Genome-Wide Screens for Molecular Convergent Evolution in Mammals

Jun-Hoe Lee and Michael Hiller

Abstract Convergent evolution can occur at both the phenotypic and molecular level. Of particular interest are cases of convergent molecular changes that underlie convergent phenotypic changes, as they highlight the genomic differences that underlie phenotypic adaptations and can inform us on why evolution has repeatedly chosen the same solution in lineages that have evolved independently. Many approaches to identify convergent molecular evolution have focused on candidate genes with known functions as well as lineages with known convergent phenotypes. The growing amount of genomic sequence data makes it now possible to systematically detect molecular convergence genome-wide. Here, we highlight the advantages and drawbacks of using genomic screens to identify molecular convergence. We present our method to detect convergent substitutions between any pair of lineages in a genome-wide manner, ways of enriching for convergence that are more likely to affect protein function, and present novel cases of convergence in echolocating mammals. Our results suggest that genomic screens have the potential to generate new hypotheses of associations between molecular convergence and phenotypic convergence. Together with experimental assays to test for functional convergence, this will contribute to revealing the genomic changes that underlie convergent phenotypic changes.

1 Convergent Molecular Evolution

Convergent evolution, in the most basic sense, refers to the acquisition of similar traits in independent lineages. Some well-known examples include the wings that birds and bats use for powered flight, the highly streamlined body form of dolphins and fish that allows for efficient movement in an aquatic environment, or adapta-

J.-H. Lee · M. Hiller
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

J.-H. Lee · M. Hiller (✉)
Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
e-mail: hiller@mpi-cbg.de

P. Pontarotti (ed.), *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*,
DOI 10.1007/978-3-319-61569-1_16

tions in ant-foraging mammals across different continents that have evolved powerful digging forelimbs and a long, sticky tongue. These examples of convergent phenotypic evolution document the power of natural selection to repeatedly result in highly similar adaptations that are extremely unlikely to arise by neutral evolution.

From a molecular perspective, convergence can also be observed at different levels of hierarchy, such as pathways, structures, and genes. For instance, many different plants including maize and sugarcane have convergently evolved the C_4 photosynthesis pathway (Williams et al. 2013). Similarly, many distantly related fish and insects have independently evolved antifreeze proteins that share similar structural attributes (Chen et al. 1997; Davies et al. 2002). Convergence can also occur in the same gene through different mutations that confer a similar functional change. This is illustrated by the higher oxygen affinity of hemoglobin in independent bird species that have adapted to high-altitude environments. A recent study compared 56 pairs of high- and low-altitude birds and found that amino acid substitutions at multiple sites can increase oxygen affinity, suggesting that there can be multiple solutions for the same problem (Natarajan et al. 2016). However, this study also revealed several independent high-altitude lineages, where higher oxygen affinity can be traced to identical amino acid substitutions. In the following, we focus on these particular cases of molecular convergence. Such cases where the same nucleotide or amino acid substitution occurs in independent lineages have sometimes been divided into parallel and convergent substitutions, depending on whether the inferred ancestral residues in both lineages are the same (parallel) or different (convergent) (Zhang and Kumar 1997). For simplicity, we refer here to both cases as convergent substitutions.

2 Convergent Molecular Evolution Can Underlie Convergent Phenotypic Evolution

A well-studied example of convergent molecular evolution that contributes to phenotypic convergence is prestin, a motor protein critical for high-frequency hearing in mammals (Dallos and Fakler 2002). High-frequency hearing is essential for echolocating bats and toothed whales as it allows them to detect small, moving prey in conditions with poor visibility. Given the convergence in high-frequency hearing between independent echolocating mammalian lineages, *Slc26a5* encoding prestin was a promising candidate gene to examine whether molecular convergence has occurred. Phylogenetic analysis of the prestin protein sequence clustered the echolocating lineages (dolphin and two independent bat lineages) together (Li et al. 2008, 2010). Thus, the topology of the phylogenetic tree inferred from the prestin protein differs from the generally accepted mammalian phylogeny and suggested molecular convergence between these lineages. Furthermore, a tree that clustered the echolocating mammals was also obtained using the first and second codon

positions that mainly determine the encoded protein sequence. In contrast, a tree computed from only third codon positions that are mostly synonymous has a topology consistent with the mammalian phylogeny (Li et al. 2010). A closer examination of the prestin sequence alignment identified several parallel substitutions in the echolocating mammals (Fig. 1). Subsequent in vitro experiments demonstrated that some of these parallel substitutions (N7T and I384T) alter the voltage-dependent properties of prestin, consistent with a functional change that contributes to hearing higher frequencies (Liu et al. 2014). Thus, the function of prestin in echolocating mammals exemplifies that convergent molecular evolution can be involved in convergent phenotypic changes.

