Frontiers in Zoology

RESEARCH   **Open Access**

Gene expression vs. sequence divergence: comparative transcriptome sequencing among natural *Rhinolophus ferrumequinum* populations with different acoustic phenotypes

Hanbo Zhao[1], Hui Wang[1], Tong Liu[1], Sen Liu[2], Longru Jin[1], Xiaobin Huang[3], Wentao Dai[1], Keping Sun[1*] and Jiang Feng[1,4*]

Abstract

Background: Although the sensory drive hypothesis can explain the geographic variation in echolocation frequencies of some bat species, the molecular mechanisms underlying this phenomenon are still unclear. The three lineages of greater horseshoe bat (*Rhinolophus ferrumequinum*) in China (northeast, central-east, and southwest) have significant geographic variation in resting frequencies (RF) of echolocation calls. Because their cochleae have an acoustic fovea that is highly sensitive to a narrow range of frequencies, we reported the transcriptomes of cochleae collected from three genetic lineages of *R. ferrumequinum*, which is an ideal organism for studying geographic variation in echolocation signals, and tried to understand the mechanisms behind this bat phenomenon by analyzing gene expression and sequence variation.

Results: A total of 8190 differentially expressed genes (DEGs) were identified. We identified five modules from all DEGs that were significantly related to RF or forearm length (FL). DEGs in the RF-related modules were significantly enriched in the gene categories involved in neural activity, learning, and response to sound. DEGs in the FL-related modules were significantly enriched in the pathways related to muscle and actin functions. Using 21, 945 single nucleotide polymorphisms, we identified 18 candidate unigenes associated with hearing, five of which were differentially expressed among the three populations. Additionally, the gene *ERBB4*, which regulates diverse cellular processes in the inner ear such as cell proliferation and differentiation, was in the largest module. We also found 49 unigenes that were under positive selection from 4105 one-to-one orthologous gene pairs between the three *R. ferrumequinum* lineages and three other Chiroptera species.

Conclusions: The variability of gene expression and sequence divergence at the molecular level might provide evidence that can help elucidate the genetic basis of geographic variation in echolocation signals of greater horseshoe bats.

Keywords: *Rhinolophus ferrumequinum*, Echolocation, Geographic evolution, Adaptation, Transcriptome

* Correspondence: sunkp129@nenu.edu.cn; fengj@nenu.edu.cn
[1]Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun 130117, China
Full list of author information is available at the end of the article

Background

Many evolutionary biologists strive to explain the mechanisms of phenotypic divergence [1]. However, what drives phenotypic divergence remains one of the least understood biological phenomena [2]. Some evidence has suggested that different habitats are likely to impose different selective pressures on isolated populations and result in geographic variation of phenotypic traits, such as morphological, physiological, behavioral, and sensory traits [3, 4]. Elucidating the genetic mechanisms behind phenotypic divergence that are driven by adaptation is a primary mission of modern evolutionary biology [5].

Sensory traits in animals directly impact individual fitness by affecting resource acquisition, orientation, mate choice, and species recognition. Geographic variation in these traits is usually mediated by adaptive processes rather than random processes like genetic drift [6]. Therefore, the sensory drive hypothesis, which predicts a close association between the geographic variation of sensory signals and environmental variables, has been proposed to explain how environments affect signal traits and sensory systems [7].

Acoustic signals, which are an important sensory characteristic, have long attracted numerous researchers who mainly focused on the geographic variation of calls in birds, anurans, insects, and bats [8–10]. In particular, the echolocation calls of bats have drawn attention from an increasing number of researchers as a new research model for testing the sensory drive hypothesis. The high variability in echolocation calls over the distributional ranges of many lineages is likely due to numerous factors, including differences in environmental conditions, prey size, variation in body size, age, and sexual dimorphism [11–14]. Besides those external factors, internal molecular mechanisms also play an important role in the geographic variation of acoustic signals, because phenotypic changes indicate changes in gene expression, and genotypes can affect gene expression based on interactions between genotypes and the environment [15]. Although studies that evaluated intraspecific gene expression and genotype variation are scarce, some studies have found that several genes, such as *prestin* [16], *TMC1* [17], *KCNQ4* [18], *CDH23*, *PCHH15*, and *OTOF* [19], were related to the adaptation of echolocation and convergently evolved in some animals that use echolocation. However, the molecular mechanism underlying environmentally driven adaptive trait divergence of acoustic signals within species remains unclear.

The greater horseshoe bat (*Rhinolophus ferrumequinum*), a widespread, constant frequency–frequency modulating (CF–FM) bat, is an ideal model organism to help elucidate the molecular mechanisms underlying acoustic geographic variation and adaptive evolution of Chiroptera for several reasons. First, *R. ferrumequinum*

has a broad distribution in the Old World and exhibits acoustic geographic variation. Different populations have different dominant frequencies of echolocation calls that have been reported [20, 21]. Second, Sun et al. [22] revealed a geographic pattern of *R. ferrumequinum* echolocation pulses emitted at rest (resting frequencies, RF), and, interestingly, they found that RF corresponded to genetic differentiation. Three genetic lineages, northeast (NE), central-east (CE), and southwest (SW), were identified based on several neutral genetic markers [22–24] and had significantly different echolocation calls. Third, combined with molecular data, acoustic parameters, and climatic factors, Sun et al. [22] inferred that ecological selection and cultural drift were most likely related to the variation of *R. ferrumequinum* calls across China. However, the internal molecular mechanisms of acoustic variation are still not known. For horseshoe bats, RF is largely genetically determined [25], so the divergence of echolocation calls between geographical populations might be related to genotype changes and differential gene expression.

Furthermore, the mammalian cochlea is an exceptionally sensitive organ that detects a very wide variety of sound intensities. The strongest intensity that does not damage the ear is 10^{12} larger than the threshold level of detectible sound, and can discriminate both infrasonic and ultrasonic sounds in different species [26]. Therefore, the cochlea is an important organ that participates in the hearing process by receiving acoustic signals. For bats, the specialization of cells in the ears and brain improves the response to the frequencies of the sounds they emit. *Rhinolophus ferrumequinum*, as a CF–FM bat, has cochleae with a specific area known as an acoustic fovea. The acoustic fovea is well-known for its key role in the process of Doppler-shift compensation behavior, because it makes the bats highly sensitive to a narrow range of frequencies [27, 28]. Although no study can definitively support that bat populations with different RF have different genetic features of the cochleae, significant genetic differences were detected between bat species with different RF in the cochleae [29]. Moreover, a study that focused on the cochleae of inbred mice with hearing variation revealed the existence of anatomical- and frequency-specific genes, and also demonstrated that there is at least one specific genome-wide locus that is significantly associated with each frequency in inbred mouse strains [30]. Even though there was no clear evidence that difference in RF of populations was mirrored by genetic differences in the cochlea, it would be appropriate to test the cochlea to reveal molecular mechanisms of acoustic variation of bats because of the role of the cochlea in hearing.

To explore the molecular mechanisms of the diversity of echolocation calls, we conducted RNA sequencing (RNA-Seq) in this study. Among novel sequencing

Zhao *et al. Frontiers in Zoology*   (2019) 16:37

Page 3 of 15

technologies, RNA-Seq is outperforming the traditional hybridization-based microarray method, because it can evaluate gene expression without prior knowledge of the sequence [31, 32]. Furthermore, RNA-Seq can reveal genotype information and does not require prior genomic or genetic resources, and can be used as a cheaper alternative that enables identification of single nucleotide polymorphisms (SNPs) located in transcribed regions [33]. Recently, it has become easier to apply this approach to many organisms because of the development of appropriate statistically grounded pipelines for data analysis [34, 35]. Moreover, a recent study have demonstrated that small sample sizes are sufficient to assess interpopulation divergence when thousands of biallelic SNP markers were used [36]. Therefore, high-throughput genomic sequencing of RNA has great potential to elucidate evolution in non-model organisms.

In this study, we performed transcriptome sequencing of *R. ferrumequinum* cochlea samples collected from representative populations of three genetic lineages in China. We propose that the diversity of echolocation calls among *R. ferrumequinum* populations is, to a large extent, related to (i) regulation of gene expression and (ii) sequence divergence. Therefore, we analyzed gene expression and orthologous gene sequences to identify genes that may be involved in shaping the geographic variation of acoustic signals. Our results will provide insight into the underlying mechanism that produces geographic variation in echolocation calls.

Methods

Sample collection and phenotyping

We set mist nets to capture bats at dusk in unproductive habitats and collected 14 individuals in total from all three populations, which included representatives of all three clusters based on our previous study [22]. To reduce the impact of gender and age on gene expression, we sampled adult males of similar sizes and weights. We captured five individuals from a representative population of the NE genetic lineage in Jilin Province (sample IDs JL01–JL05, E 125.89°, N 43.28°) and one population of the CE lineage in Henan Province (sample IDs HN01–HN05, E 113.94°, N 35.71°). In Yunnan Province, we collected four individuals from a representative population of the SW lineage (sample IDs YN01–YN04, E 99.86°, N 26.53°). All sampling was conducted during the summer (late July to late August 2017).

The echolocation calls of greater horseshoe bats emitted at rest have simple structures and are ideal for sonographic analysis. Additionally, these calls have a narrow range of frequencies that are consistent with the maximal sensitivity of frequencies of the cochlear acoustic fovea. Therefore, we recorded calls emitted at rest using Avisoft-UltraSoundGate (Avisoft Bioacoustics, Glienicke,

Germany) with a sample rate of 441 kHz at 16 bits/sample. We put the microphone (CM16/CMPA, Avisoft Bioacoustics, Berlin, Germany; flat frequency response: 10 Hz–200 kHz, ± 3 dB) approximately 30 cm in front of bats at rest, and the recordings were transferred to and saved on a computer.

We selected high-quality calls based on the criteria proposed by Russo et al. and Jiang et al. [37, 38] and analyzed them using Avisoft-SASLab Pro v 5.2.12 [39]. More specifically, initial calls were not considered for analysis, and only the second call per group was chosen when calls were emitted in groups. Moreover, we excluded calls with a CF portion that lasted less than 10 ms. For each bat, 30 high-quality calls were arbitrarily selected and measured for the CF components in the dominant second harmonic from the power spectra of a call. Then, the mean RF value of each individual was used in the analysis. Given that body size may act on acoustic features of bat echolocation calls, we also measured the forearm length (FL) of each bat. Then, we pairwise compared average RF and FL values among the three populations by Mann–Whitney U test. After we finished recording echolocation sounds and measuring FL, bats were sacrificed, and their whole cochleae were separately preserved in RNAlater (Tiangen Biotech, Beijing, China) within 25 min post-mortem for RNA-Seq data generation. All samples were subsequently stored at − 80 °C until RNA extraction.

