Functional Genomics of Air Breathing in Catfish

by

Xiaoli Ma

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 2, 2020

Key words: Transcriptome, Catfish, development, hypoxia, air breathing

Approved by

Rex Dunham, Chair, Professor of Fisheries, Aquaculture and Aquatic Sciences
Charles Y. Chen, Professor of Crop, Soil and Environmental Sciences
Joanna W. Diller, Associate Professor of Biological Sciences
Anthony G. Moss, Associate Professor of Biological Sciences

Abstract

Channel catfish (*Ictalurus punctatus*), tra catfish (*Pangasianodon hypophthalmus*) and walking catfish (*Clarias macrocephalus*) all belong to order Siluriformes, but *I. punctatus* does not possess an air breathing organ (ABO) and thus cannot breath in air, and *P. hypophthalmus* is a facultative air-breather and uses the swim bladder as its air breathing organ, which can help it conduct aerial breathing in the low oxygen conditions. *C. macrocephalus* have both gills and modified gill structures serving as an ABO, allowing them to aerial breathing (AB). *C. macrocephalus* can live in muddy marshes and burrows inside the mudflat during summer periods through air-breathing. These three species serve as a great model for studying the transition of life from water to terrestrial living, as well as understanding the genes which are critical for the functioning of air breathing.

In this study, seven early developmental stages in *I. punctatus* were selected for transcriptome analysis, 22,635 genes were covered with 590 million high-quality RNA-seq reads. Differential expression analysis between neighboring developmental timepoints revealed that the most enriched biological categories were associated with growth, development and morphogenesis. A gene co-expression network constructed using WGCNA approach identified four critical modules. Among some the candidate hub genes, GDF10, FOXA2, HCEA and SYCE3 were closely related with head formation, egg development and transverse central element of synaptonemal complexes. CK1, OAZ2, DARS1 and UBE2V2 were mainly associated with regulation of cell cycle, cell growth, brain development, differentiation and proliferation of enterocytes. IFI44L and ZIP10 were considered to regulate immune activity and control ion transport. TCK1 and TGFB1 were involved in transferring phosphate and regulating

cell proliferation. All these genes play vital roles in studying the early development in channel catfish and teleost fish.

Seven time points in *P. hypophthalmus* and *C. macrocephalus* were selected for RNA-Seq analysis based on their transition to a stage that could live at 0 ppm oxygen. More than 587 M clean reads were ultimately generated in *P. hypophthalmus*, and 504 M clean reads were retained in *C. macrocephalus*. A total of 21,448 and 25,239 unique genes were detected in *P. hypophthalmus* and *C. macrocephalus,* respectively. Through comparative genomic analysis with *I. punctatus*, 109 genes were identified to be *P. hypophthalmus*-specific genes, while 1,458 genes were detected to be specific in *C. macrocephalus*. Gene expression and network analysis were performed for these specific genes. Hypoxia challenge and microtomy experiments collectively suggested the timepoints for the functioning of air breathing in *P. hypophthalmus* and *C. macrocephalus*. Fourteen genes were detected to be important to the functioning of air breathing in *P. hypophthalmus*, in which, *hrg, grp and cx3cl1* genes were ultimately identified to be most related to the formation of air breathing ability in *P. hypophthalmus*. In addition, twenty-six genes were selected to be candidate genes involved in the formation of air-breathing function in *C. macrocephalus*, including *mb, ngb, hbae* genes. This study provides a large data resource for functional genomic studies in air breathing function in *P. hypophthalmus* and *C. macrocephalus*, and sheds light on the adaption of aquatic organisms to the terrestrial environment.

Acknowledgments

I would like to express my sincere and deep appreciation to my major advisor, Dr. Rex Dunham, for his guidance, continuous support and advice. Throughout my study in lab, Dr. Dunham has provided strong guidance for my research, and has always been there to give us support, help and encouragement. Without his guidance, support and encouragement, I would never have been able to finish my dissertation and degree work. His shared experience has a tremendous impact both on my research and life. I owe my gratitude to my committee members, Dr. Charles Chen, who always encouraged me with his experience and gave me career guidance. I would like to thank my committee member, Joanna Wysocka Diller, who always encouraged me to keep learning and be mindful when doing research. I would like to thank my committee member, Anthony G. Moss, who always be there to help me when I encounter problems for histology, and was always so encouraging. My thanks also extend to the outside reader for my dissertation, Dr. Xu Wang for his great help and advice, and his spirit of scientific research has always inspired and encouraged me.

I would like to extend my appreciation to my lab fellows, and all the people who helped me when I was in Auburn.

I would like to express my deepest appreciation and love to my parents, Xiantao Ma and Hua Zhang, for their encouragement, support, and thanks for their unconditional, endless love.

Table of Contents

List of Tables

List of Figures

List of Supplementary Tables

List of Supplementary Figures

List of Abbreviations

RNA-Seq	RNA-sequencing
WGCNA	Weighted gene co-expressed network analysis
FPKM	Fragments per kilobase of exon model per million reads mapped
WGS	Whole-genome sequencing
DEGs	Differentially expressed genes
PE	Paired-End
AA	Amino acid
log2FC	Log2-fold change
DPF	Days post fertilization
GO	Gene ontology
PPI	Protein-protein network
HCS	HiSeq Control Software
SCS	Synaptonemal complexes
ZFIN	Zebrafish Information Network
MSRs	Module-stage relationships
ABO	Air breathing organ
pident	Percentage of identical matches
qcovs	Query coverage per subject

PCA	Principal component analysis
PNEC	Neuroendocrine cells

Chapter 1 General introduction

1.1 Problem statement

Catfish (order Siluriformes) is one of the most taxonomically diverse orders representing over 12% of all fish species (about 33,000 species total) and 6.2% of all vertebrates (around 64,000 total species of vertebrates) [1], representing 36 families, 478 genera and more than 3,000 species [2]. Characteristic morphological phenotypes of catfish have a roughly cylindrical, muscular body, at least one pair of barbels attached to a large mouth, and dorsal and/or pectoral fins with spines [3]. The vast majority of this order are scaleless, except some families armoured with bony dermal plates, such as Loricariidae and Callichthyidae [3].

The catfish farming is the largest aquaculture industry in the United States, peaking at 350 million kg (kilograms) of catfish produced in 2003 before dramatically declining due to the recession, increased costs, inefficiencies and competition from imported products [4]. Farm-raised catfish was listed sixth in the 2010 "Top 10" fish and seafood consumption in the United States, at about 214 million kg of catfish being processed in 2010 [4]. Since then the production dropped to 138 million kg in 2011 [5]. Catfish production increased slightly in 2015-2017 and reached 150 million kg in 2017 [6], and continued to grow slowly with 158 million kg catfish produced in 2019 [5]. Alabama, Arkansas and Mississippi are the top three catfish production states in the US [4].

The interest in stocking channel catfish (*Ictalurus punctatus*) began in the United States [7]. Channel catfish were native to east of the Rocky Mountains, south of Canada, north of Mexico and south of Delaware. Then they were widely introduced to Europe, Russian Federation, Cuba and portions of Latin America, Thailand and China [7]. Globally, 432, 931 metric tons of channel catfish were produced in 2016 [7].

The culture of tra catfish *(Pangasianodon hypophthalmus)* accounts for 2/3 of Vietnam's overall aquaculture yields [7]. Due to the ideal climate, plenty of rivers and streams, as well as the developing technologies [8], the Mekong Delta, Vietnam has become one of the largest aquaculture producers with the annual output of tra catfish reaching 1.14 million metric tonnes and export income of about 1.4 billion dollars [9].

The walking catfish is widely distributed in Asia, the Indian subcontinent and Africa. It has great economic value as a food fish [10]. The Asian catfish, *Clarias batrachus* is one of the most cultured food fish of India, Bangladesh, Thailand and Philippines [11], *Clarias macrocephalus* is an important freshwater fish distributed throughout Southeast Asia [12]. The main cultured species in Thailand is the *Clarias*, with a current production of about 11,000 metric tons/year [7]. The hybrid between *Clarias macrocephalus* female × *Clarias gariepinus* male is the most preferred species that has been cultured for more than 20 years in Thailand [13]. This species is well adapted to a variety of environmental conditions, such as hypoxia [14].

When comparing the morphological traits of these three species, channel catfish have the largest head and a fusiform body, while walking catfish have relatively smaller heads and longer,

elongated body shape [15]. Tra catfish have a very small head compared to their body size [16]. In addition, the air-breathing capability also varies greatly among these three species: channel catfish does not possess an air breathing organ (ABO), and thus, cannot breathe air [17]. *Clarias macrocephalus*, has both gills and modified gill structures that serve as an air breathing organ, which allows aerial breathing (AB). Walking catfish can also live in muddy marshes through air-breathing and can sometimes travel across land, [18, 19]. The tra catfish are facultative air-breathers and utilize the swim bladder as an air breathing organ, which allows them to breathe air in low oxygen environments [20]. The ability of air-breathing is a big advantage for these species to combat and survive in the hypoxic environment.

1.2 Air-breathing in fish

Physiologists have been interested in the transition from aquatic to aerial gas exchange in vertebrates for centuries. Fishes that perform aerial respiration provide critical evidence of evolution during the transition from life in the ocean to terrestrial living [21]. The transition of water breathing to air breathing was one of the most important events in terms of the evolution of vertebrate life. About 350 million years ago, the water oxygen level gradually declined due to high temperatures and the decay of dead organic components that consumed the dissolved oxygen in swamps, rivers and lakes [22]. Severely reduced oxygen level in water forced certain fish ancestors to develop air-breathing organs. Some fish left aquatic environments and colonized the land; progeny of others are air-breathing species [23]. Air breathing fishes can

conduct gas exchange for respiration directly with the aerial environment rather than being limited to respiration through water. Air breathing fish may rise to the surface of the water, gulp air, and even crawl onto land and survive for a long time [24]. There are estimated to be more than 370 extant air breathing fish species in the world, distributed among 49 families, routinely involved in some level of aerial respiration [25]. During the hypoxic period, air breathers can air breathe to increase the gill oxygen uptake ability, which may involve removing the O_2 from air through increasing vascularization of the buccal cavity. However, in some extreme circumstances, even when water is enriched with oxygen, air-breathers which are completely dependent on aerial respiration may drown without access to the water surface [21].

The evolutionary transition from aquatic to aerial breathing involves changes in gas transport, ion regulation physiology and nitrogenous waste excretion. The air-breathing ability among fish also varies greatly [24]. Oxygen uptake directly from the air above surface water requires a radical transformation of their respiratory organs [26]. The air breathing organ (ABO) varies from species to species. Gills are typical fish breathing organ. In bony fish, capillary blood flow in operculum covered gills if efficient for gas exchange in the water. However, the gill is not adapted to non-aquatic air exchange in land environment [27]. In order to breathe oxygen from the air, certain fish species became air-breathers by evolving different kinds of air-holding chambers. The most common structure for air breathing is the swim bladder. In water breathing fish, this chamber is used to balance the hydrostatic pressure as well as navigating sounds. In some air-breathing species, it evolved to a fully functional lung [26]. Besides the swim bladder,

air-breathing fish utilize a diverse structure to perform the aerial-respiratory function, including but not limited to structured gills, modified intestine, muscle, skin, pharyngeal pouches, stomach and the skeleton [28] [22].

1.3 Function of air breathing organ

For species with modified intestine as an ABO, the air is taken in the mouth with unidirectional ventilation of the posterior region of the intestine and continuous exhaust gas from the vent [29]. In this region, the mucosa surface is lined with respiratory epithelium as well as capillary networks. Respiratory epithelium cells stagger with capillaries, serving as a tiny air-blood barrier (0.24-3.00 um) for air exchange [30]. Obviously, the intestinal gas exchange and digestion in fish are not mutually exclusive processes. The eating and breathing behavior can alternate in a very short time in intestine-breathing fish [29]. Apart from the capability of respiration, buoyancy regulation and the gas exchange in intestinal fish are supposed to be similar to the aerial gas exchange in other facultative air-breathing fish [29].

In teleost, gills serve as one of the most important organs. The functions of gills include O_2 and CO_2 exchange, acid balance, ionic homeostasis and ammonia excretion [31]. For the species that use modified gills as ABO, the efferent branchial arteries of anterior (first and second) gill arches serve as the accessory air breathing organ and are also the site for gas exchange. The ventral aorta originates from the heart and splits into a ventral branch and a dorsal branch. The ventral branch pumps blood to anterior gill arches, flows through the accessory ABO, then the

blood returns to the heart. The dorsal branch distributes blood to posterior (third and fourth) gill arches and proceeds to the circulatory system, which can help transport oxygen-rich blood to other tissues [32].

Specifically, *Clarias macrocephalus* is a continuous, facultative air breather, possessing an accessory respiratory organs (ARO) similar to those for *Clarias batrachus.* The ARO of *Clarias batrachus* are derived from gill tissue and consist of four sections: (1) supra-branchial chamber, (2) gill fans, (3) the dendritic organ or the respiratory tree, and (4) respiration epithelium [33]. Gill fans from the second and third gill arches extend to the supra-branchial chamber and split it into anterior and posterior recess, the dendritic organ from the second gill arch and fourth gill arch extends into the anterior recess and the posterior recess, respectively [32]. The respiration epithelium lines surround the supra-branchial chamber and maintain the structural characteristics as well as arterioarterial vessels similar to those of branchial filaments. Gill and ARO vessels align in parallel as integrated parts of the circulation system [32]. Although there are differences in the gross features of ABO in *Clarias* and *Heteropneustes fossilis*, the vascular organization and respiration vessel structure are similar between them, which may indicate that the air-breathing organ originated in a common *Siluridae* ancestor [34, 35]

The modified swim bladder is another form of the air breathing organ. The swim bladder in teleost is a large, trabeculated, well vascularized organ, which is widely considered as homologous with lung of immemorial Osteichthyes [36, 37]. Swim bladder has been reported as the organ involved in aerial respiration, includes *Gonorynchiformes* [38], *Characiformes* and

some species in *Siluriformes*, such as suborder *Gymnotus* and the *Pangasianodon hypophthalmus* (tra catfish) [39, 40]. The swim bladder in *Characiformes* and *P. hypophthalmus* is double chambered [41]. *P. hypophthalmus* is a facultative air breather, and the swim bladder of *P. hypophthalmus* extends from the posterior part of the head to the tail beyond the anus. This structure is supported by subdivided collagenic fibrous walls. There are two types of epithelium on the surfaces of the fibrous walls inside the swim bladder, one is the thin respiration type, covering the majority of the surface and highly vascularized with a large amount of red blood cells. This feature makes it a major place for gas exchange between air and blood in the swim bladder. The second type is thicker with a brush border [42].

