Current Zoology, 2024, **70**, 575–588
https://doi.org/10.1093/cz/zoad045
Advance access publication 30 September 2023
Original Article

The roles of different gene expression regulators in acoustic variation in the intermediate horseshoe bat revealed by long-read and short-read RNA sequencing data

Qianqian Li, Jianyu Wu, and Xiuguang Mao*

School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200062, China

*Address correspondence to Xiuguang Mao. E-mail: xgmao@sklec.ecnu.edu.cn
Handling editor: Rüdiger Riesch

Abstract

Gene expression changes contribute greatly to phenotypic variations in nature. Studying patterns of regulators of gene expression is important to fully understand the molecular mechanism underlying phenotypic variations. In horseshoe bats, the cochleae are finely tuned to echoes of call frequency. Here, using 2 recently diverged subspecies of the intermediate horseshoe bat (*Rhinolophus affinis hainanus* and *R. a. himalayanus*) with great acoustic variations as the system, we aim to explore relative roles of different regulators of gene expression (differential gene expression, alternative splicing (AS) and long non-coding RNAs (lncRNAs)) in phenotypic variation with a combination of Illumina short-read and Nanopore long-read RNA-seq data from the cochlea. Compared to *R. a. hainanus*, *R. a. himalayanus* exhibited much more upregulated differentially expressed genes (DEGs) and multiple of them may play important roles in the maintenance and damage repair of auditory hair cells. We identified 411 differentially expressed lncRNAs and their target DEGs upregulated in *R. a. himalayanus* were also mainly involved in a protective mechanism for auditory hair cells. Using 3 different methods of AS analysis, we identified several candidate alternatively spliced genes (ASGs) that expressed different isoforms which may be associated with acoustic divergence of the 2 subspecies. We observed significantly less overlap than expected between DEGs and ASGs, supporting complementary roles of differential gene expression and AS in generating phenotypic variations. Overall, our study highlights the importance of a combination of short-read and long-read RNA-seq data in examining the regulation of gene expression changes responsible for phenotypic variations.

Key words: alternative splicing, lncRNAs, noise-induced hearing loss, phenotypic variation, repair of hair cell.

Differences in gene expression contribute greatly to phenotypic variations in nature (Harrison et al. 2012). This was supported by numerous studies using transcriptome sequencing (i.e., RNA-seq) (reviewed in Alvarez et al. 2015), for example, melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*, Abolins-Abols et al. 2018), seasonal changes of coat color molt in the snowshoe hare (*Lepus americanus*, Ferreira et al. 2017), changes of bill morphology and plumage in Holarctic redpoll finches (Genus: *Acanthis*, Mason and Taylor 2015), and skin pigmentation variation in the Virginia opossum (*Didelphis virginiana*, Nigenda-Morales et al. 2018). Thus, regulations of gene expression can be considered as drivers of phenotypic evolution (Carroll 2008; Romero et al. 2012; Mank 2017). Non-coding RNAs, such as microRNAs and long non-coding RNAs (lncRNAs) have been shown to be key regulators of gene expression (Chen and Rajewsky 2007; Kittelmann and McGregor 2019; Gil and Ulisky 2020). Recently, the role of microRNAs in phenotypic evolution has been studied in a wide range of organisms (e.g., Maida cichlids, Franchini et al. 2016, 2019; and other teleost fishes, Desvignes et al. 2021; insects, Ma et al. 2021). However, little is known about the role of lncRNAs in phenotypic divergence to date.

Apart from gene expression, alternative splicing (AS) is another major form of transcriptional regulation to generate transcriptome diversity and hence phenotypic diversity by producing multiple isoforms from a single gene. AS is widespread in eukaryotes (Bush et al. 2017) and its frequency and abundance might be positively correlated with the level of phenotypic complexity (Xing and Lee 2006; Keren et al. 2010; Kornblihtt et al. 2013). In addition, AS has been suggested to underlie phenotypic and functional novelty in nature (Bush et al. 2017; Wright et al. 2022), such as the cryptic coloration in *Peromyscus* mice (Mallarino et al. 2017), difference in the spine length in the 3 spine sticklebacks *Gasterosteus aculeatus* (Howes et al. 2017) and the caste system plasticity in the bumble bee *Bombus terrestris* (Price et al. 2018). In contrast to differential gene expression, the role of AS in phenotypic diversity is understudied possibly due to the unavailability of high-quality genomes and annotations for most non-model species.

Short-read RNA-seq has been widely used to uncover the genetic basis of phenotypic variations in natural organisms by investigating expression changes of thousands of genes (Todd et al. 2016). However, short-read RNA-seq cannot accurately capture all transcripts, in particular for lowly expressed transcripts or large transcripts (>10k bp; Conesa et al. 2016; Stark et al. 2019). The recent development of long-read RNA sequencing (e.g., nanopore sequencing) can overcome

Received 29 May 2023; accepted 27 September 2023

the shortcomings of short-read RNA-seq by sequencing full-length (FL) transcripts at a single molecular level (Stark et al. 2019; Workman et al. 2019). This technology made it possible to quantify alternative transcripts instead of genes with high accuracy (Wang et al. 2019).

Acoustic signals produced by many animals (e.g., insects, birds, and mammals) are essential for survival and communication (Chen and Wiens 2020). For example, many bats use echolocation calls to explore their environment and detect prey (Schnitzler et al. 2003) and the communication potential of echolocation calls has also been suggested and studied (Schuchmann and Siemers 2010; Jones and Siemers 2011). Considerable divergence of echolocation calls has been reported in bats (Luo et al. 2019), in particular in horseshoe bats (*Rhinolophus*) (Chen et al. 2009; Zhang et al. 2009; Sun et al. 2013; Odendaal et al. 2014; Mutumi et al. 2016; Jacobs et al. 2017; Xie et al. 2017). Horseshoe bats have evolved a derived form of echolocation in which their hearing is narrowly tuned to that of the emitted call (Jones 2010). So variation of echolocation call frequency in *Rhinolophus* reflects variation of the hearing frequency in cochlea. In addition, the echolocation call frequency of horseshoe bats can be regarded as a morphological character related to their acoustic foveae (Siemers et al. 2005). As such, compared to the organs that are responsible for sound production (larynx) and sound interpretation (brain), cochlea tissues have been commonly used to investigate the genetic basis of echolocation and also echolocation call frequency variation in bats using comparative transcriptomics (Dong et al. 2013; Wang et al. 2018; Zhao et al. 2019; Sun et al. 2020; Li et al. 2022).

In this study, we focus on 2 recently diverged subspecies of the intermediate horseshoe bat *Rhinolophus affinis* showing acoustic divergence. Specifically, among the 3 *R. affinis* subspecies, 2 of them (*R. a. himalayanus* and *R. a. macrurus*) are from the mainland of China and a third (*R. a. hainanus*) is from Hainan Island. Previous phylogeographic studies on this species have shown that *R. a. himalayanus* was the first to diverge and a sister relationship was observed between the other 2 subspecies (Mao et al. 2010, 2013, 2014). Although the 3 subspecies were recently diverged (less than 1 million years ago, Mao et al. 2010), *R. a. himalayanus* emits much higher echolocation call (ultrasound) frequencies than the other 2 subspecies which show similar ultrasound frequencies (*R. a. himalayanus*, 87.12 kHz; *R. a. macrurus*, 73.68 kHz; *R. a. hainanus*, 70.85 kHz; Mao et al. 2010, 2013, 2014). Despite this high acoustic divergence, extensive introgression of mitochondrial DNA has been detected between *R. a. himalayanus* and *R. a. macrurus* in the eastern region of China (Mao et al. 2010; Mao and Rossiter 2020). Because sequence introgression might affect patterns of gene expression (e.g., Dannemann et al. 2017; McCoy et al. 2017), in this study we chose 2 allopatric subspecies (*R. a. himalayanus* and *R. a. hainanus*) as the system to conduct comparative transcriptomics. This system has been recently used to explore the role of gene expression changes and microRNAs in ultrasound frequency differences across the 3 *R. affinis* subspecies (Sun et al. 2020; Li et al. 2022). However, in Sun et al. (2020), de novo transcriptomes assembled based on short-reads RNA-seq were used as the reference to quantify the level of gene expression. Further studies using the high-quality genome or FL transcriptome as the references are needed to test for the previous results. More importantly, the role of AS in generating ultrasound frequency variation is still unknown so far.

In this study, our main aim is to investigate patterns of differential gene expression and AS and their relative roles in ultrasound frequency divergences between *R. a. himalayanus* and *R. a. hainanus*. To this aim, we used both short-read and long-read RNA-seq of the cochlear tissue from the same individuals of the 2 focal subspecies. Based on datasets of these 2 sequencing platforms, we identified candidate differentially expressed genes (DEGs) and pathways putatively associated with ultrasound frequency variations between the 2 subspecies. We also explore the role of lncRNAs in the regulation of expression changes of these genes. In mammals, it is now known that high-frequency sound can damage cochlear hair cells (Kurabi et al. 2017; Wagner and Shin 2019) although a recent study revealed that echolocating bats might be one of the few exceptions (Liu et al. 2021). We propose that higher frequency sound may cause more serious damage to cochlear hair cells compared with lower ones. The findings in Liu et al. (2021) have supported the existence of a protective mechanism in echolocating bats to defend cochlear hair cells against intense sound exposure. Thus, under our proposal, we further predict that genes involved in improving the survival of cochlear hair cells or repair of hair cell damage could be highly expressed in echolocating bats emitting higher ultrasound frequency (*R. a. himalayanus*) compared with ones with lower frequency (*R. a. hainanus*). Lastly, we examine the relative roles of differential gene expression and AS in generating phenotypic variations.

