

ORIGINAL ARTICLE

Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence

KTJ Davies¹, JA Cotton^{1,3}, JD Kirwan², EC Teeling² and SJ Rossiter¹

Recent findings of sequence convergence in the *Prestin* gene among some bats and cetaceans suggest that parallel adaptations for high-frequency hearing have taken place during the evolution of echolocation. To determine if this gene is an exception, or instead similar processes have occurred in other hearing genes, we have examined *Tmc1* and *Pjvk*, both of which are associated with non-syndromic hearing loss in mammals. These genes were amplified and sequenced from a number of mammalian species, including echolocating and non-echolocating bats and whales, and were analysed together with published sequences. Sections of both genes showed phylogenetic signals that conflicted with accepted species relationships, with coding regions uniting laryngeal echolocating bats in a monophyletic clade. Bayesian estimates of posterior probabilities of convergent and divergent substitutions provided more direct evidence of sequence convergence between the two groups of laryngeal echolocating bats as well as between echolocating bats and dolphins. We found strong evidence of positive selection acting on some echolocating bat species and echolocating cetaceans, contrasting with purifying selection on non-echolocating bats. Signatures of sequence convergence and molecular adaptation in two additional hearing genes suggest that the acquisition of high-frequency hearing has involved multiple loci.

Heredity advance online publication, 14 December 2011; doi:10.1038/hdy.2011.119

Keywords: parallel sequence evolution; bats; high-frequency hearing

INTRODUCTION

Examples of adaptive phenotypic convergence are widespread in nature and are usually considered to arise from similar selection pressures acting on unrelated taxa (see, for example, Packard, 1972; Nevo, 1979). In comparison, there are few documented cases of convergence acting at the sequence level, although this appears to be steadily changing with the proliferation of genetic and genomic data (as reviewed in Christin *et al.*, 2010). Some of the earliest and best examples of convergent sequence evolution include the stomach lysozymes of langurs and cows (Stewart *et al.*, 1987), and the peptide-binding regions of major histocompatibility complex class Ib genes in primates and rodents (Yeager *et al.*, 1997). More recently, the mitochondrial genomes of snakes and agamid lizards have also been found to contain numerous convergent sites distributed across all protein-coding genes, leading to a conflict with the true species phylogenetic signal obtained from nuclear genes (Castoe *et al.*, 2009).

The 'hearing gene' *Prestin* was recently shown to have undergone unprecedented levels of sequence convergence between lineages of echolocating mammals (Li *et al.*, 2008, 2010; Liu *et al.*, 2010a). *Prestin* encodes a motor protein of the outer hair cell and is thought to drive the cochlear amplifier that gives mammalian hearing its characteristically high sensitivity and selectivity (Zheng *et al.*, 2000). Phylogenies based on *Prestin* lead to erroneous groupings of unrelated echolocating

bats (Li *et al.*, 2008) and also horseshoe bats and echolocating cetaceans (Li *et al.*, 2010; Liu *et al.*, 2010a,b). These results as well as observed associations between numbers of replacements and auditory sensitivity (Liu *et al.*, 2010b; Rossiter *et al.*, 2011) all suggest a strong link between *Prestin* and high-frequency hearing. Indeed, detected positive selection in *Prestin* in bats (Li *et al.*, 2008) and toothed whales (Liu *et al.*, 2010b) reinforces the adaptive significance of these changes.

The molecular basis of mammalian hearing involves over 50 candidate genes identified via studies of mutagenesis and non-syndromic hearing loss (Accetturo *et al.*, 2010; Dror and Avraham, 2010). The mammalian hearing apparatus has evolved into a wide range of auditory systems, the most specialised of which arguably occur in echolocating bats and cetaceans (Vater and Kossel, 2004). In the bats (order Chiroptera), laryngeal echolocation is shared by all members of the suborder Yangochiroptera, but only some members of the other suborder, the Yinpterochiroptera. In this latter group, the Old World fruit bats (family Pteropodidae) do not possess laryngeal echolocation, indicating that echolocation (and associated ultrasonic hearing) has either evolved separately in the Yangochiroptera and Yinpterochiroptera, or has been lost in the Old World fruit bats (see, for example, Teeling *et al.*, 2002; Miller-Butterworth *et al.*, 2007). Notwithstanding this debate, bats undoubtedly do show convergence in particular forms of echolocation, such as constant-frequency (CF) echolocation

¹School of Biological and Chemical Sciences, Queen Mary University of London, London, UK and ²UCD School of Biology and Environmental Science, UCD Science Centre West, University College Dublin, Belfield, Ireland

³Current address: Wellcome Trust Sanger Institute, Cambridge, UK.

Correspondence: Dr SJ Rossiter, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK.

E-mail: s.j.rossiter@qmul.ac.uk or Dr EC Teeling, UCD School of Biology and Environmental Science, UCD Science Centre West, University College Dublin, Belfield, Dublin 4, Ireland.

E-mail: emma.teeling@ucd.ie

Received 25 May 2011; revised 2 September 2011; accepted 18 October 2011

seen in horseshoe bats and the neotropical species *Pteronotus parnellii* (see, for example, Jones and Teeling, 2006).

Results from *Prestin* all suggest that echolocating mammals might be an especially useful group in which to test for evidence of genetic convergence. Extending this work, here we investigate whether adaptive convergence in *Prestin* in mammals with high-frequency hearing is exceptional, or whether similar patterns are detectable in other candidate hearing genes. We focus on two genes that have specific roles in hair cell function: *Tmc1* (transmembrane cochlear-expressed gene 1) and *Pjvk* (Pejvakin). *Tmc1* encodes a transmembrane protein of the inner and outer hair cells that might either traffic molecules to the plasma membrane or act as an intracellular regulatory signal during hair cell maturation (Marcotti *et al.*, 2006). In mice, *Tmc1* is expressed from early development and is needed for normal hair cell function (Marcotti *et al.*, 2006), whereas the human gene (*TMCI*) is associated with non-syndromic hearing loss with over 20 documented pathogenic mutations (Kurima *et al.*, 2002; Kitajiri *et al.*, 2007). *Pjvk* (also known as *Dfnb59*) encodes a protein called pejvakin. Missense and stop mutations in the human form have both been linked to deafness—in the former case caused by auditory neuropathy (Delmaghani *et al.*, 2006)—whereas in mice, premature stop codons disrupt hair cell activity and also cause vestibular defects (Schwander *et al.*, 2007).

Using a comparative analysis, we tested whether each gene (*Tmc1* and *Pjvk*) shows evidence of convergence and/or molecular adaptation associated with the independent evolution of high-frequency hearing. First, we used a phylogenetic approach to test whether, if present, sequence convergence would cause unrelated echolocating taxa (that is, divergent lineages of bats, and bats and cetaceans) to group together, as previously reported for *Prestin*. We expected that, if adaptive, any gene–species tree conflicts would be more evident in coding DNA. Second, we predicted that within the species tree, greater support for convergent changes between branches would correspond to the inferred origins of echolocation and high-frequency hearing. Stronger evidence of convergence between the ancestral branches of the two main groups of laryngeal echolocators would support an independent origin of echolocation (and high-frequency hearing) in bats. Third, we predicted that both candidate hearing genes would show evidence of positive selection in key branches and clades of echolocating taxa, and that the distribution of selection would correspond to convergence.

