marine species, because two of them did not show genetic lesions in any marine species, β-myosin heavy chain (*MYH7*) and a muscle-specific glycolysis enzyme (*PGAM2*) (supplementary table S1 and fig. S1, Supplementary Material online alignments in supplementary data). The other muscle genes, nebulin and myoferlin, had early stop codons only in the dolphin sequence. The reason behind the relative rate increases for these muscle genes could be related to the relatively large body size of marine mammals, because a control analysis analyzing convergent rate changes for a set of large-bodied mammals (alpaca, camel, elephant, horse, and rhino) also showed significant rate increases for genes in this category (supplementary table S7, Supplementary Material online).

The strongest case of gene loss was seen in three digestion and absorption genes that, in addition to their marine-acceleration, showed obvious genetic lesions in 60% of marine species on average, compared with 7% in terrestrial species (permutation $P$ value $< 0.00001$; supplementary table S6, Supplementary Material online). These genes were involved in transport in gastric surface mucus cells (*CAPN8*), salivary gland aldehyde dehydrogenase activity (*ALDH3B2*), and intestinal water permeability (aquaporin, *AQP10*). Many marine-accelerated genes were gustatory and olfactory genes encoding chemoreceptors, G-protein signaling proteins, and neuronal proteins specialized for taste and smell (table 1 and fig. 3; supplementary table S2, Supplementary Material online). Almost all of these olfactory and gustatory marine-accelerated genes showed evidence of relaxed constraint in marine species (supplementary table S1, Supplementary Material online), and a high number of genetic lesions, 29% in marine species versus 12% in terrestrial species (permutation $P$ value $< 0.00001$). Taste-related genes showed strongly convergent rate increases in that two of the top three marine-accelerated genes were specific to gustatory signaling: G-

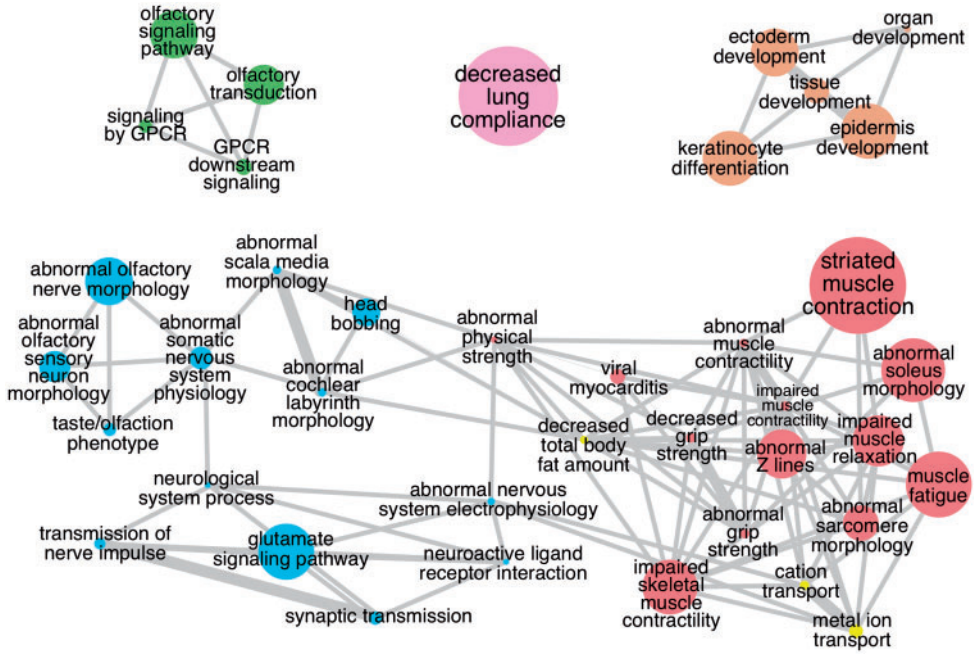


**Fig. 3.** Marine-accelerated genes are strongly enriched for interrelated functional categories. Many functions enriched in the top 500 marine-accelerated genes are interconnected through shared genes, as reflected in this network. Each node represents a significantly enriched category with its node diameter proportional to the degree of enrichment (range = 2.3- to 17.7-fold). Displayed categories were restricted to those with enrichment above 2-fold and a FDR q-value below 5%. Edges between nodes reflect shared accelerated genes between categories, and their width is proportional to the degree of sharing (range = 21%–100% shared gene content). The resulting networks group broad biological functions in to olfaction, lung, skin and connective tissue, nervous system, and muscle.

