EVOLUTIONARY BIOLOGY

Genomic and functional evidence reveals molecular insights into the origin of echolocation in whales

Zhen Liu[1]*, Fei-Yan Qi[1,2]*, Dong-Ming Xu[1]*, Xin Zhou[1], Peng Shi[1,3,4]†

Echolocation allows toothed whales to adapt to underwater habitats where vision is ineffective. Because echolocation requires the ability to detect exceptional high-frequency sounds, fossils related to the auditory system can help to pinpoint the origin of echolocation in whales. However, because of conflicting interpretations of archaeocete fossils, when and how whales evolved the high-frequency hearing correlated with echolocation remain unclear. We address these questions at the molecular level by systematically investigating the convergent evolution of 7206 orthologs across 16 mammals and find that convergent genes between the last common ancestor of all whales (LCAW) and echolocating bats are not significantly enriched in functional categories related to hearing, and that convergence in hearing-related proteins between them is not stronger than that between nonecholocating mammalian lineages and echolocating bats. However, these results contrast with those of parallel analyses between the LCA of toothed whales (LCATW) and echolocating bats. Furthermore, we reconstruct the ancestral genes for the hearing protein *prestin* for the LCAW and LCATW; we show that the LCAW *prestin* exhibits the same function as that of nonecholocating mammals, but the LCATW *prestin* shows functional convergence with that of extant echolocating mammals. Mutagenesis shows that functional convergence of prestin is driven by convergent changes in the prestins S392A and L497M in the LCATW and echolocating bats. Our results provide genomic and functional evidence supporting the origin of high-frequency hearing in the LCAW, not the LCATW, and reveal molecular insights into the origin and evolutionary trajectories of echolocation in whales.

INTRODUCTION

The living cetaceans (Neoceti) evolved from archaeocetes (ancestral fossil whales) are divided into two highly distinct suborders: Mysticeti (baleen whales) and Odontoceti (toothed whales) (*1*–*3*). Unlike baleen whales, toothed whales locate, range, and hunt in dark and turbid aquatic environments using echolocation, an ability to perceive the environment that largely depends on high-frequency hearing (*4*). The monophyly of toothed whales implies that echolocation evolved either in the lineage of the last common ancestor of toothed whales (LCATW; branch IV in Fig. 1A) or in the lineage of the LCA of all whales (LCAW; branch II in Fig. 1A) and was subsequently lost in the lineage of the LCA of baleen whales (LCABW). Although fossil features strongly support the evolution of ultrasonic hearing that is closely related to echolocation in the basal odontocetes (*5*–*7*), it remains unclear whether high-frequency hearing originated in the archaeocetes (*8*, *9*) because on the basis of fossil records, some studies suggest that archaeocetes were capable of high-frequency hearing (*5*, *10*), whereas others support the hypothesis that archaeocetes had low-frequency hearing (*6*, *11*, *12*).

This fundamental question in cetacean biology could be tested by systematically searching for convergent amino acid substitutions associated with convergent phenotypes of high-frequency hearing because several studies have observed molecular convergence/parallelism in some hearing genes between echolocating bats and toothed whales, in which echolocation independently evolved (*13*–*17*). The phylogenetic reconstructions of the protein sequences for these genes generate monophyletic clades of all echolocators that exclude their more closely related, nonecholocating taxa (*13*, *18*), suggesting that these hearing genes may play important roles in the occurrence and/or development of echolocation in mammals. In addition, functional examinations of the echolocation-related genes have demonstrated significant differences between echolocating and nonecholocating mammals (*19*), suggesting that functional alterations of echolocation-related genes from ancestral whales can offer promising evidence to address when and how high-frequency hearing evolved in whales.

By reconstructing ancestral gene sequences, we screened molecular convergence between the LCAW and echolocating bats on a genome-wide scale and found no significant enrichment of convergent genes in functional categories related to hearing and no more significant convergent signatures in hearing-related proteins relative to a control; but these results were opposite to the patterns that we observed between the LCATW and echolocating bats. Furthermore, we resurrected ancestral *prestin* genes from the key nodes on the species tree of whales and found that the LCAW and LCABW *prestin* genes exhibited functional similarity with those from the extant nonecholocating mammals; however, the LCATW *prestin* displayed functional convergence with those from living echolocating mammals that was driven by adaptive convergent substitutions. Therefore, our findings from computational and experimental analyses strongly support the hypothesis that high-frequency hearing more likely originated in the LCATW rather than in the LCAW.

RESULTS

Accessing levels of genomic convergence in hearing-related proteins

Echolocation has evolved independently in bats and whales as an adaptation to their respective environments. Previous studies have shown that some hearing genes have convergently evolved in echolocating bats and toothed whales (*13*–*15*, *17*) as well as among echolocating bats (*18*, *20*). Hence, if echolocation originated in the LCAW, then we

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650223, China. [3]School of Future Technology, University of Chinese Academy of Sciences, Beijing, China. [4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China.
*These authors contributed equally to this work.
†Corresponding author. Email: ship@mail.kiz.ac.cn

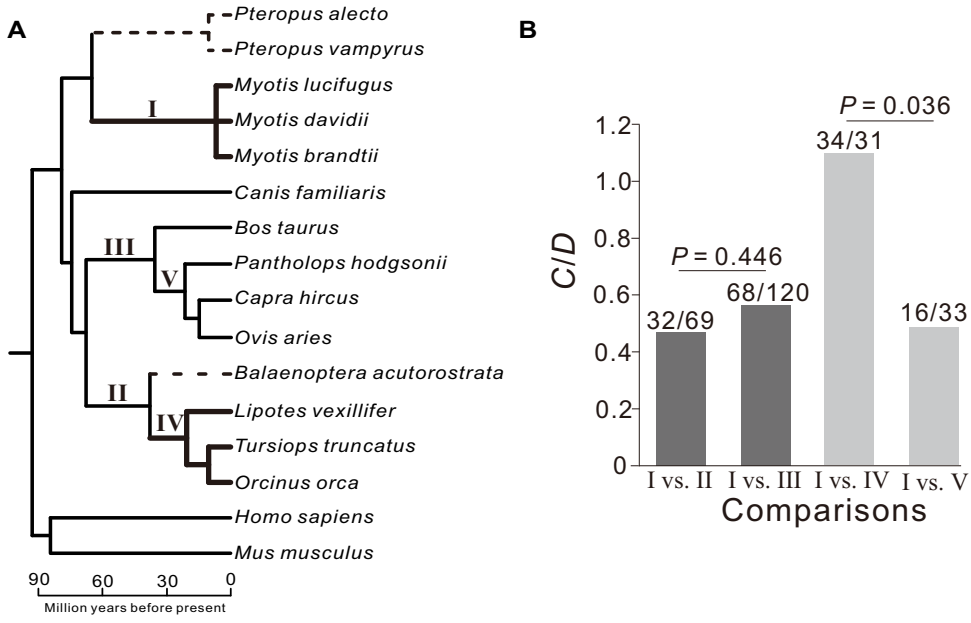

Fig. 1. Detection of convergence between echolocating bats and ancestral whales. (**A**) Phylogeny of the mammalian species used to detect molecular convergence in this study. Bold lineages indicate toothed whales and echolocating bats with high-quality genomic data, and dashed lineages indicate the nonecholocating baleen whales and Old World fruit bats. The branches labeled I, II, III, IV, and V denote where convergent sites are counted. (**B**) Different comparisons of *C/D*s of hearing-related genes based on the data set containing the inferred amino acids with ≥0.95 posterior probabilities. The numbers of convergent and divergent sites are given on the bars. *C/D* indicates the ratio of the number of convergent sites to the number of divergent sites. The *P* values are from two-tailed χ^2 tests.

