

## EVOLUTIONARY BIOLOGY

## Molecular parallelism in fast-twitch muscle proteins in echolocating mammals

Jun-Hoe Lee<sup>1,2,3\*</sup>, Kevin M. Lewis<sup>4</sup>, Timothy W. Moural<sup>4</sup>, Bogdan Kirilenko<sup>1,2,3</sup>, Barbara Borgonovo<sup>1</sup>, Gisa Prange<sup>5</sup>, Manfred Koessl<sup>5</sup>, Stefan Huggenberger<sup>6</sup>, ChulHee Kang<sup>4</sup>, Michael Hiller<sup>1,2,3†</sup>

Detecting associations between genomic changes and phenotypic differences is fundamental to understanding how phenotypes evolved. By systematically screening for parallel amino acid substitutions, we detected known as well as novel cases (*Strc*, *Tecta*, and *Cabp2*) of parallelism between echolocating bats and toothed whales in proteins that could contribute to high-frequency hearing adaptations. Our screen also showed that echolocating mammals exhibit an unusually high number of parallel substitutions in fast-twitch muscle fiber proteins. Both echolocating bats and toothed whales produce an extremely rapid call rate when homing in on their prey, which was shown in bats to be powered by specialized superfast muscles. We show that these genes with parallel substitutions (*Casq1*, *Atp2a1*, *Myh2*, and *Myl1*) are expressed in the superfast sound-producing muscle of bats. Furthermore, we found that the calcium storage protein calsequestrin 1 of the little brown bat and the bottlenose dolphin functionally converged in its ability to form calcium-sequestering polymers at lower calcium concentrations, which may contribute to rapid calcium transients required for superfast muscle physiology. The proteins that our genomic screen detected could be involved in the convergent evolution of vocalization in echolocating mammals by potentially contributing to both rapid  $\text{Ca}^{2+}$  transients and increased shortening velocities in superfast muscles.

## INTRODUCTION

An important aspect to understanding how nature's phenotypic diversity evolved is to detect the genomic differences that are associated with phenotypic differences (1). Despite numerous sequenced genomes, detecting such associations remains a challenge. Convergent evolution, which refers to the repeated evolution of similar phenotypes in independent lineages, offers a paradigm to computationally screen genomes for molecular changes that evolved in parallel in these lineages and thus could be involved in the phenotypic difference (2–5).

A fascinating example of phenotypic convergence is the ability of bats and toothed whales to use biosonar (echolocation) for navigating and hunting. Both independent mammalian lineages have evolved a highly sophisticated echolocation system that involves producing biosonar signals (calls or clicks), listening for the returning echoes, and converting this information into an auditory representation of their environment. Both lineages have converged on the ability to hear high-frequency sounds, which are particularly suited for the detection of small prey. Furthermore, despite having evolved two different vocalization systems (laryngeal in bats, nasal in toothed whales), both lineages have converged on adopting a similar strategy for foraging. Bats and toothed whales are able to precisely control the rate of echolocation calls and drastically increase this rate to 160 or more calls per second in the final moments of capturing prey, a feature known as the “terminal buzz.” This ability allows the echolocators to receive near-instantaneous feedback on the position of their prey and home in with high precision (6, 7).

To detect genomic changes involved in the convergent hearing properties of these echolocating lineages, several studies have searched and identified parallel amino acid substitutions in candidate genes with a known hearing-related function such as *Dfnb59* and *Prestin* (8–10). Subsequently, *in vitro* assays demonstrated that parallel substitutions in *Prestin*, a protein known to be critical for high-frequency hearing, account for the functional convergence in key parameters of *Prestin* (11). This finding shows that searching for molecular parallelism in candidate genes provides insights into the convergent adaptations to high-frequency hearing; however, the mechanisms underlying the convergent vocalization aspect of echolocation remain unknown.

Here, we systematically searched for parallel molecular evolution to possibly find new associations with convergent phenotypic adaptations. To this end, we systematically screened one-to-one orthologous proteins for parallel amino acid substitutions in independent lineages. This screen uncovered known as well as novel cases of parallelism between echolocating bats and toothed whales in proteins that could contribute to high-frequency hearing adaptations. Unexpectedly, we found an unusually high number of parallel substitutions between these echolocators in proteins crucial for fast muscle activity. We found that these proteins are expressed in superfast muscles of an echolocating bat and show functional convergence of the calcium storage protein *Casq1*. Our study raises the possibility that these proteins are involved in the evolution of superfast muscles that power the terminal buzz, and provides a general framework for using computational approaches to potentially discover the genomic basis of novel phenotypes.

## RESULTS

To systematically search for parallel molecular evolution, we applied a computational method that identifies parallel amino acid substitutions occurring in pairs of independent lineages (fig. S1A). First, we computed a multiple protein alignment using the phylogeny-aware

<sup>1</sup>Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

<sup>2</sup>Max Planck Institute for the Physics of Complex Systems, Dresden, Germany.

<sup>3</sup>Center for Systems Biology Dresden, Dresden, Germany. <sup>4</sup>Department of Chemistry, Washington State University, Pullman, WA 99164–4630, USA. <sup>5</sup>Institute for Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany. <sup>6</sup>Department II of Anatomy—Neuroanatomy, University of Cologne, Cologne, Germany.

\*Present address: Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

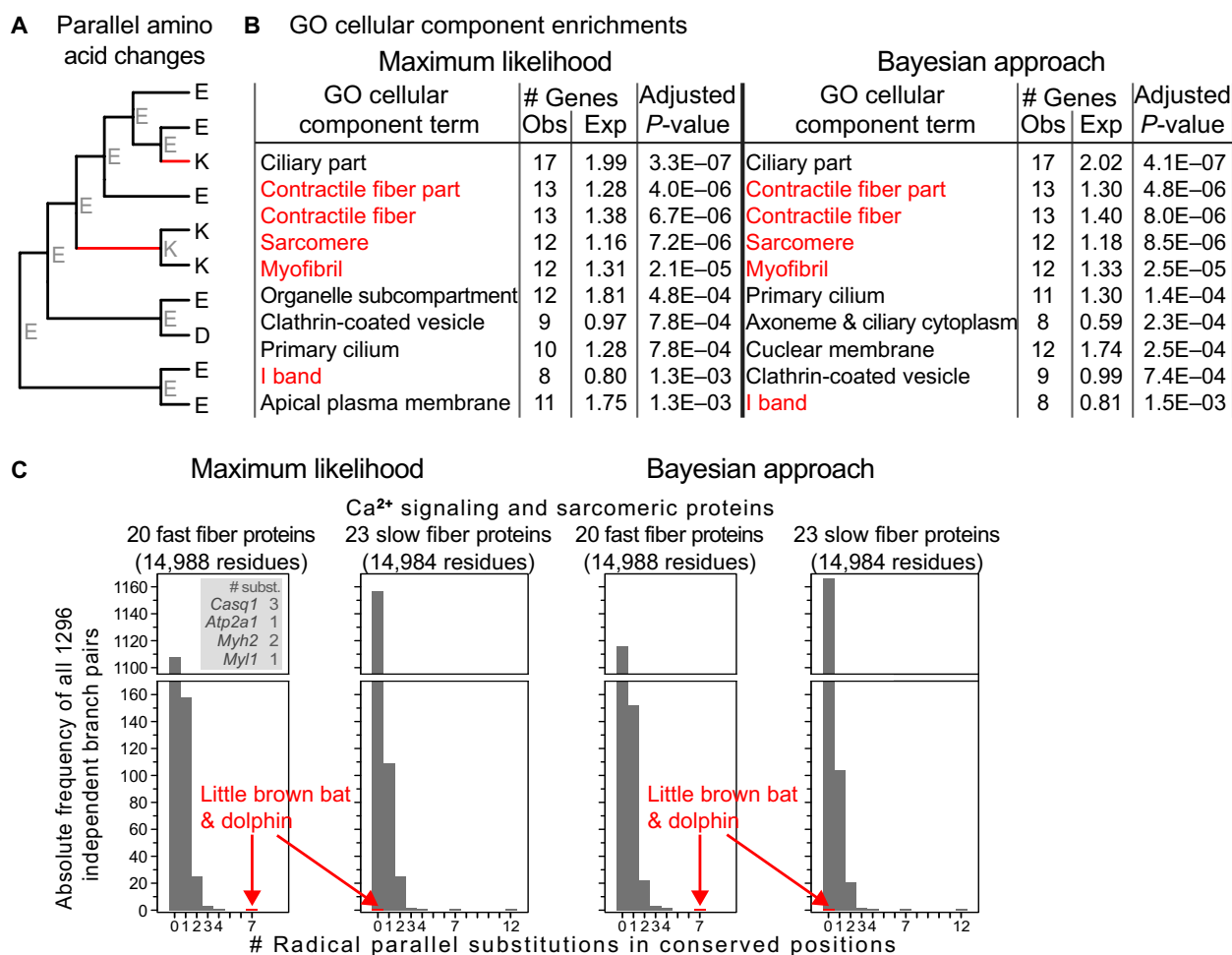
†Corresponding author. Email: hiller@mpi-cbg.de

Prank method (12). Second, we inferred the most likely ancestral protein sequence at every internal node in the phylogenetic tree using both maximum likelihood and Bayesian approaches. Third, considering each protein position and all possible pairs of branches, we extracted those pairs in which one or more parallel amino acid substitutions have occurred (Fig. 1A and fig. S1, B to E). The result is a list of lineage pairs that share at least one parallel substitution.

