

SCIENTIFIC REPORTS



OPEN

Developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function of Dojo loach, *Misgurnus anguillicaudatus*

Weiwei Luo¹, Xiaojuan Cao^{1,2}, Xiuwen Xu¹, Songqian Huang¹, Chuanshu Liu¹ & Tea Tomljanovic³

Received: 10 May 2016
Accepted: 26 July 2016
Published: 22 August 2016

Dojo loach, *Misgurnus anguillicaudatus* is a freshwater fish species of the loach family Cobitidae, using its posterior intestine as an accessory air-breathing organ. Little is known about the molecular regulatory mechanisms in the formation of intestinal air-breathing function of *M. anguillicaudatus*. Here high-throughput sequencing of mRNAs was performed from six developmental stages of posterior intestine of *M. anguillicaudatus*: 4-Dph (days post hatch) group, 8-Dph group, 12-Dph group, 20-Dph group, 40-Dph group and Oyd (one-year-old) group. These six libraries were assembled into 81300 unigenes. Totally 40757 unigenes were annotated. Subsequently, 35291 differentially expressed genes (DEGs) were scanned among different developmental stages and clustered into 20 gene expression profiles. Finally, 15 key pathways and 25 key genes were mined, providing potential targets for candidate gene selection involved in formation of intestinal air-breathing function in *M. anguillicaudatus*. This is the first report of developmental transcriptome of posterior intestine in *M. anguillicaudatus*, offering a substantial contribution to the sequence resources for this species and providing a deep insight into the formation mechanism of its intestinal air-breathing function. This report demonstrates that *M. anguillicaudatus* is a good model for studies to identify and characterize the molecular basis of accessory air-breathing organ development in fish.

Dojo loach, *Misgurnus anguillicaudatus*, is a freshwater fish species of the loach family Cobitidae¹. The species is native to East Asia, but is also popular as an aquarium fish and has been introduced elsewhere in Asia and to Europe and North America². The loach inhabits mud, ponds, and rice fields, which are subject to periodic drying³. During drying, loaches burrow under the mud. In such poor dissolved oxygen conditions, their respiratory system has been the focus of much interest. The loaches have three forms of respiration: branchial, cutaneous and intestinal. The first two types are common to other fish and amphibian. Intestinal respiration, however, is found in Cobitidae⁴. The intestine of loaches can be divided into anterior, mid, and posterior sections. The anterior intestine and middle intestine are adapted to serve major role of digestion and absorption. The posterior intestine is highly modified, being well vascularized with intraepithelial capillaries, which makes it well suited for gas exchange⁵. Studies of intestinal air-breathing of the loach have focused on histological description of the intestine and physiological aspect of gas exchange^{1,3}. Little is known about the molecular regulatory mechanisms in occurrence and formation of intestinal air-breathing function of *M. anguillicaudatus*.

With the advantage of low cost and high speed, massively parallel sequencing (Illumina) RNA-Seq analysis is now the most convenient method to find out new genes and investigate gene expression patterns of non-model

¹College of Fisheries, Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 437000, Hubei, People's Republic of China. ²Freshwater Aquaculture Collaborative Innovation Center of Hubei Province, Hubei, People's Republic of China. ³Department for Fisheries, Beekeeping, Game management and Special Zoology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia. Correspondence and requests for materials should be addressed to X.C. (email: caoxiaojuan@mail.hzau.edu.cn)

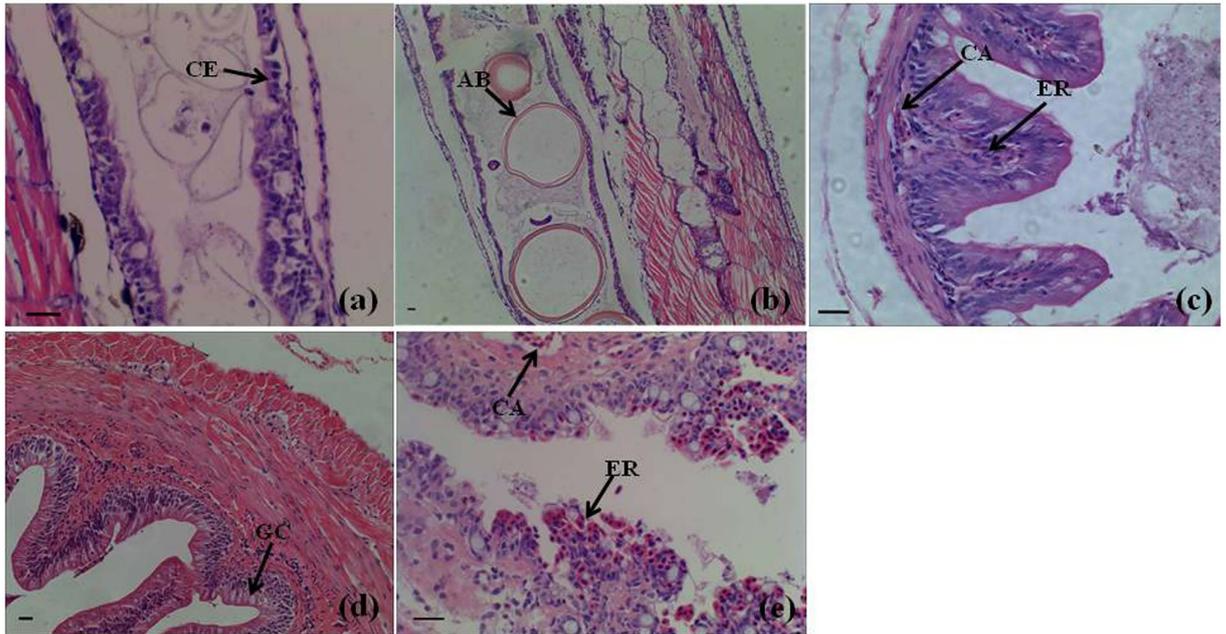


Figure 1. Histological observations of intestine tissues of *M. anguillicaudatus* during different developmental stages. (a) Longitudinal section of intestine from 8-Dph loach. (b) Longitudinal section of intestine from 10-Dph loach. (c) Cross section of posterior intestine from 40-Dph loach. (d) Cross section of anterior intestine from Oyd. (e) Cross section of posterior intestine from Oyd. Abbreviations: air bubble (AB); capillary vessel (CA); simple columnar epithelium (CE); erythrocyte (ER); goblet cell (GC). Bar indicated 20 μm .

organisms, especially for species of which the whole genome sequence is not yet available⁶. To date, several studies have reported on the gene expression profiles of different intestine tissues of fishes like Atlantic cod⁷, tilapia⁸ and zebrafish⁹. However, studies about the dynamic expression profiles on the intestine of *M. anguillicaudatus* remain scarce. Here we used high-throughput sequencing of mRNAs from posterior intestine of loaches at different developmental stages to research the genes involved in formation of intestinal air-breathing function of *M. anguillicaudatus*. We constructed six libraries including 4-Dph (days post hatch) group, 8-Dph group, 12-Dph group, 20-Dph group, 40-Dph group and Oyd (one year old) group. Hence, the major objective of this study was to investigate genes related to the formation of intestinal air-breathing function of *M. anguillicaudatus* using high-throughput sequencing and bioinformatic analysis.

Results

Histological observations of intestine tissues in *M. anguillicaudatus* during different developmental stages. The newly hatched loach larva had a long straight intestinal tube with a very simple structure. In order to uncover the relationship between the occurrence and formation of intestinal air-breathing function and the changes of intestinal structure, microstructures of intestine tissues in loach during different developmental stages were observed by using Delafield's haematoxylin, counterstained with eosin (HE) staining. Intestinal epithelium of 8-Dph loach was simple columnar epithelium (Fig. 1a). Intestinal lumen of 10-Dph loach showed numerous air bubbles (Fig. 1b). The numerous erythrocytes and capillary vessels were showed in posterior intestine of 40-Dph loach (Fig. 1c). Histological observations of posterior intestines suggested the initial formation period of intestinal air-breathing function of loach began at around 10 days post hatch. With the development of posterior intestine of loach, especially adults, the obvious structural differences between anterior intestine and posterior intestine were present (Fig. 1d,e). Mucosal folds of anterior intestine of Oyd loach exhibited a more complicated, interconnecting arrangement, and were more extensive than those in the posterior intestine of Oyd loach. The anterior intestine of Oyd loach possessed the typical characteristics of an absorptive intestinal epithelium composed of columnar epithelial cells with abundant goblet cells and a brush border (Fig. 1d). While, the typical characteristics of posterior intestine of Oyd loach was abundant in blood capillaries within the mucosal epithelium, which made it well suited for gas exchange (Fig. 1e).

