The genomes of two bat species with long constant frequency echolocation calls

Dong Dong[1^]*, Ming Lei[1^], Panyu Hua[1], Yi-Hsuan Pan[1], Shuo Mu[1], Guantao Zheng[1], Erli Pang[2], Kui Lin[2] and Shuyi Zhang[3]

[1] Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai 200241, China;

[2] School of Life Sciences, Beijing Normal University, Beijing 100875, China;

[3] Key Laboratory of Zoonosis of Liaoning Province, College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, Shenyang 110866, China.

[^] These authors contributed equally to this work

*To whom correspondence should be addressed. Email: ddong.ecnu@gmail.com

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Abstract

Bats can perceive the world by using a wide range of sensory systems, and some of the systems have become highly specialized, such as auditory sensory perception. Among bat species, the Old World leaf-nosed bats and horseshoe bats (rhinolophoid bats) possess the most sophisticated echolocation systems. Here, we reported the whole-genome sequencing and *de novo* assembles of two rhinolophoid bats – the great leaf-nosed bat (*Hipposideros armiger*) and the Chinese rufous horseshoe bat (*Rhinolophus sinicus*). Comparative genomic analyses revealed the adaptation of auditory sensory perception in the rhinolophoid bat lineages, probably resulting from the extreme selectivity used in the auditory processing by these bats. Pseudogenization of some vision-related genes in rhinolophoid bats was observed, suggesting that these genes have undergone relaxed natural selection. An extensive contraction of olfactory receptor gene repertoires was observed in the lineage leading to the common ancestor of bats. Further extensive gene contractions can be observed in the branch leading to the rhinolophoid bats. Such concordance suggested that molecular changes at one sensory gene might have direct consequences for genes controlling for other sensory modalities. To characterize the population genetic structure and patterns of evolution, we re-sequenced the genome of 20 great leaf-nosed bats from four different geographical locations of China. The result showed similar sequence diversity values and little differentiation among populations. Moreover, evidence of genetic adaptations to high altitudes in the great leaf-nosed bats was observed. Taken together, our work provided a useful resource for future research on the evolution of bats.

Introduction

Bats (order Chiroptera) are one of the largest monophyletic clades in mammals (order Chiroptera), and constitute nearly 20% of living mammalian species. They can perceive their surroundings using a wide range of sensory systems, and have long been regarded as the most unusual and specialized species of all mammals. Most bats are sophisticated echolocators and rely on their echolocation systems for navigation. However, Old World fruit bats have no laryngeal echolocating ability, and navigate largely by vision. Based on overwhelming molecular genetic evidence, it has been proposed that echolocating bats are paraphyletic (Teeling, et al. 2005). Old World fruit bats and some laryngeal echolocators (including Rhinolophidae, Hipposideridae, Craseonycteridae, Megadermatidae, and Rhinopomatidae families) are a natural group – the suborder Yinpterochiroptera, and the remaining laryngeal echolocating bats are grouped to another suborder Yangochiroptera .

Two distinct navigation approaches can be employed by echolocating bats: low duty cycle (LDC) echolocation and high duty cycle (HDC) echolocation (Teeling 2009). LDC echolocators can separate pulse and echo in time to avoid forward masking, whereas some species of HDC echolocators separate pulse and echo in frequency. It has been documented that rhinolophoid bats might possess the most sophisticated echolocation systems (Jones and Teeling 2006). Recently, results from some hearing-related genes suggested sequence convergence in laryngeal echolocating bats (Li, et al. 2008; Davies, et al. 2012). We attempted to investigate whether similar patterns can be detectable in other hearing-related genes. Furthermore, a sensory trade-off between investment in vision and echolocation has been identified (Dechmann and Safi 2009). Loss-of-function in short-wave sensitive opsin (*Sws1* gene) occurred in rhinolophoid bats, which use HDC echolocation and can emit long constant frequency calls (Zhao, et al. 2009). Although several bat genomes have been sequenced (Zhang, et al. 2013), the evolutionary mechanisms of the rhinolophoid bats remains unclear. Comparative genomics will provide us opportunities to investigate

whether similar patterns can be detectable in other sensory genes.

The great leaf-nosed bat (*Hipposideros armiger*) and the Chinese rufous horseshoe bat (*Rhinolophus sinicus*) are two important species of rhinolophoid bats. First, these are model organisms with remarkable HDC echolocation ability and can emit continuous ultrasonic calls of long constant frequency with remarkable acoustic features (Doppler-shift compensation) (Schnitzler, et al. 2003). We can comprehensively explore how rhinolophoid bats evolved a specialized form of echolocation. Second, they are important reservoir hosts of emerging viruses, and the Chinese rufous horseshoe bat has been suggested to carry the direct ancestor of severe acute respiratory syndrome (SARS) coronavirus (Ge, et al. 2013). In this work, we presented the genomes of the great leaf-nosed bat and the Chinese rufous horseshoe bat using the next generation sequencing platform (Illumina Hiseq 2500). The result revealed the adaptation of auditory sensory perception in HDC echolocators, and showed an extensive contraction of olfactory receptor gene repertoires as well as pseudogenization of some vision-related genes. Furthermore, we performed genome re-sequencing to analyze the population genetic structure of the great leaf-nosed bats. The genomic data provide genetic evidence of adaptive evolution in rhinolophoid bats.

Materials and Methods

Genome sequencing and *de novo* assembly

A female great leaf-nosed bat (*Hipposideros armiger*) and a female Chinese rufous horseshoe bat (*Rhinolophus sinicus*) were captured from a cave (N 30°20.497′ E 117°50.104′) in Anhui Province, China on September 24, 2011. The bats were euthanized by respiratory hyperanesthesia with isoflurane, and then we performed a thoracotomy on anesthetized bats. The tissues were stored in liquid nitrogen and transported to the sequencing center. All procedures were performed in accordance with the guidelines of Regulations for the Administration of Laboratory Animals (Decree No. 2 State Science and Technology Commission of the People's Republic of China). This study was approved by the Animal Ethics Committee of East China Normal University (ID no: 20090219).

Genomic DNA was extracted from bat muscle using the Qiagen DNeasy Blood and Tissue Kit. Six paired-end libraries with insert size of 170 bp, 500 bp, 800 bp, 3k bp, 8k bp and 20k bp were constructed and sequenced for the great leaf-nosed bat and the Chinese rufous horseshoe bat, respectively. The libraries were sequenced using Illumina HiSeq2500 platform, which has a read length of 101 bp. Low quality sequencing reads were filtered out and potential sequencing errors were removed. The following filtering criteria were carried out: 1) Filter reads with >5% unidentified nucleotides; 2) Filter reads with >10 nucleotides aligned to the adapter sequence, allowing <3 mismatches; 3) remove putative PCR duplicates generated by PCR amplification in the library construction process. Finally, we generated 476.5 Gb and 288.5 Gb of sequences for the great leaf-nosed bat and the Chinese rufous horseshoe bat, respectively.

