### **Original Research**

## Transcriptome profiling reveals gene regulation programs underlying tail development in the Ornamented Pygmy frog *Microhyla fissipes*

Shouhong Wang<sup>1,2</sup>, Lusha Liu<sup>1</sup>, Yun-Bo Shi<sup>2,\*</sup>, Jianping Jiang<sup>1,\*</sup>

<sup>1</sup>Chengdu Institute of Biology, Chinese Academy of Sciences, 610041 Chengdu, Sichuan, China, <sup>2</sup>Section on Molecular Morphogenesis, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, Maryland, MD 20892, USA

### TABLE OF CONTENTS

- 1. Abstract
- 2. Introduction
- 3. Materials and methods
  - 3.1 Experimental animals
  - 3.2 Sample collection
  - 3.3 Second-generation sequencing and SMART sequencing
  - 3.4 DET analysis and GO/KEGG enrichment analysis
- 4. Results
  - 4.1 Morphological changes associated with tail development in Microhyla fissipes
  - 4.2 Comparison between full length and de novo transcripts and functional annotation
  - 4.3 DETs and enriched GO terms and KEGG pathways underlying tail development
  - 4.4 Distinct genes and biological processes involved in tailbud determination and outgrowth
- 5. Discussion
- 6. Conclusions
- 7. Author contributions
- 8. Ethics approval and consent to participate
- 9. Acknowledgment
- 10. Funding
- 11. Conflict of interest
- 12. Availability of data and materials
- 13. References

### 1. Abstract

**Introduction**: Tadpole tail develops from the tailbud, an apparently homogenous mass of cells at the posterior of the embryo. While much progress has been made in understanding the origin and the induction of the tailbud, the subsequent outgrowth and differentiation have received much less attention, particularly with regard to global gene expression changes. **Methods**: By using RNA-seq with SMRT and further analyses, we report the transcriptome profiles at four key stages of tail development, from a small tailbud to the onset of feeding (S18, S19, S21 and S28) in *Microhyla fissipes*, an anuran with a number of advantages for developmental and genetic studies. **Results**: We obtained 48,826 transcripts and discovered 8807 differentially expressed transcripts (DETs, q < 0.05) among these four developmental stages. We functionally classified these DETs by using GO and KEGG analyses and revealed 110 significantly enriched GO categories and 6 highly enriched KEGG pathways (Protein digestion and absorption; ECM-receptor interaction; Pyruvate metabolism; Fatty acid degradation; Valine, leucine and isoleucine degradation; and Glyoxylate and dicarboxylate metabolism) that are likely critically involved in developmental changes in the tail. In addition, analyses of DETs between any two individual stages demonstrated the involvement of distinct biological pathways/GO terms at different stages of tail development. Furthermore, the most dramatic changes in gene expression profile are those between S28 and any of the other three stages. The upregulated DETs at S28 are highly

Submitted: 2 July 2021 Revised: 25 July 2021 Accepted: 10 August 2021 Published: 30 November 2021 This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/). © 2021 The Author(s). Published by BRI.

enriched in "myosin complex" and "potassium channel activity", which are important for muscle contraction, a critical function of the tail that the animal needs by the end of embryogenesis. Additionally, many DETs and enriched pathways discovered here during tail development, such as HDAC1, Hes1 and Hippo signaling pathway, have also been reported to be vital for the tissue/organ regeneration, suggesting conserved functions between development and regeneration. **Conclusion**: The present staudy provides a golbal overview of gene expression patterns and new insights into the mechanism involved in anuran tail development and regeneration.

### 2. Introduction

The tadpole tail is a larva-specific organ, which is twice as long as the body. It forms at the early embryonic stages and is lost at the climax of metamorphosis in anurans [1, 2]. The tail is anatomically defined by its position from posterior to the anal opening, a continuation of the structures of the main body axis, which is well patterned and contains many axial and paraxial tissues, including epidermis, connective tissue (dermis), spinal cord, notochord, dorsal aorta, and skeletal muscle [3, 4]. It has a number of functions at larval stages, such as swimming, balance, fat storage, and predator escape via autotomy [3].

There are two phases in tail development, tailbud determination and tailbud outgrowth. Studies in Xenopus laevis have shown that the tailbud is first formed at the end of gastrulation, and then commences to outgrowth at the tailbud stage [4, 5]. The molecular NMC model (N, the neural plate to anterior to M; M, tail somite; C, caudal notochord) has been proposed to explain tailbud induction [6]. In addition, various genes, such as Xlim1, Xbro, Xcad3, Xpo, BMP4, Tbx6 and Gdf11, and signaling pathways including the Wnt, BMP, and Notch pathways, have been analyzed and shown to be regulated and/or involved in tailbud induction [3, 6]. However, no studies have employed systematic, high-throughput sequencing technologies, such as RNA-Seq and single molecule long-read sequencing (SMRT sequencing), for global analyses of gene expression changes and tail development.