Materials and Methods

Sampling and collection of short-read and long-read RNA-seq data

In this study, we sampled 8 male individuals of *Rhinolophus affinis*, including 4 *R. a. himalayanus* from Anhui, China and 4 *R. a. hainanus* from Hainan Island, China (Figure 1A and Table 1). These bats have been used in our previous study to investigate the role of microRNAs in call frequency variation (Li et al. 2022). Here, we only sampled males to avoid the effect of sex differences on patterns of gene expression (Naqvi et al. 2019). In addition, no significant difference in call frequency between sexes was observed in *R. affinis* (Mao et al. 2014). Bats capture, adult bat identification, recording of the echolocation call frequency, tissue sampling, and short-read RNA-seq data collection have been described previously (Li et al. 2022). Briefly, for each adult bat, 2 cochleae were collected and 1 of them was used for short-read RNA-seq (Figure 1A). A total of 8 sequencing libraries were created using Illunima's TruSeq mRNA standard library preparation kit and sequenced on Novaseq 6000 mRNA (pair-end 150bp). Raw sequencing reads of these 8 short-read RNA-seq libraries were obtained from the NCBI's Sequence Read Archive (SRA) database under BioProject accession no. PRJNA764560.

For each subspecies, an equal amount of total RNA from 4 individuals was combined to perform long-read sequencing on Oxford Nanopore PromethION platform (long-read RNA-seq, Figure 1A). Briefly, libraries were generated according to the Nanopore community protocol using SQK-LSK109 kit and sequenced on R9 flow cells. Raw sequencing reads were then basecalled in guppy v3.2.10 with the default options to yield fastq files. Only reads with a minimum average quality score >7 were collected for subsequent analysis.

Transcript identification and annotation

We analyzed the long-read RNA-seq data of each subspecies using the same pipeline (Figure 1B). First, we assessed the quality of reads using NanoPlot v 1.8.1 (De Coster et al. 2018) and only reads with the score >7 were retained.

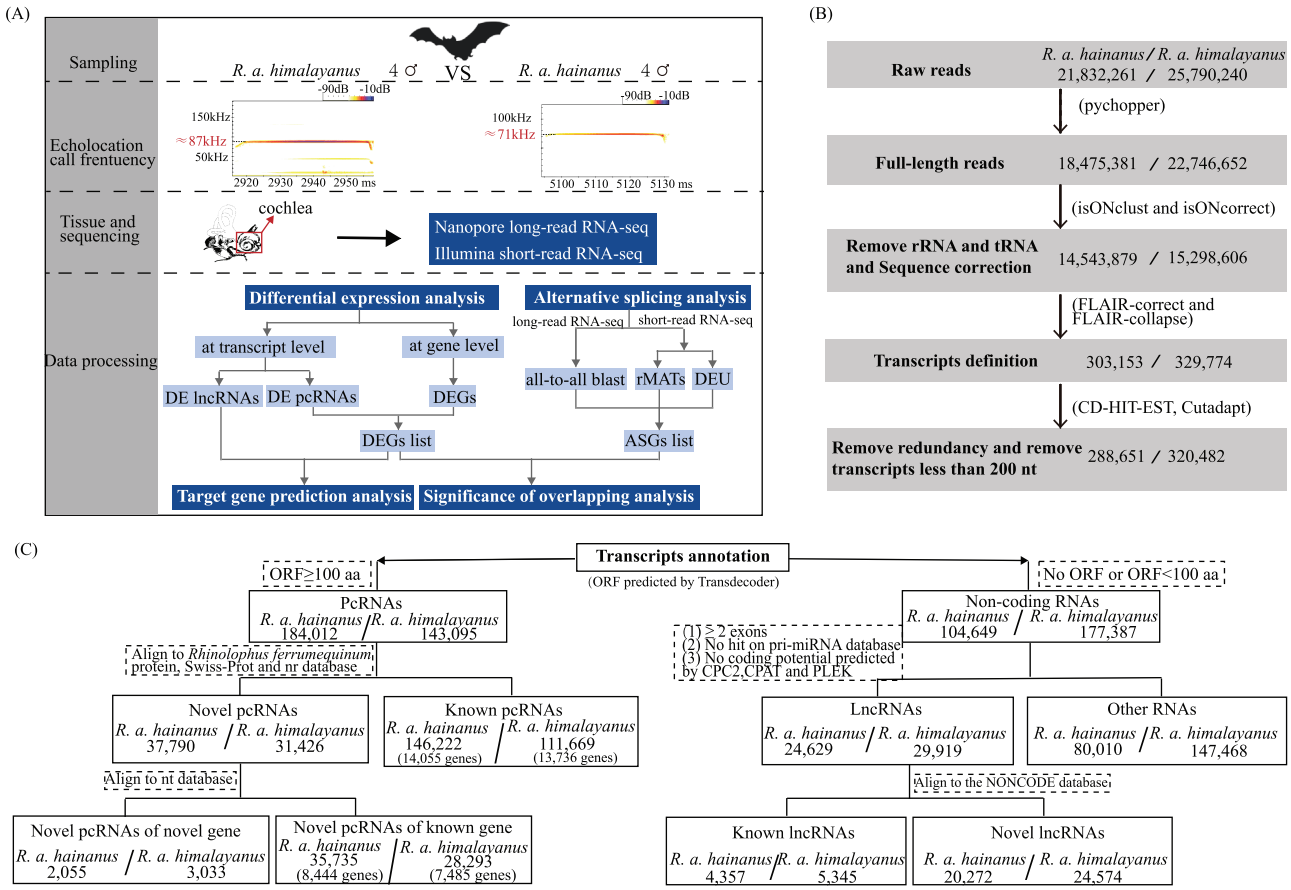

Figure 1. (A) Sampling and experimental design. Cochlea diagram was modified from the previous study (Dallos 1992). (B) Identification of transcripts in two subspecies (*R. a. hainanus* and *R. a. himalayanus*). (C) Annotation of transcripts including protein-coding transcripts (pcRNAs) and lncRNAs. The number in bracket represents the number of genes.

Table 1. Detailed information about samples and sequencing data (short-read and long-read RNA-seq) used in this study

Sample ID	Taxon	Sex	Call frequency (kHz)	Short-read RNA-seq (Li et al. 2022)	Long-read RNA-seq (this study)
Hai-05	*R. a. hainanus*	Male	71.0	7.8 Gb	52 Gb
Hai-16	*R.a. hainanus*	Male	71.4	8.2 Gb	
Hai-19	*R. a. hainanus*	Male	71.8	7.3 Gb	
Hai-20	*R. a. hainanus*	Male	70.5	7.2 Gb	
Him-12	*R. a. himalayanus*	Male	87.5	6.3 Gb	41 Gb
Him-13	*R. a. himalayanus*	Male	87.4	6.8 Gb	
Him-35	*R. a. himalayanus*	Male	87.4	6.5 Gb	
Him-36	*R. a. himalayanus*	Male	87.6	7.2 Gb	

Second, the full-length (FL) reads were identified with adapter sequences on both ends. Then, Pychopper v 2.5.0 (github.com/nanoporetech/pychopper) was used to trim adapters and orient reads. Poly A tails were identified and trimmed using Cutadapt v 2.6 (Martin 2011) with the parameter "--discard-untrimmed -a 'A{100}' -O 10." Third, by searching against the Silva rRNA (https://www.arb-silva.de) and tRNA database, transcripts of rRNAs and tRNA were removed. Fourth, we clustered reads using isONclust v 0.0.6.1 (Sahlin and Medvedev 2020) and error-correcting was conducted with isONcorrect v 0.0.8 (Sahlin and Medvedev 2021). Fifth, we defined the final transcripts using FLAIR v1.5.0 (Tang et al. 2020). Specifically, we first aligned all reads against the reference genome of *Rhinolophus affinis* (Zhao et al. unpublished manuscript) using minimap2 v 2.24-r1122 (Li 2018) with the parameter "-ax splice -k14 -secondary=no -splice-flank=yes." Then, SAMtools v 1.15.1 (Li et al. 2009) was used to generate sorted BAM files which were converted to BED12 with FLAIR. Next, we corrected the genomic alignment based on splice evidence from Illumina short-read RNA-seq using FLAIR-correct. FLAIR-collapse was used to generate a first-pass isoforms. Then, reads were realigned to the first-pass isoforms and filtered with stringent parameters. We removed transcripts with fewer than 3 supporting reads. In addition,

all supporting reads need to cover over 80% of the isoform and include more than 25 nt of the first and last exons. The resulting transcripts with less than 200 nt were further filtered out and redundancy was removed using CD-HIT-EST v 4.8.1 (Fu et al. 2012) with an identity of 99% and coverage of 90%.

To annotate transcripts, we identified Open Reading Frames (ORFs) of each transcript using TransDecoder v 5.5.0 (http://transdecoder.github.io) with default parameters. Transcripts with ORFs longer than 100 amino acids were considered as protein-coding RNAs (pcRNAs) and others were non-coding RNAs (Figure 2C). The longest ORF of each transcript was extracted and searched against the *Rhinolophus ferrumequinum* protein database (GCF_004115265.1_mRhiFer1_v1.p_protein.faa.gz), the Swiss-prot (Swiss-Prot protein sequence database) and nr (non-redundant protein sequence database) using BLASTP with an *E*-value of 10^{-6} and identity of 75%. Transcripts annotated in the above protein databases were considered as known pcRNAs and the rest were novel pcRNAs. The latter ones were further searched against the nt (Non-Redundant nucleotide sequence) database using BLASTN with *E*-value of 10^{-6} and identity of 75%. Then, transcripts with annotation in nt database were considered as novel pcRNAs from known genes and the rest were novel pcRNAs from unknown genes.

To identify lncRNAs, the non-coding RNAs were aligned to miRBase (release 22.1) and Rfam databases (release 14.8) to remove pri-miRNAs and other non-coding RNAs. Then, coding potential of the remaining lncRNAs were further assessed using Coding Potential Calculator (CPC2, https://github.com/gao-lab/CPC2_standalone), Coding-Non-Coding Index (CNCI, https://github.com/www-bioinfo-org/CNCI) and PLEK v 1.2 (Li et al. 2014) and those without coding potential predicted by all 3 softwares were considered as lncRNAs. Finally, we only retained lncRNA with more than 2 exons. The defined lncRNAs were searched against the known lncRNAs sequences in RNAcentral (https://rnacentral.org/) database using BLASTN with *E*-value of 10^{-6} and identity of 90%. Those with and without hits were considered as known and novel lncRNAs, respectively.