The genome sequences were assembled using ALLPATHS software (Butler, et al. 2008). Briefly, contigs were generated by constructing a *de Bruijin* graph with the sequencing reads from short-insert library data. The graph was simplified to generate

the contigs by removing tips, merging bubbles and solving repeats. The sequencing reads were mapped to the assembled contigs, and the scaffolds were constructed by weighting the rates of consistent and conflicting paired-end relationships. At last, we retrieved the read pairs with one end that uniquely mapped to the contigs and the other end located in the gap region, a local assembly for these collected reads was performed to fill the gaps. A more detailed genome assembly method is provided in Supplementary methods.

Transcriptomic sequencing

Total RNAs of the two bats were extracted from brain, cerebellum, heart, liver, stomach, kidney, lung and muscle tissues for the generation of transcriptome data. Paired-end libraries for RNA sequencing were constructed using the Illumina mRNA-seq Prep Kit. The quality and integrity of the RNA samples were determined using the Agilent2100 Bioanalyzer. Poly(A) mRNAs were isolated using oligo(dT) beads, fragmented, and converted to cDNAs followed by end repair, adaptor ligation, and PCR amplification. The libraries thus generated were sequenced using the Illumina HiSeq2500 platform as described above.

Genome annotation

We searched for tandem repeats across the genomes using Tandem Repeats Finder. Transposable elements were predicted in the genomes by homology search against the known transposable elements (TE) in RepBase (Jurka, et al. 2005) (version 20110920) using RepeatMasker version 3.3.0 (Tarailo-Graovac and Chen 2009). The protein-coding genes of the bat genomes were annotated by combining homology-based, *ab initio* and RNA-seq gene prediction methods. At first, RNA-seq data were assembled using the Trinity package (Trapnell, et al. 2013). PASA (version r2012-06-25) (Haas, et al. 2003) was then used to map the assembled transcripts. Based on the set of gene models, a training set was constructed for *de novo* predictors by selecting the genes with complete structures and at least 100% mapping rate for UniProt vertebrate proteins. For the *ab initio* prediction, Augustus (Stanke and Waack

2003) and GenScan (Burge and Karlin 1997) were used to predict protein-coding genes with the training set generated by PASA. For homology-based gene prediction, the protein sequences of human, mouse, dog, cow, little brown bat and large flying fox were downloaded from Ensembl Release 72 and mapped onto the repeat-masked genome using GenBlastA (She, et al. 2009). RNA-seq data were mapped to the genome using TopHat (Trapnell, et al. 2009), and the transcription-based gene structure were generated by Cufflinks (Trapnell, et al. 2013). The final gene set was generated by merging all genes predicted using GLEAN software (http://sourceforge.net/projects/glean-gene/). To infer gene function, it was based on the best match of the alignment to the Swissprot and translated EMBL nucleotide sequence data library databases using BLASTP. InterproScan (Mulder and Apweiler 2007) was used to determine motifs and domains in the final gene set. To evaluate completeness of the genomes and annotations, CEGMA method (Parra, et al. 2007) was employed.

Gene families

We used the TreeFam methodology (Li, et al. 2006) to define gene families in 14 mammalian genomes (human, macaque, mouse, rat, dog, cat, horse, rhinoceros, cow, pig, little brown bat, large flying fox, great leaf-nosed bat and Chinese rufous horseshoe bat). The protein sequences of other 12 mammalian species were obtained from Ensembl database (Release 72). Gene family expansion and contraction analysis was performed by CAFÉ software (De Bie, et al. 2006). A random birth and death model was proposed to study gene gain and loss in the gene families across a user-specified phylogenetic tree. A global parameter λ (lambda), which described both gene birth (λ) and death ($\mu = -\lambda$) rates across all branches of all gene families was estimated using the maximum likelihood method. A conditional p-value was calculated for each gene family, and families with conditional p-values less than 0.05 were considered to have a significantly accelerated rate of expansion and contraction.

Evolutionary analyses

Protein sequences of the aforementioned 14 mammals were aligned using MUSCLE software (Edgar 2004). All orthologous genes were concatenated to one super gene for each species. RAxML (Stamatakis 2014) was applied to build phylogenetic trees. We partitioned the data by coding genes, and evaluated the model parameter independently for each partition. In all partitioned analyses, the empirical base frequencies and the evolutionary rates were estimated independently for every partition. Bootstrap support was obtained by repeating the original partitioned ML RAxML analysis on 100 bootstrap replicates for each dataset using different random number seeds in each repetition. Next, we inferred the species tree using coalescent method: maximum pseudo-likelihood estimation of species tree (MP-EST) (Liu, Yu, et al. 2010). Individual gene tree for each gene was estimated using the maximum-likelihood method and rooted by an outgroup (human). Species trees were estimated from the rooted gene trees in the program MP-EST with 100 bootstrap replicates. The results supported that bats are member of scrotifera (Chiroptera + Carnivores + Perissodactyla + Cetartiodactyla) with bat lineage diverging from Fereuungulata (Carnivores + Perissodactyla + Cetartiodactyla).

The values of *K*a, *K*s, the *K*a/*K*s ratio were estimated for each gene using the Codeml programs nested in the PAML package (Yang 2007). In order to detect positively selected genes, optimized branch-site likelihood model (Zhang, et al. 2005) was used. We separately explored the positively selected genes in the great leaf-nosed bat and the Chinese rufous horseshoe bat. For each analysis, only one bat species was selected as foreground branches, and all other species were regarded as the background branches. The revised branch-site model A was employed, which attempts to identify positive selection acting on some sites on the "foreground branches". Using an likelihood ratio test (LRT), the alternative hypothesis that positive selection occurs on the foreground branches (*K*a/*K*s > 1) is compared with the null hypothesis (*K*a/*K*s=1). Bayesian empirical Bayes values were used to identify sites under positive selection. Then, branch two-ratio model was applied to detect accelerated evolved genes in specific lineage. The one-ratio model assumed an equal *K*a/*K*s ratio for all

lineages in the phylogeny, and the two-ratio model assumed two *K*a/*K*s ratios: one branch for the background, one for the foreground branch leading to the specific species. Then, Clade model C was employed to test for positive selection along the rhinolophoid bats. The two clades were assumed to share sites under purifying selection and neutral evolution, but to differ at a third site partitions under divergent selection. The null model used for the Clade model C was M2a_rel (Weadick and Chang 2012), whose LRT has a relatively lower false-positive rate. GO annotations were downloaded from Ensembl databases and were assigned to these orthologous genes. The binomial test was used to identify GO categories with more than 20 gene that had an excess of non-synonymous changes in bat lineages. Next, we used the program MAPP (multivariate analysis of protein polymorphism) (Stone and Sidow 2005) to evaluate the physicochemical impact of these convergent amino acid substitutions on bats. Physicochemical variations can be used to predict how these particular convergent amino acid substitutions might affect protein function.

In this work, we performed a probabilistic analyses of the sequence convergence in echolocating bats. A maximum likelihood approach, implemented in the software package Codeml ancestral, was used. We compared the pair-wise branches of two echolocating bat in the phylogeny, and posterior probabilities of all possible amino acid substitutions were calculated. The probabilities of divergent and convergent substitutions were calculated as the sum of joint probabilities of substitutions between the two branches of echolocating bats. Convergence and divergence estimates were based on posterior distributions of ancestral states and substitutions. The same state (same amino acid) represents convergent substitutions, and the different state represents divergent substitutions. Finally, to further validate that the convergence between two branch pairs of echolocating bats was significant, we performed the simulation analysis to compare the observed probabilities against that of the null hypothesis. Simulated sequences were generated using Evolver, another package from PAML package (Yang 2007). The branch-wise convergence probabilities were calculated with 1,000 replicates.

