Building superfast muscles: insights from molecular parallelism in fast-twitch muscle proteins in echolocating mammals

Jun-Hoe Lee[1,2], Kevin M Lewis[3], Timothy W Moural[3], Bogdan Kirilenko[1,2], Barbara Borgonovo[1], Gisa Prange[4], Manfred Koessl[4], Stefan Huggenberger[5], ChulHee Kang[3] and Michael Hiller[1,2*]

[1] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.
[2] Max Planck Institute for the Physics of Complex Systems, Dresden, Germany.
[3] Department of Chemistry, Washington State University, Pullman, USA.
[4] Goethe University, Frankfurt, Germany.
[5] University of Cologne, Cologne, Germany.

*To whom correspondence should be addressed:
Michael Hiller
Computational Biology and Evolutionary Genomics, Max Planck Institute of Molecular Cell Biology and Genetics & Max Planck Institute for the Physics of Complex Systems, Dresden, Germany.
Tel: +49 351 210 2781
Fax: +49 351 210 1209
Email: hiller@mpi-cbg.de

Running title: Parallel molecular evolution and superfast muscles

Keywords: comparative genomics and parallel evolution, echolocation, muscle physiology, calcium signaling, calsequestrin, myosin

Abbreviations: SR, sarcoplasmic reticulum

Abstract

Detecting associations between genomic changes and phenotypic differences is fundamental to understanding how phenotypes evolved. By developing a method to systematically screen for parallel amino acid substitutions, we discovered that echolocating bats and dolphins exhibit more parallel substitutions than expected in fast-twitch but in not slow-twitch muscle fiber proteins. Both bats and dolphins rely on specialized superfast muscles to produce an extremely rapid call rate when homing in on their prey. We show that these genes with parallel substitutions are expressed in the superfast call-producing muscle of bats and demonstrate that the calcium storage protein calsequestrin 1 of bats and dolphins form calcium-sequestering polymers at lower calcium concentrations, consistent with adaptations to rapid calcium transients required for superfast muscle physiology. Our results provide molecular insights into the convergent evolution of vocalization in echolocating mammals and demonstrate the potential of systematic screens to uncover associations between genomic changes and phenotypic differences.

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, moving prey. Furthermore, despite having evolved a different vocalization system (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 allows the echolocators to receive near instantaneous feedback on the position of their prey and home in with high precision [6-8].

In order 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, Otof, and prestin [9-12]. Subsequently, *in vitro* assays demonstrated that parallel substitutions in prestin, a protein known to be critical for high-frequency hearing [13], account for the functional convergence in key parameters of prestin [14]. This shows that searching for molecular parallelism in candidate genes provided insights into the convergent adaptations to high frequency hearing; however, the mechanisms underlying the convergent vocalization aspect of echolocation remain unknown.

The increasing number of mammalian genomes provides an unprecedented opportunity to systematically search for parallel molecular evolution and find new associations to convergent phenotypic adaptations. To this end, we

developed an approach to systematically screen one-to-one orthologous proteins for parallel amino acid substitutions in independent lineages. Our approach 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 discovered an unusually high number of parallel substitutions between these echolocators in proteins crucial for fast muscle activity, suggesting that these proteins are involved in the evolution of superfast muscles that power the terminal buzz. We show that these proteins are expressed in superfast muscles and demonstrate functional convergence of the calcium storage protein Casq1 that is concordant with adaptations to superfast muscle physiology. Therefore, our study provides insights into the mechanisms involved in the convergent evolution of vocalization in echolocating mammals and provides a general framework for using computational approaches to discover the genomic basis of novel phenotypes.

Results

To systematically search for parallel molecular evolution, we developed a computational method that identifies parallel amino acid substitutions occurring in pairs of independent lineages (Figure S1). First, we used the phylogeny-aware Prank method [15] to compute a multiple alignment of a protein. 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, the method extracts those pairs in which one or more parallel amino acid substitutions have occurred (Figures 1A, S2). The result is a list of lineage pairs that share at least one parallel substitution.

We used this approach to conduct a genome-wide screen of 14,406 genes across 30 placental mammals (Figure S3). As expected, we found that the vast majority (~90%) of the parallel substitutions detected in our screen are conservative (similar physicochemical properties of amino acids) or occur at alignment positions that are poorly conserved in evolution (Table S1), consistent with previous results [16]. Since such substitutions are less likely to affect protein function, we filtered for proteins with at least one parallel substitution that occurs in a conserved alignment position and is radical (changes the physicochemical group, Table S2, Figure S2).

As a proof of concept, we first tested whether our method could detect known parallel substitutions. To this end, we extracted all proteins from our data that contain radical parallel substitutions in conserved positions between the echolocating microbat *Myotis lucifugus* (little brown bat) and the bottlenose dolphin (*Tursiops truncatus*). With Maximum Likelihood and Bayesian 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 (Figure S4A) that have been identified in previous candidate gene studies [4,9-11]. Importantly, our datasets also included three additional hearing-related proteins that have not been associated with the evolution of echolocation: stereocilin (Strc), tectorin alpha (Tecta) and calcium binding protein 2 (Cabp2) (Figure S4B). The function of Strc is to connect stereocilia to each other as well as to the tectorial membrane [17], which contains Tecta as a major component [18]. Cabp2 plays a role in regulating the activity of voltage-gated calcium channels that open in response to stereocilia deflection [19]. Thus, our method discovered additional hearing-related proteins that could contribute to convergent adaptations for high-frequency hearing in echolocating mammals.

Next, we searched for functional enrichments in these sets of 409/416 proteins to explore the functions of the remaining non-hearing related proteins. Surprisingly, we found that the proteins with parallel substitutions between microbat and dolphin are enriched for gene ontology terms relating to the contractile machinery of skeletal muscle fibers (Tables S3, S4). These proteins include key components of the contractile machinery expressed specifically in fast-twitch skeletal muscle fibers [20,21]: calsequestrin 1 (Casq1), ATPase sarcoplasmic/endoplasmic reticulum calcium transporting 1 (Atp2a1), myosin heavy chain 2 (Myh2) and myosin light chain (Myl1) (Figures 1 and 2). Casq1 adsorbs large amounts of Ca^{2+} ions in the sarcoplasmic reticulum (SR) [22-24]. Atp2a1 pumps Ca^{2+} into the SR [25]. Both Myh2 and Myl1 are components of the force-generating molecular motor complex [20,26].

