



OPEN

Comparative transcriptomes of nine tissues for the Heilongjiang brown frog (*Rana amurensis*)

Wanyu Li^{1,2}, Yue Lan¹, Lei Wang^{2,3}, Lewei He¹, Ruixiang Tang¹, Megan Price¹, Bisong Yue^{1,2,3}✉ & Zhenxin Fan^{1,2}✉

The Heilongjiang brown frog (*Rana amurensis*) is widely used in traditional Chinese medicine. In particular, the oviduct and skin have been developed into various health products. However, limited numbers of complete genomes of amphibian species have been reported, excluding the Heilongjiang brown frog. Here, the transcriptomes of 45 samples from the liver, spleen, heart, ovaries, thigh muscles, skin, oviduct, stomach and intestine of five Heilongjiang brown frog were reassembled and analyzed. A total of 1,085,532 unigenes with an average length of 676.6 bp and N50 of 722 bp were obtained. Comparative transcriptomics of different tissues detected tissue-specific expression. There were 3248 differentially expressed genes (DEGs) in the ovary, and the number of unique DEGs between the ovary and spleen was the largest. The results of DEGs enrichment showed there were many pathways and items related to protein synthesis and metabolism in the oviduct. The DEGs of the skin were enriched with many bacterial defense items, indicating that there were a large number of antimicrobial peptides in the skin. Thus, these were suitable as biological sources for the development and extraction of antimicrobial peptides. Through the assembly of transcriptome sequencing data and functional annotation of the Heilongjiang brown frog genome, this study provides reference materials for further exploring and utilizing functional gene resources of frogs and lays a foundation for medical research and the development of new products.

Amphibian research has received more attention in the last few decades, and thus species descriptions have increase with roughly three new amphibian species being described per week¹. As of 2020, approximately 8000 amphibian species have been described, but because their genomes are large and complex there are only ~ 13 published amphibian genomes and most assemblies are highly fragmented². Only one genome is available from Ranidae (true frogs), the North American bullfrog (*Rana [Lithobates] catesbeiana*), which is surprising as it is the largest frog family with species found on every continent except Antarctica³. Sequenced frog genomes may be limited by the high cost of large-genome sequencing, yet transcriptome analysis is economically viable for studying the large and complex genomes of amphibians⁴. Indeed, transcriptome sequencing technology has been used in many amphibian species⁵⁻⁷, and has provided novel insights into their gene expression profiles and the evolution of their immune systems.

The Heilongjiang brown frog (*Rana amurensis*) belongs to Ranidae (Anura Amphibia) and is mainly distributed in Northeastern China. The frog is widely used in traditional Chinese medicine and is considered an invaluable economic animal with high medical value^{8,9}. Its products have anti-fatigue, anti-lipid peroxidation, antitussive and expectorant effects on immune function and stress performance, and have been widely accepted by Chinese, and its comprehensive development prospect is broad as a traditional Chinese medicine¹⁰. Recent studies have focused on the frog's oviduct, skin, kidney, muscle and other tissues, but mainly regarding nutritive material component analysis in these tissues¹¹⁻¹³, scarcely any study about the genome sequence. Without its full genome sequence, the mechanism of action, gene expression mechanism and molecular regulation mechanism remain largely unknown. Here, we performed large-scale transcriptome sequencing of nine tissues (liver, spleen, heart, ovary, thigh muscles, skin, oviduct, stomach and intestine) from five Heilongjiang brown frogs using Illumina NovaSeq 6000 platform. We aimed to (1) perform de novo assembly to provide high quality transcriptome

¹Key Laboratory of Bioresources and Eco-Environment (Ministry of Education), College of Life Sciences, Sichuan University, Chengdu 610064, Sichuan, China. ²Sichuan Key Laboratory of Conservation Biology On Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu 610064, Sichuan, China. ³Sichuan Engineering Research Center for Medicinal Animals, Xichang 615000, Sichuan, China. ✉email: bsyue@scu.edu.cn; zxfan@scu.edu.cn

resources; (2) detect the gene expression and identify the differentially expressed transcripts between tissues; (3) investigate the pathways and functional terms related to tissue-specific characters. Our results exhibited tissue-specific gene expressions and provided a reference gene set with high quality, which would serve as a valuable resource for further genetic, genomic and medical research of the Heilongjiang brown frog and other frogs.

Materials and methods

Experimental design and sample collection. Five non-breeding adult wild female Heilongjiang brown frog were obtained from the foot of the mountain in Jiaohe, Jilin City, Jilin Province, China. The frogs were placed in a airtight container with CO₂ for anesthesia and euthanasia. The liver, spleen, heart, ovaries, thigh muscles, skin, oviduct, stomach and intestine were dissected after anesthetized and euthanasia. The 45 samples from nine tissues were collected in cryotubes and immediately stored in liquid nitrogen for later RNA extraction. All animal experiments in this study were approved by the Ethics Committee of the College of Life Sciences, Sichuan University, number 20210309009. All experiments were performed in accordance with relevant guidelines and regulations. Our reporting of research involving animals follows the recommendations of the ARRIVE guidelines.

Library preparation and RNA sequencing. Total RNA was isolated separately from the 45 samples using a TRIzol Kit (Invitrogen, CA, USA) according to the manufacturer's instructions. The RNA integrity was checked using an Agilent 2100 BioAnalyzer (Agilent Technologies, CA, USA), and the mRNA was purified from the total RNA using the PolyA Tract mRNA Isolation Systems Kit (Promega) following the manufacturer's instructions. Fragmentation was undertaken using divalent cations under elevated temperature in NEB Next First Strand Synthesis Reaction Buffer. First strand cDNA was synthesized using a random hexamer primer and M-MuLV reverse transcriptase (RNase H-). The second strand cDNA synthesis was then performed using DNA polymerase I and RNase H. In order to select cDNA fragments that were preferentially 150–200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). The adaptor modified fragments were selected by gel purification and amplified through PCR to create the final cDNA library. Transcriptome sequencing was conducted using Illumina NovaSeq 6000 according to the manufacturer's instructions.