RNA-Seq library preparation and sequencing

We homogenized each cochlea sample and extracted total RNA using a TRIzol Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. After checking the amount of RNA degradation by RNase-free agarose gel electrophoresis and capillary electrophoresis with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), we performed a poly (A)-capture to remove rRNA; then, mRNA was reverse-transcribed into first-strand cDNA using random primers. Next, we synthesized the second-strand cDNA using DNA polymerase I, RNase H, dNTP, and second-strand buffer. After purification using a QiaQuick PCR Extraction kit (Qiagen, Hilden, Germany), the cDNA fragments were end-repaired, poly(A)-tailed, and ligated to Illumina sequencing adapters. The ligation products were -elected by agarose gel electrophoresis and PCR amplification. Sequencing was carried out on an Illumina HiSeq X Ten (Illumina, San Diego, CA, USA) with 2 × 150-bp paired-end reads. All raw reads were deposited in the NCBI Short Read Archive (SRA) Database under SRA accession PRJNA515764.

Transcript assembly, quantification, annotation, and differential expression testing

First, we filtered and trimmed barcoded RNA-Seq reads and low-quality reads. We removed reads that were

contaminated by Illumina adapters, contained more than 10% unknown nucleotides (N), or had more than 40% low-quality (Q-value ≤10) bases using trimmomatic v 0.36 [40]. Next, filtered reads from all 14 individual cDNA libraries, which included individuals of all the three groups, were loaded into Trinity v 2.4.0 to assemble a de novo reference transcriptome with default parameters. To avoid redundant transcripts, we kept the longest isoform for each "trinity gene" identified by Trinity using a Perl script and defined the longest transcript as a unigene. This de novo reference was used to obtain expression profiles. Moreover, we also assembled a reference transcriptome for each population in the same way for positive selection analysis. To evaluate and compare the completeness of the gene set of our four transcriptomes, we used Benchmarking Universal Single-Copy Orthologs (BUSCO v 3.1.0) to search for orthologs in the "laurasiatherian_odb9" database, which includes a collection of 6253 single-copy Laurasiatherian orthologs. All generated unigenes in these four unigene sets were aligned with an E-value of 1E− 5 to the following protein databases: Nr [41], Swiss-Prot [42], KEGG [43], and COG/KOG [42] by BLASTx [44].

We then performed pairwise comparisons to identify differentially expressed genes (DEGs) in the Bioconductor package edgeR with default parameters [45], which scored well in a recent comparison [46]. In this study, unigenes with false discovery rate (FDR) ≤ 0.05 and an absolute value of a log_2-fold change > 1 were considered DEGs. More specifically, the de novo reference transcriptome assembled with all of the samples was used as the reference sequence, and we calculated the reads per kilobase per million mapped reads (RPKM) of each sample by mapping high-quality clean reads to the reference transcriptome using the short read alignment tool in Bowtie2 with default parameters [47]. According to the plot showing the sample relationships based on multidimensional scaling (MDS), we filtered outlier samples. Then, we performed pairwise comparisons for the three populations to select DEGs. For each comparison (HN vs. JL, HN vs. YN, and YN vs. JL), the former population always had a higher call frequency than the latter. We considered a unigene with higher expression in the former population to be upregulated and vice versa to be downregulated. The *P*-value was corrected by the Benjamini and Hochberg method [48]. We visualized the gene expression profiling of all DEGs by heatmap analyses of hierarchical clustering.

Weighted gene co-expression network analysis

A weighted gene co-expression network analysis (WGCNA) [49] was used to identify unigenes and gene networks associated with RF and FL. First, we extracted the expression values of DEGs obtained from three pairwise comparisons

and calculated Pearson's correlation matrix for all genes. Next, the correlation matrix was transformed by raising all values to a soft-thresholding power (β = 9 in this study) to obtain an adjacency matrix. Then, we transformed the adjacency matrix into a topological overlap matrix (TOM) and identified modules using 1-TOM as the distance measure with a deepSplit value of 2 and a minimum size cut-off of 50 for the resulting dendrogram. After identification by clustering, groups of co-regulated genes (modules) were merged with a height cut-off of 0.25. Then, we searched biologically meaningful modules by evaluating Pearson's correlation between modules and phenotypic features. We focused on those modules that were strongly correlated with the phenotypic features. Subsequently, to get a better understanding of how those genes impact phenotypic features, genes in all modules were studied using heatmap analyses of hierarchical clustering, GO functional enrichment, and KEGG pathway analysis (P < 0.01, FDR < 0.01). We also merged those modules that were significantly correlated with RF or FL to perform GO term enrichment analysis (P < 0.01, FDR < 0.01).

SNP identification and filtering

As described in detail in Maestre et al. [34], SNPs were called from RNA-Seq data without a reference genome using KisSplice v 2.4.0 [35] and KisSplice2RefTranscriptome v 1.3.2 [34]. To remove sequencing errors, we filtered rare variants by setting the cut-off to 5%. Then, we aligned SNPs to the de novo reference transcriptome assembled with all of the samples using BLAT suite.34 [50] with default parameters to predict amino acid changes. The output file was manually converted to a VCF file, and the following filters were applied: 1) SNPs that were located in the noncoding region of transcripts were removed; 2) rare variants were filtered out, as they are more likely to be sequencing error, and only markers with MAF > 5% were retained; 3) only those SNPs with less than 10% missing data across all sites were included in further analyses; and 4) to reduce linkage, SNPs were pruned using the "snpgdsLDpruning" function in the R package SNPRelate v 1.18.1 with a linkage disequilibrium threshold of 0.1. The filtered VCF file was converted into the formats supported by software used in further analyses with PGDSpider v 2.1.1.5 [51].

Environmental variables

Climate variable layers were obtained from the CHELSA dataset [52], which includes 35 years of high spatial and temporal accuracy bioclimatic data (from 1979 to 2013). Given the effect of temperature and relative humidity on echolocation call frequency [22, 53], 19 climate variables related to temperature and precipitation (BIO01–BIO19, Additional file 1: Table S1) were extracted by the R package 'raster' [54] for each sampling site. First, we

performed an association analysis between climate variables. Because of the high correlation between many of the climate variables, a principal component analysis (PCA) was performed to summarize the environmental data and identify the data that explained the highest proportion of the environmental variance. The first principal component (PC) for climate variables was used as the climate variables in gene–environment association (GEA) analyses.

Discovery of SNPs putatively under divergent selection

Detected SNPs under selection are likely to have high false-positive rates caused by inaccurate population structure control or spatial autocorrelation of allele frequencies. Recently, use of a combined analysis approach has been increasing to obtain more robust results [55–58]. We used two different genome scan methods to detect loci putatively under divergent selection: outlier tests and GEA analyses. An outlier test was implemented in the program PCAdapt v 3.1.0 [59]. After choosing the number of PCs, the test statistic was computed based on the number K of PCs (in our case, K = 2). Based on the correlations between SNPs and the first two PCs, we computed and corrected *P*-values by Bonferroni correction ($\alpha < 0.1$) [60].

The programs Bayenv2 and LFMM v 1.5 were used to perform the GEA test. For Bayenv2 [61], 100,000 MCMC cycles were first run to obtain a covariance matrix. Bayenv2 used five independent runs with 50,000 iterations to estimate Bayes factors (BF) for each SNP, and the SNPs with an average BF > 3 were considered to be associated with the first climate variable PC. In LFMM [62], the number of latent factors that best described the population structure in the dataset was set based on the results of STRUCTURE v 2.3.4 [63, 64], Tracy–Widom tests were implemented in SmartPCA in EIGENSOFT v 6.1.4 [65] using the mean genomic inflation factor (λ) [66]. LFMM was run five times, and, as suggested in Frichot & François 2015 [67], we re-adjusted the *P*-values using the expected value of the FDR equal to q = 10%. Finally, we applied functional enrichment analyses for the outliers predicted using OmicShare tools [68] to test for over-representation of GO and KEGG functional categories.

Identification of one-to-one orthologous genes

Genomic protein sequences of *R. sinicus*, *Hipposideros armiger*, and *Pteropus alecto* were downloaded from NCBI. We obtained the longest transcript of each gene using a Python script and identified one-to-one orthologous genes among three *R. ferrumequinum* populations and three other species using the best reciprocal hit method (E-value, 1E– 5). For each predicted single-copy orthologous gene, we performed multiple alignments with PRANK v 170,427 (parameters: -f = FASTA -F -codon -noxml -notree -post) [69]. To reduce the rate of

false-positive predictions, Gblocks v 0.91b [70] was used to filter out sequencing errors, incorrect alignments, and non-orthologous regions based on codons with the following parameters: -t = C -b3 = 1 -b4 = 6 -B5 = N. After trimming, only the alignments with lengths greater than 100 bp were used for further analysis.

Phylogenetic tree and positive selection analyses

The phylogenetic tree was constructed using concatenated sequences of all filtered single-copy orthologous genes common to the three populations of greater horseshoe bat and three other bat species. Maximum likelihood analyses were run using the JTT + I + G + F model in RAxML v 8.2.10 [71], and relative support of internal nodes was assessed by rapid bootstrap (–f a −× 12,345) of 1000 replicates.

Using our tree topology as the guide tree, branch-site model in PAML4 (parameters: null hypothesis: model = 2, NSsites = 2, fix_omega = 1, omega = 1; alternative hypothesis: model = 2, NSsites = 2, fix_omega = 0, omega = 1) [72] was used to detect positive selection in the one-to-one orthologous genes. The three *R. ferrumequinum* populations were used as the foreground branches. We used a likelihood ratio test to detect positive selection on each foreground branch, and the genes with FDR less than 0.05 were considered positively selected genes (PSGs). After identifying PSGs, the empirical Bayes method was implemented to calculate posterior probabilities and record positively selected sites. We also performed enrichment analysis on the PSG dataset.

Results

Transcriptome assembly, quantification, and annotation

After transcriptome quality filtering, we obtained approximately 4 Gb of clean data for each of the 14 cDNA samples from the Illumina platform. For each sample, we obtained 26–36 million paired-end reads. After quality filtering, approximately 96% of the raw reads remained, and the sequencing result details are provided in Additional file 1: Table S2. We used Trinity methods with standard parameter values for the de novo assemblies of all samples and the three separate populations. Assembly of all samples yielded 70,704 transcripts that ranged from 201 to 27,462 bp with an N50 of 2532 bp. For the assembly of the three separate populations, we obtained 53,558–103,932 transcripts with N50 s from 2250 to 2811. Annotation results of unigenes in the four reference transcriptomes are provided in Additional file 1: Table S3. The completeness of the four transcriptomes was assessed using the BUSCO pipeline, which revealed that the majority of the Laurasiatherian core genes had been successfully recovered in all four assemblies. Of the 6253 single-copy Laurasiatherian orthologs, 78.0–84.0% were complete, 6.7–10.3% were fragmented, and 9.3–

11.7% were missing (Table 1). The good coverage indicated that all four assemblies were high-quality. For all four unigene sets, almost half had annotation hits to *R. sinicus* (Additional file 1: Table S4).