In an earlier study [7], we took advantages of SMRT sequencing and RNA-Seq methods to generate a full length and high-quality reference transcriptome and analyzed the global gene expression changes associated with tail resorption during the metamorphosis of *Microhyla fissipes*, an anuran with a number of advantages for developmental and genetic studies, such as transparent tadpoles and fast development [8, 9]. Here, we have used the same approach to investigate the global gene regulation profiles involved in the initial formation of the tail during embryogenesis in *Microhyla fissipes*. We have obtained 48,826 transcripts and identified 8807 differentially expressed transcripts (DETs). These DETs are further analyzed with re-

gard to enriched Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, revealing 110 significantly enriched GO categories and 6 significantly enriched KEGG pathways that are highly correlated with the developmental changes in the tail. Our results provide a global overview of gene expression patterns and new insights into the possible mechanism involved in tail development of frogs.

### 3. Materials and methods

#### **3.1 Experimental animals**

*Microhyla fissipes* breeding adults were collected from Shuangliu, Chengdu, China (30.5825°N, 103.8438°E) in June 2016. All animal care and treatments were done as approved by Animal Care Committee, Chengdu Institute of Biology (Permit Number: 20150121003). All animals were obtained and maintained as previously described [7]. Developmental stages were determined as described [9].

#### 3.2 Sample collection

Tadpoles at stage 18 (S18), S19, S21, and S28 were randomly collected and anesthetized in 0.01% MS222 for tail morphology observation. For gene expression analysis, three biological replicates, each containing 10–30 animal tails, were used for RNA-seq at each developmental stage. To maximize the discovery of *M. fissipes* gene pools, heart, liver, spleen, lung, kidney, skin, ovary, testis from adult *M. fissipes*, tails (at S38, S40, S41, S43) and dorsal muscle (at S36, S40, S43, S45) were dissected for SMRT sequencing. All dissected tissue samples were immediately frozen in liquid nitrogen and stored at -80 °C.

# 3.3 Second-generation sequencing and SMART sequencing

The detailed method for RNA-Seq and SMRT sequencing, and functional annotation of transcripts were described previously [7]. All raw data from Illumina sequencing have been deposited in the NCBI Short Read Archive (SRA) database with the accession number PRJNA504611.

#### 3.4 DET analysis and GO/KEGG enrichment analysis

To compare the transcript levels in the tail among the four developmental stages, FPKM (fragments per kilobase of exon model per million mapped reads) was used to evaluate the transcript levels among the four stages. Then, we used DESeq R package (1.10.1) which is available at https://www.huber.embl.de/users/anders/DESeq/ to identify DETs. The resulting DETs were subjected to GO and KEGG pathway enrichment analyses. The details for these analyses were as reported before [7].



**Fig. 1. Morphological changes associated with tail development of** *M. fissipes* **embryos/tadpoles at different stages used in our study.** S18: the neural tube stage when tail bud is small but visible; S19: tail bud I; S21: tail bud III when tail fin and medial cloacal tail piece are distinct; S28: end of embryogenesis and onset of feeding (see more detail at [9]). Red dashed line marks the point of dissection for tissue isolation. The time (hr) from fertilization to each of the four stages are also indicated.

### 4. Results

# 4.1 Morphological changes associated with tail development in *Microhyla fissipes*

We studied two phases of tail development: tailbud determination and tailbud outgrowth, in *M. fissipes* (Fig. 1, Ref. [9]). Tailbud determination from S18 (neural tube) to S19 (elongated tailbud) is defined by tailbud elongation while tailbud outgrowth refers to stages from S19 to S28, when a functional tail is formed [5]. At S21, tail fin and medial cloacal tail piece were distinct and the tail starts to outgrow. S28 is the end of embryogenesis, when the tadpole swims freely and start to feed, while at S18–S21, the tadpoles hardly move. Tail development is a fast process in *M. fissipes*, needing only 15 hours from fertilization to neural tube (tail determination) and 3.5 days to the end of embryogenesis (S28). At each sampling stage, the tailbud was cut from posterior to cloaca as indicated by the red dash line in Fig. 1 and used for RNA isolation.

# **4.2** Comparison between full length and *de novo* transcripts and functional annotation

Total RNA used for SMRT sequencing was isolated from heart, liver, spleen, lung, kidney, skin, ovary, and testis from two adults M. fissipes as well as from tadpole tail including at four stages (S38, S40, S41, S43) and tadpole dorsal muscle at three stages (S36, S43, S45). The RNA was sequenced on a PacBio RSII platform. All highquality short reads from the RNA-Seq from the tadpole tails at eight developmental stages (S18, S19, S21, S28, S39, S40, S41, S43) were used to error-correct the SMRT sequences as previously described [7] and analyzed by using the Proovread correction software [10]. We found that *de* novo transcripts in the 200-500 bp range account for the most, 60.11% (360,893) of the total reads (Table 1) while only 0.22% (149) of that in SMRT sequencing data [7]. However, 30.78% of full-length transcripts had the length >2000 bp while only 3.85% of that in *de novo* transcripts (Table 1).