Apart from high-frequency hearing, several other convergent phenotypes have been linked to convergent amino acid substitutions. For example, spectral tuning has been reported in the opsin of different cichlid populations, which enables these species to detect light from different wavelengths at different water depths (Nagai et al. 2011). Adaptations to a herbivorous diet in ruminants and leaf-eating monkeys have been linked to convergent mutations in lysozyme and RNAse1 (Stewart et al. 1987; Zhang 2006). Toxin resistance against cardenolide in insects, reptiles, amphibians, and mammals is associated with various convergent mutations in a sodium–potassium pump (Zhen et al. 2012; Dobler et al. 2012; Ujvari et al. 2015), while resistance against tetrodotoxin in various snakes is mediated by convergent mutations in a voltage-gated sodium channel (Geffeney et al. 2005; Feldman et al. 2012). Recently, it was suggested that convergent evolution in limb development genes is involved in the development of the pseudothumb in the giant and the red panda, two independent bamboo-eating lineages (Hu et al. 2017).

The various examples of convergent molecular evolution involving identical mutations raise the question of why evolution has repeatedly chosen the same

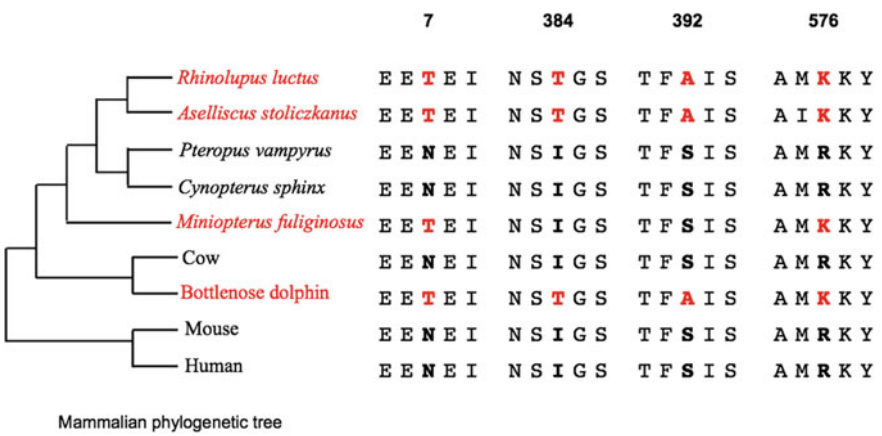

Fig. 1 Simplified mammalian phylogeny and selected regions of the prestin sequence alignment. The echolocating mammals as well as several of the sites showing convergence (N7T, I384T, S392A, and R576 K) are highlighted in *red* (Li et al. 2010). The N7T and I384T sites have been experimentally tested to affect prestin function (Liu et al. 2014)

solution. One possible explanation is that there could be very few amino acid changes that shift the functional activity of a protein in a specific direction (Christin et al. 2010). Consequently, the number of solutions can be very limited. For example, it appears that there are only three critical substitutions (S180A, Y277F, and T285A) in the primate M/LWS opsin that lead to specific color vision changes by causing additive (and reversible) spectral shifts of the wavelength of maximal absorbance (λ_{max}) by −7, −8, −15 nm, respectively (reviewed in Kawamura et al. 2012). In addition, the order of mutations can further constrain the evolutionary trajectory, as shown for the β-lactamase protein (Weinreich 2006). There are five amino acid mutations in β-lactamase leading to increased bacterial resistance. These five mutations could be derived from 120 possible evolutionary trajectories. However, many of these trajectories do not follow a continuous increase in bacterial resistance and are therefore less likely to be favored in evolution. This shows that the effect of mutations is not independent of other mutations. Such nonadditive effects of mutations on protein structure and function are called epistasis (Starr and Thornton 2016). Thus, even if several solutions exist, epistasis within a gene can restrict the possible evolutionary trajectories.

Convergent sequence evolution is particularly interesting as it has the potential to highlight the genomic changes that underlie phenotypic adaptations. Proteins with convergent changes represent candidates that can be experimentally tested for convergence in protein function. Furthermore, since one often lacks a good understanding how changes in sequence affect protein structure and function, it is difficult to narrow down the list of individual amino acid changes to be experimentally tested for their effect on function. The identification of convergent substitutions can provide a starting point for experiments to determine the specific sequence changes that contribute toward convergence in protein function.

3 Do the Convergent Substitutions Always Have the Same Effect?

Although functional experiments on selected convergent sites in prestin demonstrated a similar functional change in echolocating bats and toothed whales, these results cannot be generalized (Liu et al. 2014). In the case of RNAse1, the specific substitutions that modify the catalytic activity of the enzyme in ruminants can have a different effect if these substitutions are introduced into the orthologous RNAse1 of leaf-eating langurs (Zhang 2003). Specifically, even though the Q28L substitution increases the enzymatic activity in cows, it led to decreased enzymatic activity in the langur protein. This suggested that the same substitutions can have different and possibly even opposing effects in homologous proteins because of the different evolutionary history of individual proteins. Due to epistasis, a recent substitution could modify the rate of mutations at other sites to preferentially accommodate substitutions that yield a new stable conformation, a phenomenon that has been termed as Stokes shift (Pollock et al. 2012). Thus, the longer the divergence time

between two lineages, the less likely it is that the same mutation in a homologous protein has the same effect, which decreases the rate of convergence (Goldstein et al. 2015; Zou and Zhang 2015a). Notably, closely-related lineages can exhibit a high level of non-adaptive convergence, i.e., sequence similarity by chance (also known as background convergence), which decreases as divergence time increases. Therefore, caution must be taken when trying to infer adaptive convergence from identical substitutions observed in independent lineages.