Sequence preprocessing and transcripts assembly. The 150-bp paired-end (PE) short reads were obtained after sequencing. We used NGS QC Toolkit v2.3.3¹⁴ with stringent criteria (reads with more than 90% bases within ≥ 20 base quality (Q-value)) to remove the low-quality paired-end reads or reads containing adaptors resulting in the clean reads. Samples of different tissues were combined, and de novo assembled with Trinity separately using the following parameters: -seqType fq -JM 200G -CPU 24 -min_contig_length 300, while the remaining were analyzed using default parameters¹⁵. The assembled transcripts of the nine tissues were clustered with CD-HIT-EST¹⁶, obtaining the final assembled unigenes. To evaluate the completeness, we employed Benchmarking Universal Single-Copy Orthologs (BUSCO; <http://busco.ezlab.org/>) to evaluate the gene set of the Heilongjiang brown frog using vertebrata data¹⁷. The protein coding regions prediction of assembled unigenes was carried out using TransDecoder-v5.5.0 and the longest open reading frame (ORF) predicted for each contig sequence with a minimum of 100 amino acids long¹⁵.

Gene annotation. Unigenes from the de novo assembly were aligned to the NCBI (<http://blast.ncbi.nlm.nih.gov/>) non-redundant protein sequence (Nr) database¹⁸, Swiss-Prot (<http://www.uniprot.org/>)¹⁹ and Clusters of Orthologous Groups (COG; <http://www.ncbi.nlm.nih.gov/COG/>)²⁰ by BLASTX with E-value cutoff of 1×10^{-5} . Gene names were assigned to each protein sequence based on the best hit from all the alignment results. The resultant unigenes were searched against the Gene Ontology (GO; <http://www.geneontology.org/>) database²¹. To generate the Web Gene Ontology Annotation Plot (WEGO) software²² chart, the results from Blast2GO annotation were exported in WEGO native format. The WEGO chart was then generated by uploading to <http://wego.genomics.org.cn/>. Kyoto Encyclopaedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/pathway.html>) databases²³ annotation was performed using the single-directional best-hit (SBH) method in KEGG automatic annotation server (KAAS). KEGG Mapper* was used to reconstruct pathways.

Differentially expressed gene and gene enrichment analysis. Processed reads from each sample were mapped to the assembled unigenes using HISAT2²⁴. To obtain the expression raw read counts for each gene, expression abundance of genes was generated using Salmon²⁵ with the SAM files created by HISAT2. In order to remove the calculation error, we removed the low expression genes from the final results, which met all the following conditions. (1) The number of samples with non-zero expression is less than 1/3 of the total number of samples. (2) The total expression raw read counts of the gene in all samples were less than 10. Based on the expression abundance of removing low expression genes, principal component analysis (PCA) from the plotPCA of DESeq2 R package²⁶ was used to visualize the correlation of 45 RNA-seq samples. The DESeq2 R package was used to identify the different expression genes (DEGs) between the 9 different tissues, with raw read counts as the input. P values were corrected for multiple testing with the Benjamini–Hochberg false discovery rate (FDR ≤ 0.05) and an absolute value of log₂ fold change ≥ 2 was used to determine the significant differences in gene expression. GO and KEGG enrichment were performed by R package clusterProfiler²⁷ with GO and KEGG annotation results of all genes of the Heilongjiang brown frog as the background gene set.

Construction of co-expression network. Gene expression data after normalization by R of all 45 samples were used to construct multi-tissue co-expression networks that simultaneously captured intra- and inter-tissue gene–gene interactions. We ranked all genes according to MAD (median absolute deviation) value and selected the first 10,000 as representative genes²⁸. These genes were clustered according to their dissimilarity and were divided into different modules. Modules with high similarity were merged and a topological overlap matrix (TOM) was constructed based on the merged modules. For each tissue, we calculated the connectivity of all genes in the modules with significant module–trait relationship (>0.8) to identify hub genes. Cytoscape²⁹ was then used to visualize the weighted correlation network for hub genes and find the key regulator (driver) genes. Meanwhile, GO and KEGG enrichment analysis of these hub genes were performed using the same method.

Results

Sequencing and de novo assembly. A total of 1,182,707,807 150 bp PE raw reads were generated. The sequencing depth is 6×. After quality-trimming and adaptor-clipping, 1,160,223,961 clean reads remained. The average Q20 and GC contents for the samples were 97.46% and 46.68% (respectively; Table S1). The assembled transcripts of the nine tissues were combined and clustered, and a total of 1,085,532 unigenes with an average length of 676.56 bp (N50 = 722 bp) and a total length of 734,428,057 nt were generated after splicing and removing redundancy (Table S2). BUSCO and homologous assessment showed that nearly 28.7% of total complete and single-copy BUSCOs were identified in this gene set and the fragmented BUSCOs and missing BUSCOs was nearly zero. However, the ratio of complete and duplicated BUSCOs was 68.5% (Fig. S1). The complete BUSCO is the sum of complete and single-copy BUSCO and complete and duplicated BUSCO (97.2%). A total of 1,239,626 ORFs were identified from 363,378 of the 1,085,532 assembled unigenes by TransDecoder, and 149,093 of the unigenes contained only one ORF (Fig. 1A, B).

Functional annotation of unigenes. There were 357,673 unigenes obtained from the annotated samples based on sequence similarity. Among these, 310,369 (28.59%), 194,620 (17.93%), 115,785 (10.67%), 113,311 (10.44%) and 232,297 (21.34%) unigenes were aligned to NR, Swiss-Prot, KEGG, COGs and GO databases, respectively (Fig. 1C). The species with the highest similarity were obtained according to the NR annotation results (Fig. 1D), of which the top three species in Anura were: bullfrog (*Rana [Lithobates] catesbeiana*), tropical clawed frog (*Xenopus tropicalis*), African clawed frog (*Xenopus laevis*). Notably, the number of unigenes aligned to the bullfrog was greater than the sum of the other species.

The GO annotation showed that the major subcategories were “Cell (GO:0005623)”, “Cell part (GO:0044464)”, “binding (GO:0005488)”, “catalytic activity (GO:0003824)”, “cellular process (GO:0009987)” and “metabolic process (GO:0008152)” (Fig. S2). COG assignments were performed to predict and classify the possible gene functions (Fig. S3). Analysis of the COG annotations showed that there were 113,311 (10.44%) hits assigned by 1,085,532 unigenes. The hits from the COG prediction were classified into 26 functional categories, where the most enriched terms were general function (20,573 unigenes), followed by amino acid transport and metabolism (8,851 unigenes) and translation, ribosomal structure and biogenesis (8,829 unigenes).