We observed a total of seven radical parallel substitutions in conserved positions between microbat and dolphin in these four proteins (Casq1, Atp2a1, Myh2, Myl1). Only one of these seven substitutions (K49Q in Myl1) could potentially be explained by GC-biased gene conversion (a process which biases G/C over A/T alleles during recombination repair [27]), suggesting that this mutational process is not a major contributing factor (Table S5). Since the extremely rapid echolocation calls during the terminal buzz are powered by unique, sound-producing “superfast” muscles that consist of specialized fast-twitch fibers [28,29], we hypothesized that these proteins could

be involved in molecular optimizations that convert fast fibers into superfast fibers, and thus underlie the convergent vocalization aspect of echolocation.

If this hypothesis is true, we would expect that microbat and dolphin exhibit more parallel substitutions in fast-twitch than in slow-twitch muscle proteins, as superfast muscles are composed of specialized fast-twitch fibers [29]. Indeed, compared to the seven parallel substitutions in fast-twitch fiber proteins, we did not detect any parallel substitution in Ca^{2+} signaling and sarcomeric slow-twitch fiber proteins (Figure 1B). As our method detects parallel substitutions between any pair of lineages, we utilized it to evaluate whether this difference of seven vs. zero parallel substitutions in fast- vs. slow-twitch fiber proteins is unusually high compared to all other branch pairs in the 30-mammal phylogeny. Indeed, for the 1,295 other branch pairs, we observed a maximum difference of four (Table S6), showing that the seven parallel substitutions that the microbat and dolphin pair has in excess within the fast-twitch fiber set is an outlier observation (Figure 1B). Finally, a test that permutated alignment columns of the fast- and slow-twitch proteins estimates a low probability of 0.031 that any branch pair exhibits a difference of seven or more parallel substitutions (Figure S5). 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 and higher than the random expectation, consistent with the hypothesis that these proteins are involved in adaptations related to superfast muscle physiology.

If the four fast-twitch fiber proteins (Casq1, Atp2a1, Myh2, and Myl1) are involved in adaptations related to superfast muscles that power the terminal buzz, these proteins should be expressed in the specific sound-producing 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-seq, we found that all four proteins were expressed in this superfast sound-producing muscle [28] (Figure 3). 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 (Figure 3). We further compared expression of the three fast-twitch fiber myosin heavy chains (Myh1/2/4) that differ in their fiber shortening velocity, which decreases in the order: Myh4 > Myh1 > Myh2 [26]. 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. Interestingly, *Myh2* is the second most abundant myosin heavy chain (9%) in anterior cricothyroid muscle despite 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, which is in line with adaptations for the superfast muscles that power the terminal buzz.

If the proteins exhibiting parallel substitutions are involved in adaptations that convert fast fibers into superfast fibers, we 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 Ca^{2+} transport from the sarcoplasm into the SR [29,30], where it is adsorbed to Casq1. The adsorption of Ca^{2+} depends on the ability of Casq1 to form dimers and tetramers in a Ca^{2+}-dependent manner, resulting in polymers that adsorb large amounts of Ca^{2+} on the negatively charged surface and between the interfaces [23,31]. Since the superfast contraction/relaxation cycles are too short to pump all of the released Ca^{2+} back into the SR, the Ca^{2+} concentration in the SR decreases despite less Ca^{2+} being released into the sarcoplasm in subsequent cycles [32]. Therefore, we expect that Ca^{2+}-dependent polymerization of Casq1 in microbat and dolphin is adapted to the lower concentrations in the SR of repeatedly-stimulated superfast muscles.

To explore whether the parallel substitutions in Casq1 could affect Ca^{2+}-dependent polymerization, we examined their positions in the three-dimensional structure (Figure 2). We found that the E35Q substitution is located at the N-terminus, which participates in an arm exchange to stabilize dimer formation [31,33]. 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 of the high-affinity Ca^{2+} binding residues (D114, Figure 2), which interacts with the tetrameric partner. This suggests that all three radical substitutions could affect Ca^{2+}-dependent polymerization, which is crucial for the high capacity Ca^{2+} storage function of Casq1.

To experimentally investigate whether microbat and dolphin Casq1 polymerization are adapted to lower Ca^{2+} concentrations, we first conducted turbidity assays to compare Ca^{2+} concentration-dependent polymerization of Casq1 from mouse, microbat and dolphin (Figure 4A). For both microbat and dolphin Casq1, turbidity of the solution started to increase significantly at a Ca^{2+} concentration between 1.5 mM and 2.0 mM, respectively, and plateaued

at ~3 mM Ca^{2+}. This is in contrast to Casq1 from mouse where substantially higher Ca^{2+} concentrations are required to increase turbidity significantly. Second, to directly monitor Ca^{2+}-dependent dimerization at low Ca^{2+} concentrations, we used a multiangle light scattering experiment, which confirmed that microbat and dolphin Casq1 dimerizes at lower Ca^{2+} concentrations compared to mouse Casq1 (Figure 4B). In summary, these results demonstrate that microbat and dolphin Casq1 are able to form polymers at lower Ca^{2+} concentrations, which could facilitate 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, which powers the terminal buzz in bats. Superfast muscles consist of specialized fibers capable of contracting and relaxing at a rate that is an order of magnitude higher compared to the fastest locomotor muscles [30]. 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 Ca^{2+} transients and (ii) a higher fiber shortening velocity due to myosin motors with a fast cross-bridge detachment rate [30,34,35], suggesting evolutionary tinkering with various components of the Ca^{2+} signaling and molecular motor machinery. The proteins that our genomic screen detected could be involved in the evolution of superfast muscles by contributing to both rapid Ca^{2+} transients and high shortening velocities.