MATERIALS AND METHODS

Taxonomic coverage

We obtained *Tmc1* and *Pjvk* coding region sequences for all available mammals from Ensembl (<http://www.ensembl.org>). In addition, we used BlastN (Altschul *et al.*, 1997) to obtain the bat sequences, either from GenBank for the species *Myotis lucifugus* and *Pteropus vampyrus*, or from assembled short read Solexa (Illumina, Inc., San Diego, CA, USA) data for *Eidolon helvum*, *P. parnellii* and *Rhinolophus ferrumequinum*. For gene searches, the alpaca *Vicugna pacos* was used as a reference, and expected value thresholds of 10^{-6} were used for sequences of ~100 base pairs (bp) or more. The final *Tmc1* data set contained 23 species and covered 2139 bp, corresponding to 713 out of the total 760 amino acids (that is, from amino acids 25–736 using human *TMCI* as a reference). For *Pjvk*, sequences from 21 species were collated, which covered the entire 1059 bp or 352 amino acids. Thus, in total we obtained 93.8% and 100% of coding sequence of *Tmc1* and *Pjvk*, respectively, although both are henceforth referred to as ‘complete’ to distinguish them from the partial sequences (see below). Both of these data sets included three echolocating bats (two Yangochiroptera and one Yinpterochiroptera), two Old World fruit bats (Yinpterochiroptera), one echolocating cetacean and a range of other mammals with varying auditory thresholds (see Supplementary Table S1a for species list).

To improve taxonomic coverage, we also generated partial *Tmc1* and *Pjvk* gene sequences that contained coding (exonic) and non-coding (intronic) DNA. Sequence data from a total of 64 mammals included 39 laryngeal echolocating bats (from 12 families), 10 Old World fruit bats (1 family), 3 echolocating toothed whales, 1 non-echolocating baleen whale and 11 other mammals from 6 other orders. We also obtained partial gene sequences of 29 species (13 orders) from Ensembl. Our total coverage comprised 93 mammal species from 14 mammalian orders (see Supplementary Table S1b). This second data set of partial sequences included bat species that exhibit a wide range of echolocation call types and auditory characteristics (Jones and Teeling, 2006). From the Yinpterochiroptera, all six families were represented; the Old World fruit bats (Pteropodidae) do not possess laryngeal echolocation, the horseshoe (Rhinolophidae) and roundleaf bats (Hipposideridae) have narrowband pure CF echolocation and the other families (Megadermatidae, Rhinopomatidae and Craseonycteridae) possess more broadband echolocation calls with a range of bandwidths. From the Yangochiroptera, seven families were included, all characterised by broadband frequency-modulated echolocation calls, with the exception of *P. parnellii* that has independently evolved narrowband CF echolocation (see Supplementary Figure S1 and Supplementary Table S2).

DNA isolation, primer design, amplification and sequencing

For new sequences generated in this study, DNA was extracted using DNeasy kits (Qiagen, Crawley, UK). To amplify regions of interest across the study species, degenerate primers were designed using ‘Uniprime’ (Bekaert and Teeling, 2008) (see Supplementary Materials and Methods for details). For *Tmc1*, the primers amplified a region (corresponding to human exons 16 to 17) that has been implicated in deafness in mice (Kurima *et al.*, 2002). For *Pjvk*, the primers amplified a region (corresponding to human exons 5 to 6) that contains the functionally important putative nuclear localisation signal and zinc-binding motif (Delmaghani *et al.*, 2006). This region also contains a premature stop codon in mutant mice (Schwander *et al.*, 2007).

Data analysis

For each gene, separate alignments of complete and partial sequences were conducted using ClustalW2 (Larkin *et al.*, 2007) and checked by eye. For the partial sequences that contained non-coding DNA, the *Homo sapiens* (EMBL-EBI) gene was used to identify the exon–intron boundaries and identify open reading frames for generating alignments of in-frame exons only. Before analyses, we removed a three-codon insertion that was present in two species of *Hipposideros* (X497E, X498M and X499A) and five representatives of the family Phyllostomidae (X497Q, X498L and X499S). The same positional insertion (S, G and L) is also seen in the mouse (*Mus musculus*). To test for adaptive sequence convergence associated with the acquisition of high-frequency hearing, we undertook analyses that comprised three steps, described below (for full details, see Supplementary Materials and Methods). For both genes, each analysis was repeated separately for the alignments of complete and partial sequences.

Tests of phylogenetic signal and hypotheses. We first tested whether phylogenies based on the complete coding sequence of each of the two genes were concordant with the species relationships or the published *Prestin* trees in which echolocators erroneously group together. Maximum likelihood and Bayesian trees were reconstructed with RAXML-7.0.3 (Stamatakis, 2006) and MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003), respectively. In both trees, laryngeal echolocating bats formed a monophyletic group (see Results). Therefore, to compare statistical support for these observed trees versus the constrained species tree, we used the approximately unbiased test (AUT) (Shimodaira, 2002). To address our prediction that such phylogenetic conflicts would arise because of adaptive changes, we repeated these analyses using the second data set, for both the exon-only data and the intron plus exon data. To compare how support for the convergent versus species tree topologies was distributed between exonic and intronic sequence, we also used Lento plots visualised in SPECTRONET (Huber *et al.*, 2002). Here, reduced sequence alignments were converted into a series of taxonomic splits, and support and conflict values for each of these splits were calculated.

Probabilistic analyses of sequence convergence. For each gene, we characterised the distribution of sequence convergence between pairs of branches in the species phylogeny following the approach described by Castoe *et al.* (2009) implemented in the package Codeml ancestral. We predicted greatest levels would occur between taxa known to have independently evolved echolocation or particular types of echolocation (see specific hypotheses in Introduction); in the case of Yangochiroptera versus Yinpterochiroptera, evidence of convergence would add weight to the hypothesis that laryngeal echolocation in bats has evolved multiple times. In this method, posterior probabilities of all possible amino-acid substitutions were calculated along each branch under a JTT+F+G model of amino-acid substitution. We used tree topologies based on published data (Csorba *et al.*, 2003; Nishihara *et al.*, 2006; Miller-Butterworth *et al.*, 2007; Murphy *et al.*, 2007; McGowen *et al.*, 2009; Khan *et al.*, 2010) and we estimated branch lengths using MrBayes v3.1.2. For all pair-wise branch comparisons where both branches followed divergent paths, the sum of the joint probabilities of all possible pairs of convergent substitutions (same amino acid) and divergent substitutions were calculated. Levels of convergence for pairs of branches of interest were examined.

To test whether convergence between pairs of focal branches was significant, we compared observed probabilities against null distributions based on simulations. We first generated simulated sequences in EVOLVER (Yang, 2007) using the species tree topology under a substitution model based on the empirical codon frequencies. Sequences of 5000 and 10000 codons (for partial and complete data sets, respectively) were analysed in Codeml ancestral to obtain pools of site-wise posterior probabilities of convergence. A total of 1000 replicates of branch-wise convergence probabilities were calculated, with each value obtained by summing a random sample of site-wise probabilities in which sample size equalled the observed number of values in the equivalent branches. These tests were run in the R package (R Development Core Team, 2011).

Finally, we also examined whether the location of high probabilities of convergence along the protein was associated with particular functional domains. The positions of protein domains were taken from literature sources (Kurima *et al.*, 2002; Delmaghani *et al.*, 2006; Schwander *et al.*, 2007) and confirmed using SMART (Schultz *et al.*, 1998).

Tests for selection. To determine whether observed convergent substitutions were associated with molecular adaptation or neutral evolution, we used branch-site and clade models of selection, implemented in Codeml in the package PAML 4.4 (Yang, 2007). We derived maximum likelihood estimates of non-synonymous (d_N) and synonymous (d_S) substitution rates. A d_N/d_S ratio (termed omega, ω) of >1 signifies positive selection, a ratio of ~ 1 signifies neutrality and a ratio of <1 signifies purifying selection.