Identification of DEGs among tissues. RNA-seq data were first normalized to reduce the influence of technical noise, and then we performed a PCA on gene expression from the 45 samples. We found that the nine tissues clearly clustered by the first Principal Component (PC1) and PC2, which explained 48% of the variance (Fig. S4). To identify genes that displayed significantly different expression between different tissues, DEGs were analyzed with pairwise comparison of the nine tissues. A total of 82,903 significant DEGs were detected by DESeq2 (Table 1). DEGs with higher expression levels in one sample when compared to another were denoted as “upregulated”, while those with lower expression levels were termed as “downregulated”. In general, the number of DEGs (29,414) between ovary and spleen was the largest, while the number of DEGs (3917) between intestine and stomach was the smallest. The intersection of DEGs between one and the other eight tissues was infested as a unique DEG (UEG) of this tissue. We found that the stomach had the fewest UEGs (209) and the ovaries had the highest number (3248). We also detected 779 UEGs in the skin and 752 in the oviduct. Meanwhile these two tissues are key to the medicinal and edible value of Heilongjiang brown frog, and also illustrate the special characteristics of the skin and oviduct of Heilongjiang brown frog compared with other tissues at the molecular level, which will provide a foundation for further in-depth mining and development of Heilongjiang brown frog economic value and research on its medicinal mechanism.

GO category and KEGG pathway enrichment analyses. To gain insight into the biological roles of UEGs, we performed GO (Gene ontology) category and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses.

The oviduct is one of the tissues we are most interested in. For the 752 UEGs of the oviduct, the upregulated genes of the oviduct were mainly enriched in regulation of endopeptidase activity (GO:0052548), microtubule polymerization (GO:0046785) and O-glycan processing (GO:0016266) (Table S18). The downregulated genes of the oviduct were mainly enriched in activation of adenylate cyclase activity (GO:0007190) and response to jasmonic acid (GO:0009753) (Table S19). Mucin type O-glycan biosynthesis (ko00512), glycosphingolipid biosynthesis—lacto and neolacto series (ko00601) and other glycan degradation (ko00511) were significantly enriched in the oviduct’s upregulated genes in KEGG enrichment (Table S18).

Likewise, we performed GO and KEGG enrichment with the 779 UEGs of the skin. The skin’s upregulated genes were mainly enriched in epidermis development (GO:0008544), antibacterial humoral response (GO:0019731) and defense response to gram-negative bacterium (GO:0050829) (Table S12). The skin’s downregulated genes were enriched in digestion (GO:0007586), taurine transport (GO:0015734) and modulation

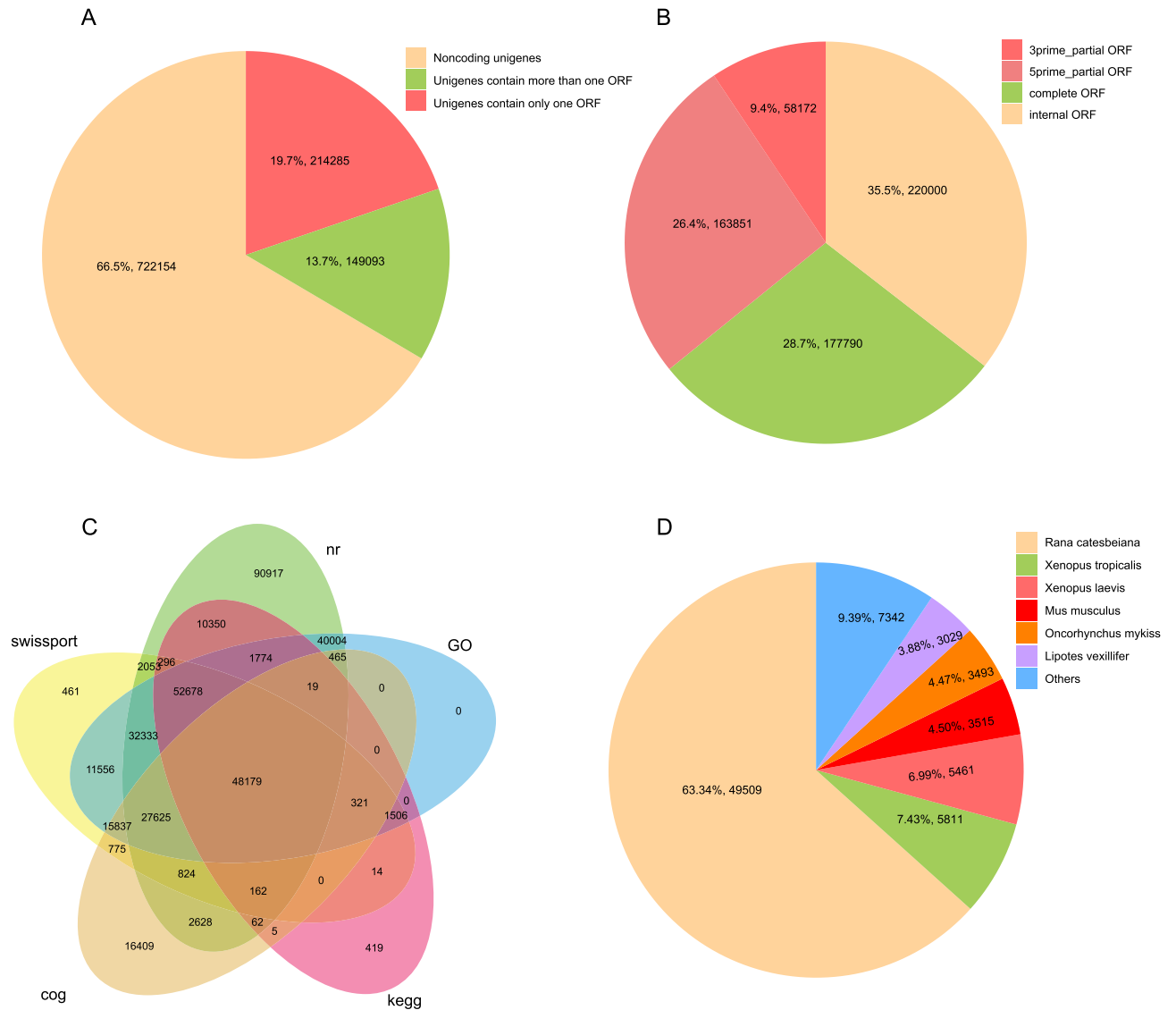


Figure 1. Coding gene prediction and annotation of unigenes. **(A)** Piechart showing the identification results of assembled unigenes coding ability by TransDecoder. **(B)** Piechart showing the classification of all identified open reading frames (ORFs). **(C)** Venn-diagram showing the overlap between results of unigenes aligned to NR, Swiss-Prot, KEGG, COGs and GO databases. **(D)** Species annotation results from unigenes aligned to NR database.

of chemical synaptic transmission (GO:0050804) (Table S13). Arachidonic acid metabolism (ko00590), drug metabolism—cytochrome P450 (ko00982) and isoquinoline alkaloid biosynthesis (ko00950) were significantly enriched in the skin's upregulated genes highlighted in the KEGG enrichment analysis (Table S12).