To systematically implement this approach, it is essential to have high-quality sets of one-to-one orthologous genes that also exclude any similar paralogs. Therefore, we used Ensembl Compara (13), which provided ortholog data for 30 different placental mammals, comprising 10 primates, the tree shrew, 5 glires, 11 laurasiatherians, and 3 afrotherians (fig. S2). A total of 14,406 ortholog sets were obtained after removing poorly aligned regions and spurious sequences. We found that the vast majority (~90%) of the detected parallel substitutions occur at alignment positions that are poorly conserved in

evolution or are conservative replacements (similar physicochemical properties of amino acids) (table S1). In contrast to conservative substitutions, radical substitutions that change physicochemical properties of amino acids such as charge, polarity, or aromaticity occur at a lower rate in evolution, because these substitutions are more likely to affect protein function and are thus under stronger purifying selection. Aiming at detecting parallel substitutions that may affect function, we decided to focus on proteins with at least one parallel substitution that occurs in a conserved alignment position and that results in a radical amino acid replacement (fig. S1, B to E).

As our data set contained two echolocating mammals, the microbat *Myotis lucifugus* (little brown bat) and the bottlenose dolphin (*Tursiops truncatus*), we tested whether our screen detected known parallel substitutions. To this end, we extracted all proteins that contain radical parallel substitutions in conserved positions between microbat and dolphin. With maximum likelihood and Bayesian



**Fig. 1. Screen for parallel molecular evolution in mammals.** (A) Illustration of parallel amino acid substitutions, defined as the same amino acid substitution from an identical ancestral state. Inferred ancestral amino acids are in gray font. Red branches have a parallel substitution from E (Glu) to K (Lys) at this alignment position. (B) Functional enrichments of 409 (416) proteins, where maximum likelihood (Bayesian) ancestral reconstruction infers parallel substitutions between the microbat (little brown bat) and the bottlenose dolphin, using GeneTrail2 (<https://genetrail2.bioinf.uni-sb.de/>). Ontology terms relating to muscle fiber components are in red font. Only the top 10 enrichments are shown for cellular component (65 significant enrichments). Muscle-related and other non-muscle-related enrichment terms for GO biological process and molecular function are shown in table S2. (C) Histograms showing the number of radical parallel amino acid substitutions in conserved positions of fast-twitch versus slow-twitch Ca<sup>2+</sup> signaling and sarcomeric proteins for all 1296 independent pairs of branches. Ancestral sequence reconstruction with maximum likelihood (left) and a Bayesian approach (right) consistently shows that microbat and dolphin have seven parallel substitutions exclusively in fast-twitch fiber proteins (inset) and that no other branch pair has an excess of seven or more substitutions in the fast-twitch fiber set.

ancestral reconstruction, this resulted in sets of 409 and 416 proteins, respectively. For 392 proteins (>94% of 409/416), both approaches consistently detected parallel substitutions. These 392 proteins included the hearing-related proteins Prestin, Dfnb59, and Slitrk6 (fig. S3A) that have been identified in previous candidate gene studies (4, 8–10).

Our screen also detected previously unknown parallel substitutions in three additional hearing-related proteins: stereocilin (*Strc*), tectorin  $\alpha$  (*Tecta*), and calcium binding protein 2 (*Cabp2*) (fig. S3B). The function of *Strc* is to connect stereocilia to each other as well as to the tectorial membrane (14), which contains *Tecta* as a major component (15). *Cabp2* plays a role in regulating the activity of voltage-gated calcium channels that open in response to stereocilia deflection (16). Thus, our screen discovered additional hearing-related proteins that may contribute to convergent adaptations for high-frequency hearing in echolocating mammals.

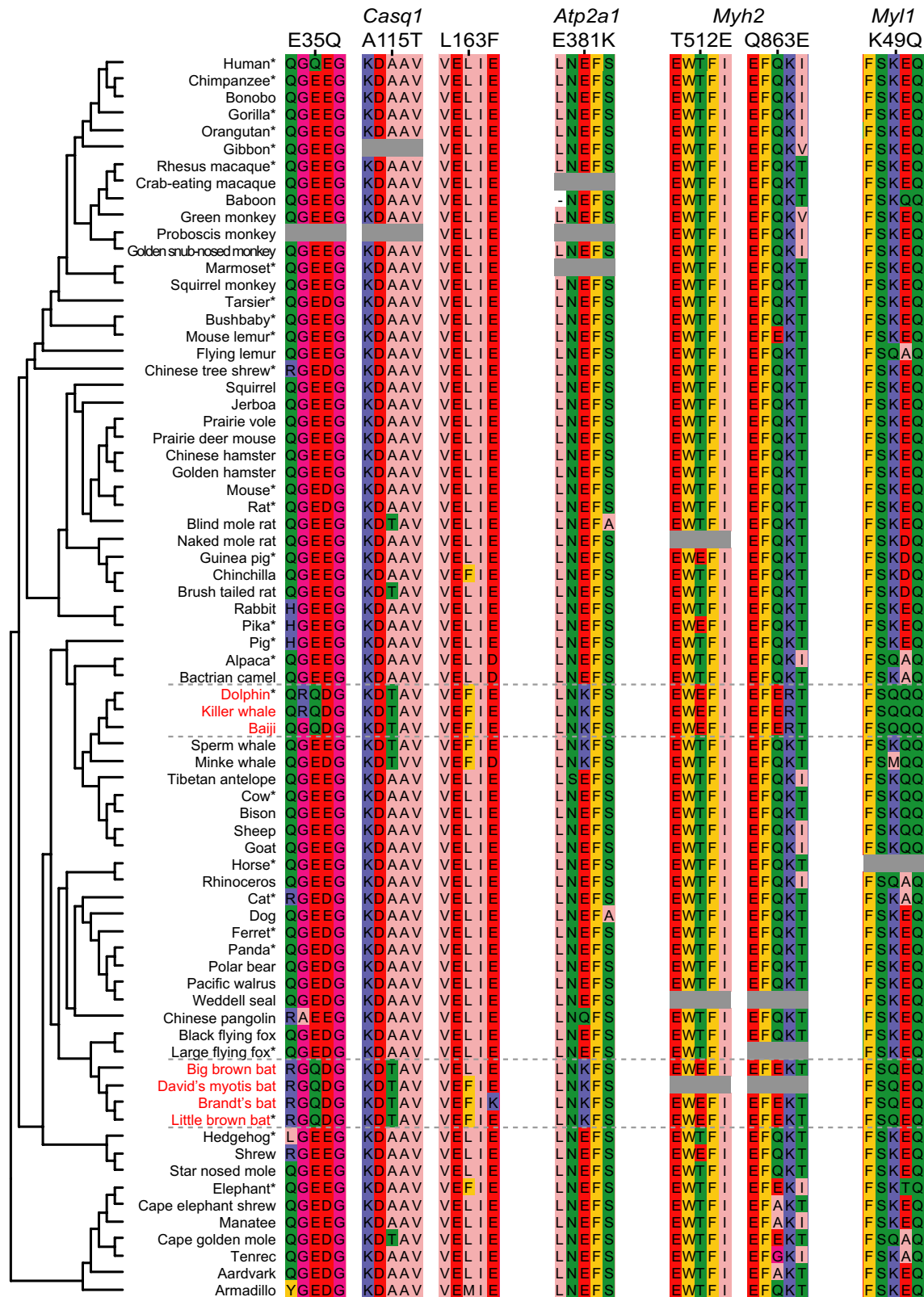
To explore the functions of the remaining non-hearing-related proteins with parallel substitutions between the microbat and bottlenose dolphin, we searched for functional enrichments in these sets of 409/416 proteins. Surprisingly, we found that the proteins with parallel substitutions between microbat and dolphin are, along with other gene ontology (GO) terms, enriched for terms relating to components of the contractile machinery of muscle fibers and calcium signaling (Fig. 1B and table S2). We examined the proteins across the different GO terms for muscle and calcium signaling, and observed that while most of them also have functional roles outside of muscle fibers or are part of cardiac muscles, the following four proteins have specific roles in skeletal muscle fibers (17, 18): calsequestrin 1 (*Casq1*), adenosine triphosphatase (ATPase) sarcoplasmic/endoplasmic reticulum calcium transporting 1 (*Atp2a1*, synonym *Serca1*), myosin heavy chain 2 (*Myh2*), and myosin light chain 1 (*Myl1*) (Figs. 2 and 3). *Casq1* adsorbs large amounts of  $\text{Ca}^{2+}$  ions and serves as the  $\text{Ca}^{2+}$  ion storage protein in the sarcoplasmic reticulum (SR) (19–21). *Atp2a1* pumps  $\text{Ca}^{2+}$  into the SR to achieve muscle relaxation (22). Both *Myh2* and *Myl1* are components of the force-generating molecular motor complex (17, 23). In these four proteins, we observed a total of seven radical parallel substitutions in conserved positions between microbat and dolphin.