Transcriptome sequencing and *de novo* assembly. Approximately 38.35 million, 40.18 million, 36.59 million, 37.79 million, 32.89 million and 58.15 million raw reads were sequenced using RNA-Seq technique in 4-Dph, 8-Dph, 12-Dph, 20-Dph, 40-Dph and Oyd libraries, respectively (Table 1). These raw reads were further filtered by removing adaptors, ambiguous nucleotides and low-quality sequences to generate clean reads. After trimming the raw reads, there were 37.79 million, 39.66 million, 36.03 million, 37.28 million, 32.45 million and 48.37 million clean reads generated from 4-Dph, 8-Dph, 12-Dph, 20-Dph, 40-Dph and Oyd, respectively (Table 1).

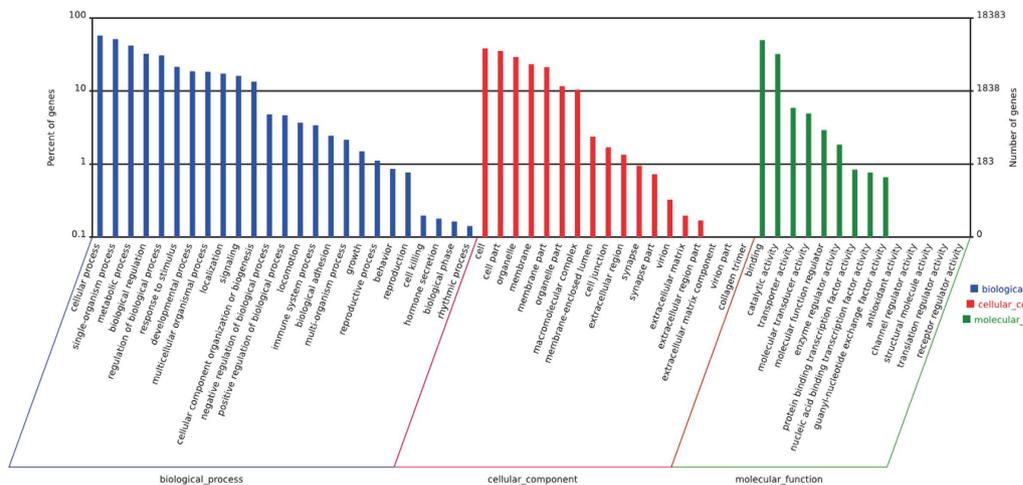


Figure 2. GO classifications of all unigenes from posterior intestine of *M. anguillicaudatus*.

Samples	4-Dph	8-Dph	12-Dph	20-Dph	40-Dph	Oyd
Raw read number	38347890	40177226	36590380	37786998	32886306	58152842
Clean read number	37787284	39660682	36031650	37275524	32448448	48367330
Raw read base pairs	4793486250	5022153250	4573797500	4723374750	4110788250	5873437042
Clean read base pairs	4723410500	4957585250	4503956250	4659440500	4056056000	5136684000
Q20%	97.41%	97.67%	97.50%	97.50%	97.67%	97.74%
Q30%	94.64%	95.14%	94.83%	94.81%	95.18%	91.79%
GC%	48.42%	48.33%	47.43%	47.22%	46.58%	46.98%

Table 1. Overview of the sequencing reads. Q20: The percentage of bases with quality value larger than 20. Q30: The percentage of bases with quality value larger than 30. GC%: The percentage of proportion of guanidine and cytosine nucleotides among total nucleotides.

The clean reads obtained from the six different transcriptome libraries were pooled and assembled full-length transcripts without reference genomes by Trinity software. After the elimination of redundant transcripts, 81300 unigenes were achieved with an average length of 879.78 base pairs and N50 of 1662 bp (Supplementary Table S1). The size distribution of the unigenes is shown in Supplementary Fig. S1. The all unigenes provided a sequence basis for analysis of gene expression in posterior intestine of loach during different developmental stages.

Unigene annotation to public databases. In order to maximize the information of assembled unigenes, all of the 81300 unigene sequences were searched against four public databases: NCBI non-redundant protein (Nr) database, Swiss-Prot protein database (Swissprot), euKaryotic Ortholog Groups (KOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Among these unigenes, 40563 (49.89%), 31588 (38.85%), 24837 (30.55%), 17077 (21.00%) of the unigenes were identified in the Nr, Swissprot, KOG and KEGG database, respectively (Supplementary Table S2). Totally 40757 unigenes (50.13% of all the 81300 unigene sequences) were annotated in the four public databases. Among all the 40757 annotated unigenes, 14102 (27.98%) were simultaneously annotated by the four databases, and 32747 (80.35%) showed homology to the known sequences deposited in at least one database (Supplementary Fig. S2).

Functional annotation and classification of unigenes. GO assignment was performed to classify functions of the predicted loach genes. Based on sequence homology, annotated unigenes were classified into three major functional categories: biological process, cellular component and molecular function (Fig. 2). Under the biological process category, the major GO terms were cellular process (10512 unigenes), single-organism process (9414 unigenes), metabolic process (7709 unigenes), biological regulation (5919 unigenes), regulation of biological process (5634 unigenes), response to stimulus (3938 unigenes), developmental process (3417 unigenes) and multicellular organismal process (3365 unigenes). Within the cellular component category, a significant percentage of genes were clustered into cell (7023 unigenes), cell part (6520 unigenes), organelle (5358 unigenes), membrane (4274 unigenes) and membrane part (3892 unigenes). In the molecular function category, most genes were assigned to binding (9144 unigenes) and catalytic activity (5869 unigenes) followed by transporter activity (1081 unigenes), molecular transducer activity (907 unigenes) and molecular function regulator (537 unigenes).

To classify orthologous proteins, the assembled unigenes were compared against KOG. 24837 unigenes were clustered into 25 categories (Fig. 3). Among them, the cluster for signal transduction mechanisms represented the