Differential expression analysis

Prior to differential expression analysis, we first generated a collection of transcripts expressed in both subspecies which were used as the reference transcriptome (Figure 2A). Specifically, in each subspecies short-read RNA-seq data were mapped to the transcripts identified with the long-read RNA-seq above using Kallisto v 0.48.0 (Bray et al. 2016). Then, mapped reads were quantified with Kallisto and counts of transcripts across samples were normalized using DESeq method in DESeq2 v1.30.1. (Love et al. 2014). Transcripts with CPM (counts per million) >1 in at least 2 samples of each subspecies were retained. Then, by mapping the short-read RNA-seq data of *R. a. hainanus* (or *R. a. himalayanus*) on the expressed transcripts of *R. a. himalayanus* (or *R. a. hainanus*), we determined the transcripts expressed in both subspecies and specific transcripts in each subspecies. Transcripts expressed in both subspecies were combined and redundancy was removed, resulting in a reference transcriptome used in the differential expression analysis below.

With transcripts generated using both long-reads and short-reads RNA-seq datasets, we conducted differential expression analysis at both transcript and gene levels. First, at the transcript-level we identified differentially expressed (DE) transcripts, including DE pcRNAs and DE lncRNAs. Specifically, short-reads of each sample were mapped to the reference transcriptome. Then, Kallisto was used to quantify and counts of transcripts across samples were normalized using DESeq2. Principal component analysis (PCA) based on normalized transcripts expression matrix of all samples revealed that samples of each subspecies were clustered into 2 groups (Supplementary Figure S1). Following Li et al. (2022), we used both DESeq2 and edgeR (Robinson et al. 2010) to perform differential expression analysis. DE transcripts were determined by both methods with significant results (Benjamini and Hochberg 1995, *P*adj < 0.05 in DESeq2 and FDR < 0.05 in edgeR) and $|\log_2(\text{fold change})| > 1$.

Second, differential expression analysis was conducted at the gene-level to identify differentially expressed genes (DEGs). Based on the pcRNA quantification above, gene quantification was conducted by combining counts of all transcripts of the focal gene (see also Zhu et al. 2021). PCA based on the normalized gene expression matrix of all samples also separated samples of each subspecies into different clusters (Supplementary Figure S1). Similarly, DEGs were identified using both DESeq2 and edgeR with the same procedures as above.

Identification of DE lncRNAs *cis*- and *trans*-target genes

Genes within 100k bp up- and down-stream of DE lncRNAs were identified as potential *cis*-target genes. To identify *trans*-target genes, RIblast v 1.2.0 (Fukunaga and Hamada 2017) was used to predict lncRNA-mRNA interactions by means of free energy minimization. Then, we calculated the Pearson's correlations of expression value between DE lncRNAs and 1,069 DEGs identified by at least 2 methods of differential expression analysis (see Results). Genes with Pearson's correlation coefficients >0.95 or <–0.95 were determined as *trans*-target genes. Finally, the lncRNA-mRNA co-expression networks were generated using Cytoscape (version 3.9.1, http://www.cytoscape.org/).

Alternative splicing analysis

Based on long-read RNA-seq, we conducted AS analysis using all-vs.-all BLAST methods (Liu et al. 2017). Specifically, all transcripts of 1 subspecies were searched against ones of the other subspecies using BLASTn with an *E*-value of 1e–20. The alignment results were processed using a series of Python scripts (https://github.com/liuxiaoxian/IsoSeq_AS_de_novo; Liu et al. 2017). Genes whose alignments meet all following criteria were considered as alternatively spliced genes (ASGs): 1) 2 transcripts from each subspecies were both longer than 1,000 bp; 2) the alignment includes 2 high-scoring segment pairs (HSPs) with the same direction; 3) 2 HSPs on 1 transcript should be conjuncted or with a small overlap no more than 5 bp, whereas 2 HSPs on the other transcript should show a gap with at least 100 bp (see details in Liu et al. 2017).

Based on short-read RNA-seq, we used 2 methods to identify AS events between the 2 subspecies. Short-reads of RNA-seq from each sample were filtered using TRIMMOMATIC version 0.38 (Bolger et al. 2014) with the parameters of SLIDINGWINDOW:4:20. We removed reads shorter than 150 bp to ensure the same length of reads as the input required by rMATs (see below). The processed reads were then aligned to the reference genome of *Rhinolophus affinis* using HISAT2

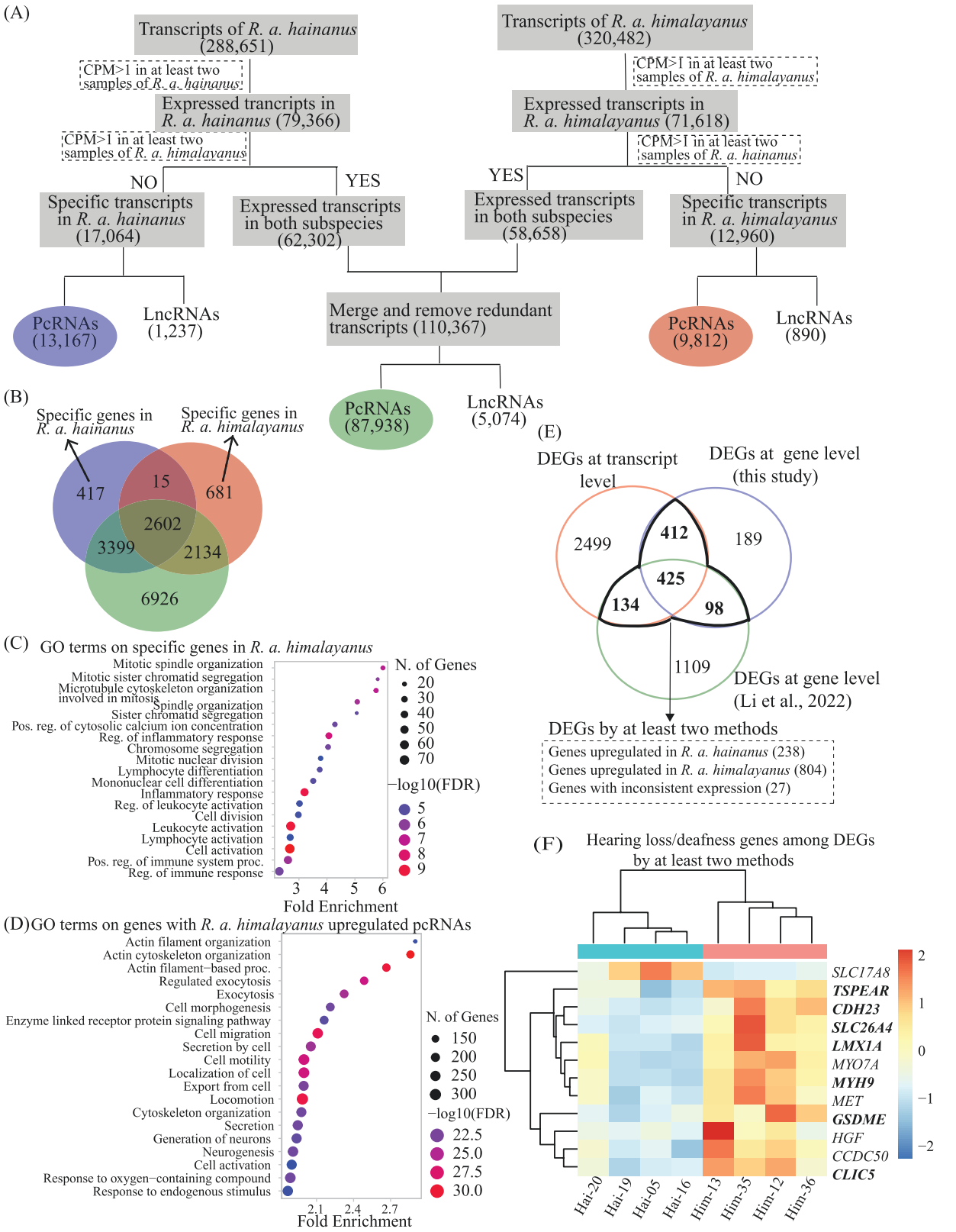

Figure 2. Results of differential expression analysis on protein-coding transcripts (pcRNAs) and genes. (A) Identification of specific and shared pcRNAs between the two subspecies. (B) Venn diagrams showing specific genes in each subspecies. (C) The top 20 significant Gene Ontology (GO) terms enriched on specific genes in *R. a. himalayanus*. (D) The top 20 significant GO terms enriched on genes with upregulated differentially expressed pcRNAs (DETs) in *R. a. himalayanus*. (E) Venn diagram showing differentially expressed genes (DEGs) identified by 3 methods (DETs at the transcript level; DEGs at the gene level by a combination of short-read RNA-seq and either FL transcriptome in this study or reference genome in Li et al. 2022). The numbers of DEGs identified by at least two methods were shown in bold. (F) Hierarchical clustering and heatmaps showing expression patterns of hearing loss/deafness genes among DEGs by at least two methods. DEGs by all 3 methods were shown in bold.

v2.2.0 (Kim et al. 2015) with default parameters. SAMtools v1.11 (Li et al. 2009) was used to convert the resulting SAM files to sorted BAM files. The mRNA alignments in sorted BAM files were used in the following analyses.

First, rMATs (v4.1.0) (Shen et al. 2014) were used to detect AS events including skipped exon, alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), mutually exclusive exons (MXE), and retained intron. This analysis was performed by comparing the PSI value (percent spliced in value) between the 2 subspecies. To reduce the false positives, AS events were determined with $0 < PSI < 1$ in at least half of the samples in each subspecies. Then, the difference in inclusion level (IncLevelDifference or ΔPSI) between the 2 subspecies was calculated for each AS event, and significance was determined using the false discovery rate (FDR) < 0.05 and IncLevelDifference > 0.1. Second, we used DEXSeq v 1.42.0 (Anders et al. 2012) to determine differential exon usage (DEU). The genome annotation file was flattened using the Python script "dexseq_prepare_annotation.py" in DEXSeq package. Then, the Python script "dexseq_count.py" in DEXSeq package was used to quantify exon-specific read counts and generate the count table for each sample. Significance difference in exon usage between the 2 subspecies was determined with $|\log_2 FC| > 1$ and $P\text{adj} < 0.05$.