Branch-site models

Branch-site models (Zhang *et al.*, 2005) are suitable for detecting specific episodes of positive selection that may have acted over a small number of branches. Using the complete sequences, we estimated the selection pressures acting at sites along the ancestral bat branches, as well as the branches leading to echolocating taxa. Estimates of site-wise ω values in these so-called foreground branches were then compared with estimates across the remaining background branches in the species phylogeny. The ω values were assigned to four predefined site classes under model A. The first site class was estimated from the data but constrained ($0 < \omega_0 < 1$), the second class $\omega_1 = 1$, the third class ω_{2a} could exceed 1 on the foreground but is constrained to be under purifying selection on the background, and the final class ω_{2b} could exceed 1 on the foreground but not on the background. This model was compared with the null model A where $\omega_{2a} = 1$, with a likelihood ratio test (LRT) with one degree of freedom (d.f.). If the alternative model was a better fit (that is, positive selection was detected along a focal branch), Bayes empirical Bayes was used to quantify the probability that particular site was under positive selection.

Clade models

To test for divergent selection acting on echolocating versus non-echolocating taxa, we compared ω averaged across branches within focal clades (foreground) with ω estimated for the rest of the tree (background) (Bielawski and Yang, 2004). These analyses were conducted on the partial sequences that included

more echolocating species. For foreground and background, three site classes were modelled (model C), ω was constrained in two of these ($0 < \omega_0 < 1$ and $\omega_1 = 1$) and in the third class, ω_2 and ω_3 were modelled for foreground and background, respectively. Site-wise mean ω were estimated by multiplying resulting ω within each class by its corresponding posterior probability. For each clade, model C was compared with model M1a (nearly neutral) using a LRT.

To visualise selection pressures acting along sites in those species found to be associated with high probabilities of convergence, we repeated our clade models for several families of echolocating bats. In these models, we removed all other echolocating bats and cetaceans from the background, to avoid underestimating divergent selection.

RESULTS

We combined bioinformatics approaches and molecular methods to obtain both complete and partial sequences of two candidate hearing genes: *Tmc1* and *Pjvk*. The complete sequences provided more power for studying site-wise patterns of molecular adaptation in a few key species, whereas our partial sequence data sets included many more taxa, and thus provided better power for examining molecular evolution within echolocating mammals.

Tests of phylogenetic signal and hypotheses

Bayesian and maximum likelihood trees based on complete *Tmc1* and *Pjvk* coding regions recovered consistent topologies that showed some strong similarities to previous results reported for *Prestin*; that is, all laryngeal echolocating bats erroneously formed a monophyletic clade with the exclusion of Old World fruit bats. However, both gene trees recovered all bats as monophyletic, hence supporting the accepted deeper level phylogeny. At the same time, the bottlenose dolphin *Tursiops truncatus* was not seen to group with the bats. Support for the node of laryngeal echolocating bats was relatively high for both *Tmc1* (Bayesian posterior probability (BPP)=0.83, bootstrap (BS)=63%) and *Pjvk* (BPP 0.91, BS 51%; see Figure 1). AUTs indicated that for both genes, neither 'convergent tree' topology (*Tmc1*: $P=0.291$; *Pjvk*: $P=0.136$, AUT) was significantly less supported than either the unconstrained tree (*Tmc1*: $P=0.804$; *Pjvk*: $P=0.295$, AUT), or the constrained 'species tree' (*Tmc1*: $P=0.102$; *Pjvk*: $P=0.830$, AUT).

When we repeated the maximum likelihood and Bayesian *Tmc1* and *Pjvk* reconstructions using partial sequences from a greater range of taxa, we again recovered the monophyly of laryngeal echolocating bats based on coding regions (that is, exons only). In both cases, node support for this monophyletic group was low (*Tmc1*: BPP 0.73, BS 36%; *Pjvk*: BPP 0.71, BS 25%) and none of the alternative tree topologies could be rejected (for AUT results, see Supplementary Results). However, trees based on exon+intron alignments did not recover this clade. Instead, the *Tmc1* tree contained the correct subordinal groupings of Yangochiroptera and Yinpterochiroptera, whereas in the *Pjvk* tree, the Old World fruit bats clustered with the Yangochiroptera to the exclusion of the laryngeal echolocating Yinpterochiroptera. As with the data set of complete sequences, the cetaceans did not cluster with the bats (for full tree results from partial sequences, see Supplementary Results and Supplementary Figure S2).

Lento plots undertaken to visualise the numbers of splits that grouped echolocating species of Yinpterochiroptera with Yangochiroptera revealed that intron data contained fewer bifurcating splits than exon data (see Supplementary Figure S3). The smaller number of supported splits in the introns points to a clearer phylogenetic signal, compared with a more conflicting signal found in the exons.

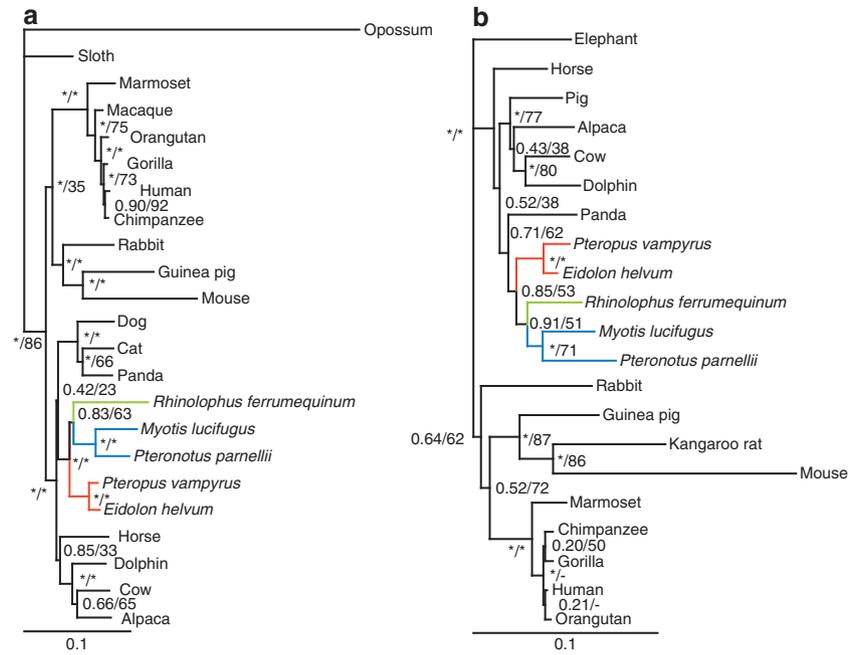


Figure 1 Tree topologies recovered by Bayesian and maximum likelihood (ML) analyses for (a) 2139 bp of *Tmc1* and (b) 1059 bp of *Pjvk*. Nodal support values are Bayesian posterior probabilities and bootstrap values based on 1000 replicates (- indicates topology differed in ML, */ indicates BPP > 0.95 and /* indicates BS > 95%). Branch colours are as follows: non-bat species (black); Old World fruit bats (red); laryngeal echolocating Yinypterochiroptera (green); and Yangochiroptera (blue).