could predict that (i) convergent genes between echolocating bats and the LCAW (branch I versus branch II; Fig. 1A) should be enriched in the functional categories associated with hearing compared to those identified from their equally phylogenetically distant control (branch I versus branch III; Fig. 1A), and (ii) more convergent amino acids in hearing-related proteins should be observed between branches I and II than between branches I and III. To verify these predictions, we inferred ancestral amino acid sequences and counted the number of convergent sites for the relevant lineages as a direct measurement of molecular convergence. In total, 7206 one-to-one orthologous proteins across 16 mammalian species were analyzed (Fig. 1A). On the basis of the data set containing the inferred amino acids with ≥0.95 posterior probabilities (PP0.95), we identified 932 and 1924 genes with convergence in comparison I versus II and comparison I versus III, respectively, and found that no functional category associated with hearing is significantly enriched for both data sets of convergent genes (table S1). Next, 104 hearing-related genes were classified from 7206 orthologs (see Materials and Methods and table S2 for details), and we counted the numbers of their convergent sites and divergent sites that we used as a control because the number of convergent sites is expected to be proportional to the number of divergent sites under no adaptive convergence (*21*). In comparison, the ratio of the number of convergent sites to the number of divergent sites (*C/D*) between branches I and II is not significantly different from that between branches I and III ($P = 0.446$, two-tailed χ^2 test; Fig. 1B). To further confirm this result, we randomly picked 104 genes from the remaining 7102 genes that are unrelated to hearing and compared *C/D* for these genes with that for hearing-related genes. This comparison was repeated 1000 times and showed that convergence in the genes related to hearing is not significantly different from that in the genes unrelated to hearing in both comparison I and II ($P = 0.608$; Fig. 2A) and comparison I and III ($P = 0.281$; Fig. 2B).

Using the same methods, we examined molecular convergence between echolocating bats and the LCATW (branches I and IV) as well as its equally distant control group (branches I and V; Fig. 1A), and we identified a total of 762 convergent sites in 388 genes between branches I and IV and 877 convergent sites in 436 genes between branches I and V across 7602 orthologous genes from 16 mammals. The number of convergent genes between echolocating bats and toothed whales is not significantly different from that between their comparable nonecholocating lineages ($P = 0.085$, two-tailed χ^2 test). Similarly, the *C/D* ratio between branches I and IV (762:1501) is not significantly different from that between branches I and V (877:1853; $P = 0.247$, two-tailed χ^2 test). These results strongly support the hypothesis that there is no greater convergence on a genome-wide scale for the species with convergent phenotypes than for their equally distant controls (*22*–*24*).

In functional enrichment analyses, three categories related to hearing are significantly enriched for the 388 convergent genes between branches I and IV, but no such categories are significantly enriched for the 436 convergent genes between branches I and V (Table 1). In addition, significantly more convergent sites are identified in hearing-related proteins between branches I and IV than between branches I and V when given their respective numbers of divergent sites ($P = 0.036$, two-tailed χ^2 test; Fig. 1B). Moreover, the convergence between branches I and IV is significantly stronger in hearing-related proteins than in the proteins unrelated to hearing ($P = 0.009$; Fig. 2C), but the convergence in hearing-related proteins between branches I and V is not significantly different from that in the proteins unrelated to hearing ($P = 0.467$; Fig. 2D).

All the above results were derived from the analyses based on the PP0.95 data set. To evaluate the impacts of different cutoffs of posterior probabilities on our results, we generated two additional data sets containing the inferred amino acids with ≥0.7 and ≥0.5 posterior

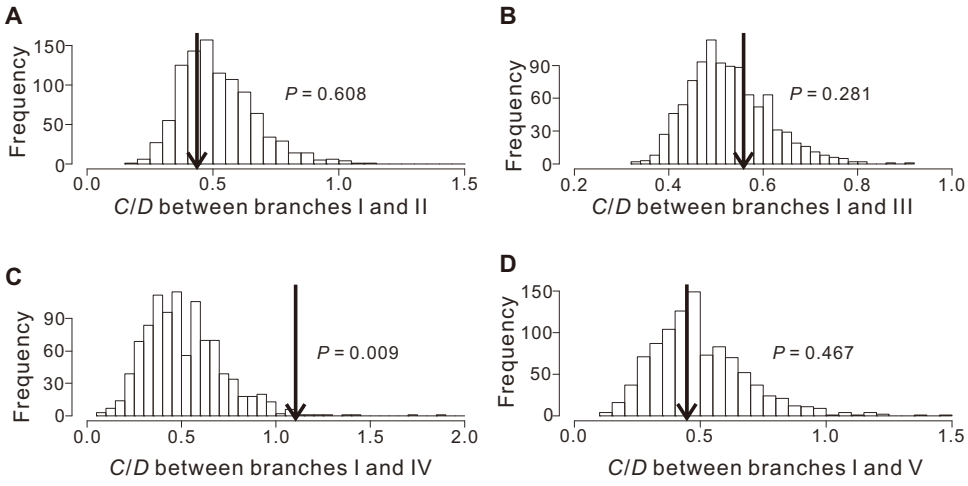

Fig. 2. Evolutionary convergence is significantly higher for hearing genes than for nonhearing genes between LCATW and echolocating bats. Frequency distributions of *C/D*s from 104 genes unrelated to hearing for a total of 1000 random sets in the I and II (**A**), I and III (**B**), I and IV (**C**), and I and V (**D**) comparisons based on the data set containing the inferred amino acids with ≥0.95 posterior probabilities. The arrow in each panel indicates the ratio of the number of convergent sites to the number of divergent sites (*C/D*) identified from 104 hearing-related genes from different comparisons. *C/D* indicates the ratio of the number of convergent sites to the number of divergent sites.

probabilities (PP0.7 and PP0.5), respectively, and performed the same analyses as those of the PP0.95 data set for each of them. For the three data sets, the observations are highly similar in terms of the comparisons between branch sets (I and II) and branch sets (I and III) and between branch sets (I and IV) and branch sets (I and V). Briefly, the hearing-related genes have no greater convergence both in the numbers of convergent genes and convergent sites for the LCAW than for its equally distant control lineage; instead, these results are opposite to the patterns that we observed from the comparisons between the LCATW and its equally distant control (tables S3 to S6 and figs. S1 and S2), suggesting the independence of our results from the cutoffs of posterior probabilities. It should be noted that, although the convergent genes between branches I and IV that were identified from the PP0.7 and PP0.5 data sets are not significantly enriched in the functional categories related to hearing after multiple tests, these categories rank much higher than those involved in the convergent genes between branches I and V (tables S4 and S6).