Rapid Ca^{2+} transients require a fast decrease of the Ca^{2+} concentration in the sarcoplasm, which is achieved through a high amount of Atp2a1 and parvalbumin, a protein that temporarily binds Ca^{2+} in the sarcoplasm until it can be pumped back into the SR by Atp2a1 [30]. As expected, we observed a high expression level of *Atp2a1* and parvalbumin in the bat anterior cricothyroid muscle (Figure 3, Table S7). In addition, our experiments demonstrate that both microbat and dolphin Casq1 are able to form polymers at lower Ca^{2+} concentrations, while retaining a maximum binding capacity that is similar or possibly higher to Casq1 of other species. This functional convergence suggests that microbat and dolphin Casq1 are adapted to adsorbing Ca^{2+} under conditions of lower Ca^{2+} concentrations in the SR. Such conditions are likely present in superfast muscles, where the

contraction/relaxation cycles are too short to fully restore the basal Ca^{2+} concentration in the SR. Therefore, the ability of Casq1 to adsorb Ca^{2+} under lower concentrations could be a novel mechanism that contributes to the rapid calcium transients required for superfast muscle physiology.

Besides rapid 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 [30,35]. Fiber shortening velocity is determined by the myosin relay domain [36,37]. Among the parallel substitutions in Myh2, the T512E substitution is located in the relay domain (Figure 2) 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. 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^{2+} release/sequestration cycle. The myosin isoform that is exclusively expressed in flight muscles has an extremely fast cross-bridge detachment rate [35] and, as shown in Figure S6, 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 for 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 [36]. Second, mammalian Myh4 myosin, which confers the highest muscle fiber shortening velocity [26], also exhibits a negatively-charged residue (E) at position 512 (Figure S7). 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 (Figure S7). Thus, it appears that Myh2 in microbat and dolphin have converged towards a higher similarity with Myh4. In addition, genome alignments and analysis of sequencing reads show that cetaceans have lost *Myh4* (Figures S8, S9), raising the possibility that sequence convergence in the dolphin *Myh2* compensates for the loss of *Myh4*, the gene encoding the fastest muscle myosin heavy chain.

Our study highlights the utility of comparative genomics approaches to provide insights into the genomic basis of complex phenotypes. The rapidly increasing number of sequenced genomes will provide new opportunities to extend candidate gene studies and systematically search for parallel

molecular evolution. The detection of novel patterns of molecular parallelism has the potential to uncover new associations to convergent phenotypes and ultimately, provide insights into the mechanisms underlying convergent phenotypic adaptations.

Materials and Methods

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

Figure S1 gives an overview of the steps. We used Ensembl Compara (release 75) [38] 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 80-120% relative to the human or mouse sequence. All ortholog sets that contained less than 15 sequences were discarded. This resulted in 14,406 sets.

In order to reconstruct ancestral protein sequences, we first aligned each ortholog set using Prank (version 10603, parameters: +F) with the mammalian phylogeny (Figure S3) as a fixed guide tree [39] since the phylogeny-aware Prank alignment method was reported to consistently outperform other methods [40,41]. 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) [42]. This alignment trimming was shown to improve the accuracy in phylogenetic reconstructions [42].

Third, we estimated ancestral sequences using either Maximum Likelihood and 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) [43,44], 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 a wrapper for the codeML functions in PAML [45,46], 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 [47] and allowed for rate heterogeneity (RAxML parameter “–m PROTCATLG”, Lazarus parameter “fixed_asrv=False, model=LG”). For the Bayesian approach, we used PhyloBayes (version 4.1b) [48], constraining the tree topology by the mammalian phylogeny (Figure S3) and allowing PhyloBayes to estimate branch lengths using the default CAT substitution model [49]. Two independent Monte Carlo Markov chains were then run, sampling every 100 steps until both chains converged. Based on 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. Since 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 out 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 (Table S2). 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 (Figure S2). 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. All primary data (alignments, parallel substitutions and their classification) are available at https://bds.mpi-cbg.de/hillerlab/ParallelEvolution/.

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

Fast- and slow-twitch muscle fiber proteins were taken from Dataset S1 in [21], which used gene expression data from isolated mouse fast-twitch and slow-twitch muscle fibers to obtain sarcomeric and Ca^{2+} signaling genes with a higher expression level in fast-twitch (slow-twitch) fibers, and from [20], 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.

To estimate the probability that any pair of branches would exhibit an excess of seven or more parallel substitutions in fast- vs. slow-twitch fiber proteins, we used a permutation procedure. We randomly permutated the alignment columns of the fast- and slow-twitch proteins such that the number of alignment columns is identical to the real data sets (14,988 vs. 14,984). Then, using these two permutated sets, we determined for each pair of independent

branches the difference between the number of parallel substitutions observed in the fast-twitch and slow-twitch fiber sets. Only radical parallel amino acid substitutions in conserved positions were considered. This procedure was repeated 100,000 times with different random seeds and the maximum difference observed between any branch pair (including the microbat – dolphin pair) was recorded. We estimated a probability of 0.031 (Maximum Likelihood: 3,092 of 100,000; Bayesian: 3,142 of 100,000 iterations) for observing a maximum difference of at least seven, which is the difference observed in the real data for microbat and dolphin (Figure S5).

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 *Pteronotus parnellii* bats, kept in a colony in our animal house. The bats were euthanized through an overdose of Narcoren (Merial GmbH, solution contained 400 mg/kg pentobarbital) injected intraperitoneally. The anterior cricothyroid muscle was anatomically defined according to [50]. 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 Scientific) and approximately 15 µg of RNA were found in each sample. RNA integrity was assessed using a Bioanalyzer 2100 (Agilent Technologies) and samples with RNA integrity number (RIN, a measurement of RNA degradation [51]) > 7 were selected for RNA sequencing. An Illumina HiSeq 2500 was used to generate 75 bp paired-end reads with insert size of approximately 250 bp.