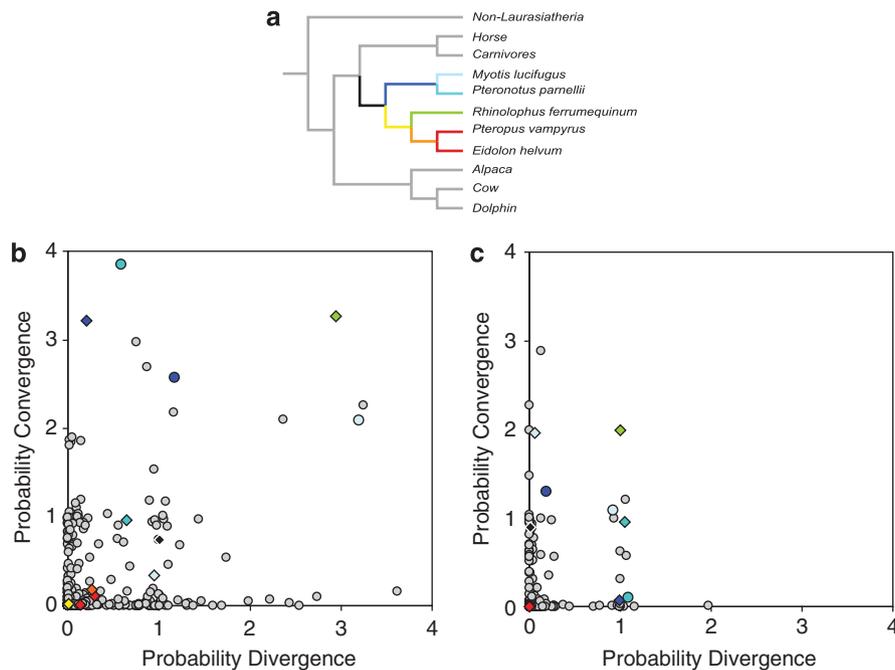


Figure 2 (a) Simplified species tree including taxa involved in the focal bat-dolphin branch comparisons. Branch-pair plots of total posterior probability divergence vs total posterior probability convergence for (b) *Tmc1* and (c) *Pjvk*. *T. truncatus*-bat comparisons are indicated by diamonds and *R. ferrumequinum*-Yangochiroptera comparisons by circles. Points are coloured according to the second branch in the comparison and follow the species tree (a). The remaining points (grey circles) correspond to comparisons between the remaining non-echolocating mammal species.

Probabilistic analyses of sequence convergence

Using the complete sequences we tested for convergence associated with the acquisition of echolocation and high-frequency hearing in the two suborders of echolocating bats, and between bats and the dolphin.

Plots of branch-wise probabilities of convergence versus divergence for all placental mammals in our tree showed that comparisons between echolocating taxa had among the highest values across all species for both genes (Figure 2).

For *Tmc1* the highest convergence probability occurred between *Rhinolophus ferrumequinum* and *P. parnellii*, both of which have independently evolved CF echolocation. *R. ferrumequinum* also showed high convergence probabilities with *M. lucifugus* and with the ancestral branch of the Yangochiroptera. The second and third highest values overall were seen between *T. truncatus* and, respectively, *R. ferrumequinum* and the ancestral branch of the Yangochiroptera. In contrast, we found no evidence for high levels of convergence between *T. truncatus* and either the ancestral Yinpterochiroptera branch or the ancestral bat branch (see Figure 2a). All of these high convergence probabilities between branches were significant based on simulations, with the exception of the comparison between *P. parnellii* and *T. truncatus* (see Table 1a).

To examine the *Tmc1* sites driving the observed branch-wise convergence between echolocating taxa, we plotted site-wise convergence probabilities and related these to functional domains along the protein (Figure 3). We found that *R. ferrumequinum* showed multiple convergent sites with *P. parnellii* (4 sites), *M. lucifugus* (2) and the

ancestral Yangochiroptera (2). In total, therefore, eight convergent sites occurred between three branch-wise comparisons involving Yangochiroptera and echolocating Yinpterochiroptera; in contrast, three sites were found among nine branch-wise comparisons involving the Yangochiroptera and the Old World fruit bats (Figure 2a). Between bats and the dolphin, three convergent sites were detected between *T. truncatus* and *R. ferrumequinum*, one of which was also shared by *P. parnellii* (see above). Three further convergent sites were detected between *T. truncatus* and the ancestral Yangochiroptera branch and one between *T. truncatus* and the common ancestral bat branch (see Table 1a and Figure 3a). Taking both bat–bat and bat–dolphin convergence into account, the majority of sites supporting convergence were located outside of transmembrane domains (with one site located in the highly conserved TMC domain) and are, therefore, either intra- or extra-cellular residues (Figure 3). We found that all *Tmc1* sites with high probabilities of convergence had arisen from the same ancestral state, and can thus be classified as parallel changes.

Table 1 Sites along (a) *Tmc1* and (b) *Pjvk* with >0.5 PP total convergence along key branch pairs

Branch pairs	Site (BPP convergence)	Branch-wise significance
(a)		
<i>R. ferrumequinum</i> vs <i>P. parnellii</i>	N 50 S (0.84), S 128 N (0.98), A 154 V (0.83), M 269 L (0.99)	<0.001
<i>T. truncatus</i> vs <i>R. ferrumequinum</i>	A 154 V (0.98), C 613 G (0.97), K 691 M (1.00)	<0.001
<i>T. truncatus</i> vs Ancestral Yangochiroptera	K 76 R (0.93), L 192 M (0.97), C 613 G (0.96)	<0.001
<i>R. ferrumequinum</i> vs <i>M. lucifugus</i>	K 263 Q (0.97), K 298 Q (0.99)	<0.001
<i>R. ferrumequinum</i> vs Ancestral Yangochiroptera	P 501 A (0.85), C 613 G (0.93)	<0.001
<i>T. truncatus</i> vs <i>P. parnellii</i>	A 154 V (0.84)	0.072
<i>T. truncatus</i> vs Ancestral bat	A 710 V (0.66)	<0.001
<i>T. truncatus</i> vs Ancestral Yinpterochiroptera	None	<0.001
<i>T. truncatus</i> vs Ancestral Pteropodidae	None	0.067
<i>T. truncatus</i> vs <i>E. helvum</i>	None	<0.001
<i>T. truncatus</i> vs <i>P. vampyrus</i>	None	<0.001
<i>T. truncatus</i> vs <i>M. lucifugus</i>	None	0.079
Ancestral Pteropodidae vs <i>M. lucifugus</i>	I 204 V (0.88)	NA
Ancestral Pteropodidae vs <i>P. parnellii</i>	L 481 I (0.87)	NA
<i>P. vampyrus</i> vs <i>P. parnellii</i>	V 358 I (0.92)	NA
(b)		
<i>T. truncatus</i> vs <i>R. ferrumequinum</i>	I 72 V (0.99), K 193 R (0.98)	0.007
<i>T. truncatus</i> vs <i>M. lucifugus</i>	A 181 S (0.99), K 193 R (0.96)	0.002
<i>R. ferrumequinum</i> vs Ancestral Yangochiroptera	A 262 V (0.73), L 282 F (0.55)	<0.001
<i>R. ferrumequinum</i> vs <i>M. lucifugus</i>	K 193 R (0.96)	<0.001
<i>T. truncatus</i> vs <i>P. parnellii</i>	I 346 V (0.94)	0.028
<i>T. truncatus</i> vs Ancestral bat	I 110 V (0.89)	<0.001
<i>T. truncatus</i> vs Ancestral Yangochiroptera	None	0.057
<i>R. ferrumequinum</i> vs <i>P. parnellii</i>	None	0.031
<i>T. truncatus</i> vs Ancestral Yinpterochiroptera	None	<0.001
<i>T. truncatus</i> vs Ancestral Pteropodidae	None	0.066
<i>T. truncatus</i> vs <i>E. helvum</i>	None	1.000
<i>T. truncatus</i> vs <i>P. vampyrus</i>	None	1.000
Ancestral Pteropodidae vs <i>M. lucifugus</i>	None	NA
Ancestral Pteropodidae vs <i>P. parnellii</i>	T 105 A (0.69)	NA
<i>P. vampyrus</i> vs <i>P. parnellii</i>	None	NA

Abbreviations: BPP, Bayesian posterior probability; NA, not applicable; PP, posterior probability. Significance levels of branch-wise convergence are based on 1000 simulated data sets. Amino-acid numbers refer to the human peptide.

