

Evidence of Echolocation in the Common Shrew from Molecular Convergence with Other Echolocating Mammals

Simin Chai¹, Ran Tian¹, Xinghua Rong¹, Guiting Li¹, Bingyao Chen¹, Wenhua Ren¹, Shixia Xu^{1,*}, and Guang Yang^{1,*}

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China.

*Correspondence: E-mail: xushixia78@163.com (Xu); gyang@njnu.edu.cn (Yang)

E-mail: chaisiminendeavor@163.com (Chai); tianrannjnu@163.com (Tian); 821393856@qq.com (Rong); 936877504@qq.com (Li); chby2632@163.com (Chen); 08162@njnu.edu.cn (Ren)

Received 5 November 2019 / Accepted 22 January 2020 / Published 25 February 2020
Communicated by Benny K.K. Chan

Along with sophisticated echolocation found in bats and toothed whales, the common shrew (*Sorex araneus*) was confirmed to possess echolocation ability based on behavioral and experimental evidence such as high-frequency twittering and close-range spatial orientation. However, whether echolocation in the common shrew is convergent with bats and dolphins at the molecular level remains poorly understood. In this study, we gathered the coding region sequences of 11 hearing-related genes from genome data and previous studies. Convergent evolutionary analyses identified 13 amino acid residues (seven in CDH23, five in OTOF, and one in PRESTIN) under strong convergent evolution shared among the common shrew and other echolocating mammals (bats and dolphins). Furthermore, a phylogenetic tree was constructed based on the combined amino acid dataset of convergent/parallel substitutions, sites with parallel radical property changes, and sites supporting echolocator-convergence; it supported the converged topology of the simple echolocator *Sorex araneus* and sophisticated echolocating bats with high posterior probability. This study gives evidence at the molecular level that the common shrew echolocate and provides novel insights into the convergent evolution between the common shrew and bats and dolphins.

Key words: Convergent evolution, Echolocation, *Sorex araneus*, Radical property change, Amino acid substitution.

BACKGROUND

Adaptive convergent evolution, an important concept in Darwin's theory of evolution, occurs when a similar trait evolves independently in more than one lineage from distantly diverged ancestors. In nature, convergent evolution commonly arises in the evolutionary trajectories of different species, includes the organic osmotic solute systems in bacteria, plants, and animals facing the challenges of water stress (Yancey et al. 1982); the ability to fly in birds and bats (Norberg and Rayner 1987); the reduced pigmentation of skin

melanin among Europeans and Asians in adaptation to high-latitude environments (Wang et al. 2014); and the shared phenotypic traits of several unrelated marine mammals adapted to aquatic environments (Foote et al. 2015; Zhou et al. 2015). It has been suggested that convergent morphological evolution is attributed to convergent substitution at the amino acid layer (Zakon 2002; Wierer et al. 2012). Convergent amino acid substitutions refer to changes from different ancestral states to the same descendent amino acid within a pair of branches, whereas parallel substitutions are defined as independent changes from the same ancestral amino

acid. Furthermore, multiple approaches to detect natural selection acting on genes and proteins have been put forward. Amino acid replacement diagnosis via testing radical physicochemical changes has been suggested as a sensitive method for detecting selection in protein-coding genes (McClellan and McCracken 2001; McClellan and Ellison 2010).

Echolocation is a biological ability in animals to emit sonar pulses via clicking sounds and then utilize the returning echoes for spatial orientation, food acquisition, and predator avoidance. It is widely accepted that bats (Chiroptera, except for Pteropodidae) and toothed whales (Odontoceti) independently acquired the ability to echolocate over their evolutionary histories (Schnitzler et al. 2003; Eick et al. 2005; Steeman et al. 2009). In earlier studies, the peak frequencies of bats' echolocation signals were found to range from 8 to > 200 kHz, with an effective working range of ≥ 20 meters (Fenton and Bell 1981; Fenton et al. 2014). As for bats' echolocating counterparts underwater, toothed whales have the highest frequency of narrow-band clicks on record (130–135 kHz) (Fenton et al. 2014; Kyhn et al. 2010).

In recent studies, some hearing-related proteins have been identified as potentially essential to echolocation, such as the motor protein PRESTIN, which is expressed in the cochlear outer hair cells (OHCs) and charges voltage motility; homodimers of protein CDH23 and PCDH15, involved in hair bundle motility; transmembrane protein TMC1, expressed in the inner hair cells (IHCs) and OHCs, which regulates hair cell maturation. These proteins have been reported to be convergent in bats and toothed whales (Li et al. 2008; Liu et al. 2010 2011; Davies et al. 2012; Ru et al. 2012; Shen et al. 2012; Kirwan et al. 2013). Similar to other convergent evolution phenomena, convergence of echolocation has been suggested to be the result of convergence and/or parallelism in protein-coding sequences that alter hearing-related genes to some extent (Zakon 2002; Parker et al. 2013; Stern 2013). Evidence that echolocation is convergent can be found in the topologies in which echolocating bats and toothed whales were grouped together based on the amino acid sequences of PRESTIN. However, these topologies of convergence were not supported after excluding convergent substitution amino acid sites (Li et al. 2008; Liu et al. 2010).

It is noteworthy that, in addition to the aforementioned well-known echolocating animals, a small insectivorous mammal, the common shrew (*Sorex araneus*), very likely possesses the ability to echolocate (Buchler 1976; Forsman and Malmquist 1988; Siemers et al. 2009). The frequency of ultrasonic calls from the common shrew was recorded at 50 kHz (Thomas

and Jalili 2004). The low-amplitude, broadband, and frequency-modulated high-pitched laryngeal twittering ultrasonic calls from the common shrew are proposed to be used for simple and close-range spatial orientation, such as to detect habitat type and locate protective cover to minimize the risk of predation (Forsman and Malmquist 1988; Siemers et al. 2009).

With the exception of behavioral experiments, ours is the first molecular study address the ultrasonic call emitting capability or the underlying mechanism of echolocation in the common shrew. We collected 11 hearing-related genes from different mammalian lineages under an enlarged toothed whale dataset. We present this evolutionary study with the goal of investigating whether the common shrew, as an ultrasonic call emitter and potential echolocator, has experienced similar adaptive evolution at the molecular level. To our knowledge, this is the first study to provide a novel insight into the molecular advancement of unexplored echolocation behavior in the common shrew. Further, it could expand our understanding of the evolution of complex diverse mechanisms involved in mammal echolocation.