To demonstrate that convergent substitutions in a protein are adaptive, Zhang (2006) has proposed 4 criteria: i. The convergent substitutions are observed in lineages that have evolved independently, ii. the convergent substitutions were driven by a common selective pressure, iii. there is convergence in protein function, and iv. the change in protein function can be clearly linked to the convergent substitution. With few exceptions, such as prestin in echolocating mammals, most of the studies that reported on associations between molecular convergence and potentially adaptive convergent phenotypes do not fulfill all criteria, in particular criteria iii and iv that require functional experiments. The fourth criterion is more difficult to satisfy as it requires introducing amino acid mutations in the orthologous protein in other species, followed by functional assays. Even if the convergent substitutions are important for convergence in protein function in the particular lineages, introducing the same mutations in the orthologous protein of another species that has a different evolutionary background might not result in a similar functional change, due to epistasis and/or Stokes shift. Alternatively, the convergent mutations could be introduced in the reconstructed ancestral version of the protein, using the ancestor that predates the adaptive phenotype. However, if several mutations have occurred on the branch descending from this ancestor, the convergent mutation might not be the first mutation that had occurred and experimental results might again depend on the background of other mutations. Thus, demonstrating that the change in protein function is caused directly by the convergent substitutions can be difficult if the effect of these mutations is influenced by the evolutionary background of the protein.

4 Genome-Wide Screens for Convergent Molecular Evolution

The identification of molecular convergence in prestin led to a search for molecular convergence between echolocating mammals in additional candidate genes with known hearing-related functions, based on knockout studies in mice or associations with deafness and hearing disorders in humans. These candidate gene approaches detected molecular convergence in several hearing-related proteins in echolocating mammals: Kcnq4, Tmc1, Pjvk, Cdh23, Pcdh15, and Otof (Liu et al. 2011; Davies et al. 2012; Shen et al. 2012). However, functional convergence of these six proteins or the effect of the convergent substitutions (criteria iii and iv according to Zhang (2006)) has not been experimentally explored. Consequently, it is currently

unknown whether there is convergence in protein function and whether the convergent mutations are involved.

The main limitation of the candidate gene approach to discover molecular convergence is the requirement for functionally well-characterized genes that could be associated with a well-characterized convergent phenotype in selected species. Advancements in sequencing technologies have led to the sequencing of hundreds of genomes, and the number continues to grow rapidly. This wealth of genomic data has made it possible to carry out genomic screens to detect genes with molecular convergence in the selected species. For example, Parker et al. (2013) sequenced the genomes of four bats and screened 2326 orthologous proteins in 22 mammals for signatures of molecular convergence in the echolocating bats and dolphin. For each gene, Parker et al. (2013) calculated the difference in the site-specific likelihood between a null tree (the accepted mammalian phylogeny) and hypothetical trees in which the echolocating mammals are artificially clustered together. A higher value for the hypothetical trees was taken as support for convergence, based on comparisons with a simulated null distribution of the site under the same parameters. In this way, Parker et al. (2013) suggested that 117 proteins exhibit signatures of convergence, particularly those linked to hearing and vision. Besides echolocating mammals, a genomic screen in three marine mammalian lineages (cetaceans, pinnipeds, sirenia) reported positive selection and molecular convergence in a small subset of proteins that could be linked to marine adaptations (Foote et al. 2015).

However, the use of site-specific support values and simulations by Parker et al. (2013) was subsequently criticized by several studies, mostly because many of the reported proteins do not exhibit convergent amino acid substitutions between echolocators. Out of the 117 candidate genes identified by Parker et al. (2013), only 19 were found to exhibit convergent substitutions between the echolocating lineages (Thomas and Hahn 2015). Although convergent mutations can result in a higher likelihood for the hypothetical tree, there are other factors unrelated to convergence that affect the likelihood of the null tree and hypothetical tree. As succinctly stated by Zou and Zhang (2015b), "convergence does not necessarily result in a wrong phylogeny and a wrong phylogeny is not necessarily caused by convergence." Furthermore, a re-examination of the 22 novel candidate genes related to hearing (Parker et al. 2013) found that 45% (10 of 22) did not pass statistical tests to demonstrate a higher than expected amount of convergence (Zou and Zhang 2015b). Using a set of 6400 orthologous proteins from 9 mammals, Thomas and Hahn (2015) also found 1951 genes that show convergence between microbat and cow, which is higher than the 1372 genes that show convergence between microbat and dolphin. In addition, the majority of the candidate genes exhibit convergent substitutions in other non-echolocating lineages; thus, the observed convergence cannot be wholly attributed to the adaptive phenotype (echolocation). Similarly, Foote et al. (2015) noted in their study of marine mammals that the terrestrial sister taxa of the marine mammals, which were used as control, exhibited a higher amount of convergent substitutions. Nonetheless, both follow-up studies agreed that the existence of background convergence in many

species does not exclude the possibility of adaptive molecular convergence in a small number of proteins in specific lineages, such as prestin and possibly other hearing-related genes in echolocating mammals (Thomas and Hahn 2015; Zou and Zhang 2015b).