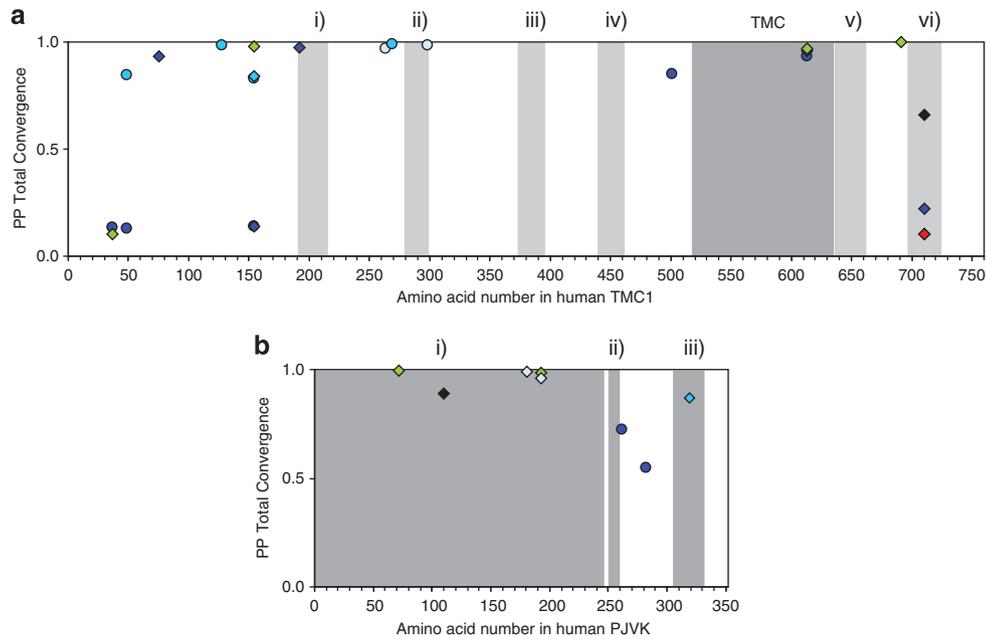


Figure 3 Distribution of sites along (a) TMC1 and (b) PJKV with posterior probability (PP) of convergent substitutions, >0.1 , for *T. truncatus*-bat comparisons (diamonds) and *R. ferrumequinum*-Yangochiroptera comparisons (circles). Colours are according to Figure 2a. Functional peptide domains of each protein were identified from published sources; in (a) the TMC1 transmembrane domains are numbered (i-vi) and shown as pale grey blocks, and the highly conserved TMC domain (TMC) is shown as a dark grey block. In (b) dark grey blocks indicate the (i) highly conserved Gasdermin domain, (ii) the putative nuclear localisation signal and (iii) the zinc-binding motif.

Table 2 Amino-acid sites identified as undergoing convergent substitutions in (a) *Tmc1* and (b) *Pjvk*

<i>Tmc1</i> AA site	<i>Yinpterochiroptera</i>	<i>Est. ω</i> (PP third site class)	<i>Yangochiroptera</i>	<i>Est. ω</i> (PP third site class)	PP convergence	Branch-wise significance
(a)						
I 481 L	Hipposideridae	4.49 (0.93)	Vespertilionidae	1.17 (0.94)	0.88	0.036
G 500 A	Rhinolophidae and Hipposideridae	2.39 (0.96)	Internal Phyllostomidae	0.82 (0.93)	0.98	<0.001
F 504 Y	Internal Hipposideridae	4.70 (0.99)	Internal Kerivoulinae	2.09 (0.77)	0.79	<0.001
<i>Pjvk</i> AA site	<i>Yinpterochiroptera</i>	<i>Est. ω</i> (PP third site class)	<i>Yangochiroptera</i>	<i>Est. ω</i> (PP third site class)	PP convergence	Branch-wise significance
(b)						
L 245 F	Hipposideridae	1.02 (0.93)	Phyllostomidae	0.63 (0.89)	0.99	0.002
A 262 V	Rhinolophidae and Hipposideridae	0.94 (0.92)	Phyllostomidae	0.64 (0.89)	0.92	0.018
A 262 V	Rhinolophidae and Hipposideridae	0.94 (0.92)	Vespertilionidae	0.62 (0.88)	0.92	0.014
F 282 L	Hipposideridae	1.04 (0.90)	Internal Kerivoulinae	0.54 (0.83)	0.94	0.012

Again, only sites with posterior probability (PP) >0.5 are shown, and branch-wise significance levels are also shown. Estimated (*Est.*) ω values are calculated using clade models and the posterior probability of the site being in the third site class is shown in brackets.

Focussing on bats with a range of echolocation call types, we repeated the calculations of the posterior probabilities of convergence for all ancestral branch pairs in our wider taxonomic data set of partial sequences. For comparisons among bats, evidence of high branch-wise convergence was restricted to echolocating members of the two suborders, whereas only weak evidence of convergence was found among lineages within either the Yangochiroptera or Yinpterochiroptera (see Supplementary Figure S4). These cases were found to be driven by three amino acids with high site-wise convergence posterior probabilities (>0.5 ; see Table 2a).

For *Pjvk*, plots of branch-wise probabilities of convergence for all placental mammals, based on the complete gene sequence, revealed fewer cases of convergence among echolocating taxa (Figure 2b). Similarly, divergence probabilities were also lower than for *Tmc1*, probably reflecting the overall greater sequence conservation. Among echolocating bats, three significant cases of convergence were found, between *R. ferrumequinum* and the ancestral Yangochiroptera, *R. ferrumequinum* and *M. lucifugus* and *R. ferrumequinum* and *P. parnellii*. The two former cases were characterised by two convergent sites and one convergent site, respectively (Table 1b). In comparison,

greater convergence was detected between the echolocating bats and the dolphin, of which the highest pair-wise convergence values were recorded between *T. truncatus* and both *R. ferrumequinum* and *M. lucifugus* associated with two sites in each case (Figure 2b and Table 1b). Weaker evidence of branch-wise convergence was seen between *T. truncatus* and the ancestral branches of the main suborders and all bats. In total, three sites were identified with a high probability of convergence between the two echolocating bat suborders, whereas, only one was identified between Yangochiroptera and the Old World fruit bats (Table 1b). The majority of the Pjvk protein is made up of the highly conserved Gasdermin domain (see Figure 3b), which was seen to contain a total of five sites driving bat–dolphin and bat–bat convergence. Once again, all convergent sites were parallel changes.

Posterior probabilities of convergence calculated among branches in the wider taxonomic data set revealed a similar small number of high values between echolocating bat branches as between echolocating bats and other taxa (see Supplementary Figure S4). The former case was associated with four amino acids with high site-wise convergence posterior probabilities (>0.5 ; see Table 2b).

Tests for selection

For *Tmc1*, branch-site selection models based on the complete sequences detected positive selection in both clades of echolocating bats (Supplementary Table S3a and b). First, in the Yangochiroptera, positive selection was found on the ancestral branch ($\omega=9.36$, LRT=4.83, d.f.=1, $P=0.028$) with a total of 16 sites identified from Bayes empirical Bayes analyses. Positive selection was also detected on the *M. lucifugus* branch ($\omega=33.36$, LRT=5.53, d.f.=1, $P=0.019$) with eight sites identified, but not on the *P. parnellii* branch. Second, in the Yinpterochiroptera, positive selection was detected on the branch leading to the echolocating species *R. ferrumequinum* ($\omega=3.88$, LRT=5.20, d.f.=1, $P=0.023$), with a total of 25 sites identified. Positive selection was not detected on the branches leading to either all Yinpterochiroptera or the dolphin.