**Table 2.** Prevalence of Positive Selection and Relaxed Constraint in Major Functional Categories.

| Class | Function | Relaxed Selection | Positive Selection |
|---|---|---|---|
| Sensory | Taste | 2 | 0 (0%) |
| | Olfaction | 29 | 1 (3%) |
| | Nervous system | 14 | 0 (0%) |
| Structural | Cartilage | 2 | 1 (33%) |
| | Skin | 3 | 2 (40%) |
| | Skin and Hair | 0 | 2 (100%) |
| | Hair | 3 | 0 (0%) |
| Metabolism | Lipid | 2 | 0 (0%) |
| | Digestion, Absorption | 3 | 0 (0%) |
| | Glycolysis | 1 | 0 (0%) |
| | Detoxification | 2 | 0 (0%) |
| | Muscle | 4 | 0 (0%) |
| | Lung surfactant | 1 | 0 (0%) |
| | Immune system | 1 | 0 (0%) |
| | Unknown | 41 | 1 (2%) |

protein subunit *GNAT3* (1) and gustatory ion channel *CALHM1* (3) (table 1). This is consistent with recent reports of taste receptor gene loss in toothed and baleen whales (Feng et al. 2014; Zhu et al. 2014). Similarly, we found that *GNAT3* contained obvious genetic lesions in both the killer whale and bottlenose dolphin.

## Uncharacterized Neuronal Genes Likely Contribute to Sensory Functions

There were also several poorly characterized genes that evolved convergently in marine species. Many of these are in gene families affecting axon guidance (*UNC5C*, *EPHA8*, *FEZ1*, *ULK2*), synaptic development and neurotransmission (*CDH12*, *GRIN2C*, *GRIN3B*), and other nervous system processes (*TMEM25*, *OTOF*, *GLRA4*, *CLSTN2*), yet they are not currently ascribed to organism-level neurological processes (supplementary table S1, Supplementary Material online). Given the strong signal of convergent relaxation of constraint in marine sensory systems (supplementary tables S1 and S6, Supplementary Material online), we propose that these genes function in olfactory and/or gustatory neurons in terrestrial mammals. These genes could also be involved in other sensory modalities as we observe acceleration in genes known to affect temperature (*TRPV3*) and pain sensation (*ASIC4* and *TACR2*).

Alternatively, two of the marine lineages (cetaceans and pinnipeds) are vocal learners and convergence acceleration in neuronal genes may represent a shift in selection pressure due to that trait (Petkov and Jarvis 2012; Janik 2014; Jarvis et al. 2014; Reichmuth and Casey 2014). Although it is difficult to advance this hypothesis given that we do not know whether the ancestors of cetaceans or pinnipeds were vocal learners.

## Marine-Accelerated Branches Do Not Contain an Excess of Site-Specific Convergent Amino Acid Changes

We tested the hypothesis that convergent or parallel amino acid changes at specific sites contributed to the reported marine-associated accelerations. We used the branch model of CODEML to reconstruct the substitutions on all branches in a mammalian phylogeny containing up to 49 placental