MATERIALS AND METHODS

Data collection and sequence alignment

The coding sequences of mammalian hearing-related genes from previous studies (Liu et al. 2010 2011; Davies et al. 2012; Ru et al. 2012; Shen et al. 2012; Kirwan et al. 2013) were obtained according to accession numbers. Sequences of the common shrew (*Sorex araneus*) and hedgehog (*Erinaceus europaeus*) were downloaded from the Ensembl Genome Database (<http://www.ensembl.org>) and GenBank (<http://www.ncbi.nlm.nih.gov>). Sequences of the cetaceans used in this study—killer whale (*Orcinus orca*), Yangtze finless porpoise (*Neophocaena asiaeorientalis*), Yangtze river dolphin (*Lipotes vexillifer*), sperm whale (*Physeter microcephalus*), bowhead whale (*Balaena mysticetus*), minke whale (*Balaenoptera acutorostrata*)—were retrieved by BLAST using the cow (*Bos taurus*) sequence as the query, from the available genome data, Ensembl Genome Database and GenBank, or The Bowhead Whale Genome Resource (<http://www.bowhead-whale.org/>) (Keane et al. 2015). All taxonomy and sequence information are in table S1. Sequence alignments were done using the MUSCLE algorithm (Edgar 2004) in MEGA5 (Tamura et al. 2011). DAMBE (Xia 2013) was employed to examine the degree of nucleotide substitution saturation.

Convergent and parallel amino acid substitutions

Convergent and parallel amino acid substitutions between the common shrew and branches of sophisticated echolocators (echolocating bats and toothed whales) were conducted separately through pairwise comparison with their most recent common ancestors. Sequences of all ancestral nodes were inferred by the maximum likelihood method in PAML 4.7 (Yang 2007). To maximize the effectiveness of the control group and take phylogenetic relationships into consideration, we chose the hedgehog (*Erinaceus europaeus*), cow (*Bos taurus*), and Pteropodidae bats as representative non-echolocators, *i.e.*, the control group. We employed the same analytical method to investigate the convergent and parallel amino acid substitutions between representative non-echolocating branches and two groups of advanced echolocating mammals.

The true positive convergent and parallel substitutions denote substitutions that were only shared between the common shrew and advanced echolocating lineages, and were missing in the control group.

Site-wise log-likelihood support

Site-wise differences were measured for different topology hypotheses—*i.e.*, hypotheses of the well-accepted species tree and constrained echolocator-converged trees—using RAxML version 8 (Stamatakis 2014) (Fig. 1). In this study, we computed respective site-wise log-likelihood support (SSLS) for a commonly accepted species tree (referred as H_0) and three alternative constrained echolocator-converged topologies H_A (H_1 corresponds to all convergences in echolocators, H_2 to the shrew convergences with the toothed whales, and H_3 to the convergence between the common shrew and echolocating bats). Site-wise

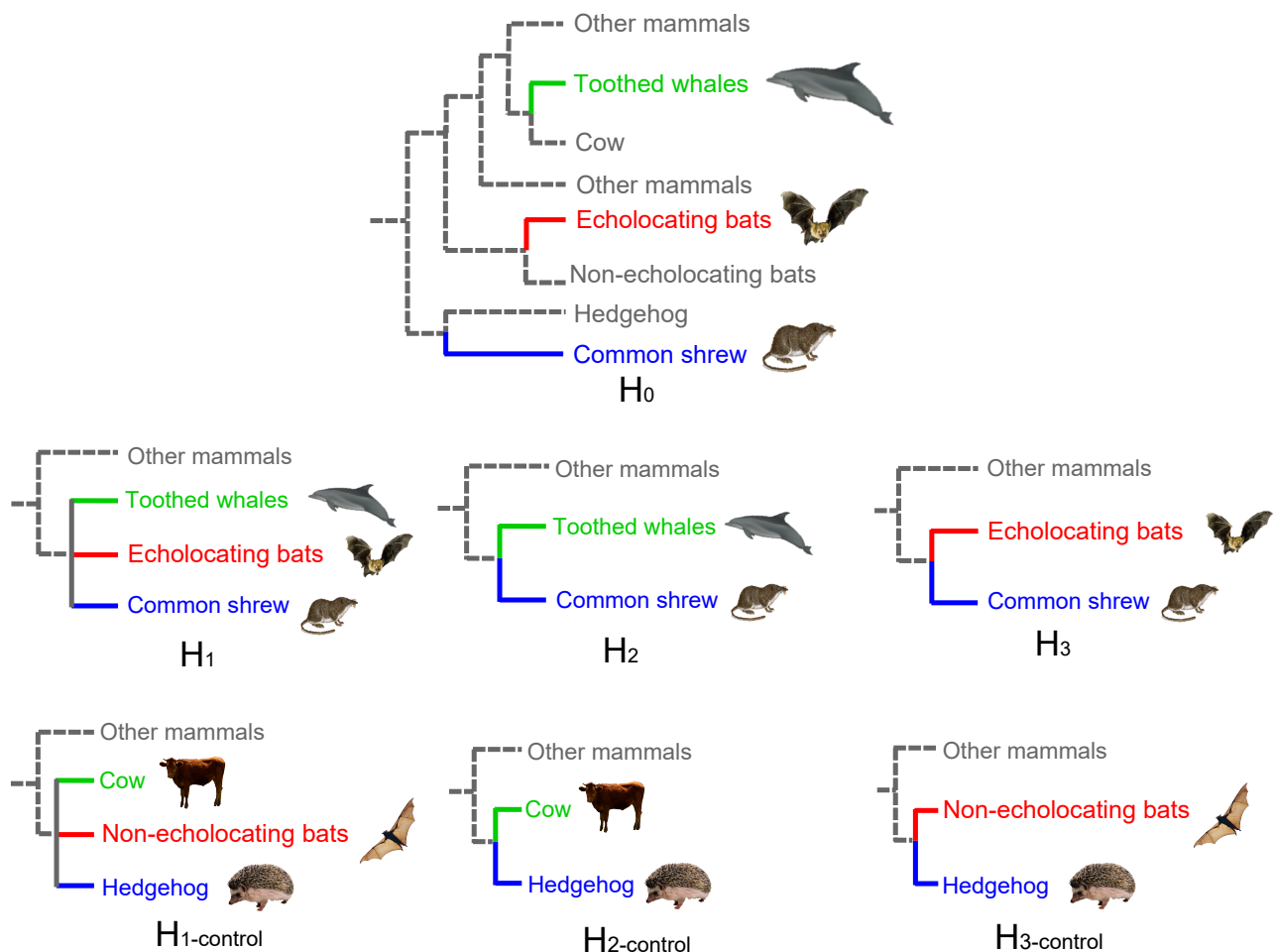


Fig. 1. Species trees and alternative constraint convergence topologies. H_0 is the well-accepted species tree. H_1 , H_2 , and H_3 are three alternate echolocator-converged topologies. $H_{1\text{-control}}$, $H_{2\text{-control}}$ and $H_{3\text{-control}}$ refer to the constraint convergent topologies of representative non-echolocators (cow, hedgehog, and non-echolocating bats).