These results indicate that there is significant convergence in hearing genes between the LCATW and echolocating bats but not between the LCAW and echolocating bats, suggesting that the high-frequency hearing or even ultrasonic hearing that is closely related to echolocation more likely occurred in the LCATW than in the LCAW.

Functional differences in *prestin* between echolocating and nonecholocating whales

Unsurprisingly, the first identified echolocation-related gene, *prestin* (*13*, *14*, *19*, *25*), is found to be on the list of hearing genes convergent between the LCATW and echolocating bats. Although the *prestin* of an echolocating toothed whale [bottlenose dolphin (*Tursiops truncatus*)] functions differently from that of a nonecholocating baleen whale [fin whale (*Balaenoptera physalus*)] (*19*), more species are required to determine whether there is a general functional disparity in prestin between echolocating and nonecholocating whales. Thus, we examined the function of *prestin* from two additional species from phylogenetically distant families (Fig. 3A), Blainville's beaked whale (*Mesoplodon densirostris*) and the pygmy sperm whale (*Kogia breviceps*). The nonlinear capacitance (NLC) of positive cells transfected by the *prestin* genes was measured using whole-cell patch-clamp recordings, which display a robust bell-shaped dependence on membrane potential (Fig. 3B) as observed in other mammals (*19*, *26*, *27*). After fitting the NLC responses with a two-state Boltzmann function, the related functional parameters $1/\alpha$, $V_{1/2}$, and Q_{max}/C_{lin} are obtained. In combination with previous data (*19*), we found that the $1/\alpha$ values of the echolocating whales (pygmy sperm whale, 54.49 ± 2.38 mV; Blainville's beaked whale, 53.19 ± 2.35 mV; bottlenose dolphin, 48.53 ± 1.99 mV) are consistently significantly larger than those of the nonecholocating fin whale (38.21 ± 1.04 mV) and their cow outgroup (34.05 ± 1.16 mV; $P < 0.001$, Student's *t* test; Fig. 3C). In contrast, the $V_{1/2}$ value of dolphin prestin is not significantly different from that of cow prestin but significantly smaller than those from other toothed whales and the fin whale ($P < 0.05$, Student's *t* test). The charge density Q_{max}/C_{lin} values from the pygmy sperm whale and Blainville's beaked whale are significantly larger than those from cow, fin whale, and dolphin ($P < 0.001$, Student's *t* test; Fig. 3C). In other words, the functional parameters $V_{1/2}$ and Q_{max}/C_{lin} show species-specific variation in whales, and only $1/\alpha$ can distinguish the echolocating and nonecholocating whales in the function of *prestin*.

To further explore the relationship between the functional parameter $1/\alpha$ of prestin and whale hearing, we collected data on the estimated frequency of the best hearing sensitivity from whales and other mammals (*28*). The plot of the estimated frequency of best hearing sensitivity versus the size of $1/\alpha$ showed a positive and statistically significant association ($R = 0.77$, $P = 0.015$, *F* test; Fig. 4). Notably, when we corrected for phylogeny using an independent contrast test (*29*), the correlation was consistently positive but no longer significant ($R = 0.34$, $P = 0.37$, *F* test), which is probably due to the small data set available for this analysis and/or more genes involved in the hearing capabilities of mammals.

Functional consistency of *prestin* between the LCAW and nonecholocating mammals

On the basis of the functional differences in *prestin* between echolocating and nonecholocating whales, we hypothesized that if the LCAW

Table 1. Top 10 of GO enrichment categories for convergent genes based on the PP0.95 data set.

GO ID	Description	*P* value	*q* value
Enrichment analysis for the convergent genes between branches I and IV			
GO:0007283	Spermatogenesis	2.79E-05	0.0119
GO:0048232	Male gamete generation	3.04E-05	0.0119
GO:0009913	Epidermal cell differentiation	4.43E-05	0.0119
GO:0060113	**Inner ear receptor cell differentiation**	4.50E-05	**0.0119**
GO:0060119	**Inner ear receptor cell development**	6.79E-05	**0.0147**
GO:0042490	Mechanoreceptor differentiation	1.24E-04	0.0241
GO:0051321	Meiotic cell cycle	1.83E-04	0.0338
GO:0030855	Epithelial cell differentiation	2.71E-04	0.0476
GO:0008544	Epidermis development	3.19E-04	0.0486
GO:0007605	**Sensory perception of sound**	3.30E-04	**0.0486**
Enrichment analysis for the convergent genes between branches I and V			
GO:0060271	Cilium morphogenesis	9.66E-05	0.0523
GO:0042384	Cilium assembly	2.19E-04	0.0833
GO:0044782	Cilium organization	2.22E-04	0.0833
GO:0003351	Epithelial cilium movement	3.93E-04	0.1131
GO:0045766	Positive regulation of angiogenesis	4.24E-04	0.1133
GO:0010927	Cellular component assembly involved in morphogenesis	5.30E-04	0.1322
GO:0042312	Regulation of vasodilation	9.00E-04	0.2105
GO:1904018	Positive regulation of vasculature development	1.25E-03	0.2531
GO:0035082	Axoneme assembly	1.28E-03	0.2531
GO:0007017	Microtubule-based process	1.49E-03	0.2539

had had similar high-frequency or ultrasonic hearing as the extant echolocating whales, then its prestin should most likely act as those of extant echolocating whales. To test this hypothesis, we first inferred the LCAW *prestin* sequence using the maximum likelihood method (*30*). Then, we synthesized the LCAW *prestin* and performed the functional experiments. The NLC of the LCAW prestin exhibits the same robust bell-shaped curve as those of living whales (Fig. 5A), and the yielded $1/\alpha$ value is 38.09 ± 1.73 mV, which is comparable to those from the extant nonecholocating mammals but significantly smaller than those from extant echolocating whales ($P < 0.001$, Student's *t* test; Fig. 5B). On the basis of the close relationship between $1/\alpha$ and high-frequency hearing, the empirical results for the LCAW *prestin* suggest that the LCAW might not have had the same high-frequency or ultrasonic hearing ability as the extant echolocating whales.

Functional convergence of *prestin* between the LCATW and echolocating whales

Next, we examined whether high-frequency or ultrasonic hearing evolved in the LCATW and whether its *prestin* acts as those of the extant echolocating whales. The LCATW *prestin* sequence was inferred and synthesized, and the NLC curve displays the same robust bell shape as those from other extant species (Fig. 5A). The value $1/\alpha = 54.16 \pm 2.47$ mV derived from the LCATW *prestin* is highly significantly larger than those from the LCAW and the nonecholocating mammals ($P < 0.001$, Student's *t* test; Fig. 5B). This value is comparable to those of the living echolocating whales and other echolocating mammals (*19*). As a result, the functional analyses of the ancestral *prestin* genes of the LCAW and the LCATW strongly support that high-frequency or ultrasonic hearing in living toothed whales most likely originated in the LCATW rather than in the LCAW.