The main challenge, as discussed by the studies above, is the lack of an accurate probabilistic model of sequence evolution that applies to all sites in a protein. Most sequence evolution models tend to underestimate or do not take into account background convergence (Castoe et al. 2009). This is problematic as several studies have reported that background convergence is more prevalent than previously expected (Stayton 2008; Rokas and Carroll 2008; Thomas and Hahn 2015; Zou and Zhang 2015a). Several solutions have been proposed to address these issues, such as tools that employ different statistical methods to determine whether convergent substitutions in a pair of lineages are higher than expected under certain models or tests that compare the number of divergent to the number of convergent amino acid changes between all lineage pairs (Castoe et al. 2009; Qian et al. 2015). Nevertheless, determining whether adaptive molecular convergence has occurred remains an active area of research.

5 Development of a Pipeline to Search for Convergent Molecular Evolution in an Unbiased Fashion

Despite the lack of a proper statistical framework to assess the significance of convergent mutations, the rapidly increasing number of sequenced genomes provides an unprecedented opportunity to carry out genomic screens on a multi-species scale to develop new hypotheses about associations between molecular convergence and phenotypic convergence. This approach is analogous to the use of genomic screens for positively selected proteins or proteins with lineage-specific amino acid mutations, which uncovered compelling examples that could be related to phenotypic changes. For example, proteins related to DNA replication and repair such as APEX1 and RFC1 exhibit amino acid changes that are unique to the naked mole rat and did not occur in other mammals (Kim et al. 2012). These naked mole rat-specific changes might be linked to the extraordinarily long life span and cancer resistance in that species. Another study identified a number of aging and cancer-related genes that were under positive selection such as SOCS2, APTX, NOG, and LEP in the long-lived bowhead whale (Keane et al. 2015).

There are two advantages in using genomic screens. First, they have the potential to uncover sets of functionally related genes that exhibit a higher number of convergent substitutions in particular lineages. The effect of molecular convergence on phenotypic convergence might be easier to interpret within such a gene set, even though some of those genes could exhibit convergence due to random chance. Second, the genomes of many species provide an opportunity to not only search for convergence in species known to exhibit convergent phenotypes, but to detect molecular convergence between any pair of independent lineages. This is

advantageous as it allows for a thorough test on whether a set of functionally-related genes (such as hearing-related genes) also exhibits a similar number of convergent mutations in other lineages (such as non-echolocating mammals).

To this end, we have developed a pipeline to look for genes exhibiting convergent evolution in a systematic, genome-wide manner (Lee et al. 2017). In our approach, the first step involved obtaining one-to-one orthologous protein sequences of 31 mammalian species from Ensembl (Hubbard et al. 2009; Kinsella et al. 2011), followed by filtering of ambiguous and poor-quality sequences. Next, we carried out a multiple sequence alignment for all ortholog sets, followed by ancestral reconstruction using a maximum likelihood approach. Subsequently, we iterated over every position in the alignment and systematically detected all convergent amino acid substitutions in all independent pairs of lineages (Fig. 2). Applying this pipeline to 14,407 sets of orthologous proteins across 31 mammals required approximately 180 CPU hours, though this can be greatly reduced through parallelization on a computer cluster.

We found 13,330 proteins that have at least one convergent substitution in one or more independent pairs of lineages. In total, we found over a million entries that consist of a protein with one or more convergent substitutions between a lineage pair. This finding is consistent with studies that reported on widespread background convergence in various genomes (Rokas and Carroll 2008; Thomas and Hahn 2015; Zou and Zhang 2015a). We further observed that most of the convergent substitutions are either conservative (substitution to another amino acid with similar physicochemical properties) or occur at alignment positions that are poorly conserved (Fig. 3). Both factors indicate that the majority of convergent amino acid substitutions are less likely to affect protein function. Indeed, the convergent changes in prestin that affect protein function (N7T and I384T) occur at highly conserved positions and are radical substitutions i.e., a replacement of an amino acid with another residue that has different physicochemical properties (Li et al. 2010). To enrich for molecular convergence that is more likely to affect function, we specifically filtered for convergent substitutions that are both radical and occur