All four proteins (*Casq1*, *Atp2a1*, *Myh2*, and *Myl1*) are known to be predominantly expressed in fast-twitch skeletal muscle fibers, while respective paralogs of these proteins are usually expressed at higher levels in slow-twitch or cardiac muscle fibers. Therefore, we examined whether these parallel substitutions are specific to fast-twitch muscle proteins or whether microbat and dolphin also exhibit parallel substitutions in  $\text{Ca}^{2+}$  signaling or sarcomeric proteins that are predominantly expressed in slow-twitch muscle fibers. To this end, we used available mouse gene expression data (17, 18) and obtained 20 proteins (*Atp2a1*, *Calm3*, *Casq1*, *Mybph*, *Myh1*, *Myh13*, *Myh2*, *Myh8*, *Myl1*, *Myl7*, *Myom2*, *Myoz1*, *Pak1*, *Pdlim3*, *Ppp3ca*, *Pvalb*, *Tmod1*, *Tnnc2*, *Tnnt3*, and *Vcl*) with higher expression levels in fast-twitch muscle fibers and 23 proteins (*Acta1*, *Actn2*, *Atp2a2*, *Casq2*, *Cryab*, *Hspb1*, *Itgb1bp2*, *Itpr1*, *Murc*, *Myh3*, *Myh6*, *Myh7*, *Myl2*, *Myl3*, *Myot*, *Myoz2*, *Smpx*, *Smtnl1*, *Tnnc1*, *Tnni1*, *Tnnt1*, *Tnnt2*, and *Tpm2*) with higher expression levels in slow-twitch muscle fibers. In contrast to the fast-twitch muscle proteins, which included the aforementioned 4 proteins, we did not detect a single parallel substitution between microbat and dolphin in any of these 23 slow-twitch fiber proteins (Fig. 1C). Next, we used our screen that detected parallel substitutions between any pair of lineages to

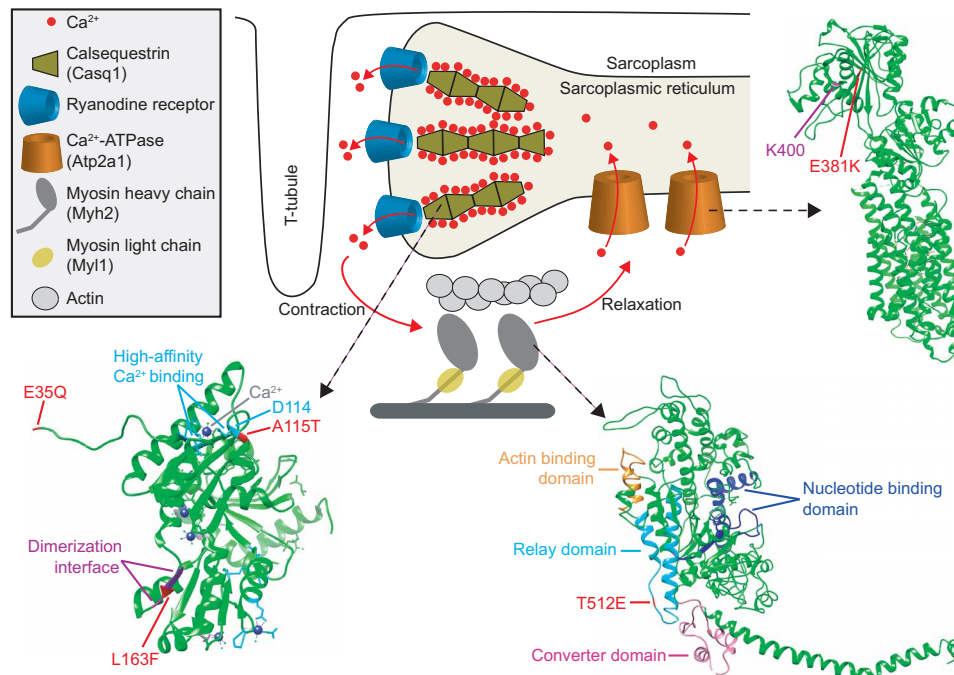
test whether other mammals also have seven or more parallel substitutions in the 20 fast-twitch fiber proteins. Among all other 1295 pairs of independent branches, we observed a maximum of four parallel substitutions (table S3), showing that the seven parallel substitutions between microbat and dolphin are an outlier observation (Fig. 1C). Furthermore, while microbat and dolphin have no parallel substitutions in the 23 slow-twitch proteins, other pairs of lineages have up to 12 such substitutions in this set. This shows that the number of parallel substitutions that occur specifically in fast-twitch fiber proteins in echolocators is the highest compared to other lineage pairs. Next, we determined whether GC-biased gene conversion (a process that biases G/C over A/T alleles during recombination repair) may be a potential explanation for the observed excess of parallel substitution between dolphin and microbat in the fast-twitch fiber proteins. We found that only one of the seven substitutions (K49Q in *Myl1*) could potentially be explained by GC-biased gene conversion (table S4). In summary, we observed that (i) the echolocating microbat and dolphin have more parallel substitutions in fast-twitch fiber proteins than all other lineages, (ii) both species have no parallel substitutions in slow-twitch fiber proteins, and (iii) GC-biased gene conversion is not a major contributing factor explaining the presence of these parallel substitutions in fast-twitch fiber proteins. While parallel substitutions often occur by chance (24, 25), these observations suggest that the seven parallel substitutions in the fast-twitch muscle proteins of echolocators, as a whole, may not have arisen due to chance alone.

In bats, the extremely rapid echolocation calls during the terminal buzz are powered by unique, sound-producing “superfast” muscles, which consist of specialized fast-twitch fibers that display contraction rates similar to the call rates during the terminal buzz (~200 times per second) (26, 27). Just like bats, toothed whales such as bottlenose dolphin, harbour porpoise, and false killer whale use terminal buzz in the final moments before capturing prey and are able to dynamically control the echolocation click rates based on the distance to their prey (7, 28). This behavioral similarity suggests that superfast sound-producing muscles may also exist in these toothed whales. Thus, we hypothesized that these proteins could be involved in functional changes that convert fast muscle fibers into superfast fibers and thus potentially contribute to the convergent vocalization aspect of echolocation.

Because our orthologous gene sets comprised only 30 placental mammals including 2 echolocators, we manually analyzed 46 additional mammals to determine whether the parallel substitutions in these proteins also occurred in related echolocating bats and toothed whales. We found that most of the parallel amino acid changes are also observed in other echolocating mammals (big brown bat, additional *Myotis* species, killer whale, and baiji), though not exclusively in them (Fig. 2). A further manual examination of the six aforementioned hearing-related proteins (*Prestin*, *Dfnb59*, *Slitrk6*, *Strc*, *Tecta*, and *Cabp2*) also shows that none of the parallel substitutions occur specifically in these echolocating mammals, with the exception of *Strc* H320Q that our screen uncovered (fig. S3). Notably, the N7T and P26L substitutions that were experimentally shown to affect *Prestin* function also occur in the non-echolocating elephant shrew and cat/minke whale, respectively (fig. S3A). This pattern suggests that the observation of a derived substitution in a background species does not necessarily preclude a functional effect on the protein and that pure patterns of convergence rarely exist when taxonomic sampling depth increases (29). Furthermore, in various examples such as *Prestin*, *RNAse1*, and  $\text{Na}^+/\text{K}^+$ -ATPase, functional convergence is



**Fig. 2. Parallel molecular evolution in fast-twitch muscle fiber proteins in echolocating mammals.** In addition to species included in the genome-wide screen (marked by asterisk), the alignments contain 46 additional mammals, which shows that parallel amino acid substitutions also occur in other echolocating mammals that emit calls at a high repetition rate (big brown bat, other *Myotis* species, killer whale, and baiji) but not exclusively in them. Sequences of additional mammals were extracted from a recent multiple genome alignment (46). Gray box, missing sequence.



**Fig. 3. Simplified diagram showing the skeletal muscle contraction-relaxation cycle and 3D structures of proteins with parallel substitutions.**  $\text{Ca}^{2+}$  bound to Casq1 is released from the SR through ryanodine receptors into the sarcoplasm to initiate the cross-bridge cycle during which myosin complexes generate mechanical work. For relaxation,  $\text{Ca}^{2+}$  is actively pumped back into the SR by Atp2a1, where majority of  $\text{Ca}^{2+}$  ions are bound again to polymerized Casq1. The 3D protein structures of Casq1 (Protein Data Bank: 3TRP; residues that bind  $\text{Ca}^{2+}$  with high affinity are in light blue), Atp2a1 (4NAB), and Myh2 (2MYS; Q863E is in the myosin light chain binding region but not part of this structure) are shown. Parallel substitutions are indicated in red. Residues involved in Casq1 dimerization and  $\text{Ca}^{2+}$  binding (21) and the Atp2a1 residue K400 that binds the inhibitor phospholamban (47) are labeled. The Myh2 relay domain that determines the actin sliding velocity (39) by communicating conformational changes between the converter domain, nucleotide-binding site, and actin-binding site (38, 48) is indicated. K49Q is not part of any available Myl1 structure.