1.4 The timing of the development for air breathing organ in embryos/fry

Anabas testudineus is an air breathing climbing perch which can survive in swamps containing low dissolved oxygen [43]. Aquatic respiration was found to be obligatory for young larvae with the young fish hatching from the egg membrane 10 h after fertilization. In the yolk-sac stages all the gas exchange processes take place in the well vascularized skin, none of gill or labyrinthine organs could be seen during this stage, and later the skin alone was not sufficient to sustain the growing demand for oxygen. The gill arches and filaments start differentiating at about 24 h and become functional at about 40 h after hatching [44]. The secondary lamellae with marginal channels were recognizable 60 h after hatching, and well-organized gill system (arches,

filaments, lamellae) forms around the same time. The labyrinthine organs differentiated at 51 h, but the hatchlings did not aerial breathe until 13-14 days [44, 45].

The Indian catfish *Heteropneustes fossili*, known as the stinging catfish, inhabits ponds, swamps and marshes, can survive the high density stocking due to its aerial respiration ability and thus is an ideal aquaculture species, especially for the Indian subcontinent and Southeast Asian [46]. The embryo differentiation starts about 10 h after fertilization, the gills begin to develop at 48 h. However, the air breathing behavior of larvae was not observed until 10 days after hatching [47, 48]. African catfish, *Clarias gariepinus*, is a widely cultured fish in many parts of Africa and Asia due to its rapid growth rate as well as tough vitality. Hatching starts at 17 h after fertilization at 28.5 ± 0.5◦C. The circulatory system was activated at 13 h after hatching. The rudimentary operculum was established about 43 h and it was not well developed until 59 h after hatching. The buccal and branchial systems are fully vascularized at about 83 h after hatching [49, 50]. In the case of *Clarias macrocephalus*, the gill was observed shortly after hatching, the suprabranchial organ begins to develop as the fish size reaches 11 mm, and this was also the initiation point of aerial breathing [51].

Pangasianodon hypophthalmus is a large freshwater catfish native to Thailand and the Mekong in Vietnam [52]. It possess a swim bladder for aerial gas exchange, but in contrast to most other air breathing species that typically have reduced gills, *P. hypophthalmus* is endowed with fully-developed gills, which makes it an excellent species that can adapt to varied environments through both aerial respiration and aquatic respiration [53]. This species can

survive in water temperatures from 24°C to a maximum of 32°C [54]. The diameter of *P. hypophthalmus* eggs post fertilization is 1.2-1.8mm, the eggs will hatch within 24 hours after fertilization at 26°C and above. At 5 d post fertilization, the yolk sac is totally absorbed and the larvae have the capability to swim [55]. The air breathing organ – swim bladder also develops at this time [56].

Pangasius sutchi is a continuous, obligatory air-breather and is an important Asian commercial species using the swim bladder as an air-breathing organ [57]. Which is an alveolus at the middle of esophagus. The fertilized eggs of *P. sutchi* are yellow to green-brown in color and adhesive. Nine hours after fertilization, the head, tail, and embryonic shield could be observed clearly, at 24-30 h post fertilization, the larvae were hatched with a body length of about 3 mm, and the heart became functional about 12-14 h post hatching. About 3 days post hatched, the yolk sac is fully absorbed [55]. The swim bladder with double chambers appeared at 8-10 days post hatching. Twelve days post hatching, the lumen of the swim bladder is formed and the inner surfaces of alveoli are covered by a network of capillary, conferring aerial breathing ability at this stage for *P. sutchi*. [57, 58].

1.5 Functional genomics

Functional genomics is the study of how genes and intergenic regions of the genome contribute to different biological processes [59]. A researcher in this field typically studies genes or regions on a “genome-wide” scale with the purpose of narrowing them down to several

candidate genes or regions to analyze in more detail [60]. As a new branch of science, functional genomics, is trying to better understand the genetics of an organism, such as the linear order of the nucleotide bases, the expression of all genes as a dynamic system, how the genes interact and influence biological pathways, and the evolutionary role of the organism [61, 62]. Generally, there are several specific functional genomics methods: (1) DNA level (genomics and epigenomics); (2) RNA level (transcriptomics); (3) protein level (proteomics); and (4) metabolite level (metabolomics) [60]. The introduction focus on two of the most widely used techniques: transcriptome sequencing (RNA-sequencing) and genome sequencing (DNA-sequencing).

1.5.1 RNA sequencing

Unlike the genome, the transcriptome is dynamic with developmental stages, biological conditions, and environmental effects at a specific moment of time [63]. RNA-Seq analysis contains the expression profile changes of each gene with the development of different stages, as well as the transcriptional structure of genes, which is of great value to the annotation of functional elements in the genome [64]. Sanger sequencing, also considered as "first-generation sequencing" [65], was widely used in screening of cDNA and EST libraries at the beginning of sequencing-based transcriptome analysis technologies [66]. However, although it is still widely used today to sequence smaller gene regions, identify site directed mutagenesis, verify the results of gene editing and others, traditional Sanger sequencing is limited by the discovery of substitutions, small insertions and deletions, and it is impossible to use sanger sequencing for

large genomic sequencing such as the human genome project [65, 67]. With rapid advances in sequencing platforms, sequencing-based technology has reached a new level. The arrival of next-generation sequencing (NGS) allows an entire genome to be sequenced in less than one day and offers generous volumes of data at a much lower cost - in some cases in excess of ten billion short reads per run for the Illumina NovaSeq system [65, 68]. Another advantage of NGS is that we can verify and quantify transcripts without prior knowledge of particular genes, which is a huge advantage compared with microarray approach [69].

There are three main types of platforms for NGS: Roche/454, Illumina/HiSeq 2000, and Life/SOLID [70]. RNA sequencing uses numbers of methods which can be roughly summarized as template preparation, library construction and sequencing, and data analysis. The type of data generated is different from one platform to another, and it is also different when comparing the data quality and cost [65]. Currently, the Illumina NovaSeq 6000 dominates the NGS market. Illumina yields longer and more accurate contigs, it also has less average sequencing error in the raw reads compared with other two platforms, moreover, Illumina platform is much more economical (only 1/4 of the cost of Roche 454 platform) [71].

The regular bioinformatic procedures for downstream RNA-Seq are summarized in Figure 1. As shown in Figure 1, the next-generation sequencing raw data is stored using the FASTQ format. The major steps involved in a basic RNA-seq analysis are transcriptome assembly, differential expression analysis, functional annotation and identification of key genes of relevant biological function. Sequencing adapters and low-quality bases were removed by trimmomatric

version 0.36 [18]. Based on whether a high quality reference genome is available, there are two strategies for transcriptome analysis. If a reference genome for the organism of interest is not available, de novo assembly is needed to generate a collection of transcript contigs. Several software are available for de novo assembly, such as ABySS [72], Trans-ABySS [73], QSRA [74], SOAPdenovo and Trinity [75, 76]. Among these de novo assembly software packages, Trinity is the most efficient and widely used one [77]. Trinity software includes almost all the packages needed for a complete transcriptome analyses without a reference genome, including RSEM for transcript abundance estimation, R/Bioconductor packages for differential expression analysis and TRANSDECODER for protein-coding gene prediction [76]. If a reference genome exists, reference-based assembly is a better choice. There are many software packages for reference-based datasets, such as Tophat-Cufflinks-Cuffdiff and their upgraded version Hisat2-StringTie-Ballgown [78, 79]. STAR-HTseq2-DEseq2 also has good performance for analyzing the RNA data with the aid of reference genomes [80-82]. We can choose the combination of software based on our research purpose and specific objective.

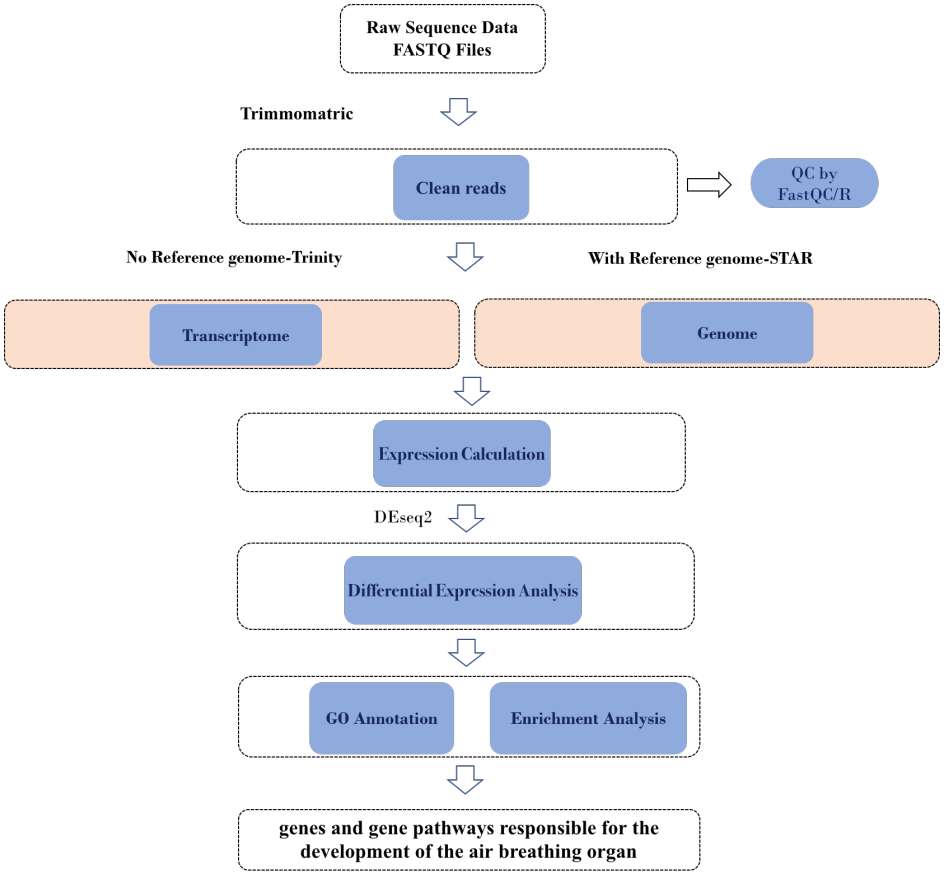


Figure 1. RNA Sequencing workflow

1.5.2 DNA sequencing

As the application of RNA sequencing, in recent years, the development of science and technologies, especially the high-throughput sequencing ("Next-generation" sequencing technology) can also be applied to DNA sequencing, and it has drastically improved genome studies, and many genomic resources have been developed [83]. Massive DNA sequencing platforms have been constructed in several years, among which, Illumina genome analyzer is the most widely used platform. The development of sequencing platforms can promote the

biological and biomedical research by enabling the comprehensive analysis of genomes, and it also provides a broad range of biological phenomena [84]. There is a list of software provided for genome analysis, ABySS [85], SOAPdenovo2 [75], and velvet are used for genome assembly [86]; BUSCO and Quast are widely used software to access the completeness of assembled genome [87, 88]. As shown in Figure 2, after quality control and the Trimmomatric step to remove adaptor and low-quality reads, there are two kinds of genome sequencing: one is whole genome sequencing (WGS), and the other is whole genome resequencing (WGRS). Genome sequencing investigates the order of DNA nucleotides, or bases, in a genome. Allowing the research in gene interaction and their functions in growth, development and maintenance of the organism [89, 90]. Whole genome resequencing can be used to examine the differences between a specific individual and the reference genome in the DNA sequence. It also performs massively parallel sequencing for retrieving enough DNA fragments to cover the whole span of the genome of the organism [91, 92].

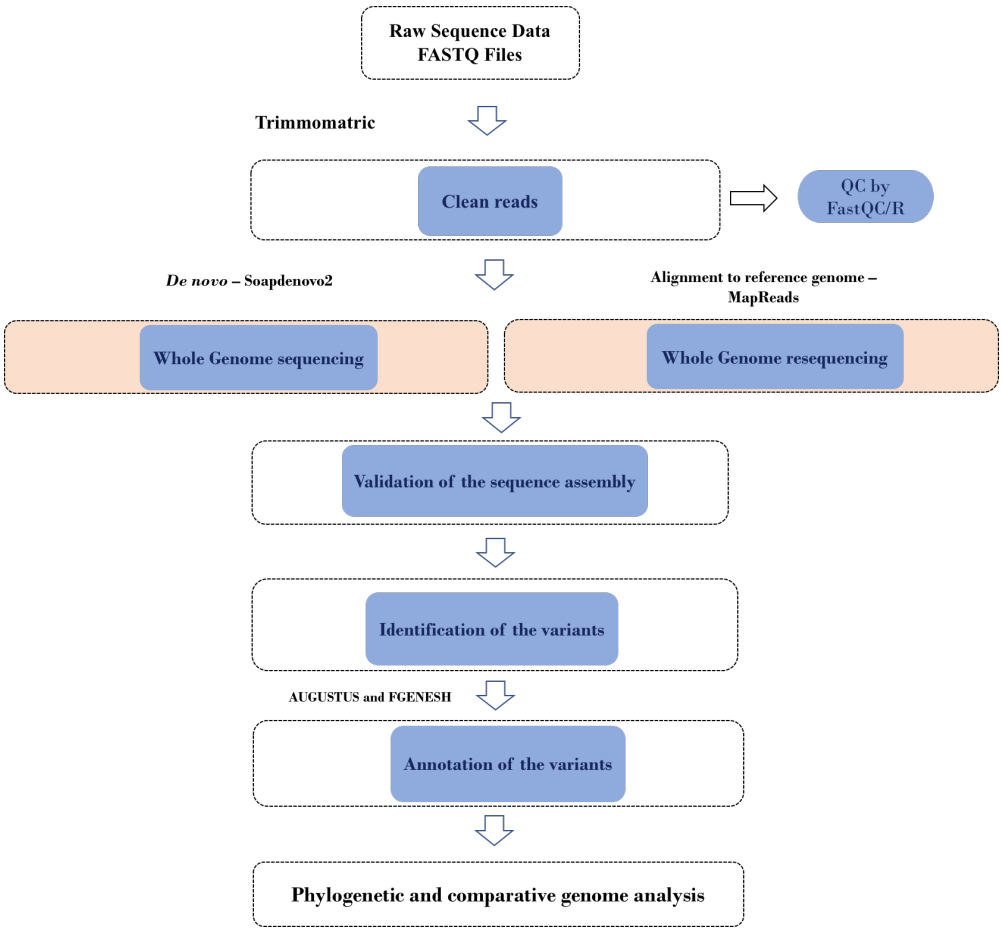


Figure 2. DNA sequencing workflow

1.6 The latest developments in DNA technology.

There has been a rapid development of DNA sequencing technology in the past few decades. Sanger sequencing, known as the first-generation sequencing, began in 1970. Throughout the 1990, next generation sequencing (NGS, also known as second-generation sequencing) nearly superseded sanger sequencing [44]. Except for the well-known DNA-seq and RNA-seq, some new technologies were rapidly developed based on the NGS technologies in recent years, such as BS-seq, ChIP-seq, small RNA-seq, single-cell sequencing and others. Shortly after that, a new

sequencing technology-third generation sequencing (TGS) has emerged, which did not require PCR amplification and achieved the individual sequencing of each DNA molecule. Over the last 10 years, Genome Wide Association Studies (GWAS) also presented a powerful tool to link a biological trait back to its underlying genetics [93]. Except for the sequencing-based technologies, the DNA-editing and DNA transfer technologies also play an essential role in recent years, especially the CRISPR/Cas9.