Clade models for *Tmc1* using the partial sequences from the wider taxonomic sample were conducted to test hypotheses of divergent selection between echolocating and non-echolocating taxa. Significant divergent selection pressures were detected in all focal groups examined, as indicated by better fit than their corresponding M1a models (Supplementary Table S4a). Estimates of ω were >1 (positive selection) for $\sim 16\%$ of sites when the foreground clade was defined as either all echolocating species of Yinpterochiroptera ($\omega=1.37$) or all Rhinolophidae and Hipposideridae ($\omega=2.48$). In comparison, the foreground ω was ~ 0 (purifying selection) in non-echolocating fruit bats and 0.86 (purifying selection) in the Yangochiroptera. In echolocating cetaceans, our clade model also showed evidence of positive selection ($\omega=1.13$) at 16% of sites.

For the complete coding sequence of Pjvk, branch-site models revealed no evidence of positive selection on any of the bat or dolphin branches tested (Supplementary Table S3c and d). Clade models based on the partial sequences once again revealed divergent selection pressures in all groups of echolocating taxa examined (Supplementary Table S4b). However, for the bats, ω values estimated for these divergent sites typically ranged from zero (purifying) to one (neutral), although we cannot rule out the possibility of a burst of positive selection in the past. For the echolocating whales, the clade model revealed strong positive selection ($\omega=7.40$), detected at 16% of sites.

Association between convergence and positive selection

Comparing the results of convergence and selection tests revealed some clear patterns. Based on the complete *Tmc1* sequences, eight of

the sites identified as undergoing convergent substitutions between echolocating taxa were also found to be under positive selection in at least one of the taxa concerned (Figure 4a). Similarly, from the *Tmc1* partial sequences, sites identified with high posterior probabilities of convergent substitutions corresponded to positive selection pressures in at least one of the clades involved (Supplementary Table S4a and Supplementary Figure S6a). Finally, separate *Tmc1* clade models conducted for key groups of bats, in which echolocating taxa were excluded from the background, revealed evidence of positive selection in four main groups (Supplementary Figure S5). In each of these models, sites with the ω values of >1 were seen to correspond broadly to high posterior probabilities of convergence.

For *Pjvk*, branch-site models of complete sequences showed no evidence of significant positive selection, and hence it is not possible to compare selection pressures with convergence in these taxa (Figure 4b). From the clade models of partial sequences, two sites in the Hipposideridae were estimated to have ω values marginally above one (Supplementary Table S4b and Supplementary Figure S6b); however, no sites were found to be under positive selection in any Yangochiroptera species.

DISCUSSION

Results from two independent hearing genes, *Tmc1* and *Pjvk*, revealed strong evidence of positive selection in echolocating whales, and also in some echolocating bats; however, we found no evidence of positive selection acting on the ancestral bat branch in either gene.

In *Tmc1*, positive selection was detected along the branch leading to the Yangochiroptera, and also to the echolocating Yinpterochiroptera. These trends were further supported by clade models performed on a greater number of taxa. We also detected evidence that convergent substitutions between echolocating taxa in the *Tmc1* gene corresponded to sites under selection. Thus, these results support our expectation that molecular adaptation in the hair cell protein Tmc1 is associated with the evolution of high-frequency hearing. The functional significance of the convergent substitutions in *Tmc1* is currently unknown; however, most occurred in either intra- or extra-cellular residues (Keresztes *et al.*, 2003), and thus these substitutions may well have an adaptive role. No obvious sequence convergence was found between dolphin and non-echolocating fruit bats. Taken together, these results from *Tmc1* suggest that functional adaptations for high-frequency hearing arose after the split between the Yangochiroptera and Yinpterochiroptera, hence supporting multiple independent origins of laryngeal echolocation.

In *Pjvk*, we found no evidence of positive selection in either bat branches or clades; however, strong positive selection was detected in the clade comprising four echolocating cetaceans. Compared with *Tmc1*, levels of convergence among echolocating bats, and between bats and cetaceans, were lower. Although the exact function of Pjvk remains unknown, it has been predicted to contain a zinc-binding motif and also the highly conserved Gasdermin domain and, therefore, of the two proteins studied here, Pjvk is likely to be under higher structural constraint.

In addition to the convergence observed between the two suborders of bats, both genes also showed evidence of convergent amino-acid substitutions between specific clades of echolocating bats. In particular, comparisons of the family Hipposideridae versus the subfamily Kerivoulinae, and also between the families Rhinolophidae versus Phyllostomidae, showed increased levels of sequence convergence. These results are especially intriguing given that species of *Hipposideros* and *Kerivoula* possess the highest echolocation call frequencies recorded to date (Fenton and Bell, 1981; Schmieider

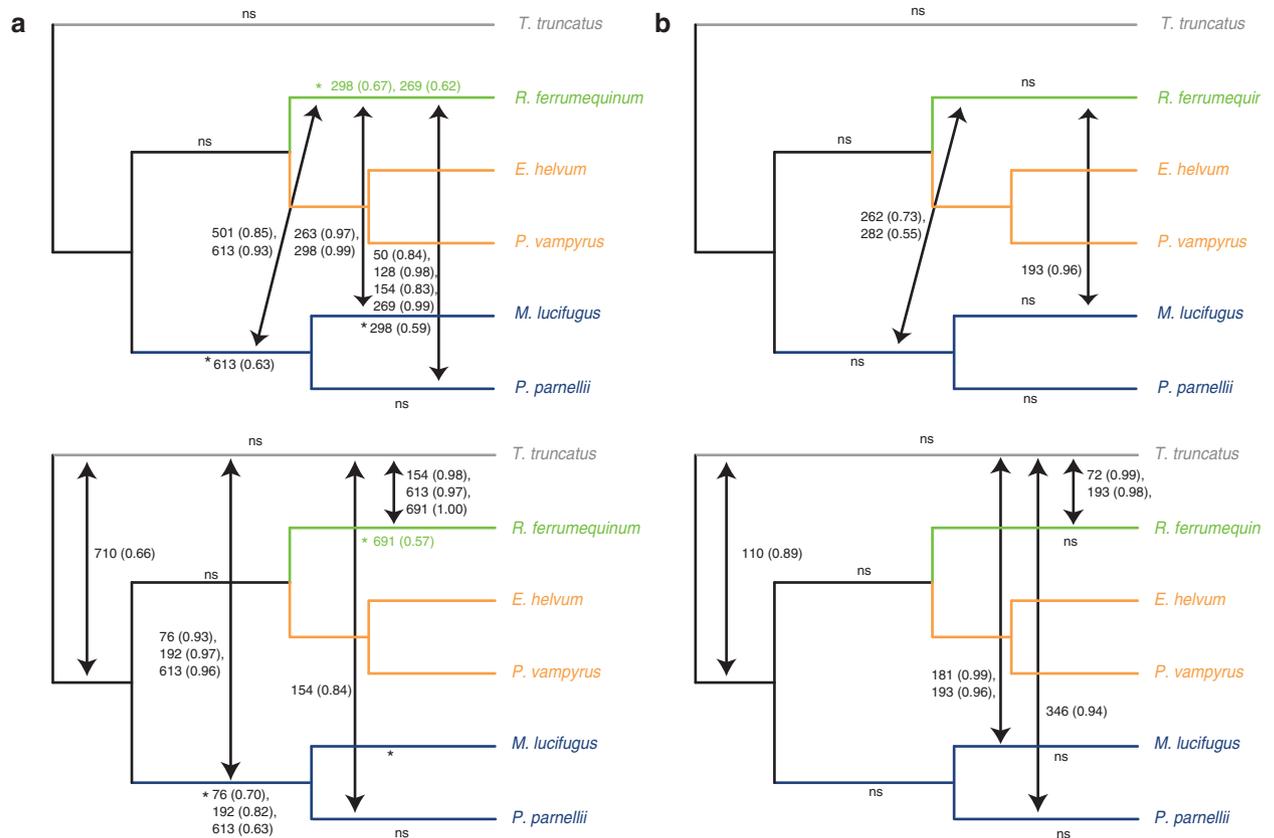


Figure 4 Positive selection and convergent substitutions in (a) *Tmc1* and (b) *Pjvk*. Sites with probability of convergent substitutions, in parentheses, are shown in black text; arrows indicate branches that share the convergence. Significant branch-site models are indicated by coloured text, site-wise probability of being under positive selection shown in parentheses. Sites are listed if identified as undergoing convergent substitutions and positive selection. For each gene the upper tree displays convergence between echolocating Yinypterochiroptera and Yangochiroptera, and the lower between the dolphin and echolocating bats. NS, not significant; * $P < 0.05$.