mammals. We then tallied all convergent amino acid substitutions on marine branches for the top 45 marine-accelerated genes, including both strictly convergent and parallel changes. Strictly convergent refers to a scenario where the ancestors of two marine species had different amino acids at a given site that changed to the same amino acid on both marine branches. Parallel changes refer to two marine ancestors having the same amino acid at a given site which evolved to the same, new amino acid in both extant species. The proportion of all changes that was convergent on marine branches was generally low, with 42% of branches having none (supplementary fig. S3*A*, Supplementary Material online). Convergent changes on marine branches could have resulted from neutral processes or alternatively due to positive selection for the same amino acid variant. For the latter case, we would expect there to be more convergent amino acid changes in these genes for marine species compared to negative control species. We tallied convergent changes within a control set of species not expected to show convergence and chosen to match the topology and branch lengths of the marine species (supplementary fig. S3*C*, Supplementary Material online) (alpaca, camel and their ancestral branch, bushbaby, human, and aardvark). As an additional control, we tallied the convergent changes between the marine and control branches (supplementary fig. S3*D*, Supplementary Material online). The marine branch data set did not show an excess of convergent amino acid sites (mean proportion = 0.086) compared to the control data sets (0.088 and 0.052, for control branches and marine- vs. -control branches, respectively) (supplementary fig. S3, Supplementary Material online). Furthermore, the proportion of convergent changes on marine branches is small compared with the amount of excess changes that led to the acceleration in relative rate (mean proportion = 0.549; supplementary fig. S3*B*, Supplementary Material online). Overall, we found no evidence for adaptively driven site- and amino acid-specific convergence in marine-accelerated genes.

## Discussion

Across three independent lineages, the shift from a terrestrial to marine environment dictated convergent shifts in evolutionary selective pressures for hundreds of genes, resulting in marine-specific accelerations and decelerations. Namely, we observed genes shift specifically in marine species to evolutionary modes of low constraint, higher constraint, and even adaptive evolution. These marine-accelerated and marine-decelerated genes were present in statistical excess and were highly enriched for functions that agree with and extend our knowledge of physiological adaptations in marine mammals (figs. 1*B* and 3). Particularly, strong categories were sensory genes mediating smell and taste, genes forming skin and connective tissue, and genes involved in muscle structure and metabolism.

We observed strong evidence of positive selection and a marine-acceleration in a large number of skin-associated proteins. For example, there was marine-acceleration and positive selection in representatives of nearly all components of the desmosome, a structure that forms cell-cell junctions in skin and muscle epithelium. A likely hypothesis for this positive selection is that the remodeling of skin and connective tissue in marine mammals required additional amino acid changes. The selected changes may be due to either specific anatomical change to skin and connective tissue (Meyer et al. 2011) or alternatively, to exposure to a greater number and diversity of pathogens in the marine environment. Indeed, the oceans harbor an enormous diversity and number of viruses and prokaryotes, which could frequently come in to contact with marine mammal skin (Whitman et al. 1998; Suttle 2007). In a similar case, there was also marine-acceleration and positive selection for a gene encoding a lung surfactant protein (*SFTPB*). This membrane-bound protein ensures rapid distribution of pulmonary surfactant, and so it is plausible that its rate increase was due to respiratory adaptations for diving; however, it must also be considered that its rate increase may have been due to pathogen pressures as well.

On the other hand, marine-accelerated genes participating in olfaction, gustation, and muscle function exhibited overwhelming evidence of relaxation of constraint. These observations included greatly accelerated rates consistent with neutral evolution and even obvious genetic lesions and pseudogenization (e.g., *GNAT3*). While the loss of taste in cetaceans is well established, its state in pinnipeds and manatees is still debated (Bills 2011; Sato and Wolsan 2012). Taste-specific genes *GNAT3* and *CALHM1* were both highly accelerated and found to be under relaxed constraint in pinnipeds and the manatee, suggesting that these marine taxa also have a reduced sense of taste. Furthermore, consistent with our findings of such convergence in the marine environment, a recent report found loss of taste receptors in penguins (Zhao et al. 2015). This convergent reduction of taste in marine mammals and birds has been suggested to result from change in prey, swallowing food whole, and the masking of tastes by seawater (Sato and Wolsan 2012). We found olfactory genes were also marine-accelerated and under relaxed constraint in marine species, which is consistent with reports of the loss of olfactory receptors, the olfactory bulb, and the cranial nerve in toothed whales (Odontoceti) (Marino 2007; Oelschläger et al. 2008; Oelschläger et al. 2010) and a reduced olfactory epithelium in manatees relative to terrestrial mammals (Mackay-Sim et al. 1985; Bills 2011). Finally, the convergent relaxation of constraint on muscle-specialized proteins may have resulted from changes in contractile properties, body size, or metabolic demands in periods of anoxia during dives.