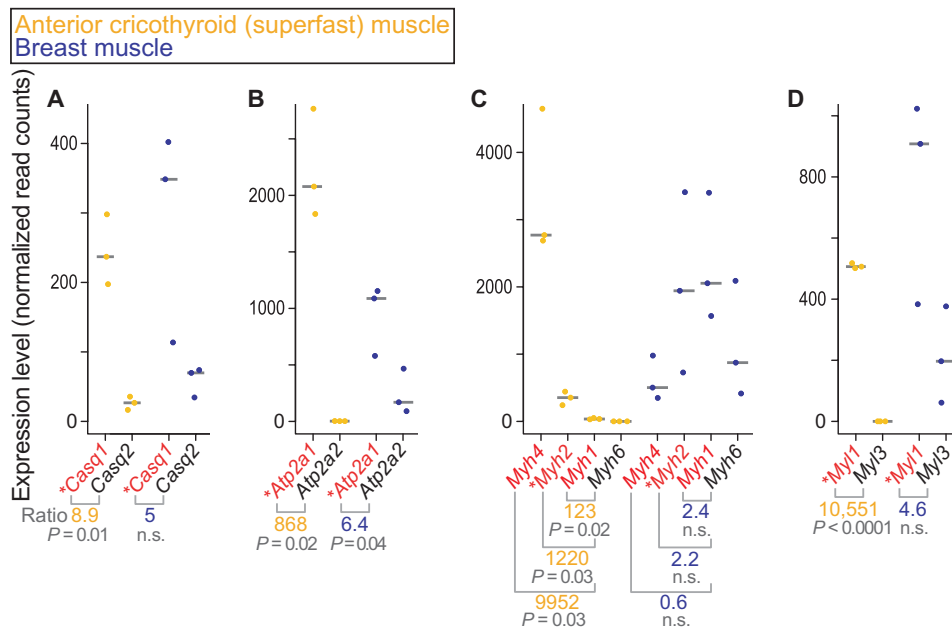
observed along with the presence of multiple sites displaying parallel substitutions (11, 30, 31).

Next, we examined whether the four fast-twitch fiber proteins (Casq1, Atp2a1, Myh2, and Myl1) are expressed in the specific sound-producing superfast muscles. Although our attempts to extract RNA from the nasal muscles of the harbour porpoise (a close relative of the dolphin) were unsuccessful due to poor RNA preservation in the available samples, we succeeded in extracting high-quality RNA from the laryngeal anterior cricothyroid muscle ( $n = 3$ ) of the echolocating bat *Pteronotus parnellii*. Using RNA sequencing (RNA-seq), we found that all four proteins were expressed in this superfast sound-producing muscle (Fig. 4) (26). Moreover, using breast muscle as control ( $n = 3$ ), we found that the anterior cricothyroid muscle has a substantially higher expression ratio of the four proteins relative to their slow-twitch fiber paralogs (Fig. 4). We further compared the expression of the three fast-twitch fiber myosin heavy chains (*Myh1/2/4*) whose encoded proteins differ in their fiber shortening velocity, which decreases in the following order: *Myh4* > *Myh1* > *Myh2* (23). We found that *Myh4* makes up 90% of the myosin heavy chain expression in the anterior cricothyroid muscle, consistent with the need for fast contraction. In contrast, *Myh4* has the lowest expression in breast muscle, where the expression order of *Myh4* > *Myh2* > *Myh1* observed in the anterior cricothyroid muscle is reversed. *Myh2* is the second most abundant myosin heavy chain (9%) in anterior cricothyroid muscle, despite the *Myh2* protein having a lower shortening velocity than *Myh1*. Thus, the superfast anterior cricothyroid muscle contains a higher proportion of fast-

twitch fiber muscle components and expresses all four proteins with parallel substitutions.

If the proteins exhibiting parallel substitutions contribute to the function of superfast fibers, then we would expect that protein function has converged between microbat and dolphin in a manner that helps to achieve the extremely rapid kinetics of superfast muscles. The rate-limiting step is muscle relaxation, which involves  $\text{Ca}^{2+}$  transport from the sarcoplasm into the SR (27, 32), where it is adsorbed to Casq1. The adsorption of  $\text{Ca}^{2+}$  depends on the ability of Casq1 to form dimers and tetramers in a  $\text{Ca}^{2+}$ -dependent manner, resulting in polymers that adsorb large amounts of  $\text{Ca}^{2+}$  on the negatively charged surface and between the interfaces (20, 33). Because the superfast contraction/relaxation cycles are too short to pump all of the released  $\text{Ca}^{2+}$  back into the SR, the  $\text{Ca}^{2+}$  concentration in the SR decreases despite less  $\text{Ca}^{2+}$  being released into the sarcoplasm in subsequent cycles (34). Therefore, the  $\text{Ca}^{2+}$ -dependent polymerization of Casq1 in microbat and dolphin may have changed to work under lower concentrations in the SR of repeatedly stimulated superfast muscles.

To explore whether the parallel substitutions in Casq1 could potentially affect  $\text{Ca}^{2+}$ -dependent polymerization, we examined their positions in the three-dimensional (3D) structure (Fig. 3). We found that the E35Q substitution is located at the N terminus, which participates in an arm exchange to stabilize dimer formation (33, 35). The L163F substitution changes the hydrophobicity of the dimer interface, likely influencing the calcium-induced hydrophobic interactions. The third substitution (A115T) is located adjacent to one



**Fig. 4. Comparison of gene expression of the four fast-twitch muscle fiber proteins relative to their slow-twitch paralogs.** Expression level of calsequestrin (A),  $\text{Ca}^{2+}$  ATPase (B), myosin heavy chain (C), and myosin light chain (D) genes comparing fast-twitch (red font) and slow-twitch (black font) muscle fiber components in three biological replicates of *P. parnellii* anterior cricothyroid muscle (yellow) and breast muscle (blue). Horizontal line is the median. Genes with parallel substitutions are marked by an asterisk. Expression ratios and *P* values of a two-sided *t* test are shown at the bottom and in table S6. n.s., not significant.

of the high-affinity  $\text{Ca}^{2+}$  binding residues (D114, Fig. 3), which interacts with the tetrameric partner. This suggests that all three radical substitutions could affect  $\text{Ca}^{2+}$ -dependent polymerization, which is crucial for the high-capacity  $\text{Ca}^{2+}$  storage function of Casq1.

To experimentally investigate whether microbat and dolphin Casq1 polymerization has evolved to work under lower  $\text{Ca}^{2+}$  concentrations, we first conducted turbidity assays to compare  $\text{Ca}^{2+}$  concentration-dependent polymerization of Casq1 between echolocating (dolphin and microbat) and non-echolocating mammals (mouse, the megabat *Pteropus vampyrus*, horse, and pig) (Fig. 5A). For both microbat and dolphin Casq1, turbidity of the solution started to increase significantly at a  $\text{Ca}^{2+}$  concentration between 1.5 and 2.0 mM, respectively, and plateaued at  $\sim 3$  mM  $\text{Ca}^{2+}$ . This is in contrast to Casq1 from non-echolocating species, where substantially higher  $\text{Ca}^{2+}$  concentrations are required to increase turbidity significantly. Second, to directly monitor  $\text{Ca}^{2+}$ -dependent dimerization at low  $\text{Ca}^{2+}$  concentrations, we used a multiangle light scattering experiment, which confirmed that microbat and dolphin Casq1 dimerizes at lower  $\text{Ca}^{2+}$  concentrations compared to mouse Casq1 (Fig. 5B). In summary, these results demonstrate that microbat and dolphin Casq1 are able to form polymers at lower  $\text{Ca}^{2+}$  concentrations than Casq1 of non-echolocating mammals, which could facilitate  $\text{Ca}^{2+}$  adsorption under decreasing concentrations in the SR during superfast contraction-relaxation cycles.

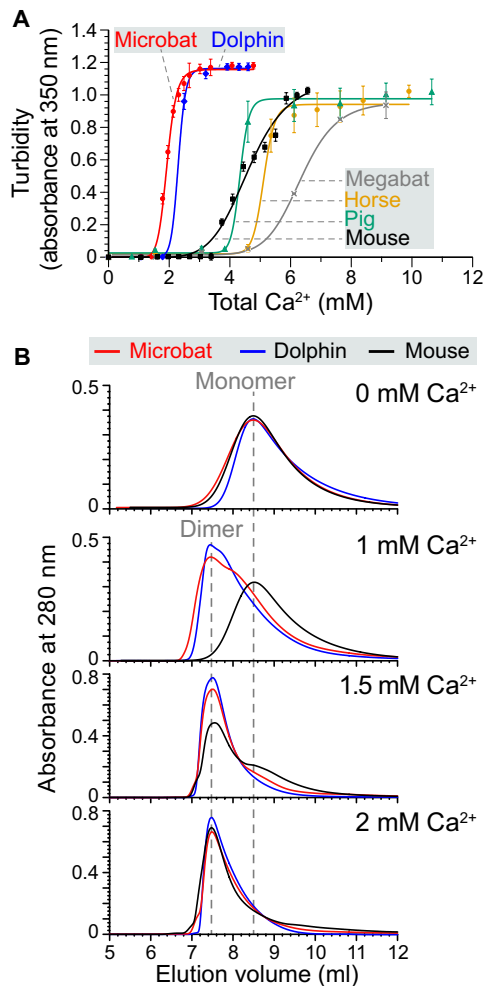
## DISCUSSION

Here, we show that echolocating bats and dolphins have an unusually high number of parallel substitutions in proteins that (i) are specific to the contractile machinery of fast-twitch muscle fibers and (ii) are expressed in the superfast muscle of *P. parnellii*, which powers the terminal buzz in bats. Superfast muscles consist of spe-

cialized fibers capable of contracting and relaxing at a rate that is an order of magnitude higher compared to the fastest locomotor muscles (32). To achieve this extraordinarily high rate, superfast muscles have evolved a number of key adaptations to greatly accelerate muscle relaxation, which is the rate-limiting step. These adaptations include (i) rapid  $\text{Ca}^{2+}$  transients and (ii) a higher fiber shortening velocity due to myosin motors with a fast cross-bridge detachment rate (32, 36, 37), suggesting evolutionary tinkering with various components of the  $\text{Ca}^{2+}$  signaling and molecular motor machinery. The proteins that our genomic screen detected could thus be potentially involved in the evolution of superfast muscles by contributing to both rapid  $\text{Ca}^{2+}$  transients and high shortening velocities.