Adaptive convergent evolution in the origin of high-frequency or ultrasonic hearing of whales

Because our results suggest that the large $1/\alpha$ value is closely related with high-frequency or ultrasonic hearing ability in whales, identifying the responsible amino acids in the LCATW *prestin* becomes a high priority for exploring how the characteristic evolved. By comparing prestin protein sequences between the LCATW and extant echolocating bats, we identified two convergent displacements of amino acids (S392A and L497M; fig. S3) that are significantly larger than expected at random

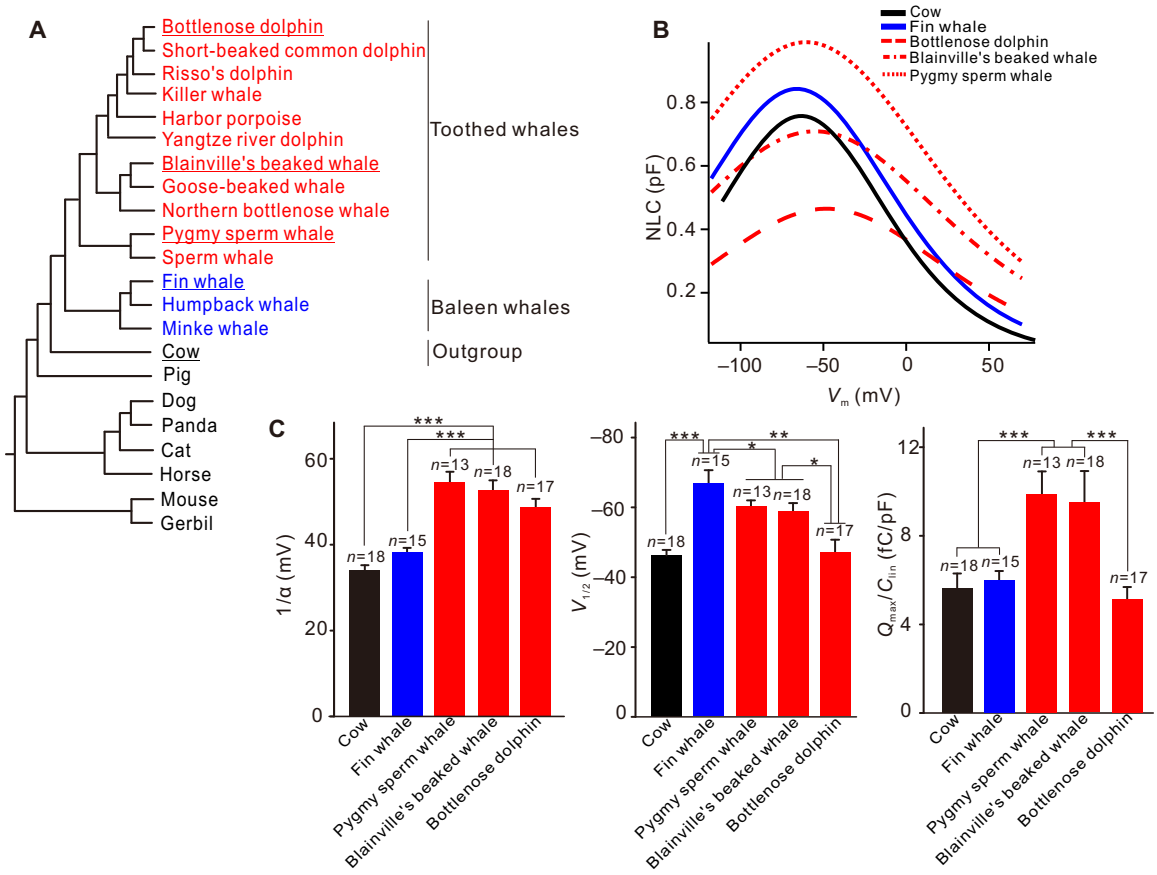

Fig. 3. Functional results of *prestin* in modern whales. (**A**) Phylogenetic relationships of whales with *prestin* sequences. Species names in red indicate echolocating toothed whales, and those in blue denote nonecholocating whales. Underlined names are representative species chosen for the functional examination of their *prestin* genes. (**B**) Representative fitting curves of nonlinear capacitance obtained from human embryonic kidney (HEK) 293 cells transfected by *prestin*. Different line types and colors indicate different species. (**C**) Comparison of three functional parameters, 1/α, $V_{1/2}$, and Q_{max}/C_{lin}, between echolocating and nonecholocating whales. All values are presented as means ± SE. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. All *P* values are from Student's *t* tests.

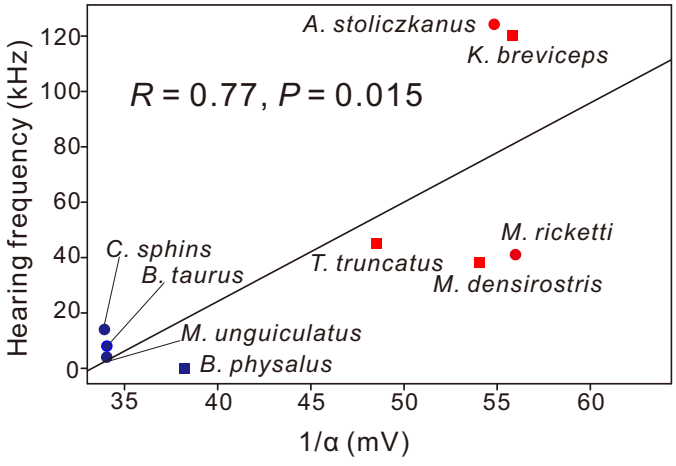

Fig. 4. Plot of 1/α versus the frequency of best hearing sensitivity showing significant relationships (*R* = 0.77, *P* = 0.015, *F* test). Squares represent whale species, and circles represent other mammals. The squares in red indicate echolocating mammals, and those in blue denote nonecholocating mammals.

under the JTT-f_{gene} neutral substitution model ($P = 7.28 \times 10^{-4}$) (*31*, *32*). However, the difference is not significant ($P = 0.0973$) when we used another neutral amino acid substitution model, JTT-f_{site}, for the test (*31*), suggesting weak selection, rather than chance, underlying the observed convergent substitutions. To determine the responsibility of the two convergent sites for the larger 1/α value of the LCATW prestin, we created a double mutant by changing Ala at position 392 to Ser and Met at position 497 to Leu on the genetic background of the LCATW *prestin*. The 1/α value of 42.03 ± 1.89 mV of this double mutant is significantly smaller than that of its wild type ($P < 0.01$, Student's *t* test; Fig. 5C). Further, we constructed another doubly mutated prestin (S392A/L497M) based on the background sequence of the LCAW *prestin* to test whether the 1/α value increases significantly. The resultant 1/α value of 48.71 ± 2.40 mV for this double mutant is highly significantly larger than that of the LCAW prestin ($P < 0.01$, Student's *t* test; Fig. 5C). Consequently, the mutational and rescue experiments show that two adaptive convergent changes in prestin between echolocating bats and the LCATW account for the functional convergence of the larger 1/α value in echolocating mammals and the LCATW.