***De novo* assembly of transcriptome**

Prior to transcriptome assembly, adapters and low-quality reads (minimum read length: 35 bp) were removed using Cutadapt (http://code.google.com/p/cutadapt/). Because the available *P. parnellii* genome [4] is highly fragmented, we used the Trinity (v 2.0.6) de novo assembly package [52] to assemble the *P. parnellii* transcriptome. Highly similar transcripts were collapsed using CD-Hit [53]. Annotation of the transcriptome was carried out using Blast+ against human cDNA 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 UCSC genome browser's Blat [54,55] 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*, *Myl1* and their paralogs align with >90% identity to the orthologous gene according to the Ensembl and CESAR annotation [56,57], 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 [58]. Consequently, the introns between the *Myh2* ortholog of both species evolved neutrally for a much shorter amount of time and thus still align (Figure S10A). 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 [59] 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 (Figure S10B). 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 *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 (v 2.2.0) [60]. Subsequently, the gene expression level based on the aligned reads was estimated using RSEM (v 1.2.31) [61]. Normalization across samples was applied using the trimmed means of M-values [62]. Differential expression analysis was performed using the DESeq2 package [63].

Functional experiments of mouse (*Mus musculus)*, dolphin (*Tursiops truncates)* and microbat (*Myotis lucifugus) calsequestrin 1*

<u>*Protein Expression and Purification*</u> – Rosetta(DE3)pLysS *E. coli* cells containing the Casq1 vectors from were grown in LB media at 37 °C until an OD_{600} 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-1-thiogalactopyranoside (IPTG) and left to grow overnight. Following induction, the cells were harvested by centrifugation at 4,000g for 30 mins and then froze at -20°C for later use.

For protein purification, cell pellets were suspended in 20 mM tris, 0.5 g/L NaN_3, pH 7.5 and sonicated at 8000 rpm using a 450 Sonifier® (Branson Ultrasonics), until they reached apparent homogeneity. The resulting lysate was clarified by centrifugation at 20,000 × *g* and loaded onto a GE Healthcare AKTA^{TM} pure FPLC with an attached Toyopearl DEAE-650M (Tosoh Biosciences) column equilibrated with DEAE loading buffer (20 mM tris, 0.5 g/L NaN_3, pH 7.5). Casq1 was eluted from the DEAE column between 12.5% to 25% buffer B, by running a linear gradient of Buffer A (20 mM tris, 0.5 g/L NaN_3, pH 7.5) to buffer B (20 mM tris, 0.5 g/L NaN_3, 2M NaCl, pH 7.5). The fractions containing Casq1 were buffer-exchanged to buffer C (5 mM sodium phosphate, 0.5 g/L NaN_3, 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 pre-equilibrated with buffer C. Casq1 eluted from the HA column between 50 to 100 % buffer D, by running a linear gradient of buffer C to buffer D (0.5 M sodium phosphate, 0.5 g/L NaN_3, pH 6.8). For the final purification step, the HA Casq1 elution was buffer-exchanged to buffer E (20 mM tris, 0.5 g/L NaN_3, pH 8.5) and loaded onto a pre-equilibrated Mono Q^{TM} column (GE Healthcare). Casq1 was eluted form the Mono Q^{TM} column by running a linear gradient from buffer E to buffer F (20 mM tris, 0.5 g/L NaN_3, 2M 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/L, pH 7.2). Throughout the purification process fractions containing

Casq1 were identified with SDS-PAGE. Protein concentrations were determined using the bicinchoninic acid (BCA) assay (Thermo Scientific).

Molecular mass determination by multiangle light scattering

We injected 100 µl of each protein at 1 mg/ml concentration, which was pre-equilibrated with chromatography running buffer (20 mM Tris-HCl pH 7.5, 300 KCl and either with or without 1 mM $CaCl_2$), onto a Bio-Sep S-2000 column (Phenomenex). The chromatography was carried out at a flow rate of 0.5 ml/min using Acuflow series IV pump (Analytical Scientific Inst.). The eluate was passed in tandem through an UV detector (GILSON), a refractometer (Optilab DSP, Wyatt Tech.) and a multiangle laser light scattering detector (Dawn EOS, Wyatt Tech.). All the chromatography experiments were performed at 25 °C. Scattering data were analyzed using software, ASTRA (Wyatt Tech.) 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 $4\pi^2n^2(\mathrm{d}n/\mathrm{d}c)^2/\lambda_0^4N_A$ (where n = solvent refractive index, $\mathrm{d}n/\mathrm{d}c$ = refractive index increment of scattering sample, λ_0 = wavelength of scattered light and N_A = 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 (i.e., absorbance at 350 nm) as a function of Ca^{2+} concentration was monitored using a Genesys 10S UV-Vis Spectrophotometer (Thermo Scientific). Assays were performed using 1.3 mL of 15 µM mouse, bat, and dolphin Casq1 in the assay buffer. Concentrated Ca^{2+} solutions (0.10 M, 0.25 M, 0.50 M, and 1.0 M) were added in 1.0 µL to 2.0 µL aliquots to the 1.3 mL Casq1 solutions in a quartz cuvette to achieve the proper calcium concentration. Upon addition of each concentrated Ca^{2+} aliquot, the samples were mixed by stirring with a small stir bar and allowed to equilibrate (dA350/dt = 0) before addition of the next aliquot. Dilutions from adding Ca^{2+} aliquots were included in data analysis.

Acknowledgment

We are grateful to Holger Brandl for help with transcriptome analysis, Julia Jarrells and Andreas Dahl for RNA-sequencing, and Jochen Rink for helpful discussions. We thank the Computer Service Facilities of the MPI-CBG and MPI-PKS, and the Scientific Computing Facility and Gene Expression Facility of the MPI-CBG for their support. This work was supported by National Institutes of Health Grant 1R01GM11125401 and the M. J. Murdock Charitable Trust to CHK, and the Max Planck Society to MH.

References

1 Wray, G. A. Genomics and the Evolution of Phenotypic Traits. *Annual Review of Ecology, Evolution, and Systematics* **44**, 51-72 (2013).

2 Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61, doi:10.1038/nature10944 (2012).

3 Hiller, M. *et al.* A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep* **2**, 817-823, doi:10.1016/j.celrep.2012.08.032 (2012).

4 Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228-231, doi:10.1038/nature12511 (2013).

5 Levin, L. & Mishmar, D. The genomic landscape of evolutionary convergence in mammals, birds and reptiles. *Nat Ecol Evol* **1**, 41, doi:10.1038/s41559-016-0041 (2017).