### Comparison of Approaches to Identify Convergent Evolution

A previous study of molecular convergence in marine mammals employed the hypothesis that convergence would be seen as single amino acid sites changing to the same amino acid. They did not find an excess of such changes over background (Foote et al. 2015). We also asked if our marine-accelerated genes experienced such amino acid site convergence. We did not find an excess of convergent changes on marine branches when compared to negative control branches (supplementary fig. S3, Supplementary Material online). It has been reported that proteome-wide searches for

convergent amino acids should have low power and that observed convergent changes can be sufficiently explained by neutral models that do not rely on adaptively driven convergence (Zhang and Kumar 1997; Thomas and Hahn 2015; Zou and Zhang 2015a). As an alternative approach, we introduced a novel analytic framework that focuses on those genes showing convergent acceleration or deceleration in their whole-gene evolutionary rates. This strategy represents a more course-grained perspective to identify common sets of genes and pathways that experienced a shift in their evolutionary pressures during phenotypic convergence. Our application of this gene-centric strategy to marine mammals uncovered a strong statistical excess of convergent shifts in evolutionary rate (fig. 1*B*). This suggests that convergent changes in selection at the gene level is relatively common compared to convergence at single amino acids, and that its signal is more readily detected. These two approaches operated at distinct levels but employed the same basic data set. Accordingly, we asked if they detected a shared set of genes. We compared our top 690 marine-accelerated genes with nominal *P* values below 0.05 to the 15 genes reported by Foote et al. in their table 1 as being both under positive selection in marine lineages and containing convergent amino acid changes. There was an overlap of two muscle-related genes, *MYH7B* and *SMPX*. While an overlap of two genes (13%) is at higher frequency than expected (3.8%), it was not entirely unlikely by chance ($P = 0.11$, binomial test). However, the comparison suffered from low power due to a low number of genes identified by Foote et al., so it is still plausible that the two approaches could be recovering similar targets in their quest for convergent evolution.

Foote et al. also reported genes under positive selection in one or multiple marine branches using the branch-site models of CODEML (their supplementary tables S3 and S9–S11, Supplementary Material online) (Yang 2007; Foote et al. 2015). However, the tests in Foote et al. did not exclude the presence of positive selection in terrestrial species. To illustrate, the gene *ANPEP*, which was featured in Foote et al., yields strong evidence for positive selection across all branches in the mammalian tree (CODEML sites models M1 vs. M2, $P = 1.8 \times 10^{-21}$). When all marine species are removed, *ANPEP* still shows strong evidence of positive selection, indicating that positive selection in *ANPEP* is not restricted to marine species (M1 vs. M2, $P = 2.7 \times 10^{-10}$; supplementary table S8, Supplementary Material online). Four of the five genes that passed Foote et al.'s correction for multiple testing (*ANPEP*, *CD48*, *EMP1*, and *MUC1*) show the same evidence of positive selection outside of marine species (supplementary table S8, Supplementary Material online). The exception was *ZNF582*, for which our alignment showed no evidence of positive selection in any model comparison, including the marine-species branch-site test. Our method, on the other hand, uses a different criterion that specifically tests for convergent acceleration or deceleration in marine branches when compared with terrestrial branches. The resulting gene set contains genes that convergently experienced positive selection or relaxed constraint (marine-accelerated) or additional constraint (marine-decelerated) in the marine environment. For these reasons, our studies report very different genes.