differences between each pair of null and alternative hypotheses were calculated using:

$$\Delta SLS_i = SLS_{i, H_A} - SLS_{i, H_0}$$

where positive ΔSLS for the i^{th} amino acid in a locus indicates better fitness for the convergence in echolocators hypotheses. Likewise, there were three alternative hypotheses for non-echolocators in the control group, collectively referred to as $H_{A\text{-control}}$. $H_{1\text{-control}}$ corresponds to hedgehog, cow, and non-echolocating bat convergence. $H_{2\text{-control}}$ is hedgehog and cow convergence, and $H_{3\text{-control}}$ is convergence of hedgehog and non-echolocating bat. Positive $\Delta SLS_{i\text{-control}}$ indicates more support for the convergence hypotheses in the control group, and negative $\Delta SLS_{i\text{-control}}$ indicates no support.

$$\Delta SLS_{i\text{-control}} = SLS_{i, H_{A\text{-control}}} - SLS_{i, H_0}$$

Amino acid sites with positive ΔSLS_i and negative $\Delta SLS_{i\text{-control}}$ values are regarded as true positive sites supported the echolocators clustering topologies.

Analysis of radical property changes in amino acid residues

Amino acid residues that might undergo adaptive changes were identified from radical changes in physiochemical properties (McClellan and McCracken 2001). The 31 distinguished structural and physicochemical properties of amino acid residues were analyzed in the software TreeSAAP under the goodness-of-fit test (Woolley et al. 2003). All 31 structural and physiochemical properties were graded into eight magnitude categories, and amino acid substitutions with changes in magnitude from six to eight with P values ≤ 0.05 were generally considered radical property changing residues (McClellan et al. 2005). Amino acid residues with radical changes between each of the two echolocating branches were counted.

Phylogenetic reconstruction

Phylogenetic reconstruction was done using the maximum likelihood (ML) method in RAxML version 8 and Bayesian inference (BI) in MrBayes version 3.2 (Ronquist et al. 2012) based on entire nucleotide sequences and amino acid sequences of each gene separately. We chose the best fitting model for each coding region sequence in accordance with AIC criterion by ModelTest version 3.7 (Posada and Crandall 1998) for nucleotide evolution and ProtTest version 3 (Darriba et al. 2011) for amino acid sequence

evolution. Trees were built based on the third codon position and the first plus second codon positions, separately, by partitioning each codon. After that, ML and BI trees were constructed independently based on nonsynonymous amino acid substitutions and synonymous substitutions to reveal the variation in evolutionary signals. Multilocus concatenated tree reconstruction was conducted in species with at least five genes. ML bootstrap support was obtained from a rapid bootstrap analysis, and the best-scoring ML tree was identified from 1000 replicates under the best-fitting evolutionary model for each dataset. As for the BI method, the best tree of each gene was derived from running at least 1,000,000 generations of Metropolis-coupled Markov chains with a burn-in of 2500 replications until the standard deviation of split frequencies was below 0.01.

Three-dimensional prediction and distribution of sites on protein structures

We mapped the amino acids that might affect molecular convergent evolution of echolocation onto three-dimensional protein structures by PyMOL (The PyMOL Molecular Graphics System, Version 1.4 Schrödinger, LLC.). To mark the amino acids on a structurally obtained protein, we searched for gene names on the RCSB Protein Data Bank (<http://www.rcsb.org/>) (Bernstein et al. 1977). If tertiary structures were not obtained, the most likely three-dimensional model was predicted using the I-TASSER web tool according to the given quality criteria (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Zhang 2008; Roy et al. 2010; Yang et al. 2015). We used human sequences of each protein gathered from UniProt as templates (<http://www.uniprot.org/>) (Apweiler et al. 2004).

RESULTS

Sequence convergence of echolocation in the common shrew

Of the 11 hearing-related genes examined in this study, *Kcnq4* (in cats, bottlenose dolphins, and finless porpoises) and *Gjb6* (in cats) were excluded from our study due to premature stop codons. Besides, there was no substitution saturation in any sequence used in this study.

A total of 93 true positive convergent or parallel amino acid substitutions were identified between the common shrew and other sophisticated echolocators (Table S2). Of them, four substitutions were convergent, and the remaining (89 of 93) were parallel. Except for

Tmc1, 10 other hearing-related genes within our study were separately detected, and at least one convergent or parallel amino acid substitution was shared between the common shrew and sophisticated echolocators. Two convergent substitutions existed in *CDH23* (G/S1057N in the common shrew and the ancestor of Vespertilionoidea, T/R1693S in the common shrew and *Chaerephon plicatus*), and two occurred in *SLC4A11* (N/D453S in the common shrew and the ancestor of Yangochiroptera, N/Q522H and *Myotis lucifugus*). There were 45 parallel amino acid substitutions in the protein *CDH23*, four parallel substitutions in *DFNB59*, three in *GJB6*, three in *KCNQ4*, four in *MYO6*, 17 in *OTOF*, three in *PCDH15*, one in *POU3F4*, eight in *PRESTIN*, and five in *SLC4A11*. In addition, four parallel substitutions in *CDH23* (T449A, F1420V, I2105V, D3060E), one parallel substitution in *DFNB59* (I72V), and two amino acids in *OTOF* (T1041S, S1312G) were shared by the common shrew, echolocating bats, and toothed whales. It was inferred that the parallel substitutions of all echolocating lineages might add weight to the molecular evidence that echolocation underwent convergent evolution. Figure 2 shows the distribution of convergent and parallel substitutions among echolocating mammals.

In addition to the test of convergent and parallel amino acid substitutions, site-wise log-likelihood detected amino acids with a positive ΔSSLS_i value, which suggests higher fitness to the topologies of echolocator-convergence. First, for the nested hypotheses of H_0 and H_A in most genes, there was

no significant difference between the ΔSSLS_i and $\Delta\text{SSLS}_{i\text{-control}}$ in each hearing-related gene (Fig. S1). After ranking all the amino acid sites of each gene by ΔSSLS_i , we observed that nearly half of the amino acid residues of each locus had positive ΔSSLS_i , which supports the echolocator-converged hypotheses. To search for true positive amino acid sites supporting the echolocator-converged hypotheses, we screened out sites with positive $\Delta\text{SSLS}_{i\text{-control}}$ that indicated support for non-echolocator-converged hypotheses. The true positive sites could offer more evidence for the molecular convergence of echolocation. Consequently, table S3 illustrates true positive amino acid sites with positive ΔSSLS_i , supporting echolocators convergence in the top 5% within each locus. In the top 5% of amino acids ranked by ΔSSLS_i of 11 hearing-related genes, 5.6% to 69.8% were regarded as true positive echolocator-convergence supporting sites.