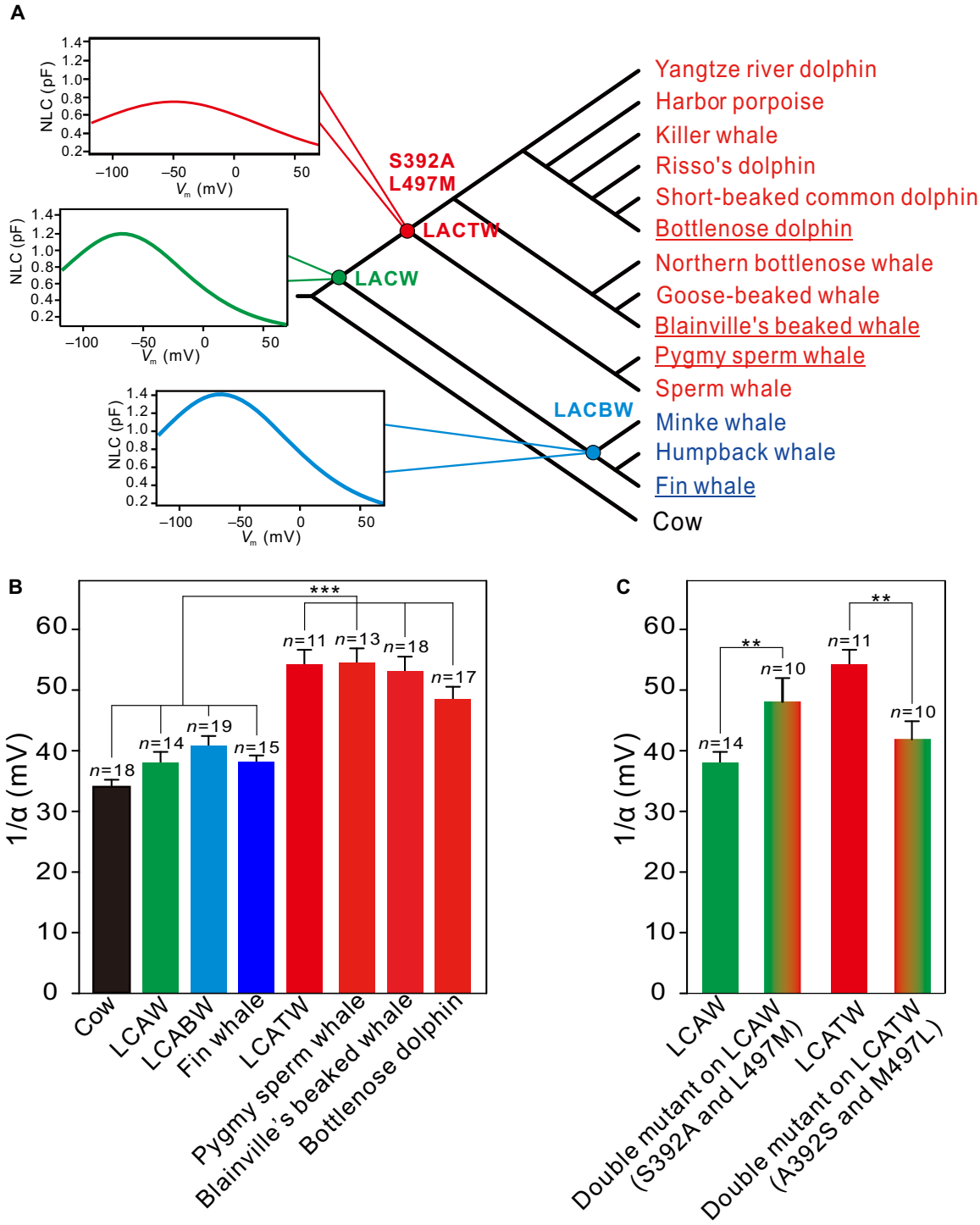

Fig. 5. Functional tests for resurrected ancestral *prestin* genes. (**A**) Schematic phylogenetic tree showing the nodes where ancestral *prestin* genes are examined. Representative fitting curves of NLC derived from ancestral prestins are shown. (**B**) Comparison of 1/α values of prestin among the ancestral and living whales as well as their outgroup. (**C**) Convergent sites (S392A and L497M) between the LCATW and echolocating bats account for enhancement of the functional parameter 1/α. Both convergent sites have mutations based on the *prestin* backgrounds of the LCAW and the LCATW, respectively. Values of 1/α significantly increase in the LCAW double mutant and decrease in the LCATW double mutant when compared to their respective wild-type controls. All values are given as means ± SE. $^{**}P < 0.01$ and $^{***}P < 0.001$. All *P* values are from Student's *t* tests.

DISCUSSION

Here, we investigated convergent evolution between the LCAW and echolocating bats on a genome-wide scale and found that (i) there is no significant enrichment in functional categories related to hearing for the convergent genes and (ii) stronger convergence is not observed in hearing-related proteins. The functional experiments show that the LCAW prestin has similar functions as those of nonecholocating mammals. These findings, in combination with the opposite results of the same genomic evolutionary analyses between the LCATW and echolocating bats as well as experimental results of the LCATW prestin, strongly suggest that high-frequency or ultrasonic hearing most likely evolved in the LCATW rather than in the LCAW. Newly discovered fossils of hearing organs have showed that the ancient protocetid whales and their terrestrial kin had similar hearing abilities (*33*), indicating consistency with our results of no high-frequency hearing in the LCAW. Nevertheless, our results cannot completely exclude the possibility that the LCAW with high but less ultrasonic-frequency or even relatively low-frequency hearing had the capability for rudimentary echolocation as the Old World fruit bats might do (*34*) because the molecular-level evolutionary and experimental analyses were derived from comparisons between the LCAW and echolocating bats that already had ultrasonic hearing and sophisticated echolocation.

Although several hearing-related genes exhibit various degrees of molecular convergence between toothed whales and echolocating bats (*15*, *17*, *18*), *prestin* is one of the most representative echolocation-related genes because (i) its molecular evolution is linked directly with the evolution of high-frequency or ultrasonic hearing in whales (*28*); (ii) the number of observed convergent amino acid replacements in echolocators is significantly larger than that expected from chance (*13*); (iii) functional experiments have demonstrated that *prestin* functionally converged among echolocating bats and toothed whales (*19*, *35*); and (iv) the evolution of *prestin* involves multiple adaptive changes early in the evolution of high-frequency hearing in vertebrates (*26*). These results strongly suggest a close relationship between high-frequency or ultrasonic hearing and *prestin*; thus, it is reasonable to use the functional changes in ancestral *prestin* genes to trace the occurrence of high-frequency or ultrasonic hearing in whales. Actually, it is not uncommon to resurrect ancestral genes to reveal the origin and evolution of related traits by examining their functional changes (*36*–*39*).

Regardless of the origination of high-frequency or ultrasonic hearing in the LCAW or the LCATW, the LCABW should have no high-frequency or ultrasonic hearing ability. Thus, the LCABW prestin is expected to yield the same value of the functional parameter $1/\alpha$ as those of nonecholocating mammals, which could further confirm the reliability of the results of our functional experiments for ancestral *prestin* genes. The obtained $1/\alpha$ value of 40.93 ± 1.58 mV of the LCABW prestin does not significantly differ from the values of the LCAW and living nonecholocating mammals, but it is significantly smaller than those of echolocating whales ($P < 0.01$, Student's t test; Fig. 5B).