Through basic alignments to public databases, in total, 132,414 (22.05%), 82,804 (13.79%), 46,184 (7.69%), 116,599 (19.42%), 118,544 (19.74%), 121,829 (20.29%)

Index	Illumina RNA-Seq	Database	Annotation number of RNA-Seq (Annotation rate)
Total Nucleotides	374444178	NR	132414 (22.05%)
All Unigenes	600432	NT	82804 (13.79%)
Length (200–500 bp)	360893	KEGG	46184 (7.69%)
Length (500–1k bp)	157528	SwissProt	116599 (19.42%)
Length (1k–2k bp)	58874	PFAM	118544 (19.74%)
Length (>2k bp)	23137	GO	121829 (20.29%)
Mean length	624	KOG	42249 (7.04%)
Max length	28898	Annotated in all databases	15384 (2.56%)
N50	769	Annotated in at least one database	196154 (32.67%)

and 42,249 (7.04%) *de novo* transcripts were annotated in the NR, NT, KEGG, SwissProt, PFAM, GO and KOG databases, respectively (Table 1). However, the annotation rate of *de novo* transcripts in each database was much lower than that of SMRT transcripts, such as 47,112 (69%) and 41,254 (61%) in the NR and GO, respectively [7]. Totally, there are only 32.67% of *de novo* transcripts were annotated in at least one database while the annotation rate of fulllength transcripts were 78% [7]. These data indicated that transcripts obtained from SMRT sequencing were highly improved compared to *de novo* RNA-seq data.

# 4.3 DETs and enriched GO terms and KEGG pathways underlying tail development

In order to identify the gene expression program underlying tail development, we mapped RNA-Seq clean reads at four indicated stages (S18, S19, S21, S28) to proofread-corrected reference sequences. Totally 329,650,058 (50.11%) reads were mapped (Supplementary Table 1). Then, the expression level was determined by using RSEM software [11]. 48,826 expressed transcripts with FPKM >0.3 were identified in the tail among the four stages. Among of them, 22,208 transcripts were commonly expressed at all four stages while 2313, 854, 2576 and 5305 transcripts were uniquely expressed at S18, S19, S21 and S28, respectively (Fig. 2). Next, pair-wise comparisons of gene expression among the four stages were carried out, revealing 8807 DETs in the tail during tailbud determination and outgrowth (Fig. 3, Supplementary Table 2). Among the six pair-wise groups, most DETs were found when any of the other three stages were compared with S28 while relatively few DETs were found when comparisons were done between any two stages among S18, S19 and S21 (Fig. 3). Overall, we identified 4284 DETs (3695 upregulated, 589 downregulated), 5311 DETs (4075 upregulated, 1236 downregulated) and 3093 DETs (2624 upregulated, 469 downregulated) for comparisons between S28 vs S18, S28 vs S19 and S28 vs S21, respectively. These results indicated that most of the DETs were upregulated at S28 comparing to other three stages during tail development.

Hierarchical clustering of totally 8807 DETs were performed and the results were shown in a heatmap of the



**Fig. 2. Venn diagram analyses reveal stage-specific of gene expression.** Venn diagrams showing the number of expressed transcripts (FPKM >0.3) at different stages during tail development. Note that totally 48,826 transcripts were expressed, among which 22,208 transcripts were expressed at all four stages, and that each stage had some uniquely expressed transcripts, with S28 having the most (5305). FPKM: fragments per kilobase of exon model per million mapped reads.

expression levels of each DET at four stages in Fig. 4. The DETs fell largely into two groups with highest or lowest expression levels at S28, and there were more DETs in the group with the highest expression level at S28 (Fig. 4). Thus, tail maturation at S28 involves more DETs than the other developmental stages and that most of the DETs involved in tail maturations are upregulated by S28. In addition, the tail gene expression patterns at S18, S19, and S21 were very similar to each other, especially between S18 and S19, consistent with the lack of significant morphology changes in the tail between S18 to S19 but drastic changes at S28 (Fig. 1).

To understand the global gene functional categories and biological pathways during tail development, GO and KEGG analyses were performed on the 8807 DETs. These analyses revealed 110 significantly enriched GO categories and 6 highly enriched KEGG pathways (q < 0.05) (**Supplementary Fig. 1, Supplementary Table 3**). Inter-



Fig. 3. Pair-wise comparisons reveal that S28 tail has the highest numbers of differentially expression transcripts (DETs, q < 0.05) during tail development. Volcano plot of DETs in the tail for six different comparison groups. X axis represents fold-change (log<sub>2</sub>FC) of the DETs and the Y axis represents the –log10 padj- (adjusted P- or q-) value of the DET, with 0.05 set as the significance level cutoff. Up-regulated DETs are shown in red while down-regulated ones in green; non-significantly DETs are presented in blue.

estingly, 5 of the 6 enriched KEGG pathways were different metabolic processes while the last one was ECM-receptor interaction, highlighting the importance of metabolism during tailbud growth period. In support of this, the most significantly enriched GO terms were related to oxidoreductase activity and metabolism/catabolism.

## 4.4 Distinct genes and biological processes involved in tailbud determination and outgrowth

To correlate the gene expression changes with the two phase of tail development, we first used K-means algorithm to cluster the DETs based on gene expression profiles across the four developmental stages. The DETs were scattered into eighteen clusters (Supplementary Fig. 2). To explore the gene regulation programs underlying different developmental stages, we focused on cluster 15, cluster 13, and cluster 3, in which the expression of the DETs were upregulated to high levels by S19, S21 and S28, respectively, suggesting their involvement in the respective transitions in the developmental stages. The DETs in each of the three clusters were classified into three GO categories by the Blast2GO, including biological process (BP), cellular component (CC), and molecular function (MF), and directed acyclic graph (DAG) was used to display the GO structures (Fig. 5). Top 10 GO term levels were selected as the master nodes in which only highly enriched GO term levels, such as levels 1, 6, 7, 8, 9, and 10, were displayed, with no significantly enriched GO terms were found in the GO

term levels 2–5 (Fig. 5A, note that a node represents a GO term, an arrow represented a parent node, and the branches represent the containment relationships, with the range of function decreasing in size from top to bottom). For cluster 15, only DETs within the BP category were significantly enriched and the DAG showed that all enriched GO terms were eventually linked to three GO nodes: leucine catabolic process, valine catabolic process and tryptophan metabolic process (level 10 GO terms, Fig. 5A), suggesting an important role of amino acid metabolism in the tail development at S19 and later.