1.6.1 BS-seq

DNA methylation refers to the binding of a methyl group to the C5 carbon residue of cytosines through DNA methyltransferases [94]. DNA methylation is a critical epigenetic mechanism for the eukaryotic genome [95], involving several key physiological processes, such as X chromosome inactivation, parental allele-specific imprinting, imprinting disorder and cancer [95]. Bisulfite sequencing (BS-seq) relies on bisulfite conversion to detect the methylation without routine sequencing. Treatment of DNA with sodium bisulfite can covert cytosine into uracil, but leaves 5-methylcytosine residues unmodified. Uracil is identified as thymine by DNA polymerase, as amplifying bisulfite-treated DNA by PCR yields products in which unmethylated cytosines appear as thymine. By comparing the modified DNA sequence with the original sequence, the methylation state of the original DNA sequence can be inferred [94, 96]

1.6.2 ChIP-seq

Chromatin immunoprecipitation followed by sequencing (ChIP–seq) is a technique combining chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing, which can be used to analyze the positions where proteins interact with DNA, histone modifications or nucleosomes [97]. With the advantages of relevant low cost, high resolution, great coverage and large dynamic range, ChIP-seq thus became a consummate tool for the understanding of gene regulation and epigenetic mechanisms [97, 98]

1.6.3 Single-cell seq

Single cell sequencing (Single-cell seq) provides an unbiased method for the study of extent, basis and function of gene expression variation between seemingly identical cells [99], by deep sequencing DNA and RNA from a single cell, cellular functions can be investigated extensively [100]. The application of single-cell seq has a great impact on our conceptual understanding of various biological processes with broad implications for both basic and clinical research [101]. A typical Single-cell seq involves the isolation of a single cell, then implementing whole-genome-amplification (WGA), sequencing libraries construction and sequencing the DNA using a next-generation sequencer [101].

1.6.4 Third-Generation Sequencing

About 40 years ago, the emergence of Sanger sequencing was a revolutionary as it made genome sequencing possible for the first time. Next-generation sequencing (NGS) is a second revolution, which made genome sequencing much cheaper and efficient [102]. But at the same time, NGS also have several shortcomings, and short reads are the most notable drawbacks. In recent years, the appearance of third-generation sequencing (TGS), also known as long-read sequencing or single-molecule sequencing (SMS) technology, makes it possible for dramatically reading lengths within a short time [103]. Additionally, this new technology can identify epigenetic modifications on native DNA directly [103]. Table 1 compares first-generation sequencing, next-generation sequencing and Third-generation sequencing [102, 104, 105].

Table 1. Comparison of first-generation sequencing (FGS), next-generation sequencing (NGS) and third-generation sequencing (TGS)

	First generation	**Second generation**	**Third generation**
RNA-sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing and cDNA sequencing
Resolution	Averaged across many copies of the DNA	Averaged across many copies of the DNA molecule being sequenced	Single-molecule resolution

	molecule being sequenced		
Current read length	Moderate (800–1000 bp)	Short, generally much shorter than Sanger sequencing	Long, 1000 bp and longer in commercial systems
Current cost	Low cost per run	High cost per run	Low cost per run
Sample preparation	Moderately complex, PCR amplification not required	Complex, PCR amplification required	Ranges from complex to very simple depending on technology
Data analysis	Routine	Complex because of large data volumes and because short reads complicate assembly and alignment algorithms	Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values, potentially other

			base information such as kinetics

There are now three commercial SMS technologies, the first is the Illumina Tru-seq Synthetic Long-Read technology, this method is relatively slow and expensive. The second technology is termed "single-molecule real-time" (SMRT) released by Pacific Biosciences (PacBio). The third one is Nanopore sequencing introduced by Oxford Nanopore Technology (ONT) [104]. With this new third-sequencing technology, genome regions that remained ambiguous to date can now be resolved, and the complexity of transcriptomes can be inferred in unprecedented detail [102]. Long-read technology have now become a new revolution in genomics research.

1.6.5 CRISPR/Cas9

Various artificial nuclease systems have been produced for genome editing. Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) were the first two widely used engineered nucleases [106, 107]. Recently, genome editing using clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) supplied an alternative to ZFNs and TALENs, and became the most popular genome editing technology. CRISPR/Cas9 is a prokaryotic immune system that can be used to edit genes within organisms [108]. In general, CRISPR/Cas9 has three major types: type I, type II and type III, and 12

subtypes based on the Cas genes and the proteins they encode [109]. CRISPR/Cas9 consists of a nonspecific Cas9 nuclease and a list of programmable sequence-specific CRISPR RNA (crRNA), which can be used to guide Cas9 to DNA cleavage and produce double-strand breaks at target sites, followed by cellular DNA repair, leading to desired insertions, deletions or substitutions at specific sites [110]. The rapid development of CRISPR/Cas9 tools benefit many cell and molecular biology studies, and also drives innovative applications from basic biology research and medical research on human diseases [108]. The development of gene editing technologies offers a more comprehensive model to rationalize the balance between target recognition and nuclease activation, thus improving the targeting accuracy of CRISPR-Cas9 [111]. In addition, the development of multiple engineered sgRNAs with various target sequences can guide Cas9 to the corresponding sites within the same cells [112], which has many potential applications, such as the mutation of genes which control complex traits, as well as multiple members of gene families [113, 114].

Reference

1. Eschmeyer, W. and J. Fong, *Species by family/subfamily in the Catalog of Fishes.* San Francisco, CA: California Academy of Sciences, 2015.

2. Ferraris, C.J. and M.C. de Pinna, *Higher-level names for catfishes (actinopterygii, ostariophysi, siluriformes)*. 1999.

3. Armbruster, J.W. *Global catfish biodiversity*. in *American Fisheries Society Symposium*. 2011.

4. University, M.S., *Catfish farming.* Mississippi State University Extension, 2020.

5. NASS, *Catfish farming.* United States Department of Agriculture, 2020.

6. University, M.S., *Catfish production.* Mississippi State University Extension, 2017.

7. Nguyen, A.L., et al., *Simulated impacts of climate change on current farming locations of striped catfish (Pangasianodon hypophthalmus; Sauvage) in the Mekong Delta, Vietnam.* Ambio, 2014. **43**(8): p. 1059-1068.

8. Nguyen, P.T.H., et al., *Experimental assessment of the effects of sublethal salinities on growth performance and stress in cultured tra catfish (Pangasianodon hypophthalmus).* Fish physiology and biochemistry, 2014. **40**(6): p. 1839-1848.

9. De Silva, S.S. and N.T. Phuong, *Striped catfish farming in the Mekong Delta, Vietnam: a tumultuous path to a global success.* Reviews in Aquaculture, 2011. **3**(2): p. 45-73.

10. Teugels, G.G. and D. Adriaens, *Taxonomy and phylogeny of Clariidae: an overview.* Catfishes, 2003. **1**: p. 465-487.

11. Saxena, S., *Impact of different feeds on growth of catfish Clarias batrachus (Gunther).* International Journal of Science and Research, 2014. **3**(8): p. 1875-1878.

12. Kanjanaworakul, P., et al., *cDNA structure and the effect of fasting on myostatin expression in walking catfish (Clarias macrocephalus, Günther 1864).* Fish physiology and biochemistry, 2015. **41**(1): p. 177-191.

13. Koolboon, U., et al., *Effects of parental strains and heterosis of the hybrid between Clarias*

macrocephalus and Clarias gariepinus. Aquaculture, 2014. **424**: p. 131-139.

14. Chatchaiphan, S., et al., *De novo transcriptome characterization and growth-related gene expression profiling of diploid and triploid bighead catfish (Clarias macrocephalus Günther, 1864).* Marine biotechnology, 2017. **19**(1): p. 36-48.

15. Raji, A.R. and E. Norouzi, *Histological and histochemical study on the alimentary canal in Walking catfish (Claris batrachus) and piranha (Serrasalmus nattereri).* Iranian Journal of Veterinary Research, 2010. **11**(3): p. 255-261.

16. Roberts, T.R., *Pangasius bedado, a new species of molluscivorous catfish from Sumatra (Pisces, Siluriformes, Pangasiidae).* Nat. Hist. Siam Soc, 1999. **47**: p. 109-115.

17. Burleson, M.L. and N.J. Smatresk, *Branchial chemoreceptors mediate ventilatory responses to hypercapnic acidosis in channel catfish.* Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 2000. **125**(3): p. 403-414.

18. Bruton, M.N., *The survival of habitat desiccation by air breathing clariid catfishes.* Environmental Biology of Fishes, 1979. **4**(3): p. 273-280.

19. Belão, T., et al., *Cardiorespiratory responses to hypoxia in the African catfish, Clarias gariepinus (Burchell 1822), an air-breathing fish.* Journal of Comparative Physiology B, 2011. **181**(7): p. 905.

20. Lefevre, S., et al., *Effects of nitrite exposure on functional haemoglobin levels, bimodal respiration, and swimming performance in the facultative air-breathing fish Pangasianodon hypophthalmus.* Aquatic toxicology, 2011. **104**(1-2): p. 86-93.

21. Brauner, C., et al., *Transition in organ function during the evolution of air-breathing; insights from Arapaima gigas, an obligate air-breathing teleost from the Amazon.* Journal of Experimental Biology, 2004. **207**(9): p. 1433-1438.

22. Johansen, K., *9 Air Breathing in Fishes*, in *Fish physiology*. 1970, Elsevier. p. 361-411.

23. Hsieh, S.-T.T., *A locomotor innovation enables water-land transition in a marine fish.* PloS one, 2010. **5**(6): p. e11197.

24. Graham, J.B., *Air-breathing fishes: evolution, diversity, and adaptation*. 1997: Elsevier.

25. Graham, J.B. and H.J. Lee, *Breathing air in air: in what ways might extant amphibious fish biology relate to prevailing concepts about early tetrapods, the evolution of vertebrate air breathing, and the vertebrate land transition?* Physiological and Biochemical Zoology, 2004. **77**(5): p. 720-731.

26. Johansen, K., *Air-breathing fishes.* Scientific American, 1968. **219**(4): p. 102-111.

27. Jesse, M.J., C. Shub, and A.P. Fishman, *Lung and gill ventilation of the African lung fish.* Respiration physiology, 1967. **3**(3): p. 267-287.

28. Burggren, W.W. and K. Johansen, *Circulation and respiration in lungfishes (Dipnoi).* Journal of Morphology, 1986. **190**(S1): p. 217-236.

29. McMAHON, B.R. and W.W. BURGGREN, *Respiratory physiology of intestinal air breathing in the teleost fish Misgurnus anguillicaudatus.* Journal of Experimental Biology, 1987. **133**(1): p. 371-393.

30. Podkowa, D. and L. Goniakowska-Witalińska, *Adaptations to the air breathing in the posterior intestine of the catfish (Corydoras aeneus, Callichthyidae). A histological and ultrastructural study.* Folia biologica, 2002. **50**(1-2): p. 69-82.

31. Rombough, P., *The functional ontogeny of the teleost gill: which comes first, gas or ion exchange?* Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 2007. **148**(4): p. 732-742.

32. Olson, K., et al., *Gill microcirculation of the air-breathing climbing perch, Anabas testudineus (Bloch): Relationships with the accessory respiratory organs and systemic circulation.* American journal of anatomy, 1986. **176**(3): p. 305-320.

33. Munshi, J.D., *The accessory respiratory organs of Clarias batrachus (Linn.).* Journal of morphology, 1961. **109**(2): p. 115-139.

34. Bevan, D.J. and D.L. Kramer, *The respiratory behaviour of an air-breathing catfish, Clarias macrocephalus (Clariidae).* Canadian journal of zoology, 1987. **65**(2): p. 348-353.

35. Olson, K., et al., *Microcirculation of gills and accessory respiratory organs of the walking catfish Clarias batrachus.* The Anatomical Record, 1995. **242**(3): p. 383-399.

36. Alexander, R.M., *Physical aspects of swimbladder function.* Biological Reviews, 1966.

41(1): p. 141-176.

37. Nguyen, T.H.P., *Effects of temperature and salinity on growth performance in cultured Tra catfish (Pangasianodon hypophthalmus) in Vietnam*. 2015, Queensland University of Technology.

38. Motta, P.J., *Mechanics and functions of jaw protrusion in teleost fishes: a review.* Copeia, 1984: p. 1-18.

39. Graham, J.B., D.L. Kramer, and E. Pineda, *Comparative respiration of an air-breathing and a non-air-breathing characoid fish and the evolution of aerial respiration in characins.* Physiological Zoology, 1978. **51**(3): p. 279-288.

40. Liem, K.F., B. Eclancher, and W.L. Fink, *Aerial respiration in the banded knife fish Gymnotus carapo (Teleostei: Gymnotoidei).* Physiological Zoology, 1984. **57**(1): p. 185-195.

41. Podkowa, D. and L. Goniakowska-Witalinska, *The structure of the airbladder of the catfish Pangasius hypophthalmus Roberts and Vidthayanon 1991 (previously P. sutchi Fowler 1937).* FOLIA BIOLOGICA-KRAKOW-, 1998. **46**: p. 189-196.

42. Phuong, L.M., et al., *Ontogeny and morphometrics of the gills and swim bladder of air-breathing striped catfish Pangasianodon hypophthalmus.* Journal of Experimental Biology, 2018. **221**(3): p. jeb168658.

