Comparative cochlear transcriptomics of echolocating bats provides new insights into different nervous activities of CF bat species

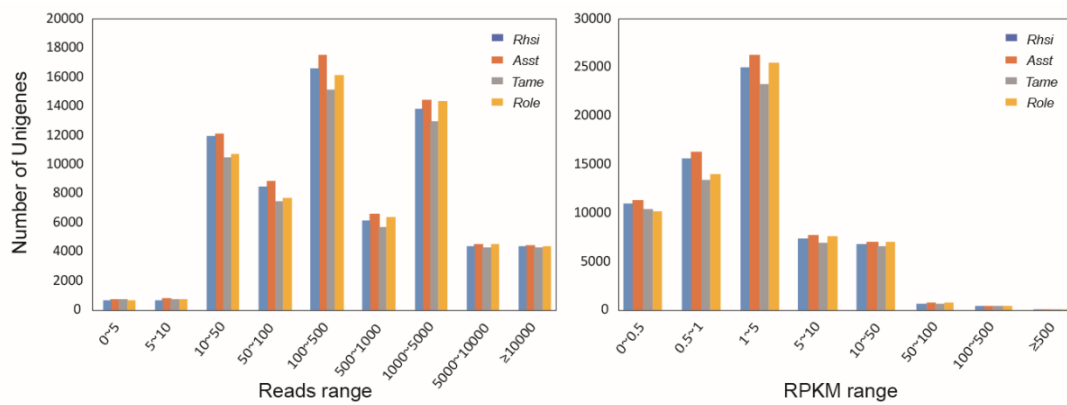**Hui Wang, Hanbo Zhao, Xiaobin Huang, Keping Sun & Jiang Feng**

Figure S1. Interval distribution of unigene expression abundance vs. read range (left) and RPKM range (right).
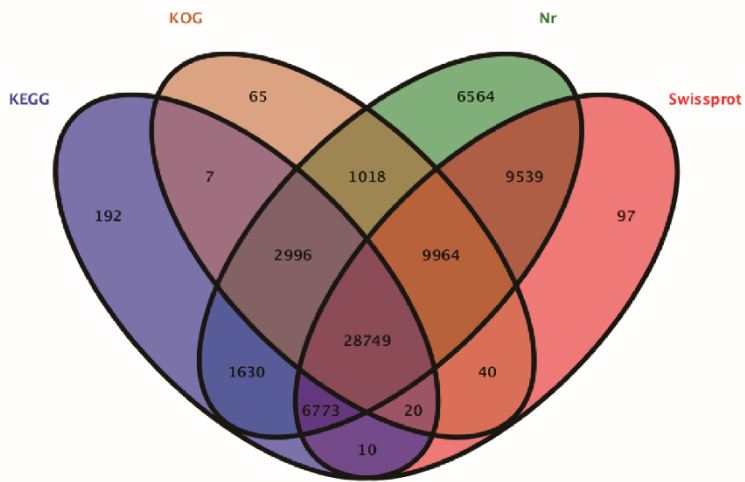
Figure S2. Venn diagram of Nr, Swiss-Prot, KOG and KEGG databases, showing homology

sequence numbers of the unigenes in the four different databases.