For cluster 13, we found significantly enriched GO terms in three GO categories, BP, CC and MF and that there was a single node within each GO category linked to all enriched GO terms (Fig. 5B). These single nodes were GO term "cytoskeletal anchoring at plasma membrane" within BP, "dystrophin-associated glycoprotein complex" within CC, and "hyaluronic acid binding" within MF (Fig. 5B), indicating an important role of these process for the transition from tailbud determination to tailbud outgrowth (S19–S21).

Finally, the DETs upregulated by S28 (cluster 3) also had significantly enriched GO terms in all three GO categories, BP, CC and MF. There was a single node, the GO term "potassium ion ( $K^+$ ) transport" within BP. There were two nodes, the GO terms "myosin complex" and "intermediate filament cytoskeleton" within CC, and two nodes, the GO terms "potassium channel activity" and "fructose 1,6-bisphosphate 1-phosphatase activity" within



**Fig. 4. A heatmap showing the global expression patterns of the 8807 DETs at the four stages.** The intensity of color indicates relative expression levels. Colors from red to blue correspond to high to low levels of expression. Note that the DETs fall into largely two categories with highest or lowest expression at stage 28, respectively. In addition, the expression levels for the majority of the DETs are most different between S18/19 and S28.

MF, respectively (Fig. 5C). These processes are likely important for the function of the tail as embryogenesis is complete by S28. Interestingly, a K<sup>+</sup>-related GO term was a node within both BP and CC, implicating a role of K<sup>+</sup>-dependent processes in tail maturation and/or function. Notably, the most significant difference between S28 and S18–21 is that tail can move freely at S28 but not at S18–21, and muscle activity was reported to associate with potassium displacements. Importantly, muscle K<sup>+</sup> homeostasis

contributes to muscle fatigue and recovery [12, 13]. It is intriguing that tadpole tail muscle can keep K<sup>+</sup> homeostasis under the almost nonstop contracting. Our finding that the genes in potassium ion processes are involved in tail development may have important implications for understanding muscle fatigue in human.

When we carried out KEGG pathway analysis on these three clusters (#15, 13, 3), we found 2, 7 and 11 KEGG pathways highly enriched (q < 0.05) for the three clusters, respectively (Supplementary Fig. 3, Supplementary Table 4; Supplementary Fig. 4, Supplementary Table 5; Supplementary Fig. 5, Supplementary Table 6). The two KEGG pathways for cluster 15 were "Pyruvate metabolism" and "Citrate cycle (TCA cycle)" (Supplementary Fig. 3, Supplementary Table 4), both of which are related to the energy metabolism, likely important during tailbud outgrowth because energy metabolism is essential for the biosynthesis of amino acids, nucleotide and lipid during development. Both clusters 13 and 3 highly enriched (q < 0.05) KEGG pathways "ECM-receptor interaction" and "Focal adhesion", implicating a critical involvement of cell-cell and cell-EMC interactions during late stages of tail development as genes in these clusters were upregulated by stage 21 and 28, respectively.

In contrast to the above three clusters, in which the DETs were upregulated to highest levels at S28, two clusters 4 and 11, with a combined 1485 DETs, were DETs whose expression was downregulated to lowest levels by S28 (Supplementary Fig. 2). Analyses of these 1485 DETs revealed the enrichment of 31 GO terms (Supplementary Fig. 6A, Supplementary Table 7) and 9 KEGG pathways (Supplementary Fig. 6B, Supplemen**tary Table 7**) with q < 0.05. Within the GO category Biological Process, there were 20 organ development GO terms highly enriched (Supplementary Table 7). To further investigate the potential function of the DETs involved in tail development, we identified the common DETs present in each of these 20 GO terms and found 29 DETs, including Xld3, HDAC1, dkk1, 14-3-3, Ruvb-like2, Foxc1, Xiro3, Arih2, Hes1, etc. (Fig. 6, Supplementary Table 8). All these genes were highly expressed at S18 and/or S19, but were gradually repressed from S21 to S28, suggesting that these genes are important for early tail development, i.e., tail determination period (Fig. 6), but not for tail outgrowth or that their repression is important for tail outgrowth.

The KEGG pathways enriched among these 29 downregulated DETs included TGF- $\beta$  signaling pathway, signaling pathways regulating pluripotency of stem cells and Notch signaling pathway. Common among these pathways was the regulation of *smad1/5/8*, *Id*, *Hes1/5* and *HDAC* genes.