Cardiac muscle contraction (GO:0060048), cardiac muscle tissue morphogenesis (GO:0055008) and regulation of heart rate (GO:0002027) were mainly enriched in the heart's upregulated genes, found via GO enrichment analysis (Table S24). The UEGs of other tissues were also analyzed separately for GO and KEGG enrichment, refer to Tables S3–S29.

Construction of gene co-expression modules and weighted correlation network analysis. We applied weighted gene co-expression network analysis (WGCNA) to gain further insight into the molecular mechanisms of each tissue²⁸. After merging modules with high similarity, 10,000 selected genes were clustered into 29 co-expressed modules and the modules were denoted using different colors in Fig. S5. Each tissue showed significant module–trait relationship among the different traits (Fig. 2A). The relationships between module membership were shown in Fig. 2B, which also highlighted the connectivity of the most significant modules were correlated with tissues. For example, the turquoise module was strongly correlated with ovary tissue, which had the largest number of genes. DNA replication, meiosis—yeast and p53 signaling pathway were significantly enriched by the highly expressed genes (hub genes) in this module. The dark turquoise module was strongly cor-

	Intestine		Muscle		Liver		Ovary		Skin		Spleen		Oviduct		Stomach		Heart	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
Intestine	-	-	7856	12,014	4617	7867	7371	17,150	4963	6708	6274	9747	6396	9298	1750	2167	6383	9694
Muscle	12,014	7856	-	-	9600	8697	10,815	17,819	10,228	7852	10,299	8732	11,175	8716	9836	5930	4435	3670
Liver	7867	4617	8697	9600	-	-	9307	15,761	7032	5492	5295	6679	7842	7330	7316	4422	6576	6335
Ovary	17,150	7371	17,819	10,815	15,761	9307	-	-	17,573	7648	20,309	9105	17,456	8316	17,780	7436	18,581	8984
Skin	6708	4963	7852	10,228	5492	7032	7648	17,573	-	-	4331	6662	5634	6641	4690	3209	4836	6231
Spleen	9747	6274	8732	10,299	6679	5295	9105	20,309	6662	4331	-	-	9894	10,011	8606	4763	4429	3657
Oviduct	9298	6396	8716	11,175	7330	7842	8316	17,456	6641	5634	10,011	9894	-	-	6526	3148	8022	8415
Stomach	2167	1750	5930	9836	4422	7316	7436	17,780	3209	4690	4763	8606	3148	6526	-	-	4115	7414
Heart	9694	6383	3670	4435	6335	6576	8984	18,581	6231	4836	3657	4429	8415	8022	7414	4115	-	-
UEGs	537	21	1056	208	916	217	2003	1245	745	34	694	63	615	137	197	12	346	7
Hub-genes	37		154		125		273		21		64		86		19		49	

Table 1. The numbers of differential expressed genes between samples.

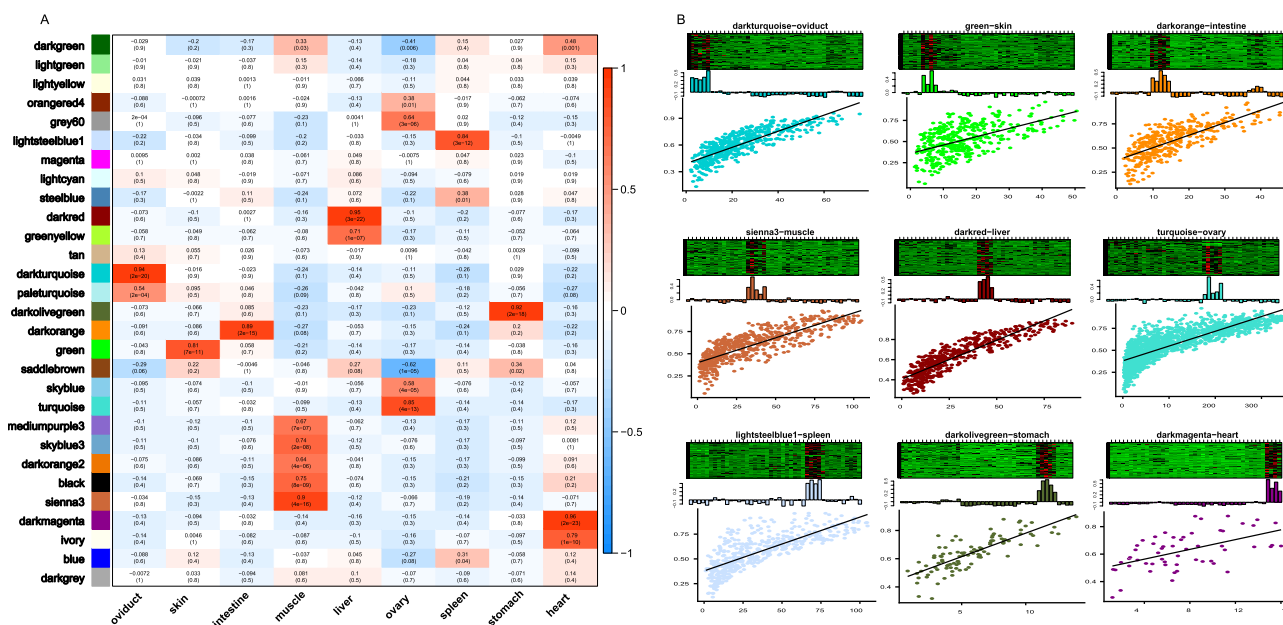


Figure 2. Analysis of gene co-expression modules. (A) Heatmap illuminates the association between modules and tissues. Each row corresponds to a module. Each column corresponds to a specific tissue. The color of each cell at the row-column intersection indicates the correlation coefficient between the module and the tissue. Red means significant correlation and blue means no significant correlation. (B) The internal situation of each module significantly related to the tissue (module – trait relationship > 0.8). The heatmap and histogram show the expression of all the genes belong to this module in each sample. The point graph visually shows the correlation between the module and tissue. (WGCNA_1.69: https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/WGCNA_1.69-81.tar.gz).

related with oviduct tissue. Mucin type O-glycan biosynthesis and other glycan degradation were significantly enriched by the highly expressed genes in this module. (Table 1, Table S20).