Identification of olfactory receptor (OR) repertoires

We used the similar *in silico* method as previously reported in Dong et al. (Dong, et al. 2009). At first, we used previously published OR genes in vertebrates as query sequences (Niimura and Nei 2007) and conducted a TBlastN search against the genome sequences with a cutoff E value of 1e-10 to identify the OR gene repertoires. Here, we totally identified OR gene repertories from eight mammalian genomes (the great leaf-nosed bat, Chinese rufous horseshoe bat, little brown bat, large flying fox, human, mouse, dog and cow). The draft genome sequences of human (Homo sapiens, hg38; GCA_000001405.15), mouse (Mus musculus, Mouse Build 38, GCA_000001635.2), dog (Canis lupus familiaris, canFam3), cow (Bos taurus, bosTau8), large flying fox (Pteropus vampyrus, pteVam1) and little brown bat (Myotis lucifugous, Myoluc2.0) were downloaded from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu). Then, the non-redundant blast-hits were extended to the 5' and 3' directions along the genome sequences, and the potential coding regions were extracted from these sequences. The chemosensory receptor genes in mammals have high sequence similarity. Here, we re-performed a TBlastN against the genome sequences using OR coding genes identified from each species, and the non-redundant blast-hits were used to identify the OR pseudogenes containing interrupting stop codons or frameshifts. To identify partial OR genes from these sequences, we extracted the sequences that did not have any nonsense or frameshift mutations. We then constructed a multiple alignment of these sequences together with functional OR genes by the program E-INS-i in MAFFT version 5.8 (Katoh, et al. 2002). From those alignments, we extracted partial OR sequences that meet the following criteria. When the C-terminal region of an OR gene is missing from the genome sequence, the N-terminal region should contain an initiation codon at a proper position and should not contain any nonsense mutations, frameshifts, or long gaps. When the N-terminal region is missing, the C-terminal portion should have a stop codon at a proper position and should not contain any nonsense mutations, frameshifts, or long gaps. We also identified 6 and 10 sequences with nonsense stop codon in the great leaf-nosed

bat and Chinese rufous horseshoe bats, which miss both a start and stop codons. However, These sequences were removed because they have relatively short sequence length (~400 bp) and have strong sequence similarity with bitter taste receptor genes. To assign identified OR genes into distinct OR gene families, a collection of protein sequences from HORDE database version 42 (Safran, et al. 2003) was used.

To detect the extensive gain and lose of OR gene repertories, we employed the reconciled tree method (Nam and Nei 2005), in which the topology of a gene tree is reconciled with that of a species tree. An in-house program was applied. Briefly, based on the phylogenetic tree of OR genes, we compared the condensed gene tree and the species tree under the parsimony principle. The number of ancestral genes can be estimated, and the information of the past occurrence of gene expansion and contraction. Here, we used a 70% condensed tree of OR genes for analyses.

Visual perception pseudogene identification

A list of vision-related genes were obtained from GO category of visual perception (GO:0007601). We subjected human vision-related proteins to TblastN against the genomes with cutoff threshold of e-value 1e-5. We found that best-hits for each human protein by using the criteria that more than 30% of the aligned sequences showed an identity above 30%. GeneWise algorithm was employed to identify potential pseudogenes with parameters –genesf –for –quiet. Those genes with frame shifts or pre-mature stop codons were considered as candidates. We then filtered them as follows: 1) we aligned all human proteins to their corresponding genomic loci, and those genes with frameshifts or premature stop codons in human-to-human alignments were removed; 2) as for the human-to-human alignments, those genes with obvious splicing errors near their frameshifts or premature stop codons were removed; 3) candidate pseudogenes with a low number of sequencing reads covering their frameshift or premature stop codon sites were regarded assembly error. Those genes with a number of reads containing genotype variations at these sites were considered as heterozygous and were also removed.

Rapid evolution

We used a method based on *Ka/Ks* to identify GO categories that significantly above average in the great leaf-nosed bat genome and Chinese horseshoe bat genome. At first, the *Ka* and *Ks* rates are calculated by PAML package from all aligned bases with quality score larger than 20 in orthologs, using the F3x4 codon frequency model and the REV substitution matrix. In order to examine the evolution function catalog, we downloaded the GO annotation of human gene from the Ensembl biomart database (release-71). We estimated the average *Ka* and *Ks* values for those genes which have annotated GO as following equations (S1, S2).

$$k_a = \frac{\sum_{ieT} a_i}{\sum_{ieT} A_i} \qquad \text{(S1)}$$

$$k_s = \frac{\sum_{ieT} s_i}{\sum_{ieT} S_i} \qquad \text{(S2)}$$

Where T is the number of annotated genes within GO categories, a_i and A_i are the numbers of non-synonymous substitutions and sites, s_i and S_i are the numbers of synonymous substitutions and sites in gene *i*, as estimated by PAML, respectively. The expected proportion of non-synonymous substitutions p_A in a GO category was then calculated (S3).

$$p_A = \frac{k_a \sum_{ieC} A_i}{k_a \sum_{ieC} A_i + k_s \sum_{ieC} S_i} \qquad \text{(S3)}$$

For a given GO category C, the probability p_c of observing an equal or higher number of non-synonymous substitutions and synonymous substitutions was calculated assuming a binominal distribution (S4).

$$p_c = \sum_{j-a_c}^{a_c+s_c} \binom{a_c+s_c}{j} p_A^j \left(1-p_A\right)^{a_c+s_c-j} \qquad \text{(S4)}$$

Where a_C and s_C are the total number of non-synonymous and synonymous substitutions in GO category *C*, respectively.

We applied an approach to the binomial test described above to identify GO categories that have an excess of non-synonymous changes on one lineage. For lineages x and y, the average proportion of non-synonymous substitutions were calculated by the following formula (S5).

$$p_x = \frac{x}{x+y} \text{ , which } x = \sum_i x_i \text{ , } y = \sum_i y_i \quad \text{(S5)}$$

Where $x = \sum_i x_i, y = \sum_i y_i$, x is the total number of non-synonymous substitutions in the x lineage, y is the total number of non-synonymous substitutions in the y lineage, and the divergence of the proportion of non-synonymous substitution numbers in different lineages between the observed and expected obeys binomial distribution, the formula is as in the following equation (S6).

$$p_c = \sum_{j-x_c}^{x_c+y_c} \binom{x_c+y_c}{j} p_x^j \left(1-p_x\right)^{x_c+y_c-j} \quad \text{(S6)}$$

As described for the absolute rate tests, we then computed this statistic for every GO category, as well as for every category in 10,000 randomly permuted data sets.