Phenotypic variation and differential gene expression between populations

The significant differences in RF were detected among the three *R. ferrumequinum* populations (Mann–Whitney U test, $P < 0.05$ in all cases). The HN population had the highest RF (mean = 75.5 kHz, SD = 0.36), followed by the YN population (mean = 73.0 kHz, SD = 0.93); the JL population (mean = 68.2 kHz, SD = 0.32) had the lowest RF (Additional file 2: Figure S1 a). For FL, the YN population had the longest FL (mean = 61.22 mm, SD = 1.90), followed by the HN population (mean = 60.08 mm, SD = 0.55) and then the JL population (mean = 58.60 mm, SD = 0.58). A significant difference was detected between HN and JL (Mann–Whitney U test, $P = 0.008$) (Additional file 2: Figure S1 b).

The pairwise comparisons showed significant gene expression differences among the three populations. Three samples, JL3, HN4, and YN2, were removed as outliers based on the MDS plot (Additional file 3: Figure S2). Based on the remnant samples, 8190 total DEGs were obtained from pairwise differential expression analysis, and the heatmap of the hierarchical clustering of all DEGs indicated that HN was more similar to YN than to JL based on gene expression patterns (Fig. 1a). For each comparison, a total of 7039, 4337, and 1611 DEGs in HN vs. JL, YN vs. JL, and HN vs. YN were detected, respectively (Fig. 1b, c).

Correlation between gene network modules and phenotype

Based on all 8190 DEGs (Fig. 1c), a gene co-expression network was constructed, and 12 gene modules were created by WGCNA (Fig. 2a). Two of them (M1, M6) were most significantly correlated with RF (both $P < 0.001$), and three modules (M3, M4, and M5) were significantly correlated with FL (all $P < 0.01$, Fig. 2b). M1,
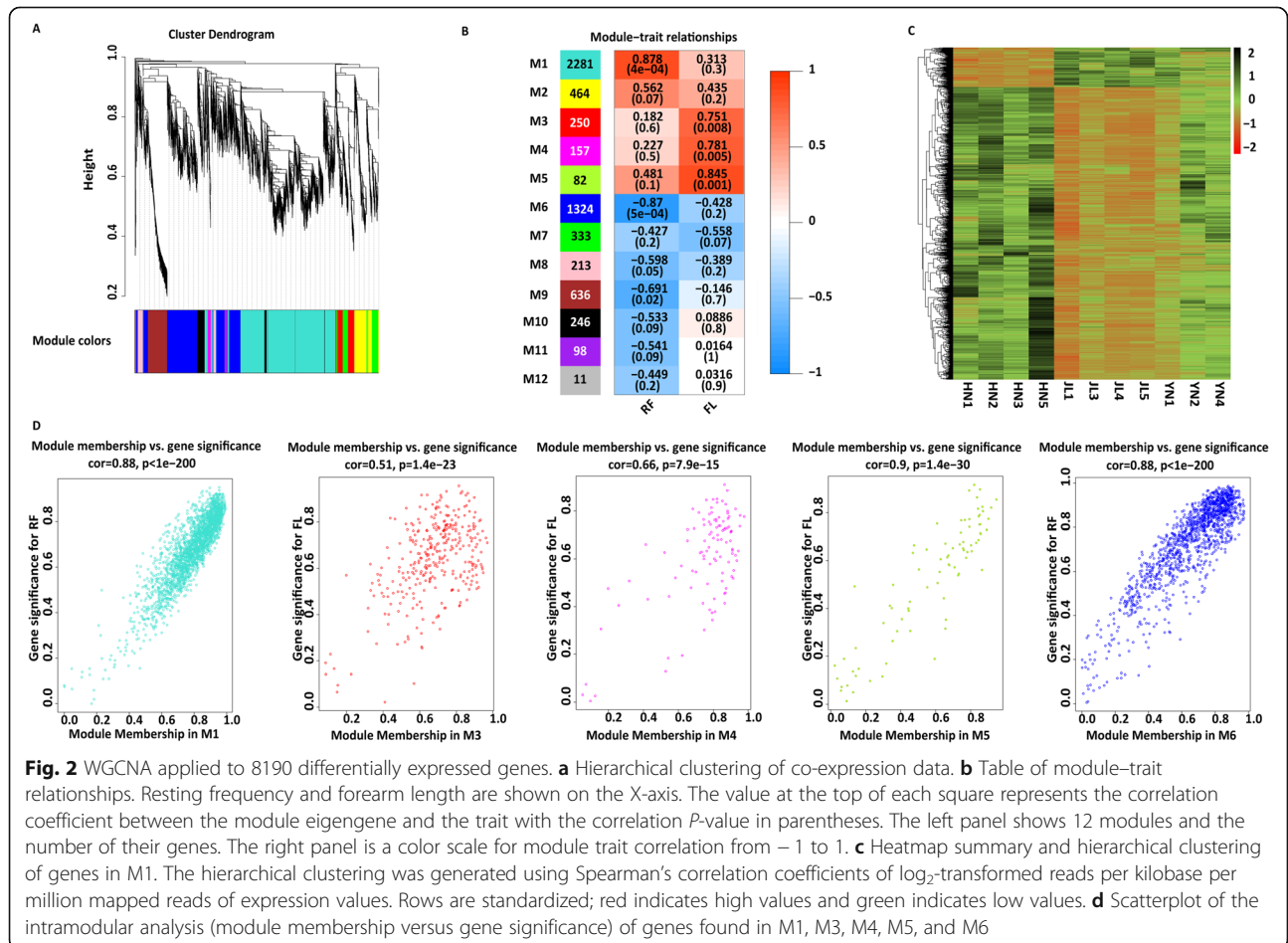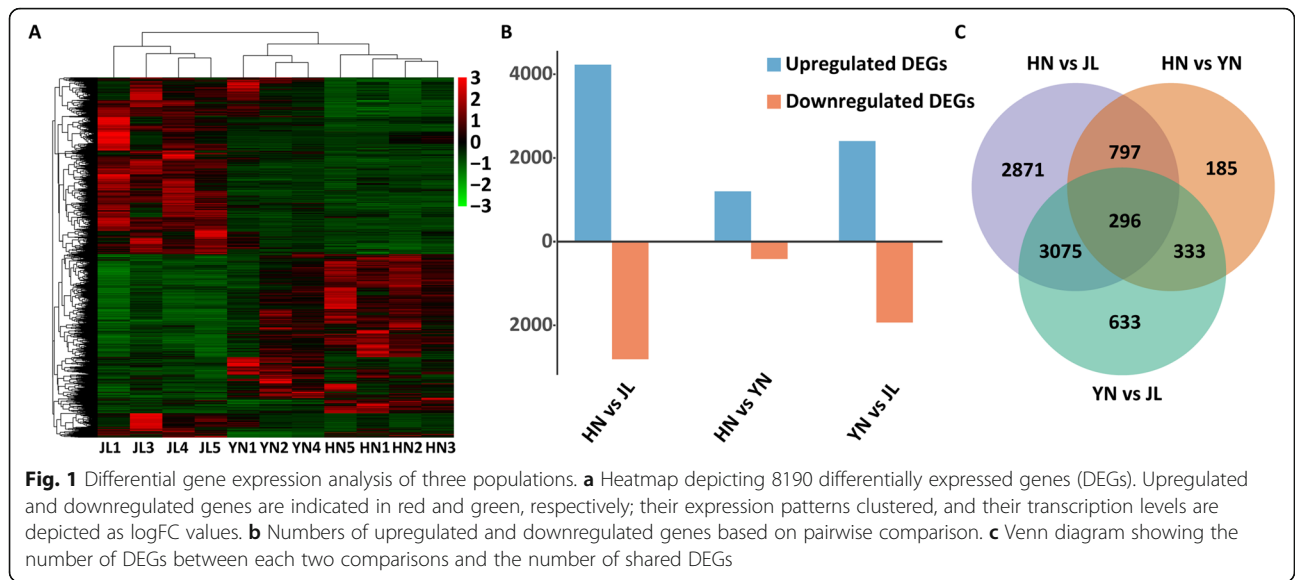
with 2281 unigenes, had the highest number of DEGs, and the heatmap of hierarchical clustering indicated large differences in gene expression between the three populations (Fig. 2c). We plotted scaled connectivity on the X-axis and gene significance (absolute value of the correlation coefficient, r, between gene expression and RF or FL) on the Y-axis for each module to visualize the relationships and significant positive correlations between gene significance and intramodular connectivity (Fig. 2d).
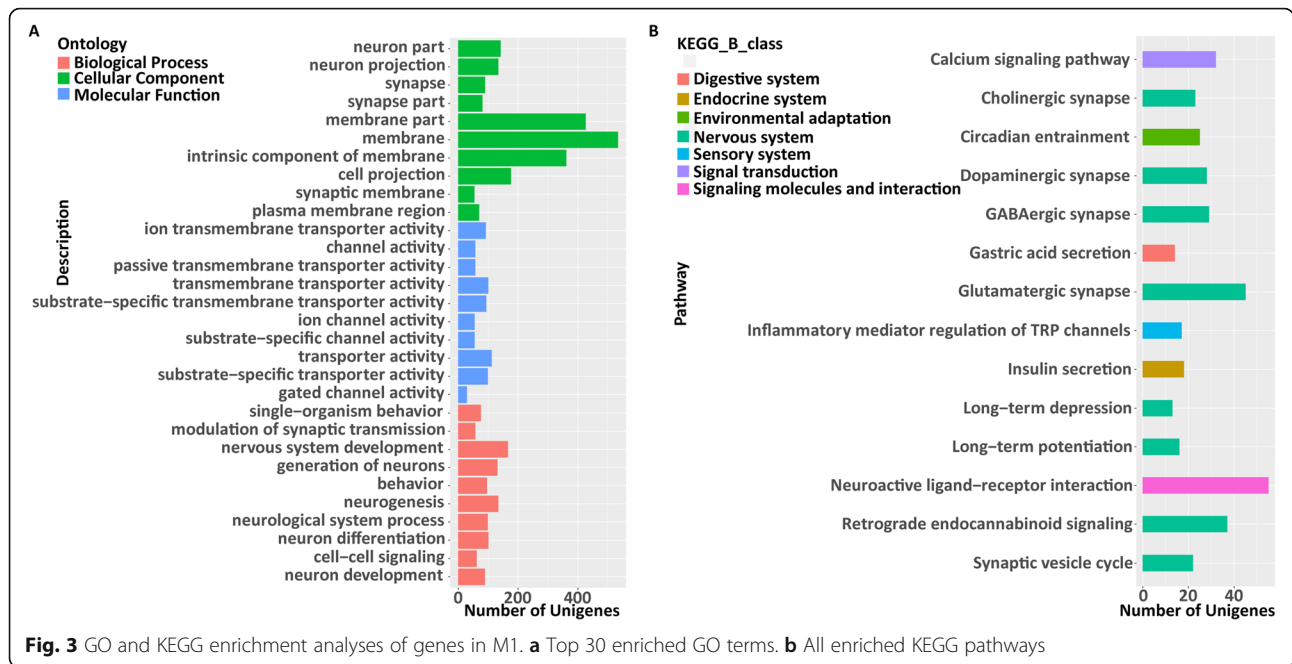
To obtain a clearer understanding of how differential gene expression patterns affect acoustic characteristics and body size, we then performed GO term and KEGG pathway enrichment analyses for each gene module. For unigenes included in M3, M4, M5, and M6, GO and KEGG enrichment analysis did not reveal any significantly enriched gene profile directly related to hearing (Additional file 1: Table S5 and Table S6). Unigenes in the M1 were significantly enriched in GO classifications that covered all three domains of ontology and were related to synaptic, neuron, membrane, and ion transporter functions, such as "synapse" (GO:0045202, FDR = 1.17E− 35), "neuron part" (GO:0097458, FDR = 1.25E− 43), "membrane part" (GO:0044425, FDR = 1.11E− 33), and "ion channel complex" (GO:0034702, FDR = 1E− 15) (Fig. 3a and Additional file 1: Table S5), and the KEGG pathway was also significantly enriched in synaptic functions (Fig. 3b and Additional file 1: Table S6). We also found that GO items related to learning were significantly enriched, such as "learning or memory" (GO:0007611, FDR = 1.01E− 17) and "learning" (GO:0007612, FDR = 2.83E− 07).