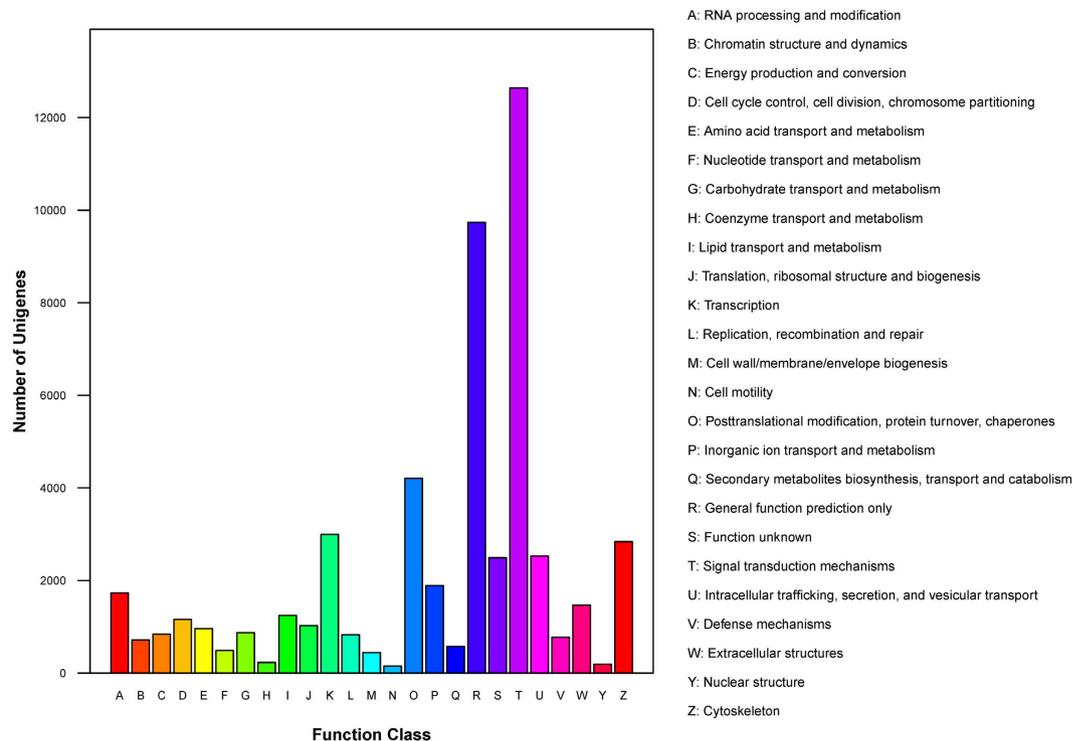


Figure 3. KOG classifications of all unigenes from posterior intestine of *M. anguillicaudatus*.

largest group (12641 unigenes, 50.90%), followed by general function prediction only (9738 unigenes, 39.21%), posttranslational modification, protein turnover, chaperones (4209 unigenes, 16.95%), transcription (2993 unigenes, 12.05%), cytoskeleton (2841 unigenes, 11.44%), intracellular trafficking, secretion, vesicular transport (2530 unigenes, 10.19%), function unknown (2491 unigenes, 10.03%), inorganic ion transport and metabolism (1893 unigenes, 7.62%), RNA processing and modification (1732 unigenes, 6.97%) and extracellular structures (1466 unigenes, 5.90%).

KEGG pathway-based analyses help to identify the biological pathways that are related to unigenes. Totally, 17077 (41.90%) out of 40757 were assigned 240 KEGG pathways. Among these unigenes, top ten enrichment of KEGG pathway were metabolic pathways (ko01100, 1799 unigenes), pathways in cancer (ko05200, 763 unigenes), regulation of actin cytoskeleton (ko04810, 622 unigenes), neuroactive ligand-receptor interaction (ko04080, 616 unigenes), MAPK signaling pathway (ko04010, 616 unigenes), dilated cardiomyopathy (ko05414, 582 unigenes), hypertrophic cardiomyopathy (HCM) (ko05410, 563 unigenes), focal adhesion (ko04510, 546 unigenes) and calcium signaling pathway (ko04020, 524 unigenes).

Identification and functional analysis of differentially expressed genes. The DEGs during the course of posterior intestine development were explored by using DEGseq with the criteria of $|\log_2\text{Ratio}| \geq 1$ and $\text{FDR} \leq 0.01$. In total, the number of DEGs between two consecutive stages were as follows: 10279 between stages Dph-4 and Dph-8 (23.94% up- and 76.06% down-regulated in stage Dph-8), 9875 between stages Dph-8 and Dph-12 (74.10% up- and 25.90% down-regulated in stage Dph-12), 7107 between stages Dph-12 and Dph-20 (71.34% up- and 28.66% down-regulated in stage Dph-20), 11130 between stages Dph-20 and Dph-40 (69.20% up- and 30.80% down-regulated in stage Dph-40), and 21461 between stages Dph-40 and Oyd (53.04% up- and 46.96% down-regulated in stage Oyd). Details on the number of DEGs at each two consecutive time points are presented in Supplementary Table S3.

To better understand the dynamic changes of gene expression in posterior intestine during all the six developmental stages, further analyses of the DEGs were also performed between the two stages of Dph-4-VS-Dph-8, Dph-4-VS-Dph-12, Dph-4-VS-Dph-20, Dph-4-VS-Dph-40 and Dph-4-VS-Oyd, respectively (Supplementary Fig. S3). Finally, a total of 35291 genes were identified as DEGs among the six developmental stages.

To investigate the biological events that the DEGs mainly involved during posterior intestine development, GO term enrichment analysis were conducted (Supplementary Fig. S4). Under the biological process category, the major GO terms were cellular process (5603 unigenes), single-organism process (5064 unigenes), metabolic process (4190 unigenes), biological regulation (3175 unigenes), regulation of biological process (3021 unigenes), response to stimulus (2122 unigenes), developmental process (1862 unigenes) and multicellular organismal process (1810 unigenes). In the cellular component category, a significant percentage of genes were clustered into cell (3700 unigenes), cell part (3424 unigenes), organelle (2817 unigenes), membrane (2132 unigenes) and membrane part (1937 unigenes). Within the molecular function category, most genes were assigned to binding (4863 unigenes) and catalytic activity (3261 unigenes), followed by transporter activity (483 unigenes), molecular transducer activity (453 unigenes) and molecular function regulator (304 unigenes). These suggested that

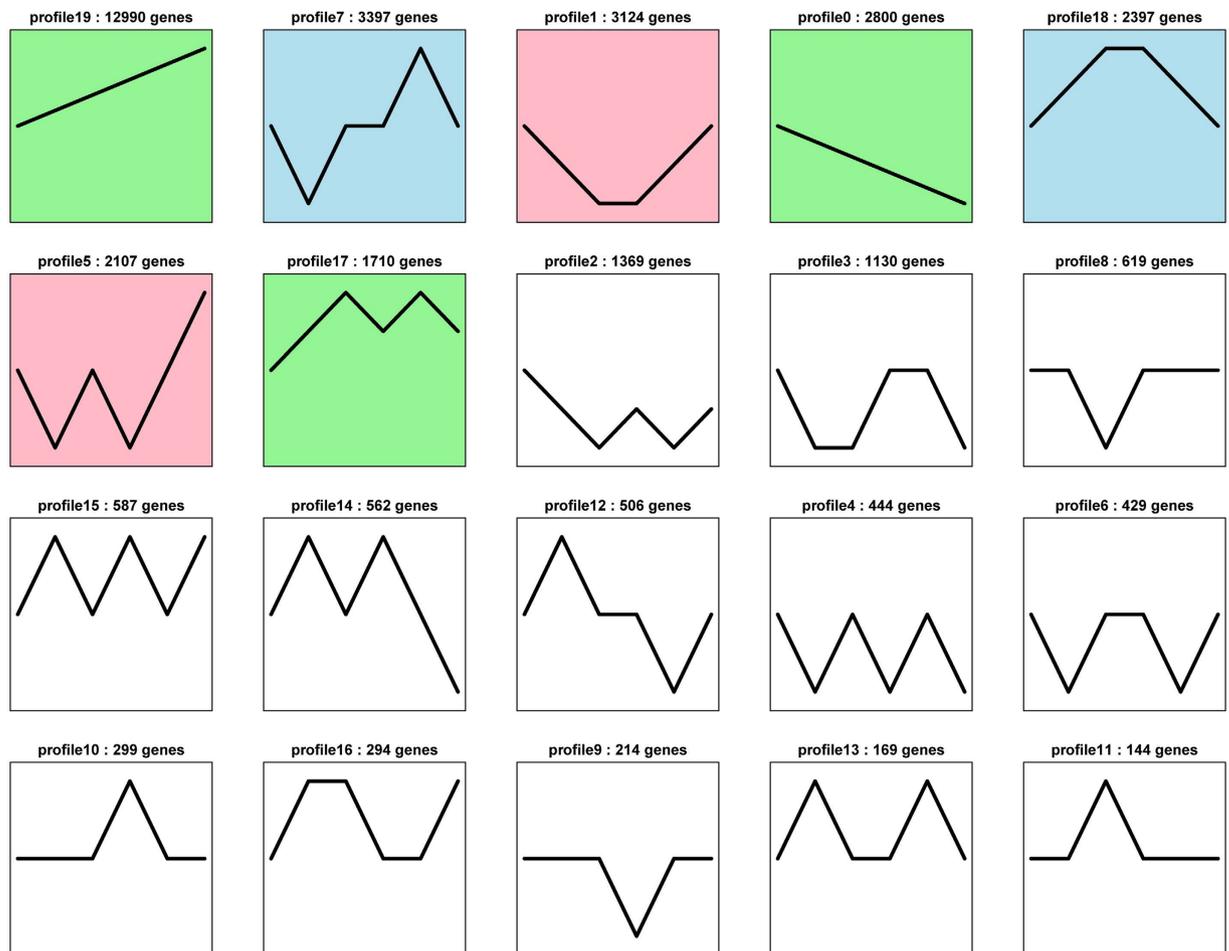


Figure 4. Expression profiles and clusters of DEGs obtained from the STEM clustering. Numbers indicated profiles or gene numbers. Significantly different profiles were represented by different background colors.

gene regulation was crucial for achieving various physiological functions associated to the development of loach posterior intestine.