6 Kalko, E. K. V. & Schnitzler, H. U. The echolocation and hunting behavior of Daubenton's bat, Myotis daubentoni. *Behavioral Ecology and Sociobiology* **24**, 225–238 (1989).

7 Surlykke, A. & Moss, C. F. Echolocation behavior of big brown bats, Eptesicus fuscus, in the field and the laboratory. *The Journal of the Acoustical Society of America* **108**, 2419-2429 (2000).

8 Wisniewska, D. M., Johnson, M., Nachtigall, P. E. & Madsen, P. T. Buzzing during biosonar-based interception of prey in the delphinids Tursiops truncatus and Pseudorca crassidens. *The Journal of experimental biology* **217**, 4279-4282, doi:10.1242/jeb.113415 (2014).

9 Li, Y., Liu, Z., Shi, P. & Zhang, J. The hearing gene Prestin unites echolocating bats and whales. *Current biology : CB* **20**, R55-56, doi:10.1016/j.cub.2009.11.042 (2010).

10 Liu, Y., Rossiter, S. J., Han, X., Cotton, J. A. & Zhang, S. Cetaceans on a molecular fast track to ultrasonic hearing. *Current biology : CB* **20**, 1834-1839, doi:10.1016/j.cub.2010.09.008 (2010).

11 Davies, K. T., Cotton, J. A., Kirwan, J. D., Teeling, E. C. & Rossiter, S. J. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. *Heredity (Edinb)* **108**, 480-489, doi:10.1038/hdy.2011.119 (2012).

12 Shen, Y. Y., Liang, L., Li, G. S., Murphy, R. W. & Zhang, Y. P. Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLoS Genet* **8**, e1002788, doi:10.1371/journal.pgen.1002788 (2012).
13 Zheng, J. *et al.* Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**, 149-155, doi:10.1038/35012009 (2000).
14 Liu, Z., Qi, F. Y., Zhou, X., Ren, H. Q. & Shi, P. Parallel sites implicate functional convergence of the hearing gene prestin among echolocating mammals. *Molecular biology and evolution* **31**, 2415-2424, doi:10.1093/molbev/msu194 (2014).
15 Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635, doi:10.1126/science.1158395 (2008).
16 Rokas, A. & Carroll, S. B. Frequent and widespread parallel evolution of protein sequences. *Molecular biology and evolution* **25**, 1943-1953, doi:10.1093/molbev/msn143 (2008).
17 Verpy, E. *et al.* Stereocilin connects outer hair cell stereocilia to one another and to the tectorial membrane. *The Journal of comparative neurology* **519**, 194-210, doi:10.1002/cne.22509 (2011).
18 Legan, P. K. *et al.* 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).
19 Schrauwen, I. *et al.* A mutation in CABP2, expressed in cochlear hair cells, causes autosomal-recessive hearing impairment. *Am J Hum Genet* **91**, 636-645, doi:10.1016/j.ajhg.2012.08.018 (2012).
20 Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiological reviews* **91**, 1447-1531, doi:10.1152/physrev.00031.2010 (2011).
21 Chemello, F. *et al.* Microgenomic analysis in skeletal muscle: expression signatures of individual fast and slow myofibers. *PloS one* **6**, e16807, doi:10.1371/journal.pone.0016807 (2011).
22 MacLennan, D. H. & Wong, P. T. Isolation of a calcium-sequestering protein from sarcoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 1231-1235 (1971).
23 Park, H. *et al.* 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, doi:10.1074/jbc.M311553200 (2004).
24 Sanchez, E. J., Lewis, K. M., Danna, B. R. & Kang, C. High-capacity Ca2+ binding of human skeletal calsequestrin. *J Biol Chem* **287**, 11592-11601, doi:10.1074/jbc.M111.335075 (2012).
25 MacLennan, D. H. Purification and properties of an adenosine triphosphatase from sarcoplasmic reticulum. *J Biol Chem* **245**, 4508-4518 (1970).
26 Bottinelli, R., Betto, R., Schiaffino, S. & Reggiani, C. 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).
27 Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* **10**, 285-311, doi:10.1146/annurev-genom-082908-150001 (2009).

28 Elemans, C. P., Mead, A. F., Jakobsen, L. & Ratcliffe, J. M. Superfast muscles set maximum call rate in echolocating bats. *Science* **333**, 1885-1888, doi:10.1126/science.1207309 (2011).

29 Rome, L. C. Design and function of superfast muscles: new insights into the physiology of skeletal muscle. *Annual review of physiology* **68**, 193-221, doi:10.1146/annurev.physiol.68.040104.105418 (2006).

30 Rome, L. C., Syme, D. A., Hollingworth, S., Lindstedt, S. L. & Baylor, S. M. The whistle and the rattle: the design of sound producing muscles. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8095-8100 (1996).

31 Wang, S. *et al.* Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nature structural biology* **5**, 476-483 (1998).

32 Harwood, C. L. *et al.* Paying the piper: the cost of Ca2+ pumping during the mating call of toadfish. *The Journal of physiology* **589**, 5467-5484, doi:10.1113/jphysiol.2011.211979 (2011).

33 Park, H., Wu, S., Dunker, A. K. & Kang, C. Polymerization of calsequestrin. Implications for Ca2+ regulation. *J Biol Chem* **278**, 16176-16182, doi:10.1074/jbc.M300120200 (2003).

34 Rome, L. C. *et al.* Trading force for speed: why superfast crossbridge kinetics leads to superlow forces. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5826-5831 (1999).

35 Swank, D. M., Vishnudas, V. K. & Maughan, D. W. An exceptionally fast actomyosin reaction powers insect flight muscle. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17543-17547, doi:10.1073/pnas.0604972103 (2006).

36 Bloemink, M. J. *et al.* Alternative exon 9-encoded relay domains affect more than one communication pathway in the Drosophila myosin head. *Journal of molecular biology* **389**, 707-721, doi:10.1016/j.jmb.2009.04.036 (2009).

37 Kronert, W. A., Melkani, G. C., Melkani, A. & Bernstein, S. I. Alternative relay and converter domains tune native muscle myosin isoform function in Drosophila. *Journal of molecular biology* **416**, 543-557, doi:10.1016/j.jmb.2011.12.044 (2012).