For unigenes in RF-related modules (M1 and M6), the GO term enrichment analysis results were similar, but two of those GO terms were associated with the response to sound: "response to auditory stimulus" (GO:0010996, FDR = 6.10E− 3) and "auditory behavior" (GO:0031223, FDR = 6.10E− 3) (Additional file 1: Table S7). For unigenes in FL-related modules (M3, M4 and M5), statistically significant items were related to a variety of functions, and many items were related to muscle or actin, such as

Table 1 Transcriptome completeness inferred from Benchmarking Universal Single-Copy Orthologs (BUSCO) search. ALL, JL, YN, and HN represent de novo references used to obtain expression profiles, single nucleotide polymorphisms (SNPs), and the reference transcriptome for each separate population. The number and percentage represent the number of genes inferred from BUSCO and their percentage of all 6253 single-copy Laurasiatherian orthologs

BUSCO statistic	ALL	JL	YN	HN
Complete BUSCOs	5252 (84.0%)	4880 (78.0%)	5067 (81.0%)	5159 (82.5%)
Complete - single-copy BUSCOs	5170 (82.7%)	4734 (75.7%)	4996 (79.9%)	5027 (80.4%)
Complete – duplicated BUSCOs	82 (1.3%)	146 (2.3%)	71 (1.1%)	132 (2.1%)
Fragmented BUSCOs	416 (6.7%)	647 (10.3%)	489 (7.8%)	427 (6.8%)
Missing BUSCOs	585 (9.3%)	726 (11.7%)	697 (11.2%)	667 (10.7%)

Fig. 1 Differential gene expression analysis of three populations. **a** Heatmap depicting 8190 differentially expressed genes (DEGs). Upregulated and downregulated genes are indicated in red and green, respectively; their expression patterns clustered, and their transcription levels are depicted as logFC values. **b** Numbers of upregulated and downregulated genes based on pairwise comparison. **c** Venn diagram showing the number of DEGs between each two comparisons and the number of shared DEGs

Fig. 2 WGCNA applied to 8190 differentially expressed genes. **a** Hierarchical clustering of co-expression data. **b** Table of module–trait relationships. Resting frequency and forearm length are shown on the X-axis. The value at the top of each square represents the correlation coefficient between the module eigengene and the trait with the correlation *P*-value in parentheses. The left panel shows 12 modules and the number of their genes. The right panel is a color scale for module trait correlation from − 1 to 1. **c** Heatmap summary and hierarchical clustering of genes in M1. The hierarchical clustering was generated using Spearman's correlation coefficients of $\log_2$-transformed reads per kilobase per million mapped reads of expression values. Rows are standardized; red indicates high values and green indicates low values. **d** Scatterplot of the intramodular analysis (module membership versus gene significance) of genes found in M1, M3, M4, M5, and M6

Fig. 3 GO and KEGG enrichment analyses of genes in M1. **a** Top 30 enriched GO terms. **b** All enriched KEGG pathways

"muscle system process" (GO:0003012, FDR = 6.30E– 14), "striated muscle cell development" (GO:0055002, FDR = 8.91E– 12), and "actin filament-based movement" (GO: 0030048, FDR = 8.01E– 7) (Additional file 1: Table S8).

Adaptive loci

After applying the filtering criteria as described in the Methods, we identified 21,945 SNPs distributed across 17, 772 unigenes for use in further analyses. PCA of the 21,945 SNPs revealed two dimensions that clustered by population (JL, HN, and YN; Fig. 4a). The Δ(K) from the STRUCTURE run showed that lnP(D) was highest when K = 2 (Fig. 4b), but the bar-plot for K = 2–4 showed that K = 3 represented the best possible number of populations (Fig. 4c). In LFMM, we chose K = 2 based on the Tracy–Widom results, because K = 2 had λ estimates closer to 1.0 than K = 3.

To explain most of the climate variation, PCA was performed to reduce the variables into fewer components. In our case, climate variations were highly consistent, and the contribution of the first PC was over 85% (Additional file 4: Figure S3). The association analysis of SNP markers and the first PC identified 250 and 948 unique SNP markers using Bayenv2 and LFMM, respectively. We also identified an extensive list of 1713 outliers using PCAdapt. Only 349 significant SNP markers were identified by at least two of the three outlier tests (Fig. 5a), and this finding indicated that some of these SNP markers were significantly associated with environmental variation. We predicted that these SNPs fell within 349 unigenes, and 310 of these significant SNP markers had annotations in the greater horseshoe bat transcriptomes (Additional file 1: Table S9). The identified unigenes represented a broad range of biological
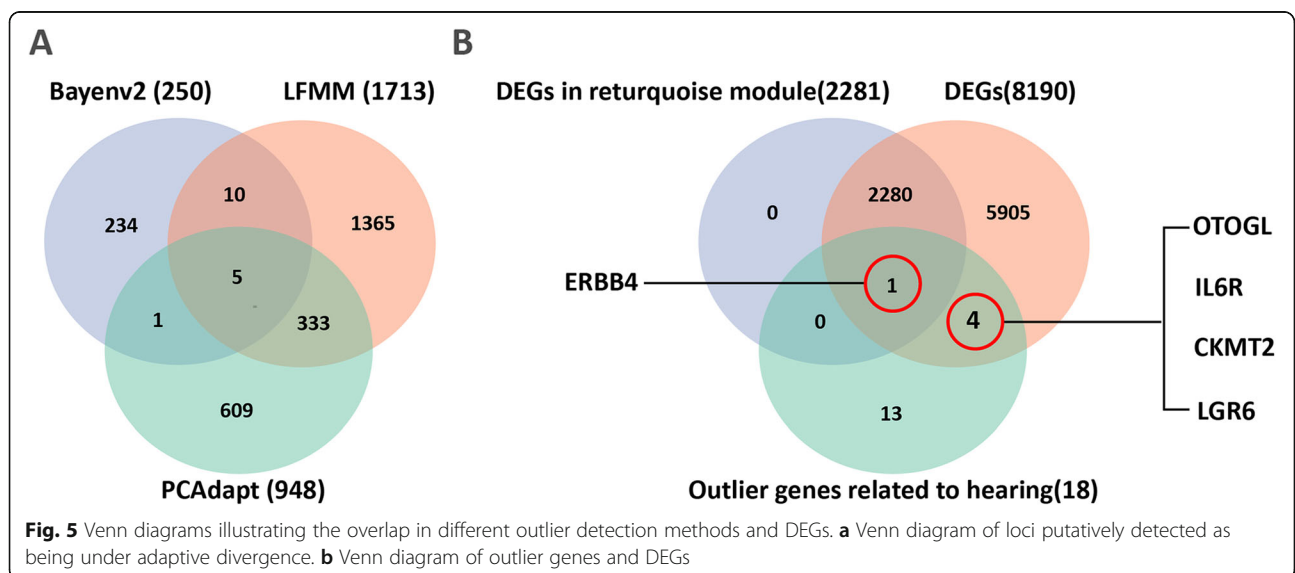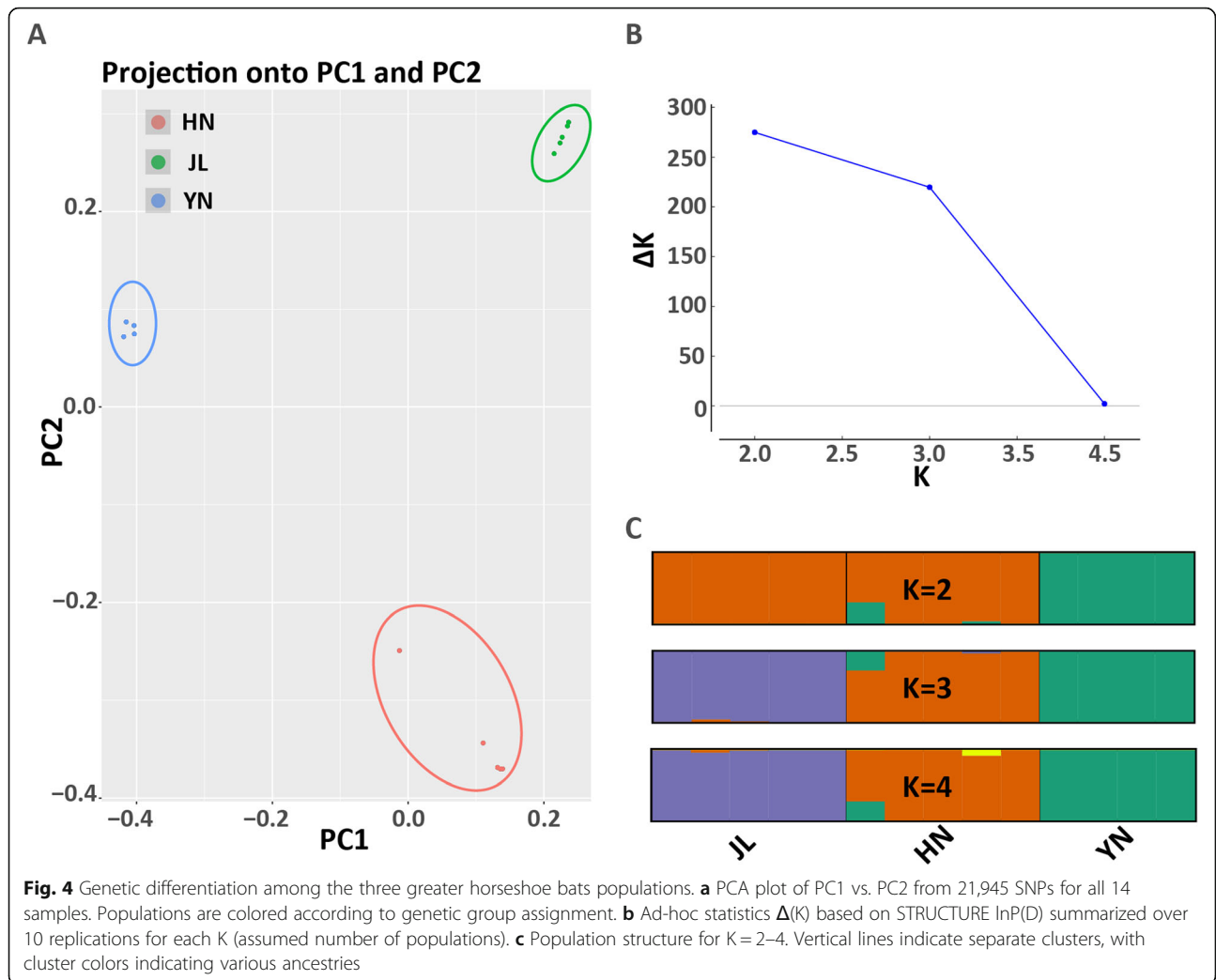
processes, such as immune (i.e., immunoglobulin superfamily member 3-like), transcription (i.e., aquaporin-3), and synaptic (i.e., synaptotagmin-16-like) processes. Furthermore, 18 of these unigenes were annotated in 18 genes that were thought to be involved in hearing-related processes or development of the hearing organ. Five of these auditory genes, *ERBB4* (Unigene0004747), *OTOGL* (Unigene0036784), *IL6R* (Unigene0051178), *CKMT2* (Unigene0004892), and *LGR6* (Unigene0005403), were significantly differentially expressed among the three populations, and *ERBB4* (Unigene0004747) was in the M1 of WGCNA (Fig. 5b).