Population-based re-sequencing and SNP calling

We sampled a total of 20 great leaf-nosed bats distributed in four different locations. Genomic DNA was extracted from wing membranes of each individual. Paired-end sequencing library with an insert size of 500 bp was constructed for each sample, and sequenced on the Illumina Hiseq 2500 platform with 2×101 bp mode. Duplicate sequencing reads were filtered out according to the following criteria: 1) any reads with >10% unidentified nucleotides; 2) reads with >10 nt aligned to the adapter sequence, allowing <10% mismatches; 3) reads with 50% bases having phred quality <5. The filtered reads were mapped to the genome using BWA, and SAMtools were used to call SNPs. Then, we filtered SNPs using VCFtools and GATK under the

following criteria: 1) coverage depth >4 and <10000; 2) root mean square mapping quality >10; 3) the distance of adjacent SNPs >10 bp; 4) the distance to a gap > 10; 5) read quality value >30.

Phylogenetic and population genetic analyses

To estimate phylogenetic relationships, the genetic distances were calculated among all samples to generate a neighbor-joining (NJ) tree using PHYLIP. We performed a principal component analysis using the package GCTA. The population structure was inferred using frappe (v1.1) with a maximum likelihood method (Tang, et al. 2005).

Identification of selected regions

Sliding-window approach (10 kb window sliding in 10 kb step) was employed to quantify polymorphism levels (θ_π, pairwise nucleotide variation as a measure of diversity) and genetic differentiation (F*st*) between the high altitude region (DQ) and low altitude regions (TW, JX and GZ). To detect significant signatures of selective sweep, Z-transformed F*st* values was calculated.

Results

Genome sequencing and assembly

Next-generation genome sequencing was carried out, generating 476.5 Gb and 288.5 Gb of sequences for the great leaf-nosed bat and the Chinese rufous horseshoe bat (Supplementary Table S1), respectively. The genome size was estimated to be 2.18 Gb and 2.07 Gb for the great leaf-nosed bat and the Chinese rufous horseshoe bat (Supplementary Fig.1), which suggests the genome coverage of 218.6 × for the great leaf-nosed bat and 146.4 × for the Chinese rufous horseshoe bat. Genome assembly resulted in the contig and scaffold N50 values of 34 kb and 2.3 Mb for the great leaf-nosed bat and 17.5 kb and 3.75 Mb for the Chinese rufous horseshoe bat (Table 1, Supplementary Table S2,3 Supplementary Fig.2), respectively. A total of 827,722 and

8,110,439 heterozygous single nucleotide polymorphisms (SNPs) were identified in the great leaf-nosed bat and the Chinese rufous horseshoe bat genomes, respectively. The great leaf-nosed bat has a relatively lower number of heterozygous SNP rate at 0.037% than that of the Chinese rufous horseshoe bat (0.355%, Supplementary Table S4, Supplementary Fig.3).

Genome annotation

Known transposon-derived repeats account for 25.8% and 28.5% of the genomes in the great leaf-nosed bat and the Chinese rufous horseshoe bat, respectively, which are lower than other non-bat mammalian species (Supplementary Table S5). To facilitate the genome annotation, we generated a high-depth transcriptome data from these two rhinolophoid bats. With repeats masked, the genome was annotated by integrating the homologous prediction, *ab initio* prediction and transcription-based prediction methods. As a result, a non-redundant reference gene set of 22,009 and 23,152 protein-coding genes were generated for the great leaf-nosed bat and the Chinese rufous horseshoe bat (Supplementary Fig.4), respectively. We employed CEGMA method to evaluate the completeness of genome annotation. The result showed that the vast majority of the core genes were present in our predicted gene sets (97.08% for the great leaf-nosed bat and 96.14% for the Chinese rufous horseshoe bat), indicating the completeness of gene sets identification. Next, we aligned the transcriptome sequencing reads to the predicted genes, and the result showed that approximately 96% of exons are accurately covered (96.8% for the great leaf-nosed bat and 97.1% for the Chinese rufous horseshoe bat). Comparative analysis showed a high gene sequence similarity between them (91%, Supplementary Fig.5). We next examined the level of homology between our predicted genes and sequences in the Uniprot database. The result showed that >92% of the genes were functionally annotated (94% for the great leaf-nosed bat and 92.2% for the Chinese rufous horseshoe bat).

Compared with the gene families in other three mammalian species – the little

brown bat, large flying fox and human, we identified 8,792 homologous gene families shared by five species. A total of 975 gene families were specific to the rhinolophoid bats (**Fig. 1**). Further functional annotation indicated that the rhinolophoid bats specific gene families were significantly over-represented in two major functional categories: ATP binding (43 genes, *F.D.R.*= 0.0002) and immunity and host defense (25 genes, *F.D.R.*= 0.0051; Supplementary Table S6).

Phylogenetic reconstruction and evolutionary analysis

Until now, the relationship of bats to other members of superorder Laurasiatheria has proven difficult to resolve. Some studies insisted that bats belong to the clade of Pegasoferae which comprises chiroptera, carnivores and odd-toed ungulates (Lindblad-Toh, et al. 2011; Meredith, et al. 2011; McCormack, et al. 2012), whereas others proposed that bats are a sister group to the clade comprising carnivores and euungulata (Pumo, et al. 1998; Murphy, et al. 2001; Murphy, et al. 2007; Song, et al. 2012; Zhang, et al. 2013). To determine the phylogenetic position of bats within the superorder Laurasiatheria, a total of 4,569 single-copy 1:1 orthologous genes were used from 14 genomes (see Materials and Methods) to estimate the phylogenetic relationship. At first, phylogenetic reconstruction was based on concatenated data. The data were partitioned by coding genes, and estimated model parameters independently for each partition. In the bat subordinal relationship analysis, all phylogenetic evidence vigorously support the reciprocal monophyly of the proposed suborders Yinpterochiroptera and Yangochiroptera (bootstrap value = 100%, Supplementary Fig.6), with two rhinolophoid bats being more closely related to the non-echolocating large flying fox. Considering the position of bats in the phylogenetic tree, conflicting results were obtained using concatenated nucleotides and amino acids (Supplementary Fig.6). The result based on nucleotide data was in line with previous analysis that bats are a sister group to odd-toed ungulates, whereas the result based on amino acid data supported that bat bats are sister group to the Fereuungulata (Carnivores + Perissodactyla + Cetartiodactyla). To account for the tree discordance among loci, coalescent method was applied. Coalescent trees were highly consistent

with the result inferred from amino acid data using partitioned method (Supplementary Fig.7). To dissect the phylogenetic signal, previously published eight different phylogenetic hypotheses (Supplementary Fig.8) were proposed (Waddell, et al. 1999; Madsen, et al. 2001; Murphy, et al. 2001; Nishihara, et al. 2006; Prasad, et al. 2008; Lindblad-Toh, et al. 2011; Meredith, et al. 2011; McCormack, et al. 2012). We calculated the log-likelihood support for the nucleotides and amino acids datasets under the different proposed species tree topologies, and computed the approximately unbiased (AU) test statistic to compare the alternative topologies using CONSEL (Shimodaira and Hasegawa 2001). Only the hypotheses that bats are member of scrotifera (Chiroptera + Carnivores + Perissodactyla + Cetartiodactyla) with bat lineage diverging from Fereuungulata (Carnivores + Perissodactyla + Cetartiodactyla) are statistically supported by both nucleotide and amino acid evidence (Supplementary Table S7, Supplementary Fig.9). The result is consistent after incorporating the data from Eulipotyphyla group (Supplementary Fig.10). We subsequently estimated the divergence time among these 14 mammalian species. The bat lineage seems to be diverged from Fereuungulata around 81 million years ago, and the rhinolophoid bats seem to be diverged from the Old World fruit bats around 68 million years ago.