In this study, we focus not only on the convergent changes to the exact same amino acid, but also on the convergent amino acid physicochemical property changes. Changes in convergent amino acid physicochemical properties allow for detailed structural and functional contributions to the convergent evolution of echolocation. Analytical results of TreeSAAP implied a set of amino acids with convergent radical property changes in echolocating species. Twenty-four amino acid residues (13 in protein *CDH23*, one in *GJB6*, eight in *OTOF*, and two in *PRESTIN*) were found with radical property changes occurring only in the common shrew and either echolocating bats or dolphins

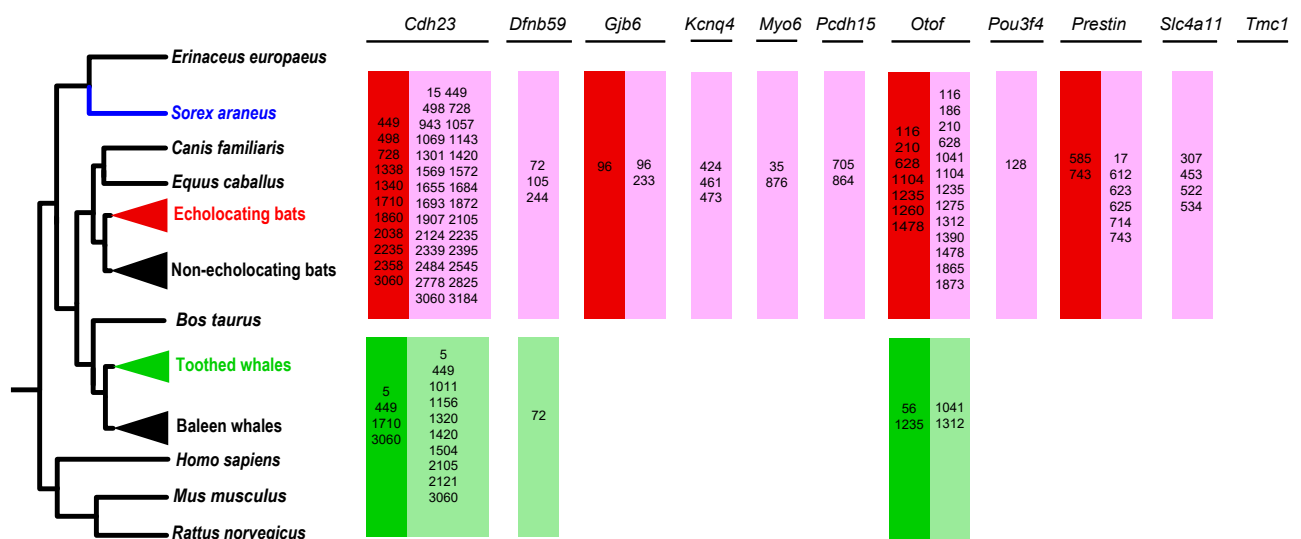


Fig. 2. Radical property changes and convergent/parallel amino acid substitutions in echolocating mammals. Lineages in red, green, and blue indicate echolocating bats, toothed whales, and the common shrew, respectively. Convergent and parallel amino acid substitutions in common shrew and echolocating bat lineages are in pink boxes. Sites with radical property changes in the common shrew and echolocating bats are depicted in red boxes. Sites in light green indicate the convergent and parallel substitutions between common shrew and toothed whales, and sites in green boxes indicate radical property changes.

(Table S4). All 24 amino acids with radical property changes in echolocators showed various quantities of altered properties ranging from one to thirteen. Of these, residues 449, 1710, and 3060 in CDH23, and 1235 in OTOF possessed the radical property changes shared by all the ultrasound users, including the common shrew, echolocating bat, and toothed whale branches. Co-occurrences of the four amino acids in CDH23 and OTOF identified in echolocators might be particularly strong evidence for convergent evolution of echolocation. Figure 2 shows the distribution of amino acid sites with radical property changes among echolocating mammals.

We also found that 13 amino acid sites—5, 449, 498, 728, 2235, 3060, and 3184 in protein CDH23; 116, 210, 1104, 1235, and 1478 in protein OTOF; and 743 in PRESTIN—overlapped in the detection of convergent molecular shifting in echolocators, which supported the common shrew clusters with other echolocating mammals (Table 1). Together, convergent/parallel amino acid substitutions in hearing-related genes characterized by radical property changes as well as echolocator-converged topology support are speculated to contribute to the convergent evolution between echolocators.

Phylogenetic reconstruction

The ML and BI trees based on complete nucleotide alignments of each gene basically agreed with the well-accepted species tree. Neither the ML nor BI trees based on entire amino acid sequences, nor the trees based on different codon position alignments, could group the three subsets of ultrasonic users

together. The trees based on synonymous or non-synonymous amino acid substitutions did not cluster the common shrew with other sophisticated echolocators. The multilocus concatenated tree based on amino acid alignment clustered echolocating bats of two suborders and grouped toothed whales with baleen whales, which was the same as in the species tree (Fig. 3B). Surprisingly, the BI tree based on the combined datasets of sites possessing the radical property changes, true positive convergent and parallel substitutions, and true positive echolocator-converged hypotheses supporting amino acids of 11 genes grouped the common shrew and echolocating bats with a notably high posterior probability (Fig. 3A).

DISCUSSION

Convergent evolution of echolocation acts as a complex and specialized adaptation, and studies on bats and dolphins have given molecular evidence for natural selection in hearing-related genes (Li et al. 2008; Liu et al. 2010 2011; Davies et al. 2012; Ru et al. 2012; Shen et al. 2012; Kirwan et al. 2013). For instance, the motor protein PRESTIN expressed in OHCs was speculated to undergo sequence convergence, and therefore the lineages of echolocating bats were clustered with dolphins in gene trees based on amino acid alignment (Liu et al. 2010; Li et al. 2010). Similar to the environmental challenges that bats and dolphins face, such as low luminosity and dysphotia, the common shrew is characterized as a nocturnal vocal mammal with poor eyesight that lives exclusively in dense

Table 1. Amino acids show evidence for hypotheses that echolocation is convergent. Sites overlapped in analyses of convergent/parallel substitutions, radical property changes, and site-wise log-likelihood support

Gene	Site*	Convergent/parallel substitution	Radical property changes	SSLS _i		
				H ₁	H ₂	H ₃
<i>Cdh23</i>	5	√	√		√	
	449	√	√	√		
	498	√	√	√	√	
	728	√	√	√	√	
	2235	√	√		√	√
	3060	√	√	√	√	
	3184	√	√		√	
<i>Otof</i>	116	√	√	√	√	
	210	√	√	√	√	
	1104	√	√		√	
	1235	√	√		√	
	1478	√	√	√	√	√
<i>Prestin</i>	743	√	√		√	

Site*: With human protein sequence as a reference.

vegetation (Merritt et al. 1994). Although the laryngeal-emitted broadband ultrasonic clicks of the common shrew do not emit as much high-sound pressure as the sophisticated echolocation clicks from bats or toothed whales, they are necessary and effective in probing protective cover and avoiding predators (Forsman and Malmquist 1988). However, echolocation in the common shrew has so far not been addressed through molecular evidence.