Rapid  $\text{Ca}^{2+}$  transients require a fast decrease of the  $\text{Ca}^{2+}$  concentration in the sarcoplasm, which is achieved through a high amount of Atp2a1 and parvalbumin (Pvalb), a protein that temporarily binds  $\text{Ca}^{2+}$  in the sarcoplasm until it can be pumped back into the SR by Atp2a1 (32). As expected, we observed a high expression level of Atp2a1 and Pvalb in the bat anterior cricothyroid muscle (Fig. 4 and table S5). In addition, our experiments show that both microbat and dolphin Casq1 are able to form polymers at lower  $\text{Ca}^{2+}$  concentrations than mouse Casq1. This functional convergence suggests that microbat and dolphin Casq1 could adsorb  $\text{Ca}^{2+}$  under conditions of lower  $\text{Ca}^{2+}$  concentrations in the SR. These conditions are likely present in superfast muscles, where the contraction/relaxation cycles are too short to fully restore the basal  $\text{Ca}^{2+}$  concentration in the SR. Therefore, the ability of Casq1 to adsorb  $\text{Ca}^{2+}$  under lower concentrations may increase the SR storage capacity during superfast contraction/relaxation cycles and thus contribute to the rapid calcium transients required for superfast muscle physiology.

Besides rapid  $\text{Ca}^{2+}$  transients, a high fiber shortening velocity due to myosin motors with a fast cross-bridge detachment rate is also a distinguishing feature of superfast muscles (32, 37). Fiber shortening



**Fig. 5. Microbat and dolphin Casq1 functionally converged in the ability to form Ca<sup>2+</sup>-sequestering polymers at lower Ca<sup>2+</sup> concentrations.** (A) Ca<sup>2+</sup>-dependent turbidity shows that microbat and dolphin Casq1 oligomerizes at lower Ca<sup>2+</sup> concentrations. Error bars represent standard deviations of triplicate measurements. (B) Multiangle light scattering experiments show that microbat and dolphin Casq1 dimerizes at lower Ca<sup>2+</sup> concentrations compared to mouse Casq1. While Casq1 from all three species is monomeric at a Ca<sup>2+</sup> concentration of 0 mM, both microbat and dolphin Casq1 dimerizes at a concentration of 1 mM, in contrast to mouse Casq1 that remains monomeric.

velocity is determined by the myosin relay domain (38, 39). Among the parallel substitutions in Myh2, the T512E substitution is located in the relay domain (Fig. 3) and appears as a promising candidate to increase actin sliding velocity for two reasons. First, we found that the myosin isoform present in the *Drosophila* indirect flight muscles exhibits a substitution similar to T512E (Fig. 6). These asynchronous flight muscles are capable of contracting ~200 times per second; however, each contraction is controlled by a stretch-activated mechanism instead of a Ca<sup>2+</sup> release/sequestration cycle. The myosin isoform that is exclusively expressed in flight muscles has an extremely fast cross-bridge detachment rate (37) and, as shown in Fig. 6, exhibits a negatively charged residue (Asp) at the position corresponding to Myh2 512. In contrast to the polar Thr (T), a negatively charged residue [Glu (E) in Myh2 or Asp in *Drosophila* Myh] allows the formation of a salt bridge with a positively charged Arg in the converter domain, and this interaction specifically occurs during the post-power stroke conformation that precedes cross-bridge detachment (38). Second, mammalian Myh4 myosin, which confers the highest muscle fiber shortening velocity (23), also exhibits a negatively charged residue (E) at position 512 (fig. S4). Therefore, the presence of a negatively charged residue in the fastest myosins in both *Drosophila* and mammals supports the hypothesis that T512E is involved in increasing actin sliding velocity of microbat and dolphin Myh2.

Furthermore, we observed that the derived amino acids of additional parallel substitutions in Myh2 are also found at the corresponding Myh4 positions (fig. S4). Thus, it appears that Myh2 in microbat and dolphin has converged toward a higher similarity with Myh4. If this sequence convergence is associated with functional convergence, then the higher expression of Myh2 over Myh1 in the sound-producing muscle of *P. parnellii* would lead to muscle fibers that consist almost exclusively (99%, Fig. 4) of myosin motors with a very fast actin sliding velocity. The higher expression level of Myh2 versus Myh1 in the cricothyroid muscle, coupled with the higher fatigue resistance of Myh2 compared to Myh1 and particularly Myh4 (due to the physiological trade-off between myosin speed and fatigue resistance), could potentially explain why the parallel changes occurred in Myh2 instead of Myh1. Thus, given that a higher fatigue resistance characterizes laryngeal muscle fibers (17), it remains to be tested whether the parallel changes in Myh2 increase speed while maintaining the fatigue resistance aspect of this myosin motor.



**Fig. 6. Both microbat/dolphin Myh2 and the fastest *Drosophila* myosin exhibit a negatively charged residue at position 512.** The amino acid sequence of exon 15 of the mammalian Myh2 and the homologous alternative exons 9a/9b/9c of the *Drosophila* Myh are shown. In contrast to mammals, *Drosophila* has only a single Myh gene, but different muscles produce functionally different isoforms through alternative splicing. The *Drosophila* indirect flight muscle expresses a Myh isoform with an extremely fast cross-bridge detachment rate that exclusively includes exon 9a (blue font). In contrast to exons 9b (A, Ala) and 9c (T, Thr), exon 9a exhibits a negatively charged residue (D, Asp, red arrow) that resembles the parallel substitution T512E (Thr to Glu, black arrow) in microbat/dolphin Myh2 (blue font). A negatively charged residue (Asp or Glu) at this position likely modulates the actin sliding velocity by dynamically forming a salt bridge with the positively charged Arg<sup>759</sup> in the converter domain during the post-power stroke conformation that precedes cross-bridge detachment (39). Note that mammalian Myh4 myosin, which confers the highest muscle fiber shortening velocity, also exhibits a negatively charged residue (Glu) at position 512 (fig. S4).

Notably, we found that toothed whales have a large ~50-kb deletion that covers the *Myh4* gene. This deletion has the same breakpoints in toothed and baleen whales, strongly suggesting that this deletion already occurred in the cetacean ancestor (fig. S5). We were able to validate *Myh4* deletion by analyzing sequencing reads of these species (fig. S6). This observation raises the possibility that sequence convergence in dolphin *Myh2* compensates for the ancestral loss of *Myh4*, the gene encoding the fastest muscle myosin heavy chain.

In summary, our study highlights the utility of comparative genomic approaches to generate new hypotheses of the genomic basis underlying complex phenotypes, in this case, the genomic changes that contribute to superfast muscle physiology, which has remained largely unknown. The parallel amino acid changes between echolocating mammals in several proteins, which are expressed in the fast-twitch fibers of the bat larynx, may contribute toward the functional changes required to generate exceptional speed. Although the complete morphology of sound production in toothed whales has yet to be fully understood, the molecular parallels hint toward a similar role of those proteins in muscles or tissues that control sound production in these aquatic mammals. Nonetheless, additional experiments, both in vitro and in vivo, will be necessary to fully determine whether the parallel amino acids could affect functional changes and, consequently, how these changes could play a role in building superfast muscles.

## MATERIALS AND METHODS

### Genome-wide screen for parallel amino acid substitutions in placental mammals

Figure S1A gives an overview of the steps. While we initially used the ortholog data from various platforms such as OrthoDB, OrthoMCL, and MetaPhOrs, we found that these protein sets often contain both orthologous and paralogous proteins (which could lead to spurious parallel substitutions), as well as low coverage of complete sets of one-to-one orthologs. Because high-quality sets of one-to-one orthologs are essential for a genome-wide screen, we used Ensembl Compara (release 75) (13) to obtain comprehensive and accurate lists of one-to-one orthologous genes between human and other mammals and retrieved the corresponding protein sequences. We removed ambiguous amino acids (letter X) from the sequences and only kept sequences where the length is within 80 to 120% relative to the human or mouse sequence. All ortholog sets that contained less than 15 sequences were discarded. This resulted in 14,406 sets.

To reconstruct ancestral protein sequences, we first aligned each ortholog set using Prank (version 10603, parameters: +F) with the mammalian phylogeny (fig. S2) as a fixed guide tree (12), because the phylogeny-aware Prank alignment method was reported to consistently outperform other methods. Second, to focus on protein regions with a higher chance of being reliably aligned, we automatically removed poorly aligned regions and spurious sequences from each alignment by applying trimAl (version 1.4.rev15, “gappyout” trimming mode, <http://trimal.cgenomics.org/>).