43. Santhakumar, M., M. Balaji, and K. Ramudu, *Effect of sublethal concentrations of monocrotophos on erythropoietic activity and certain hematological parameters of fish Anabas testudineus (Bloch).* Bulletin of environmental contamination and toxicology, 1999. **63**(3): p. 379-384.

44. Hughes, G. and J.D. Munshi, *Fine structure of the respiratory organs of the Climbing perch, Anabas testudineus (Pisces: Anabantidae).* Journal of Zoology, 1973. **170**(2): p. 201-225.

45. Hughes, G., J.D. Munshi, and J. Ojha, *Post-embryonic development of water-and airbreathing organs of Anabas testudineus (Bloch).* Journal of fish biology, 1986. **29**(4): p. 443-450.

46. Puvaneswari, S., et al., *Early embryonic and larval development of Indian catfish, Heteropneustes fossilis.* EurAsian Journal of BioSciences, 2009. **3**.

47. Ratmuangkhwang, S., P. Musikasinthorn, and Y. Kumazawa, *Molecular phylogeny and biogeography of air sac catfishes of the Heteropneustes fossilis species complex (Siluriformes: Heteropneustidae).* Molecular phylogenetics and evolution, 2014. **79**: p. 82-91.

48. Singh Kohli, M. and S. Vidyarthi, *Induced breeding, embryonic and larval development in Heteropneustes fossilis (Bloch) in the agro-climatic conditions of Maharashtra.* Journal of the Indian Fisheries Association, 1990. **20**: p. 15-19.

49. Olaniyi, W.A. and O.G. Omitogun, *Stages in the early and larval development of the African catfish Clarias gariepinus (Teleostei, Clariidae).* Zygote, 2014. **22**(3): p. 314-330.

50. Van Snik, G., J. Van Den Boogaart, and J. Osse, *Larval growth patterns in Cyprinus carpio and Clarias gariepinus with attention to the finfold.* Journal of Fish Biology, 1997. **50**(6): p. 1339-1352.

51. Morioka, S., et al., *Growth and morphological development of laboratory-reared larval and juvenile bighead catfish Clarias macrocephalus (Siluriformes: Clariidae).* Ichthyological research, 2013. **60**(1): p. 16-25.

52. Singh, P., et al., *Isolation and characterisation of collagen extracted from the skin of striped catfish (Pangasianodon hypophthalmus).* Food chemistry, 2011. **124**(1): p. 97-105.

53. Lefevre, S., et al., *Partitioning of oxygen uptake and cost of surfacing during swimming in the air-breathing catfish Pangasianodon hypophthalmus.* Journal of Comparative Physiology B, 2013. **183**(2): p. 215-221.

54. Li, S., X. Lu, and R.T. Bush, *CO2 partial pressure and CO2 emission in the Lower Mekong River.* Journal of hydrology, 2013. **504**: p. 40-56.

55. Islam, A., *Embryonic and larval development of Thai Pangas (Pangasius sutchi Fowler, 1937).* Development, growth & differentiation, 2005. **47**(1): p. 1-6.

56. Morioka, S., et al., *Growth and morphological development of laboratory-reared larval and juvenile Pangasianodon hypophthalmus.* Ichthyological research, 2010. **57**(2): p. 139-147.

57. Liu, W., *Development of the respiratory swimbladder of Pangasius sutchi.* Journal of fish

biology, 1993. **42**(2): p. 159-167.

58. Zheng, W. and W. Liu, *Morphology and histology of the swimbladder and infrastructure of respiratory epithelium in the air-breathing catfish, Pangasius sutchi (Pangasiidae).* Journal of fish biology, 1988. **33**(1): p. 147-154.

59. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets—10 years on.* Nucleic acids research, 2010. **39**(suppl_1): p. D1005-D1010.

60. EMBL-EBI. *What Is Functional Genomics?* 2016 [cited 2019 14 April]; Available from: https://www.ebi.ac.uk/training/online/course/functional-genomics-i-introduction-and-designing-e/what-functional-genomics.

61. Bashinskaya, V., et al., *A review of genome-wide association studies for multiple sclerosis: classical and hypothesis-driven approaches.* Human genetics, 2015. **134**(11-12): p. 1143-1162.

62. Nussbaum, R.L., R.R. McInnes, and H.F. Willard, *Thompson & Thompson genetics in medicine e-book*. 2015: Elsevier Health Sciences.

63. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nature Reviews Genetics, 2011. **12**(10): p. 671.

64. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature reviews genetics, 2009. **10**(1): p. 57.

65. Metzker, M.L., *Sequencing technologies—the next generation.* Nature reviews genetics, 2010. **11**(1): p. 31.

66. Chen, S., et al., *454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in Panax ginseng.* Plant cell reports, 2011. **30**(9): p. 1593.

67. Behjati, S. and P.S. Tarpey, *What is next generation sequencing?* Archives of Disease in Childhood-Education and Practice, 2013. **98**(6): p. 236-238.

68. Grada, A. and K. Weinbrecht, *Next-generation sequencing: methodology and application.* The Journal of investigative dermatology, 2013. **133**(8): p. e11.

69. Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends in

genetics, 2008. **24**(3): p. 133-141.

70. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nature Reviews Genetics, 2016. **17**(6): p. 333.

71. Luo, C., et al., *Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample.* PloS one, 2012. **7**(2): p. e30087.

72. Birol, I., et al., *De novo transcriptome assembly with ABySS.* Bioinformatics, 2009. **25**(21): p. 2872-2877.

73. Robertson, G., et al., *De novo assembly and analysis of RNA-seq data.* Nature methods, 2010. **7**(11): p. 909.

74. Bryant, D.W., W.-K. Wong, and T.C. Mockler, *QSRA–a quality-value guided de novo short read assembler.* BMC bioinformatics, 2009. **10**(1): p. 69.

75. Luo, R., et al., *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.* Gigascience, 2012. **1**(1): p. 18.

76. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.* Nature protocols, 2013. **8**(8): p. 1494.

77. Evans, V.C., et al., *De novo derivation of proteomes from transcriptomes for transcript and protein identification.* Nature methods, 2012. **9**(12): p. 1207.

78. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols, 2012. **7**(3): p. 562.

79. Pertea, M., et al., *Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.* Nature protocols, 2016. **11**(9): p. 1650.

80. Dobin, A. and T.R. Gingeras, *Mapping RNA-seq reads with STAR.* Current protocols in bioinformatics, 2015. **51**(1): p. 11.14. 1-11.14. 19.

81. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.

82. Anders, S., *Analysing RNA-Seq data with the DESeq package.* Mol Biol, 2010. **43**(4): p. 1-17.

83. Morozova, O. and M.A. Marra, *Applications of next-generation sequencing technologies in functional genomics.* Genomics, 2008. **92**(5): p. 255-264.

84. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nature biotechnology, 2008. **26**(10): p. 1135.

85. Simpson, J.T., et al., *ABySS: a parallel assembler for short read sequence data.* Genome research, 2009. **19**(6): p. 1117-1123.

86. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome research, 2008. **18**(5): p. 821-829.

87. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-3212.

88. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-1075.

89. Consortium, I.H.G.S., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860.

90. Consortium, M.G.S., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520.

91. Huang, X., et al., *High-throughput genotyping by whole-genome resequencing.* Genome research, 2009. **19**(6): p. 1068-1076.

92. Rubin, C.-J., et al., *Whole-genome resequencing reveals loci under selection during chicken domestication.* Nature, 2010. **464**(7288): p. 587.

93. Korte, A. and A. Farlow, *The advantages and limitations of trait analysis with GWAS: a review.* Plant methods, 2013. **9**(1): p. 29.

94. Krueger, F., et al., *DNA methylome analysis using short bisulfite sequencing data.* Nature methods, 2012. **9**(2): p. 145.

95. Smith, Z.D., et al., *High-throughput bisulfite sequencing in mammalian genomes.* Methods, 2009. **48**(3): p. 226-232.

96. Schaefer, M., et al., *RNA cytosine methylation analysis by bisulfite sequencing.* Nucleic acids research, 2008. **37**(2): p. e12-e12.

97. Park, P.J., *ChIP–seq: advantages and challenges of a maturing technology.* Nature reviews genetics, 2009. **10**(10): p. 669.

98. Pepke, S., B. Wold, and A. Mortazavi, *Computation for ChIP-seq and RNA-seq studies.* Nature methods, 2009. **6**(11s): p. S22.

99. Shalek, A.K., et al., *Single-cell RNA-seq reveals dynamic paracrine control of cellular variation.* Nature, 2014. **510**(7505): p. 363.

100. Brennecke, P., et al., *Accounting for technical noise in single-cell RNA-seq experiments.* Nature methods, 2013. **10**(11): p. 1093.

101. Saliba, A.-E., et al., *Single-cell RNA-seq: advances and future challenges.* Nucleic acids research, 2014. **42**(14): p. 8845-8860.

102. van Dijk, E.L., et al., *The third revolution in sequencing technology.* Trends in Genetics, 2018.

103. Schadt, E.E., S. Turner, and A. Kasarskis, *A window into third-generation sequencing.* Human molecular genetics, 2010. **19**(R2): p. R227-R240.

104. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nature methods, 2007. **5**(1): p. 16.

105. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nature methods, 2008. **5**(1): p. 16-18.

106. Wood, A.J., et al., *Targeted genome editing across species using ZFNs and TALENs.* Science, 2011. **333**(6040): p. 307-307.

107. Gupta, R.M. and K. Musunuru, *Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9.* The Journal of clinical investigation, 2014. **124**(10): p. 4154-4161.

108. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-1278.

109. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity.* Molecular cell, 2014. **54**(2): p. 234-244.

110. Zhang, F., Y. Wen, and X. Guo, *CRISPR/Cas9 for genome editing: progress, implications and challenges.* Human molecular genetics, 2014. **23**(R1): p. R40-R46.

111. Chen, J.S., et al., *Enhanced proofreading governs CRISPR–Cas9 targeting accuracy.* Nature, 2017. **550**(7676): p. 407-410.

112. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-823.

113. Xie, K., B. Minkenberg, and Y. Yang, *Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system.* Proceedings of the National Academy of Sciences, 2015. **112**(11): p. 3570-3575.

114. Ma, X., et al., *CRISPR/Cas9 platforms for genome editing in plants: developments and applications.* Molecular plant, 2016. **9**(7): p. 961-974.

Chapter 2 Deep transcriptomic analysis reveals dynamic developmental progression during early development of channel catfish (*Ictalurus punctatus*)

2.1 Abstract

The transition from fertilized egg to larva is accompanied with various biological processes and gene expression changes. In this study, seven early developmental stages in channel catfish, *Ictalurus punctatus*, were selected for transcriptome analysis, and 22,635 expressed genes were covered with 590 million high-quality RNA-seq reads. Differential expression analysis between neighboring developmental timepoints revealed that the most enriched biological categories were associated with growth, development and morphogenesis, which was most evident at 2 vs. 5 days post fertilization (dpf) and 5 vs. 6 dpf. A gene co-expression network constructed using WGCNA approach identified four critical modules. Among some the candidate hub genes, GDF10, FOXA2, HCEA and SYCE3 were closely related with head formation, embryonic development, egg development and the transverse central element of synaptonemal complexes. CK1, OAZ2, DARS1 and UBE2V2 were mainly associated with regulation of cell cycle, cell growth, brain development, differentiation and proliferation of enterocytes. In addition, IFI44L and ZIP10 were thought to be critical for the regulation of immune activity and ion transport. Similarly, TCK1 and TGFB1 were involved in phosphate transport and regulating cell proliferation. All these genes play vital roles in embryogenesis and regulation of early

development in channel catfish. Critically, in this project, data resources for functional genomic studies were generated. Our work reveals new insights for exploring the underlying mechanisms of channel catfish early development.

2.2 Background

Catfish (order Siluriformes) is one of the most taxonomically diverse orders, which includes over 12% of all teleost species (about 33,000 teleost species) and 6.2% of all vertebrates (64,000 vertebrates) [1]. Channel catfish (*Ictalurus punctatus*) and its hybrid from mating with blue catfish (*I. furcatus*), males are the most extensively cultured type of fish in the USA, accounting for a farmgate revenue of $185 million in 2015 [2].

The study of embryogenesis is critical for a comprehensive understanding of the gene expression patterns and underlining biological changes during early embryonic developmental stages of an organism. There are many studies concerning the early embryonic development in model species using RNA-seq, such as mouse (*Mus musculus*), fruit fry (*Drosophila melanogaster*) and zebrafish (*Danio rerio*) [3-6]. However, the early development in channel catfish has not been studied at the level of transcriptome, which is a large void to further comprehend the differentiation and growth mechanisms of this commonly domesticated order. Channel catfish is one of the most studied catfish species, with the first genome assembly released in 2016 [7], which provides a vast resource for functional genomic studies and biological research.

Unlike the genome, the transcriptome is dynamic and could be a good reflection of the cellular states with developmental stages, biological conditions, and the environment in a specific moment [8]. RNA-Seq analysis could provide the gene expression profiles of developing embryos at different stages, as well as the transcriptional structures of genes, which is of great value to the annotation of functional elements of the genome [9]. Weighted gene co-expression network analysis (WGCNA) is a system biological method and usually utilized to correlate modules as well as to concatenate external traits, WGCNA is also used for exploring the genes within the same modules, which are thought to be highly connective, and the genes inside the same module potentially have similar functions [10]. WGCNA analysis can further identify hub genes, which is critical for a specific trait or biological process [10]. WGCNA is a powerful method to conduct sequencing analysis and has been widely applied in studying many different biological contexts, including cancerogenesis, brain imaging, and early development in various species [11-13].

In this study, we used deep RNA sequencing to investigate the gene expression profiles of channel catfish during early development. Seven early developmental stages, including 2 days post fertilization (dpf), 5 dpf, 6 dpf, 7 dpf, 8 dpf, 9 dpf and 10 dpf, were selected for transcriptome sequencing and analysis. A transcriptome dynamic progression is provided, which may serve as a blueprint for future investigation of early development and organogenesis. The aim of this study was to verify gene expression in a stage-specific manner and compare the expression profiles of seven early developmental stages to identify the differentially expressed

genes (DEGs) in each stage. The DEGs were used to conduct Gene Ontology (GO) enrichment analysis to study the biological functions and also, they were utilized for WGCNA analysis to verify the co-expressed modules. Cytoscape was utilized to identify the candidate hub genes within eight stage-specific modules, which may be closely involve in early development in channel catfish.