KEGG pathway annotation enabled us to assign 35291 DEGs to 238 pathways. Supplementary Fig. S5 showed the top 20 significantly enriched KO pathways. The result revealed significant enrichment for 20 pathways with 17 metabolism pathways, two genetic information processing pathways and one organismal system pathway. These observations disclosed the vital implications of amino acid metabolism, lipid metabolism, metabolism of cofactors and vitamins, genetic information processing and organismal system in developmental processes of loach posterior intestine.

Dynamic expression profiles of DEGs. To get dynamic expression patterns of DEGs during development, Short Time-series Expression Miner (STEM) software was performed to classify all the DEGs according to their abundance changes. The 35291 DEGs were classified into 20 clusters according to their expression patterns (Fig. 4). Seven significant expression profiles (profile 0, profile 1, profile 5, profile 7, profile 17, profile 18 and profile 19) were identified. As shown in Fig. 4, significantly different profiles were represented by different background colors. Respectively, 2800, 3124, 2107, 3397, 1710, 2397 and 12990 co-expressed DEGs were identified for profile 0, profile 1, profile 5, profile 7, profile 17, profile 18 and profile 19. Of these, the most abundant group was profile 19, with 12990 genes whose expression showed a positive slope, which was reversed to the trend in profile 0. Interestingly, reversed trends were also observed between profile 1 and profile 18, profile 5 and profile 17, respectively. In order to obtain a glimpse of the biological functions of these 7 clusters of genes, GO enrichment analysis was conducted. The genes co-expressed in profile 0, profile 1, profile 5, profile 7, profile 17, profile 18 and profile 19, correspondingly participated in 1470, 1897, 1222, 760, 386, 801 and 2368 biological functions (GO terms).

Profile 18 and profile 19 were the two represented patterns. Profile 18 firstly increased, and reached to top in Dph-12 and Dph-20, and then decreased. It was consistent with the trend of occurred and formed process of intestinal air-breathing function in the loach. The GO functional analysis of profile 18 showed: under the biological process category, the significantly representatives included anatomical structure development, single-organism developmental process, anatomical structure morphogenesis, single-organism cellular process and multiple processes related to development; Within the cellular component category, a significant percentage

#Pathway	Pathway ID
Regulation of actin cytoskeleton	Ko04810
MAPK signaling pathway	Ko04010
VEGF signaling pathway	Ko04370
Notch signaling pathway	Ko04330
Vascular smooth muscle contraction	Ko04270
Wnt signaling pathway	Ko04310
Insulin signaling pathway	Ko04910
ECM-receptor interaction	Ko04512
Jak-STAT signaling pathway	Ko04630
TGF-beta signaling pathway	Ko04350
ErbB signaling pathway	Ko04012
Oxidative phosphorylation	Ko00190
Hedgehog signaling pathway	Ko04340
Osteoclast differentiation	Ko04380
Cell cycle	Ko04110

Table 2. Key pathways associated with the formation of intestinal air-breathing function of *M. anguillicaudatus*.

of genes were clustered into cell and cell part; In the molecular function category, most genes were assigned to binding, catalytic activity and ion binding. The expression levels of DEGs of profile 19 was constantly increased across posterior intestine development stages, suggesting more and more genes were involved in the occurred and formed process of intestinal air-breathing function. The GO functional analysis of profile 19 showed: in the biological process category, cellular process, single-organism process, biological regulation, regulation of biological process and single-organism cellular process represented the majorities of this category; among cellular component terms, cell and cell part were observed to be the most abundant classes; binding, catalytic activity and molecular transducer activity showed a high percentage of genes in the category of molecular function. It was considered that DEGs of profile 18 and profile 19 were closely involved in formation of intestinal air-breathing function of the loach.

Individual key pathways and DEGs related to formation of loach intestinal air-breathing function.

Fifteen key pathways (Table 2) and twenty-five key DEGs associated with the formation of intestinal air-breathing function in the loach were identified from profile 18 and profile 19. These key pathways were mainly involved in development, angiogenesis and cytoskeleton. Fifty important enzymes of the 15 key pathways were selected by three strategies simultaneously which were considered to be related to the occurred and formed process of intestinal air-breathing function: (i) 25 important enzymes that showed high participation frequency in the 15 key pathways of profile 18 and profile 19; (ii) another 20 important enzymes which appeared in some individual key pathways (10/15 pathways) of both profile 18 and profile 19; (iii) 5 important enzymes which showed in the other key pathways (5/15 pathways) of profile 18 or profile 19. Then, the 50 important enzymes in total obtained from these 15 key pathways of profile 18 and profile 19 constructed the protein-protein interaction (PPI) network by the Search Tool for the Retrieval of Interacting Genes (STRING) (Fig. 5). Based on the connectivity degree of each node, 25 key enzymes counted high connectivity degree of each node were finally identified: gelsolin (GSN), YES proto-oncogene 1 (YES1), cytokine inducible SH2-containing protein (CISH), ras homolog family member A (RHOA), protein kinase C, epsilon (PRKCE), Raf-1 proto-oncogene, serine/threonine kinase (RAF1), BCL2 associated agonist of cell death (BAD), suppressor of cytokine signaling 3 (SOCS3), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), v-akt murine thymoma viral oncogene homolog 1 (AKT1), kinase insert domain receptor (KDR), epidermal growth factor receptor (EGFR), tumor protein p53 (TP53), jun proto-oncogene (JUN), SMAD family member 4 (SMAD4), 3-keto-acyl-CoA thiolase 2 (PKT2), mitogen-activated protein kinase 14 (MAPK14), growth factor receptor bound protein 2 (GRB2), vascular endothelial growth factor A (VEGFA), v-myc avian myelocytomatosis viral oncogene homolog (MYC), tumor necrosis factor (TNF), dishevelled segment polarity protein 2 (DVL2), ROS proto-oncogene 1, receptor tyrosine kinase (ROS1), ets variant 5 (ETV5) and frizzled class receptor 10 (FZD10). Key DEGs encoding these 25 enzymes need further studies to clarify the associations between genes and the formation of intestinal air-breathing function in the loach.

Validation of RNA-Seq results by quantitative real-time RT-PCR (qRT-PCR). To validate the RNA-seq data obtained in the current study, qRT-PCR of 9 randomly selected DEGs were performed, including intraflagellar transport protein 22 homolog (*IFT22*), serine/threonine-protein phosphatase 2A catalytic subunit beta isoform (*PP2AB*), *VEGFAa*, glutamate dehydrogenase (*gdh*), annexin A1a (*anxa1a*), vasoactive intestinal peptide (*VIP*), metallothionein (*Mt*), protein phosphatase 1 (*PPI*), and spondin 1b (*Spon1b*). Results obtained by qRT-PCR were compared to data obtained by RNA-seq (Supplementary Fig. S6). The expression patterns for the 9 DEGs obtained using the two methods were similar, confirming the reliability of the data obtained by RNA-seq.