Third, we estimated ancestral sequences using either maximum likelihood or a Bayesian Monte Carlo Markov chain sampling approach. For maximum likelihood, ancestral reconstruction was split into two separate steps to improve computational runtime. Using RAXML (version 7.9.5) (40), we first estimated for each alignment

the branch lengths in a tree whose topology is constrained by the mammalian phylogeny (RAXML parameter “-g”). Then, we used Lazarus (version 2.79), which acts as a wrapper for the codeML functions in PAML (41), to estimate ancestral sequences using the fixed tree with estimated branch lengths as input. For both RAXML and PAML, we used the LG amino acid substitution model and its amino acid frequencies and allowed rate heterogeneity (RAXML parameter “-m PROTCATLG,” Lazarus parameter “fixed\_asrv=False, model=LG”). For the Bayesian approach, we used PhyloBayes (version 4.1b) (42), constraining the tree topology by the mammalian phylogeny (fig. S2) and allowing PhyloBayes to estimate branch lengths using the default CAT substitution model. Two independent Monte Carlo Markov chains were then run, sampling every 100 steps until both chains converged. On the basis of the PhyloBayes manual, an acceptable convergence level is reached when the maximal difference between the observed bipartitions (maxdiff) is less than 0.3 and the minimum effective sample size is larger than 50. After convergence, we proceeded to carry out ancestral reconstruction, with a burn-in of the first 100 steps and sampling every 10 steps. Because Bayesian approaches can have long runtimes (in total 390 CPU days for 14,396 proteins), we excluded proteins for which the chains did not converge after 2 weeks of runtime, which was the case for only 10 of the 14,406 proteins. Finally, for each alignment with its reconstructed ancestral sequences that are assigned to each node in the tree, we considered all internal and terminal branches, where an amino acid substitution occurred and identified all pairs of branches that exhibit exactly the same (parallel) amino acid substitution.

### Identifying radical substitutions in conserved positions

Given an observed parallel substitution between a pair of branches, we classified it as a radical amino acid substitution if the ancestral and the derived amino acid belong to a different physicochemical group, using an amino acid classification that is based on charge, polarity, and aromaticity: aliphatic and nonpolar (A,I,L,M,V,G), polar and uncharged (S,T,C,P,N,Q), positive charge (R,K,H), negative charge (E,D), and nonpolar and aromatic (F,W,Y). A parallel substitution in a conserved position satisfies the following two criteria. First, all species that descend from both branches share the derived amino acid. Second, at least 90% of the species outside of the two convergent lineages share the ancestral amino acid (fig. S1, B to E). By filtering for radical substitutions in conserved positions, we enrich for parallel substitutions that are more likely to affect protein function, regardless of whether these amino acid substitutions involve a single or more than one nucleotide mutation.

### Parallel substitutions in fast- and slow-twitch muscle fiber proteins

Fast- and slow-twitch muscle fiber proteins were taken from data set S1 in (18), which used gene expression data from isolated mouse fast-twitch and slow-twitch muscle fibers to obtain sarcomeric and Ca<sup>2+</sup> signaling genes with a higher expression level in fast-twitch (slow-twitch) fibers, and from (17), which specifies the fiber type expression of myosin-associated genes. We then iterated over all pairs of independent branches, counted the number of radical parallel amino acid substitutions in conserved positions, and plotted a histogram of this number. The branches in an independent pair do not share a direct common ancestor, and either branch is not a descendant of the other.



### RNA-seq in *P. parnellii* muscle tissue

The use of bats complies with all current laws for animal use and experimentation in Germany and with the Declaration of Helsinki. Anterior cricothyroid and breast muscle tissue were extracted from three adult female *P. parnellii* bats, kept in a colony in our animal house. The bats were euthanized through an overdose of Narcoren [Merial GmbH, solution contained pentobarbital (400 mg/kg)] injected intraperitoneally. The anterior cricothyroid muscle was anatomically defined according to (43). The muscle tissue was excised using surgical instruments that had been cleaned with RNaseZAP and was immediately transferred into RNAlater stabilization buffer (Qiagen).

Muscle tissue samples were approximately 1 mg and were homogenized using a pestle grinder, followed by total RNA isolation according to a standard TRIzol isolation protocol. The amount of isolated total RNA was measured with the NanoDrop (Thermo Fisher Scientific), and approximately 15 µg of RNA was found in each sample. RNA integrity was assessed using a Bioanalyzer 2100 system (Agilent Technologies), and samples with RNA integrity number (RIN; a measurement of RNA degradation) > 7 were selected for RNA-seq. An Illumina HiSeq 2500 instrument was used to generate 75–base pair (bp) paired-end reads with an insert size of approximately 250 bp.

### De novo assembly of transcriptome

Before transcriptome assembly, adapters and low-quality reads (minimum read length, 35 bp) were removed using Cutadapt (<https://cutadapt.readthedocs.io/en/stable/>). Because the available *P. parnellii* genome (4) is highly fragmented, we used the Trinity (version 2.0.6) de novo assembly package (<https://github.com/trinityrnaseq/trinityrnaseq>) to assemble the *P. parnellii* transcriptome. Highly similar transcripts were collapsed using CD-Hit (<http://weizhongli-lab.org/cd-hit/>). Annotation of the transcriptome was carried out using Blast+ against human complementary DNA sequences from Ensembl (version 86). We kept the best Blast hit with the lowest *E* value and highest percent identity, which associates each transcript to a human Ensembl gene.

### Assembly of myosin heavy chain transcripts

To check assembly correctness, we used the University of California, Santa Cruz (UCSC) genome browser's Blat to align the assembled *P. parnellii* transcripts against the high-quality genome assembly of the little brown bat (*myoLuc2*). The longest transcripts of *Casq1*, *Atp2a1*, and *Myh1* and their paralogs align with >90% identity to the orthologous gene according to the Ensembl and CESAR annotation (44, 45), showing that they are correctly assembled. In contrast, we found *Myh2* to be misassembled as its transcript aligns to various parts of different myosin heavy chain genes. This misassembly is caused by the very high sequence similarity between the coding exons of the different myosin heavy chain genes, which is likely due to purifying selection.

Therefore, we used the following rationale to manually assemble the different myosin heavy chain (*Myh*) genes. The different fast-twitch fiber *Myh* genes were already present before the divergence of the amniote ancestor approximately 310 million years ago. Thus, intronic sequences that typically evolve neutrally will be diverged between *Myh2* and other paralogs. In contrast, speciation between the little brown bat and *P. parnellii* happened much more recently around 53 million years ago (<http://timetree.org/>). Consequently,

the introns between the *Myh2* ortholog of both species evolved neutrally for a much shorter amount of time and thus still align (fig. S7A). Therefore, the amount of aligning intronic sequence distinguishes a genomic locus encoding an orthologous *Myh* gene from a locus that encodes a paralogous *Myh* gene.

To quantify the amount of aligning intronic sequence, we built alignment chains, as previously described (46), between the little brown bat genome (*myoLuc2*, reference) and the *P. parnellii* genome. Then, we selected the alignment chains that best align the intronic regions to obtain *P. parnellii* loci encoding parts of an orthologous *Myh* gene (fig. S7B). By extracting the aligning exons from these loci, we manually assembled the *Myh1/2/4/6* genes. Blat of the translated transcript sequences against the little brown bat genome shows clear alignments to the orthologous genes (Ensembl and CESAR annotation), confirming the correct assembly of these four *Myh* genes. These assembled transcripts were used to quantify the *Myh* expression below.

### Quantification of gene expression

We used the utility scripts provided by Trinity to automate the alignment of the reads from each tissue to the transcriptome using Bowtie2 (version 2.2.0, <https://github.com/BenLangmead/bowtie2>). Subsequently, the gene expression level based on the aligned reads was estimated using RSEM (version 1.2.31, <https://github.com/deweylab/RSEM>). Normalization across samples was applied using the trimmed means of *M* values, and differential expression analysis was performed using the DESeq2 package (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>).

### Functional experiments of mouse (*Mus musculus*), dolphin (*Tursiops truncatus*), and microbat (*M. lucifugus*) *calsequestrin 1*

#### Protein expression and purification

Rosetta(DE3)pLysS *E. coli* cells containing the Casq1 vectors were grown in LB media at 37°C until an OD<sub>600</sub> (optical density at 600 nm) of 0.6 was reached. Next, the incubation temperature was reduced to 25°C, and the cells were induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside and left to grow overnight. Following induction, the cells were harvested by centrifugation at 4000g for 30 min and then frozen at –20°C for later use.