Gene evolution

Comparative genome analyses were carried out to assess the evolution and innovation within the rhinolophoid bats. We next determined the expansion and contraction of gene orthologous clusters during evolution. The result identified 48 significantly expanded and 65 significantly contracted gene families in the great leaf-nosed bat, 46 significantly expanded and 54 significantly contracted gene families in the Chinese rufous horseshoe bat (**Fig. 2)**. Functional annotation showed that gene family contraction mainly included many olfactory receptor gene families in both rhinolophoid bat lineages (Supplementary Table S8), which is consistent with the result that the olfactory system is aberrant in some echolocating bats. Many of the expanded gene families in both rhinolophoid bats are significantly enriched in

immune-related functional categories (Supplementary Table S9). Moreover, we identified 577, 453 and 182 positively selected genes in the great leaf-nosed bat, the Chinese rufous horseshoe bat and the large flying fox, (Supplementary Tables S10, 11, 12), respectively.

Evolution of olfactory receptor (OR) genes

Olfaction is of great importance in the lives of bats. Many bats can use olfaction for mother-pup recognition, find food and avoid danger. In Old World fruit bats, olfaction appears to be of particular importance, and fruit bats can find food from scent cues. Animals that rely heavily on the sense of smell tend to have large numbers of OR genes, while species that always use other senses have fewer functional OR genes (Niimura and Nei 2007). It has been suggested that bats displayed a diverse olfaction abilities. In order to describe the diversity of bat OR gene repertoires, we identified the entire set of OR genes of four bat species (Supplementary methods, Supplementary Table S13). In line with previous work (Hayden, et al. 2014), we observed that echolocating bats have less fraction of OR pseudogenes (18% for the great leaf-nosed bat, 16% for the Chinese rufous horseshoe bat and 14% for the little brown bat) than non-echolocating bats (26% for the large flying fox). However, further analysis showed that the large flying fox and little brown bat have more than 400 intact OR genes while these two rhinolophoid bats only have <300 intact OR genes. This finding is consistent with the result that rhinolophoid bats have a relatively small olfactory epithelium than the frugivorous Pteropodidae (Neuweiler 2000). Next, we reconstructed a protein neighbor-joining tree of all newly identified intact OR genes in bats (**Fig. 3a**). It is obvious that OR genes can be classified into two distinct classes based on sequence similarity: Class I, postulated to bind to water-borne molecules, and Class II, hypothesized to bind to airborne molecules. The exact number of OR genes in each class/OR family are shown (Supplementary Table S14, Table S15). It seems that four bat species contain similar number of OR genes in Class I, while OR gene contraction occurred in two rhinolophoid bats in Class II .

Previous works have documented that the number of OR genes varies extensively among mammalian species, and extensive gains and losses of OR genes have been observed (Niimura and Nei 2007). To further understand the evolutionary changes of OR gene repertoires, we estimated the gains and losses of the OR genes in a diverse range of mammals (Supplementary methods). Evolutionary changes in the number of OR genes in mammals have been shown in **Fig. 3b**. We can clearly identify an extensive OR gene contraction events occurred to the branch leading to the common ancestor of bats. Further extensive gene contractions can be observed in the branch leading to the rhinolophoid bats. This finding also suggests massive "birth-and-death" of OR genes in the bat species.

Evolution of hearing-related genes

Comparative analysis of the auditory perception can help to elucidate the molecular basis that underpin different auditory capabilities. Accelerated gene evolution between lineages could reflect an association between genetic changes and the evolution of specific traits. Compared with the large flying fox, analysis of non-synonymous to synonymous substitution (Ka/Ks) ratios of orthologous for different GO categories revealed an enrichment of elevated Ka/Ks values in the inner ear development (GO:0048839, P = 8.97E-10 for the great leaf-nosed bat and P = 1.16E-03 for the Chinese rufous horseshoe bat) and sensory perception of sound (GO:0007605, P = 1.11E-16 for the great leaf-nosed bat and P = 2.41E-08 for the Chinese rufous horseshoe bat) categories (Supplementary Table S16, 17). Of the 53 genes examined in the category of sensory perception of sound, 14 (P = 1.5E-5, the great leaf-nosed bat), 13 (P = 6.7E-4, the Chinese rufous horseshoe bat) and 12 (P = 6.6E-4, the little brown bat) genes showed evidence of positive selection in the echolocating bats. Meanwhile, only 2 genes were positively selected in the large flying fox (Supplementary Table S18). Since high omega may be due to stochastic effect caused by extremely small sample size, we removed these genes with omega value of 999. The result is also stable that more positively selected genes were detected in the branches leading to echolocating bats (12 genes, great leaf-nosed bat,

P = 9.1E-5; 10 genes, Chinese rufous horseshoe bat, *P* = 8.8E-3; 10 genes, little brown bat, *P* = 0.011). Next, branch model (two-ratio model) was carried out with the attempt to detect genes with accelerated evolution in the bat species. The result further indicated that more hearing-related genes have higher ω values on the branches leading to echolocating bats than all other lineages (Supplementary Table S19). Clade model C implemented in PAML was employed (Weadick and Chang 2012), and the result also persisted that more positively selected genes were detected in the branches leading to echolocating bats (Supplementary Table S20). Moreover, a significant association between the average number of non-synonymous substitutions for all the hearing-related genes leading to each mammalian species and the estimated frequency of best hearing sensitivity for that species (*r* = 0.84, *P* = 0.00032, **Fig. 4**) was observed. No significant correlation between such hearing frequencies and number of synonymous changes was observed (*P* = 0.132). A significant association between the number of non-synonymous changes between sister taxa was observed (*r* = 0.67, *P* = 0.006). It is obvious that echolocating bats have typically undergone many more non-synonymous changes in the hearing-related genes than non-echolocating mammals. These results indicated the evolution of ultrasonic hearing in the rhinolophoid bats has involved in adaptive amino acid replacements in the hearing-related genes, which provided evidence conferring greater auditory sensitivity to ultrasonic frequency.