Comparison of differential expression and AS

To test the relative roles of differential gene expression and AS in ultrasound frequency divergences between the two subspecies, we estimated the extent of overlap between 1,069 DEGs and 1,538 ASGs identified by at least 2 methods (see Results). Specifically, the expected number of genes that are both DEGs and ASGs were calculated as (total no. DEGs × total no. ASG)/total no. expressed genes. Then, we estimated the representation factor (RF) as the ratio of the observed number of overlapped genes and the expected number. The significance of more overlap than expected (RF > 1) or less overlap than expected (RF < 1) was determined using a hypergeometric test in R version 4.0.5 with the cutoff *P*-value of 0.05.

Functional enrichment analysis

We performed gene ontology (GO) term enrichment analysis for genes identified in differential expression and AS analyses, and also target genes of DE lncRNA using ShinyGo v0.77 (Ge et al. 2020). Significantly enriched GO biological processes were determined with an FDR cutoff = 0.05 and the top 20 enriched terms were shown. In addition, we also used a candidate gene approach by searching against the list of hearing loss/deafness genes in humans (122 genes from https://hereditaryhearingloss.org/, accessed on 6 February 2023).

Results

Characterization of the transcripts with long-read and short-read RNA-seq

Using Oxford Nanopore long-read sequencing technology, we generated a total of 21,832,261 (52 Gb) and 25,790,240 (41 Gb) high-quality raw reads for *R. a. hainanus* and *R. a. himalayanus*, respectively (Figure 1B and Table 2). These raw reads included 18,475,381 and 22,746,652 FL reads with an average length of 1,168 bp and 820 bp in *R. a. hainanus* and *R. a. himalayanus*, respectively (Figure 1B and Table 2). Length distribution revealed similar numbers of FL reads in each category of sequence length in each subspecies (Supplementary Figure S2). After a serial of filtering procedures, a final of 288,651 and 320,482 transcripts were retained in *R. a. hainanus* and *R. a. himalayanus*, respectively (Figure 1B and Table 2). Although *R. a. himalayanus* has more transcripts than *R. a. hainanus*, the average length of transcripts was longer in the latter than in the former (2,507 bp in *R. a. hainanus*; 1,502 bp in *R. a. himalayanus*). In addition, the maximum length of transcript in *R. a. hainanus* was almost twice the 1 in *R. a. himalayanus* (Table 2). Length distributions of the total transcripts also showed that *R. a. hainanus* had more number of long length transcripts (>3,000 bp) than *R. a. himalayanus*, whereas the latter had more number of transcripts with 1,001–2,000 bp (Supplementary Figure S2).

By predicting ORFs, we found 184,012 and 143,095 protein-coding transcripts (pcRNAs) with an average length of

Table 2. Detailed information about the long-read RNA-seq data and transcripts defined in each subspecies. PcRNAs represent protein-coding transcripts

Subspecies	*R. a. hainanus*	*R. a. himalayanus*
Number of raw reads	21,832,261	25,790,240
Number of full-length reads	15,602,610	16,421,998
Average length (bp)	1,168	820
Minimum length (bp)	200	200
Maximum length (bp)	17,055	102,875
N50 length (bp)	1,676	1,085
Number of total transcripts	288,651	320,482
Average length (bp)	2,507	1,502
Minimum length (bp)	200	200
Maximum length (bp)	14,483	7,289
N50 length (bp)	3,454	1,930
Number of pcRNAs	184,012	143,095
Average length (bp)	2,507	1,501
Minimum length (bp)	200	200
Maximum length (bp)	14,483	7,289
N50 length (bp)	3,454	1,930
Top 5 large pcRNAs (bp)	14,483; 13,773; 13,177; 13,103; 13,050	7,289; 6,778; 6,669; 6,510; 6,333
Number of genes	14,055	13,736
Number of lncRNAs	24,629	29,919
Average length (bp)	1,605	1,151
Minimum length (bp)	200	200
Maximum length (bp)	11,579	5,980
N50 length (bp)	2,505	1,471
Top 5 large lncRNAs (bp)	11,579; 11,348; 10,998; 10,430; 10,372	5,980; 5,093; 5,012; 4,839; 4,806

2,811 and 1,845 bp in *R. a. hainanus* and *R. a. himalayanus*, respectively (Figure 1C and Table 2). In contrast to the case of the total transcripts, *R. a. hainanus* had almost twice number of transcripts with 1,001–2,000 bp than *R. a. himalayanus* (Supplementary Figure S2), which may explain why the former had more number of pcRNAs than the latter. By searching against multiple protein databases, we found 146,222 and 111,669 known transcripts, encoded by 14,055 and 13,736 genes, in *R. a. hainanus* and *R. a. himalayanus*, respectively. Novel transcripts were further searched against the nt database, which revealed that a majority of novel transcripts (35,735 of 37,790 in *R. a. hainanus*; 28,293 of 31,426 in *R. a. himalayanus*) were from known genes and a minor of them (2,055 in *R. a. hainanus*; and 3,133 in *R. a. himalayanus*) were from novel genes (Figure 1C). In this study, we identified 5 large pcRNAs (>13 kb, Table 2) which are annotated as *ANK3*, *HUWE1*, *KIAA1109*, *DLG1*, and *SOGA1*.

After filtering out pri-miRNAs, other small RNAs and potentially coding transcripts, we detected 24,639 and 29,919 lncRNAs with an average length of 1,605 bp and 1,150 bp in *R. a. hainanus* and *R. a. himalayanus*, respectively (Figure 1C and Table 2). In contrast to the case of pcRNAs, *R. a. himalayanus* had a greater number of lncRNAs with 1,001–2,000bp than *R. a. hainanus* (Supplementary Figure S2). By searching against the current lncRNAs database, we found that only 17.69% and 17.86% of the lncRNAs were known (4,357 in *R. a. hainanus* and 5,345 in *R. a. himalayanus*) and a majority of them were novel (20,272 in *R. a. hainanus* and 24,574 in *R. a. himalayanus*). This might be because most available lncRNA annotations are from humans and other few model species, and lncRNAs show poor conservation across species (Derrien et al. 2012; Hezroni et al. 2015). Similar to pcRNAs above, we also identified multiple large lncRNAs (>10 kb). However, all of them are unknown and need to be annotated in the future.

With short-read RNA-seq data from the 4 individuals of each subspecies (see Supplementary Table S1 in Li et al. 2022), we quantified and confirmed the transcripts identified using the long-read RNA-seq above (Figure 2A). A total of 79,366 (67,046 pcRNAs and 3,331 lncRNAs) and 71,618 (53,820 pcRNAs and 4,029 lncRNAs) transcripts were confirmed to be expressed in *R. a. hainanus* and *R. a. himalayanus*, respectively. Among them, 17,064 (13,167 pcRNAs and 1,237 lncRNAs) and 12,960 (9,812 pcRNAs and 890 lncRNAs) transcripts were specifically expressed in *R. a. hainanus* and *R. a. himalayanus*, respectively (Supplementary Table S1). We combined transcripts expressed in both subspecies and removed redundancy, resulting in 110,367 transcripts (87,938 pcRNAs and 5,074 lncRNAs) which were used in the following differential expression analysis (Figure 2A and Supplementary Table S1).

To explore the functional roles of genes showing specifically expressed transcripts in each subspecies, we performed enrichment analysis on them. Prior to this analysis, we first excluded the genes whose transcripts were in the lists of expressed transcripts either in *R. a. hainanus* (or *R. a. himalayanus*) or shared in both subspecies (Figure 2B). A total of 417 genes (458 pcRNAs, Supplementary Table S1) were found to be specifically expressed in *R. a. hainanus* but we identified zero significant GO terms on these genes. In contrast, we detected 681 genes (801 pcRNAs, Supplementary Table S1) specifically expressed in *R. a. himalayanus* and they were mainly enriched into GO terms involved in the mitotic cell cycle (e.g., mitotic spindle organization, mitotic sister chromatid segregation, and microtubule cytoskeleton organization involved in mitosis) and regulation of immune response (Figure 2C and Supplementary Table S2). It was notable that 2 deafness genes were found to be specific genes in *R. a. himalayanus* (*ILDR1* and *LRTOMT*).

Differential expression analysis

With 2 different methods (DESeq2 and edgeR), we detected a total of 5,734 differentially expressed protein-coding transcripts (DETs) between the 2 subspecies, including 5,143 known and 591 novel transcripts (Supplementary Figure S3a and Supplementary Table S3). Among them, more DETs were found to be upregulated in *R. a. himalayanus* (3,529 DETs including 3,217 known and 312 novel) than in *R. a. hainanus* (2,205 DETs including 1,926 known and 279 novel; Supplementary Figure S3b,c and Supplementary Table S3). After annotation on these DETs, 1,515 and 2,235 genes were found to be upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively (Supplementary Table S3). It was notable that 62.9% of genes (2,359) had 1 DET (988 and 1,371 upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively). Among the genes with multiple DETs, 833 (248 in *R. a. hainanus* and 585 in *R. a. himalayanus*) showed consistent expression patterns of DETs. However, we found that DETs in 278 genes exhibited inconsistent expression patterns (i.e., some DETs of a gene were upregulated in a subspecies whereas other DETs of this gene were downregulated in the same subspecies). It was notable that among the 278 genes (Supplementary Table S3), 3 (*ATP2B2*, *COCH*, and *TJB2*) were found as hearing loss/deafness genes in humans and 2 of them have been recently recognized as echolocation genes in bats (*TJP2* in Jebb et al. 2020) and whales (*ATP2B2* in Yuan et al. 2021). Specifically, *ATP2B2* had 4 DETs with 3 highly expressed in *R. a. himalayanus* and 1 in *R. a. hainanus*; *COCH* had 8 DETs with 7 highly expressed in *R. a. hainanus* and 1 in *R. a. himalayanus*; *TJP2* had 2 DETs and each of them was highly expressed in each of the 2 subspecies.

By subtracting the 278 genes whose transcripts exhibited inconsistent expression patterns between subspecies, we identified 1,957 genes with upregulated DETs only in *R. a. himalayanus* and these genes were mostly enriched into GO terms related to actin filament-based process, exocytosis, cell morphogenesis, and cell migration (Figure 2D and Supplementary Table S4). Among them, 20 hearing loss/deafness genes were observed and 4 of them (*CCDC50*, *CDH23*, *GJB2*, and *MYH14*) were suggested to be echolocation genes (Yuan et al. 2021). We also found 1,237 genes with upregulated DETs only in *R. a. hainanus* and most of them were enriched into GO terms associated with regulation of mRNA processing and RNA splicing (Supplementary Figure S3d and Supplementary Table S4). In contrast, only 4 hearing loss/deafness genes were found among these genes and none of them has been recognized as echolocation genes to date. Transcripts encoded by the 24 hearing loss/deafness genes exhibited 2 clusters corresponding to each subspecies (Supplementary Figure S3k).