Orthologous gene prediction and constructed phylogenetic tree

After removing low-quality sequences, we identified 4151 one-to-one orthologous gene pairs in the three populations, black fly fox (*P. alecto*), great roundleaf bat (*H. armiger*), and Chinese horseshoe bat (*R. sinicus*). After multiple sequence alignment and filtering, we retained 4105 one-to-one orthologous genes for model selection. All of these genes were concatenated into a single supergene dataset for model selection. The generated maximum likelihood tree of the three greater horseshoe bat lineages and other bat species was determined to be well resolved based on the high bootstrap value (100%). The phylogenetic relationship of those species indicated that JL was more closely related to HN than YN, and this was consistent with the findings of previous research [22] (Fig. 6).

Positive selection

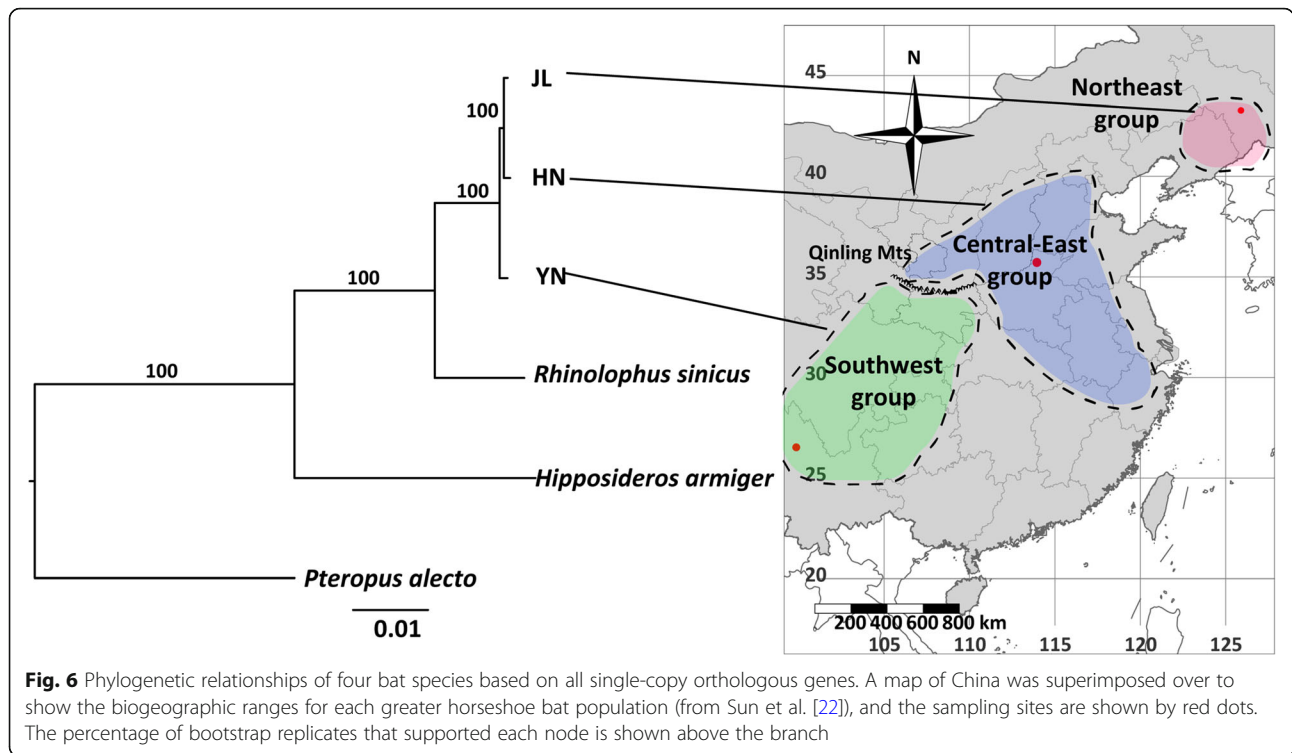We detected PSGs among the 4105 orthologous genes along the JL, HN, and YN branches using our

Fig. 4 Genetic differentiation among the three greater horseshoe bats populations. **a** PCA plot of PC1 vs. PC2 from 21,945 SNPs for all 14 samples. Populations are colored according to genetic group assignment. **b** Ad-hoc statistics **Δ**(K) based on STRUCTURE lnP(D) summarized over 10 replications for each K (assumed number of populations). **c** Population structure for K = 2–4. Vertical lines indicate separate clusters, with cluster colors indicating various ancestries

Fig. 5 Venn diagrams illustrating the overlap in different outlier detection methods and DEGs. **a** Venn diagram of loci putatively detected as being under adaptive divergence. **b** Venn diagram of outlier genes and DEGs

Fig. 6 Phylogenetic relationships of four bat species based on all single-copy orthologous genes. A map of China was superimposed over to show the biogeographic ranges for each greater horseshoe bat population (from Sun et al. [22]), and the sampling sites are shown by red dots. The percentage of bootstrap replicates that supported each node is shown above the branch

reconstructed tree topology, and detected 31, 6, and 12 PSGs, respectively (Additional file 1: Table S10). The PSGs had a variety of functions, such as in cilia formation, oxidoreductase activity, and immune process. The enrichment analyses of candidate PSGs showed that no category was significantly enriched.

Discussion

Recently, geographic variation in echolocation frequencies of bats has been frequently observed and widely investigated in the context of allopatric differentiation [4]. According to the sensory drive hypothesis, acoustic signals vary in association with local climatic conditions, and animal communication systems are adapted to local environments. Different environmental conditions may cause phenotypic and gene expression diversity between different populations, such as humans (*Homo sapiens*) [73] and fruit flies (*Drosophila melanogaster*) [74]. However, we know little about the molecular mechanisms underlying geographic variation of acoustic signals, especially in non-model organisms. Here, we report high-quality cochlea transcriptomic data for three representative populations of three Chinese *R. ferrumequinum* lineages. The quality of our sequencing allowed us obtain precise information about gene expression, sequence variation, and orthologous genes (Table 1). Our study will help elucidate genes or sequences with differential expression among populations that might be closely related to bat RF and could provide insight into the genetic basis of geographic variation in acoustic signals.

DEGs related to acoustic signal traits

In this study, we found that gene expression diversity among different populations and gene expression of cochleae could contribute to the RF variation of *R. ferrumequinum* echolocation calls. First, we found a significant change in gene expression patterns among the three populations (Fig. 1a), and we obtained a large number of DEGs using pairwise comparison methods (Fig. 1b and c). Second, HN and YN populations had the lowest difference in RF (difference of average RF = 2.5 kHz) and more similar gene expression patterns than the other two pairwise comparisons (HN vs JL and YN vs JL; Fig. 1a), even though this finding was not consistent with the phylogenetic relationship inferred from orthologous genes (Fig. 6) and the STRUCTURE results based on SNPs (Fig. 4b). Third, we used WGCNA, a very widely used R software package, to identify groups, known as modules, of correlated genes in suitable data [75]. We found strong associations between phenotypic traits (RF and FL) and five different expression gene modules (Fig. 2b). The unigenes with higher connectivity tended to be more strongly correlated with RF and FL. This finding indicated that these unigenes might play potentially important roles in the acoustic signal phenotype of bats.