2.3 Materials and methods

2.3.1 Ethics statement

All of the experimental protocols involved in animal care and sample collection were approved by the Auburn University Institutional Animal Care and Use Committee. All samples were collected after euthanization with buffered MS-222 (200 mg/L). All animal handling procedures were performed following the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States.

2.3.2 Sampling of channel catfish

Kansas random strain of channel catfish was raised in earthen ponds at the Genetics Research Unit, E.W. Shell Research Center, Auburn University. The Kansas strain was derived from the Ninnescah River in Pratt, Kansas in 1911 [14, 15], and is the oldest domestic channel catfish strain in the US.

After harvesting, females and males were mated and spawned at the genetics facility greenhouse. Channel catfish embryos and larvae were obtained by artificially spawning brood stock. Fertilized eggs were incubated in a hatching trough at 25 to 26 degrees Celsius, with water hardness above 40 ppm and dissolved oxygen was 5 mg/L. If fungal infection occurred, a 15-minute static treatment of 100 ppm formalin was administered. Treatments ceased 24 hours before the expected hatch date. From 7 dpf, swim-up fry were fed to satiation six times per day using a powdered 50% protein starter diet from Purina® AquaMax® [15]. Channel catfish samples were obtained at 2, 5, 6, 7, 8, 9 and 10 dpf. A total of 20-50 embryos/larvae were collected at each sampling, and 200 ppm buffered MS-222 was utilized to euthanize larvae [15]. The samples were placed into 1.5 mL centrifuge tubes, flash-frozen in liquid nitrogen, and stored in -80°C for RNA extraction.

2.3.3 Microscopic anatomy

At each sampling timepoint, another 20-50 embryos/larvae were fixed with 10% phosphate-buffered formalin in 1.5 mL centrifuge tubes and sealed for microscopic analysis. For microscopic anatomical observations, samples were transferred to a 75% ethanol solution. Observations were conducted with a MEIJI TECHNO anatomy microscope and images were photographed with a Canon DS126311 camera.

2.3.4. RNA isolation, library construction and sequencing

In order to examine changes in expression through out early stages development, at each time point, 2, 5, 6, 7, 8, 9 and 10 dpf, eight embryos/larvae were randomly selected and divided into two replicate pools (four embryos/larvae each). For each replicate, samples of four embryos/larvae were homogenized in liquid nitrogen with a mortar and pestle. RNA extraction was conducted using the RNeasy Plus Kit (Qiagen) following the standard protocols. The concentration and quality of RNA were measured using RNA NanoDrop spectrophotometer (NanoDrop Technologies). Equal amounts of RNA from the two replicates for each sample were pooled together for RNA-Seq library construction and Illumina sequencing [16].

Library construction and sequencing reactions were conducted at GENEWIZ, LLC. (South Plainfield, NJ, USA). The RNA integrity was checked with 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Ribosomal RNA depletion was conducted using Ribozero rRNA Removal Kit (Illumina, San Diego, CA, USA). RNA sequencing library preparation was performed using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA). Enriched RNAs were fragmented for 15 minutes at 94 °C. The first strand and second strand cDNA were subsequently synthesized. The cDNA fragments were end-repaired and adenylated at 3'ends, and a universal adapter was ligated to cDNA fragments, followed by index addition and library enrichment with limited cycle PCR. Sequencing libraries were validated using the Agilent Tapestation 4200 (Agilent Technologies, Palo Alto, CA, USA) and quantified by using

Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). The samples were sequenced using a 2x150 Paired-End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS).

2.3.5. Reads mapping and differential expression

The channel catfish genome was obtained from the NCBI database (https://www.ncbi.nlm.nih.gov) and used as a reference for mapping reads. The channel catfish genome assembly included 783 Mb of 9,974 scaffolds with an N50 of 7.73 Mbp [7]. Channel catfish transcriptome data (.bcl files) generated from Illumina HiSeq were converted into fastq files and de-multiplexed using Illumina's bcl2fastq software (version 2.17). The quality of raw reads was controlled by FASTQC [17], reads were filtered by removing low-quality reads, adapters, and reads with length shorter than 36 bases using Trimmomatic v0.36 [18]. The resulting clean reads were quality-controlled again and aligned to the channel catfish genome by STAR software (version 2.7.0) [19], allowing less than a 4-bases mismatch. HTSeq-count [20] was conducted to calculate the number of aligned reads of each gene overlapping its exons. To perform the differential expression analysis with the embryonic development of channel catfish, an R package DESeq2 [21] was employed to calculate the log2-fold change (log2FC), a criterion with $|\log2FC| \geq 1$ and p-value < 0.05 was used as the threshold for evaluating the DEGs. When $\log2FC > 1$, DEGs were considered to be up-regulated; while $\log2FC < -1$, DEGs were considered to be down-regulated.

2.3.6 Enrichment analysis

For every differential expression comparison, the Gene Ontology (GO) terms of each gene were assigned by using zebrafish annotations for the unigene set. Enrichment analysis was also performed using the ClusterProfiler R package (version 3.6) [22] to profile their major biological processes, molecular functions and cellular components. The threshold of significance criteria was set at 0.05 for p-value and q-value cutoffs and the enriched GO terms were ranked by p-value.

2.3.7 Gene co-expression network construction

To verify the interesting gene modules and network properties of the gene expression profile in the early development of channel catfish, an R package named weighted gene co-expressed network analysis (WGCNA) [10] was employed following the standard protocol. Then, the intramodular connectivity and gene significance were applied to verify key co-expressed genes in the network and correlate the identified modules to external information, development stages. A total of 8,504 DEGs from 7 timepoints were used to calculate the correlation between samples. These DEGs were used for hierarchical clustering analysis.

2.3.8 Identification of development-related modules and visualization

After the co-expression network was constructed, the developmental stages-responsive modules and genes were selected based on the correlation coefficient between the modules and

developmental stages. The genes within the same modules are highly connective, and the genes inside the same module potentially have similar functions. To verify the biological function of the specific modules as well as the correlation between the modules and different development stages, genes inside the same modules were selected to perform GO enrichment analysis using the R package ClusterProfiler (version 3.6) [22]. Hub genes are defined as genes inside co-expression modules with high correlation. To further verify the hub genes and their possible roles in early development stages, the top 200 ranked genes within each module were extracted according to the intra-modular connectivity with module eigengenes, which were used to construct a protein-protein network (PPI). The R package Cytoscape (version 3.7.2) [23] was employed to identify genes of the highest node degree, which may have a critical function in the PPI.

2.4 Results

2.4.1 Morphology of the channel catfish embryos/larvae during early development

At 2 dpf, the embryos, which were developing within the chorion, were approximately 4 mm in length and appeared oval with no eyes (Figure 3). At 5 dpf, the average length was 9 mm, and tail buds were free from the yolk sac. The larval head was close to the yolk sac, eyes were observed, larvae laid on their side in the water, and activity of the tail allowed the fish to move slowly. At 6 dpf, the yolk sac was partially absorbed, and the embryo looked more like a larva. The larvae were able to swing their tails slowly, propelling them to the surface of the water.

Their average length was 9 mm. At 7 dpf, length had increased to 11 mm, the dorsal fin had started to develop, and bone could be observed clearly through the translucent body. From days 8 to 10, and especially at 10 dpf, the yolk sac was almost entirely absorbed. Total length was approximately 12-14 mm. The external features such as the, head, fins, musculature, mouth and barbels, the adult-like appearance of the larvae progressively developed between 8, 9 and 10 dpf.

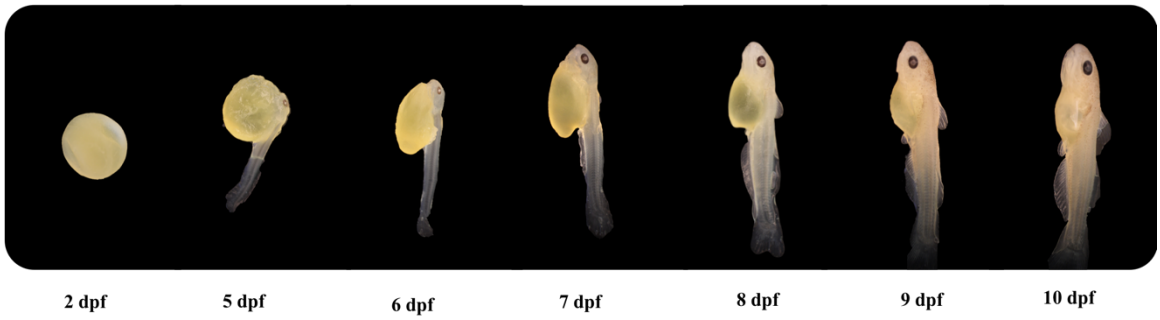


Figure 3. Channel catfish, *Ictalurus punctatus*, morphology at different development stages. dpf: days post fertilization.

2.4.2 Global analysis of channel catfish early development transcriptome

Initially, the transcriptome sequencing resulted in a total of 1,259 million raw reads for all samples. After removing low-quality reads with a quality score < 25 and very short reads of < 36 bases, more than 591 million clean reads were retained for further analysis. The clean reads for each sample were aligned to the channel catfish genome (Supplementary Table 1 and Supplementary Figure 1). The average number of raw, filtered reads number, GC content, number of mapped reads and mapping rate for the samples of each development stage are shown in

Supplementary Table 1. From each stage, a total number of 72.26% to 95.06% reads were successfully mapped. Pair-end sequencing (2×150 bp) was generated with an average read length ranging from 36 to 135 bp.

To profile channel catfish gene expression, clean reads were mapped to the channel catfish genome and normalization was employed using the FPKM (fragments per kilobase of exon model per million reads mapped) method. Genes with a FPKM value smaller than 0.1 were removed. In total, the expression of 22,635 distinct genes was identified. The highest number of expressed genes (22,266) occurred on day 9 post fertilization, while the 2 dpf sample contained the lowest number of expressed genes (20,670) (Figure 4a). A total of 19,415 of the same genes were expressed on each day (Supplementary Table 2).

2.4.3 Identification of DEGs among different channel catfish early developmental stages

To investigate genes related to the early development of channel catfish, differential gene expression analysis was conducted among the seven developmental stages in channel catfish using the software package DESeq2. The DEGs were generated by comparing two continuous developmental stages in the current study and were annotated. The number of DEGs varied from 690 (355 up-regulated, 335 down-regulated), between 10 dpf and 9 dpf, to 6,700 (4,298 up-regulated, 2,402 down-regulated) between 5 dpf and 2 dpf (Supplementary Table 3). Overall, the number of DEGs decreased over time of development (except a slight increase between 10 dpf and 9 dpf) of which 2,370 (1,663 up-regulated, 707

down-regulated), 1,146 (710 up-regulated, 436 down-regulated), 875 (504 up-regulated, 371 down-regulated) and 486 (371 up-regulated, 115 down-regulated). DEGs were distributed between 6 dpf and 5 dpf, 7 dpf and 6 dpf, 8 dpf and 7 dpf, 9 dpf and 8 dpf, respectively (Figure 4b). When all of the DEGs at various stages were combined, a high percentage (91.33%) of DEGs was detected between the first two pair comparisons, i.e., 7,767 (DEGs exclusively identified from 5 dpf vs. 2 dpf and 6 dpf vs. 5 dpf) out of 8,504 DEGs (all DEGs exclusively identified through development), indicating that great changes occurred in these stages. These stages may be the critical stages for the transitions from fertilized embryos to larvae in channel catfish, which was consistent with the observations of morphological changes. Moreover, MA plot was constructed to identify the transcripts significantly changed during the early development in channel catfish (Figure 5). Venn diagram (Supplementary Figure 2) indicated that 728 genes were differentially expressed in both 5 dpf vs. 2 dpf and 6 dpf vs. 5 dpf, 37 genes were differentially expressed in all stages.

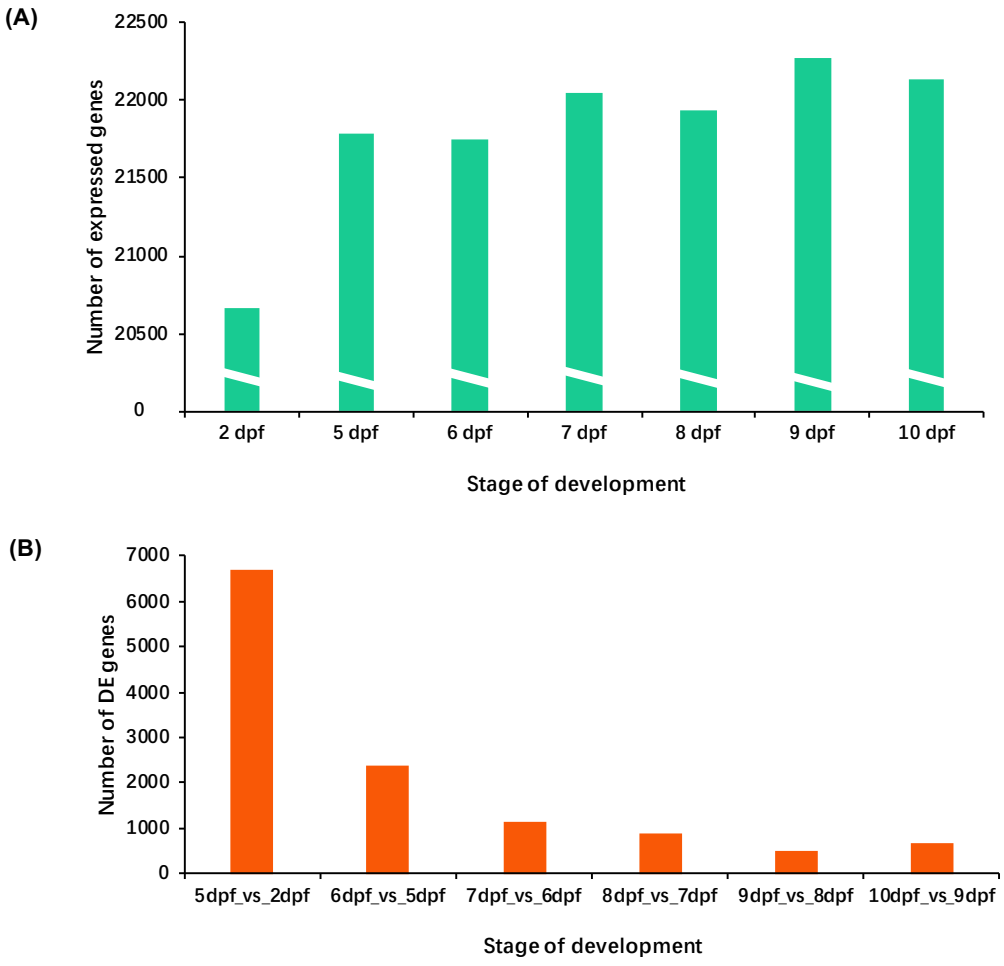


Figure 4. Gene expression during early development in channel catfish, *Ictalurus punctatus*. **(A)** Mean number of expressed genes of two replicates identified at each development stage; (**B**) The number of differentially expressed genes (DEGs) for comparison of each stage with the previous stage.