KOG Function Classification of all assembled unigenes

A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
G: Carbohydrate transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
O: Posttranslational modification, protein turnover, chaperones
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
T: Signal transduction mechanisms
U: Intracellular trafficking, secretion, and vesicular transport
V: Defense mechanisms
W: Extracellular structures
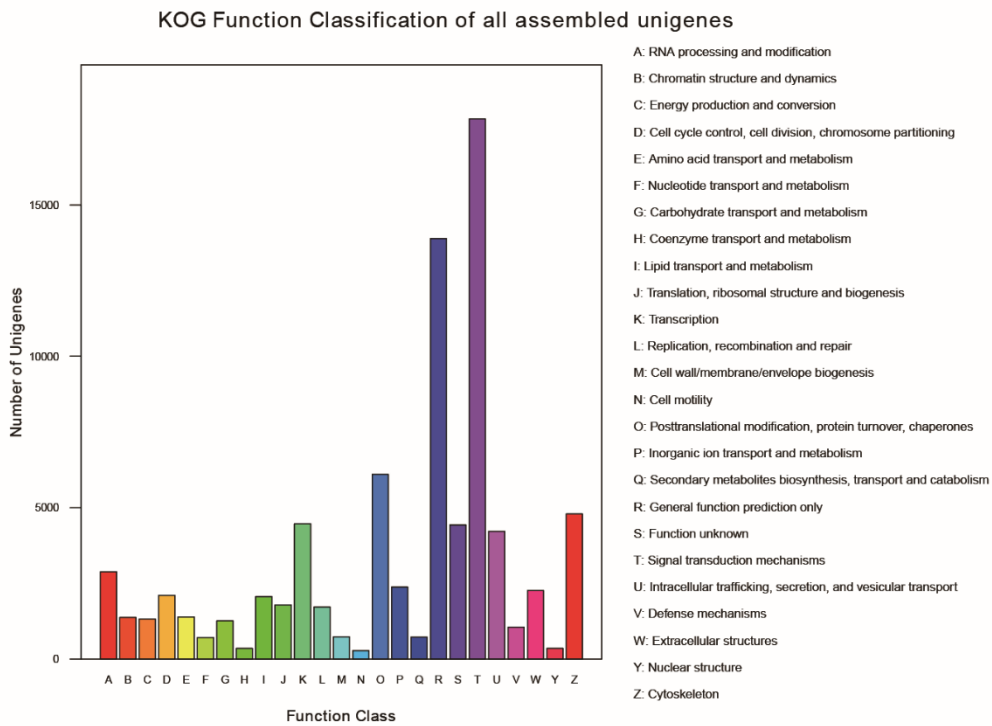Y: Nuclear structure
Z: Cytoskeleton

Figure S3. KOG functional classification of the unigenes. A total of 25 categories were obtained.

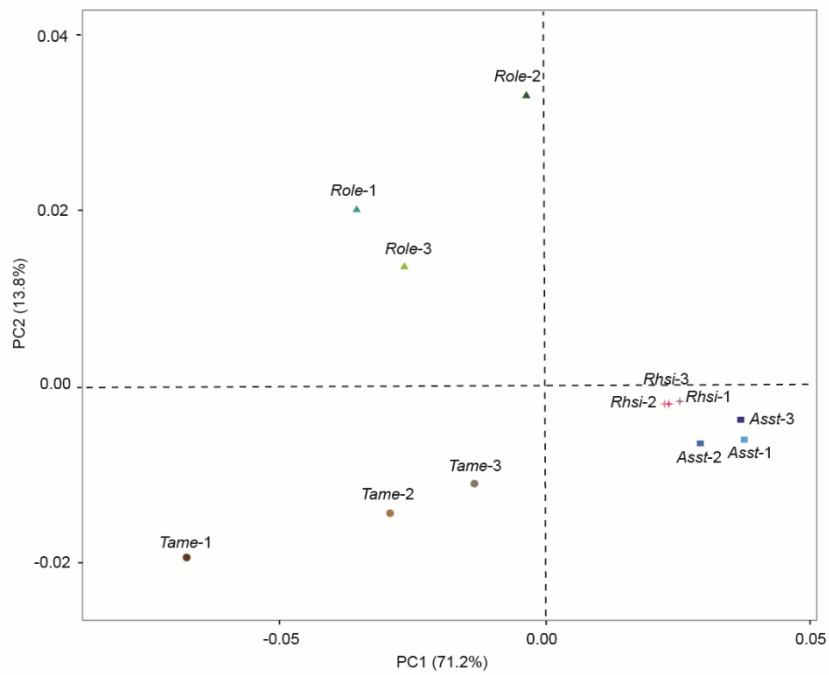The x-axis indicates the categories and the y-axis indicates the number of unigenes.