Subsequently, hub genes in the modules that correlated with tissues were analyzed by co-expression network analysis (Fig. 3). The weighted correlation network showed that *CLDN1* (claudin 1) was the key gene of skin tissue, which functions as a major constituent of the tight junction complexes that regulate the permeability of epithelia³⁰. *FUCA1* (Alpha-L-Fucosidase 1) was the key gene of oviduct tissue, which encodes a lysosomal enzyme involved in the degradation of fucose-containing glycoproteins and glycolipids³¹. *LMAN2L* (Lectin, Mannose Binding 2 Like) was the most important hub gene in the ovary, which functions in the mammalian early secretory pathway³². Within heart tissue, *CORIN* (Corin, Serine Peptidase) was in the most important position and is involved in the regulation of blood volume and pressure by encoding a member of the type II transmembrane serine protease class of the trypsin superfamily³³.

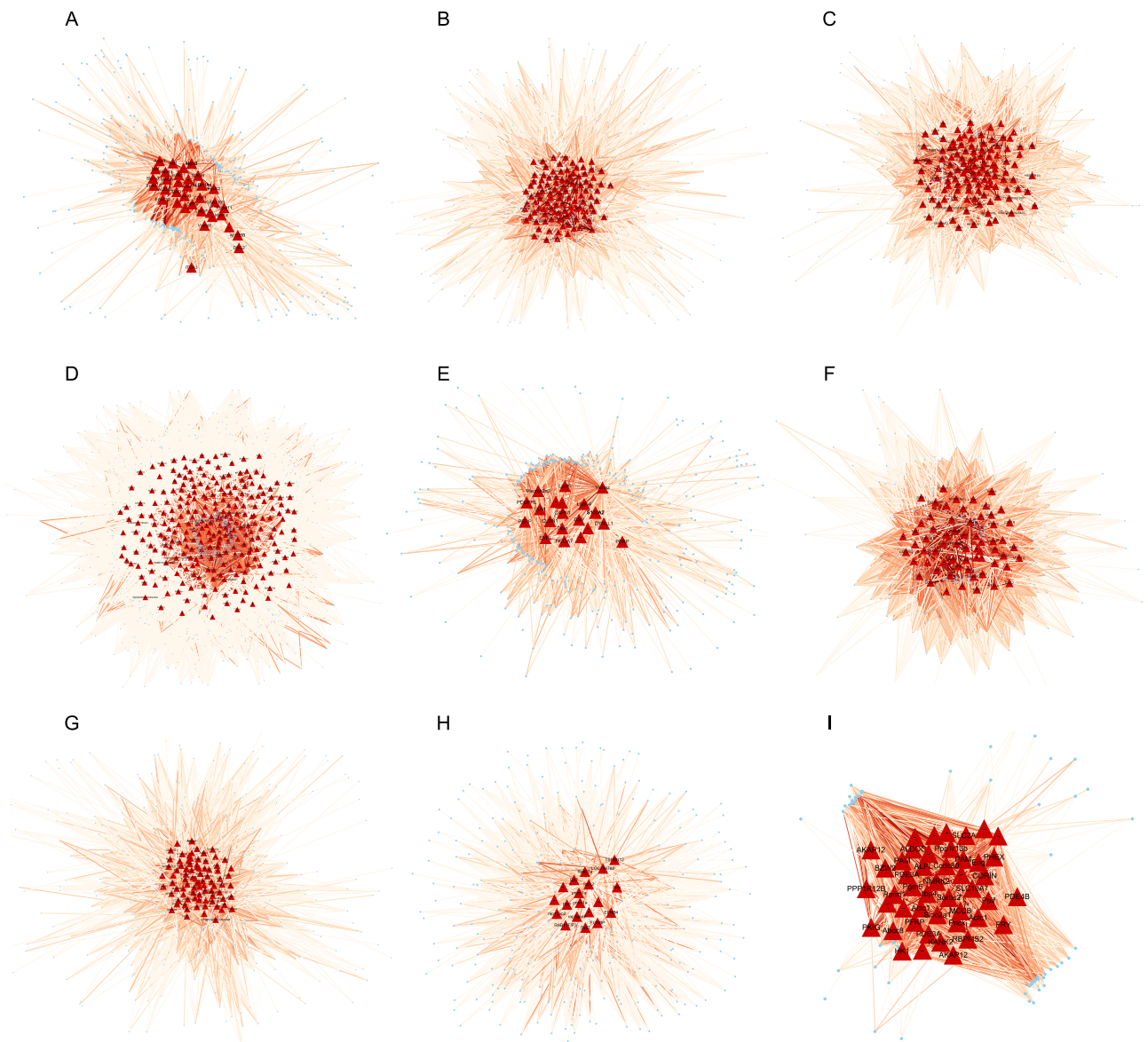


Figure 3. Weighted gene co-expression network of modules significantly correlate with tissues. The genes in network are not illuminated when the weight is less than 0.1. The triangle represents hub genes, which is statistically analyzed by WGCAN using cutoff of P value $< 1e-8$ and kME > 0.85 in each module. (A) Intestine. (B) Muscle. (C) Liver. (D) Ovary. (E) Skin. (F) Spleen. (G) Oviduct. (H) Stomach. (I) Heart.

Discussion

RNA-seq analysis is an ideal high-throughput sequencing method, which provides high efficiency and rapidity for gene discovery and is useful for studying the large and complex genomes of amphibians⁴. In this study, RNA-seq technology was used to sequence the cDNA library of 45 samples from 9 tissues of five Heilongjiang brown frogs to construct a high-quality reference gene set and further analyze its tissue-specific expression. We obtained a relatively large transcriptome dataset, the large transcriptome may be the result of mixed splicing of different tissues. Additionally, it may be due to the complexity of the amphibian genome and more gene copies were produced with transposon element expansion and low loss rate³⁴. In addition, our work also verified that RNA-seq can generate more new transcripts compared to traditional methods and more economical than genome sequencing, especially for species with large genomes³⁵.