The NLC is often used to evaluate the function of *prestin* by fitting a two-state Boltzmann function with three parameters: $1/\alpha$, $V_{1/2}$, and Q_{max}/C_{lin} (*40*, *41*). Why do the *prestin* genes of echolocating mammals consistently generate much larger $1/\alpha$ values than those of nonecholocating mammals (*20*)? Because the functional parameter $1/\alpha$ represents the reciprocal of the fraction of an elementary charge moving across the cell membrane (*42*), a large value of $1/\alpha$ suggests that prestins of the echolocating mammals transfer less charge across the cell membrane than those of the nonecholocating mammals. However, little is known about the biological and physiological implications of this difference for mammalian high-frequency or ultrasonic hearing. The value of $1/\alpha$ in the outer hair cells of the more basal cochlea, which is sensitive to high-frequency sound, is larger than that in the cells that are sensitive to low-frequency sound (*43*). Nevertheless, more experimental research on *prestin* or other echolocation-related genes is necessary to reveal the molecular mechanisms underlying mammalian high-frequency or ultrasonic hearing and echolocation.

MATERIALS AND METHODS

Genomic data

The coding gene sequences (CDSs) of the 16 species used in this study, including *Homo sapiens*, *Mus musculus*, *Canis familiaris*, *Bos taurus*, *Ovis aries*, *Capra hircus*, *Pantholops hodgsonii*, *T. truncatus*, *Orcinus orca*, *Lipotes vexillifer*, *Balaenoptera acutorostrata*, *Pteropus alecto*, *Pteropus vampyrus*, *Myotis brandtii*, *Myotis davidii*, and *Myotis lucifugus* were downloaded from the National Center for Biotechnology Information database (www.ncbi.nlm.nih.gov/nuccore). After selecting the longest CDS for each gene of each species, we used Inparanoid software (version 4.1) (*44*) with default parameters to identify one-to-one orthologs among these species and finally obtained 7206 one-to-one orthologous genes.

Identification of convergent and divergent sites

The coding sequences for each orthologous gene in our data set was aligned using PRANK (*45*) and translated into amino acid sequences. On the basis of the species tree (Fig. 1A), we inferred ancestral sequences for each gene using maximum likelihood and empirical Bayesian approaches (*30*), counting the numbers of convergent and divergent substitutions along the branch pairs in which we were interested (Fig. 1A). To exclude the effects of sequencing errors, incorrect alignments, and non-orthologous regions in the alignments on the identified convergent and divergent sites, we deleted a convergent/divergent site if its flanking sequences ±10 amino acids met one of the following criteria: (i) mean sequence similarity < 0.7; (ii) lowest similarity < 0.35 between any two sequences; and (iii) >5 successive indels in more than two species. We defined the convergent substitutions at a site as those inferred substitutions that resulted in the same amino acid along the branch pairs examined for convergence, thus including both convergent and parallel substitutions. If a gene contained at least one convergent substitution, then it was defined as a convergent gene. To ensure the robustness of our analyses, the inferred sites were respectively included into our analyses with three cutoffs of posterior probabilities: ≥0.95, ≥0.7, and ≥0.5.

Gene ontology enrichment analysis

We used the enrichGO model in the clusterProfiler (version 2.4.7) (*46*) to analyze gene ontology (GO) enrichment for the genes with convergence. Fisher's exact test was used to calculate a P value for overrepresented GO categories when comparing the target and control lineages. To decrease the possibility of false positive results, we performed multiple testing using the false discovery rate correction.

Set of genes associated with hearing

We searched our orthologous data set for loci involved in hearing, including those linked to deafness and/or ear development. All GO annotations for each gene in our data set were obtained by searching the DAVID 6.7 database (http://david.abcc.ncifcrf.gov) (*47*), and we identified a total of 104 hearing-related genes with the annotations containing the words "auditory," "sound," or "ear" (table S2). After identifying convergent and divergent substitutions, we compared the ratios of the

number of convergent substitutions to the number of divergent substitutions between branches I and II and branches I and III as well as between branches I and IV and branches I and V, respectively. We also used the bootstrap approach to test whether the *C*/*D* value of hearing genes is significantly different from that of nonhearing genes for the same number as hearing genes for each branch pair examined for convergence.

Reconstruction of prestin sequences for ancestral whales

To more reliably infer the ancestral *prestin* sequences in whales, we added another 10 whale species with intact coding sequences of *prestin* in GenBank (www.ncbi.nlm.nih.gov/genbank/). The final data set included 11 echolocating toothed whales and 3 nonecholocating baleen whales, covering the main lineages of living whales (Fig. 3A and table S7). After alignment, we inferred the *prestin* sequences of ancestral whales using the same approaches described above. A statistical test and its improved version (*31*, *32*) were used to calculate the probability that the number of observed convergent sites exceeded that expected by chance.

Sequence synthesis of *prestin*, site-directed mutations, and transient transfection

Bottlenose dolphin, Blainville's beaked whale, pygmy sperm whale, and fin whale were chosen as representative living species for the functional examination of *prestin*; the outgroup was cow. The entire coding regions of *prestin* from these species and the ancestral *prestin* sequences inferred above were synthesized (Generay) and cloned into the expression vector pEGFP-N1 (Clontech). This yielded C-terminal green fluorescent protein (GFP) fusion constructs that were used to confirm whether *prestin* expresses and locates in the cell membrane. HEK 293 cells were cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum, and cells were plated for 24 hours before transient transfection. Then, 4-μg prestin-GFP plasmids were transfected to HEK 293 cells using 10 μl of Lipofectamine 2000 (Invitrogen) in 500-μl Opti-MEM (Life Technologies). After 24 to 48 hours of incubation, successfully transfected cells were used for NLC measurements. We used polymerase chain reaction (PCR) technology to produce the mutants. Prestin-GFP plasmid DNA served as the templates, and the PCR primers designed to create the mutant constructs were as follows: S392A_F, CTCACTCTTCCAGACTTTTGCAATTTCATGCTCCTTGTC and S392A_R, GACAAGGAGCATGAAATTGCAAAAGTCTGGAAGAGTGAG; L497M_F, TTCTGTAAATCACAGTCATCAGTGCGATGATCACAGC and L497M_R, GCTGTGATCATCGCACTGATGACTGTGATTTACAGAA; A392S_F, CTCACTCTTCCAGACTTTTTCAATTTCATGCTCCTTGTC and A392S_R, GACAAGGAGCATGAAATTGAAAAAGTCTGGAAGAGTGAG; and M497L_F, TTCTGTAAATCACAGTCAGCAGTGCGATGATCACAGC and M497L_R, GCTGTGATCATCGCACTGCTGACTGTGATTTACAGAA. All mutational constructs were sequenced to confirm the nucleotide exchanges and sequence correction.