There are clear advantages and limitations to our gene-centric strategy. Generally, we found that a search for convergent rates in protein-coding genes can excel at revealing adaptations in physiological and structural genes. However, genes controlling convergent morphological adaptations were not identified. We expect this limitation is due to the fact that morphological changes typically involve regulatory rather than protein-coding changes (Carroll 2008); however, this suggests a potential extension to study rate changes in *cis*-regulatory regions as well. A notable advantage to our method is that it reveals convergently evolving biological processes even when nothing is known about adaptation *a priori*. Our approach applied no prior hypotheses about which genes or functions would be marine-accelerated or decelerated, yet it recovered genes significantly enriched for functions that are altered in marine mammals (fig. 3). Furthermore, it may be useful to assign probable functions to poorly characterized but convergently evolving genes, as we demonstrated for 11 neuronal genes likely participating in sensory perception. Overall, searching for convergent shifts in evolutionary rate represents a promising strategy to identify candidate genes underlying convergent phenotypic traits, as demonstrated in this and past studies (Goodman et al. 2009). The study of marine mammals is just the first application of our relative evolutionary rates approach. It can be readily applied to essentially any environmental or phenotypic variable with sufficient range in the mammalian phylogeny, or other taxonomic group for that matter.

## Materials and Methods

### Calculating Correlations of Protein-Coding Genes with the Marine Environment

Amino acid alignments of orthologous genes were derived from the "100-way" 100 vertebrate species genomic multiz alignment available at the UCSC genome browser (Blanchette et al. 2004), we studied those alignments with a minimum of 30 species. For each amino acid alignment we estimated branch lengths using the AAML program from the phylogenetic analysis using maximum likelihood (PAML) package (Yang 2007). Branch lengths were estimated under an empirical model of amino acid substitution rates with rate variability between sites modeled as a gamma distribution approximated with four discrete classes (for computational efficiency) and an additional class for invariable sites (AAML model "Empirical + F") (Yang 1996; Whelan and Goldman 2001). Branch lengths were estimated by AAML on a published mammalian species tree topology (Murphy et al. 2004). Raw branch lengths were transformed into relative rates using a projection operator method (Sato et al. 2005). This method begins with each gene's branch lengths in the form of a vector that we then scale to a vector length of 1. The next step is to transform these branch lengths in to relative rates. Relative rates quantify how much faster or slower this gene changed on a given branch after factoring out the divergence on that branch resulting from parameters affecting all genes (e.g., the

time since speciation, effective population size, mutation rate). Removing these genome-wide effects will leave only the gene-specific relative rate for a given branch. Relative rates are calculated using a normalization vector that is calculated as the average amount of divergence on each branch across all scaled genes. Finally, we determine the relative rate of each branch as the residual of that branch after factoring out the normalization vector. A relative rate vector is thus defined as the component of the branch-length vector that is orthogonal to the normalization vector. These branch-specific relative rates were then used to perform a Wilcoxon rank sum test and correlation analysis over the binary variable of "terrestrial" or "marine" branches (fig. 1A). Marine branches are those leading to the manatee (*Trichechus manatus latirostris*), Weddell seal (*Leptonychotes weddellii*), walrus (*Odobenus rosmarus divergens*), killer whale (*Orcinus orca*), bottlenose dolphin (*Tursiops truncatus*), and the cetacean ancestral branch leading to the killer whale and dolphin. Control branches for *P* value enrichments in figure 1 were those leading to the aardvark (*Orycteropus afer afer*), alpaca (*Vicugna pacos*), camel (*Camelus ferus*), microbat (*Myotis lucifugus*), and David's myotis bat (*Myotis davidii*), as well as the ancestral branch leading to alpaca and camel. Large-body control branches were those leading to alpaca (*V. pacos*), camel (*C. ferus*), elephant (*Loxodonta africana*), horse (*Equus caballus*), and white rhino (*Ceratotherium simum*).