It is of interest that of the 9 highly enriched (q < 0.05) KEGG pathways among downregulated DETs in clusters 4 and 11 were Cell cycle and Hippo signaling pathway (**Supplementary Fig. 6B**, **Supplementary Table 7**)



**Fig. 5. Directed acyclic graph (DAG) displaying the results of DEG enrichment according to GO analysis in three different DEG clusters.** "N" in the left panel represents the number of DEGs in each cluster. For each cluster, the significantly enriched GO terms were arranged according to the GO levels, with the first level being either BP (Biological Process), CC (Cellar component), or MF (Molecular Function). Each box and circle indicate a GO term and the descriptions inside of which, from top to bottom, show GO term id, GO description, q value and the enriched DETs numbers/background gene numbers in the indicated GO term, respectively. Color depth represents the enrichment degree (based on q value, red is the most significantly enriched). Note that some levels had no GO terms shown as there were no GO terms enriched among the DEGs. (A) A cluster where the DETs were upregulated by S19 and remained highly expressed subsequently. These DETs were only significantly enriched. (B) A cluster where the DETs were upregulated by S21 and remained highly expressed subsequently. The GO terms "cytoskeletal anchoring at plasma membrane" of BP and "dystrophin-associated glycoprotein complex" of CC, and "hyaluronic acid binding" of MF are the most significantly enriched. (C) A cluster where the DETs were upregulated by S28. The GO terms "potassium ion transport" of BP and "myosin complex" of CC, and "potassium channel activity" of MF are the most significantly enriched.



Fig. 6. A heatmap showing the developmental expression patterns of the 29 EDTs present in total 20 organ development GO terms enriched among genes downregulated between stage 18 and stage 28. Most of the genes are highly expressed at S18 and S19 while downregulated by S28. The colors red to blue corresponds to high to low levels of expressions.

(Fig. 7). Both Cell cycle and Hippo pathways are very important for the organ development. The downregulation of genes in these pathways, including *Skp2*, *CDK2-cyclinA*, *Mps1*, *HDAC*, *Cdh1* and *Yap/TAZ*, *Slug*, etc. (Fig. 7), suggest that they are involved in tailbud determination and perhaps early tailbud outgrowth but are no longer needed as tail development approaches completion by S28.

### 5. Discussion

RNA-Seq has been widely applied to studies in various species, especially non-model organisms, by offering a novel and efficient approach to generate transcriptome sequence [14, 15]. In addition, it allows researchers to focus on highly functional parts of the genome while avoiding complicated analyses of introns. However, it has 1009



**Fig. 7.** Selected pathways enriched among DETs that are highly expressed between S18 and S19 but downregulated by S28. (A) Cell cycle pathway. Note that *Skp2*, *CDK2-cyclinA*, *HDAC*, *Mps1* and *Cdh1*, etc., were among the regulated genes. (B) Hippo signaling pathway. Note that BMPs, Wnt and *YAP/TAZ* were among the regulated genes. The red color indicates the genes that significantly enriched of the pathways.

many challenges in the transcriptome assembly and related analyses [16, 17]. In this study, we used SMRT sequencing with assistance of RNA-seq to provide a comprehensive transcriptome of *M. fissipes*. The average length and N50 length of full length transcripts, 1599 and 1956 bases, respectively, were much longer than corresponding values for *de novo* assembled transcripts of (624 and 769 bases, respectively) in this study or those reported for the same

species from a previous study (732 and 1330 bases, respectively) [18]. In addition, the percentages of annotated transcripts from our SMRT sequencing analyses were also much higher than those reported earlier based on short-read RNA-Seq analysis. For example, 69% of the full length transcripts were annotated in NR database in our study compared to only 24.77% in the previous study [18]. Thus, SMRT sequencing with assistance of RNA-seq is more advantageous in obtaining real transcriptome transcripts for global gene expression studies.

Making use of the *M. fissipes* transcriptome thus obtained, we analyzed the gene expression changes during tail development from S18 to S28. S18 is the tail determination stage when just a small tailbud is present at the posterior end of the body. S19 to S28 are the tailbud outgrowth stages with the fin, muscle vascular, etc., being formed during this period. We discovered that there were 48,826 expressed transcripts and 8807 DETs based on pair-wise comparisons of gene expression among these four stages. These DETs should be valuable resource for study gene regulation and function during frog tail development. Indeed, our GO analysis of the DETs upregulated between S18 and S19 revealed that amino acid metabolic processes are highly enriched. Amino acid metabolism, such as BCAAs (branched-chain amino acids, leucine, isoleucine and valine) and tryptophan metabolism (Fig. 5A), is likely critical for building muscle tissue by participating in increased protein synthesis as the tail transitions from tailbud determination to outgrowth. Previous studies have shown that the metabolites of BCAAs play an important role in protein synthesis, such as branched-chain keto acids (BCKA),  $\beta$ -aminoisobutyric acid (BAIBA),  $\beta$ -hydroxy- $\beta$ methylbutyrate (HMB) and glutamine [19]. For example, HMB (the metabolite of leucine) stimulates protein synthesis through upregulation of mTOR signaling pathway, which is a central regulator of mammalian cell growth, proliferation and migration, and HMB is much more effective than leucine in increasing protein synthesis through the mTOR system in rat L6 myotubes [20]. In addition,  $\beta$ aminoisobutyric acid (BAIBA), the metabolite of valine, that is increased by exercise, is secreted by the skeletal muscle, and acts on other tissues, such as white adipose tissue, to increase energy expenditure. Furtherly, BCAA metabolism and proteolysis can contribute to changes in BCAA concentration. Thus, catabolism and balance of BCAAs are closely associated with development, body health and disease [21, 22]. In addition, tryptophan and its metabolites such as serotonin, melatonin, kynurenine or quinolinic acid, are known to be important for fetal development and immune regulation during pregnancy [23]. Additionally, tryptophan may modulate gene expression and nutrient metabolism to impact whole-body homeostasis in organisms [24]. Our data suggests these four essential amino acids and their metabolic processes are involved in tail development, particularly in tail transitions from tailbud determination to outgrowth. Interestingly, the DETs upregulated at S28 are highly enriched in "myosin complex", "potassium channel activity" and fructose 1,6-bisphosphate 1-phosphatase activity, which are important for muscle contraction as embryogenesis is completed by S28 and tail begins to function. Notably, fructose 1,6 bisphosphates is an gluconeogenesis enzyme, suggesting that gluconeogenesis is involved in tail development, which may help to remove the excess lactate formed by Glycolysis when tail movement since the accumulation of lactate may occur lactic acidosis and deteriorate pyruvate utilization, and to convert lactate for the synthesis of glucose for tail utilization when tail function is required for tadpole movement [25].