For protein purification, cell pellets were suspended in 20 mM tris and NaN<sub>3</sub> (0.5 g/liter) (pH 7.5) and sonicated at 8000 rpm using 450 Sonifier (Branson Ultrasonics) until they reached apparent homogeneity. The resulting lysate was clarified by centrifugation at 20,000g and loaded onto a GE Healthcare AKTA pure fast protein liquid chromatography (FPLC) with an attached Toyopearl DEAE-650M (Tosoh Biosciences) column equilibrated with DEAE loading buffer [20 mM tris, NaN<sub>3</sub> (0.5 g/liter) (pH 7.5)]. Casq1 was eluted from the DEAE column between 12.5 and 25% buffer B by running a linear gradient of buffer A [20 mM tris, NaN<sub>3</sub> (0.5 g/liter) (pH 7.5)] to buffer B [20 mM tris, NaN<sub>3</sub> (0.5 g/liter), 2 M NaCl (pH 7.5)]. The fractions containing Casq1 were buffer-exchanged to buffer C [5 mM sodium phosphate, NaN<sub>3</sub> (0.5 g/liter) (pH 6.8)] and further purified with FPLC using a ceramic hydroxyapatite (HA) column (Bio-Rad Laboratories). Casq1 protein was loaded onto an HA column preequilibrated with buffer C. Casq1 was eluted from the HA column between 50 and 100% buffer D by running a linear gradient of buffer C to buffer D [0.5 M sodium phosphate, NaN<sub>3</sub> (0.5 g/liter) (pH 6.8)]. For the final purification step, the HA Casq1 elution was

buffer-exchanged to buffer E [20 mM tris,  $\text{NaN}_3$  (0.5 g/liter) (pH 8.5)] and loaded onto a preequilibrated Mono Q column (GE Healthcare). Casq1 was eluted from the Mono Q column by running a linear gradient from buffer E to buffer F [20 mM tris,  $\text{NaN}_3$  (0.5 g/liter), 2 M NaCl (pH 8.5)]. Fractions containing Casq1 were then buffer-exchanged into Casq1 assay buffer (20 mM Mops, 0.3 M KCl (0.5 g/liter) (pH 7.2)). Throughout the purification process, fractions containing Casq1 were identified with SDS–polyacrylamide gel electrophoresis. Protein concentrations were determined using the bicinchoninic acid assay (Thermo Fisher Scientific).

#### Molecular mass determination by multiangle light scattering

We injected 100  $\mu\text{l}$  of each protein at a concentration of 1 mg/ml, which was preequilibrated with chromatography running buffer [20 mM tris-HCl (pH 7.5), 300 mM KCl, and either with or without 1 mM  $\text{CaCl}_2$ ], onto a Bio-Sep S-2000 column (Phenomenex). The chromatography was carried out at a flow rate of 0.5 ml/min using an Acuflo series IV pump (Analytical Scientific Instruments). The eluate was passed in tandem through an ultraviolet detector (Gilson), a refractometer (Optilab DSP, Wyatt Technology), and a multiangle laser light scattering detector (Dawn EOS, Wyatt Technology). All the chromatography experiments were performed at 25°C. Scattering data were analyzed using ASTRA software (Wyatt Technology) supplied with the instrument. Relative weight-averaged molecular masses were determined from the scattering data collected for a given condition using the Zimm fitting method, in which  $K^*c/R(Q)$  is plotted against  $\sin^2(Q/2)$ , where  $Q$  is the scattering angle,  $R(Q)$  is the excess intensity ( $I$ ) of scattered light at the angle  $Q$ ,  $c$  is the concentration of the sample, and  $K^*$  is a constant equal to  $4\pi^2 n^2 (dn/dc)^2 / \lambda_0^4 N_A$  (where  $n$  is the solvent refractive index,  $dn/dc$  is the refractive index increment of scattering sample,  $\lambda_0$  is the wavelength of scattered light, and  $N_A$  is Avogadro's number). Extrapolation of a Zimm plot to zero angle was used to estimate the weight-averaged molecular mass.

#### Turbidity assays

The turbidity of Casq1 solutions (that is, absorbance at 350 nm) as a function of  $\text{Ca}^{2+}$  concentration was monitored using a Genesys 10S UV-Vis spectrophotometer (Thermo Fisher Scientific). Assays were performed using 1.3 ml of 15  $\mu\text{M}$  mouse, bat, and dolphin Casq1 in the assay buffer. Concentrated  $\text{Ca}^{2+}$  solutions (0.10, 0.25, 0.50, and 1.0 M) were added in 1.0 to 2.0  $\mu\text{l}$  aliquots to the 1.3 ml of Casq1 solutions in a quartz cuvette to achieve the proper calcium concentration. Upon addition of each concentrated  $\text{Ca}^{2+}$  aliquot, the samples were mixed by stirring with a small stir bar and allowed to equilibrate ( $dA_{350}/dt = 0$ ) before addition of the next aliquot. Dilutions from adding  $\text{Ca}^{2+}$  aliquots were included in the data analysis.

#### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/9/eaat9660/DC1>

Fig. S1. Computational approach to identify parallel amino acid substitutions between pairs of independent lineages.

Fig. S2. Phylogeny of the 30 placental mammals.

Fig. S3. Radical parallel substitutions in conserved positions between microbat and dolphin in hearing-related genes.

Fig. S4. Convergence of microbat and dolphin *Myh2* toward *Myh4*.

Fig. S5. *Myh4* is deleted in the cetacean lineage.

Fig. S6. Unassembled genomic and RNA reads confirm the absence of *Myh4* in the cetacean lineage.

Fig. S7. The amount of aligning intronic sequence distinguishes *Myh* orthologs from paralogs.

Table S1. Classification of all parallel amino acid substitutions detected by reconstructing ancestral sequences with a maximum likelihood and a Bayesian approach.

Table S2. Functional enrichments of proteins with parallel substitutions between the microbat (little brown bat) and the bottlenose dolphin.

Table S3. Pairs of independent branches and their number of parallel substitutions in fast-twitch versus slow-twitch fiber proteins.

Table S4. GC-biased gene conversion alone can potentially explain only one of the seven parallel substitutions in the fast-twitch muscle proteins.

Table S5. Top 30 genes with a significantly higher expression level in the anterior cricothyroid muscle compared to breast muscle of *P. pamellii*.

Table S6. Expression of calsequestrin,  $\text{Ca}^{2+}$  ATPase, myosin heavy chain, and myosin light chain genes in the anterior cricothyroid muscle and the breast muscle of the echolocating *P. pamellii* bat.

References (49–55)

#### REFERENCES AND NOTES

- G. A. Wray, Genomics and the evolution of phenotypic traits. *Annu. Rev. Ecol. Evol. Syst.* **44**, 51–72 (2013).
- F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
- M. Hiller, B. T. Schaar, V. B. Indjejan, D. M. Kingsley, L. R. Hagey, G. Bejerano, A “forward genomics” approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep.* **2**, 817–823 (2012).
- 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).
- L. Levin, D. Mishmar, The genomic landscape of evolutionary convergence in mammals, birds and reptiles. *Nat. Ecol. Evol.* **1**, 0041 (2017).
- E. K. V. Kalko, H.-U. Schnitzler, The echolocation and hunting behavior of Daubenton's bat, *Myotis daubentonii*. *Behav. Ecol. Sociobiol.* **24**, 225–238 (1989).
- D. M. Wisniewska, M. Johnson, P. E. Nachtigall, P. T. Madsen, Buzzing during biosonar-biased interception of prey in the delphinids *Tursiops truncatus* and *Pseudorca crassidens*. *J. Exp. Biol.* **217**, 4279–4282 (2014).
- Y. Li, Z. Liu, P. Shi, J. Zhang, The hearing gene *Prestin* unites echolocating bats and whales. *Curr. Biol.* **20**, R55–R56 (2010).
- 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).
- 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 (2012).
- 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).
- A. Löytynoja, Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
- A. J. Vilella, J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, E. Birney, EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
- E. Verpy, M. Leibovici, N. Michalski, R. J. Goodyear, C. Houdon, D. Weil, G. P. Richardson, C. Petit, Stereocilin connects outer hair cell stereocilia to one another and to the tectorial membrane. *J. Comp. Neurol.* **519**, 194–210 (2011).
- P. K. Legan, V. A. Lukashkina, R. J. Goodyear, M. Kössl, I. J. Russell, G. P. Richardson, A targeted deletion in  $\alpha$ -tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback. *Neuron* **28**, 273–285 (2000).
- I. Schrauwen, S. Helfmann, A. Inagaki, F. Predoehl, M. A. Tabatabaiefar, M. M. Picher, M. Sommen, C. Z. Seco, J. Oostrik, H. Kremer, A. Dheedene, C. Claes, E. Franssen, M. H. Chaleshtori, P. Coucke, A. Lee, T. Moser, G. Van Camp, A mutation in *CABP2*, expressed in cochlear hair cells, causes autosomal-recessive hearing impairment. *Am. J. Hum. Genet.* **91**, 636–645 (2012).
- S. Schiaffino, C. Reggiani, Fiber types in mammalian skeletal muscles. *Physiol. Rev.* **91**, 1447–1531 (2011).
- F. Chemello, C. Bean, P. Cancellara, P. Laveder, C. Reggiani, G. Lanfranchi, Microgenomic analysis in skeletal muscle: Expression signatures of individual fast and slow myofibers. *PLOS ONE* **6**, e16807 (2011).
- D. H. MacLennan, P. T. S. Wong, Isolation of a calcium-sequestering protein from sarcoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1231–1235 (1971).
- H. Park, Y. Park, E. Kim, B. Youn, K. Fields, A. K. Dunker, C. Kang, Comparing skeletal and cardiac calsequestrin structures and their calcium binding: A proposed mechanism for coupled calcium binding and protein polymerization. *J. Biol. Chem.* **279**, 18026–18033 (2004).
- E. J. Sanchez, K. M. Lewis, B. R. Danna, C. Kang, High-capacity  $\text{Ca}^{2+}$  binding of human skeletal calsequestrin. *J. Biol. Chem.* **287**, 11592–11601 (2012).