### Functional Enrichment Analysis

Functional information for marine-accelerated and marine-decelerated genes in table 1 was taken from the Uniprot and RefSeq databases, and from literature cited directly (Pruitt et al. 2007; Uniprot Consortium 2007). Computational tests for functional enrichment were performed using the hypergeometric test with the background set of genes restricted to genes that were tested for marine convergence and had at least one annotation in the corresponding annotation file. Correction for multiple testing was performed using false discovery rate (FDR) q-values (Storey 2002). We used two sources of annotations for computational tests for functional enrichment, the "canonical pathways" from MSigDB (Liberzon et al. 2011) and mammalian phenotypes from MGI (Smith and Eppig 2009). The mammalian phenotype annotations were compiled by associating gene symbols listed in the genotype name to the reported phenotypes and all their ancestors in the mammalian phenotype ontology (https://bioportal.bioontology.org/ontologies/MP, last accessed June 6, 2016). Enriched categories are displayed in a network in figure 3. In this network we required a minimum of 2-fold enrichment within the 500 most marine-accelerated genes and an FDR q-value below 5%. Edges were drawn between categories if they shared a minimum of 20% of their accelerated gene content as calculated using the smaller of the two categories. Accordingly, a value of 100% results when all of a category's accelerated genes are found in the other category. The network in figure 3 was visually laid out by an edge-weighted force-directed algorithm in Cytoscape (Killcoyne et al. 2009).

### FDR Analysis

FDR analysis was performed on probabilities resulting from the Wilcoxon rank sum test. FDRs and the proportion of tests conforming to the null hypothesis were estimated by a modified version of the "qvalue" function in the corresponding "qvalue" R module (Storey and Tibshirani 2003; Storey 2015). The q-value method first estimates $\pi_0$ (the proportion of tests for which the null hypothesis holds) by comparing the cumulative distribution of observed *P* values to a uniform distribution. The $\pi_0$ value is subsequently used to calculate a FDR for each *P* value threshold. We modified the $\pi_0$ calculation by replacing expected uniform quantiles with the observed null distribution quantiles calculated with the control species set. This approach parallels the empirical permutation-based FDR calculations that are often employed in genomic data analysis which are typically more conservative as they preserve dependence in the data structure. In our case this produced FDR estimates that are slightly more conservative because the *P* value distribution from our control species was not perfectly uniform. The slight departure from uniformity in the control species was seen as a slight enrichment of small *P* values due to nonindependence arising from phylogenic relationships.

### Phylogenetic Models of Selective Pressure

Of the marine-accelerated genes, the top 115 were subjected to phylogenetic models of codon evolution in order to test for significant evidence of relaxation of constraint or positive selection over the marine mammal branches. We used CODEML of the PAML package to run three models: the branch-site neutral model (BS1), the branch-site selection model (BS2), and a neutral sites model (M1) (Yang 2007). Likelihood ratio tests (LRT) between BS1 and its nested null model M1 were used to assign significance of relaxation of constraint on marine mammal branches specifically. LRTs between BS2 and its null BS1 were used to infer positive selection on the marine mammal branches (Zhang et al. 2005). Probabilities were assigned for each of these two LRTs using the chi-square distribution with 1 degree of freedom. In a similar way, mammal-wide positive selection was inferred using the M8 (selection) and M8A (neutral) sites models and their respective LRT (Swanson et al. 2004). Significance for this LRT was assigned with the chi-square distribution with 1 degree of freedom, which is a conservative assumption (Wong et al. 2004).

## Supplementary Material

Supplementary materials is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

This study was supported by a Charles E. Kaufman New Investigator Award from The Pittsburgh Foundation to N.L.C., an National Institutes of Health grant from the National Human Genome Research Institute (U54HG008540) to M.C. and N.L.C., and a Research Experience for Undergraduates grant funded by National

Science Foundation and the Department of Defense (NSF-DBI 1263020). The authors thank the Vertebrate Genome Biology Group at the Broad Institute for allowing us to use the unpublished Weddell seal genome for this study.

## References


Andersen HT. 1966. Physiological adaptations in diving vertebrates. *Physiol Rev*. 46:212–243.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*. 18:1585–1592.