Figure S4. Principal component analysis (PCA) of the transcriptome of four bat species. The numbers in parentheses represent the proportion of variance explained by that principal component. PC1 and PC2 represent the top two dimensions of the genes showing differential expression among these samples, which account for 71.2% and 13.8% of the expressed genes, respectively.
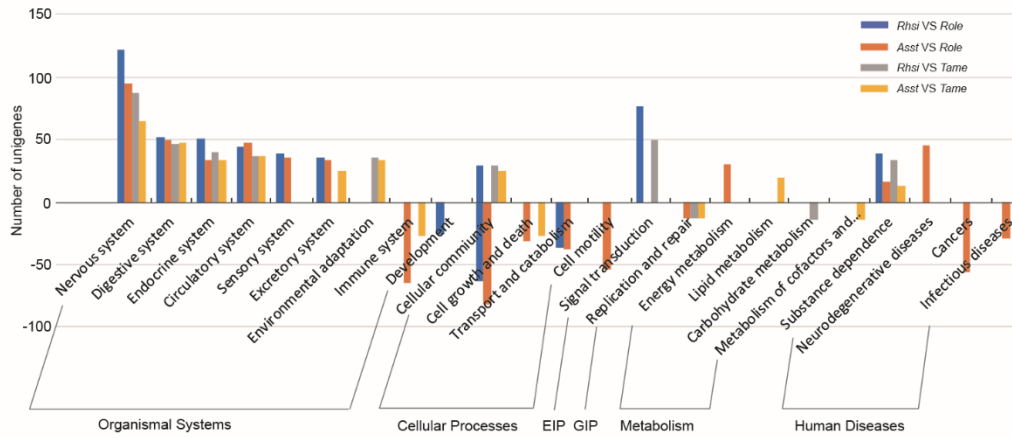
Figure S5. KEGG pathways significantly (*p* < 0.01) enriched for DEGs from four comparisons:

Rhsi vs. *Role* (blue), *Asst* vs. *Role* (orange), *Rhsi* vs. *Tame* (grey) and *Asst* vs. *Tame* (yellow). EIP

stands for environment information process; GIP stands for genetic information process.
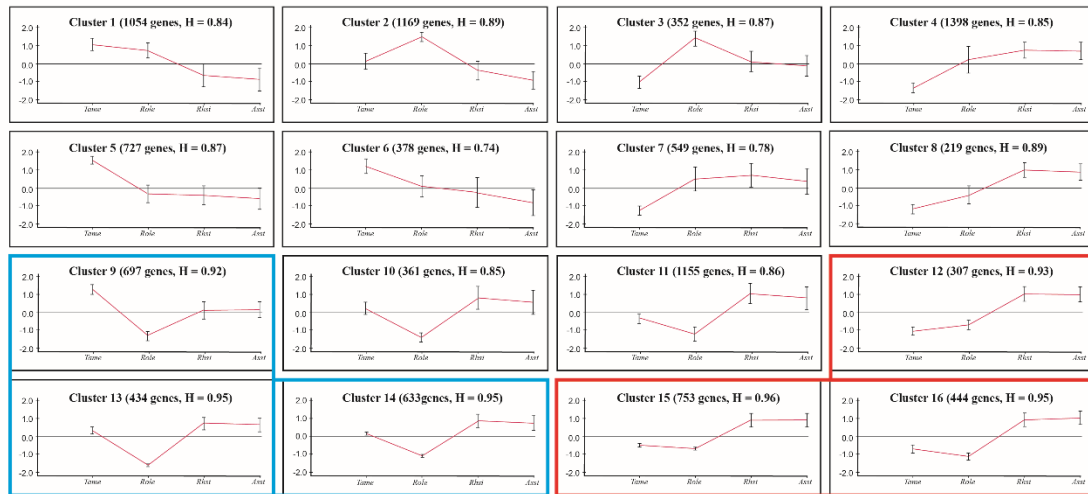
Figure S6. Sixteen clusters of 10,630 DEGs based on their gene expression patterns in the order of their dominant frequency from low to high: *Tame*, *Role*, *Asst* and *Rhsi*. In red super-group 1 (clusters 12, 15 and 16) and blue super-group 2 (clusters 9, 13 and 14), colours represent gene clusters displaying similar expression with homogeneity (H) values larger than 0.9. The number of genes in each cluster and the H value are shown on the top of the curve. Each cluster graph displays the mean pattern of expression (red lines) of the genes in it. The x-axis represents species in order of their dominant frequency from low to high and the y-axis represents $\log_2$ fold change in gene expression.