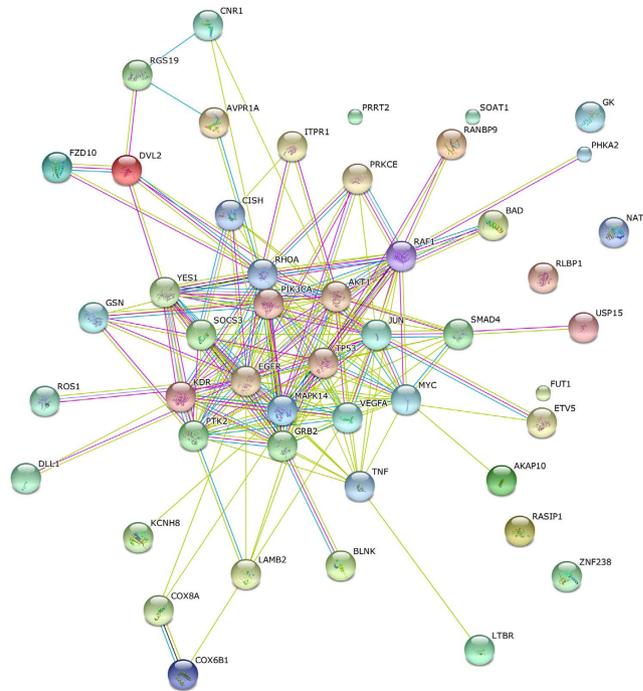


Figure 5. Protein-protein interaction network for the 50 important enzymes encoded by key DEGs of key pathways. Different nodes represented different enzymes. The interactions among these enzymes were represented by different colorful lines.

Discussions

The acquisition of atmospheric O₂ shows important difference from water breathing. *M. anguillicaudatus* can exchange gases both with water, *via* gills and skin, and with air, *via* the posterior region of the intestine¹. The first two types are common to other fish and amphibian. Intestinal respiration, however, is found in Cobitidae and because of this accessory type of respiration, *M. anguillicaudatus* is adapted to the low oxygen tensions (hypoxia) common in its habitat⁴. Besides, *M. anguillicaudatus* is widely distributed in China, Korea, Japan, and other countries in southeastern Asia². Therefore, the loach can be used as an excellent animal for exploration the formation mechanisms of intestinal air-breathing function. Structures of the loach intestine in relation to respiration have been studied many times³. The posterior intestine with a thin muscular wall is rich in erythrocytes and blood vessels, while the anterior intestine possesses a thick muscular wall and abundant mucosal folds⁴, which was consistent with histological observations of the anterior intestine and posterior intestine in this study. And more importantly, in this study, 10-Dph loach showed numerous air bubbles in the intestinal lumen, suggesting the initial formation period of intestinal air-breathing function of loach began at around 10 days post hatch. All the time, scientists have been attracted by structures of the posterior intestine, respiratory zone of loach¹. Very little is known on the development and molecular aspects of the intestinal air-breathing function. The expansion of sequencing technologies has sped up the generation of transcriptomics data even for scarce species⁶. Here, high-throughput sequencing of posterior intestine mRNAs at different development stages, including 4-Dph, 8-Dph (before the formation of intestinal air-breathing function) and 12-Dph, 20-Dph, 40-Dph and Oyd (with intestinal air-breathing function) were used to identify genes involved in formation of intestinal air-breathing function in the loach.

As expected, the significantly enriched GO terms that were identified in the present study were mainly associated with metabolic process and molecular function including binding, catalytic activity and substance transport, which was similar to the findings for gene expression profile in the intestine of sea cucumber¹⁰. As we known, transition from larval to juvenile stage in fish involves development of most organs and tissues, and maturation of different physiological functions¹¹. Therefore, GO enrichment analysis of DEGs during the posterior intestine developmental process (4-Dph group, 8-Dph group, 12-Dph group, 20-Dph group, 40-Dph group and Oyd group) of loach was complicated. DEGs obtained here mapped to the enriched GO term cellular process, single-organism process, biological regulation, developmental process and multicellular organismal process, which was also found in European seabass during its early development¹². These suggested gene regulation was crucial for achieving various physiological functions associated to the development of loach posterior intestine, which was similar to the intestine regeneration process in sea cucumbers since they endured both proliferation and development¹³.

Oxidative phosphorylation¹⁴ can produce energy for the whole development of loach posterior intestine. Wnt signaling is a highly conserved cellular communication system that is involved in a wide variety of developmental and adult processes in all metazoan organisms¹⁵. Its mediated effects are diverse, ranging from proliferation, apoptosis, migration, polarization, to stem cell maintenance and differentiation¹⁶. Notch genes encode transmembrane

receptors that are highly conserved from invertebrates to mammals¹⁷. Notch-mediated signals regulate cell fate decisions in a large number of developmental systems¹⁸. Hedgehog (Hh) proteins are fundamental to animal development and are conserved in species ranging from *Drosophila melanogaster* to humans¹⁹. The mammalian family of Hh proteins includes sonic hedgehog (Shh), desert hedgehog (Dhh), and indian hedgehog (Ihh). These ligands signal through a mechanism involving two transmembrane proteins: patched homolog 1 (Ptch1) and smoothened (Smo)²⁰. Mitogen-activated protein kinases (MAPKs) are serine-threonine protein kinases that play an important role in the regulation of many cellular processes including cell growth and proliferation, differentiation, and apoptosis. MAPKs consist of growth factor-regulated extracellular signal-related kinases (ERKs), and the stress-activated MAPKs, c-jun NH2-terminal kinases (JNKs) and p38 MAPKs²¹. Genes encoding components of TGF- β signaling pathway, including ligands TGF- β 1 and TGF- β 2 as well as receptor TGFBR1, are functionally polymorphic in humans²². TGF- β can regulate such diverse processes as cell proliferation, differentiation, motility, adhesion, organization, and programmed cell death¹⁷. Wnt signaling pathway, Notch signaling pathway, Hedgehog signaling pathway, MAPK signaling pathway and TGF- β signaling pathway are also involved in vasculogenesis and angiogenesis^{17,21}. Besides, the VEGF pathway is recognized as the key regulators of the angiogenic process²³. Previous studies have also presented vascular smooth muscle contraction signaling pathway²⁴, JAK/STAT1 signaling pathway²⁵, insulin signaling pathway²⁶ and ErbB signaling pathway²⁷ played important roles in angiogenesis. To adapt to the development of loach posterior intestine, such as cell size growing, cell proliferation and migration, the pathways associated with cytoskeleton were found. Regulation of actin cytoskeleton signaling pathway²⁸, ECM-receptor interaction²⁹, osteoclast differentiation³⁰ and cell cycle³¹ are mostly associated with the process of cell functions, cell proliferation, differentiation and migration.

To further identify key DEGs which were related to the occurred and formed process of intestinal air-breathing function in loach, the STRING was used for constructing PPI network. Finally, 25 genes were identified: *GSN*³², *YES1*³³, *CISH*³⁴, *RHOA*³⁵, *PRKCE*³⁶, *RAF1*³⁷, *BAD*³⁸, *SOCS3*³⁹, *PIK3CA*⁴⁰, *AKT1*⁴¹, *EGFR*⁴⁰, *TP53*⁴², *JUN*⁴³, *SMAD4*⁴⁴, *PKT2*⁴⁵, *MAPK14*⁴⁶, *GRB2*⁴⁷, *VEGFA*⁴⁸, *MYC*⁴⁹, *TNF*³⁴, *DVL2*⁵⁰, *ROS1*⁵¹, *ETV5*⁵², *FZD10*⁵³. These 25 genes were seldom mentioned in the intestinal transcriptomes of the fishes without intestinal air-breathing function, such as zebrafish⁹, European sea bass⁵⁴ and tilapia⁸, while these intestinal transcriptomes mainly told the roles of intestine in nutrition, immune and osmoregulatory functions. In this study, these 25 genes which were mainly involved in development, angiogenesis and cytoskeleton, were thus considered to be related to the intestinal air-breathing function occurred and formed process during posterior intestine development of loach. Further functional characterization of these genes determined by using transgenic, over expression, knockout and knockdown strategies may help elaborate the molecular mechanisms for intestinal air-breathing function formation in the loach.