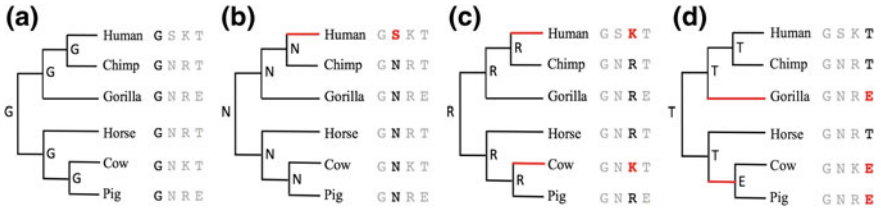

Fig. 2 Illustration of the systematic search for convergent substitutions in all pairs of lineages. First, ancestral reconstruction was carried out for every internal node in the species phylogeny. Next, we screen for convergent substitutions in independent lineages in each column of the alignment. **a** There are no substitutions observed in any lineages. **b** A substitution of N → S was observed only in human. **c** A convergent change of R → K is observed in human and cow, and this is recorded as an entry. **d** Another convergent change of T → E is observed in gorilla and the ancestral lineage of cow-pig, which is also recorded as an entry

at a highly conserved position, which greatly reduced the candidate list to 145,465 lineage pairs with one or more convergent substitutions in a protein (Lee et al. 2017).

Complex phenotypic changes likely require multiple changes in functionally related genes, exemplified by several hearing-related genes that exhibit convergence in echolocating mammals. Therefore, we proceeded to perform functional enrichment tests of all proteins that exhibit convergence between a pair of lineages. As shown in Table 1, for the echolocating microbat and dolphin, we found enrichments for hearing-related terms such as "abnormal ear morphology" that refer to genes that affect ear morphology in a mouse knockout. These genes include those that have been reported previously (e.g., prestin and Pjvk) (Li et al. 2010; Davies et al. 2012). In addition, we found several genes that have not been reported previously. For example, we detected the Bardet–Biedl syndrome 2 protein (Bbs2) and tyrosine-related protein 1 (Tyrp1), two proteins that are expressed in the cochlea (Fig. 4). Bbs2 is implicated in Bardet–Biedl syndrome, an autosomal recessive disorder that includes speech impairment and sensorineural hearing loss (May-Simera et al. 2009). Tyrp1 is a melanosomal enzyme that has been implicated in the decline of the endocochlear potential, which is one of the factors that contributes toward age-related hearing loss (Ohlemiller 2009).

The design of our genome-wide screen in detecting convergence between any pair of lineages makes it possible to test whether the observed enrichment in hearing-related genes is higher for microbat and dolphin compared to all other pairs of lineages. To this end, we iterated over all pairs of independent lineages, defined as two branches that do not share a direct common ancestor and where one branch is not a descendant of the other. For each independent pair of branches, we counted the number of convergent amino acid changes in all proteins that are associated

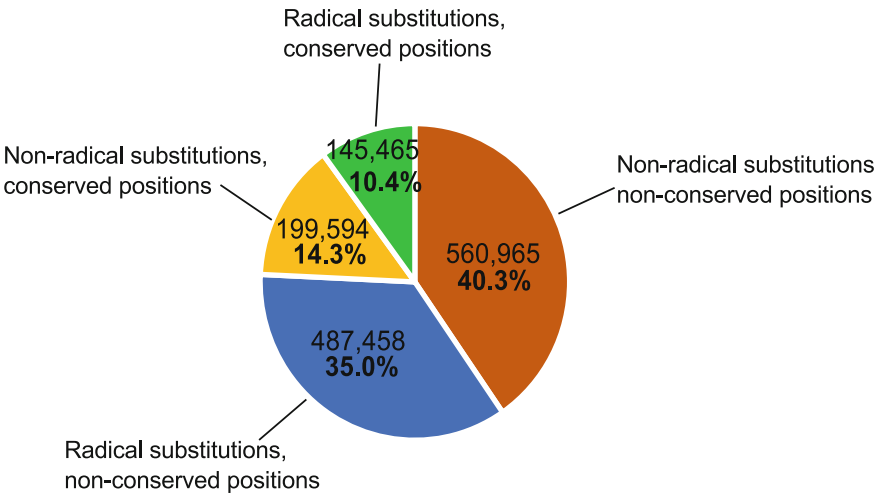

Fig. 3 Classification of all detected convergent amino acid substitutions

Table 1 Functional enrichments of the set of genes with radical convergent amino acid substitutions at conserved positions using Enrichr (Kuleshov et al. 2016) for mammalian phenotypes and GeneTrail2 (Stöckel et al. 2016) for gene ontology, ranked by P-values adjusted for multiple testing. The terms highlighted in *red* represent terms that are related to the physiology/morphology of ear and cellular components of contractile muscle fibers. The "Hits" column indicates the number of genes annotated with the phenotype or gene ontology term that has convergent amino acid substitutions