By combing reads counts of all transcripts of each gene, we conducted differential expression analysis at the gene level, resulting in a total of 1,124 differentially expressed genes (DEGs) with 220 and 904 upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively (Supplementary Figure S3e–g and Supplementary Table S3). No significant GO terms were observed for 220 upregulated DEGs in *R. a. hainanus*,

whereas 904 upregulated DEGs in *R. a. himalayanus* were mostly enriched into terms related to the immune function and exocytosis (Supplementary Figure S3h and Supplementary Table S4). Similar to genes with upregulated DETs only in *R. a. himalayanus* above, we also detected multiple significant GO terms associated with actin filament-based process and regulation of cell population proliferation (FDR = 0.000, Supplementary Table S4). In addition, we found 11 hearing loss/deafness genes among the 1,124 DEGs (1 in *R. a. hainanus* and 10 in *R. a. himalayanus*; Supplementary Figure S3l).

We compared the results of differential expression analyses at both transcript and gene levels to our previous study using the same RNA-seq data but by mapping reads to the reference genome (Li et al. 2022). A total of 1,069 DEGs were shared by at least 2 methods, including 238 upregulated in *R. a. hainanus*, 804 upregulated in *R. a. himalayanus*, and 27 with inconsistent expression patterns, respectively (Figure 2E and Supplementary Table S3). Functional enrichment analysis revealed that DEGs upregulated in *R. a. himalayanus* were enriched into multiple GO terms related to the immune function, exocytosis, actin filament-based process, and regulation of cell differentiation (Supplementary Figure S3j and Supplementary Table S5). In contrast, DEGs upregulated in *R. a. hainanus* were mainly enriched into terms related to RNA processing and splicing (Supplementary Figure S3i and Supplementary Table S5). Lastly, we also found 12 hearing loss/deafness genes among the 1,069 DEGs (1 in *R. a. hainanus* and 11 in *R. a. himalayanus*; Figure 2F).

Characterization of the lncRNAs target genes

Using DESeq2 and edgeR, we identified 411 differentially expressed (DE) lncRNAs between the 2 subspecies (Supplementary Figure S4a and Supplementary Table S6). Among them, 269 (46 known and 223 novel) and 142 (35 known and 107 novel) were upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively (Supplementary Figure S4b). It was notable that 1 known DE lncRNA (Homo sapiens nonprotein coding lnc-AMOTL2) was highly expressed in *R. a. himalayanus* and its original gene (*AMOTL2*), linked with VE-cadherin, has been reported to affect endothelial cell shape and transmit mechanical force (Hultin et al. 2014) which had been suggested to be involved in producing high frequency hearing in mammals (Kennedy et al. 2005; Fettiplace and Hackney 2006). Clustering and heatmap based on expression patterns of DE lncRNAs revealed clear separation of each subspecies (Supplementary Figure S4c).

To explore the functions of the DE lncRNAs, we identified their potential *cis*-and *trans*-target genes. For *cis*-target genes, we screened genes within 100K bp up- or down-stream of DE lncRNAs, resulting in 1,751 lncRNA-gene pairs (Supplementary Table S7). These pairs involved 102 DEGs (23 and 79 upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively). Functional enrichment analysis identified significant GO terms only on 79 target DEGs and most of them were related to immune function, NADP metabolic process, and cytoskeleton organization (Supplementary Figure S4d and Supplementary Table S8). A majority of DE lncRNAs were more than 50k bp from their target DEGs (Figure 3b) and only 10 were found to be overlapped with 9 target DEGs (Supplementary Table S7).

For *trans*-regulation analysis, we predicted the target DEGs of DE lncRNAs by expression levels. A total of 5,192 lncRNA-gene pairs were retained with Pearson's correlation coefficients >0.95 or <–0.95, involving 756 DEGs (182 and 574 upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively; Supplementary Table S7). Functional enrichment analysis on the 574 upregulated DEGs in *R. a. hainanus* identified multiple significant GO terms related to RNA splicing, cilium assembly, and RNA processing (Supplementary Figure S4f and Supplementary Table S8), whereas most of the 182 upregulated DEGs in *R. a. himalayanus* were enriched into GO terms associated with the immune response, exocytosis, regulation of cell proliferation and programmed cell death, actin filament-based process, and actin cytoskeleton organization (Supplementary Figure S4e and Supplementary Table S8). Among the *trans*-target genes we found 7 hearing loss/deafness genes and 6 of them were upregulated in *R. a. himalayanus* (Figure 3c). Expression networks between DE lncRNAs and these7 target DEGs revealed that most target genes interacted with multiple lncRNAs (Figure 3c). Interestingly, 2 DEGs (*MYH9* and *SLC17A8*; each was upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively) were found to share 1 network and interact with each other by the same lncRNA.

Alternative splicing analysis

Based on long-read RNA-seq data, we used all-vs.-all BLAST method to identify a total of 15,319 AS events, involving 5,804 genes (called alternatively spliced genes, ASGs ; Supplementary Table S9). The average length of gap between each pair was 507 bp ranging from 100 to 7,902 bp. Based on short-read RNA-seq data, we used 2 methods to identify AS events. With rMATs, a total of 3,352 AS events were identified and these AS events were found to occur in 1,932 genes (Supplementary Table S9). With the differential exon usage (DEU) method, we identified 1,849 exons showing significant differences in exon usage between subspecies, occurring in 1,267 genes (Supplementary Table S9).

By comparing the lists of ASGs identified by each of the 3 methods (all-vs.-all BLAST, rMATs, and DEU), 1,538 ASGs were overlapped by at least 2 methods (Figure 4A and Supplementary Table S9). Functional enrichment analysis identified multiple significant GO terms associated with neuron projection development, microtubule-based process, RNA processing, and cytoskeleton organization (Figure 4B and Supplementary Table S10). In addition, we found 17 hearing loss/deafness genes and 3 of them were identified as ASGs by all 3 methods (*CCDC50*, *TJP2,* and *SLC12A2*; Figure 4A). The AS results of 2 of them (*CCDC50* and *TJP2*) have been shown as examples here (Supplementary Figure S5 and Figure 4C).

Comparisons of differential expression and alternative splicing

We observed significantly less overlap than expected between DEGs and ASGs (RF = 0.72, $P = 0.0006$; Supplementary Figure S6a). Functional enrichment analysis on those only DEGs and only ASGs also identified different functional categories of significant GO terms. Specifically, most of the only DEGs were enriched into terms associated with the immune response, exocytosis, cell migration, actin cytoskeleton organization, and regulation of cell death (Supplementary Figure S6b and Supplementary Table S11), whereas the only ASGs were mostly involved in cytoskeleton-dependent intracellular transport, microtubule-based process, RNA processing, and neuron development (Supplementary Figure S6c and

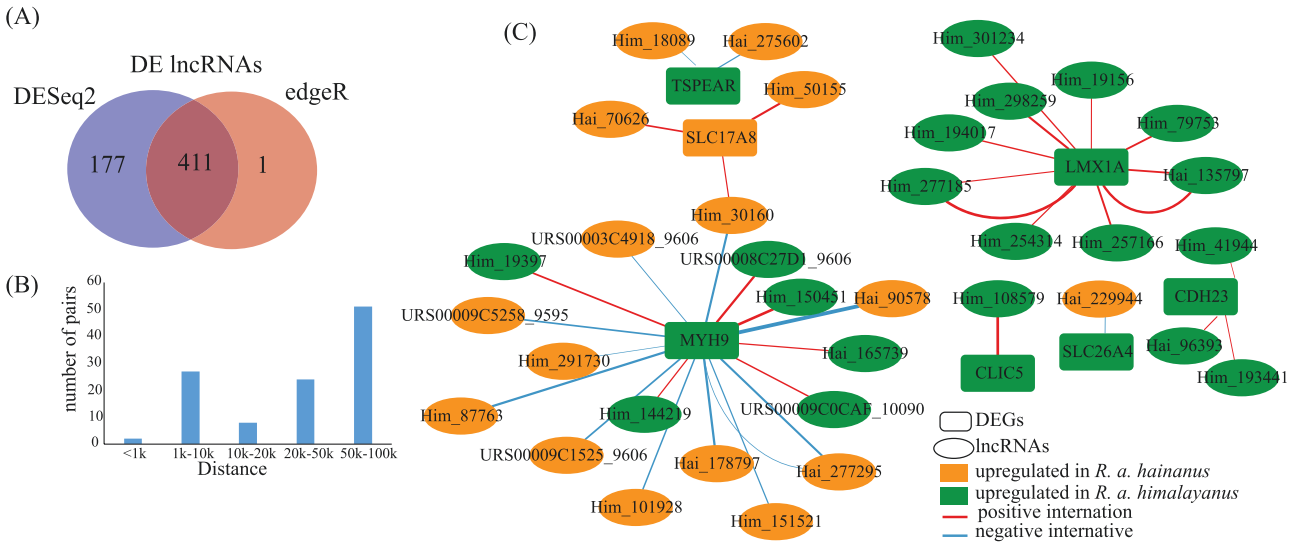

Figure 3. Results of differential expression analysis on lncRNAs and target prediction. (A) Venn diagram showing differentially expressed lncRNAs (DE lncRNAs). (B) Summary of distance between DE lncRNAs and their target DEGs. (C) Networks showing the interaction of co-expressed DE lncRNAs and their target DEGs that were hearing loss/deafness genes.

Supplementary Table S11). It was expected that the overlapped DEGs and ASGs were enriched into GO terms that were similar to ones identified on both DEGs and ASGs (Supplementary Figure S6d and Supplementary Table S11).