Comparison with NR, a non-redundant protein database, showed that a large number of annotated transcripts were aligned to proteins predicted by electronic annotation. This is due to the fact that the research on amphibians has not been in-depth yet, and the functions of a large number of proteins have not been verified by experiments. These annotation results are also inferred based on the similar functions of homologous proteins in different species. The matched number of unigenes of the bull frog was larger than the sum of all other species. The alignment results demonstrated the quality of the assembled unigenes and reflected the species similarity of relative species at the molecular level³⁶.

Many differentially expressed genes in different tissues were also expected and reflected in the PCA analysis, the more DEGs between the two tissues, the farther they are in PCA plot. Because several tissues have similar functions (e.g. stomach and intestines, heart and thigh muscle) and thus their gene expression profiles are similar^{37,38}. For example, after pairwise comparisons of nine tissues, we found that the smallest number of DEGs was between the intestine and stomach. However, some tissues are very different and thus we observed large differences in their gene expression profiles. For example, the ovary is a gonadal organ, which secretes estrogen and ovulates to produce oocytes³⁹, while the spleen's functions include digestion, absorption and transportation of food, filtration and storage of blood, and immunity⁴⁰. Our results showed that there was almost no overlap in their enrichment results, and the number of DEGs between them was relatively large. The genetic divergence among the nine tissues of the Heilongjiang brown frog at the gene expression level would be helpful for the future research of its organ development.

Different tissues that were enriched in different GO terms and KEGG pathways were closely related to specific tissue functions and was consistent with similar studies in humans⁴¹.

The first example is the oviduct. We found that the UEGs in the oviduct were enriched in the extracellular domain, suggesting that the gene expression of the oviduct is more prone to exocytosis⁴². There was also a difference in the synthesis and secretion of several oviduct proteins compared to other tissues, which was consistent with a previous report⁴³, which suggested the oviduct was a passive channel for sperm and egg transport and a highly active secretory organ. We also found that the most significant oviduct molecular function items were related to enzyme inhibitor activity and microtubule polymerization. Additionally, there were more protein synthesis activities in the oviduct compared to other tissues and that is why the oviduct is known for its strong protein secretion and metabolism⁴³. In the biological process of GO enrichment, it was found that the oviduct had more glycosylation processes than other tissues. In particular, O-Glycosylation had an impact on a diversity of biological processes covering cellular aspects (targeted transport of glycoproteins), molecular aspects (protein conformation, resistance to proteolysis) and aspects involved in cellular communication (cell–cell and cell–matrix interaction) and was enriched by the upregulated genes of the oviduct. Cellular communication plays crucial roles in human infection by viral or bacterial pathogens (innate immunity) and in the progression of cancer⁴⁴. The specific glycosylated protein may be related to the nutrition and medicinal function of the oviduct⁴⁵. A large number of development-related items were enriched in the oviduct. These specific UEGs were probably the most fundamental reason for their special effects on activating cellular functions (treating scald, improving skin, anti-aging, etc.)⁴⁶. In addition, some immune related GO terms were enriched, which could provide the support of the molecular regulation for the pharmacological mechanism of the oviduct. There were a small number of genes enriched in behavior and learning, and memory items, which indicates that the oviduct may have a special expression level in these aspects⁴⁷. The two most significant pathways of KEGG annotation in the oviduct were glycosphingolipid biosynthesis globin series and O-type mucin glycan biosynthesis. Evidence for high expression of the mucin gene in the oviduct of *R. chensinensis*⁴⁸ and *X. tropicalis*⁴⁹ has been reported before. Mucin genes and O-type ligand glycosylation genes were likely candidates for maternal effects^{50–52}. A variety of amino acid metabolic pathways, including related pathways of several essential amino acid (EAA), were found in the oviduct of the Heilongjiang brown frog. Therefore, it is likely that the oviduct has specific nutritional values and functions⁵³.

Another example is the skin. We found that various enzyme inhibitors and regulatory activities, and a large number of items related to receptor binding were enriched in the skin, suggesting the tissue has a great capacity for protein synthesis. There were many receptors involved in protein expression regulation in skin, which were related to the metabolism, synthesis and secretion of proteins⁵⁴. In the biological process of GO enrichment, the skin's UEGs were mainly associated with bacterial defense response and external stimulus response. Previous study demonstrated that antimicrobial peptides were an essential part of the innate immune system in amphibians and were produced in response to a variety of external stimuli⁵⁵. The rich stimulus response protein also further indicated that the skin of the Heilongjiang brown frog, as its own defense barrier, had a positive response to external stimuli. There were many immune defense regulatory mechanisms and expression products enriched in the skin. Jiang and Shang⁵⁶ isolated Fraction III, a mixture of peptides from Chinese forest frog skin, and demonstrated that these bioactive peptides have antimicrobial activity against bacteria. Xiao et al.⁵⁷ extracted five active peptide types from Chinese forest frog skin and found that the peptides had very high activity against gram-negative and gram-positive bacteria. These results indicate that the skin of amphibians is the main source of natural antimicrobial peptides and analogues. In the follow-up work, transcriptome data can be used to predict, screen and verify antimicrobial peptides from the skin of the Heilongjiang brown frog⁵⁸. This research may highlight new varieties of antimicrobial peptides to support antibiotic development.

In conclusion, we assembled and annotated a fairly comprehensive gene set of the Heilongjiang brown frog. We also detected the gene expression differences among different tissues. The findings will provide basic data for genomics, gene cloning and functional verification of the Heilongjiang brown frog and other closely related amphibians.

Data availability

The raw RNA-seq data were deposited in the China National GeneBank DataBase (CDGBdb) with the accession number CNP0001589.