Electrophysiological experiments for NLC measurements

We used the HEKA EPC 10 USB amplifier (HEKA Instruments Inc.) controlled by Patchmaster software (HEKA Instruments Inc.) to measure the NLC of prestin by whole-cell patch-clamp recordings at room temperature (22° to 26°C). Recording pipettes made of borosilicate glass were pulled with resistances of 2.5 to 4 megohms and filled with an internal solution containing 140 mM CsCl, 2 mM $MgCl_2$, 10 mM EGTA, and 10 mM Hepes. During the recordings, cells were bathed in an external solution containing 120 mM NaCl, 20 mM TEA-Cl, 2 mM CoCl, 2 mM $MgCl_2$, 10 mM Hepes, and 5 mM glucose. Solutions were adjusted to pH 7.2. The osmolarities of the internal and external solutions were adjusted with glucose to 300 and 320 mOsm/liter, respectively. We measured whole-cell membrane capacitance (C_m) using sine + DC software, a lock-in function of Patchmaster. Voltage-dependent NLC was assessed by recording C_m during voltage ramps, as previously described (*48*, *49*). NLC curves were quantified by fitting with the derivative of a two-state Boltzmann function

$$C_m = \frac{Q_{max}\alpha}{\exp[\alpha(V_m - V_{1/2})](1+\exp[-\alpha(V_m - V_{1/2})])^2} + C_{lin}$$

where Q_{max} is the maximum charge transfer, $V_{1/2}$ is the voltage at which the maximum charge is equally distributed across the membrane, C_{lin} is the linear capacitance, and α is the slope factor of the voltage dependence of the charge transfer. C_{lin} is proportional to the surface area of the membrane (cell size). To compare the magnitude of NLC obtained from different cells with different levels of prestin expression as a function of cell size, we normalized the NLC by the linear capacitance of the cells. Because differences in Q_{max} could have been caused by cell size, the charge movement was normalized to C_{lin}. This quantity, designated as the charge density (Q_{max}/C_{lin}), had units of femtocoulomb per picofarad. We used IgoPro software (IgoPro, WaveMetrics) for data processing and fitting.

Correlation analysis

To test for correlation between the frequency of best hearing sensitivity and the functional parameter $1/\alpha$ of prestin, we combined related data from previous studies (*19*, *28*) with those of our study. The frequencies of best hearing were from published audiograms, and if no audiogram was available, then the values were inferred on the basis of published call frequencies with maximum energy for the species (*28*). The correlation between the NLC parameters and mammalian hearing frequencies, as well as statistical tests, were carried out with program *R*. To exclude the influence of phylogeny on the correlation analysis, we conducted an independent contrasts test (*29*) based on a species tree with estimated divergence times (*28*, *50*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/10/eaat8821/DC1

Fig. S1. Detection of convergence in hearing-related genes based on the PP0.7 data set.
Fig. S2. Detection of convergence in hearing-related genes based on the PP0.5 data set.
Fig. S3. The phylogenetic tree that we used to test whether the number of the observed convergent sites exceeds the neutral expectations.
Table S1. Top 10 of GO enrichment categories for convergent genes between branches I and II and between branches I and III based on the PP0.95 data set.
Table S2. Hearing-related genes based on the annotations in the DAVID database.
Table S3. Top 15 of GO enrichment categories for convergent genes between branches I and II and between branches I and III based on the PP0.7 data set.
Table S4. Top 15 of GO enrichment categories for convergent genes between branches I and IV and between branches I and V based on the PP0.7 data set.
Table S5. Top 15 of GO enrichment categories for convergent genes between branches I and II and between branches I and III based on the PP0.5 data set.
Table S6. Top 15 of GO enrichment categories for convergent genes between branches I and IV and between branches I and V based on the PP0.5 data set.
Table S7. Whale species used in this study.

REFERENCES AND NOTES

1. M. R. McGowen, M. Spaulding, J. Gatesy, Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol. Phylogenet. Evol.* **53**, 891–906 (2009).