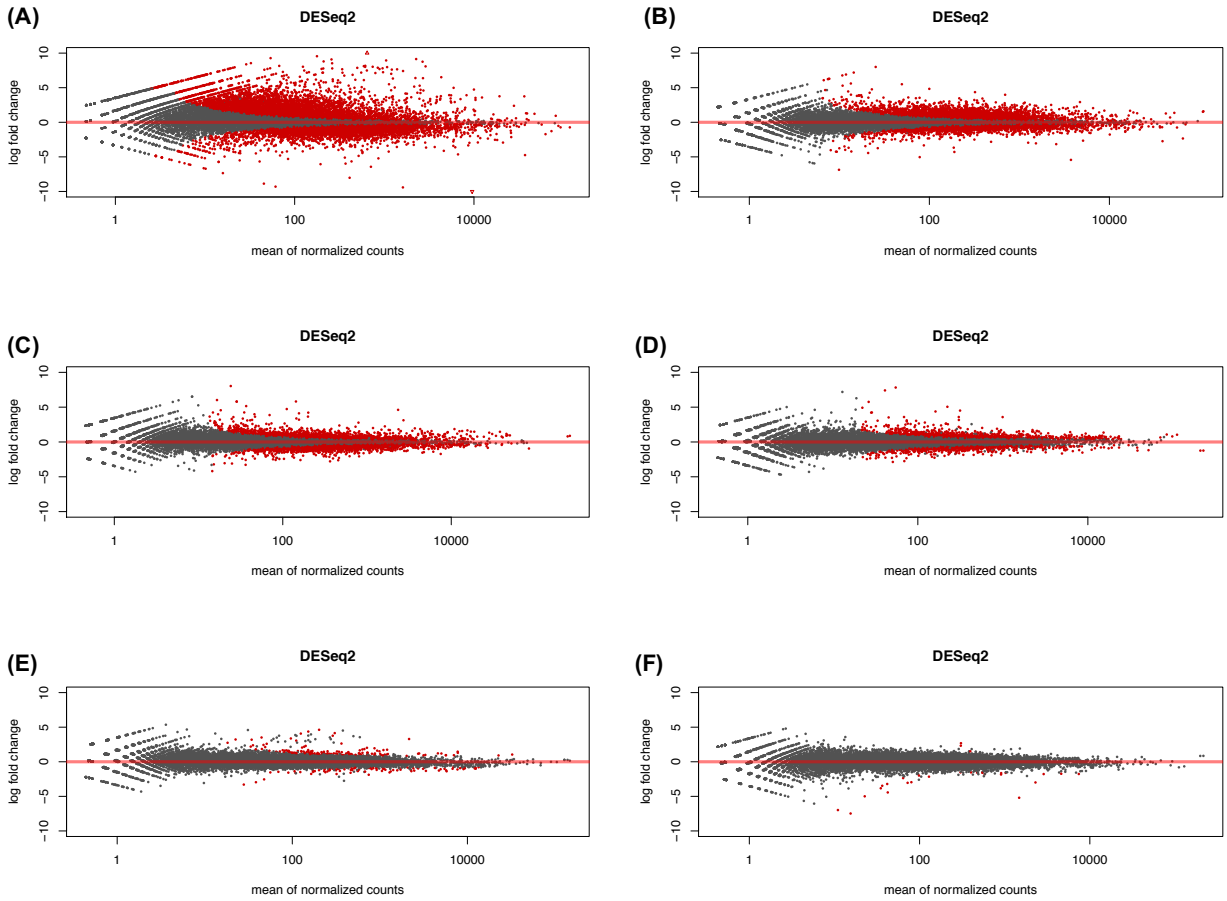


Figure 5. M-versus-A (MA) plot of the differentially expressed genes (DEGs) in different comparisons during early development in channel catfish, *Ictalurus punctatus*. Red dots indicate the downregulation (negative value) and upregulation (positive value). Black dots represent non-DEGs. (**A**) 5 days post fertilization (dpf) vs. 2 dpf; (**B**) 6 dpf vs. 5 dpf; (**C**) 7 dpf vs. 6 dpf; (**D**) 8 dpf vs. 7 dpf; (**E**) 9 dpf vs. 8 dpf; (**F**) 10 dpf vs. 9 dpf.

2.4.4 Gene ontology enrichment analysis of DEGs at different stages

Enrichment analysis of DEGs at each developmental stage was conducted, and the gene ontology (GO) categories that were significantly enriched during channel catfish early development are listed in Supplementary Table 3. The top 15 enriched categories for each stage

are listed here in Figure 6. Functional annotation of DEGs between 5 dpf and 2 dpf contained categories mainly associated with development, growth, synaptic signaling and ion transport involved in muscle development, regulation of nervous system development, chondrocyte development, bone growth, heart contraction, blood coagulation, chemical synaptic transmission and potassium ion transport. DEGs between 6 dpf and 5 dpf were mainly enriched in functions pertaining to development, morphogenesis and differentiation. These included genes involving embryonic skeletal system development, connective tissue development, gland development, neuron projection development, immune system development, heart morphogenesis, embryonic organ morphogenesis, cardiocyte differentiation and stem cell differentiation. Between 7 dpf and 6 dpf, enriched categories were most related to development and differentiation that included bone development, liver development and myeloid leukocyte differentiation. Between 8 dpf and 7 dpf, DEGs were enriched in development, organization and homeostasis activity, such as cardiac muscle fiber development, extracellular matrix organization, extracellular structure organization, sterol homeostasis and lipid homeostasis. Also, DEGs that were enriched from 9 dpf compared to 8 dpf belong to categories such as synapse activity and ion activity, including postsynaptic specialization, calcium ion transmembrane transporter activity, metal ion transmembrane transporter activity. Additionally, DEGs from 10 dpf compared to 9 dpf were enriched in regulation of some ion transport activity, including dopamine transport, voltage-gated calcium channel activity and transmitter-gated ion channel activity. It is notable that during early developmental stages of channel catfish, especially at 5 and 6 days post fertilization (the

transition stage from advanced embryo to larva), the most enriched categories were relevant to growth, development, proliferation, and morphogenesis.

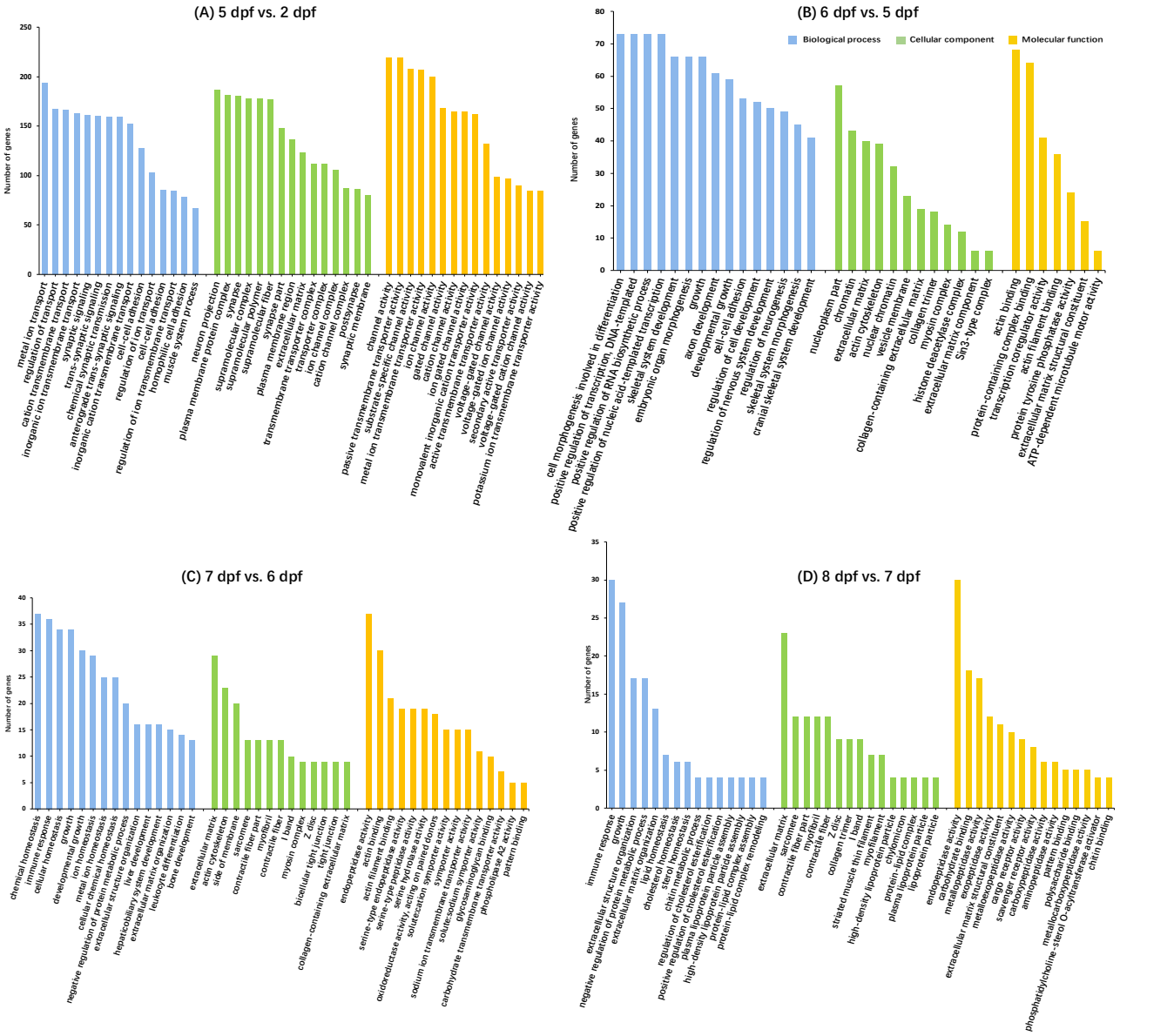

(A) 5 dpf vs. 2 dpf
Number of genes
metal ion transport
regulation of transport
cation transmembrane transport
inorganic ion transmembrane transport
synaptic signaling
chemical synaptic transmission
anterograde trans-synaptic signaling
inorganic cation transmembrane transport
cell-cell adhesion
regulation of ion transmembrane transport
homophilic cell adhesion
muscle system process
neuron projection
plasma membrane protein complex
synapse
supramolecular complex
supramolecular polymer
postsynapse part
plasma membrane region
extracellular matrix
transmembrane transporter complex
ion channel complex
cation channel complex
postsynapse
synaptic membrane
channel activity
passive transmembrane transporter activity
substrate-specific transporter activity
channel activity
metal ion transmembrane transporter activity
ion channel activity
gated channel activity
cation channel activity
ion gated channel activity
cation gated channel activity
monovalent inorganic cation transporter activity
active transmembrane transporter activity
voltage-gated channel activity
voltage-gated ion channel activity
secondary active transmembrane transporter activity
voltage-gated cation channel activity
potassium ion transmembrane transporter activity
(B) 6 dpf vs. 5 dpf
Biological process
Cellular component
Molecular function
Number of genes
cell morphogenesis involved in differentiation
positive regulation of transcription, DNA-templated
positive regulation of RNA biosynthetic process
positive regulation of nucleic acid-templated transcription
skeletal system development
embryonic organ morphogenesis
axon development
growth
developmental growth
cell-cell adhesion
regulation of cell development
regulation of nervous system development
regulation of neurogenesis
skeletal system morphogenesis
cranial skeletal system development
nucleoplasm part
chromatin
extracellular matrix
actin cytoskeleton
nuclear chromatin
vesicle membrane
collagen-containing extracellular matrix
collagen trimer
myosin complex
histone deacetylase complex
extracellular matrix component
Sin3-type complex
actin binding
protein-containing complex binding
transcription coregulator activity
actin filament binding
protein tyrosine phosphatase activity
extracellular matrix structural constituent
ATP-dependent microtubule motor activity
(C) 7 dpf vs. 6 dpf
Number of genes
chemical homeostasis
immune response
cellular homeostasis
growth
developmental growth
metal ion homeostasis
cellular chemical homeostasis
negative regulation of protein metabolic process
extracellular structure organization
live development
hepaticobiliary system development
extracellular matrix organization
leukocyte differentiation
bone development
extracellular matrix
actin cytoskeleton
side of membrane
sarcomere
contractile fiber part
myofibril
contractile fiber
I band
myosin complex
Z disc
bicellular tight junction
tight junction
collagen-containing extracellular matrix
endopeptidase activity
actin binding
actin filament binding
serine-type endopeptidase activity
serine-type peptidase activity
oxidoreductase activity, acting on paired donors
serine hydrolase activity
solutecation symporter activity
sodium ion transmembrane transporter activity
solutesodium symporter activity
glycosaminoglycan binding
carbohydrate transmembrane transporter activity
phospholipase A2 activity
pattern binding
(D) 8 dpf vs. 7 dpf
Number of genes
immune response
growth
cellular homeostasis
metal ion homeostasis
sterol homeostasis
cholesterol homeostasis
chitin metabolic process
regulation of cholesterol esterification
positive regulation of cholesterol esterification
plasma lipoprotein particle assembly
high-density lipoprotein particle assembly
protein-lipid complex assembly
protein-lipid complex remodeling
extracellular matrix
sarcomere
contractile fiber part
myofibril
contractile fiber
Z disc
collagen trimer
I band
striated muscle thin filament
high-density lipoprotein complex
myofilament
chylomicron
protein-lipid complex
plasma lipoprotein particle
lipoprotein particle
endopeptidase activity
carbohydrate binding
metallopeptidase activity
extracellular matrix structural constituent
exopeptidase activity
metallopeptidase activity
cargo receptor activity
sialic acid binding
scavenger receptor activity
carboxypeptidase activity
aminopeptidase activity
pattern binding
polysaccharide binding
metallocarboxypeptidase activity
phosphatidylcholine-sterol O-acyltransferase activator activity
chitin binding
(E) 9 dpf vs. 8 dpf
Number of genes
chitin metabolic process
plasma membrane region
synapse
synapse part
transmembrane transporter complex
transporter complex
cytosolic part
postsynapse
ribosome
ion channel complex
synaptic membrane
cytosolic ribosome
cytosolic ribosome
postsynaptic membrane
postsynaptic density
asymmetric synapse
neuron to neuron synapse
endopeptidase activity
metal ion transmembrane transporter activity
serine-type endopeptidase activity
serine-type peptidase activity
serine hydrolase activity
metallopeptidase activity
calcium ion transmembrane transporter activity
structural constituent of ribosome
voltage-gated cation channel activity
hydrolase activity, hydrolyzing O-glycosyl compounds
carboxylic ester hydrolase activity
chitin binding
phospholipase A2 activity
structural constituent of myelin sheath
ionotropic glutamate receptor binding
translation initiation factor binding
glutamate receptor binding
(F) 10 dpf vs. 9 dpf
Number of genes
regulation of transport
metal ion transport
regulation of ion transport
regulation of transmembrane transport
cellular potassium ion transport
potassium ion transmembrane transport
response to inorganic substance
response to metal ion
cellular response to calcium ion
response to calcium ion
dopamine secretion
regulation of dopamine secretion
dopamine transport
plasma membrane protein complex
supramolecular fiber
supramolecular complex
neuron projection
transporter complex
transmembrane transporter complex
ion channel complex
cation channel complex
synapse
synapse part
axon
cell projection membrane
voltage-gated potassium channel complex
calcium channel complex
channel activity
passive transmembrane transporter activity
ion gated channel activity
gated channel activity
ion channel activity
substrate-specific channel activity
cation channel activity
metal ion transmembrane transporter activity
voltage-gated ion channel activity
voltage-gated channel activity
voltage-gated cation channel activity
motor activity
microtubule motor activity
calcium-dependent phospholipid binding
ATP-dependent microtubule motor activity