Previous works have documented that seven hearing-related genes underwent convergent evolution in echolocators (Li, et al. 2008; Li, et al. 2010; Liu, Cotton, et al. 2010; Davies, et al. 2012; Liu, et al. 2012; Shen, et al. 2012). Here, genome-wide signatures of convergent evolution were examined in laryngeal echolocating bats. Except for the previously reported seven hearing-related genes, we totally identified 10 genes examined in the sound of perception category containing potential sequence convergent loci (site-wise convergence posterior probabilities > 0.5). To confirm our result, we amplified and sequenced these 10 hearing-related genes from another two echolocating bats (*Eptesicus fuscus* and *Miniopterus natalensis*). The result also

showed that these 10 genes have higher convergence probabilities occurred in echolocating bats from a wider range of taxa, and the convergence probabilities between branches were significant based on simulations (Supplementary Table S21). However, maximum likelihood trees recovered the topology that all laryngeal echolocating bats formed a monophyletic clade for only four genes (*COL1A1*, *ICAM1*, *BSND* and *STRC*, Supplementary Fig.11). Further analyses showed that echolocating bats are paraphyletic based on synonymous substitutions, whereas the non-synonymous trees revealed monophyly of laryngeal echolocators for only one hearing-related genes (*STRC* gene, Supplementary Fig.12). Next, multivariate analyses of protein polymorphism (MAPP) was employed to detect the physicochemical impact of convergent substitutions in echolocating bats. MAPP scores were estimated for the amino acid variants nested in the *STRC* gene, and the result showed that these replacements had important functional effects (MAPP score = 18.61, P = 1.44E-4 for H28Q; MAPP score = 10.33, P = 3.98E-3 for A39T; MAPP score = 7.37, P = 2.27E-2 for V169I). We further measured the number of sites with convergent amino acid substitutions along the branches as a direct measurement of sequence convergence, and found that the number of convergent sites in the branch pairs is proportional to the number of divergent sites (Supplementary Fig.13). The number of convergent sites in the laryngeal echolocating bats does not significantly exceed that between the branch pair of the little brown bat and large flying fox, given their numbers of divergent sites (Supplementary Table S22). No significant differences was observed in the total number of sites that have experienced convergent substitutions from hearing-related genes. This result indicated that there is no exceptional genomic signature indicative of adaptive convergence between laryngeal echolocating bats, and genes with adaptive convergent substitutions might confine to few specific genes.

Evolution of visual perception genes

Bats are nocturnal mammals. The eyes of most echolocating bats are relatively small and poorly developed, whereas Old world fruit bats often have excellent

eyesight (Shen, et al. 2010). Rhinolophoid bats have the most sophisticated echolocation ability, and have been proposed that some genes involved in visual perception may have undergone relaxed selection (Zhao, et al. 2009). We next examined the molecular basis for the poor visual perception in the echolocating bats. Of the 200 genes examined in the categories of visual perception (GO:0007601), a total of 12 genes have degenerated into pseudogenes in both great leaf-nosed bat and Chinese rufous horseshoe bat. Meanwhile, only three genes have degenerated into pseudogenes in the little brown bat, and none loss-of-function of visual perception genes through pseudogenization has been found in the large flying fox (*P-value* = 1.6e-4, Chi-square test, **Fig. 5**). These genes participate in phototransduction in the retina, dim-light adaptation, or color vision, etc. We next incorporated genes belonging to more visual-related GO categories, and identified more genes (15 genes, Supplementary Table S23, S24) that might have lost their functions due to a frameshift mutation or premature stop codons in rhinolophoid bats. The poor visual function of rhinolophoid bats might be explainable by deterioration of visual perception genes.

Evolution of immune-related genes

Bats have long been regarded as important reservoir hosts of emerging viruses (Calisher, et al. 2006). The coexistence of bats and virus might impose strong selective pressure on the bat genome. Compared with the large flying fox, categories associated with DNA repair (GO:0006281, $P < 0.01$), innate immune response (GO:0045087, $P < 0.01$) exhibited accelerated evolution in both rhinolophoid bats (Supplementary Table S15, 16). Furthermore, positively selected genes in the rhinolophoid bats are highly enriched in some immune related functional categories (Supplementary Table S25, 26), for example, regulation of innate immune response (GO:0045088, $P = 0.0053$, 5 genes), regulation of T cell differentiation (GO:0045580, $P = 0.033$, 6 genes), regulation of T cell activation (GO:0050863, $P = 0.044$, 8 genes) in the Chinese rufous horseshoe bat; regulation of innate immune response (GO:0045088, $P = 0.02$, 7 genes), T cell differentiation (GO:0030217, $P = 0.009$, 7

genes) in the great leaf-nosed bat. Additionally, we found that DNA repair pathway is over-represented among the positively selected genes in both rhinolophoid bats. DNA repair machinery of host cells can recognize viral genetic material as abnormal and damaged DNA. So, DNA repair apparatus becomes part of the innate intracellular defense against foreign invaders. Overall, these lines of evidence indicate a well-developed immune defense system against virus and microbial infection in rhinolophoid bats.

Genome re-sequencing analyses of the great leaf-nosed bats

To examine population dynamics and understand evolutionary processes, we sampled 20 great leaf-nosed bats from 4 major distributed locations in China, including one group from high-altitude region (**Fig. 6a**, **Table 2**, Supplementary Table S27). Whole-genome re-sequencing has been performed for each individual. It accomplished an average depth of 11.2×, and average genome coverage of 88% (Supplementary Table S27). We totally identified 27,064,431 high-quality single nucleotide polymorphisms (SNPs), most of which (78%, Supplementary Table S28) are located at the intergenic regions.

In order to resolve their phylogenetic relationships, we constructed a neighbor-joining (NJ) tree based on pairwise genetic distances (**Fig. 6b**). This result showed that the great leaf-nosed bats formed separate groups according to the different locations. Principal component analysis clearly divided these samples into four groups (DQ, GZ, JX and TW, **Fig. 6c**). These results suggested that there were significant population structures among the great leaf-nosed bat populations. Furthermore, we performed population structure analysis. When K=4, all these four populations were clearly separated (**Fig. 6d**).

Next, we measured the genetic diversity values (θ_π) of four populations, and found similar sequence diversity values (DQ: 0.0012, GZ:0.0009, JX:0.0009 and TW:0.0011, Supplementary Fig. S14). We further observed that the population

differentiation statistic (F*st*) between populations, and the result showed little differentiation among populations (F*st* ranging from 0.013 between JX and TW to 0.057 between TW and DQ, Supplementary Table S29), which suggests universal inter-region gene flows.

Since the method of population differentiation has been widely used to detect selective sweeps (Akey, et al. 2010; Axelsson, et al. 2013; Gou, et al. 2014), we focused on the regions with extremely high Z-transformed F*st* values (Z(F*st*)>5) to detect genetic adaptations in the great leaf-nosed bats that are associated with high altitudes (DQ *vs.* other locations). A total of 50 unique genome regions containing 80 candidate genes were identified (Supplementary Table S30). After performing GO enrichment analysis, a total of 16 GO categories were over-represented in these selected regions relative to the rest of the other genome regions (Supplementary Table S31). The result showed that genes related to catabolic process are likely to have been targets of recent positive selection. Interestingly, we found that five genes (*EPAS1*, *PLXND1*, *GJA1*, *SELL* and *CHDH*) belong to hypoxia response related GO categories (Pugh and Ratcliffe 2003; Storz and Moriyama 2008), including 'angiogenesis', 'blood coagulation', 'blood vessel morphogenesis' and 'oxidoreductase activity'. *EPAS1* can respond to the changes in available oxygen in the cellular environment under the high-altitude conditions. Our work suggested that *EPAS1* is involved in a selective sweep during the move of bats from low to high altitude. Although hypoxia GO categories are not over-represented, these highlighted hypoxia-related genes gave us a clue that genetic adaptations might be associated with high altitude.

Discussion

Using deep sequencing and *de novo* assembly, we generated two genomes of rhinolophoid bats. Rhinolophoid bats can perceive the world by using a wide range of sensory mechanisms, some of which have become highly specialized. These genome data provided useful resources to decipher the molecular adaptations of phenotypic traits.