In conclusion, this report represents the first application of Illumina sequencing technology for transcriptome studies in different developmental stages (4-Dph group, 8-Dph group, 12-Dph group, 20-Dph group, 40-Dph group and Oyd group) of loach posterior intestine, a typical accessory air-breathing organ. A total of 81300 unigenes were generated, and KEGG pathway annotation enabled us to assign 35291 DEGs to 238 pathways. Based on this transcriptome database, we further explored the gene expression profiles in loach posterior intestine, and identified 15 key pathways and 25 key DEGs, thus providing potential targets for candidate gene selection involved in formation of intestinal air-breathing function in *M. anguillicaudatus*. Analysis gained by this study will be useful for future studies on formation of the intestinal air-breathing function in loach. In addition, we demonstrated that *M. anguillicaudatus* is a good model for studies to identify and characterize the molecular basis of accessory air-breathing organ development in fish.

Methods

Ethics. This study was conducted in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of Huazhong Agricultural University. All experimental protocols were approved by the ethics committee of College of Fisheries, Huazhong Agricultural University. All efforts were made to minimize suffering of the loaches.

Preparation of tissues and histological observation. All the diploid loaches were obtained from our laboratory breeding populations, which were raised in the Aquaculture Base of College of Fisheries, Huazhong Agricultural University in Wuhan City, Hubei Province, China. Firstly, the posterior intestines from the one-year-old diploid loaches were used for RNA extraction and histological observation, meanwhile, the anterior intestines from them were used only for histological observation. Following, mature, healthy and two-year-old diploid loaches were served for reproduction. The pre-experiment showed that the loach would start to swallow air around 10-Dph, suggesting the initial formation period of intestinal air-breathing function of loach began at around 10 days post hatch. Therefore, offspring of two year old diploid loaches were collected at the following developmental stages for histological observation, including 4-Dph, 6-Dph, 8-Dph, 10-Dph, 12-Dph, 20-Dph and 40-Dph. With combination of histological observation of the posterior intestine and loach behavior observation, the different developmental stages (Table 3) selected for developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function, finally included 4-Dph (none), 8-Dph (before the occurrence of intestinal air-breathing), 12-Dph (early stage of the formation of intestinal air-breathing), 20-Dph (mature stage of the formation of intestinal air-breathing), 40-Dph (stable stage of the formation of intestinal air-breathing) and Oyd (stable stage of the formation of intestinal air-breathing). After anesthetizing, intestines of loaches from 4-Dph and posterior intestines of loaches from 8-Dph, 12-Dph, 20-Dph were dissected by needles under a dissecting microscope (Zeiss Stemi 2000-C), while posterior intestines of loaches from 40-Dph were dissected by scissors and tweezers. All intestine tissues (4-Dph, 8-Dph, 12-Dph, 20-Dph, 40-Dph and Oyd) were harvested, quickly frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. The experimental loaches were cultured in a flowing water with $6.5 \pm 0.7 \text{ mg.L}^{-1}$ dissolved oxygen and

Sample names	Histological characteristics of the posterior intestines	Developmental stage of intestinal air-breathing
4-Dph	The intestinal epithelium was simple columnar epithelium.	None (No intestinal air-breathing phenomenon was observed)
8-Dph	The epithelium of posterior intestine was still simple columnar epithelium.	Before the occurrence of intestinal air-breathing (No intestinal air-breathing phenomenon was observed)
12-Dph	A certain amount of erythrocytes were found in posterior intestine and the loach swam near the surface to swallow the air.	Early stage of the formation of intestinal air-breathing (The intestinal air-breathing phenomenon was observed)
20-Dph	The number of erythrocytes and blood capillaries increased in posterior intestine.	Mature stage of the formation of intestinal air-breathing (The intestinal air-breathing function obviously enhanced and matured)
40-Dph	The posterior intestine was abundant in erythrocytes and blood capillaries.	Stable stage of the formation of intestinal air-breathing
Oyd	The posterior intestine was abundant in erythrocytes and blood capillaries.	Stable stage of the formation of intestinal air-breathing

Table 3. Characteristics of *M. anguillicaudatus* during intestinal air-breathing function occurred and formed process.

25.0 ± 2 °C water temperature. Loaches during the larval and early juvenile stages were fed four times a day, and adult loaches were fed twice a day. All the samples used for histological observation were obtained and preserved in 4% paraformaldehyde at room temperature. After dehydration in a graded series of ethanol and transparency by xylene, the posterior intestine samples were embedded in paraffin and sectioned (5 µm thick), using a Leica RM 2135 rotary microtome (Leica Ltd, Wetzlar, Germany). Dewaxed serial sections were stained with HE staining. Images were taken using a Zeiss Axiovert imaging system.

RNA extraction, RNA-seq library construction and sequencing. Total RNA was extracted from approximately 15 individuals per developmental stage using Trizol (Invitrogen, CA, USA) according to the manufacturer's instructions. Then the total RNA was treated with RNase-free DNase I (Takara Bio, Japan) for 30 min at 37 °C to remove residual DNA. NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA) and RNase free agarose gel electrophoresis were performed to assess the quantity and quality of total RNA. The total RNA extracted from specimens of each development stage was pooled together as one stage-specific sample. Poly (A) mRNA was isolated using oligo-dT beads (Qiagen). All mRNA was broken into short fragments by adding fragmentation buffer. First-strand cDNA was generated using random hexamer-primed reverse transcription, followed by the synthesis of the second-strand cDNA using RNase H and DNA polymerase I. The cDNA fragments were purified using a QIAquick PCR extraction kit. These purified fragments were then washed with EB buffer for end reparation poly (A) addition and ligated to sequencing adapters. Following agarose gel electrophoresis and extraction of cDNA from gels, the cDNA fragments were purified and enriched by PCR to construct the final cDNA library. The cDNA library sequencing was carried out with Illumina HiSeq 2500 RNA-Seq (Illumina, San Diego, CA, USA) using the paired-end technology by Gene Denovo Company (Guangzhou, China).

De novo assembly and functional annotation. The high-quality sequences were required for *de novo* assembly analysis. Prior to assembly and mapping, a filter was used to remove low-quality reads, which were defined as those with adaptor or ambiguous bases ($N < 5\%$), as well as all reads with nucleotides that had Phred quality scores < 20 , where a Phred score of 20 corresponds to a 1% expected error rate. Then all clean reads of the six libraries were jointly assembled into unigenes employed by Trinity software. Since there is no reference genome of *M. anguillicaudatus*, a k-mer value cutoff of 25 was used after removing redundant nucleotide sequences by Tgicl (v2.1, <http://sourceforge.net/projects/tgicl/files/tgicl%20v2.1/>).

Finally, unigenes were aligned against the Nr protein database, Swissprot, KOG, and KEGG of protein database using BlastX with an E-value $< 10^{-5}$. The sequence direction of the unigenes was determined according to the best alignment results. When the results were conflicted among databases, the direction was determined consecutively by Nr, Swissprot, KOG and KEGG. When a unigene would not align to any database, ESTScan (<http://myhits.isb-sib.ch/cgi-bin/estscan>) was used to predict coding regions and determine sequence direction. GO annotation was analyzed by Blast2GO software (<https://www.blast2go.com/>). Functional classification of the unigenes was performed using WEGO software.