Received: 6 June 2022; Accepted: 17 November 2022

Published online: 01 December 2022

References

- Wake, D. B. & Koo, M. S. Amphibians. *Curr. Biol.* **28**(21), R1237–R1241. <https://doi.org/10.1016/j.cub.2018.09.028> (2018).
- Lu, B. *et al.* A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). *Mol. Ecol. Resour.* <https://doi.org/10.1111/1755-0998.13319> (2021).
- Hammond, S. A., Warren, R. L., Vandervalk, B. P., Kucuk, E. & Birol, I. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat. Commun.* **8**(1), 1433. <https://doi.org/10.1038/s41467-017-01316-7> (2017).
- Geng, X., Zhang, L., Zang, X., Guo, J. & Xu, C. RNA-seq analysis provides insight into molecular adaptations of *Andrias davidianus*. *Dev. Genes Evol.* **229**(5–6), 197–206. <https://doi.org/10.1007/s00427-019-00641-9> (2019).
- Li, F. G. *et al.* RNA-Seq analysis and gene discovery of *Andrias davidianus* using Illumina short read sequencing. *PLoS ONE* **10**, e0123730. <https://doi.org/10.1371/journal.pone.0123730> (2015).
- Xu, Y. G. *et al.* Transcriptome profiling and digital gene expression analysis of the skin of Dybowskii's frog (*Rana dybowskii*) exposed to *Aeromonas hydrophila*. *Appl. Microbiol. Biotechnol.* **101**(14), 5799–5808. <https://doi.org/10.1007/s00253-017-8385-3> (2017).
- Xiong, J., Lv, Y., Huang, Y. & Liu, Q. The first transcriptome assembly of yenyuan stream salamander (*Batrachuperus yenyuanensis*) provides novel insights into its molecular evolution. *Int. J. Mol. Sci.* **20**(7), 1529. <https://doi.org/10.3390/ijms20071529> (2019).
- Zhou, M. *et al.* Components of the peptidome and transcriptome persist in lin wa pi: The dried skin of the Heilongjiang brown frog (*Rana amurensis*) as used in traditional Chinese medicine. *Peptides* **27**(11), 2688–2694. <https://doi.org/10.1016/j.peptides.2006.05.009> (2016).
- Song, B. J., Li, C. S., Huang, Z. Y., Li, D. & Wang, C. Q. Comparative study of oviductus production performance on *Rana temporaria chensinensis* and *Rana amurensis*. *J. Chin. Med. Mater.* **39**(12), 2711–2715. <https://doi.org/10.13863/j.issn1001-4454.2016.12.009> (2016).
- Hu, X., Liu, C. B., Chen, X. P. & Wang, L. M. Main nourishment components of Oviductus ranae. *J. Jilin Agric. Univ.* **02**, 218–220. <https://doi.org/10.13327/j.jjlau.2003.02.027> (2003).
- Li, Y. Y., Zheng, W. X., Wang, R. X. & Su, X. R. Nutritive material of *Rana chensinensis* by multivariate analysis methods. *Food Sci.* **12**, 457–460. [https://doi.org/10.1016/S1872-583X\(07\)60011-4](https://doi.org/10.1016/S1872-583X(07)60011-4) (2017).
- Cao, Y., Zhang, R. W., Chen, X. H., Fu, C. W. & Bi, K. S. Determination of amino acids in different parts of *Rana amurensis* Boulenger. *Amino Acids Biotic Resour.* **02**, 17–19. <https://doi.org/10.1088/1674-4527/10/12/011> (2010).
- Fu, C. W. *et al.* Determination of trace elements in *Rana amurensis* Boulenger by ICP-MS. *J. Shenyang Pharmaceut. Univ.* **3**, 216–220. <https://doi.org/10.1007/s11606-010-1517-4> (2011).
- Patel, R. K., Mukesh, J. & Liu, Z. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**(2), e30619. <https://doi.org/10.1371/journal.pone.0030619> (2012).
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M. & Regev, A. D. novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocol.* **8**(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084> (2013).
- Fu, L. M., Niu, B. F., Zhu, Z. W., Wu, S. T. & Li, W. Z. CD-HIT: Accelerated for clustering the next generation sequencing data. *Bioinformatics* **23**, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565> (2012).
- Seppy, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14 (2019).
- Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D1–65. <https://doi.org/10.1093/nar/gkl842> (2007).
- Boeckmann, B. The swiss-prot protein knowledgebase and its supplement trembl in 2003. *Nucleic Acids Res.* **31**(1), 365–370. <https://doi.org/10.1093/nar/gkg095> (2003).
- Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**(2), R7–R7. <https://doi.org/10.1186/gb-2004-5-2-r7> (2004).
- Stefan, G. *et al.* High-throughput functional annotation and data mining with the blast2go suite. *Nucleic Acids Res.* **36**(10), 3420–3435. <https://doi.org/10.1093/nar/gkn176> (2008).
- Ye, J. *et al.* WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, 293–297. <https://doi.org/10.1093/nar/gkl031> (2006).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**(4), 357–360. <https://doi.org/10.1038/nmeth.3317> (2015).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**(4), 417–419. <https://doi.org/10.1038/nmeth.4197> (2017).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(12), 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omic J. Integr. Biol.* **16**(5), 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2018).
- Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
- Kirschner, N., Rosenthal, R., Furuse, M., Moll, I. & Brandner, J. M. Contribution of tight junction proteins to ion, macromolecule, and water barrier in keratinocytes. *J. Invest. Dermatol.* **133**(5), 1161–1169. <https://doi.org/10.1038/jid.2012.507> (2013).
- Occhiodoro, T., Beckmann, K. R., Morris, C. P. & Hopwood, J. J. Human alpha-L-fucosidase: Complete coding sequence from cDNA clones. *Biochem. Biophys. Res. Commun.* **164**(1), 439–445. [https://doi.org/10.1016/0006-291x\(89\)91739-7](https://doi.org/10.1016/0006-291x(89)91739-7) (1989).
- Alkhatir, R. A. *et al.* Dominant LMAN2L mutation causes intellectual disability with remitting epilepsy. *Ann. Clin. Transl. Neurol.* **6**(4), 807–811. <https://doi.org/10.1002/acn3.727> (2019).
- Yan, W., Sheng, N., Seto, M., Morser, J. & Wu, Q. Corin, a mosaic transmembrane serine protease encoded by a novel cDNA from human heart. *J. Biol. Chem.* **274**(21), 14926–14935. <https://doi.org/10.1074/jbc.274.21.14926> (1999).
- Sun, C. *et al.* LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**(2), 168–183. <https://doi.org/10.1093/gbe/evr139> (2012).
- Wilhelm, B. T. & Landry, J. R. RNA-Seq quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**(3), 249–257. <https://doi.org/10.1016/j.ymeth.2009.03.016> (2009).
- Yuan, Z. Y. *et al.* Spatiotemporal diversification of the true frogs (*Genus Rana*): A historical framework for a widely studied group of model organisms. *Syst. Biol.* **65**(5), 824–842. <https://doi.org/10.1093/sysbio/syw055> (2016).
- Callis, T. E., Chen, J. F. & Wang, D. Z. MicroRNAs in skeletal and cardiac muscle development. *DNA Cell Biol.* **26**(4), 219–225. <https://doi.org/10.1089/dna.2006.0556> (2007).
- Johansson, M. E., Sjvall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**(6), 352–361. <https://doi.org/10.1038/nrgastro.2013.35> (2013).