This study used a series of molecular evolutionary analyses of hearing-related genes and provided clues regarding echolocation in the common shrew and convergent evolution between the common shrew and sophisticated echolocating bats and dolphins. Convergent or parallel substitutions, in combination with radical amino acid property changes and site-wise log-likelihood support (SSLS) analysis, identified 13 amino acids as strong candidate convergent sites between different mammalian echolocating groups

(Table 1). These convergent sites are usually located within or near important functional domains of hearing-related genes.

For example, CDH23 is expressed in the cochlea and critical for proper morphogenesis and functioning of mechanosensitive hair bundles of the inner ear's neurosensory cells (Frolenkov et al. 2004). Great quantities of mutations in CDH23 have been associated with hearing loss phenotypes and Usher syndrome type 1D, which is a rare inherited disorder characterized by the sensory neural hearing loss (Frolenkov et al. 2004; Miyagawa et al. 2012). The convergent residue 449 identified in this study is located in cadherin domain 4, where three calcium ions are usually bound at the interface to rigidify the connections by imparting a strong curvature to the full-length ectodomain, whereas residues 498, 728, and 2235 are located in cadherin domains 5, 7, and 21, respectively, and play a role in mediating adhesive specificity in the process

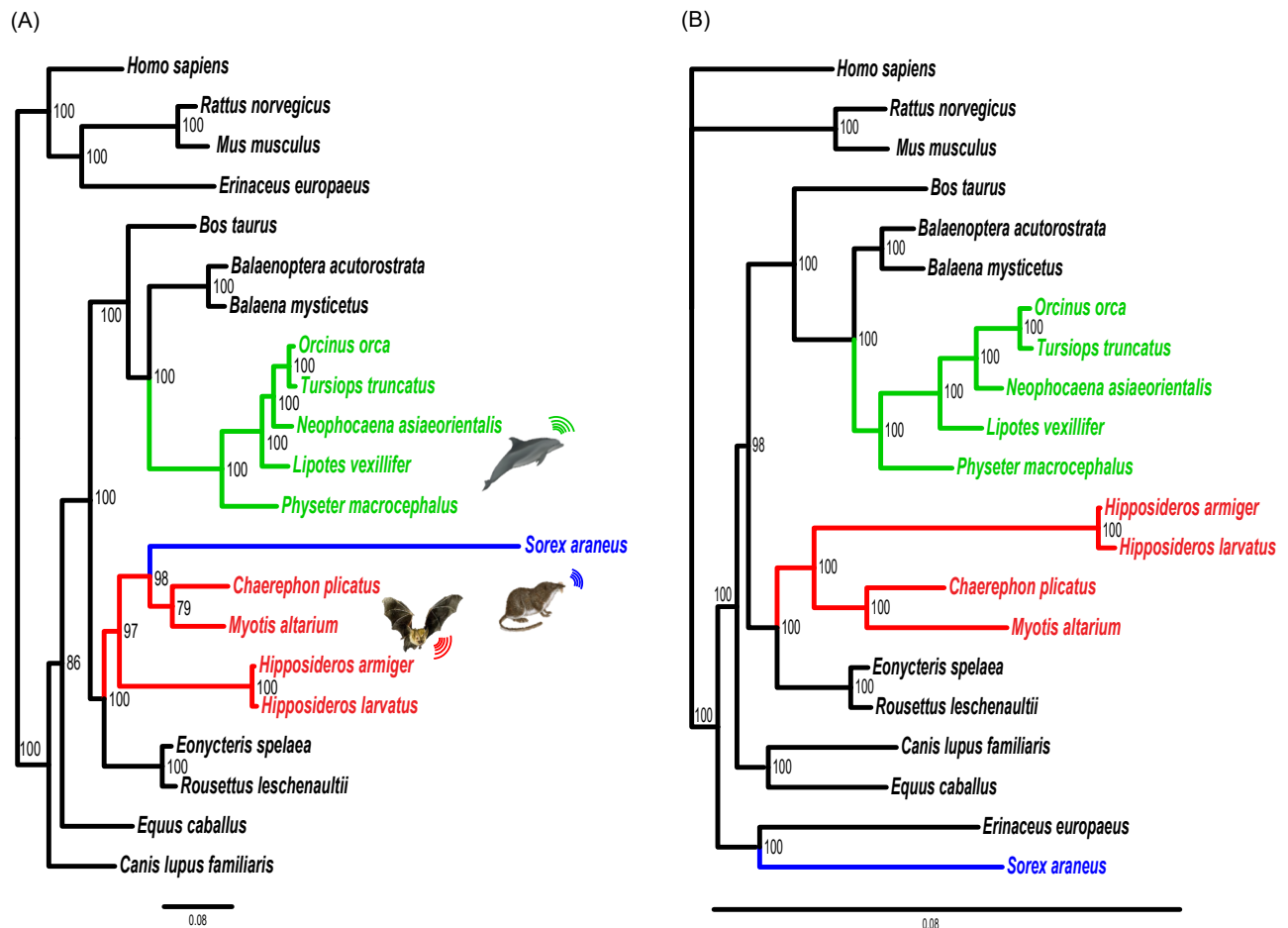


Fig. 3. Convergence of echolocation in *Sorex araneus* and sophisticated echolocators. (A) Convergence of the sophisticated echolocating bats and simple echolocator *Sorex araneus* shown by the multilocus concatenated gene tree based on 481 potential convergence contributing amino acid sites in BI method. (B) Multilocus concatenated BI tree based on the whole multilocus concatenated amino acid alignment. (Numbers at internal nodes indicate posterior probability).