MGI Mammalian Phenotype Level 3	Hits	Adjusted P-value
MP0001919 abnormal reproductive system physiology	47	1.86E-003
MP0010769 abnormal survival	95	1.86E-003
MP0001672 abnormal embryo development	46	3.00E-003
MP0002102 abnormal ear morphology	19	4.09E-003
MP0001968 abnormal touch/nociception	12	1.76E-002
MP0003632 abnormal nervous system morphology	57	1.76E-002
MP0003633 abnormal nervous system physiology	44	1.83E-002
MP0004924 abnormal behavior	67	2.44E-002
MP0004196 abnormal prenatal growth/weight/body size	28	2.59E-002
MP0009389 abnormal extracutaneous pigmentation	8	3.61E-002
MP0003878 abnormal ear physiology	16	4.98E-002
MP0005621 abnormal cell physiology	36	4.98E-002
MP0002163 abnormal gland morphology	35	5.22E-002
MP0000358 abnormal cell morphology	12	5.29E-002
MP0002106 abnormal muscle physiology	23	5.29E-002

GO – Cellular Component	Hits	Adjusted P-value
contractile fiber part	15	1.29E-007
contractile fiber	15	1.40E-007
sarcomere	14	1.40E-007
ciliary part	17	1.69E-007
myofibril	14	3.83E-007
I band	10	2.72E-005
organelle subcompartment	13	9.84E-005
clathrin coated vesicle	9	9.85E-004
primary cilium	10	9.85E-004
sarcolemma	8	9.85E-004
Golgi subcompartment	11	1.54E-003
nuclear membrane	11	1.54E-003
apical plasma membrane	11	1.55E-003
axoneme part	5	1.55E-003
axoneme & ciliary cytoplasm	7	1.55E-003

with abnormal ear morphology. This shows that out of all 1984 pairs of lineages, microbat and dolphin rank 14th with a total of 22 observed convergent changes (Fig. 5). In other words, 99.3% of all other lineage pairs exhibit fewer convergent changes, suggesting that the convergence in at least some of these proteins could play a role in the evolution of echolocation.

Fig. 4 Alignment of Bbs2 and Tyrp1, two hearing-related proteins identified to show convergent substitutions in the echolocating microbat and dolphin

6 The Evolution of Superfast Muscles in Echolocating Mammals

Apart from ear-related enrichments, the proteins with convergence between dolphin and microbat also show statistical enrichments for genes related to muscle function (Table 1). Interestingly, some of these genes are specifically related to the physiology of fast-twitch muscles. As explained below in detail and in Lee et al. (2017), this suggests that there could be a connection between molecular convergence in fast muscle proteins and the less studied convergent aspect of echolocation, which is vocalization.

The ability to echolocate is a complex phenotype that requires the ability to produce and detect high frequency calls and subsequently convert this information into an acoustic representation of their environment. Most studies on convergence have focused on the high-frequency hearing aspect of echolocation, while little attention has been directed to vocalization. This is partially due to the fact that there is no obvious convergence in vocalization between echolocating bats and toothed whales as bats produce sounds in their larynx and toothed whales in their nasal complex (Clement et al. 2006; Berta et al. 2014). Despite the different anatomical

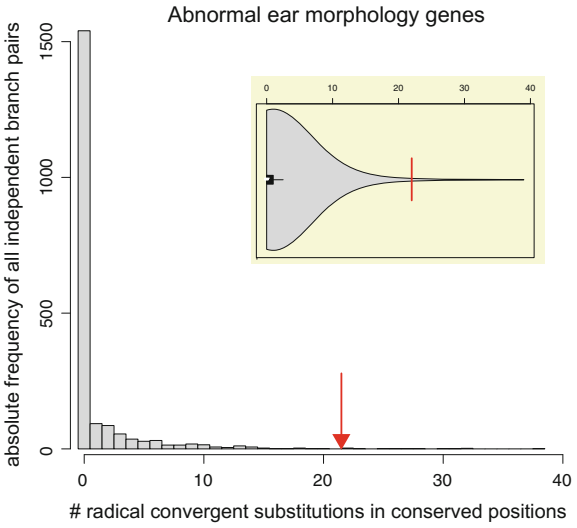

Fig. 5 Histogram of the number of convergent substitutions for each independent pair of lineages observed in 222 proteins that affect ear morphology in a mouse knockout (MGI phenotype identifier MP0002102). Only radical convergent amino acid substitutions at conserved positions are considered. The *red arrow* shows the microbat–dolphin pair. The *inset* shows the same data as a violin plot

structures used for sound production, both lineages have converged in the way clicks and calls are produced in the final moments of capturing prey. During foraging, there are two phases of call production in bats which are termed as the search phase (flying without a specific direction) and approach phase (flying toward a target). Once prey is detected, the bat will transition into the approach phase, which is accompanied by drastically increasing the repetition rate of its calls to precisely track its prey (Fenton et al. 2012). The final moments of capturing prey end with a terminal buzz, which is a period of extremely high repetition call rate up to 200 calls per second (Moss et al. 2011). The terminal buzz provides the bat with near-instantaneous feedback as it homes in on the prey's position, in case of any sudden escape maneuvers. Strikingly, the use of terminal buzz has also been observed in toothed whales, indicating that both lineages converged not only in the hearing-related aspect of echolocation but also in a similar strategy of producing calls to maximize the prey capture success rate (Johnson et al. 2006).