39. Smith, P., Wilhelm, D. & Rodgers, R. J. Development of mammalian ovary. *J. Endocrinol.* **221**(3), R145–R161. <https://doi.org/10.1530/JOE-14-0062> (2014).
40. Andrea, B., Maria-Manuela, R., Rita, C., Licia, S. & Neil, D. Development and function of the mammalian spleen. *BioEssays* **29**(2), 166–177. <https://doi.org/10.1002/bies.20528> (2010).
41. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteom.* **13**(2), 397–406. <https://doi.org/10.1074/mcp.M113.035600> (2014).
42. Hynes, H. & Richard, O. The extracellular matrix: Not just pretty fibrils. *Science* **326**(5957), 1216–1219. <https://doi.org/10.1126/science.1176009> (2019).
43. Liu, Y., Weng, J., Huang, S., Shen, Y. & Weng, Q. Immunoreactivities of PPAR γ 2, leptin and leptin receptor in oviduct of Chinese brown frog during breeding period and pre-hibernation. *Eur. J. Histochem.* **58**(3), 2422. <https://doi.org/10.4081/ejh.2014.2422> (2014).
44. Breloy, I. & Hanisch, F. G. Functional roles of O-glycosylation. *Molecules* **23**(12), 3063. <https://doi.org/10.3390/molecules23123063> (2018).
45. Liu, Y. & Liu, L. X. Ingredient and pharmacology research progress of Oviductus Ranae. *Strait Pharmaceut. J.* **19**(12), 1–3. <https://doi.org/10.3969/j.issn.1006-3765.2007.12.001> (2007).
46. Zhou, L., Du, N., Fu, T. F. & Zhao, L. X. Effect of different drying methods on protein content of *Rana chensinensis* oil. *Chin. J. Anim. Husbandry Vet. Med.* **8**(8), 32–32. <https://doi.org/10.3969/J.ISSN.1671-6027.2014.08.019> (2014).
47. Chen, M. R. *et al.* Effect of *Rana japonica* oil compound granules on learning/memory ability of rats exposed to microwave radiation under hypergravity environment. *Chin. J. Public Health* **12**, 1591–1593. <https://doi.org/10.1038/cmi.2011.4> (2011).
48. Zhang, M. *et al.* Transcriptome sequencing and de novo analysis for Oviductus Ranae of *Rana chensinensis* using illumina RNA-seq technology. *J. Genet. Genom.* **40**(3), 137–140. <https://doi.org/10.1016/j.jgg.2013.01.004> (2013).
49. Lang, T. G. *et al.* Searching the evolutionary origin of epithelial mucus protein components-mucins and FCGBP. *Mol. Biol. Evol.* **33**(8), 1921–1936. <https://doi.org/10.1093/molbev/msw066> (2016).
50. Coppin, A., Maes, E., Flahaut, C., Coddeville, B. & Strecker, G. Acquisition of species-specific O-linked carbohydrate chains from oviducal mucins in *Rana arvalis*. A case study. *Eur. J. Biochem.* **266**(2), 370–382. <https://doi.org/10.1046/j.1432-1327.1999.00862.x> (1999).
51. Guerardel, Y. *et al.* O-glycan variability of egg-jelly mucins from *Xenopus laevis*: Characterization of four phenotypes that differ by the terminal glycosylation of their mucins. *Biochem. J.* **352**(2), 449–463. <https://doi.org/10.1042/0264-6021:3520449> (2000).
52. Coppin, A., Florea, D., Maes, E., Cogalniceanu, D. & Strecker, G. Comparative study of carbohydrate chains released from the oviducal mucins of the two very closely related amphibian species *Bombina bombina* and *Bombina variegata*. *Biochimie* **85**(1/2), 53–64. [https://doi.org/10.1016/S0300-9084\(03\)00021-X](https://doi.org/10.1016/S0300-9084(03)00021-X) (2003).
53. Rondanelli, M. *et al.* Effect of essential amino acid supplementation on quality of life, amino acid profile and strength in institutionalized elderly patients. *Clin. Nutr.* **30**(5), 571–577. <https://doi.org/10.1016/j.clnu.2011.04.005> (2011).
54. Pantic, J. M. *et al.* The potential of frog skin-derived peptides for development into therapeutically-valuable immunomodulatory agents. *Molecules* **22**(12), 2071. <https://doi.org/10.3390/molecules22122071> (2017).
55. Conlon, J. M. & Mechkarska, M. Host-defense peptides with therapeutic potential from skin secretions of frogs from the family pipidae. *Pharmaceut. (Basel, Switzerl.)* **7**(1), 58–77. <https://doi.org/10.3390/ph7010058> (2014).
56. Jiang, L. & Shang, D. Antimicrobial activity of skin antimicrobial peptide in Chinese forest frogs. *Chin. J. Zool.* **9**, 70–72. <https://doi.org/10.1007/BF02911033> (2004).
57. Xiao, X., Xu, Y. & Chai, L. The isolation and extraction of active peptides from the skin of *Rana chensinensis*. *J. Northeast Forest Univ.* **33**, 44–46. <https://doi.org/10.1360/jos162021> (2005).
58. Shelomi, M., Jacobs, C., Vilcinskas, A. & Vogel, H. The unique antimicrobial peptide repertoire of stick insects. *Dev. Comp. Immunol.* **103**, 103471. <https://doi.org/10.1016/j.dci.2019.103471> (2020).

Acknowledgements

This work was supported by National Key Program of Research and Development, Ministry of Science and Technology (2016YFC0503200).

Author contributions

W.Y.L., Y.L., L.W.H., and R.X.T. performed the bioinformatics analyses; W.Y.L., R.X.T. and L.W. collected the samples; W.Y.L., Y.L., L.W. and L.W.H. wrote the manuscript; Z.X.F., M.P. and B.S.Y. revised the manuscript; Z.X.F. and B.S.Y. designed and supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24631-6>.

Correspondence and requests for materials should be addressed to B.Y. or Z.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022