of forming a coherent hair cell bundle and initiating contact between contiguous stereocilia (Boěda et al. 2002; Goodwin 2004). Residue 3184 is within a 20-residue peptide fragment corresponding to amino acid residues 3181–3200, which are crucial to form a complex structure binding to harmonin. This structure is necessary for the functional tip link of hair cells through multidentate binding modality (Pan et al. 2009; Caberlotto et al. 2011) (Fig. 4A and 4B). For the OTOF, five convergent amino acids were found in the β -strand of the C2A domain or the adjoining regions of domains C2B, C2D, and C2E, which may affect OTOF's capacity to bind to the trigger Ca^{2+} and consequently release neurotransmitters in IHCs (Roux et al. 2006; Helfmann et al. 2011). It has been reported that mutations in position 1235 might be involved in the pathogenesis of auditory neuropathy (Zadro et al. 2010) (Fig. 4C). PRESTIN is an important motor protein of OHCs (Zheng et al. 2000) and is required for electro motility and cochlear amplification (Rodríguez-Ballesteros et al. 2008). Convergent residue 743 is perched in a coil

structure of the C-terminus, which is a less conserved intracellular region, with E743D as a unique substitution in the common shrew and some echolocating bats (Fig. 4D); this suggests that it may function with other sequences or individual amino acids in the C-terminus to locate PRESTIN or sense voltage as a motor protein in cochlear OHCs (Zheng et al. 2001 2005; Pasqualetto et al. 2008). The convergent amino acid sites and their potential impacts on gene functions could provide novel insights into the evolution of echolocation in common shrews.

Convergence can also be supported by discrepant phylogenetic clustering against the accepted species tree topology. Liu et al. (2010) found that trees based on PRESTIN alignment had good support for dolphins and echolocating bats. Coincidentally, Li et al. (2010) reported bat-dolphin convergence based on protein PRESTIN sequence concurrently. Our phylogenetic reconstruction based on each coding region sequences generated topologies congruent with generally accepted species trees; however, the BI tree of combined

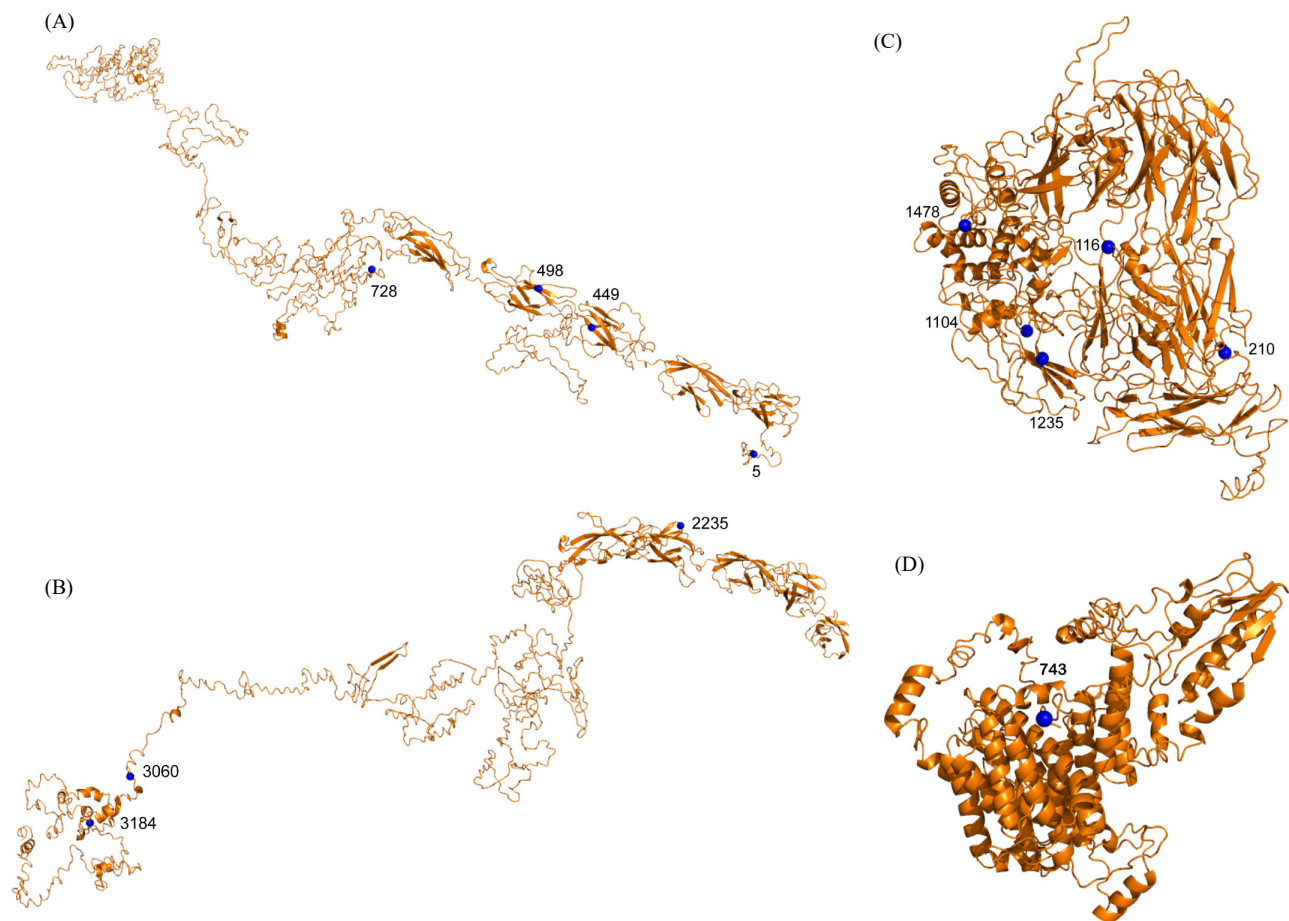


Fig. 4. Predicted three-dimensional distribution of amino acids overlapped in convergent evolutionary analyses. (A), (B) CDH23 longer than 1500 amino acids were predicted into separate parts. Overlapped sites in CDH23 are indicated in blue. (C) Overlapped sites in OTOF are indicated in blue. (D) Overlapped sites in PRESTIN are indicated in blue.

convergent amino acids (Fig. 3) had some unexpected results, such as a more significant convergence between the common shrew and echolocating bats. It is inferred that this kind of grouping might be due to the significant difference between habitat media for sound and body size. Foremost, both bats and the common shrew are mammals that hear in air, while dolphins use an entirely different medium. These differences in acoustic media lead to differences in sound speed and atmospheric absorption, as well as a different physical basis for auditory systems. Second, the high-frequency sounds emitted by echolocators are relative to the size of their transmitting aperture (Wilson et al. 2013; Madsen and Surlykke 2014). The body and aperture sizes of the common shrew are more similar to those of bats than toothed whales.