The extremely rapid call rates in bats during the terminal buzz are powered by equally rapid superfast muscles found in the larynx (Elemans et al. 2011). Isolated fibers from these highly unique muscles have been experimentally demonstrated to produce power cycles up to 180 Hz, which is close to the call rates produced during the terminal buzz. Although the bat laryngeal muscle is the first superfast muscle that was experimentally investigated in mammals, other experiments have demonstrated the existence of superfast muscles in other vertebrates including the swim bladder muscles that produce the "boatwhistle" mating call in male toadfish (Rome et al. 1996), vocal muscles in songbirds (Elemans et al. 2004), shaker muscles in rattlesnakes (Schaeffer et al. 1996), and most recently, wing muscles in manakin (Fuxjager et al. 2016).

Based on our finding that several fast muscle proteins show convergent substitutions in echolocating mammals, we hypothesized that these proteins could be involved in the physiological adaptations toward building superfast muscles that power the incredibly rapid calls during the terminal buzz. Further detailed work including functional experiments will be necessary to investigate this hypothesis.

7 Summary

The availability of many sequenced genomes has made it possible to extend candidate gene approaches and systematically screen for molecular convergence. This in turn intensifies the problem of assessing whether the observed convergence is higher than expected by chance. Despite several proposed approaches, this problem is an active area of research without a general consensus on the best solution. Nonetheless, genome-wide screens have the potential to detect novel associations between phenotypic change and genomic differences (Hiller et al. 2012; Prudent et al. 2016). As we have shown here, genome-wide screens for convergence in particular lineages can detect not only individual genes, but sets of functionally related genes or pathways by using statistical enrichment tests. Furthermore, unbiased genome-wide screens that simultaneously record convergence between any pair of lineages will allow subsequent testing on whether the convergence observed in the gene set is exceptional compared to all other pairs of lineages. Although this is not a formal proof of adaptive convergence, it can still provide new hints toward potentially unknown phenotypic convergence in these species and motivate further investigation. Ultimately, experiments are necessary to demonstrate convergence in the function of the discovered proteins and where possible, to test the role of the convergent mutations. Similar to genomic screens that searched for other molecular patterns (positive selection or lineage-specific mutations in genes, gene family expansions or gene losses), genomic screens for molecular convergence have the potential to generate new hypotheses of associations between molecular convergence and phenotypic convergence, which will help to illuminate the genomic changes that underlie nature's fascinating phenotypic diversity.

References

Berta A, Ekdale EG, Cranford TW (2014) Review of the cetacean nose: form, function, and evolution. Anat Rec 297:2205–2215

Castoe TA, de Koning APJ, Kim H-M et al (2009) Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci USA 106:8986–8991

Chen L, DeVries AL, Cheng CH (1997) Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proc Natl Acad Sci USA 94:3817–3822

Christin PA, Weinreich DM, Besnard G (2010) Causes and evolutionary significance of genetic convergence. Trends Genet 26:400–405

Clement M, Dietz N, Gupta P (2006) Audiovocal communication and social behavior in mustached bats. In: Kanwal J, Ehret G (eds) Behavior and neurodynamics for auditory communication. Cambridge University Press, Cambridge, pp 57–84

Dallos P, Fakler B (2002) Prestin, a new type of motor protein. Nat Rev Mol Cell Biol 3:104–111

Davies KTJ, Cotton JA, Kirwan JD et al (2012) Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. Heredity 108:480–489

Davies PL, Baardsnes J, Kuiper MJ, Walker VK (2002) Structure and function of antifreeze proteins. Philos Trans R Soc Lond B Biol Sci 357:927–935

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

Elemans CPH, Mead AF, Jakobsen L, Ratcliffe JM (2011) Superfast muscles set maximum call rate in echolocating bats. Science 333:1885–1888

Elemans CPH, Spierts ILY, Müller UK et al (2004) Bird song: superfast muscles control dove's trill. Nature 431:146

Feldman CR, Brodie ED, Pfrender ME (2012) Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. Proc Natl Acad Sci USA 109:4556–4561

Fenton M, Faure P, Ratcliffe J (2012) Evolution of high duty cycle echolocation in bats. J Exp Biol 215:2935–2944

Foote AD, Liu Y, Thomas GWC et al (2015) Convergent evolution of the genomes of marine mammals. Nat Genet 47:272–275

Fuxjager MJ, Goller F, Dirkse A et al (2016) Select forelimb muscles have evolved superfast contractile speed to support acrobatic social displays. Elife 5:3523–3528

Geffeney SL, Fujimoto E, Brodie ED III et al (2005) Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. Nature 434:759–763

Goldstein RA, Pollard ST, Shah SD, Pollock DD (2015) Nonadaptive amino acid convergence rates decrease over time. Mol Biol Evol 32:1373–1381

Hiller M, Schaar BT, Indjeian VB, et al (2012) A "Forward Genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep 1–7

Hu Y, Wu Q, Ma S et al (2017) Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. Proc Natl Acad Sci USA 114:201613870

Hubbard TJP, Aken BL, Ayling S et al (2009) Ensembl 2009. Nucleic Acids Res 37:D690–D697