Rhinolophoid bats arguably possess the most sophisticated echolocation systems, and can emit relatively long calls adapted to detect and classify the wing beats of insects. They are heavily reliant on hearing for a variety of ecologically important roles. Previous works have documented that hearing-related genes are predominantly evolutionarily conserved in mammals (Kirwan, et al. 2013). Here, we found evidence that some hearing-related genes have undergone Darwinian selection associated with the evolution of specialized constant frequency echolocation. Positive selection acting on hearing-related genes in rhinolophoid bats might result from the extreme selectivity used in auditory processing by these bats. Many previous works have reported the sequence convergence of some hearing-related genes reuniting echolocating bats (Li, et al. 2008; Li, et al. 2010; Liu, et al. 2011; Davies, et al. 2012; Liu, et al. 2012). We found no genome-wide sequence convergence for echolocation, indicating erroneous phylogenetic grouping are still rare

It has been suggested that the enlargement of one area of brain might be associated with the reduction in size of other brain area (Harvey and Krebs 1990). The auditory cortex and the inferior colliculus are extremely enlarged in the volume in laryngeal echolocating bats (especially in rhinolophoid bats), while visual brain areas are relatively enlarged in Old World fruit bats (Dechmann and Safi 2009). The trade-off has been proposed in investment in brain tissues because of the extreme energetic demands imposed by neural processing. Our result showed more visual perception genes have become pseudogenes in rhinolophoid bats, and it is reasonable

to speculate that some visual perception gene may have undergone relaxed natural selection in echolocating bats. Meanwhile, positive selection acting on some hearing-related genes was identified. Such concordance suggests that some genes are impacted by natural selection, which raised the possibility that changes at the sensory genes will have direct consequences for those genes controlling for other sensory modalities, perhaps via trade-offs. This finding supports the longstanding but weakly supported assumption that bats are experiencing trade-off between vision and audition .

Olfaction is of great importance in the lives of bat species. Previous works have identified olfactory receptor (OR) gene repertoire in the little brown bat and the large flying fox using the profile hidden Markov model (Hayden, et al. 2010; Hayden, et al. 2014). A profile hidden Markov model (HMM) is a probabilistic model of a multiple alignment of related proteins. The result of HMM is not considerable when identifying remote homologous. Some of the OR sequences, especially non-transmembrane regions, are extensively diverged. Here, we used the pairwise homologous search method to identify OR genes based on previously annotated OR sequences, and more intact OR genes have been detected. Our result showed dramatic changes of OR gene repertories in bats, and an extremely contraction occurred in the branches leading to rhinolophoid bats. Although the number of OR gene expanded in the lineage leading to the little brown bat, the expanded OR genes are mainly grouped in specific gene family. A possible explanation is that the little brown bat has no well-developed olfaction ability, but tends to recognize specific odorants after recent OR gene duplication. These comparative analyses have provided great insights into adaptation to their specialized sensory mechanisms.

In this work, we re-sequenced the genome of 20 great leaf-nosed bats from four distributed locations. The genome re-sequencing analysis has been performed based generally on the following considerations: 1) to characterize the genetic diversity and patterns of evolution; 2) to understand the genetic bases of adaptation to high altitude

in the great leaf-nosed bats. Efforts for the conservation measures will benefit from the knowledge of population genetic structure of the great leaf-nosed bats. Here, we found very little differentiation among populations, which suggests universal inter-region gene flows or incomplete lineage sorting. A broader geographical scale analysis is needed in the future. Furthermore, we provided evidence of genetic adaptation in the great leaf-nosed bat that are associated with high altitude. Selective sweep mapping was conducted for populations from different altitudes, and identified several hypoxia-related genes with a high extent of differentiation on the genome scale. *EPAS1* is transcription factor that respond to the changes in the available oxygen in the cellular environment under high-altitude conditions, and mutations at *EPAS1* are tightly associated with hematologic phenotypes (van Patot and Gassmann 2011). Previous works have documented that EPAS1 polymorphisms are associated with Tibetan people with lower hemoglobin concentrations (Beall, et al. 2010). A loss-of-function role of *EPAS1* might exist in high-altitude adaptation. So, our result indicated potential high-altitude hypoxia adaptation mechanisms of the great leaf-nosed bat. Our work is based on a limited genome re-sequencing resource, and data from more samples are necessary for future work. However, false positives notwithstanding, the results provided valuable staring points for experimental follow-up, and suggested an initial evolutionary scenario of bats in adaptation to high-altitude hypoxia.

To the best of our knowledge, it is the first time to report the *de novo* assembled genome and genome re-sequencing of bats with long constant frequency echolocation calls. These data are essential for us to understand the evolution of bats.

Figure legends

Figure 1 Unique and homologous gene families. The number of unique and shared gene families is shown in each of the diagram components.

Figure 2 Phylogenetic analysis. Phylogenetic tree shows the relationship among 14 mammalian species. The estimated divergent time for all nodes was shown. The number of expanded (red) and contracted (green) gene families is shown in each branch.

Figure 3 Evolution of olfactory receptor gene repertories. A) Phylogenetic tree of functional OR genes in four bat species. B) Evolutionary changes of the number of OR genes in mammals. The numbers in rectangular boxes are those of functional OR genes for the extant or ancestral species. A plus sign and a minus sign for a branch represent OR gene gains and losses. We used a 70% bootstrap condensed tree of OR genes.

Figure 4 Association between evolution of hearing-related genes and the evolution of high-frequency hearing in mammalian species. Plot of estimated frequency of best hearing sensitivity *vs.* the number of non-synonymous substitutions shows a significant relationship.

Figure 5 Visual perception genes that have been become psuedogenes in rhinolophoid bats.

Figure 6 Phylogenetic and population analyses of great leaf-nosed bats. A) the geographic distribution of the sampling locations; B) A neighbor-joining phylogenetic tree constructed using whole-genome SNPs data; C) Principal component analyses plots of the first two components. D) Population structure plots with K=2-4. The y-axis quantifies the proportion of the individual's genome from inferred ancestral

populations, and x axis represents different populations.

Tables

Table 1 Global statistics of the great leaf-nosed bat and the Chinese rufous horseshoe bat genomes.

Genomic features	great leaf-nosed bat	Chinese rufous horseshoe bat
Assembled genome size (Gb)	2.18	2.07
Genome coverage	218.6×	146.4×
Contig N50 (Kb)	34	17.5
Scaffold N50 (Mb)	2.3	3.75
GC content (%)	41.2	41.7
Repeat rate (%)	25.8	28.4
Protein-coding genes	22,009	23,152

Table 2 Sampling information for four population of the great leaf-nosed bat

Populations	**Location**	**Sample size**	**Altitude**	**Total reads (Gb)**
DQ	Diqing, Yunnan, China (N 27°83′ E 99°72′)	6	3300 m	215.26
GZ	Guiyang, Guizhou, China (N 26°63′ E106°75′)	5	900 m	146.64
JX	Fuzhou, Jiangxi, China (N 28°01′ E116°35′)	5	500 m	181.52
JX	Xinzhu, Taiwan, China (N 24°81′ E120°96′)	4	200 m	143.45

Acknowledgements

This project is supported by Key Construction Program of the National '985' project of East China Normal University to Dong Dong (79633006), and the National Natural Science Foundation of China (No. 31570382) to Shuyi Zhang. We thanks Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. for genome sequencing and Dr. Chao-Hung Lee for providing valuable advices..