Enrichment, dynamic expression profile and protein-protein interaction (PPI) network analysis of differentially expressed genes (DEGs). The mapped fragments were normalized for RNA length according to fragment per kilobase of exon model per million mapped reads (FPKM) for each gene between the six pooled samples, which facilitated the comparison of transcript levels between samples. DEGs between the two libraries (4-Dph_vs_8-Dph, 8-Dph_vs_12-Dph, 12-Dph_vs_20-Dph, 20-Dph_vs_40-Dph, 40-Dph_vs_Oyd, 4-Dph_vs_12-Dph, 4-Dph_vs_20-Dph, 4-Dph_vs_40-Dph and 4-Dph_vs_Oyd) were identified by the DEGseq package using the MA-plot-based method with Random Sampling model (MARS) method⁵⁵. DEGs between the each two libraries were selected using the following filter criteria: FDR (False discovery rate) < 0.01 and the absolute value of $\log_2\text{Ratio} \geq 1$, meaning that the expression difference for each DEG in different libraries should be at least two-fold. Go enrichment analysis (p-value ≤ 0.05) of the DEGs was performed using Goseq with the Wallenius non-central hyper-geometric distribution model to adjust gene length bias in DEGs. KEGG pathway enrichment analysis of the DEGs was done using KOBAS⁵⁶ with the hyper-geometric distribution model. The enrichment p-values were adjusted using the Benjamin and Hochberg method. In addition, the determination of expression patterns of DEGs through developmental stages was completed with clusters generated by STEM⁵⁷.

The interaction relationships of the enzymes encoded by DEGs that were obtained from STRING database (<http://string-db.org/>) providing both experimental and predicted interaction information. This database was used to analyze PPI network for DEGs by calculating their required confidence score; a score > 0.4 was chosen as the cut-off criterion.

Validation of RNA-Seq Results by qRT-PCR. To examine the reliability of the RNA-Seq results, 9 differently expressed genes were selected randomly for validation by qRT-PCR: *IFT22*, *PP2AB*, *VEGF α* , *gdh*, *anxa1a*, *VIP*, *Mt*, *PP1* and *Spon1b*. Total RNA was extracted from *M. anguillicaudatus* at six developmental stages as described above. The cDNA was synthesized from 3 μ g of total RNA for each sample using Prime-ScriptTM RT reagent Kit (TaKaRa, Dalian, China). Primers (shown in Supplementary Table S4) were designed using Primer Premier 5.0. Reactions were performed with SYBR[®] Premix Ex Taq[™] kit (TaKaRa) in Bio-Rad CFX96 Real-Time PCR system. The PCR mixture consisted of 12.5 μ l SYBR Premix Ex Taq II, 1 μ l template cDNA, 0.5 μ l forward primer (10 μ M), 0.5 μ l reverse primer (10 μ M) and 10.5 μ l ddH₂O. qRT-PCR detection system was: denature at 95 °C for 30 s, followed by 39 cycles of 5 s at 95 °C, and 58 °C for 30 s. Following the amplification, a dissociation curve was performed to verify that a single product was generated. Each sample was run in four times technical replicates. The relative expression levels were normalized to the endogenous control gene β -actin, and expression ratios were calculated by using the $2^{-\Delta\Delta Ct}$ method.

References

- McMahon, B. R. & Burggren, W. W. Respiratory physiology of intestinal air breathing in the teleost fish *Misgurnus anguillicaudatus*. *J. Exp. Biol.* **133**, 371–393 (1987).
- Wang, Y. J. *et al.* Effects of the timing of initial feeding on growth and survival of loach (*Misgurnus anguillicaudatus*) larvae. *Aquacult. Int.* **18**, 135–148 (2010).
- Park, J. Y., Kim, I. S. & Kim, S. Y. Structure and mucous histochemistry of the intestinal respiratory tract of the mud loach, *Misgurnus anguillicaudatus* (Cantor). *J. Appl. Ichthyol.* **19**, 215–219 (2003).
- Goncalves, A. F., Castro, L. F., Pereira-Wilson, C., Coimbra, J. & Wilson, J. M. Is there a compromise between nutrient uptake and gas exchange in the gut of *Misgurnus anguillicaudatus*, an intestinal air-breathing fish? *Comp. Biochem. Phys. D* **2**, 345–355 (2007).
- Ghosh, S. K., Ghosh, B. & Chakrabarti, P. Fine anatomical structures of the intestine in relation to respiratory function of air-breathing loach, *Lepidocephalichthys guntea*. *Acta Ichthyol. Piscat.* **41**, 1–5 (2011).
- Song, W., Jiang, K. J., Zhang, F. Y., Lin, Y. & Ma, L. B. Transcriptome sequencing, *de novo* assembly and differential gene expression analysis of the early development of *Acipenser baeri*. *PLoS ONE* **10**, e0137450 (2015).
- Morais, S., Edvardsen, R. B., Tocher, D. R. & Bell, J. G. Transcriptomic analyses of intestinal gene expression of juvenile Atlantic cod (*Gadus morhua*) fed diets with *Camelina* oil as replacement for fish oil. *Comp. Biochem. Phys. B* **161**, 283–293 (2012).
- Ronkin, D., Seroussi, E., Nitzan, T., Doron-Faigenboim, A. & Cnaani, A. Intestinal transcriptome analysis revealed differential salinity adaptation between two tilapia species. *Comp. Biochem. Phys. D* **13**, 35–43 (2015).
- Wang, Z. Y. *et al.* Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine. *BMC Genomics* **11**, 392 (2010).
- Zhao, Y., Yang, H. S., Storey, K. B. & Chen, M. Y. RNA-seq dependent transcriptional analysis unveils gene expression profile in the intestine of sea cucumber *Apostichopus japonicus* during aestivation. *Comp. Biochem. Phys. D* **10**, 30–43 (2014).
- Ferrarresso, S. *et al.* Exploring the larval transcriptome of the common sole (*Solea solea* L.). *BMC Genomics* **14**, 315 (2013).
- Kaitetzidou, E., Xiang, J., Antonopoulou, E., Tsigenopoulos, C. S. & Sarropoulou, E. Dynamics of gene expression patterns during early development of the European seabass (*Dicentrarchus labrax*). *Physiol. Genomics* **47**, 158–169 (2015).
- Sun, L. N., Yang, H. S., Chen, M. Y., Ma, D. Y. & Lin, C. G. RNA-Seq reveals dynamic changes of gene expression in key stages of intestine regeneration in the sea cucumber *Apostichopus japonicus*. *PLoS ONE* **8**, e69441 (2013).
- Alam, M. M., Lal, S., FitzGerald, K. E. & Zhang, L. A holistic view of cancer bioenergetics: mitochondrial function and respiration play fundamental roles in the development and progression of diverse tumors. *Clin. Trans. Med.* **5**, 3 (2016).
- Reis, M. & Liebner, S. Wnt signaling in the vasculature. *Exp. Cell Res.* **319**, 1317–1323 (2013).
- Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nat. Rev. Cancer* **8**, 387–398 (2008).
- Guo, S. C., Liu, M. L. & Gonzalez-Perez, R. R. Role of Notch and its oncogenic signaling crosstalk in breast cancer. *BBA Rev. Cancer* **1815**, 197–213 (2011).
- Lewis, J. Notch signalling and the control of cell fate choices in vertebrates. *Semin. Cell Dev. Biol.* **9**, 583–589 (1998).
- McMahon, A. P., Ingham, P. W. & Tabin, C. J. Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53**, 1–114 (2003).
- Kazmers, N. H., McKenzie, J. A., Shen, T. S., Long, F. X. & Silva, M. J. Hedgehog signaling mediates woven bone formation and vascularization during stress fracture healing. *Bone* **81**, 524–532 (2015).
- Hao, W. J. *et al.* Licochalcone A-induced human gastric cancer BGC-823 cells apoptosis by regulating ROS mediated MAPKs and PI3K/AKT signaling pathways. *Sci. Rep.* **5**, 10336 (2015).
- Beisner, J. *et al.* A novel functional polymorphism in the transforming growth factor-beta2 gene promoter and tumor progression in breast cancer. *Cancer Res.* **66**, 7554–7561 (2006).
- Lee, S. H., Jeong, D. J., Han, Y. S. & Baek, M. J. Pivotal role of vascular endothelial growth factor pathway in tumor angiogenesis. *J. Korean Surg. Soc.* **89**, 1–8 (2015).
- Chao, J. T. & Davis, M. J. The roles of integrins in mediating the effects of mechanical force and growth factors on blood vessels in hypertension. *Curr. Hypertens. Rep.* **13**, 421–429 (2011).
- Stempel, M., Keding, M., Augenlicht, L. & Klampfer, L. Essential role of the JAK/STAT1 signaling pathway in the expression of inducible nitric-oxide synthase in intestinal epithelial cells and its regulation by butyrate. *J. Biol. Chem.* **282**, 9797–9804 (2007).
- Cho, Y. L. *et al.* Specific activation of insulin-like growth factor-1 receptor by ginsenoside Rg5 promotes angiogenesis and vasorelaxation. *J. Biol. Chem.* **290**, 467–477 (2015).
- Hardy, K. M., Booth, B. W., Hendrix, M. J. C., Salomon, D. S. & Strizzi, L. ErbB/EGF signaling and EMT in mammary development and breast cancer. *J. Mammary Gland Biol.* **15**, 191–199 (2010).
- Singh, S. K., Lakshmi, M. G. M., Saxena, S., Swamy, C. V. B. & Idris, M. M. Proteome profile of zebrafish caudal fin based on one-dimensional gel electrophoresis LCMS/MS and two-dimensional gel electrophoresis MALDI MS/MS analysis. *J. Sep. Sci.* **34**, 225–232 (2011).
- Zhang, K. *et al.* Overexpression of Raf-1 and ERK1/2 in sacral chordoma and association with tumor recurrence. *Int. J. Clin. Exp. Pathol.* **8**, 608–614 (2015).
- Boyle, W. J., Simonet, W. S. & Lacey, D. L. Osteoclast differentiation and activation. *Nature* **423**, 337–342 (2003).
- Rescigno, T., Capasso, A. & Tecce, M. F. Effect of docosahexaenoic acid on cell cycle pathways in breast cell lines with different transformation degree. *J. Cell Physiol.* **231**, 1226–1236 (2016).