38 Vilella, A. J. *et al.* EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335, doi:10.1101/gr.073585.107 (2009).

39 Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods in molecular biology* **1079**, 155-170, doi:10.1007/978-1-62703-646-7_10 (2014).

40 Fletcher, W. & Yang, Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular biology and evolution* **27**, 2257-2267, doi:10.1093/molbev/msq115 (2010).

41 Jordan, G. & Goldman, N. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular biology and evolution* **29**, 1125-1139, doi:10.1093/molbev/msr272 (2012).

42 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

43 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690, doi:10.1093/bioinformatics/btl446 (2006).

44 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
45 Hanson-Smith, V., Kolaczkowski, B. & Thornton, J. W. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular biology and evolution* **27**, 1988-1999, doi:10.1093/molbev/msq081 (2010).
46 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
47 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Molecular biology and evolution* **25**, 1307-1320, doi:10.1093/molbev/msn067 (2008).
48 Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286-2288, doi:10.1093/bioinformatics/btp368 (2009).
49 Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular biology and evolution* **21**, 1095-1109, doi:10.1093/molbev/msh112 (2004).
50 Griffiths, T. A. Comparative laryngeal anatomy of the Big Brown bat, Eptesicus fuscus, and the Mustached bat, Pteronotus parnellii. *Mammalia* **47**, 377-394 (1983).
51 Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3, doi:10.1186/1471-2199-7-3 (2006).
52 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).
53 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).
54 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664, doi:10.1101/gr.229202. Article published online before March 2002 (2002).
55 Tyner, C. *et al.* The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res*, doi:10.1093/nar/gkw1134 (2016).
56 Sharma, V., Elghafari, A. & Hiller, M. Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation. *Nucleic Acids Res* **44**, e103, doi:10.1093/nar/gkw210 (2016).
57 Aken, B. L. *et al.* The Ensembl gene annotation system. *Database : the journal of biological databases and curation* **2016**, doi:10.1093/database/baw093 (2016).
58 Agnarsson, I., Zambrana-Torrelio, C. M., Flores-Saldana, N. P. & May-Collado, L. J. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Curr* **3**, RRN1212, doi:10.1371/currents.RRN1212 (2011).
59 Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11484-11489, doi:10.1073/pnas.1932072100 (2003).
60 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

61 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
62 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, doi:10.1186/gb-2010-11-3-r25 (2010).
63 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
64 Dong, X. & Thomas, D. D. Time-resolved FRET reveals the structural mechanism of SERCA-PLB regulation. *Biochemical and biophysical research communications* **449**, 196-201, doi:10.1016/j.bbrc.2014.04.166 (2014).
65 Houdusse, A., Szent-Gyorgyi, A. G. & Cohen, C. Three conformational states of scallop myosin S1. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11238-11243, doi:10.1073/pnas.200376897 (2000).
66 Koppole, S., Smith, J. C. & Fischer, S. The structural coupling between ATPase activation and recovery stroke in the myosin II motor. *Structure* **15**, 825-837, doi:10.1016/j.str.2007.06.008 (2007).