Arnason U, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, Väinölä R. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol Phylogenet Evol*. 41:345–354.

Bailey SF, Rodrigue N, Kassen R. 2015. The effect of selection environment on the probability of parallel evolution. *Mol Biol Evol*. 32:1436–1448.

Bills ML. 2011. Description of the chemical senses of the Florida manatee, Trichechus manatus latirostris, in relation to reproduction. [dissertation]. [Gainsville (FL)]: University of Florida.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 14:708–715.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36.

Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One* 3:e2201.

Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A*. 106:8986–8991.

Christin P-A, Salamin N, Muasya AM, Roalson EH, Russier F, Besnard G. 2008. Evolutionary switch and genetic convergence on rbcL following the evolution of C4 photosynthesis. *Mol Biol Evol*. 25:2361–2368.

Christin P-A, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet*. 26:400–405.

Clark NL, Alani E, Aquadro CF. 2013. Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics* 193:529–538.

Davies KTJ, Cotton JA, Kirwan JD, Teeling EC, Rossiter SJ. 2012. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. *Heredity* 108:480–489.

Dobler S, Dalla S, Wagschal V, Agrawal AA. 2012. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc Natl Acad Sci U S A*. 109:13040–13045.

Feng P, Zheng J, Rossiter SJ, Wang D, Zhao H. 2014. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol. Evol*. 6:1254–1265.

Fish FE, Howle LE, Murray MM. 2008. Hydrodynamic flow control in marine mammals. *Integr Comp Biol*. 48:788–800.

Fish FE, Hui CA. 1991. Dolphin swimming–a review. *Mamm Rev*. 21:181–195.

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

Goodman M, Sterner KN, Islam M, Uddin M, Sherwood CC, Hof PR, Hou Z-C, Lipovich L, Jia H, Grossman LI, et al. 2009. Phylogenomic analyses reveal convergent patterns of adaptive evolution in elephant and human ancestries. *Proc Natl Acad Sci U S A*. 106:20824–20829.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet*. 4:e1000303.

Janik VM. 2014. Cetacean vocal learning and communication. *Curr Opin Neurobiol*. 28:60–65.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.

Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genet*. 6:e1000924.

Killcoyne S, Carter GW, Smith J, Boyle J. 2009. Cytoscape: a community-based framework for network modeling. *Methods Mol Biol*. 563:219–239.

Kooyman GL. 2006. Mysteries of adaptation to hypoxia and pressure in marine mammals. *Mar Mam Sci*. 22:507–526.

Kulathinal RJ, Bettencourt BR, Hartl DL. 2004. Compensated deleterious mutations in insect genomes. *Science* 306:1553–1554.

Leys R, Cooper SJ, Strecker U, Wilkens H. 2005. Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. *Biol Lett*. 1:496–499.

Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Curr Biol*. 20:R55–R56.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739–1740.

Liu Y, Cotton JA, Shen B, Han X, Rossiter SJ, Zhang S. 2010. Convergent sequence evolution between echolocating bats and dolphins. *Curr Biol*. 20:R53–R54.

Mackay-Sim A, Duvall D, Graves BM. 1985. The West Indian manatee (Trichechus manatus) lacks a vomeronasal organ. *Brain Behav Evol*. 27:186–194.

Marino L. 2007. Cetacean brains: how aquatic are they? *Anat Rec*. 290:694–700.

McGowen MR, Grossman LI, Wildman DE. 2012. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc Biol Sci*. 279:3643–3651.

Meyer W, Schmidt J, Kacza J, Busche R, Naim HY, Jacob R. 2011. Basic structural and functional characteristics of the epidermal barrier in wild mammals living in different habitats and climates. *Eur J Wildlife Res*. 57:873–885.

Murphy WJ, Pevzner PA, O'Brien SJ. 2004. Mammalian phylogenomics comes of age. *Trends Genet*. 20:631–639.