While gene activation typically implicates a positive role in the subsequent developmental events, gene repression suggests that the corresponding genes may no longer be needed for and/or inhibit the subsequent development. In this regard, two downregulated DET clusters, clusters 4 and 11, are of particular interest. The DETs in these clusters are enriched with genes in "organ development" GO terms, which may be expected as tail development approaches completion between S21–S28. In addition, pathway analysis have revealed the enrichment of "Cell cycle" and "Hippo signaling pathway", which are known to be important for the organ development [26-30] and downregulation of genes in these pathways would be expected as tail development is completed by S28. Such a correlation also exists at individual gene level. For example, among the DETs include *Skp2*, *Mps1* and *Cdh1*, which are genes in Cell Cycle pathway (Fig. 7A). Skp2 positively regulates the G1-S transition and cell proliferation [31], Mps1 is a kinase required for mitotic checkpoint and Cdh1 regulates mitosis exit and G1-phase length [32, 33]. These genes are highly expressed at S18 and S19 when the tailbud starts to elongate, a process requiring extensive cell proliferation, and are then downregulated by S28 when tail development is complete.

A number of earlier studies have suggested that organ/tissue regeneration process is a part of the animal development program, especially for the re-growth phase, and thus there are various of shared genes and pathways between regeneration and normal development [34–36]. For instance, genes associated with Notch signaling, Hippo signaling, and Msx genes and Hox genes are expressed similarly during development and regeneration [35, 37]. In addition, both embryonic development and regeneration involve extensive cell proliferation, differentiation, largescale cell movement and organismal patterning. Hence, for a long time, many researchers have been comparing embryogenesis and regeneration with an intention to use knowledge on embryonic development to unlock hidden regenerative potential and eventually aid in the development of regenerative medicines [38]. Interestingly, our study has revealed here numerous genes and pathways involved in tail development, such as Hippo signaling pathway, *smad1/5/8*,

*Id*, *Hes1/5*, *Mps1*, *Slug* and *HDAC*, etc., reported to be important for regeneration [32, 39–45]. *Slug* in Hippo signaling pathway has been reported to modulate mammary epithelial stem cell activity and differentiation *in vivo* and the absence of *Slug* lead to the loss of mammary stem cell activity necessary for tissue regeneration and cancer initiation [44]. Thus, the genes/pathways that we have discovered to be significantly regulated during tail development should be a valuable resource for future research on the mechanisms of both tail development and regeneration.

### 6. Conclusions

In summary, we have obtained 48,826 highquality and non-redundant full-length expressed transcripts in *Microhyla fissipes* and discovered 8807 DETs. Our results provide a global overview of gene expression patterns and new insights into the mechanisms involved in anuran tail development. As organ/tissue regeneration bears similarities to organ/tissue development, our findings here would also help to shed light on tail regeneration, and have implications in research and development of regenerative medicine.

### 7. Author contributions

SW, LL, and JJ conceived and designed the experiment; SW and Y-BS analyzed the data and prepared the manuscript. All authors participated in the manuscript preparation and approve the final version of the manuscript.

### 8. Ethics approval and consent to participate

All experimental protocols and procedures were approved by the Experimental Animal Use Ethics Committee of the Chengdu Institute of Biology (Permit Number: 20150121003).

#### 9. Acknowledgment

Not applicable.

#### 10. Funding

This work was supported in part by the intramural Research Program of NICHD, NIH, Important Research Project of Chinese Academy of Sciences (KJZG-EW-L13), the National Natural Science Foundation of China (Grant No.32000311), Sichuan Science and Technology Project (2017JY0339), and Construction of Basic Conditions Platform of Sichuan Science and Technology Department (2019JDPT0020).

### **11. Conflict of interest**

The authors declare no conflict of interest.

### 12. Availability of data and materials

The sequencing data have been deposited in NCBI Sequence Read Archive (SRA) and the accession number is PRJNA504611. All data generated or analyzed during this study are included in this published article.