Figures

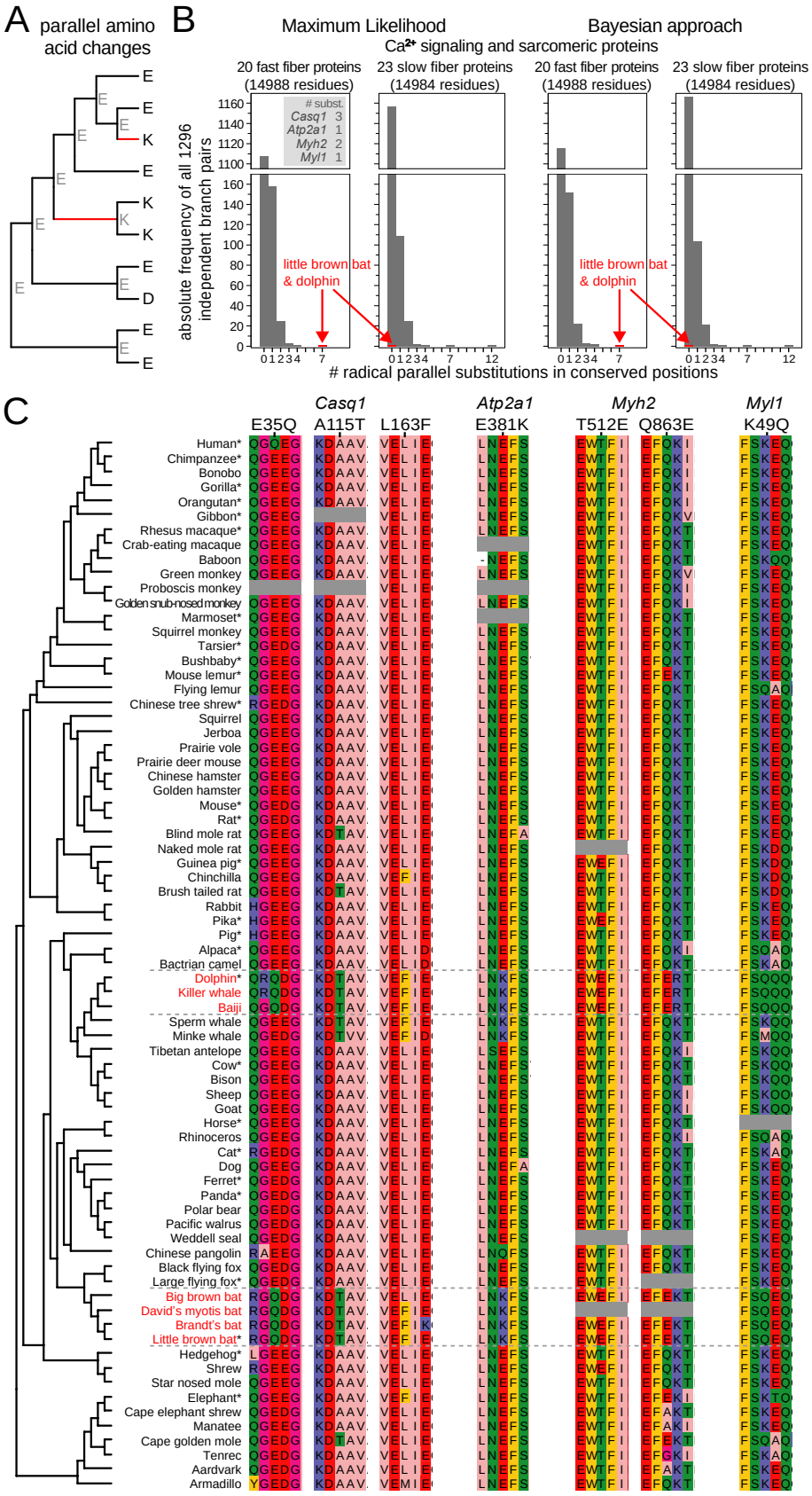

Figure 1: Parallel molecular evolution in fast-twitch muscle fiber proteins in echolocating mammals.

(A) Illustration of the screen for parallel amino acid substitutions, defined as the same amino acid substitution from an identical ancestral state. Inferred ancestral amino acids are in grey font. Red branches have a parallel substitution from E (Glu) to K (Lys) at this alignment position.
(B) Histograms showing the number of radical parallel amino acid substitutions in conserved positions of fast-twitch vs. slow-twitch Ca^{2+} signaling and sarcomeric proteins for all 1,296 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.
(C) Alignments showing radical parallel substitutions in conserved positions. 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. Grey box: missing sequence.

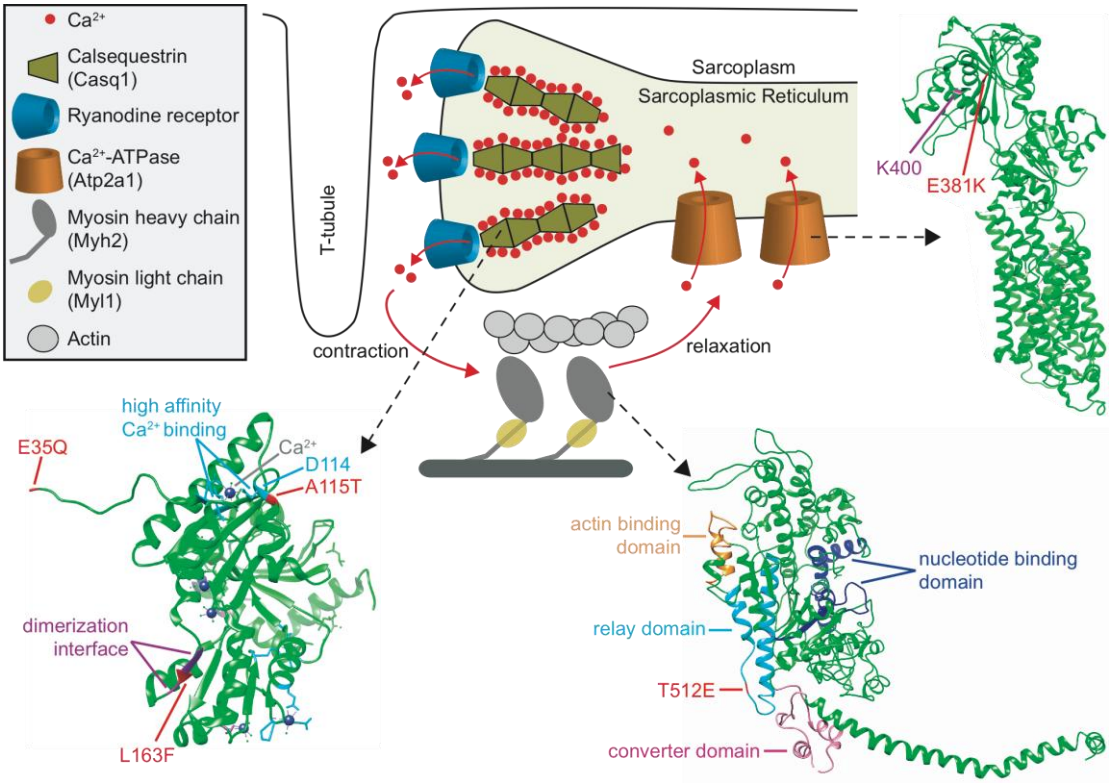

Figure 2: Simplified diagram showing the skeletal muscle contraction-relaxation cycle and 3D structures of proteins with parallel substitutions.
Ca^{2+} bound to Casq1 is released from the sarcoplasmic reticulum through ryanodine receptors into the sarcoplasm to initiate the cross-bridge cycle during which myosin complexes generate mechanical work. For relaxation, Ca^{2+} is actively pumped back into the sarcoplasmic reticulum by Atp2a1, where majority of Ca^{2+} ions are bound again to polymerized Casq1. The 3D protein structures of Casq1 (PDB: 3TRP; residues that bind 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 Ca^{2+} binding [24] and the Atp2a1 residue K400 that binds the inhibitor phospholamban [64] are labeled. The Myh2 relay domain that determines the actin sliding velocity [37] by communicating conformational changes between the converter domain, nucleotide-binding site, and actin-binding site [36,65,66] is indicated. K49Q is not part of any available Myl1 structure.