Johnson M, Madsen PT, Zimmer WMX et al (2006) Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation. J Exp Biol 209:5038–5050

Kawamura S, Hiramatsu C, Melin AD, et al (2012) Polymorphic color vision in primates: evolutionary considerations. In: Post-genome biology of primates. Springer, Tokyo, pp 269–280

Keane M, Semeiks J, Webb AE et al (2015) Insights into the evolution of longevity from the bowhead whale genome. Cell Rep 10:112–122

Kim EB, Fang X, Fushan AA et al (2012) Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479:223–227

Kinsella RJ, Kähäri A, Haider S, et al (2011) Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database 2011:bar030

Kuleshov MV, Jones MR, Rouillard AD et al (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90–W97

Lee J-H, Lewis KM, Moural TW, et al (2017) Building superfast muscles: insights from molecular parallelism in fast-twitch muscle proteins in echolocating mammals. Submitted

Li G, Wang J, Rossiter SJ et al (2008) The hearing gene *Prestin* reunites echolocating bats. Proc Natl Acad Sci USA 105:13959–13964

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

Liu Z, Li S, Wang W et al (2011) Parallel evolution of *KCNQ4* in echolocating bats. PLoS ONE 6: e26618
Liu Z, Qi F-Y, Zhou X et al (2014) Parallel sites implicate functional convergence of the hearing gene *prestin* among echolocating mammals. Mol Biol Evol 31:2415–2424
May-Simera HL, Ross A, Rix S et al (2009) Patterns of expression of Bardet-Biedl syndrome proteins in the mammalian cochlea suggest noncentrosomal functions. J Comp Neurol 514:174–188
Moss CF, Chiu C, Surlykke A (2011) Adaptive vocal behavior drives perception by echolocation in bats. Curr Opin Neurobiol 21:645–652
Nagai H, Terai Y, Sugawara T et al (2011) Reverse evolution in *RH1* for adaptation of cichlids to water depth in Lake Tanganyika. Mol Biol Evol 28:1769–1776
Natarajan C, Hoffmann FG, Weber RE et al (2016) Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354:336–339
Ohlemiller KK (2009) Mechanisms and genes in human strial presbycusis from animal models. Brain Res 1277:70–83
Parker J, Tsagkogeorga G, Cotton JA et al (2013) Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502:1–9
Pollock DD, Thiltgen G, Goldstein RA (2012) Amino acid coevolution induces an evolutionary Stokes shift. Proc Natl Acad Sci USA 109:E1352–E1359
Prudent X, Parra G, Schwede P et al (2016) Controlling for hylogenetic relatedness and evolutionary rates improves the discovery of associations between species' phenotypic and genomic differences. Mol Biol Evol 33:2135–2150
Qian C, Bryans N, Kruykov I, de Koning APJ (2015) Visualization and analysis of statistical signatures of convergent molecular evolution. University of Calgary. http://lab.jasondk.io
Rokas A, Carroll SB (2008) Frequent and widespread parallel evolution of protein sequences. Mol Biol Evol 25:1943–1953
Rome LC, Syme DA, Hollingworth S et al (1996) The whistle and the rattle: the design of sound producing muscles. Proc Natl Acad Sci USA 93:8095–8100
Schaeffer PJ, Conley KE, Lindstedt STL (1996) Structural correlates of speed and endurance in skeletal muscle: the rattlesnake tailshaker muscle. J Exp Biol 199:351–358
Shen YY, Liang L, Li GS et al (2012) Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet 8:e1002788
Starr TN, Thornton JW (2016) Epistasis in protein evolution. Protein Sci 25:1204–1218
Stayton CT (2008) Is convergence surprising? An examination of the frequency of convergence in simulated datasets. J Theor Biol 252:1–14
Stewart CB, Schilling JW, Wilson AC (1987) Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330:401–404
Stöckel D, Kehl T, Trampert P et al (2016) Multi-omics enrichment analysis using the GeneTrail2 web service. Bioinformatics 32:1502–1508
Thomas GWC, Hahn MW (2015) Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. Mol Biol Evol 32:1232–1236
Ujvari B, Casewell NR, Sunagar K et al (2015) Widespread convergence in toxin resistance by predictable molecular evolution. Proc Natl Acad Sci USA 112:201511706
Weinreich DM (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114
Williams BP, Johnston IG, Covshoff S, Hibberd JM (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. Elife 2:e00961
Zhang J (2006) Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat Genet 38:819–823
Zhang J (2003) Parallel functional changes in the digestive RNases of ruminants and colobines by divergent amino acid substitutions. Mol Biol Evol 20:1310–1317
Zhang J, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequence level. Mol Biol Evol 14:527–536

Zhen Y, Aardema ML, Medina EM et al (2012) Parallel molecular evolution in an herbivore community. Science 337:1634–1637

Zou Z, Zhang J (2015a) Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? Mol Biol Evol 32:2085–2096

Zou Z, Zhang J (2015b) No genome-wide protein sequence convergence for echolocation. Mol Biol Evol 32:1237–1241