32. Mazur, A. J. *et al.* Analysis of gelsolin expression pattern in developing chicken embryo reveals high GSN expression level in tissues of neural crest origin. *Brain Struct. Funct.* **221**, 515–534 (2016).
33. Li, L. *et al.* MiR-17-5p up-regulates YES1 to modulate the cell cycle progression and apoptosis in ovarian cancer cell lines. *J. Cell Biochem.* **116**, 1050–1059 (2015).
34. Maehr, T., Vecino, J. L. V., Wadsworth, S., Wang, T. & Secombes, C. J. Four CISH paralogues are present in rainbow trout *Oncorhynchus mykiss*: differential expression and modulation during immune responses and development. *Mol. Immunol.* **62**, 186–198 (2014).
35. Xu, X. O. *et al.* Retraction notice to “Swap70b is required for convergent and extension cell movement during zebrafish gastrulation linking Wnt11 signalling and RhoA effector function” YDBIO 386/1(2013)191–203. *Dev. Biol.* **410**, 131 (2016).
36. Kumar, S. *et al.* IPS-1 differentially induces TRAIL, BCL2, BIRC3 and PRKCE in type I interferons-dependent and-independent anticancer activity. *Cell Death Dis.* **6**, e1758 (2015).
37. Zhang, K. *et al.* Overexpression of Raf-1 and ERK1/2 in sacral chordoma and association with tumor recurrence. *Int. J. Clin. Exp. Pathol.* **8**, 608–614 (2015).
38. Winter, P. S. *et al.* RAS signaling promotes resistance to JAK inhibitors by suppressing BAD-mediated apoptosis. *Sci. Signal.* **7**, ra122 (2014).
39. Elsaedi, F., Bembem, M. A., Zhao, X. F. & Goldman, D. Jak/Stat signaling stimulates zebrafish optic nerve regeneration and overcomes the inhibitory actions of socs3 and sfpq. *J. Neurosci.* **34**, 2632–2644 (2014).
40. Zhang, Q. *et al.* Combined analysis of rearrangement of ALK, ROS1, somatic mutation of EGFR, KRAS, BRAF, PIK3CA, and mRNA expression of ERCC1, TYMS, RRM1, TUBB3, EGFR in patients with non-small cell lung cancer and their clinical significance. *Cancer Chemoth. Pharm.* **77**, 583–593 (2016).
41. Atsaves, V. *et al.* Constitutive control of AKT1 gene expression by JUNB/CJUN in ALK plus anaplastic large-cell lymphoma: a novel crosstalk mechanism. *Leukemia*. doi: 10.1038/leu.2015.127 (2015).
42. Furth, N. *et al.* Down-regulation of LATS kinases alters p53 to promote cell migration. *Gene Dev.* **29**, 2325–2330 (2015).
43. Wei, S. N., Huang, Y. H., Huang, X. H. & Qin, Q. W. Characterization of c-Jun from orange-spotted grouper, *Epinephelus coioides* involved in SGIV infection. *Fish Shellfish Immun.* **43**, 230–240 (2015).
44. Sun, Y. H., Tseng, W. C., Fan, X., Ball, R. & Dougan, S. T. Extraembryonic signals under the control of MGA, Max, and Smad4 are required for dorsoventral patterning. *Dev. Cell.* **28**, 322–334 (2014).
45. Pye, V. E., Christensen, C. E., Dyer, J. H., Arent, S. & Henriksen, A. Peroxisomal Plant 3-Ketoacyl-CoA thiolase structure and activity are regulated by a sensitive redox switch. *J. Biol. Chem.* **285**, 24078–24088 (2010).
46. Umasuthan, N., Bathige, S. D. N. K., Noh, J. K. & Lee, J. Gene structure, molecular characterization and transcriptional expression of two p38 isoforms (MAPK11 and MAPK14) from rock bream (*Oplegnathus fasciatus*). *Fish Shellfish Immun.* **47**, 331–343 (2015).
47. Ahmed, Z. *et al.* Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity. *J. Cell. Biol.* **200**, 493–504 (2013).
48. Zhao, Y. J. *et al.* Alternative splicing of VEGFA, APP and NUMB genes in colorectal cancer. *World J. Gastroenterol.* **21**, 6550–6560 (2015).
49. Hsu, T. Y. T. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, 384 (2015).
50. Metcalfe, C. *et al.* Dvl2 promotes intestinal length and neoplasia in the *Apc^{Min}* mouse model for colorectal cancer. *Cancer Res.* **70**, 6629–6638 (2010).
51. Jin, Y. *et al.* ROS1 gene rearrangement and copy number gain in non-small cell lung cancer. *Virchows Arch.* **466**, 45–52 (2015).
52. Akagi, T. *et al.* ETS-related transcription factors ETV4 and ETV5 are involved in proliferation and induction of differentiation associated genes in Embryonic Stem (ES) Cells. *J. Biol. Chem.* **290**, 22460–22473 (2015).
53. Garcia-Morales, C., Liu, C. H., Abu-Elmagd, M., Hajihosseini, M. K. & Wheeler, G. N. Frizzled-10 promotes sensory neuron development in *Xenopus* embryos. *Dev. Biol.* **335**, 143–155 (2009).
54. Sarrapoulou, E., Galindo-Villegas, J., García-Alcázar, A., Kasapidis, P. & Mulero, V. Characterization of European sea bass transcripts by RNA-seq after oral vaccine against *V. anguillarum*. *Mar. Biotechnol.* **14**, 634–642 (2012).
55. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
56. Mao, X. Z., Cai, T., Olyarchuk, J. G. & Wei, L. P. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–93 (2005).
57. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).

Acknowledgements

This study was financially supported by National Natural Science Foundation of China (Project Number: 31201719) and the Fundamental Research Funds for the Central Universities of China (Project Number: 2662015PY033).

Author Contributions

X.C. conceived and designed the experiments; W.L., X.X., S.H. and C.L. performed the experiments; W.L. analyzed the data; W.L. wrote the paper; T.T. and X.C. modified the paper. All authors read and approved the final manuscript.

Additional Information

Accession codes: All transcriptome data are available in the NCBI Short Read Archive (SRA) database under accession SRP072927.

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Luo, W. *et al.* Developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function of Dojo loach, *Misgurnus anguillicaudatus*. *Sci. Rep.* **6**, 31845; doi: 10.1038/srep31845 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016