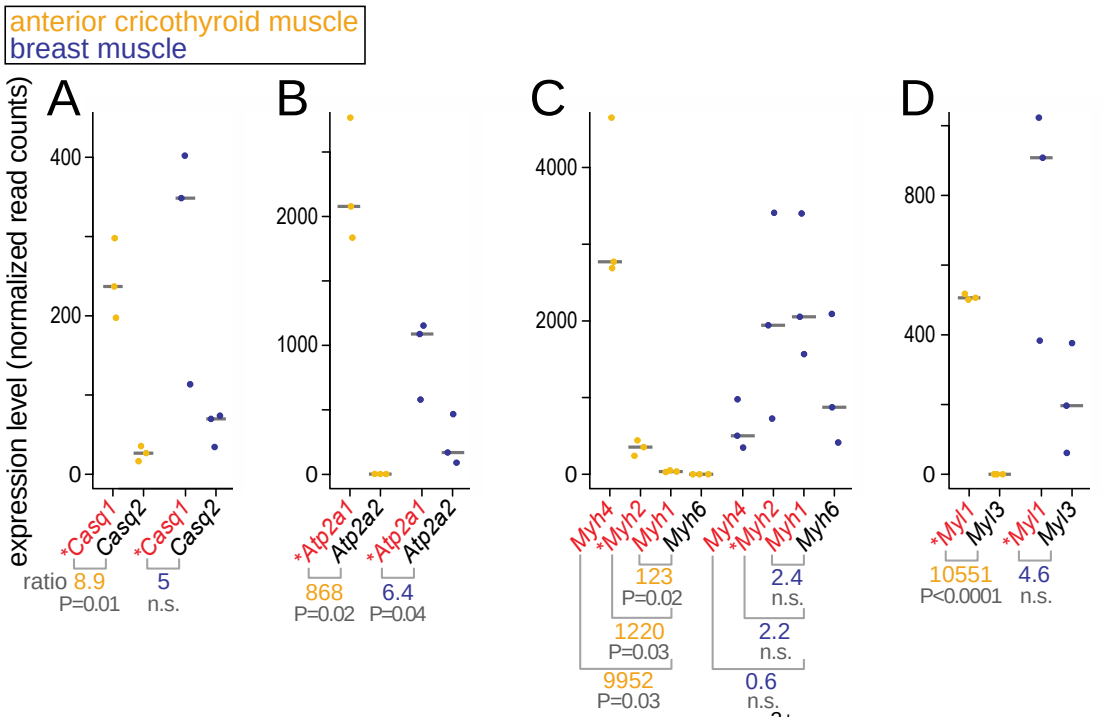

Figure 3: Expression level of calsequestrin (A), Ca^{2+} ATPase (B), myosin heavy chain (C) and myosin light chain (D) genes comparing fast-twitch (red font) and slow-twitch (black font) muscle fiber components in three biological replicates of *P. parnellii* anterior cricothyroid muscle (yellow) and 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 S8. n.s., not significant.

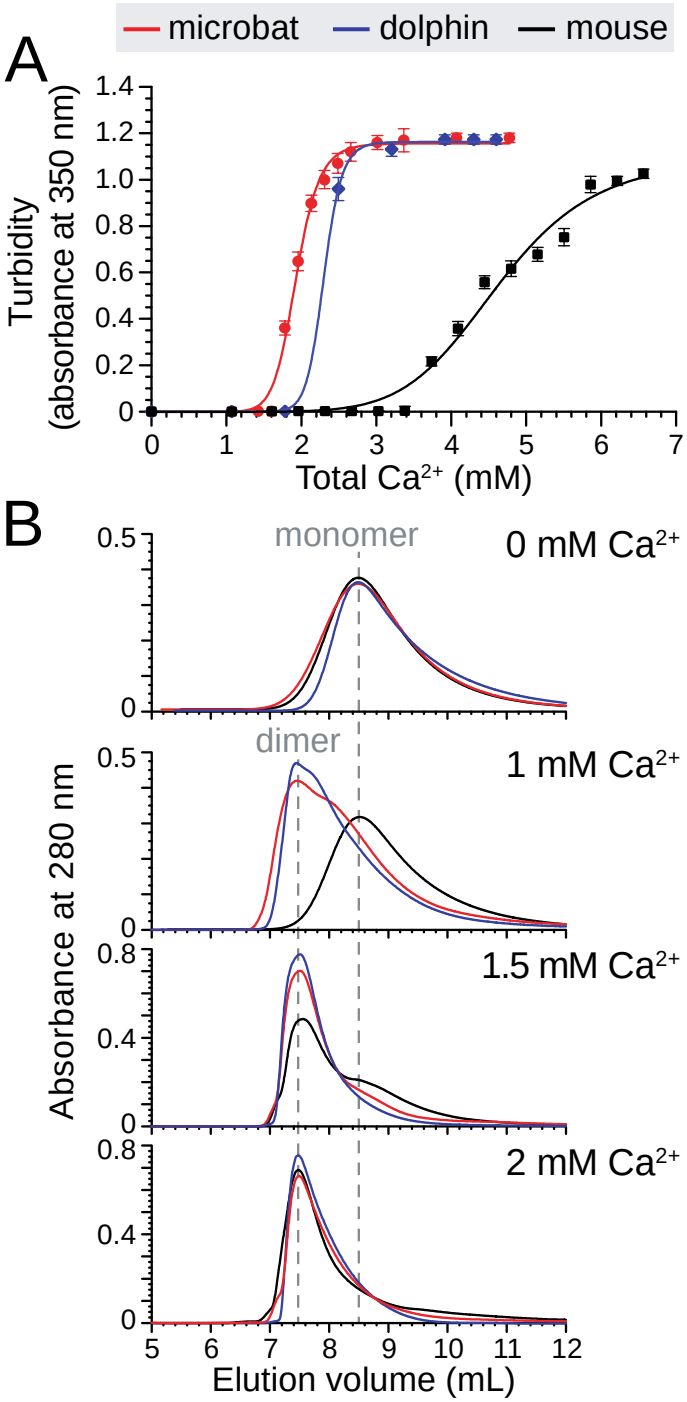

Figure 4: Microbat and dolphin Casq1 functionally converged in the ability to form Ca^{2+}-sequestering polymers at lower Ca^{2+} concentrations.

(A) Ca^{2+}-dependent turbidity shows that microbat and dolphin Casq1 oligomerizes at lower Ca^{2+} concentrations. Error bars represent standard deviations of triplicate measurements.

(B) Multiangle light scattering experiments show that microbat and dolphin Casq1 dimerizes at lower Ca^{2+} concentrations compared to mouse Casq1. While Casq1 from all three species is monomeric at a Ca^{2+} concentration of 0 mM, both microbat and dolphin Casq1 dimerizes at a concentration of 1 mM, in contrast to mouse Casq1 that remains monomeric.