Although we could not confirm which unigenes affect RF variation, our enrichment results showed that unigenes in M1 were enriched in pathways related to nervous system activity, such as a neuron, synapse, membrane, transporter, and channel activity, which are crucial in nerve

transmission in processes such as hearing. Several studies have shown that the same neural circuit, "the song system," controls the song of all songbirds, and genes activate the singing pattern [76–78]. Therefore, our results could indicate that the variable echolocation calls among different populations are regulated by the expression of genes related to the nervous system that adapt to different environments. We also identified potentially new GO categories for geographic variation in echolocation frequencies (learning or memory, and learning), as they were significantly enriched in our differentially expressed RF-related gene modules. Although there is no direct evidence that the greater horseshoe bat has song-learning abilities, Sun et al. suggested the possibility of cultural drift of the greater horseshoe bat [22]. Therefore, we cannot rule out the possibility that these bats have song-learning abilities. Although unigenes in the other four modules were significantly enriched in several GO items or KEGG pathways, their functions were not directly related to hearing, potentially because environmental factors and the state of the individuals could influence the expression of those unigenes. However, the expression changes of those unigenes might also affect hearing in some unknown ways. Therefore, we should not exclude the possibility that those unigenes could affect hearing before performing additional function analyses.

Interestingly, the enrichment results of unigenes in RF-related modules (M1 and M6) indicated another possibility: that the expression of genes related to sound response could affect bat RF. In channel catfish (*Ictalurus punctatus*), which can hear at higher frequencies, the genes in the pathway related to sound response were highly expressed in auditory organs, and could be associated with hearing [79]. Although experimental studies on the genes of these GOs are still limited, it would be interesting to study how the differential expression of those genes affect hearing. For unigenes in FL-related modules, many of the significantly enriched pathways were related to muscle and actin. Though we have no direct proof that FL is related to muscle mass, the regulation of skeletal muscle cell growth and proliferation can cause variation in body size [80], which could influence bat RF.

Outlier loci and genes under positive selection

We also attempted to identify the genetic basis of geographic variation in echolocation frequencies of bats based on nucleotide variations. In general, larger sample sizes are thought to be better for population genetic studies [81]. However, another studied showed that, even with a very small sample size (i.e., two individuals), accurate estimates of Fst could be obtained with a large number of SNPs (≥ 1500) [82]. For threatened species, species with reduced population sizes, and species that are difficult to obtain, such as bats, it is hard and detrimental to their populations to obtain enough samples.

Therefore, high-throughput screening technologies are promising for estimating genetic diversity and differentiation in such populations from very small sample sizes or populations undergoing reduction.

In this study, we found 349 candidate loci identified by at least two outlier tests. The outliers from Bayenv2, LFMM, and PCAdapt included 1.14, 4.31, and 7.81% of the total number of loci analyzed from our RNA-Seq dataset, respectively; these percentages are similar to the candidate coverage reported in other studies (2.52, 4.87, and 7.52%, respectively) using at least one of the same software programs [83, 84]. The overall percentage of outliers may be affected by species and the dataset. More specifically, the factors that determine the dataset, such as false-positive rate, sampling, genome size, power, and composition, would influence the numbers of SNPs and then affect the coverage of outliers [85]. The three methods may yield different results because they theoretically differ. Therefore, those genes recognized by more than one method are more likely to be under selection. Although there were not many candidate loci detected by overlap of the three outlier tests, especially between Bayenv2 vs. LFMM, Bayenv2 vs. PCAdapt, and all three methods, those 349 candidate loci detected by more than one method are likely important in adaptation.

Of all 349 candidate loci, 18 were in the unigenes annotated in 18 genes that are related to maintaining normal cochlea or neurofunction. Of the 18 genes, *PSAP* (Unigene0000367) [86], *ERBB4* (Unigene0004747) [87, 88], *LGR6* (Unigene0005403) [89], *BAK1* (Unigene0040895) [90], *OSTF1* (Unigene0005552) [91], and *TSHZ1* (Unigene0008250) [92] are necessary for hearing, with functions involved in the maintenance of nerve cell functions, hair cell protection, and skeleton formation. Mutations, deletions, or gene expression changes of *MUC19* (Unigene0008431) [93], *IL-6* (Unigene0051178) [94], *ZNF469* (Unigene0005986) [95], *CKMT2* (Unigene0004892) [96], *TMC2* (Unigene0035276) [97], *PYCR2* (Unigene0032863) [98], and *OTOGL* (Unigene0036784) [99] could lead to hearing loss. *ATE1* (Unigene0043244) [100], *TPRN* (Unigene0030527) [101], *TRIOBP* (Unigene0033824) [102], *GREB1* (Unigene0021168) [103], and *DPY19L2* (Unigene0017291) [104] are thought to be related to nonsyndromic hearing impairment by damaging structures in the inner ear. Moreover, five of these unigenes were differentially expressed, and Unigene0004747 (annotated *ERBB4*) was in the turquoise module and significantly related to RF (Fig. 5b). Although the mechanisms underlying how the environment affects those genes among bat populations remain unclear, functions of the genes indicated their importance in hearing, especially those differentially expressed mutant genes. In the echolocation process, those genes might affect sensitivity to calls of specific frequencies, their echolocation signals, by regulating neural

activity or inner ear structure, which receives these signals. Moreover, it cannot be ruled out that the geographic variation of acoustic signals is directly or indirectly influenced by other candidate loci.

The three populations might have undergone differential selection in different environments, because genes under positive selection varied in each lineage (Additional file 1: Table S10). For the genes under positive selection, we did not identify any enriched GO term or pathway. Moreover, we do not have direct evidence that those PSGs had functions related to hearing, but we cannot rule out the possibility that the candidate genes directly or indirectly influence greater horseshoe bat hearing.

Limitations of this study

Our results from several analytical methods demonstrated that there are distinct DEGs and sequence divergences among populations that exhibit geographic variation of acoustic signals, and the expression or sequence divergences of these genes may be related to the geographic variation in echolocation frequencies. However, as is typical for many non-model studies, our results were constrained by several technical and statistical factors. For bats, we had to collect samples from the wild where neither ecological conditions nor age can be strictly controlled. Moreover, the particularity of bats limits both sampling design and sample size, which can affect the identification of candidate genes, and many methods are known to be susceptible to false positives. By taking these factors into account, we were unable to definitively determine which candidate genes directly or indirectly affect bat hearing and RF without further verification. Although we can only speculate until further physiological studies are conducted, our evidence indicates that these genes and gene groups may be important in shaping the geographic variation of signal structure, and those genes will be the focus of a future study.

Conclusions

Using representative populations of the greater horseshoe bat in China, we obtained transcriptome data of *R. ferrumequinum* cochleae and analyzed gene expression and sequence data to examine gene expression changes and genotypes that contribute to geographic variation in echolocation calls. We identified some DEG modules and genomic variation among different populations related to RF or environment variables that could influence calling. The DEGs in modules that were significantly related to RF or FL were significantly associated with neurological and muscular functions, such as synaptic function, neuronal function, ion channel function, response to sound, learning behavior, muscle, and actin functions. As researchers pointed out in a previous study [22], both environmental adaptation and song learning are related to acoustic characteristics of greater horseshoe bats. Additionally, muscle mass, which could be related to body size, might also affect RF. Although specific genes could not be pinpointed, our results still indicated the potential importance of those genes. Genotype variation provided another way to understand the geographic variation of acoustic signals, and outlier loci and genes under positive selection might also be related to features of echolocation calls. Some of those genes have been reported to be related to hearing; five of them were significantly differentially expressed among populations, and *ERBB4* was in the module that was significantly correlated with RF. This finding indicated the importance of these genes, although the function of particular amino acid substitutions in these genes is unknown. In this study, even if RF could partially explain the changes in gene expression and sequence variation, future functional experiments will be necessary to validate their importance. Although there were limitations of our sampling strategy and analyses, our results partially explained the intraspecific geographic variation of acoustic signals and provided a direction for future research.

Supplementary information

Additional file 1: Table S1. Summary of investigated climate and environmental variables. **Table S2.** Summary of Trinity de novo transcriptome assembly statistics of *Rhinolophus ferrumequinum*. **Table S3.** Annotation result statistics between unigenes and databases. **Table S4.** Nr functional annotation of four unigene datasets hits to *Rhinolophus sinicus* and other species. **Table S5.** Complete results of the Gene Ontology (GO) enrichment analysis for gene sets in modules defined by Weighted gene co-expression network analysis (WGCNA). **Table S6.** Complete results of the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for gene sets in modules defined by WGCNA. **Table S7.** Complete results of the GO enrichment analysis for the merged gene set in modules significant associated with RF. **Table S8.** Complete results of the GO enrichment analysis for the merged gene set in modules significantly associated with FL. **Table S9.** Outlier loci ($N = 349$) identified in two tests for selection. 'SNP position' shows the position in the contig that contains the outlier loci. 'Tests' shows which tests identified the SNP as an outlier. **Table S10.** Detail of genes under positive selection.

Additional file 2: Figure S1. Boxplot of RF (a) and FL (b) between three populations. For each box plot, the box represents the 0.25 quantile, median, and 0.75 quantile. On either side of the box, the whiskers extend to the minimum and maximum values. *, indicates a statistically significant difference between populations. ° and ◇ represent outliers and means, respectively.

Additional file 3: Figure S2. Multidimensional scaling plot of 14 samples generated with edgeR. The axes represent gene expression levels among the three different populations. The samples in the red circles were removed from gene expression analysis.

Additional file 4: Figure S3. Scree plot of the principal component analysis. The first dimension explained 87.2% of the variance.

Abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs; CE: Central-east; CF–FM: Constant frequency–frequency modulating; DEGs: Differentially expressed genes; FL: Forearm length; GEA: Gene-environment association; MDS: Multidimensional scaling; NE: Northeast; PC: Principal Component; PCA: Principal component analysis; PSGs: Positively selected genes;

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate
All sampling was approved by the forestry departments throughout China and were conducted following regulations of Wildlife Conservation of the People's Republic of China (Chairman Decree [2004] No. 24).

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun 130117, China. [2]Institute of Resources & Environment, Henan Polytechnic University, Jiaozuo 454000, China. [3]Vector Laboratory, Institute of Pathogens and Vectors, Branch of Yunnan Provincial Key Laboratory for Zoonosis Control and Prevention, Dali University, Dali 671003, China. [4]College of Life Science, Jilin Agricultural University, Changchun 130118, China.