Author contributions

DD designed the study, and DD, ML, PH, YP, SM, GZ, EP, KL and SZ carried out the data analysis. DD wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Accession codes: The whole genome sequences of great leaf-nosed bat and Chinese rufous horseshoe bat have been deposited in the GenBank Bioproject database under the accession code PRJNA267039 and PRJNA294852, respectively. Raw genomic sequence data has been deposited in the GenBank sequence read archive under the accession code SRP049752 and SRP063386, respectively. Genome assembles have been deposited in GenBank under the accession code JXIK00000000 and LVEH00000000. All short reads of transcriptome data have been deposited into the Short Read Archive under the accession code SRP049970 and SRP063381. The Genome re-sequencing data have been deposited in the GenBank Bioproject database under the accession code PRJNA309742.

Competing financial interests

The authors declare no competing financial interests.

References

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. Proc Natl Acad Sci U S A 107:1160-1165.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360-364.

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, et al. 2010. Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci U S A 107:11459-11464.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 268:78-94.

Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Research 18:810-820.

Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. 2006. Bats: important reservoir hosts of emerging viruses. Clin Microbiol Rev 19:531-545.

Davies KTJ, Cotton JA, Kirwan JD, Teeling EC, Rossiter SJ. 2012. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. Heredity 108:480-489.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22:1269-1271.

Dechmann DK, Safi K. 2009. Comparative studies of brain evolution: a critical insight from the Chiroptera. Biol Rev Camb Philos Soc 84:161-172.

Dong D, He G, Zhang S, Zhang Z. 2009. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. Genome Biol Evol 1:258-264.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, et al. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535-538.

Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, Yang S, Jia J, Kong X, Wei Z, et al. 2014. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Research 24:1308-1315.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Jr., Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research 31:5654-5666.

Harvey PH, Krebs JR. 1990. Comparing brains. Science 249:140-146.

Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. Genome Research 20:1-9.

Hayden S, Bekaert M, Goodbla A, Murphy WJ, Davalos LM, Teeling EC. 2014. A cluster of olfactory receptor genes linked to frugivory in bats. Molecular Biology and Evolution 31:917-927.

Jones G, Teeling EC. 2006. The evolution of echolocation in bats. Trends in Ecology & Evolution 21:149-156.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462-467.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059-3066.

Kirwan JD, Bekaert M, Commins JM, Davies KT, Rossiter SJ, Teeling EC. 2013. A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. Evol Appl 6:412-422.

Li G, Wang JH, Rossiter SJ, Jones G, Cottont JA, Zhang SY. 2008. The hearing gene Prestin reunites echolocating bats. Proceedings of the National Academy of Sciences of the United States of America 105:13959-13964.

Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, et al. 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Research 34:D572-580.

Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. Current Biology 20:R55-56.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478:476-482.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol Biol 10:302.

Liu Y, Cotton JA, Shen B, Han XQ, Rossiter SJ, Zhang SY. 2010. Convergent sequence evolution between echolocating bats and dolphins. Current Biology 20:R53-R54.

Liu Y, Han NJ, Franchini LF, Xu HH, Pisciottano F, Elgoyhen AB, Rajan KE, Zhang SY. 2012. The Voltage-Gated Potassium Channel Subfamily KQT Member 4 (KCNQ4) Displays Parallel Evolution in Echolocating Bats. Molecular Biology and Evolution 29:1441-1450.

Liu Z, Li SD, Wang W, Xu DM, Murphy RW, Shi P. 2011. Parallel Evolution of KCNQ4 in Echolocating Bats. Plos One 6.

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong WW, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409:610-614.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research 22:746-754.

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334:521-524.

Mulder N, Apweiler R. 2007. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396:59-70.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294:2348-2351.

Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Research 17:413-421.

Nam J, Nei M. 2005. Evolutionary change of the numbers of homeobox genes in bilateral animals.

Molecular Biology and Evolution 22:2386-2394.
Neuweiler G. 2000. The biology of bats. New York: Oxford University Press.
Niimura Y, Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. Plos One 2:e708.
Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc Natl Acad Sci U S A 103:9929-9934.
Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061-1067.
Prasad AB, Allard MW, Green ED, Program NCS. 2008. Confirming the phylogeny of mammals by use of large comparative sequence data sets. Molecular Biology and Evolution 25:1795-1808.
Pugh CW, Ratcliffe PJ. 2003. Regulation of angiogenesis by hypoxia: role of the HIF system. Nat Med 9:677-684.
Pumo DE, Finamore PS, Franek WR, Phillips CJ, Tarzami S, Balzarano D. 1998. Complete mitochondrial genome of a neotropical fruit bat, Artibeus jamaicensis, and a new hypothesis of the relationships of bats to other eutherian mammals. J Mol Evol 47:709-717.
Safran M, Chalifa-Caspi V, Shmueli O, Olender T, Lapidot M, Rosen N, Shmoish M, Peter Y, Glusman G, Feldmesser E, et al. 2003. Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. Nucleic Acids Research 31:142-146.
Schnitzler HU, Moss CF, Denzinger A. 2003. From spatial orientation to food acquisition in echolocating bats. Trends in Ecology & Evolution 18:386-394.
She R, Chu JS, Wang K, Pei J, Chen N. 2009. GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Research 19:143-149.
Shen YY, Liang L, Li GS, Murphy RW, Zhang YP. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet 8:e1002788.
Shen YY, Liu J, Irwin DM, Zhang YP. 2010. Parallel and Convergent Evolution of the Dim-Light Vision Gene RH1 in Bats (Order: Chiroptera). Plos One 5.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.
Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Natl Acad Sci U S A 109:14942-14947.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 Suppl 2:ii215-225.
Stone EA, Sidow A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Research 15:978-986.
Storz JF, Moriyama H. 2008. Mechanisms of hemoglobin adaptation to high altitude hypoxia. High Alt Med Biol 9:148-157.
Tang H, Peng J, Wang P, Risch NJ. 2005. Estimation of individual admixture: analytical and study design considerations. Genet Epidemiol 28:289-301.
Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4:Unit 4 10.
Teeling EC. 2009. Hear, hear: the convergent evolution of echolocation in bats? Trends in Ecology & Evolution 24:351-354.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science 307:580-584.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology 31:46-+.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111.

van Patot MC, Gassmann M. 2011. Hypoxia: adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2alpha. High Alt Med Biol 12:157-167.

Waddell PJ, Okada N, Hasegawa M. 1999. Towards resolving the interordinal relationships of placental mammals. Systematic Biology 48:1-5.

Weadick CJ, Chang BS. 2012. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. Molecular Biology and Evolution 29:1297-1300.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.

Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, Wynne JW, Xiong Z, Baker ML, Zhao W, et al. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science 339:456-460.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution 22:2472-2479.

Zhao H, Rossiter SJ, Teeling EC, Li C, Cotton JA, Zhang S. 2009. The evolution of color vision in nocturnal mammals. Proc Natl Acad Sci U S A 106:8980-8985.