### 13. References

- [1] Ishizuya-Oka A, Hasebe T, Shi Y. Apoptosis in amphibian organs during metamorphosis. Apoptosis. 2010; 15: 350–364.
- [2] Nakajima K, Yaoita Y. Dual mechanisms governing muscle cell death in tadpole tail during amphibian metamorphosis. Developmental Dynamics. 2003; 227: 246–255.
- [3] Hoff KV, Wassersug RJ. Tadpole Locomotion: Axial Movement and Tail Functions in a Largely Vertebraeless Vertebrate. American Zoologist. 2000; 40: 62–76.
- [4] Beck CW. Development of the vertebrate tailbud. Wiley Interdisciplinary Reviews Developmental Biology. 2015; 4: 33–44.
- [5] Beck CW, Slack JM. A developmental pathway controlling outgrowth of the *Xenopus* tail bud. Development. 1999; 126: 1611– 1620.
- [6] Beck CW, Slack JM. Analysis of the developing *Xenopus* tail bud reveals separate phases of gene expression during determination and outgrowth. Mechanisms of Development. 1998; 72: 41–52.
- [7] Wang S, Liu L, Liu J, Zhu W, Tanizaki Y, Fu L, et al. Gene Expression Program Underlying Tail Resorption during Thyroid Hormone-Dependent Metamorphosis of the Ornamented Pygmy Frog *Microhyla fissipes*. Frontiers in Endocrinology. 2019; 10: 11.
- [8] Liu LS, Zhao LY, Wang SH, Jiang JP. Research proceedings on amphibian model organisms. Zoological Research. 2016; 37: 237–245.
- [9] Wang SH, Zhao LY, Liu LS, Yang DW, Khatiwada JR, Wang B, et al. A Complete Embryonic Developmental Table of *Microhyla fissipes* (Amphibia, Anura, Microhylidae). Asian Herpetological Research. 2017; 8: 108–117.
- [10] Hackl T, Hedrich R, Schultz J, Förster F. Proovread: largescale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014; 30: 3004–3011.
- [11] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323.
- [12] Lindinger MI, Sjøgaard G. Potassium regulation during exercise and recovery. Sports Medicine. 1991; 11: 382–401.
- [13] Kristensen M, Hansen T, Juel C. Membrane proteins involved in potassium shifts during muscle activity and fatigue. American Journal of Physiology. Regulatory, Integrative and Comparative Physiology. 2006; 290: R766–R772.
- [14] Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, *et al*. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature Biotechnology. 2012; 30: 693–700.
- [15] Song H, Yu Z, Sun L, Gao Y, Zhang T, Wang H. De novo transcriptome sequencing and analysis of *Rapana venosa* from six different developmental stages using Hi-seq 2500. Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics. 2016; 17: 48–57.
- [16] Mahajan S, Wei KH, Nalley MJ, Gibilisco L, Bachtrog D. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. PLoS Biology. 2018; 16: e2006348.

1012

- [17] Feng X, Jia Y, Zhu R, Chen K, Chen Y. Characterization and analysis of the transcriptome in Gymnocypris selincuoensis on the Qinghai-Tibetan Plateau using single-molecule long-read sequencing and RNA-seq. DNA Research. 2019; 26: 353–363.
- [18] Zhao L, Liu L, Wang S, Wang H, Jiang J. Transcriptome profiles of metamorphosis in the ornamented pygmy frog *Microhyla fissipes* clarify the functions of thyroid hormone receptors in metamorphosis. Scientific Reports. 2016; 6: 27310.
- [19] Nie C, He T, Zhang W, Zhang G, Ma X. Branched Chain Amino Acids: Beyond Nutrition Metabolism. International Journal of Molecular Sciences. 2018; 19: 954.
- [20] Holeček M. Beta-hydroxy-beta-methylbutyrate supplementation and skeletal muscle in healthy and muscle-wasting conditions. Journal of Cachexia, Sarcopenia and Muscle. 2017; 8: 529–541.
- [21] Kamei Y, Hatazawa Y, Uchitomi R, Yoshimura R, Miura S. Regulation of Skeletal Muscle Function by Amino Acids. Nutrients. 2020; 12: 261.
- [22] Ten Have GAM, Jansen L, Schooneman MG, Engelen M, Deutz NEP. Metabolic flux analysis of branched-chain amino and keto acids (BCAA, BCKA) and beta-hydroxy beta-methylbutyric acid across multiple organs in the pig. American Journal of Physiology-Endocrinology and Metabolism. 2021; 320: E629– E640.
- [23] Karahoda R, Abad C, Horackova H, Kastner P, Zaugg J, Cerveny L, *et al.* Dynamics of Tryptophan Metabolic Pathways in Human Placenta and Placental-Derived Cells: Effect of Gestation Age and Trophoblast Differentiation. Frontiers in Cell and Developmental Biology. 2020; 8: 574034.
- [24] Yao K, Fang J, Yin Y, Feng Z, Tang Z, Wu G. Tryptophan metabolism in animals: important roles in nutrition and health. Frontiers in Bioscience. 2011; 3: 286–297.
- [25] Adeva-Andany M, López-Ojén M, Funcasta-Calderón R, Ameneiros-Rodríguez E, Donapetry-García C, Vila-Altesor M, *et al.* Comprehensive review on lactate metabolism in human health. Mitochondrion. 2014; 17: 76–100.
- [26] Awad MM, Gruppuso PA. Cell cycle control during liver development in the rat: evidence indicating a role for cyclin D1 posttranscriptional regulation. Cell Growth & Differentiation. 2000; 11: 325–334.
- [27] Ryan AF. The cell cycle and the development and regeneration of hair cells. Current Topics in Developmental Biology. 2003; 57: 449–466.
- [28] Ahuja P, Sdek P, MacLellan WR. Cardiac myocyte cell cycle control in development, disease, and regeneration. Physiological Reviews. 2007; 87: 521–544.
- [29] Zeng Q, Hong W. The emerging role of the hippo pathway in cell contact inhibition, organ size control, and cancer development in mammals. Cancer Cell. 2008; 13: 188–192.
- [30] Fu V, Plouffe SW, Guan K. The Hippo pathway in organ development, homeostasis, and regeneration. Current Opinion in Cell Biology. 2017; 49: 99–107.
- [31] Wirbelauer C, Sutterlüty H, Blondel M, Gstaiger M, Peter M, Reymond F, *et al.* The F-box protein Skp2 is a ubiquitylation target of a Cul1-based core ubiquitin ligase complex: evidence for a role of Cul1 in the suppression of Skp2 expression in quiescent fibroblasts. The EMBO Journal. 2000; 19: 5362–5375.
- [32] Poss KD, Nechiporuk A, Hillam AM, Johnson SL, Keating MT. Mps1 defines a proximal blastemal proliferative compartment essential for zebrafish fin regeneration. Development. 2002; 129: 5141–5149.
- [33] Delgado-Esteban M, García-Higuera I, Maestre C, Moreno S, Almeida A. APC/C-Cdh1 coordinates neurogenesis and cortical size during development. Nature Communications. 2013; 4: 2879.
- [34] Gardiner DM, Endo T, Bryant SV. The molecular basis of amphibian limb regeneration: integrating the old with the new. Seminars in Cell & Amp; Developmental Biology. 2002; 13: 345–352.