References
1. Ghalambor CK, McKay JK, Carroll SP, Reznick DN. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Funct Ecol. 2007;21(3):394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x.
2. Coyne JA, Orr HA. Speciation. Sinauer Associates, Inc: Sunderland; 2004.
3. Sobel JM, Chen GF, Watt LR, Schemske DW. The biology of speciation. Evolution. 2010;64(2):295–315. https://doi.org/10.1111/j.1558-5646.2009.00877.x.
4. Jiang T, Wu H, Feng J. Patterns and causes of geographic variation in bat echolocation pulses. Integr Zool. 2015;10(3):241–56. https://doi.org/10.1111/1749-4877.12129.
5. Shaw KL, Mullen SP. Genes versus phenotypes in the study of speciation. Genetica. 2011;139(5):649–61. https://doi.org/10.1007/s10709-011-9562-4.
6. Boughman JW. How sensory drive can promote speciation. Trends Ecol Evol. 2002;17(12):571–7. https://doi.org/10.1016/s0169-5347(02)02595-8.
7. Endler JA. Signals, signal conditions, and the direction of evolution. Am Nat. 1992;139:S125–S53. https://doi.org/10.1086/285308.
8. Ey E, Fischer J. The "acoustic adaptation hypothesis"—a review of the evidence from birds, Anurans and Mammals. Bioacoustics. 2009;19(1–2):21–48. https://doi.org/10.1080/09524622.2009.9753613.
9. Pröhl H, Hagemann S, Karsch J, Höbel G. Geographic variation in male sexual signals in strawberry poison frogs (Dendrobates pumilio). Ethology. 2007;113(9):825–37. https://doi.org/10.1111/j.1439-0310.2007.01396.x.
10. Zuk M, Rotenberry JT, Simmons LW. Geographical variation in calling song of the field *cricket Teleogryllus* oceanicus: the importance of spatial scale. J Evol Biol. 2008;14(5):731–41. https://doi.org/10.1046/j.1420-9101.2001.00329.x.
11. Mutumi GL, Jacobs DS, Winker H. Sensory drive mediated by climatic gradients partially explains divergence in acoustic signals in two horseshoe bat species, *Rhinolophus swinnyi* and *Rhinolophus simulator*. PLoS One. 2016; 11(1):e0148053. https://doi.org/10.1371/journal.pone.0148053.
12. Jacobs DS, Catto S, Mutumi GL, Finger N, Webala PW. Testing the sensory drive hypothesis: geographic variation in echolocation frequencies of Geoffroy's horseshoe bat (Rhinolophidae: *Rhinolophus clivosus*). PLoS One. 2017;12(11):e0187769. https://doi.org/10.1371/journal.pone.0187769.
13. Waters DA, Rydell J, Jones G. Echolocation call design and limits on prey size: a case study using the aerial-hawking bat Nyctalus leisleri. Behav Ecol Sociobiol. 1995;37(5):321–8. https://doi.org/10.1007/bf00174136.
14. Barclay RMR, Fullard JH, Jacobs DS. Variation in the echolocation calls of the hoary bat (*Lasiurus cinereus*): influence of body size, habitat structure, and geographic location. Can J Zool. 1999;77(4):530–4. https://doi.org/10.1139/cjz-77-4-530.
15. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. Nat Rev Genet. 2010;11(12):855–66. https://doi.org/10.1038/nrg2897.
16. Li G, Wang J, Rossiter SJ, Jones G, Cotton JA, Zhang S. The hearing gene Prestin reunites echolocating bats. Proc Natl Acad Sci U S A. 2008;105(37): 13959–64. https://doi.org/10.1073/pnas.0802097105.
17. Dong D, Lei M, Liu Y, Zhang S. Comparative inner ear transcriptome analysis between the Rickett's big-footed bats (*Myotis ricketti*) and the greater short-nosed fruit bats (*Cynopterus sphinx*). BMC Genomics. 2013;14(1):916. https://doi.org/10.1186/1471-2164-14-916.
18. Liu Y, Han N, Franchini LF, Xu H, Pisciottano F, Elgoyhen AB, et al. The voltage-gated potassium channel subfamily KQT member 4 (KCNQ4) displays parallel evolution in echolocating bats. Mol Biol Evol. 2012;29(5): 1441–50. https://doi.org/10.1093/molbev/msr310.
19. Shen YY, Liang L, Li GS, Murphy RW, Zhang YP. Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet. 2012;8(6): e1002788. https://doi.org/10.1371/journal.pgen.1002788.
20. Huihua Z, Shuyi Z, Mingxue Z, Jiang Z. Correlations between call frequency and ear length in bats belonging to the families Rhinolophidae and Hipposideridae. J Zool. 2003;259(2):189–95. https://doi.org/10.1017/s0952836902003199.
21. Wang J, Gao W, Wang L, Metzner W, Ma J, Feng J. Seasonal variation in prey abundance influences habitat use by greater horseshoe bats (*Rhinolophus ferrumequinum*) in a temperate deciduous forest. Can J Zool. 2010;88(3): 315–23. https://doi.org/10.1139/z10-005.
22. Sun K, Luo L, Kimball RT, Wei X, Jin L, Jiang T, et al. Geographic variation in the acoustic traits of greater horseshoe bats: testing the importance of drift and ecological selection in evolutionary processes. PLoS One. 2013;8(8): e70368. https://doi.org/10.1371/journal.pone.0070368.
23. Flanders J, Jones G, Benda P, Dietz C, Zhang S, Li G, et al. Phylogeography of the greater horseshoe bat, *Rhinolophus ferrumequinum*: contrasting results from mitochondrial and microsatellite data. Mol Ecol. 2009;18(2):306–18. https://doi.org/10.1111/j.1365-294X.2008.04021.x.
24. Flanders J, Wei L, Rossiter SJ, Zhang S. Identifying the effects of the Pleistocene on the greater horseshoe bat, *Rhinolophus ferrumequinum,* in East Asia using ecological niche modelling and phylogenetic analyses. J Biogeogr. 2011;38(3): 439–52. https://doi.org/10.1111/j.1365-2699.2010.02411.x.
25. Rübsamen R. Ontogenesis of the echolocation system in the rufous horseshoe bat, Rhinolophus rouxi (audition and vocalization in early postnatal development). J Comp Physiol A. 1987;161(6):899–913. https://doi.org/10.1007/978-1-4684-7493-0_37.
26. Basch ML, Brown RM 2nd, Jen HI, Groves AK. Where hearing starts: the development of the mammalian cochlea. J Anat. 2016;228(2):233–54. https://doi.org/10.1111/joa.12314.
27. Schnitzler HU. Control of doppler shift compensation in the greater horseshoe bat, *Rhinolophus ferrumequinum*. J Comp Physiol. 1973;82(1):79–92. https://doi.org/10.1007/bf00714171.
28. Schuller G, Pollak G. Disproportionate frequency representation in the inferior colliculus of doppler-compensating greater horseshoe bats: evidence for an acoustic fovea. Journal of comparative physiology ? A. 1979; 132(1):47–54. https://doi.org/10.1007/bf00617731.

29. Wang H, Zhao H, Huang X, Sun K, Feng JJ Sr. Comparative cochlear transcriptomics of echolocating bats provides new insights into different nervous activities of CF bat species. Sci Rep. 2018;8(1):15934.

30. Crow AL, Ohmen J, Wang J, Lavinsky J, Hartiala J, Li Q, et al. The genetic architecture of hearing impairment in mice: evidence for frequency-specific genetic determinants. G3. 2015;5(11):2329–39.

31. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63. https://doi.org/10.1038/nrg2484.

32. Todd EV, Black MA, Gemmell NJ. The power and promise of RNA-seq in ecology and evolution. Mol Ecol. 2016;25(6):1224–41. https://doi.org/10.1111/mec.13526.

33. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet. 2013;93(4):641–51. https://doi.org/10.1016/j.ajhg.2013.08.008.

34. Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, et al. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Res. 2016;44(19):e148. https://doi.org/10.1093/nar/gkw655.

35. Sacomoto GA, Kielbassa J, Chikhi R, Uricaru R, Antoniou P, Sagot MF, et al. KISSPLICE: de-novo calling alternative splicing events from RNA-seq data. BMC Bioinf. 2012;13 Suppl 6(6):S5. https://doi.org/10.1186/1471-2105-13-S6-S5.

36. Willing EM, Dreyer C, van Oosterhout C. Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. PLoS One. 2012;7(8):e42649. https://doi.org/10.1371/journal.pone.0042649.

37. Russo D, Mucedda M, Bello M, Biscardi S, Pidinchedda E, Jones G. Divergent echolocation call frequencies in insular rhinolophids (Chiroptera): a case of character displacement? J Biogeogr. 2007;34(12):2129–38. https://doi.org/10.1111/j.1365-2699.2007.01762.x.

38. Jiang T, Metzner W, You Y, Liu S, Lu G, Li S, et al. Variation in the resting frequency of *Rhinolophus pusillus* in mainland China: effect of climate and implications for conservation. J Acoust Soc Am. 2010;128(4):2204–11. https://doi.org/10.1121/1.3478855.

39. Specht R. Avisoft-saslab pro: sound analysis and synthesis laboratory. 2002. http://avisoft.com/SASLab_deutsch.pdf. Accessed 6 May 2017.

40. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.

41. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for biotechnology information. Nucleic Acids Res. 2019;47(D1):D23–D8. https://doi.org/10.1093/nar/gky1069.

42. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41. https://doi.org/10.1186/1471-2105-4-41.

43. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47(D1):D590–D5. https://doi.org/10.1093/nar/gky962.

44. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32(Web Server issue):W20–5. https://doi.org/10.1093/nar/gkh435.

45. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616.

46. Kvam VM, Liu P, Si Y. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot. 2012;99(2):248–56. https://doi.org/10.3732/ajb.1100340.

47. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966–7.

48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

49. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008;9(1):559. https://doi.org/10.1186/1471-2105-9-559.

50. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12(4):656–64. https://doi.org/10.1101/gr.229202.

51. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298–9. https://doi.org/10.1093/bioinformatics/btr642.

52. Karger DN, Conrad O, Bohner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Data Descriptor: Climatologies at high resolution for the earth's land surface areas. Sci Dat. 2017;4:170122. https://doi.org/10.1038/sdata.2017.122.

53. Snell-Rood EC. The effect of climate on acoustic signals: does atmospheric sound absorption matter for bird song and bat echolocation? J Acoust Soc Am. 2012;131(2):1650–8.

54. Hijmans RJ, van Etten J. raster: Geographic data analysis and modeling. R package version 2014; 2(8).

55. Christmas MJ, Biffin E, Breed MF, Lowe AJ. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. Mol Ecol. 2016;25(17):4216–33. https://doi.org/10.1111/mec.13750.

56. De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. Mol Ecol. 2014;23(19):4709–21. https://doi.org/10.1111/mec.12813.