et al., 2010). Additionally, along *Tmc1* some of the best evidence of sequence convergence was seen between *R. ferrumequinum* and *P. parnellii*, two bats that are known to have independently evolved CF echolocation.

In both genes, the small number of detected convergent changes were sufficiently abundant to cause conflicts between gene trees (in which all echolocating bats cluster together) and the known species tree, in which they are paraphyletic. However, this convergence did not lead to any grouping of bats with toothed whales. Phylogenetic incongruence between data sets is not uncommon, particularly where sequence length is limited, as is the case for the wider taxonomic study. However, the specific nature of observed conflict was also evident from the longer sequences, and also from the Lento plots, both suggesting that these results were not simply because of a lack of power. These results endorse the utility of intronic DNA as putative neutral markers for recovering species relationships (see, for example, Corte-Real *et al.*, 1994), but at the same time add weight to previous warnings about the potential pitfalls of using loci under selection for phylogenetic reconstruction (Li *et al.*, 2008; Castoe *et al.*, 2009).

These results show remarkable similarities with published findings from the *Prestin* gene. All three genes encode proteins that are expressed in the cochlea and implicated in mammalian hearing, and all have mutant forms that have been linked to non-syndromic hearing loss in humans and/or mice. The proteins *Tmc1* and *Pjvk*

have vital roles in hair cell development and function, respectively, whereas *Prestin* is thought to specifically drive the motility of the outer hair cells on the basilar membrane (Zheng *et al.*, 2000; Marcotti *et al.*, 2006; Schwander *et al.*, 2007).

Echolocating mammals as potential molecular models of hearing

Given their ability to echolocate, it is perhaps not surprising that bats have long served as important models for understanding the neurophysiology of auditory processing (see, for example, Kossel *et al.*, 2003). The data of sequence convergence between taxa with ultrasonic hearing in three separate hair cell genes suggest that echolocating mammals might be equally useful for unravelling the molecular basis of hearing.

Previous studies that amplified *Tmc1* and its paralogs have found high sequence conservation across mammals, probably reflecting the TMC domain (Kurima *et al.*, 2002). It is thus noteworthy that most of the observed variable sites among the bat sequences, as well as an insertion of three amino acids in two families of bats, were located in a region homologous to the mouse exon 14, which is deleted in mutant mice with recessive deafness (Kurima *et al.*, 2002). Therefore, *Tmc1* is likely to be both important in basic hearing and a target for evolutionary adaptations for high-frequency hearing. Like *Tmc1*, published mammalian *Pjvk* gene sequences show strong amino-acid conservation (Schwander *et al.*, 2007). Our results from bats support this trend, with little amino-acid variation found in the putative nuclear localisation signal domain and zinc-binding motif (Delmaghani *et al.*, 2006).

Sequence convergence across multiple functional genes

To our knowledge, the occurrence of parallel signatures of adaptive sequence convergence across three independent loci that encode similar gene products (that is, hair cell proteins) is unique. Reported cases of convergent sequence evolution in which changes appear related to gene function are uncommon, and have nearly always focussed on single loci, often involving only a few amino acids. Examples of these include the convergent homologous sites in the visual pigments of squid and primates (Morris *et al.*, 1993), and also in the myoglobin gene of seals and cetaceans (Romeroherrera *et al.*, 1978). More extensive sequence convergence operating across several loci has been documented in the mitochondrial genomes of snakes and agamid lizards; however, mtDNA genes cannot be considered to represent independent loci, and the suggested potential adaptive role of such convergence in metabolism remains speculation (Castoe *et al.*, 2009). Several different toxin protein genes have also been shown to have undergone convergent evolution, among different frog species (Roelants *et al.*, 2010), as well as between shrews and lizards (Aminetzach *et al.*, 2009), although the latter is structural as opposed to sequence convergence. In general, surprisingly few studies that have described sequence convergence have also tested for selection and, therefore, have not explicitly been able to rule out nonadaptive homoplasy that is widespread in nature (Rokas and Carroll, 2008).

The finding that the monophyly of echolocating bats was recovered by coding regions highlights the adaptive nature of the observed convergence. Examination of the relative site-wise support for the convergent topology (data not shown), as well as the Lento plots, indicated that most support for the 'convergent topologies' was concentrated within exonic regions, whereas phylogenetic signal for the true species relationships was more obvious in intronic sections. Together with other very recent research (Li *et al.*, 2008; Rokas and Carroll, 2008; Castoe *et al.*, 2009; Li *et al.*, 2010; Liu *et al.*, 2010a), the data suggest that sequence convergence might be much more common than previously thought.

Implications for the evolution of echolocation

Although we did not aim to tackle the specific issue of whether echolocation in bats evolved more than once, the relatively high number of sites with elevated probabilities of convergence between the two main clades of echolocating bats, and also between echolocating bats and the bottlenose dolphin, could be considered as favouring multiple origins of laryngeal echolocation. Moreover, selection models showed positive selection acting only on echolocating bat species, with evidence of purifying selection rather than relaxation in Old World fruit bats. Together with *Prestin*, our results from *Tmcl* and *Pjvk* appear to indicate that functional adaptations for high-frequency hearing arose after the split between the Yangochiroptera and Yinpterochiroptera. This scenario supports at least two independent origins of high-frequency hearing (and possibly echolocation) in bats. Although these conclusions fit recent claims that the earliest fossil bat (*Onychonycteris finneyi*) could not echolocate, other Eocene fossil bats do show some traits associated with echolocation (Simmons *et al.*, 2008). Consequently, more data are needed to reconcile apparent conflicts between fossil and molecular evidence.

DATA ARCHIVING

All newly generated sequence data are available from GenBank; accession numbers JN898964–JN899071.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Michael Bekaert and Georgia Tsagkogeorga for technical help. We thank staff at the Genome Centre of the Barts and the London School of Royal Medicine and Dentistry for support. This study was funded by a Natural Environmental Research Council (UK) studentship awarded to KTJD, with additional support from the Central Research Fund (London) and Systematic Research Fund, and a Science Foundation Ireland PIYRA grant to ECT. JAC was funded by an RCUK academic fellowship. SJR was funded by a Royal Society Fellowship (UK). We are grateful to the Institute of Zoology and Taipei Zoo for donating tissue samples.