22. D. H. MacLennan, Purification and properties of an adenosine triphosphatase from sarcoplasmic reticulum. *J. Biol. Chem.* **245**, 4508–4518 (1970).
23. R. Bottinelli, R. Betto, S. Schiaffino, C. Reggiani, Maximum shortening velocity and coexistence of myosin heavy chain isoforms in single skinned fast fibres of rat skeletal muscle. *J. Muscle Res. Cell Motil.* **15**, 413–419 (1994).
24. Z. Zou, J. Zhang, No genome-wide protein sequence convergence for echolocation. *Mol. Biol. Evol.* **32**, 1237–1241 (2015).
25. 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).
26. C. P. H. Elemans, A. F. Mead, L. Jakobsen, J. M. Ratcliffe, Superfast muscles set maximum call rate in echolocating bats. *Science* **333**, 1885–1888 (2011).
27. L. C. Rome, Design and function of superfast muscles: New insights into the physiology of skeletal muscle. *Annu. Rev. Physiol.* **68**, 193–221 (2006).
28. U. K. Verfuß, L. A. Miller, P. K. D. Pilz, H.-U. Schnitzler, Echolocation by two foraging harbour porpoises (*Phocoena phocoena*). *J. Exp. Biol.* **212**, 823–834 (2009).
29. G. W. C. Thomas, M. W. Hahn, Y. Hahn, The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biol. Evol.* **9**, 213–221 (2017).
30. J. Zhang, Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat. Genet.* **38**, 819–823 (2006).
31. B. Ujvari, N. R. Casewell, K. Sunagar, K. Arbuckle, W. Wüster, N. Lo, D. O’Meally, C. Beckmann, G. F. King, E. Deplazes, T. Madsen, Widespread convergence in toxin resistance by predictable molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11911–11916 (2015).
32. L. C. Rome, D. A. Syme, S. Hollingworth, S. L. Lindstedt, S. M. Baylor, The whistle and the rattle: The design of sound producing muscles. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8095–8100 (1996).
33. S. Wang, W. R. Trumble, H. Liao, C. R. Wesson, A. K. Dunker, C. Kang, Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nat. Struct. Biol.* **5**, 476–483 (1998).
34. C. L. Harwood, I. S. Young, B. A. Tikunov, S. Hollingworth, S. M. Baylor, L. C. Rome, Paying the piper: The cost of Ca<sup>2+</sup> pumping during the mating call of toadfish. *J. Physiol.* **589**, 5467–5484 (2011).
35. H. Park, S. Wu, A. K. Dunker, C. Kang, Polymerization of calsequestrin. Implications for Ca<sup>2+</sup> regulation. *J. Biol. Chem.* **278**, 16176–16182 (2003).
36. L. C. Rome, C. Cook, D. A. Syme, M. A. Connaughton, M. Ashley-Ross, A. Klimov, B. Tikunov, Y. E. Goldman, Trading force for speed: Why superfast crossbridge kinetics leads to superlow forces. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5826–5831 (1999).
37. D. M. Swank, V. K. Vishnudas, D. W. Maughan, An exceptionally fast actomyosin reaction powers insect flight muscle. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17543–17547 (2006).
38. M. J. Bloemink, C. M. Dambacher, A. F. Knowles, G. Melkani, M. A. Geeves, S. I. Bernstein, Alternative exon 9-encoded relay domains affect more than one communication pathway in the *Drosophila* myosin head. *J. Mol. Biol.* **389**, 707–721 (2009).
39. W. A. Kronert, G. C. Melkani, A. Melkani, S. I. Bernstein, Alternative relay and converter domains tune native muscle myosin isoform function in *Drosophila*. *J. Mol. Biol.* **416**, 543–557 (2012).
40. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
41. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
42. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
43. T. A. Griffiths, Comparative laryngeal anatomy of the Big Brown bat, *Eptesicus fuscus*, and the Mustached bat, *Pteronotus parnellii*. *Mammalia* **47**, 377–394 (1983).
44. V. Sharma, A. Elghafari, M. Hiller, Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation. *Nucleic Acids Res.* **44**, e103 (2016).
45. B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. Garcia Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. Amy Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle, The Ensembl gene annotation system. *Database* **2016**, baw093 (2016).
46. V. Sharma, M. Hiller, Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res.* **45**, 8369–8377 (2017).
47. X. Dong, D. D. Thomas, Time-resolved FRET reveals the structural mechanism of SERCA-PLB regulation. *Biochem. Biophys. Res. Commun.* **449**, 196–201 (2014).
48. S. Koppole, J. C. Smith, S. Fischer, The structural coupling between ATPase activation and recovery stroke in the myosin II motor. *Structure* **15**, 825–837 (2007).
49. 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).
50. K. Verhoeven, L. Van Laer, K. Kirschhofer, P. K. Legan, D. C. Hughes, I. Schatteman, M. Verstreken, P. Van Hauwe, P. Coucke, A. Chen, R. J. H. Smith, T. Somers, F. E. Offeciers, P. Van de Heyning, G. P. Richardson, F. Wachtler, W. J. Kimberling, P. J. Willems, P. J. Govaerts, G. Van Camp, Mutations in the human  $\alpha$ -tectorin gene cause autosomal dominant non-syndromic hearing impairment. *Nat. Genet.* **19**, 60–62 (1998).
51. E. Verpy, D. Weil, M. Leibovici, R. J. Goodyear, G. Hamard, C. Houdon, G. M. Lefèvre, J.-P. Hardelin, G. P. Richardson, P. Avan, C. Petit, Stereocilin-deficient mice reveal the origin of cochlear waveform distortions. *Nature* **456**, 255–258 (2008).
52. R. Mahmood, M. Elzinga, R. G. Yount, Serine-324 of myosin’s heavy chain is photoaffinity-labeled by 3’(2’)-O-(4-benzoylbenzoyl)adenosine triphosphate. *Biochemistry* **28**, 3989–3995 (1989).
53. I. Rayment, H. M. Holden, M. Whittaker, C. B. Yohn, M. Lorenz, K. C. Holmes, R. A. Milligan, Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
54. C. Tyner, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, C. Eisenhart, C. M. Fischer, D. Gibson, J. N. Gonzalez, L. Guruvadoo, M. Haeussler, S. Heitner, A. S. Hinrichs, D. Karolchik, B. T. Lee, C. M. Lee, P. Nejad, B. J. Raney, K. R. Rosenbloom, M. L. Speir, C. Villarreal, J. Vivian, A. S. Zweig, D. Haussler, R. M. Kuhn, W. James Kent, The UCSC genome browser database: 2017 update. *Nucleic Acids Res.* **45**, D626–D634 (2017).
55. D. Stöckel, T. Kehl, P. Trampert, L. Schneider, C. Backes, N. Ludwig, A. Gerasch, M. Kaufmann, M. Gessler, N. Graf, E. Meese, A. Keller, H.-P. Lenhof, Multi-omics enrichment analysis using the GeneTrail2 web service. *Bioinformatics* **32**, 1502–1508 (2016).

**Acknowledgments:** We are grateful to H. Brandl for help with transcriptome analysis, J. Jarrells and A. Dahl for RNA-seq, and J. Rink and F. Schnorrer for helpful discussions. We thank the Computer Service Facilities of the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) and Max Planck Institute for the Physics of Complex Systems (MPI-PKS), and the Scientific Computing Facility and Gene Expression Facility of the MPI-CBG for their support. **Funding:** This work was supported by NIH grant 1R01GM11125401 and the M. J. Murdock Charitable Trust to C.K. and the Max Planck Society to M.H. **Author contributions:** J.-H.L. implemented the computational method, analyzed RNA-seq data, and performed all computational analysis. K.M.L., T.W.M., B.K., B.B., and C.K. performed calsequestrin experiments. C.K. designed and supervised calsequestrin experiments. G.P. and M.K. extracted cricothyroid and breast muscle of *P. parnellii*. S.H. extracted nasal muscles of the porpoise. M.H. conceived and supervised the study, performed data analysis, and wrote the original draft. J.-H.L., C.K., and M.H. edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All primary data (alignments, parallel substitutions, and their classification) are available at <https://bds.mpi-cbg.de/hillerlab/ParallelEvolution/>. RNA-seq data are deposited in National Center for Biotechnology Information under accession numbers SAMN09657343-SAMN09657348.

Submitted 24 April 2018

Accepted 10 August 2018

Published 26 September 2018

10.1126/sciadv.aat9660

**Citation:** J.-H. Lee, K. M. Lewis, T. W. Mural, B. Kirilenko, B. Borgonovo, G. Prange, M. Koessl, S. Huggenberger, C. Kang, M. Hiller, Molecular parallelism in fast-twitch muscle proteins in echolocating mammals. *Sci. Adv.* **4**, eaat9660 (2018).

## Molecular parallelism in fast-twitch muscle proteins in echolocating mammals

Jun-Hoe Lee, Kevin M. Lewis, Timothy W. Moural, Bogdan Kirilenko, Barbara Borgonovo, Gisa Prange, Manfred Koesl, Stefan Huggenberger, ChulHee Kang and Michael Hiller

*Sci Adv* 4 (9), eaat9660.  
DOI: 10.1126/sciadv.aat9660

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/4/9/eaat9660>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2018/09/24/4.9.eaat9660.DC1>

### REFERENCES

This article cites 55 articles, 14 of which you can access for free  
<http://advances.sciencemag.org/content/4/9/eaat9660#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.