57. Dillon S, McEvoy R, Baldwin DS, Rees GN, Parsons Y, Southerton S. Characterisation of adaptive genetic diversity in environmentally contrasted populations of *Eucalyptus camaldulensis Dehnh.* (river red gum). PLoS One. 2014;9(8):e103515. https://doi.org/10.1371/journal.pone.0103515.

58. Jordan R, Hoffmann AA, Dillon SK, Prober SM. Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: implications for adaptive potential to projected climate change. Mol Ecol. 2017;26(21):6002–20. https://doi.org/10.1111/mec.14341.

59. Luu K, Bazin E, Blum MG. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 2017;17(1):67–77. https://doi.org/10.1111/1755-0998.12592.

60. Weisstein EW. Bonferroni correction. 2004. http://mathworld.wolfram.com/BonferroniCorrection.html. Accessed 20 Aug 2018.

61. Gunther T, Coop G. Robust identification of local adaptation from allele frequencies. Genetics. 2013;195(1):205–20. https://doi.org/10.1534/genetics.113.152462.

62. Caye K, Francois O. LFMM 2.0: Latent factor models for confounder adjustment in genome and epigenome-wide association studies. Biorxiv. 2018:255893. https://doi.org/10.1101/255893.

63. Pritchard JK, Wen W, Falush D. Documentation for structure software: version 2. 2003. https://web.stanford.edu/group/pritchardlab/software/structure22/readme.pdf. Accessed 20 Aug 2018.

64. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

65. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich DJN. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904.

66. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190. https://doi.org/10.1371/journal.pgen.0020190.

67. Frichot E, François O. LEA: an R package for landscape and ecological association studies. Meth Ecol Evol. 2015;6(8):925–9. https://doi.org/10.1111/2041-210x.12382.

68. OmicShare tools. 2016. http://www.omicshare.com/tools/. Accessed 15 Dec 2018.

69. Veidenberg A, Medlar A, Loytynoja A. Wasabi: an integrated platform for evolutionary sequence analysis and data visualization. Mol Biol Evol. 2016;33(4):1126–30. https://doi.org/10.1093/molbev/msv333.

70. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

71. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033.

72. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91. https://doi.org/10.1093/molbev/msm088.

73. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013;23(7):1089–96. https://doi.org/10.1101/gr.152710.112.

74. Hutter S, Saminadin-Peter SS, Stephan W, Parsch J. Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 2008;9(1):R12. https://doi.org/10.1186/gb-2008-9-1-r12.

75. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4(1):Article17. https://doi.org/10.2202/1544-6115.1128.

76. Jarvis ED, Scharff C, Grossman MR, Ramos JA, Nottebohm F. For whom the bird sings: context-dependent gene expression. Neuron. 1998;21(4):775–88. https://doi.org/10.1016/s0896-6273(00)80594-2.

77. Maney DL, MacDougall-Shackleton EA, MacDougall-Shackleton SA, Ball GF, Hahn TP. Immediate early gene response to hearing song correlates with receptive behavior and depends on dialect in a female songbird. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2003;189(9):667–74. https://doi.org/10.1007/s00359-003-0441-z.

78. Frankl-Vilches C, Kuhl H, Werber M, Klages S, Kerick M, Bakker A, et al. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. Genome Biol. 2015;16(1):19.

79. Yang Y, Wang X, Liu Y, Fu Q, Tian C, Wu C, et al. Transcriptome analysis reveals enrichment of genes associated with auditory system in swimbladder of channel catfish. Comp Biochem Physiol Part D Genomics Proteomics. 2018;27:30–9. https://doi.org/10.1016/j.cbd.2018.04.004.

80. Wan Z, Lin G, Yue G. Genes for sexual body size dimorphism in hybrid tilapia (Oreochromis sp. x *Oreochromis mossambicus*). Aquac Fish. 2019. https://doi.org/10.1016/j.aaf.2019.05.003.

81. Manel S, Albert CH, Yoccoz NG. Sampling in landscape genomics. In: Data production and analysis in population genomics. Totowa: Humana Press; 2012. p. 3–12. https://doi.org/10.1007/978-1-61779-870-2_1.

82. Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol Ecol Resour. 2017;17(6):1136–47. https://doi.org/10.1111/1755-0998.12654.

83. Bernatchez S, Laporte M, Perrier C, Sirois P, Bernatchez LJME. Investigating genomic and phenotypic parallelism between piscivorous and planktivorous lake trout (*Salvelinus namaycush*) ecotypes by means of RAD seq and morphometrics analyses. Mol Ecol. 2016;25(19):4773–92.

84. Bowman LL, Kondrateva ES, Timofeyev MA, Yampolsky LYJM. Temperature gradient affects differentiation of gene expression and SNP allele frequencies in the dominant Lake Baikal zooplankton species. Mol Ecol. 2018;27(11):2544–59.

85. Harris SE, Munshi-South JJM. Signatures of positive selection and local adaptation to urbanization in white-footed mice (Peromyscus leucopus). Mol Ecol. 2017;26(22):6336–50.

86. Terashita T, Saito S, Nabeka H, Hato N, Wakisaka H, Shimokawa T, et al. Prosaposin-derived peptide alleviates ischaemia-induced hearing loss. Acta Otolaryngol. 2013;133(5):462–8. https://doi.org/10.3109/00016489.2012.750430.

87. Stankovic K, Rio C, Xia A, Sugawara M, Adams JC, Liberman MC, et al. Survival of adult spiral ganglion neurons requires erbB receptor signaling in the inner ear. J Neurosci. 2004;24(40):8651–61. https://doi.org/10.1523/JNEUROSCI.0733-04.2004.

88. Watanabe F, Kirkegaard M, Matsumoto S, Gont C, Mannstrom P, Ulfendahl M, et al. Signaling through erbB receptors is a critical functional regulator in the mature cochlea. Eur J Neurosci. 2010;32(5):717–24. https://doi.org/10.1111/j.1460-9568.2010.07347.x.

89. Smith ME, Groves AK, Coffin AB. Editorial: sensory hair cell death and regeneration. Front Cell Neurosci. 2016;10:208. https://doi.org/10.3389/fncel.2016.00208.

90. Falah M, Najafi M, Houshmand M, Farhadi M. Expression levels of the BAK1 and BCL2 genes highlight the role of apoptosis in age-related hearing impairment. Clin Interv Aging. 2016;11:1003–8. https://doi.org/10.2147/CIA.S109110.

91. Vermeren M, Lyraki R, Wani S, Airik R, Albagha O, Mort R, et al. Osteoclast stimulation factor 1 (Ostf1) KNOCKOUT increases trabecular bone mass in mice. Mamm Genome. 2017;28(11–12):498–514. https://doi.org/10.1007/s00335-017-9718-3.

92. Core N, Caubit X, Metchat A, Boned A, Djabali M, Fasano L. Tshz1 is required for axial skeleton, soft palate and middle ear development in mice. Dev Biol. 2007;308(2):407–20. https://doi.org/10.1016/j.ydbio.2007.05.038.

93. Kerschner JE, Khampang P, Erbe CB, Kolker A, Cioffi JA. Mucin gene 19 (MUC19) expression and response to inflammatory cytokines in middle ear epithelium. Glycoconj J. 2009;26(9):1275–84. https://doi.org/10.1007/s10719-009-9245-x.

94. Wakabayashi K, Fujioka M, Kanzaki S, Okano HJ, Shibata S, Yamashita D, et al. Blockade of interleukin-6 signaling suppressed cochlear inflammatory response and improved hearing impairment in noise-damaged mice cochlea. Neurosci Res. 2010;66(4):345–52. https://doi.org/10.1016/j.neures.2009.12.008.

95. Rohrbach M, Spencer HL, Porter LF, Burkitt-Wright EM, Burer C, Janecke A, et al. ZNF469 frequently mutated in the brittle cornea syndrome (BCS) is a single exon gene possibly regulating the expression of several extracellular matrix components. Mol Genet Metab. 2013;109(3):289–95. https://doi.org/10.1016/j.ymgme.2013.04.014.

96. Udar N, Atilano SR, Boyer DS, Chwa M, Memarzadeh M, Langberg J, et al. CKMT2 mutation in a patient with fatigue, age-related macular degeneration, deafness and atrial fibrillation. Biomed Genet Genom. 2017;2(2). https://doi.org/10.15761/bgg.1000133.

97. Kurima K, Yang Y, Sorber K, Griffith AJ. Characterization of the transmembrane channel-like (TMC) gene family: functional clues from hearing loss and epidermodysplasia verruciformis. Genomics. 2003;82(3):300–8. https://doi.org/10.1016/s0888-7543(03)00154-x.

98. Zaki MS, Bhat G, Sultan T, Issa M, Jung HJ, Dikoglu E, et al. PYCR2 mutations cause a lethal syndrome of microcephaly and failure to thrive. Ann Neurol. 2016;80(1):59–70. https://doi.org/10.1002/ana.24678.

99. Yariz KO, Duman D, Zazo Seco C, Dallman J, Huang M, Peters TA, et al. Mutations in OTOGL, encoding the inner ear protein otogelin-like, cause moderate sensorineural hearing loss. Am J Hum Genet. 2012;91(5):872–82. https://doi.org/10.1016/j.ajhg.2012.09.011.

100. Vona B, Neuner C, El Hajj N, Schneider E, Farcas R, Beyer V, et al. Disruption of the ATE1 and SLC12A1 genes by balanced translocation in a boy with non-syndromic hearing loss. Mol Syndromol. 2014;5(1):3–10. https://doi.org/10.1159/000355443.

101. Li Y, Pohl E, Boulouiz R, Schraders M, Nurnberg G, Charif M, et al. Mutations in TPRN cause a progressive form of autosomal-recessive nonsyndromic hearing loss. Am J Hum Genet. 2010;86(3):479–84. https://doi.org/10.1016/j.ajhg.2010.02.003.

102. Hilgert N, Smith RJ, Van Camp G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? Mutat Res. 2009;681(2–3):189–96. https://doi.org/10.1016/j.mrrev.2008.08.002.

103. Schrauwen I, Kari E, Mattox J, Llaci L, Smeeton J, Naymik M, et al. De novo variants in GREB1L are associated with non-syndromic inner ear malformations and deafness. Hum Genet. 2018;137(6–7):459–70. https://doi.org/10.1007/s00439-018-1898-8.

104. Waryah AM, Rehman A, Ahmed ZM, Bashir ZH, Khan SY, Zafar AU, et al. DFNB74, a novel autosomal recessive nonsyndromic hearing impairment locus on chromosome 12q14.2-q15. Clin Genet. 2009;76(3):270–5. https://doi.org/10.1111/j.1399-0004.2009.01209.x.

Publisher's Note