2. S. L. Messenger, J. A. McGuire, Morphology, molecules, and the phylogenetics of cetaceans. *Syst. Biol.* **47**, 90–124 (1998).
3. M. Nikaido, O. Piskurek, N. Okada, Toothed whale monophyly reassessed by SINE insertion analysis: The absence of lineage sorting effects suggests a small population of a common ancestral species. *Mol. Phylogenet. Evol.* **43**, 216–224 (2007).
4. G. Jones, Echolocation. *Curr. Biol.* **15**, R484–R488 (2005).
5. M. Churchill, M. Martinez-Caceres, C. de Muizon, J. Mnieckowski, J. H. Geisler, The origin of high-frequency hearing in whales. *Curr. Biol.* **26**, 2144–2149 (2016).
6. T. Park, E. M. G. Fitzgerald, A. R. Evans, Ultrasonic hearing and echolocation in the earliest toothed whales. *Biol. Lett.* **12**, 20160060 (2016).
7. J. H. Geisler, M. W. Colbert, J. L. Carew, A new fossil species supports an early origin for toothed whale echolocation. *Nature* **508**, 383–386 (2014).
8. G. Fleischer, Hearing in extinct cetaceans as determined by cochlear structure. *J. Paleontol.* **50**, 133–152 (1976).
9. D. R. Ketten, The marine mammal ear: Specializations for aquatic audition and echolocation, in *Evolutionary Biology of Hearing*, D. B. Webster, R. R. Fay, A. N. Popper, Eds. (Springer-Verlag, 1992), pp. 717–750.
10. D. R. Ketten, The cetacean ear; Form, frequency, and evolution, in *Marine Mammal Sensory Systems*, J. A. Thomas, R. A. Kastelein, A. Y. Supin, Eds. (Plenum Press, 1992), pp. 53–75.
11. E. G. Ekdale, R. A. Racicot, Anatomical evidence for low frequency sensitivity in an archaeocete whale: Comparison of the inner ear of *Zygorhiza kochii* with that of crown Mysticeti. *J. Anat.* **226**, 22–39 (2015).
12. M. D. Uhen, Form, function, and anatomy of *Dorudon atrox* (Mammalia, Cetacea): An archaeocete from the middle to late Eocene of Egypt. *Univ. Mich. Pap. Paleontol.* **34**, 1–222 (2004).
13. Y. Li, Z. Liu, P. Shi, J. Zhang, The hearing gene Prestin unites echolocating bats and whales. *Curr. Biol.* **20**, R55–R56 (2010).
14. Y. Liu, J. A. Cotton, B. Shen, X. Han, S. J. Rossiter, S. Zhang, Convergent sequence evolution between echolocating bats and dolphins. *Curr. Biol.* **20**, R53–R54 (2010).
15. K. T. J. Davies, J. A. Cotton, J. D. Kirwan, E. C. Teeling, S. J. Rossiter, Parallel signatures of sequence evolution among hearing genes in echolocating mammals: An emerging model of genetic convergence. *Heredity* **108**, 480–489 (2011).
16. J. Parker, G. Tsagkogeorga, J. A. Cotton, Y. Liu, P. Provero, E. Stupka, S. J. Rossiter, Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231 (2013).
17. Y.-Y. Shen, L. Liang, G.-S. Li, R. W. Murphy, Y.-P. Zhang, Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLOS Genet.* **8**, e1002788 (2012).
18. Z. Liu, S. Li, W. Wang, D. Xu, R. W. Murphy, P. Shi, Parallel evolution of *KCNQ4* in echolocating bats. *PLOS ONE* **6**, e26618 (2011).
19. Z. Liu, F.-Y. Qi, X. Zhou, H.-Q. Ren, P. Shi, Parallel sites implicate functional convergence of the hearing gene *prestin* among echolocating mammals. *Mol. Biol. Evol.* **31**, 2415–2424 (2014).
20. Y. Liu, N. Han, L. F. Franchini, H. Xu, F. Pisciottano, A. B. Elgoyhen, K. E. Rajan, S. Zhang, The voltage-gated potassium channel subfamily KQT member 4 (KCNQ4) displays parallel evolution in echolocating bats. *Mol. Biol. Evol.* **29**, 1441–1450 (2012).
21. T. A. Castoe, A. P. J. de Koning, H.-M. Kim, W. Gu, B. P. Noonan, G. Naylor, Z. J. Jiang, C. L. Parkinson, D. D. Pollock, Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8986–8991 (2009).
22. Z. Zou, J. Zhang, No genome-wide protein sequence convergence for echolocation. *Mol. Biol. Evol.* **32**, 1237–1241 (2015).
23. G. W. C. Thomas, M. W. Hahn, Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. *Mol. Biol. Evol.* **32**, 1232–1236 (2015).
24. A. D. Foote, Y. Liu, G. W. C. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancia, R. Nielsen, X. Qin, J. Qu, B. J. Raney, N. Vijay, J. B. W. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. Gilbert, R. A. Gibbs, Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).
25. G. Jones, Molecular evolution: Gene convergence in echolocating mammals. *Curr. Biol.* **20**, R62–R64 (2010).
26. Z. Liu, G.-H. Li, J.-F. Huang, R. W. Murphy, P. Shi, Hearing aid for vertebrates via multiple episodic adaptive events on prestin genes. *Mol. Biol. Evol.* **29**, 2187–2198 (2012).
27. X. Tan, J. L. Pecka, J. Tang, O. E. Okoruwa, Q. Zhang, K. W. Beisel, D. Z. He, From zebrafish to mammal: Functional evolution of prestin, the motor protein of cochlear outer hair cells. *J. Neurophysiol.* **105**, 36–44 (2011).
28. Y. Liu, S. J. Rossiter, X. Han, J. A. Cotton, S. Zhang, Cetaceans on a molecular fast track to ultrasonic hearing. *Curr. Biol.* **20**, 1834–1839 (2010).
29. P. E. Midford, T. Garland Jr., W. P. Maddison, PDAP package of Mesquite, version 1.07 (2005).
30. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
31. Z. Zou, J. Zhang, Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* **32**, 2085–2096 (2015).
32. J. Zhang, S. Kumar, Detection of convergent and parallel evolution at the amino acid sequence level. *Mol. Biol. Evol.* **14**, 527–536 (1997).
33. M. J. Mourlam, M. J. Orliac, Infrasonic and ultrasonic hearing evolved after the emergence of modern whales. *Curr. Biol.* **27**, 1776–1781.e9 (2017).
34. A. Boonman, S. Bumrungsri, Y. Yovel, Nonecholocating fruit bats produce biosonar clicks with their wings. *Curr. Biol.* **24**, 2962–2967 (2014).
35. Y.-Y. Li, Z. Liu, F.-Y. Qi, X. Zhou, P. Shi, Functional effects of a retained ancestral polymorphism in *prestin*. *Mol. Biol. Evol.* **34**, 88–92 (2017).
36. J. K. Hobbs, C. Shepherd, D. J. Saul, N. J. Demetras, S. Haaning, C. R. Monk, R. M. Daniel, V. L. Arcus, On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. *Mol. Biol. Evol.* **29**, 825–835 (2012).
37. J. M. Thomson, E. A. Gaucher, M. F. Burgan, D. W. De Kee, T. Li, J. P. Aris, S. A. Benner, Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**, 630–635 (2005).
38. S. Akanuma, Y. Nakajima, S. Yokobori, M. Kimura, N. Nemoto, T. Mase, K.-i. Miyazono, M. Tanokura, A. Yamagishi, Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11067–11072 (2013).
39. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
40. J. Santos-Sacchi, Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J. Neurosci.* **11**, 3096–3110 (1991).
41. P. Dallos, B. Fakler, Prestin, a new type of motor protein. *Nat. Rev. Mol. Cell Biol.* **3**, 104–111 (2002).
42. J. Ashmore, Cochlear outer hair cell motility. *Physiol. Rev.* **88**, 173–210 (2008).
43. J. Santos-Sacchi, S. Kakehata, T. Kikuchi, Y. Katori, T. Takasaka, Density of motility-related charge in the outer hair cell of the guinea pig is inversely related to best frequency. *Neurosci. Lett.* **256**, 155–158 (1998).
44. K. P. O'Brien, M. Remm, E. L. L. Sonnhammer, Inparanoid: A comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* **33**, D476–D480 (2005).
45. A. Löytynoja, N. Goldman, An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10557–10562 (2005).
46. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
47. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
48. T. J. Schaechinger, D. Gorbunov, C. R. Halaszovich, T. Moser, S. Kügler, B. Fakler, D. Oliver, A synthetic prestin reveals protein domains and molecular operation of outer hair cell piezoelectricity. *EMBO J.* **30**, 2793–2804 (2011).
49. D. Oliver, B. Fakler, Expression density and functional characteristics of the outer hair cell motor protein are regulated during postnatal development in rat. *J. Physiol.* **519** (Pt 3), 791–800 (1999).
50. E. C. Teeling, M. S. Springer, O. Madsen, P. Bates, S. J. O'Brien, W. J. Murphy, A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584 (2005).

Acknowledgments: We thank H. Yang for providing valuable comments. **Funding:** The project was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB13020400), Yunnan Province, the High-End Scientific and Technological Talents program (2013HA020), and the National Natural Science Foundation of China (31321002, 31325013, and 31301013). **Author contributions:** Z.L. and P.S. designed the study. F.-Y.Q. and X.Z. performed the experiments. Z.L. and F.-Y.Q. analyzed the experimental data. D.-M.X. analyzed the genomic data. Z.L., P.S., F.-Y.Q., and D.-M.X. wrote the paper. **Competing interests:** The authors declare that they no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 15 April 2018
Accepted 27 August 2018
Published 3 October 2018
10.1126/sciadv.aat8821

Citation: Z. Liu, F.-Y. Qi, D.-M. Xu, X. Zhou, P. Shi, Genomic and functional evidence reveals molecular insights into the origin of echolocation in whales. *Sci. Adv.* **4**, eaat8821 (2018).

Genomic and functional evidence reveals molecular insights into the origin of echolocation in whales

Zhen Liu, Fei-Yan Qi, Dong-Ming Xu, Xin Zhou and Peng Shi

Sci Adv **4** (10), eaat8821.
DOI: 10.1126/sciadv.aat8821