Discussion

To investigate the importance of differential gene expression in acoustic divergence, we identified and studied the DEGs in cochlea between the 2 subspecies. Our results showed that compared with *R. a. hainanus*, more known hearing loss/deafness genes and genes related to actin-based cytoskeleton were highly expressed in *R. a. himalayanus* and proteins encoded by some of these genes may provide protective roles for auditory hair cells, which was consistent with our proposal that higher ultrasound frequency may cause more serious damage to cochlear hair cells compared with lower one. In this study, to identify accurate DEGs, we applied 3 different analysis methods and found consistent patterns. First, we observed a greater number of genes highly expressed in *R. a. himalayanus* than in *R. a. hainanus* (Supplementary Figure S3a,e and Figure 2E). In addition, we found more numbers of hearing loss/deafness genes among the upregulated DEGs in *R. a. himalayanus* than those in *R. a. hainanus* (Supplementary Figure S3k,l and Figure 2F). Among them, *CDH23* has been shown to be associated with noise-induced hearing loss in humans and mice (Kowalski et al. 2014) and this gene is essential for the breakage and repair of tip links in mammalian hair cells (Wagner and Shin 2019). Consistent with this, *LCN2*, a gene among the top 10 highly expressed genes in *R. a. himalayanus* identified by all 3 methods (Supplementary Table S3), has also been reported to be upregulated in noised-injured cochlear cells in rats to act as a protective mechanism against apoptosis (Han et al. 2012). In addition, several DEGs, encoding mechano-electrical sensitive channels (e.g., *KCNQ1* and *PIEZO1*), were also upregulated in *R. a. himalayanus* (Supplementary Table S3). Second, we found similar functional categories on those highly expressed genes identified by each method. Specifically, genes highly expressed in *R. a. himalayanus* were mainly involved in actin filament-based process and regulation of cell population proliferation, whereas ones in *R. a. hainanus* were associated with RNA processing and RNA splicing. A complement to the results of the differential expression analysis above, we found that the genes specifically expressed in *R. a. himalayanus* were mostly involved in the mitotic cell cycle process (e.g., spindle organization, sister chromatid segregation, and nuclear division, Figure 2C and Supplementary Table S2), suggesting that regeneration might occur to replace the damaged and dead cells in *R. a. himalayanus* which exhibits a higher ultrasound frequency than *R. a. hainanus*. Consistent with this proposal, *ILDR1*, specifically expressed in *R. a. himalayanus*, is involved in deafness with cochlear outer hair cell degeneration in mice (Higashi et al. 2015; Sang et al. 2015) and has been shown to be crucial for the survival of auditory hair cells (Morozko et al. 2015). In addition, *Atoh8*, one of the Atoh family members, was found to be upregulated in *R. a. himalayanus* and protein encoded by this gene was related to human muscle fiber regeneration (Güttsches et al. 2015). Overall, our current study has identified several candidate genes that may play important roles in the maintenance and repair of mammalian hair cells and these genes can be used as therapy targets for noise-induced hearing loss in the future (see also Wagner and Shin 2019).

As for the 3 methods of differential gene expression analyses, we found that compared with methods at the gene level, more genes were identified to be differentially expressed at the transcript level, see Figure 2E), consistent with results in previous studies (e.g., Zhao et al. 2016). This might be caused by some genes with multiple transcripts that exhibit inconsistent expression patterns between groups. When measuring expression at the gene level, these genes may not be identified as DEGs. We detected 278 such genes in this study (Supplementary Table S3). Other comparative transcriptomic studies at both transcript and gene levels also revealed that quantification of RNA expression at the gene level could fail to identify DEGs when only specific transcript isoforms with significant expression differences (Zhao et al. 2016; Zhu et al. 2021; Liu et al. 2022). Thus, our current study, together with previous ones in other organisms, suggests that if FL

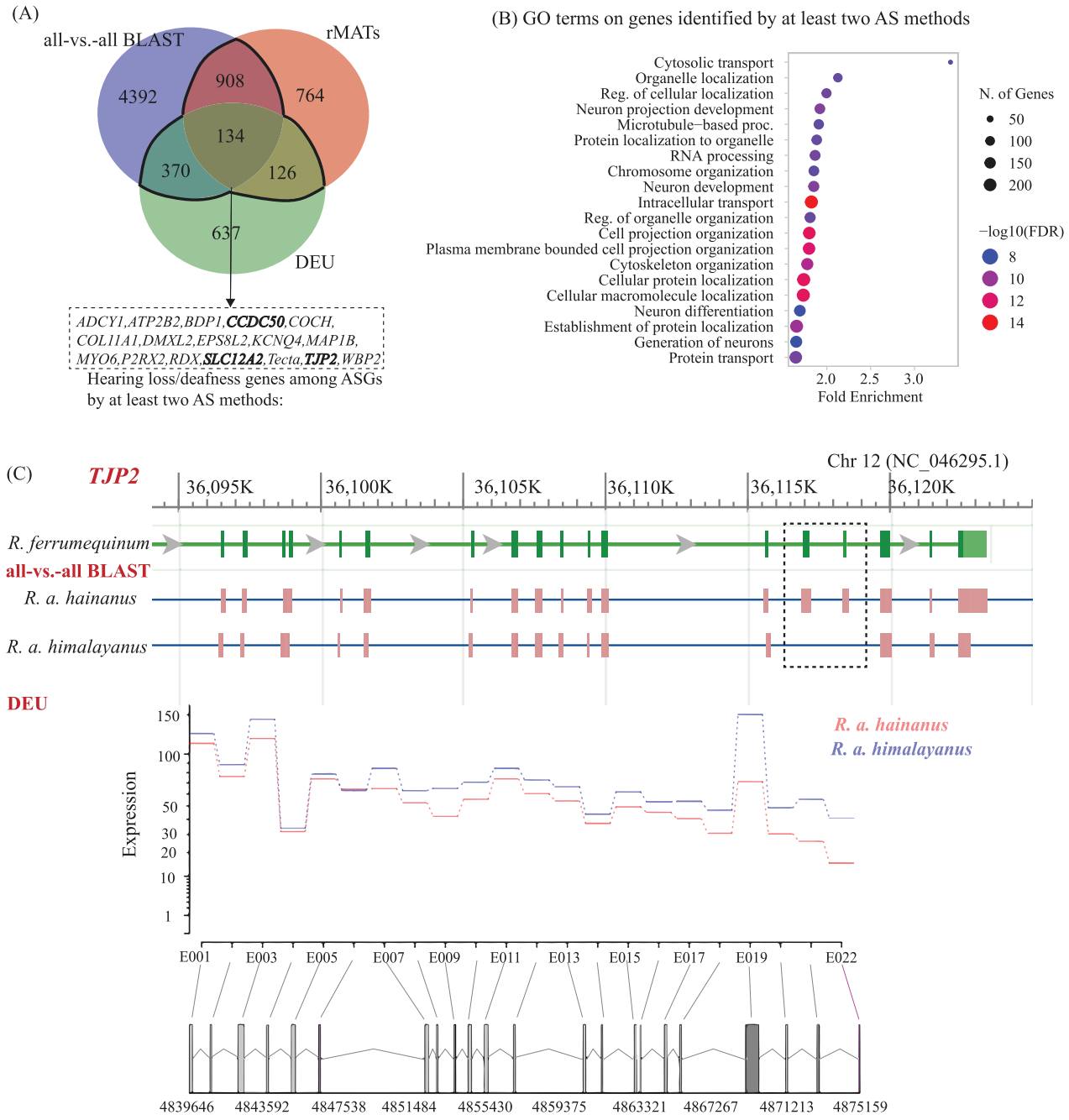

Figure 4. Results of alternative splicing analysis. (A) Venn diagram showing alternatively spliced genes (ASGs) identified using 3 different methods (all-vs.-all BLAST, rMATs, and DEU). Hearing loss/deafness genes identified by all 3 methods were shown in bold. (B) The top 20 significant GO terms enriched on ASGs were identified by at least two methods. (C) Results of all-vs.-all BLAST and DEU for *TJP2* using *R. ferrumequinum* as the reference. The dashed box represents the difference of specific exon between *R. a. hainanus* and *R. a. himalayanus* identified by all-vs.-all BLAST.

transcriptome can be available with long-read RNA-seq, it is better to perform differential expression analysis at the transcript level to identify genes and specific transcripts with significant expression changes.

To explore the role of lncRNAs in gene expression regulation, we also identified differentially expressed (DE) lncRNAs between the 2 subspecies. By examining the target genes of these DE lncRNA, we found consistent patterns with the results of the differential gene expression analysis above. Specifically, we found that target DEGs upregulated in *R. a. himalayanus*, either *cis*- or *trans*-regulated by DE lncRNAs, were mainly involved in cytoskeleton organization, cell proliferation, and programmed cell death (Supplementary Figure S4d,e and Supplementary Table S8). We also found 75 target DEGs by both *cis*- and *trans*-regulation (18 and 57 upregulated in *R. a. hainanus* and *R. a. himalayanus*, respectively). Among them, *ACTB* upregulated in *R. a. himalayanus* encodes β-actin and its deficiency has been reported to cause progressive high-frequency hearing loss in mice (Perrin et al. 2010; Patrinostro et al. 2018). Thus, our current study suggested that lncRNA might be involved in the variation of ultrasonic frequencies between the 2 subspecies by regulating the expression of target protein-coding genes. An important role of lncRNAs in phenotypic variations has also been documented in other

organisms (e.g., in variation of muscle growth and fillet quality traits in rainbow trout *Oncorhynchus mykiss*, Ali et al. 2018; in domestication-related changes in rice *Oryza sativa*, Zheng et al. 2019; in wool bending in Zhongwei goat, Li et al. 2021). In the future functional assays will be needed to further test for the interaction of these candidate lncRNAs and their target genes.