Figure 6. Gene ontology functional enrichment analysis of differentially expressed genes (DEGs) at different development stages in channel catfish, *Ictalurus punctatus*. (**A**) 5 days post fertilization (dpf) vs. 2 dpf; (**B**) 6 dpf vs. 5 dpf; (**C**) 7 dpf vs. 6 dpf; (**D**) 8 dpf vs. 7 dpf; (**E**) 9 dpf vs. 8 dpf; (**F**) 10 dpf vs. 9 dpf. The vertical axis indicates the number of DEGs between two sampling datasets, and the horizontal axis represents the GO terms significantly enriched by the DEGs.

2.4.5 Construction of gene co-expression networks

To obtain a comprehensive understanding of gene co-expression relationships in development and to characterize the genes that are highly associated with embryogenesis and organogenesis of channel catfish, weighted gene co-expression network analysis (WGCNA) approach was applied to the FPKM data resulting from RNA-Seq differential expression analysis. After removing redundant genes, there were 8,504 genes retained for further WGCNA analysis. The best soft thresholding was determined when the degree of independence was 0.8 (Supplementary Figure 3). Then the WGCNA algorithm was used to detect the co-expression modules, co-expression network was constructed relying on the assumption that highly cooperating genes were clustered into one module and contributing to the corresponding phenotype. In total, 12 distinct modules were identified and assigned different module colors (Figure 7). The interaction of these 12 co-expression modules is shown in Figure 8.

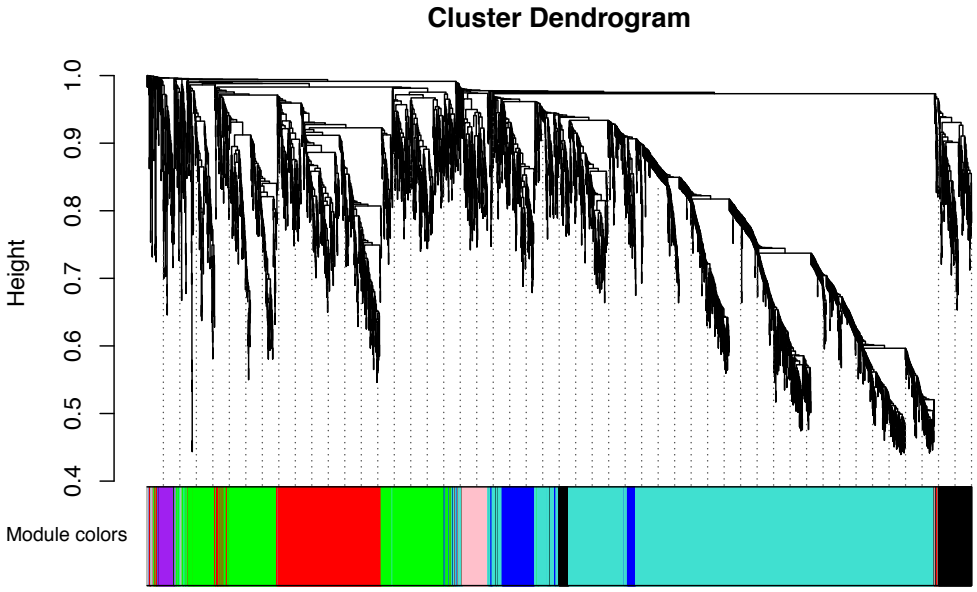


Figure 7. Hierarchical clustering dendrogram of channel catfish, *Ictalurus punctatus,* genes with dissimilarity based on topological overlap during early development. Each single leaf in the tree represents a single gene, the major tree branches constitute 12 distinct modules and are shown in different colors.

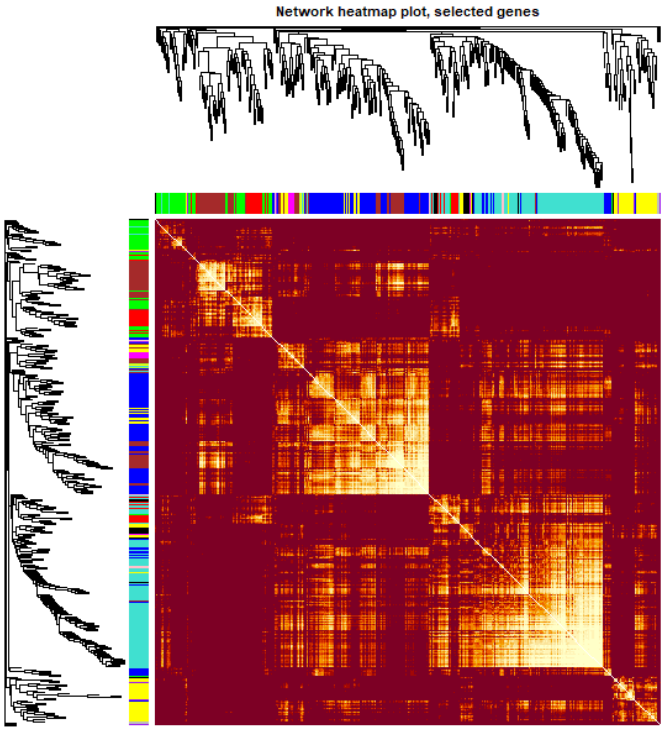


Figure 8. Heatmap plot of the gene network in channel catfish, *Ictalurus punctatus*. The heatmap shows the Topological Overlap Matrix (TOM) among all genes in the analysis. Light color represents low adjacency, and darker color represents high adjacency. The left and top sides indicate the gene dendrogram and module assignment.

2.4.6 Gene co-expression modules correspond to channel catfish early development

The 12 modules correlated with distinct developmental stages due to stage-specific profiles, and module -trait associations are shown in Figure 9. On the basis of those selection criteria ($p \leq 0.01$), eight modules (turquoise, black, blue, pink, green, grey, purple, brown modules) of interest were selected. The turquoise module, with 2,227 identified genes, was highly associated with 2 dpf stage ($r = 0.99$, $p = 1e\text{-}10$). The black, blue and pink modules were all associated with 2 dpf stage and contained 255, 1,916, 142 genes, respectively (black module: $r = -0.74$, $p = 0.002$; blue

module: r = -0.78, p = 0.001; pink module: r = - 0.68, p = 0.007). The green module, containing 910 genes, was highly associated with 5 dpf stage (r = 0.77, p = 0.001). The grey module, with 22 verified genes, was highly associated with 6 dpf stage (r = 0.88, p = 3e-05). The purple module, representing 82 genes, was highly associated with 9 dpf stage (r = 0.76, p = 0.002). The brown module, containing 1,174 genes, was highly related to 10 dpf stage (r = 0.67, p = 0.009). Correlations between the modules and the developmental traits were quantified and listed in Table 2, and all genes present in those modules are presented in Supplementary Table 4.

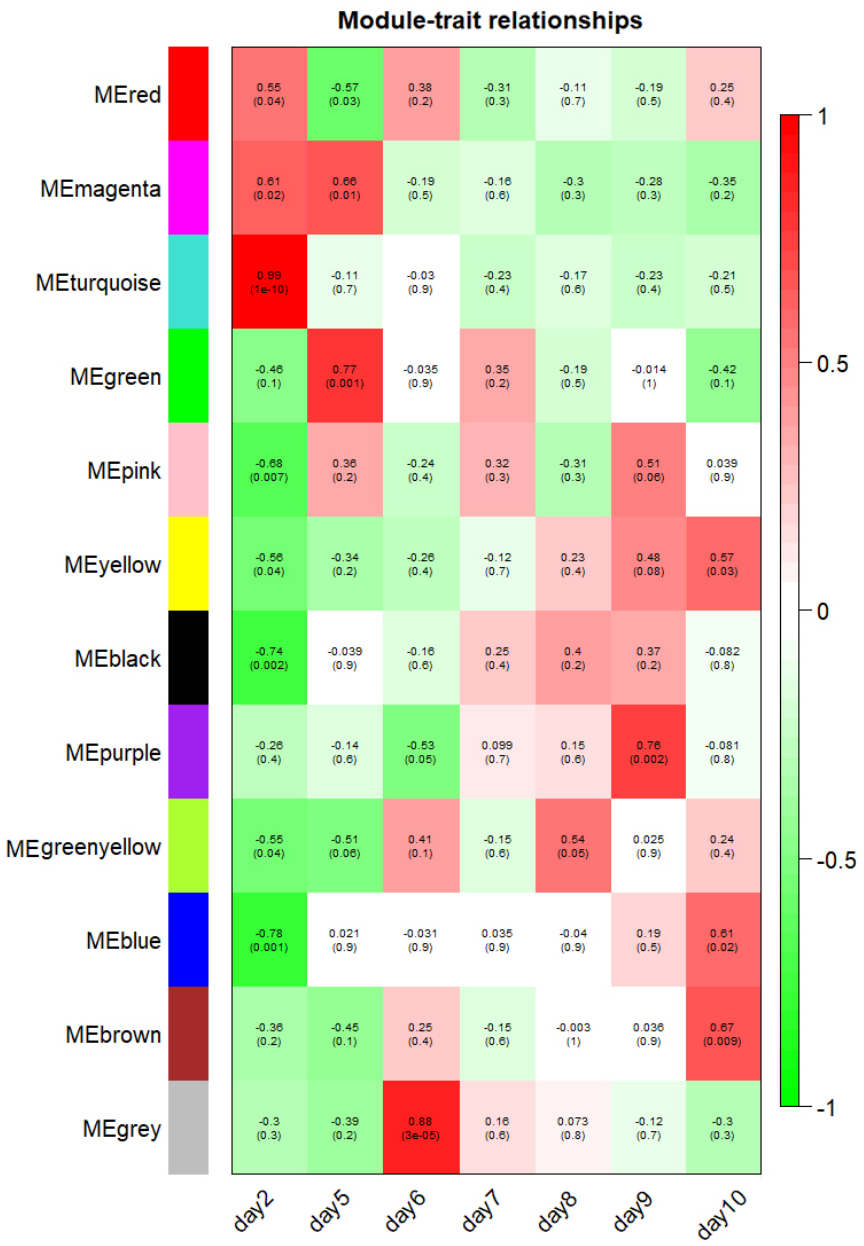


Figure 9. Module-stage relationships (MSRs) in channel catfish, *Ictalurus punctatus*. Each row corresponds to a module, and each column represents a specific development stage. The right color panel represent Pearson's r correlation coefficient. The MSRs are colored based on the correlation coefficient between the module and the developmental stages. The Pearson's r correlation coefficients and associated p-values are given in each cell.

2.4.7 Functional enrichment of genes in the eight selected modules

To investigate the biological functions related to channel catfish in early development, eight modules potentially associated with early developmental stages were selected for Gene Ontology (GO) enrichment analysis (Supplementary Table 4). Significant Gene Ontology (GO) terms in the selected modules were identified using ClusterProfiler R package (version 3.6).

The turquoise module was most correlated with the 2 dpf stage. The genes in this module were mainly enriched in four categories: development (muscle cell development, muscle structure development, cell migration involved in heart development, hematopoietic or lymphoid organ development), cell cycle process (cell proliferation, DNA replication, mitotic cell cycle process, meiotic cell cycle process), reproduction process (cellular process involved in reproduction in multicellular organism) and many transport activities (organic acid transmembrane transport, carboxylic acid transmembrane transport, amino acid transport, organic anion transmembrane transporter activity).

The genes in black, blue and pink modules were also most strongly correlated with 2 dpf stage, but they showed a negative correlation relationship with developmental stages. Genes in black module were enriched into proteasomal activity (proteasomal ubiquitin-independent protein catabolic process, proteasome complex), endopeptidase complex and endopeptidase activity. Genes in blue module were mainly related to synaptic signaling (chemical synaptic transmission, anterograde trans-synaptic signaling, regulation of trans-synaptic signaling), ion

transport (metal ion transport, cation transmembrane transport, calcium ion transmembrane transport), cell and tissue morphogenesis (cell part morphogenesis, cell morphogenesis involved in neuron differentiation, muscle tissue morphogenesis), heart process (regulation of heart rate, heart contraction, regulation of heart contraction) and nervous system development (regulation of nervous system development, nervous system process, positive regulation of nervous system development). Genes in pink module were mainly associated with cytosolic processes (cytoplasmic vesicle membrane, cytoplasmic vesicle part, cytosolic small ribosomal subunit), and vesicle membrane processes (synaptic vesicle membrane, exocytic vesicle membrane, transport vesicle membrane).