- Accetturo M, Creanza TM, Santoro C, Tria G, Giordano A, Battagliero S *et al.* (2010). Finding new genes for non-syndromic hearing loss through an in silico prioritization study. *PLoS ONE* **5**: e12742.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Aminetzach YT, Srouji JR, Kong CY, Hoekstra HE (2009). Convergent evolution of novel protein function in shrew and lizard venom. *Curr Biol* **19**: 1925–1931.
- Bekaert M, Teeling EC (2008). UniPrime: a workflow-based platform for improved universal primer design. *Nucleic Acids Res* **36**: e56.
- Bielawski JP, Yang ZH (2004). A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol* **59**: 121–132.
- Castoe TA, de Koning APJ, Kim HM, Gu WJ, Noonan BP, Naylor G *et al.* (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci USA* **106**: 8986–8991.
- Christin PA, Weinreich DM, Besnard G (2010). Causes and evolutionary significance of genetic convergence. *Trends Genet* **26**: 400–405.
- Corte-Real HBSM, Dixon DR, Holland PWH (1994). Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. *Mar Biol* **120**: 407–413.
- Csorba G, Ujhelyi P, Thomas N (2003). *Horseshoe bats of the world (Chiroptera: Rhinolophidae)*. Alana Books: Shropshire, UK.
- Delmaghani S, del Castillo FJ, Michel V, Leibovici M, Aghaie A, Ron U *et al.* (2006). Mutations in the gene encoding pejvak, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy. *Nat Genet* **38**: 770–778.
- Dror AA, Avraham KB (2010). Hearing impairment: a panoply of genes and functions. *Neuron* **68**: 293–308.
- Fenton MB, Bell GP (1981). Recognition of species of insectivorous bats by their echolocation calls. *J Mammal* **62**: 233–243.
- Huber KT, Langton M, Penny D, Moulton V, Hendy M (2002). Spectronet: a package for computing spectra and median networks. *Appl Bioinformatics* **1**: 159–161.
- Jones G, Teeling EC (2006). The evolution of echolocation in bats. *Trends Ecol Evol* **21**: 149–156.
- Keresztes G, Mutai H, Heller S (2003). TMC and EVER genes belong to a larger novel family, the TMC gene family encoding transmembrane proteins. *BMC Genomics* **4**.
- Khan FAA, Solari S, Swier VJ, Larsen PA, Abdullah MT, Baker RJ (2010). Systematics of Malaysian woolly bats (Vespertilionidae: *Kerivoula*) inferred from mitochondrial, nuclear, karyotypic, and morphological data. *J Mammal* **91**: 1058–1072.
- Kitajiri SI, McNamara R, Makishima T, Husunain T, Zafar AU, Kittles RA *et al.* (2007). Identities, frequencies and origins of *TMC1* mutations causing DFNB7/B11 deafness in Pakistan. *Clin Genet* **72**: 546–550.
- Kossli M, Foeller E, Drexler M, Vater M, Mora E, Coro F *et al.* (2003). Postnatal development of cochlear function in the mustached bat, *Pteronotus parnellii*. *J Neurophysiol* **90**: 2261–2273.
- Kurima K, Peters LM, Yang YD, Riazuddin S, Ahmed ZM, Naz S *et al.* (2002). Dominant and recessive deafness caused by mutations of a novel gene, *TMC1*, required for cochlear hair-cell function. *Nat Genet* **30**: 277–284.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Li G, Wang JH, Rossiter SJ, Jones G, Cotton JA, Zhang SY (2008). The hearing gene *Prestin* reunites echolocating bats. *Proc Natl Acad Sci USA* **105**: 13959–13964.
- Li Y, Liu Z, Shi P, Zhang JZ (2010). The hearing gene *Prestin* unites echolocating bats and whales. *Curr Biol* **20**: R55–R56.
- Liu Y, Cotton JA, Shen B, Han XQ, Rossiter SJ, Zhang SY (2010a). Convergent sequence evolution between echolocating bats and dolphins. *Curr Biol* **20**: R53–R54.
- Liu Y, Rossiter SJ, Han X, Cotton JA, Zhang S (2010b). Cetaceans on a molecular fast track to ultrasonic hearing. *Curr Biol* **20**: 1834–1839.
- Marcotti W, Erven A, Johnson SL, Steel KP, Kros CJ (2006). *Tmc1* is necessary for normal functional maturation and survival of inner and outer hair cells in the mouse cochlea. *J Physiol* **574**: 677–698.
- McGowen MR, Spaulding M, Gatesy J (2009). Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol* **53**: 891–906.

- Miller-Butterworth CM, Murphy WJ, O'Brien SJ, Jacobs DS, Springer MS, Teeling EC (2007). A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, *Miniopterus*. *Mol Biol Evol* **24**: 1553–1561.
- Morris A, Bowmaker JK, Hunt DM (1993). The molecular-basis of a spectral shift in the rhodopsins of 2 species of squid from different photic environments. *Proc R Soc Biol Sci B* **254**: 233–240.
- Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W (2007). Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* **17**: 413–421.
- Nevo E (1979). Adaptive convergence and divergence of subterranean mammals. *Annu Rev Ecol Syst* **10**: 269–308.
- Nishihara H, Hasegawa M, Okada N (2006). Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci USA* **103**: 9929–9934.
- Packard A (1972). Cephalopods and fish - limits of convergence. *Biol Rev Camb Philos Soc* **47**: 241.
- R Development Core Team (2011). R: A language and environment for statistical computing, Vienna, Austria.
- Roelants K, Fry BG, Norman JA, Clynen E, Schoofs L, Bossuyt F (2010). Identical skin toxins by convergent molecular adaptation in frogs. *Curr Biol* **20**: 125–130.
- Rokas A, Carroll SB (2008). Frequent and widespread parallel evolution of protein sequences. *Mol Biol Evol* **25**: 1943–1953.
- Romero-Herrera AE, Lehmann H, Joysey KA, Friday AE (1978). On the evolution of myoglobin. *Phil Trans R Soc Lond B* **283**: 61–163.
- Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Rossiter SJ, Zhang S, Liu Y (2011). *Prestin* and high-frequency hearing in mammals. *Comm Int Biol* **4**: 236–239.
- Schmiieder DA, Kingston T, Hashim R, Siemers BM (2010). Breaking the trade-off: rainforest bats maximize bandwidth and repetition rate of echolocation calls as they approach prey. *Biol Lett* **6**: 604–609.
- Schultz J, Milpetz F, Bork P, Ponting CP (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**: 5857–5864.
- Schwander M, Sczaniecka A, Grillet N, Bailey JS, Avenarius M, Najmabadi H *et al.* (2007). A forward genetics screen in mice identifies recessive deafness traits and reveals that pejkakin is essential for outer hair cell function. *J Neurosci* **27**: 2163–2175.
- Shimodaira H (2002). An approximately unbiased test of phylogenetic tree selection. *Syst Biol* **51**: 492–508.
- Simmons NB, Seymour KL, Habersetzer J, Gunnell GF (2008). Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* **451**: 818–U816.
- Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stewart CB, Schilling JW, Wilson AC (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* **330**: 401–404.
- Teeling EC, Madsen O, Van den Bussche RA, de Jong WW, Stanhope MJ, Springer MS (2002). Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proc Natl Acad Sci USA* **99**: 1431–1436.
- Vater M, Kossli M (2004). Introduction: The ears of whales and bats. In: Thomas JA, Moss CF, Vater M (eds). *Echolocation in Bats and Dolphins*. The University of Chicago Press: Chicago and London. pp 89–98.
- Yang ZH (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
- Yeager M, Kumar S, Hughes AL (1997). Sequence convergence in the peptide-binding region of primate and rodent MBC class Ib molecules. *Mol Biol Evol* **14**: 1035–1041.
- Zhang JZ, Nielsen R, Yang ZH (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* **22**: 2472–2479.
- Zheng J, Shen WX, He DZZ, Kevin BL, Madison LD, Dallos P (2000). Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**: 149–155.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)