To investigate the role of alternative splicing (AS) in acrostic divergence between *R. a. hainanus* and *R. a. himalayanus*, we identified ASGs using 3 different methods. We found a greater number of ASGs with the long-read RNA-seq compared with other the 2 methods (rMATs and DEU) with short-read RNA-seq (Figure 4A), supporting an advantage of long-read RNA-seq in transcript isoform identification (Stark et al. 2019; Amarasinghe et al. 2020). Consistent with the results of upregulated DEGs in *R. a. himalayanus* above, the ASGs identified between the 2 subspecies were also enriched into GO terms related to cytoskeleton organization and actin filament-based process (Figure 4B and Supplementary Table S10). In addition, we found 17 hearing loss/deafness genes among these ASGs and 5 of them were recognized as echolocation genes in recent studies (*ATP2B2*, *CCDC50*, *KCNQ4*, *TECTA*, and *TJP2*, see Jebb et al. 2020; Yuan et al. 2021). All these results support a potential role of AS in ultrasonic frequency variations of the 2 subspecies. In recent years, the role of AS in generating phenotypic diversity and adaptive evolution has been extensively reviewed (Bush et al. 2017; Mantica and Irimia 2022; Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). Now and in the near future widespread use of long-read RNA-seq will continue to fuel this field by identifying the complexity of alternatively spliced transcripts with FL transcripts sequenced directly.

To test for the relative roles of DE and AS in phenotypic variations, we compared the results of differential gene expression and AS analyses and observed significantly less than expected overlap of DEGs and ASGs (Supplementary Figure S6a). In addition, the only DEGs and only ASGs were found to be enriched into different functional categories (Supplementary Figure S6b,c). This evidence suggested that differential gene expression and AS might act independently and play different roles in phenotypic variations (see also Li et al. 2016; Grantham and Brisson 2018; Jacobs and Elmer 2021). Recent studies on sexual dimorphisms also revealed a complementary role of differential gene expression and AS in encoding sex differences (Rogers et al. 2021; Singh and Agrawal 2023). However, studies with comparative transcriptomics on adaptive evolution draw different conclusions. For example, AS has been suggested to play a more important role in cichlid adaptive radiation compared with differential gene expression (Singh et al. 2017). Another recent study in sunflowers under flooding stress revealed that AS might act by interacting with differential gene expression (i.e., by reinforcing expression differences; Lee et al. 2021). Therefore, the role of differential gene expression and AS in phenotypic variations may be different depending on different studying systems (Verta and Jacobs 2022; Wright et al. 2022).

In conclusion, using 2 recently diverged subspecies with large acoustic variations as the system, we conducted comprehensive transcriptomic analyses with the combination of short-read and long-read RNA-seq data. By performing differential expression analysis at both transcript and gene levels, we identified multiple candidate DEGs that may underlie the variations of ultrasonic frequencies. Most of these genes were nearby and therefore may be regulated by, DE lncRNAs identified between the 2 subspecies. We also identified several candidate ASGs that expressed different isoforms possibly associated with acoustic variations of the 2 subspecies. A future goal is to use quantitative reverse transcription-PCR (qPCR) to validate the candidate genes or lncRNAs identified here. In addition, high-throughput mass spectrometry proteomics can also be applied to further validate candidate DEGs and ASGs at the protein level (Ferrández-Peral et al. 2022).

Acknowledgments

We thank Sun Haijian, Wang Jiaying, and Ding Yuting for assistance with the field data collection and Chen Wenli for her help in the sequencing data collection. We also thank the handling editor and 3 anonymous reviewers whose comments significantly improved the manuscript.

Author Contributions

X.M. conceived the project; Q.L. analyzed the data with the input of J.W.; X.M. wrote the manuscript with the help of Q.L..

Funding

This work was supported by the Scientific and Technological Innovation Plan of the Shanghai Science and Technology Committee (20ZR1417000).

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

All long-read RNA-seq data generated in this study have been deposited to NCBI's Sequence Read Archive database (SRA) under BioProject accession no. PRJNA999095.

Supplementary Material

Supplementary material can be found at https://academic.oup.com/cz.

References

Abolins-Abols M, Kornobis E, Ribeca P, Wakamatsu K, Peterson MP et al., 2018. Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed junco *Junco hyemalis*. *Mol Ecol* **27**:4501–4515.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M, 2018. Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep* **8**:12111.

Alvarez M, Schrey AW, Richards CL, 2015. Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Mol Ecol* **24**:710–725.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME et al., 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**:1–16.

Anders S, Reyes A, Huber W, 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**:2008–2017.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**:289–300.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120.

Bray NL, Pimentel H, Melsted P, Pachter L, 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**:525–527.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO, 2017. Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci* **372**:20150474.

Carroll SB, 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**:25–36.

Chen SF, Jones G, Rossiter SJ, 2009. Determinants of echolocation call frequency variation in the Formosan lesser horseshoe bat *Rhinolophus monoceros*. *Proc Biol Sci* **276**:3901–3909.

Chen K, Rajewsky N, 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**:93–103.

Chen Z, Wiens JJ, 2020. The origins of acoustic communication in vertebrates. *Nat Commun* **11**:369.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A et al., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* **17**:1–19.

Dallos P, 1992. The active cochlea. *J Neurosci* **12**:4575–4585.

Dannemann M, Prüfer K, Kelso J, 2017. Functional implications of Neandertal introgression in modern humans. *Genome Biol* **18**:61.

De Coster W, D'hert S, Schultz DT, Cruts M, Van Broeckhoven C, 2018. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**:2666–2669.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S et al., 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22**:1775–1789.

Desvignes T, Sydes J, Montfort J, Bobe J, Postlethwait JH, 2021. Evolution after whole-genome duplication: Teleost microRNAs. *Mol Biol Evol* **38**:3308–3331.

Dong D, Lei M, Liu Y, Zhang S, 2013. 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* **14**:916–910.

Ferrández-Peral L, Zhan X, Alvarez-Estape M, Chiva C, Esteller-Cucala P et al., 2022. Transcriptome innovations in primates revealed by single-molecule long-read sequencing. *Genome Res* **32**:1448–1462.

Ferreira MS, Alves PC, Callahan CM, Marques JP, Mills LS et al., 2017. The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. *Mol Ecol* **26**:4173–4185.

Fettiplace R, Hackney CM, 2006. The sensory and motor roles of auditory hair cells. *Nat Rev Neurosci* **7**:19–29.

Franchini P, Xiong PW, Fruciano C, Meyer A, 2016. The role of microRNAs in the repeated parallel diversification of lineages of Midas cichlid fish from Nicaragua. *Genome Biol Evol* **8**:1543–1555.

Franchini P, Xiong P, Fruciano C, Schneider RF, Woltering JM et al., 2019. MicroRNA gene regulation in extremely young and parallel adaptive radiations of crater lake cichlid fish. *Mol Biol Evol* **36**:2498–2511.

Fu L, Niu B, Zhu Z, Wu S, Li W, 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**:3150–3152.

Fukunaga T, Hamada M, 2017. RIblast: An ultrafast RNA–RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* **33**:2666–2674.

Ge SX, Jung D, Yao R, 2020. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**:2628–2629.

Gil N, Ulitsky I, 2020. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* **21**:102–117.

Grantham ME, Brisson JA, 2018. Extensive differential splicing underlies phenotypically plastic aphid morphs. *Mol Biol Evol* **35**:1934–1946.

Güttsches AK, Balakrishnan-Renuka A, Kley RA, Tegenthoff M, Brand-Saberi B et al., 2015. ATOH8: A novel marker in human muscle fiber regeneration. *Histochem Cell Biol* **143**:443–452.

Han Y, Hong L, Zhong C, Chen Y, Wang Y et al., 2012. Identification of new altered genes in rat cochleae with noise-induced hearing loss. *Gene* **499**:318–322.

Harrison PW, Wright AE, Mank JE, 2012. The evolution of gene expression and the transcriptome–phenotype relationship. *Semin Cell Dev Biol* **23**:222–229.

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP et al., 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* **11**:1110–1122.

Higashi T, Katsuno T, Kitajiri SI, Furuse M, 2015. Deficiency of angulin-2/ILDR1, a tricellular tight junction-associated membrane protein, causes deafness with cochlear hair cell degeneration in mice. *PLoS One* **10**:e0120674.

Howes TR, Summers BR, Kingsley DM, 2017. Dorsal spine evolution in three spine sticklebacks via a splicing change in *MSX2A*. *BMC Biol* **15**:1–16.

Hultin S, Zheng Y, Mojallal M, Vertuani S, Gentili C et al., 2014. AmotL2 links VE-cadherin to contractile actin fibres necessary for aortic lumen expansion. *Nat Commun* **5**:3743.

Jacobs DS, Catto S, Mutumi GL, Finger N, Webala PW, 2017. Testing the sensory drive hypothesis: Geographic variation in echolocation frequencies of Geoffroy's horseshoe bat (Rhinolophidae: *Rhinolophus clivosus*). *PLoS ONE* **12**:e0187769.

Jacobs A, Elmer KR, 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol Ecol* **30**:4955–4969.

Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K et al., 2020. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**:578–584.

Jones G, 2010. Molecular evolution: Gene convergence in echolocating mammals. *Curr Biol* **20**:R62–R64.

Jones G, Siemers BM, 2011. The communicative potential of bat echolocation pulses. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **197**:447–457.

Kennedy HJ, Crawford AC, Fettiplace R, 2005. Force generation by mammalian hair bundles supports a role in cochlear amplification. *Nature* **433**:880–883.

Keren H, Lev-Maor G, Ast G, 2010. Alternative splicing and evolution: Diversification, exon definition and function. *Nat Rev Genet* **11**:345–355.

Kim D, Langmead B, Salzberg SL, 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* **12**:357–360.

Kittelmann S, McGregor AP, 2019. Modulation and evolution of animal development through microRNA regulation of gene expression. *Genes* **10**:321.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E et al., 2013. Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**:153–165.

Kowalski TJ, Pawelczyk M, Rajkowska E, Dudarewicz A, Sliwinska-Kowalska M, 2014. Genetic variants of *CDH23* associated with noise-induced hearing loss. *Otol Neurotol* **35**:358–365.

Kurabi A, Keithley EM, Housley GD, Ryan AF, Wong ACY, 2017. Cellular mechanisms of noise-induced hearing loss. *Hear Res* **349**:129–137.

Lee JS, Gao L, Guzman LM, Rieseberg LH, 2021. Genome-wide expression and alternative splicing in domesticated sunflowers (*Helianthus annuus* L.) under flooding stress. *Agronomy* **11**:92.

Li H, 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.

Li Q, Chen W, Mao X, 2022. Characterization of microRNA and gene expression in the cochlea of an echolocating bat *Rhinolophus affinis*. *Ecol Evol* **12**:e9025.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al., 2009. Genome project data processing subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Li X, Liu Z, Ye S, Liu Y, Chen Q et al., 2021. Integrated analysis of LncRNA and MRNA reveals novel insights into wool bending in Zhongwei goat. *Animals* **11**:3326.