- [35] Chen Y, Love NR, Amaya E. Tadpole tail regeneration in *Xeno-pus*. Biochemical Society Transactions. 2014; 42: 617–623.
- [36] Goldman JA, Poss KD. Gene regulatory programmes of tissue regeneration. Nature Reviews Genetics. 2020; 21: 511–525.
- [37] Barker DM, Beck CW. Xenopus as an Emerging Model for the Study of Regenerative Mechanisms. Developmental Dynamics. 2009; 238: 1366–1378.
- [38] Warner JF, Guerlais V, Amiel AR, Johnston H, Nedoncelle K, Rottinger E. NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone *Nematostella vectensis*. Development. 2018; 145: dev162867.
- [39] Moya IM, Halder G. Hippo–YAPTAZ signalling in organ regeneration and regenerative medicine. Nature Reviews Molecular Cell Biology. 2019; 20: 211–226.
- [40] Zheng M, Jacob J, Hung SH, Wang J. The Hippo Pathway in Cardiac Regeneration and Homeostasis: New Perspectives for Cell-Free Therapy in the Injured Heart. Biomolecules. 2020; 10: 1024.
- [41] Tseng A, Carneiro K, Lemire JM, Levin M. HDAC activity is required during *Xenopus* tail regeneration. PLoS ONE. 2011; 6: e26382.
- [42] Voss SR, Ponomareva LV, Dwaraka VB, Pardue KE, Baddar NWAH, Rodgers AK, *et al.* HDAC Regulates Transcription at the Outset of Axolotl Tail Regeneration. Scientific Reports. 2019; 9: 6751.
- [43] Suen WJ, Li ST, Yang LT. Hes1 regulates anagen initiation and hair follicle regeneration through modulation of hedgehog signaling. Stem Cells. 2020; 38: 301–314.
- [44] Phillips S, Prat A, Sedic M, Proia T, Wronski A, Mazumdar S, *et al*. Cell-state transitions regulated by SLUG are critical for tissue regeneration and tumor initiation. Stem Cell Reports. 2014; 2: 633–647.
- [45] Pentagna N, Pinheiro da Costa T, Soares Dos Santos Cardoso F, Martins de Almeida F, Blanco Martinez AM, Abreu JG, *et al.* Epigenetic control of myeloid cells behavior by Histone Deacetylase activity (HDAC) during tissue and organ regeneration in *Xenopus laevis*. Developmental & Comparative Immunology. 2021; 114: 103840.

**Supplementary material:** Supplementary material associated with this article can be found, in the online version, at https://www.fbscience.com/Landmark/articles/10. 52586/5004.

**Abbreviations:** DETs, differentially expressed transcripts; SMRT, single molecule long-read sequencing; BCAA, branched chain amino; GO, Gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; RSEM, RNA-Seq using expectation maximization; FPKM, fragments per kilobase of exon model per million mapped reads; BP, Biological process; CC, Cellar component; MF, Molecular function; DAG, Directed acyclic graph.

**Keywords:** SMRT sequencing; RNA-Seq; Regeneration; Tail development; *Microhyla fissipes* 

#### Send correspondence to:

Yun-Bo Shi, Section on Molecular Morphogenesis, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, Maryland, MD 20892, USA, E-mail: Shi@helix.nih.gov

Jianping Jiang, Chengdu Institute of Biology, Chinese Academy of Sciences, 610041 Chengdu, Sichuan, China, E-mail: jiangjp@cib.ac.cn