CONCLUSIONS

Evolutionary analyses of auditory genes on extended datasets including the common shrew and additional cetaceans give novel insights into the molecular convergence of echolocation in common shrews. Convergently evolving amino acids in the hearing-related proteins CDH23, OTOF, and PRESTIN were inferred to play a vital function in high-frequency hearing. Molecular evidence could further extend our understanding of echolocation evolution in mammals with diverse levels of echolocating capacity. To summarize, this study provides molecular evidence for the evolution of echolocation in the common shrew and potential convergence with other mammalian echolocators. Further studies are essential to clarify the inherent molecular mechanisms by which echolocation converged in these mammals, from comparative genomics to functional validation experiments.

List of abbreviations

BI, Bayesian Inference.
IHCs, Inner Hair Cells.
ML, Maximum Likelihood.
OHCs, Outer Hair Cells.
SSLS, Site-wise log-likelihood Support.

Acknowledgments: We express our great gratitude to Kaiya Zhou, Xinrong Xu, and some students for their sample collecting. We also thank Zhengfei Wang, Di Sun, and other coworkers for their help and support, and Noah Last of Third Draft Editing for editing the English language.

Authors' contributions: BC, GY, WR, SC, and SX

conceived this study. GL, SC, and XR collected and analyzed the data. SC was a major contributor in writing manuscript. GY, SX, and RT refined manuscript. All authors read and approved the final manuscript.

Competing interests: BC, GY, WR, SC, RT, SX, GL, and XR declare that they have no conflict of interest.

Availability of data and materials: The datasets supporting the conclusions of this article are included within the article and its additional files.

Consent for publication: Not applicable.

Ethics approval consent to participate: Not applicable.

REFERENCES

- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. 2004. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **32**:D115–D119. doi:10.1093/nar/gkh131.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The protein data bank. *FEBS J* **80**:319–324. doi:10.1111/j.1432-1033.1977.tb11885.x.
- Boëda B, El-Amraoui A, Bahloul A, Goodyear R, Daviet L, Blanchard S, Perfettini I, Fath KR, Shorte S, Reiners J, Houdusse A, Legrain P, Wolfrum U, Richardson G, Petit C. 2002. Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. *EMBO J* **21**:6689–6699. doi:10.1093/emboj/cdf689.
- Buchler ER. 1976. The use of echolocation by the wandering shrew (*Sorex vagrans*). *Anim Behav* **24**:858–873.
- Caberlotto E, Michel V, Foucher I, Bahloul A, Goodyear RJ, Pepermans E, Michalski N, Perfettini I, Alegria-Prévot O, Chardenoux S, Do Cruzeiro M, Hardelin JP, Richardson GP, Avan P, Weil D, Petit C. 2011. Usher type 1G protein sans is a critical component of the tip-link complex, a structure controlling actin polymerization in stereocilia. *P Natl Acad Sci USA* **108**:5825–5830. doi:10.1073/pnas.1017114108.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**:1164–1165. doi:10.1093/bioinformatics/btr088.
- 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. doi:10.1038/hdy.2011.119.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797. doi:10.1186/1471-2105-5-113.
- Eick GN, Jacobs DS, Matthee CA. 2005. A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). *Mol Biol Evol* **22**:1869–1886. doi:10.1093/molbev/msi180.
- Fenton MB, Bell GP. 1981. Recognition of species of insectivorous bats by their echolocation calls. *J Mammal* **62**:233–243.

- Fenton MB, Jensen FH, Kalko EK, Tyack PL. 2014. Sonar signals of bats and toothed whales. In: Surlykke A, Nachtigall PE, (ed) Biosonar. Springer, New York, pp. 11–59.
- Forsman KA, Malmquist MG. 1988. Evidence for echolocation in the common shrew, *Sorex araneus*. *J Zool* **216**:655–662.
- Foot AD, Liu Y, Thomas GW, Vinar T, Alföldi J, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet* **47**:272–275. doi:10.1038/ng.3198.
- Frolenkov GI, Belyantseva IA, Friedman TB, Griffith AJ. 2004. Genetic insights into the morphogenesis of inner ear hair cells. *Nat Rev Genet* **5**:489–498. doi:10.1038/nrg1377.
- Goodwin M. 2004. Classical cadherin adhesion molecules: coordinating cell adhesion, signaling and the cytoskeleton. *J Mol Biol* **35**:839–844. doi:10.1007/s10735-004-1833-2.
- Helfmann S, Neumann P, Tittmann K, Moser T, Ficner R, Reisinger E. 2011. The crystal structure of the C₂A domain of otoferlin reveals an unconventional top loop region. *J Mol Biol* **406**:479–490. doi:10.1016/j.jmb.2010.12.031.
- Keane M, Semeiks J, Webb AE, Li YI, Quesada V, et al. 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep* **10**:112–122. doi:10.1016/j.celrep.2014.12.008.
- Kirwan JD, Bekaert M, Commins JM, Davies KT, Rossiter SJ, Teeling EC. 2013. A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. *Evol Appl* **6**:412–422. doi:10.1111/eva.12047.
- Kyhn LA, Jensen FH, Beedholm K, Tougaard J, Hansen M, Madsen PT. 2010. Echolocation in sympatric Peale's dolphins (*Lagenorhynchus australis*) and Commerson's dolphins (*Cephalorhynchus commersonii*) producing narrow-band high-frequency clicks. *J Exp Biol* **213**:1940–1949. doi:10.1242/jeb.042440.
- Li G, Wang J, Rossiter SJ, Jones G, Cotton JA, Zhang S. 2008. The hearing gene *Prestin* reunites echolocating bats. *P Natl Acad Sci USA* **105**:13959–13964. doi:10.1073/pnas.0802097105.
- Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene *Prestin* unites echolocating bats and whales. *Curr Biol* **20**:R55–R56. doi:10.1016/j.cub.2009.11.042.
- 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. doi:10.1016/j.cub.2009.11.058.
- Liu Z, Li S, Wang W, Xu D, Murphy RW, Shi P. 2011. Parallel evolution of *KCNQ4* in echolocating bats. *PLoS ONE* **6**:e26618. doi:10.1371/journal.pone.0026618.
- Madsen PT, Surlykke A. 2014. Echolocation in air and water. In: Surlykke A, Nachtigall PE (ed) Biosonar. Springer, New York, pp. 257–304.
- McClellan DA, McCracken KG. 2001. Estimating the influence of selection on the variable amino acid sites of the cytochrome *b* protein functional domains. *Mol Biol Evol* **18**:917–925. doi:10.1093/oxfordjournals.molbev.a003892.
- McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsbery JK. 2005. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome *b* proteins. *Mol Biol Evol* **22**:437–455. doi:10.1093/molbev/msi028.
- McClellan DA, Ellison DD. 2010. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *Int J Bioinform Res Appl* **6**:120–133. doi:10.1504/IJBRA.2010.032116.
- Merritt JF, Kirkland GL, Rose RK. 1994. Advances in the biology of shrews. Carnegie Museum of Natural History, Pittsburgh Pennsylvania.
- Miyagawa M, Nishio SY, Usami SI. 2012. Prevalence and clinical features of hearing loss patients with *CDH23* mutations: a large cohort study. *PLoS ONE* **7**:e40366. doi:10.1371/journal.pone.0040366.
- Norberg UM, Rayner JMV. 1987. Ecological morphology and flight in bats (Mammalia: Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philos Trans R Soc Lond B: Biol Sci* **316**:335–427.
- Pan L, Yan J, Wu L, Zhang M. 2009. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. *P Natl Acad Sci USA* **106**:5575–5580. doi:10.1073/pnas.0901819106.
- Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**:228–231. doi:10.1038/nature12511.
- Pasqualetto E, Seydel A, Pellini A, Battistutta R. 2008. Expression, purification and characterisation of the C-terminal STAS domain of the SLC26 anion transporter prestin. *Protein Express Purif* **58**:249–256. doi:10.1016/j.pep.2007.12.006.
- Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics (Oxford, England)* **14**:817–818.
- Rodríguez-Ballesteros M, Reynoso R, Olarte M, Villamar M, Morera C, et al. 2008. A multicenter study on the prevalence and spectrum of mutations in the otoferlin gene (*OTOF*) in subjects with nonsyndromic hearing impairment and auditory neuropathy. *Hum Mutat* **29**:823–831. doi:10.1002/humu.20708.
- Roux I, Safieddine S, Nouvian R, Grati MH, Simmler MC, Bahloul A, Perfettini I, Le Gall M, Rostaing P, Hamard G, Triller A, Avan P, Moser T, Petit C. 2006. Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell* **127**:277–289. doi:10.1016/j.cell.2006.08.040.
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**:725–238. doi:10.1038/nprot.2010.5.
- Ru B, Han N, He G, Brayer K, Zhang S, Wang Z. 2012. Molecular Cloning and Evolutionary Analysis of *GJB6* in Mammals. *Biochem Genet* **50**:213–226. doi:10.1007/s10528-011-9463-3.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**:539–542. doi:10.1093/sysbio/sys029.
- Schnitzler HU, Moss CF, Denzinger A. 2003. From spatial orientation to food acquisition in echolocating bats. *Trends Ecol Evol* **18**:386–394. doi:10.1016/S0169-5347(03)00185-X.
- Shen YY, Liang L, Li GS, Murphy RW, Zhang YP. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLoS Genet* **8**:e1002788. doi:10.1371/journal.pgen.1002788.
- Siemers BM, Schauermaier G, Turni H, von Merten S. 2009. Why do shrews twitter? Communication or simple echo-based orientation. *Biol Letters* **5**:593–596. doi:10.1098/rsbl.2009.0378.
- Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033.
- Steele ME, Hebsgaard MB, Fordyce RE, Ho SY, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sørensen MV, Willerslev E. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst Biol* **58**:573–585. doi:10.1093/sysbio/syp060.
- Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet* **14**:751–764. doi:10.1038/nrg3483.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**:2731–2739. doi:10.1093/molbev/msr121.
- Thomas JA, Jalili MS. 2004. Echolocation in insectivores and rodents. In: Thomas JA, Moss CF, Vater M (ed) Echolocation in bats and