Li YI, Van De Geijn B, Raj A, Knowles DA, Petti AA et al, 2016. RNA splicing is a primary link between genetic variation and disease. *Science* **352**:600–604.

Li A, Zhang J, Zhou Z, 2014. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinf* **15**:1–10.

Liu Z, Chen P, Li YY, Li MW, Liu Q et al., 2021. Cochlear hair cells of echolocating bats are immune to intense noise. *J Genet Genomics* **48**:984–993.

Liu XX, Mei WB, Soltis PS, Soltis DE, Barbazuk WB, 2017. Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol Ecol Resour* **17**:1243–1256.

Liu Y, Wang J, Wu S, Yang J, 2022. A model for isoform-level differential expression analysis using RNA-seq data without pre-specifying isoform structure. *PLoS ONE* **17**:e0266162.

Love MI, Huber W, Anders S, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.

Luo B, Leiser-Miller L, Santana SE, Zhang L, Liu T et al., 2019. Echolocation call divergence in bats: A comparative analysis. *Behav Ecol Sociobiol* **73**:1–12.

Ma X, He K, Shi Z, Li M, Li F et al., 2021. Large-scale annotation and evolution analysis of MiRNA in insects. *Genome Biol Evol* **13**:evab083.

Mallarino R, Linden TA, Linnen CR, Hoekstra HE, 2017. The role of isoforms in the evolution of cryptic coloration in *Peromyscus* mice. *Mol Ecol* **26**:245–258.

Mank JE, 2017. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**:1–7.

Mantica F, Irimia M, 2022. The 3D-evo space: Evolution of gene expression and alternative splicing regulation. *Annu Rev Genet* **56**:315–337.

Mao XG, He GM, Hua PY, Jones G, Zhang SY et al., 2013. Historical introgression and the persistence of ghost alleles in the intermediate horseshoe bat *Rhinolophus affinis*. *Mol Ecol* **22**:1035–1050.

Mao XG, Rossiter SJ, 2020. Genome-wide data reveal discordant mitonuclear introgression in the intermediate horseshoe bat *Rhinolophus affinis*. *Mol Phylogenet Evol* **150**:106886.

Mao XG, Zhu GJ, Zhang SY, Rossiter SJ, 2010. Pleistocene climatic cycling drives intra-specific diversification in the intermediate horseshoe bat *Rhinolophus affinis* in Southern China. *Mol Ecol* **19**:2754–2769.

Mao XG, Zhu GJ, Zhang LB, Zhang SY, Rossiter SJ, 2014. Differential introgression among loci across a hybrid zone of the intermediate horseshoe bat *Rhinolophus affinis*. *BMC Evol Biol* **62**:405–412.

Martin M, 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**:10–12.

Mason NA, Taylor SA, 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Mol Ecol* **24**:3009–3025.

McCoy RC, Wakefield J, Akey JM, 2017. Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression. *Cell* **168**:916–927.e12.

Morozko EL, Nishio A, Ingham NJ, Chandra R, Fitzgerald T et al., 2015. ILDR1 null mice, a model of human deafness DFNB42, show structural aberrations of tricellular tight junctions and degeneration of auditory hair cells. *Hum Mol Genet* **24**:609–624.

Mutumi GL, Jacobs DS, Winker H, 2016. Sensory drive mediated by climatic gradients partially explains divergence in acoustic signals in two horseshoe bat species, *Rhinolophus swinnyi* and *Rhinolophus simulator*. *PLoS ONE* **11**:e0148053.

Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN et al., 2019. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**:eaaw7317.

Nigenda-Morales SF, Hu Y, Beasley JC, Ruiz-Piña HA, Valenzuela-Galván D et al., 2018. Transcriptomic analysis of skin pigmentation variation in the Virginia opossum *Didelphis virginiana*. *Mol Ecol* **27**:2680–2697.

Odendaal LJ, Jacobs DS, Bishop JM, 2014. Sensory trait variation in an echolocating bat suggests roles for both selection and plasticity. *BMC Evol Biol* **14**:1–19.

Patrinostro X, Roy P, Lindsay A, Chamberlain CM, Sundby LJ et al., 2018. Essential nucleotide-and protein-dependent functions of *Actb*/β-actin. *Proc Natl Acad Sci USA* **115**:7973–7978.

Perrin BJ, Sonnemann KJ, Ervasti JM, 2010. β-actin and γ-actin are each dispensable for auditory hair cell development but required for Stereocilia maintenance. *PLoS Genet* **6**:e1001158.

Price J, Harrison MC, Hammond RL, Adams S, Gutierrez-Marcos JF et al., 2018. Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*. *Mol Ecol* **27**:1036–1043.

Robinson MD, McCarthy DJ, Smyth GK, 2010. EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140.

Rogers TF, Palmer DH, Wright AE, 2021. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol Biol Evol* **38**:519–530.

Romero IG, Ruvinsky I, Gilad Y, 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**:505–516.

Sahlin K, Medvedev P, 2020. De novo clustering of long-read transcriptome data using a greedy, quality value-based algorithm. *J Comput Biol* **27**:472–484.

Sahlin K, Medvedev P, 2021. Error correction enables use of Oxford nanopore technology for reference-free transcriptome analysis. *Nat Commun* **12**:1–13.

Sang Q, Li W, Xu Y, Qu R, Xu Z et al., 2015. ILDR1 deficiency causes degeneration of cochlear outer hair cells and disrupts the structure of the organ of Corti: A mouse model for human DFNB42. *Biol Open* **4**:411–418.

Schnitzler HU, Moss CF, Denzinger A, 2003. From spatial orientation to food acquisition in echolocating bats. *Trends Ecol Evol* **18**:386–394.

Schuchmann M, Siemers BM, 2010. Variability in echolocation call intensity in a community of horseshoe bats: A role for resource partitioning or communication. *PLoS One* **5**:e12842.

Shen S, Park JW, Lu ZX, Lin L, Henry MD et al., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc Natl Acad Sci USA* **111**:E5593–E5601.

Siemers BM, Beedholm K, Dietz C, Dietz I, Ivanova T, 2005. Is species identity, sex, age or individual quality conveyed by echolocation call frequency in European horseshoe bats? *Acta Chiropterologica* **7**:259–274.

Singh A, Agrawal AF, 2023. Two forms of sexual dimorphism in gene expression in *Drosophila melanogaster*: Their coincidence and evolutionary genetics. *Mol Biol Evol* **40**:msad091.

Singh P, Ahi EP, 2022. The importance of alternative splicing in adaptive evolution. *Mol Ecol* **31**:1928–1938.

Singh P, Börger C, More H, Sturmbauer C, 2017. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol Evol* **9**:2764–2781.

Stark R, Grzelak M, Hadfield J, 2019. RNA sequencing: The teenage years. *Nat Rev Genet* **20**:631–656.

Sun HJ, Chen WL, Wang JY, Zhang LB, Rossiter SJ et al., 2020. Echolocation call frequency variation in horseshoe bats: Molecular basis revealed by comparative transcriptomics. *Proc R Soc B* **287**:20200875.

Sun K, Luo L, Kimball RT, Wei X, Jin L et al., 2013. Geographic variation in the acoustic traits of greater horseshoe bats: Testing the importance of drift and ecological selection in evolutionary processes. *PLoS One* **8**:e70368.

Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E et al., 2020. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* **11**:1–12.

Todd EV, Black MA, Gemmell NJ, 2016. The power and promise of RNA-seq in ecology and evolution. *Mol Ecol* **25**:1224–1241.

Verta JP, Jacobs A, 2022. The role of alternative splicing in adaptation and evolution. *Trends Ecol Evol* **37**:299–308.

Wagner EL, Shin JB, 2019. Mechanisms of hair cell damage and repair. *Trends Neurosci* **42**:414–424.

Wang B, Kumar V, Olson A, Ware D, 2019. Reviving the transcriptome studies: An insight into the emergence of single-molecule transcriptome sequencing. *Front Genet* **10**:384.

Wang H, Zhao H, Huang X, Sun K, Feng J, 2018. Comparative cochlear transcriptomics of echolocating bats provides new insights into different nervous activities of CF bat species. *Sci Rep* **8**:15934.

Workman RE, Tang AD, Tang PS, Jain M, Tyson JR et al., 2019. Nanopore native RNA sequencing of a human poly (A) transcriptome. *Nat Methods* **16**:1297–1305.

Wright CJ, Smith CW, Jiggins CD, 2022. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* **23**:697–710.

Xie L, Sun K, Jiang T, Liu S, Lu G et al., 2017. The effects of cultural drift on geographic variation in echolocation calls of the Chinese rufous horseshoe bat *Rhinolophus sinicus*. *Ethology* **123**:532–541.

Xing Y, Lee C, 2006. Alternative splicing and RNA selection pressure: Evolutionary consequences for eukaryotic genomes. *Nat Rev Genet* **7**:499–509.

Yuan Y, Zhang Y, Zhang P, Liu C, Wang J et al., 2021. Comparative genomics provides insights into the aquatic adaptations of mammals. *Proc Natl Acad Sci USA* **118**:e2106080118.

Zhang L, Jones G, Zhang J, Zhu G, Parsons S et al., 2009. Recent surveys of bats (Mammalia: Chiroptera) from China. I. Rhinolophidae and Hipposideridae. *Acta Chiropterologica* **11**:71–88.

Zhao X, Bergland AO, Behrman EL, Gregory BD, Petrov DA et al., 2016. Global transcriptional profiling of diapause and climatic adaptation in *Drosophila melanogaster*. *Mol Biol Evol* **33**:707–720.

Zhao HB, Wang H, Liu T, Liu S, Jin LR et al., 2019. Gene expression vs. sequence divergence: Comparative transcriptome sequencing among natural *Rhinolophus ferrumequinum* populations with different acoustic phenotypes. *Front Zool* **16**:37.

Zheng XM, Chen J, Pang HB, Liu S, Gao Q et al., 2019. Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. *Sci Adv* **5**:eaax3619.

Zhu C, Wu J, Sun H, Briganti F, Meder B et al., 2021. Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. *Nat Commun* **12**: 4203.