Oelschläger HHA, Haas-Rioth M, Fung C, Ridgway SH, Knauth M. 2008. Morphology and evolutionary biology of the dolphin (Delphinus sp.) brain–MR imaging and conventional histology. *Brain Behav Evol*. 71:68–86.

Oelschläger HHA, Ridgway SH, Knauth M. 2010. Cetacean brain evolution: dwarf sperm whale (Kogia sima) and common dolphin (Delphinus delphis) — An investigation with high-resolution 3D MRI. *Brain Behav Evol*. 75:33–62.

Petkov CI, Jarvis ED. 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front Evol Neurosci*. 4:12.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 35:D61–D65.

Reichmuth C, Casey C. 2014. Vocal learning in seals, sea lions, and walruses. *Curr Opin Neurobiol*. 28:66–71.

Rosen DAS, Renouf D. 1997. Seasonal changes in blubber distribution in atlantic harbor seals: indications of thermodynamic considerations. *Mar Mammal Sci*. 13:229–240.

Rybczynski N, Dawson MR, Tedford RH. 2009. A semi-aquatic arctic mammalian carnivore from the miocene epoch and origin of Pinnipedia. *Nature* 458:1021–1024.

Sato JJ, Wolsan M. 2012. Loss or major reduction of umami taste sensation in pinnipeds. *Naturwissenschaften* 99:655–659.

Sato T, Yamanishi Y, Kanehisa M, Toh H. 2005. The inference of protein-protein interactions by co-evolutionary analysis is improved by

excluding the information about the phylogenetic relationships. *Bioinformatics* 21:3482–3489.
Shen Y-Y, Liang L, Li G-S, Murphy RW, Zhang Y-P. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLoS Genet*. 8:e1002788.
Smith CL, Eppig JT. 2009. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med*. 1:390–399.
Spor A, Kvitek DJ, Nidelet T, Martin J, Legrand J, Dillmann C, Bourgais A, de Vienne D, Sherlock G, Sicard D. 2014. Phenotypic and genotypic convergences are influenced by historical contingency and environment in yeast. *Evolution* 68:772–790.
Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet*. 14:751–764.
Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 100:9440–9445.
Storey JD. 2002. A direct approach to false discovery rates. *J R Stat Soc Ser B (Statistical Methodol.)* 64:479–498.
Storey JD. 2015. qvalue: Q-value estimation for false discovery rate control. [Software package]. Bioconductor [cited 2016 Jun 6]. Available from: https://bioconductor.org/packages/release/bioc/html/qvalue.html.
Sun Y-B, Zhou W-P, Liu H-Q, Irwin DM, 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.
Suttle CA. 2007. Marine viruses–major players in the global ecosystem. *Nat Rev Microbiol*. 5:801–812.
Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168:1457–1465.
Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Mol Biol Evol*. 32:1232–1236.
Uniprot Consortium. 2007. The universal protein resource (UniProt). *Nucleic Acids Res*. 35:D193–D197.
Wartzok D, Ketten DR. 1999. Marine mammal sensory systems. In: Reynolds J, Rommel S, editors. Biology of marine mammals. Washington (DC): Smithsonian Institution Press. p. 117–175.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18:691–699.
Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A*. 95:6578–6583.
Whittow GC. 1987. Thermoregulatory adaptations in marine mammals: interacting effects of exercise and body mass a review. *Mar Mammal Sci*. 3:220–241.
Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051.
Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol*. 11:367–372.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK, et al. 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet*. 46:88–92.
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.
Zhao H, Li J, Zhang J. 2015. Molecular evidence for the loss of three basic tastes in penguins. *Curr Biol*. 25:R141–R142.
Zhu K, Zhou X, Xu S, Sun D, Ren W, Zhou K, Yang G. 2014. The loss of taste genes in cetaceans. *BMC Evol Biol*. 14:218.
Zou Z, Zhang J. 2015a. No genome-wide protein sequence convergence for echolocation. *Mol Biol Evol*. 32:1237–1241.
Zou Z, Zhang J. 2015b. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol Biol Evol*. 32:2085–2096.