- dolphins. University of Chicago Press, Chicago, pp. 547–564.
- Wang GD, Fan RX, Zhai W, Liu F, Wang L, Zhong L, Wu H, Yang HC, Wu SF, Zhu CL, Li Y, Gao Y, Ge RL, Wu CI, Zhang YP. 2014. Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau. *Genome Biol Evol* **6**:2122–2128. doi:10.1093/gbe/evu162.
- Wilson M, Wahlberg M, Surlykke A, Madsen PT. 2013. Ultrasonic predator-prey interactions in water-convergent evolution with insects and bats in air? *Front Physiol* **4**:137. doi:10.3389/fphys.2013.00137.
- Wierer M, Schrey A K, Kühne R, Ulbrich SE, Meyer HH. 2012. A Single Glycine-Alanine Exchange Directs Ligand Specificity of the Elephant Progesterone Receptor. *PLoS ONE* **7**:e50350. doi:10.1371/journal.pone.0050350.
- Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* **19**:671–672. doi:10.1093/bioinformatics/btg043.
- Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol* **30**:1720–1728. doi:10.1093/molbev/mst064.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586–1591. doi:10.1093/molbev/msm088.
- Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. 1982. Living with water stress: evolution of osmolyte systems. *Science* **217**:1214–1222. doi:10.1126/science.7112124.
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**:7–8. doi:10.1038/nmeth.3213.
- Zadro C, Ciorba A, Fabris A, Morgutti M, Trevisi P, Gasparini P, Martini A. 2010. Five new OTOF gene mutations and auditory neuropathy. *Int J Pediatr Otorhi* **74**:494–498. doi:10.1016/j.ijporl.2010.02.004.
- Zakon HH. 2002. Convergent evolution on the molecular level. *Brain Behav Evol* **59**:250–261. doi:10.1159/000063562.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**:40. doi:10.1186/1471-2105-9-40.
- Zheng J, Du GG, Matsuda K, Orem A, Aguiñaga S, Deák L, Navarrete E, Madison LD, Dallos P. 2005. The C-terminus of prestin influences nonlinear capacitance and plasma membrane targeting. *J Cell Sci* **118**:2987–2996. doi:10.1242/jcs.02431.
- Zheng J, Long KB, Shen W, Madison LD, Dallos P. 2001. Prestin topology: localization of protein epitopes in relation to the plasma membrane. *Neuroreport* **12**:1929–1935. doi:10.1097/00001756-200107030-00032.
- Zheng J, Shen W, He DZ, Long KB. 2000. Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**:149–155. doi:10.1038/35012009.
- Zhou X, Seim I, Gladyshev VN. 2015. Convergent evolution of marine mammals is associated with distinct substitutions in common genes. *Sci Rep-UK* **5**:16550. doi:10.1038/srep16550.

Supplementary Materials

Fig. S1. Distribution of ΔSSLS_i and $\Delta\text{SSLS}_{i\text{-control}}$. Asterisks in boxplots show the mean values. ns: $p > 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. (download)

Table S1. Taxonomy and accession numbers of sequences used in this study. (download)

Table S2. Convergent and parallel amino acid substitutions in echolocating mammals. (download)

Table S3. Amino acids with true positive ΔSSLS_i support echolocator-convergence in top 5%. (download)

Table S4. Radical amino acid property changes in echolocating mammals. (download)