The green module genes were most correlated with 5 dpf stage. Genes in this module were enriched in processes expected to be very active during development (cartilage development, skeletal system development), proliferation (positive regulation of mononuclear cell proliferation, positive regulation of T cell proliferation, positive regulation of lymphocyte proliferation, positive regulation of leukocyte proliferation) and morphogenesis (bone morphogenesis).

The grey module genes were most correlated with 6 dpf stage. This module mainly included Gene Ontology categories of extracellular matrix (extracellular matrix component, collagen-containing extracellular matrix), binding activity (GTP binding, purine nucleotide binding, carbohydrate-binding) and symporter activity (potassium:chloride symporter activity, cation:chloride symporter activity, anion:cation symporter activity).

The purple module was most correlated with the 9 dpf stage. Genes in this module were enriched in the differentiation (regulation of myeloid leukocyte differentiation, regulation of myeloid cell differentiation), homeostasis (cellular calcium ion homeostasis, cellular divalent inorganic cation homeostasis, ion homeostasis) and some transmembrane activity (urea transmembrane transporter activity, water transmembrane transporter activity).

The brown module was most related to 10 dpf stage. Genes in this module were most enriched in neuron development (regulation of neuron projection development, neuron projection development, neural crest cell development), morphogenesis of different tissues and organs (cell morphogenesis involved in neuron differentiation, cell morphogenesis involved in differentiation, sensory organ morphogenesis, dendrite morphogenesis, inner ear morphogenesis).

Notably, the enriched Gene Ontology terms for these eight modules were in agreement with our previous GO enrichment analysis and morphology observations, which related the early development with many kinds of tissues and organ differentiation, regulation of ion transportation, cell proliferation and transmembrane activities.

2.4.8 Protein-protein interaction network construction and analysis of selected modules

To further identify the function of the co-expressed genes within each module and investigate the hub genes, Cytoscape (version 3.7.2) software was used to construct a co-expression network of the top 200 ranked genes for eight selected modules (Supplementary

Table 5), including the turquoise, green, grey and purple modules. Notably, the pink, grey and purple modules only have 142, 22 and 82 genes separately, so all genes in these three modules were selected for co-expression network construction. The highest degree genes (hub genes) were illustrated with bigger size and specific color (Figure 10). For example, GDF10, FOXA2, HCEA and SYCE3 were identified as hub genes in the turquoise module. VGLL3, CELSR2 and SCARA3 were identified as hub genes in black module. ASTN1 and GAD2 were identified to be key hub genes in blue module. ARF1, NDE1 and RHOA genes were hub genes for pink module. CK1, DARS1, UBE2V2 and OAZ2 were identified as hub genes in green module. In grey module, IFI44L and ZIP10 were recognized as hub genes. TGFB1 and TCK1 were verified as hub genes in purple module. Similarly, KCNT1 and KCNC were identified as hub genes in brown module. These hub genes and their descriptions were listed in Table 3.

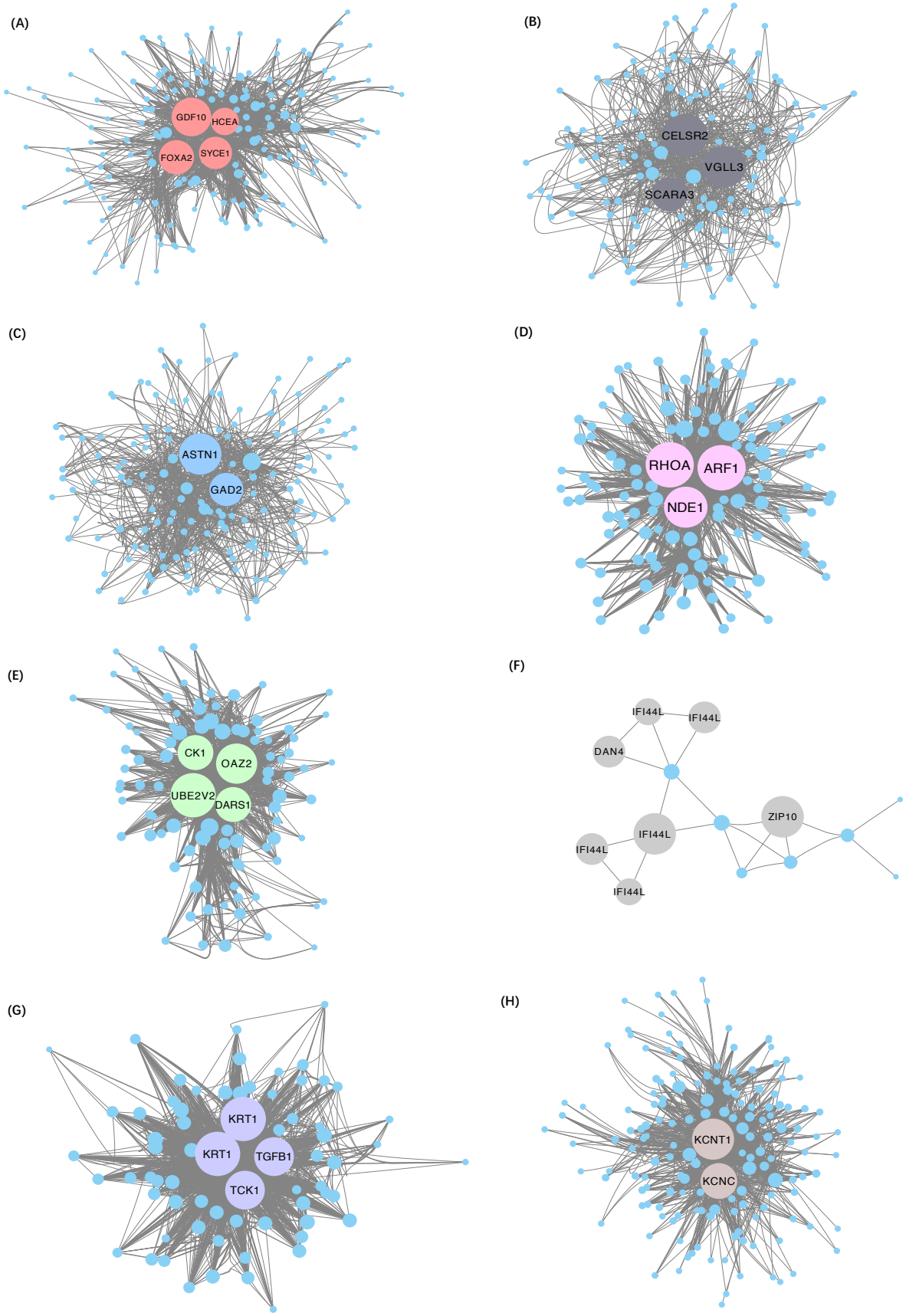

(A)
GDF10
HCEA
FOXA2
SYCE1
(B)
CELSR2
VGLL3
SCARA3
(C)
ASTN1
GAD2
(D)
RHOA
ARF1
NDE1
(E)
CK1
OAZ2
UBE2V2
DARS1
(F)
IFI44L
IFI44L
DAN4
ZIP10
IFI44L
IFI44L
IFI44L
(G)
KRT1
KRT1
TGFB1
TCK1
(H)
KCNT1
KCNC

Figure 10. Protein-protein interaction (PPI) network in channel catfish, *Ictalurus punctatus,* predicted by Cytoscape. The node degree of genes was represented using circumference of nodes. The genes with red, green, grey, purple color represent the hub genes in (**A**) turquoise module; (**B**) black module; (**C**) blue module; (**D**) pink module; (**A**) green module; (**B**) grey module; (**C**) purple module; (**D**) brown module separately.

Table 2. Correlation of module with development trait in channel catfish (*Ictalurus punctatus*). Four co-expression modules (turquoise, black, blue, pink, green, grey, purple, brown) were identified significantly corrected to channel catfish early development status, the number of genes in each module are listed here.

Module color	Number of genes	Correlation (r)	P-velue
Turquoise	2, 228	0.99	$1x10^{-10}$
Black	255	-0.74	0.002
Blue	1,916	-0.78	0.001
Pink	142	-0.68	0.007
Green	910	0.77	$1x10^{-3}$
Grey	22	0.88	$3x10^{-5}$
Purple	82	0.76	$2x10^{-3}$
Brown	1,173	0.67	0.009

Table 3. Hub genes of selected co-expression modules in channel catfish, *Ictalurus punctatus,* predicted by Cytoscape.

Module	Gene ID	Gene name	Description
Turquoise	108262421	GDF10	growth/differentiation factor 10-like
Turquoise	108257557	FOXA2	hepatocyte nuclear factor 3-beta-like
Turquoise	108259527	HCEA	high choriolytic enzyme 1-like
Turquoise	108268486	SYCE3	synaptonemal complex central element protein 3
Black	108266139	VGLL3	vestigial like family member 3
Black	108281038	CELSR2	cadherin EGF LAG seven-pass G-type receptor 2-like
Black	108258146	SCARA3	scavenger receptor class A member 3
Blue	108267173	ASTN1	astrotactin 1
Blue	108272908	GAD2	glutamate decarboxylase 2
Pink	108255676	ARF1	ADP-ribosylation factor 1-like
Pink	108277500	NDE1	nudE neurodevelopment protein 1
Pink	108254710	RHOA	transforming protein RhoA-like
Green	108274016	CK1	casein kinase I
Green	108266474	DARS1	aspartyl-tRNA synthetase 1
Green	108264029	UBE2V2	ubiquitin conjugating enzyme E2 variant 2
Green	108279393	OAZ2	ornithine decarboxylase antizyme 2-like
Grey	108266103	IFI44L	interferon-induced protein 44-like
Grey	108264165	ZIP10	zinc transporter 10-like
Purple	108279384	TGFB1	transforming growth factor beta-1-like
Purple	108262114	TCK1	creatine kinase, testis isozyme-like

Brown	108256037	KCNT1	potassium channel subfamily T member 1-like
Brown	108275314	KCNC	potassium voltage-gated channel subfamily C member 1

2.5 Discussion

Gene expression changes are complex during the transition from fertilized eggs to larvae, and the transcriptome profiles underlying these events have not been fully studied in teleost fish. The main objectives of this study were to utilize a transcriptome sequencing method to analyze the expression of seven early developmental stages in channel catfish and to construct a gene co-expression network involved in embryogenesis. In this study, we investigated a set of DEGs in each developmental stage by comparing each of the two continual stages, and found that the most numbers of DEGs occurred at 5-7 days post fertilization. The vast differences in transcript expression illustrate that these stages are of rapid, critical, expansive development in channel catfish, in accordance with morphological change observation. Gene ontology enrichment analysis revealed that during early embryogenesis, the most enriched gene ontology categories were related to development, growth, differentiation and morphogenesis, especially during 5-7 dpf. These stages are critical for the development of muscle, nerves, bone and other tissues, and cell differentiation during transformation from fertilized eggs to larvae.

Originally, WGCNA was a systems biology method conceived for exploring and describing correlation patterns among genes across microarray data [10]. RNA-sequencing (RNA-seq) is a relatively novel approach to profile the transcriptome at varieties of conditions, which can obtain the complete transcriptomic information and allow for far more extensive analyses when compared with microarray techniques [24]. However, for larger datasets, the disadvantage of RNA-seq transcriptome analysis are obvious, as grouping gene expression patterns with similar upregulation or downregulation pattern are cumbersome; furthermore, exploring gene function and constructing gene interaction network is largely dependent on known knowledge of model species from a notably accredited database, such as KEGG, which has greatly limited the potential of the prediction. WGCNA is a powerful method to identify co-expressed groups of genes from large RNA expression data sets [25], and it is widely used to explore the correlation among transcriptomic datasets, identify hub gens and find new pathways in both model and non-model species [26-28]. WGCNA has proven its superiority over partial correlation method and provided a powerful tool in identifying higher-order correlation in complexed traits of interest, by presenting simplified network on the integrated function of gene modules [29, 30].

The power of WGCNA fuels investigation of correlations, hub genes and novel pathways among different early developmental stages of channel catfish. However, no study has been done for the early development of channel catfish using WGCNA or at the transcriptome level. Previous study mainly focused on the physical development and anatomy. The first study related to the embryonic development of channel catfish was conducted from 2 days post fertilization

(dpf) through 11 dpf with a focus on the organogenesis of pronephors [31, 32]. Later on, a more detailed description when sampling was performed at 1-hr intervals after fertilization on embryogenesis was documented up to 34 hours, which corresponded to the pectoral fin-buds at water temperature from 24.7 to 26.8 °C [33]. Specific developmental stages of channel catfish embryos were for the first time defined based on the development of vascular system at 26 °C, all the stages were characterized before 5 dpf [34]. Makeeva et al. reported that stage IV (1 dpf) is mainly associated with gastrulation, stage V(~1.5 dpf) related to organogenesis, stage VI (~2 dpf) having formation of gill microstructure, eye lens, auditory vesicles and segmentation of tail, and at the end of the second day of development, they observed heartbeat of the embryo, also the rotation of embryo within the egg case, due to its tail whipping back and forth. stage VII (~ 3 dpf) basically associated with development of vascular system, development in the brain and an intestinal cavity can also be observed at this stage [34]. Our research in general agreed well with those earlier studies, the functional annotation of DEGs between 5 dpf and 2 dpf revealed that the enriched terms were most associated with muscle tissue development, bone growth, morphogenesis of different organs, chondrocyte development, heart contraction, blood coagulation and regulation of blood circulation. However, our morphology observations found that the embryo tails cannot be observed until 5 dpf, which is slightly later than previous observation in channel catfish embryonic development [34]. Our fertilized eggs were incubated at 25 to 26 degrees Celsius (mainly 25 degree Celsius) which is not that much difference than

them, maybe the slower development for our channel catfish than them is due to the strain or water quality, both could lead to the inconsistencies in development speed [34].

In this study, for the first time, we used a gene co-expression network to investigate the transcripts of the embryos and fry in early different developmental stages of channel catfish. Using contrasted biological samples at different developmental stages of channel catfish embryos, four distinct modules were identified. The genes within the same intra-module were used to perform Gene Ontology enrichment analysis, the genes most associated with 2 dpf, 5 dpf, 6 dpf and 9 dpf stages were identified to enriched in development, proliferation, morphogenesis and differentiation categories, such as muscle structure development, hematopoietic or lymphoid organ development, bone morphogenesis, regulation of myeloid cell differentiation categories. Then the Cytoscape was employed to build a PPI network for the four selected modules, and the high degree of genes (hub genes) was verified to have an essential role in the co-expression network.

2.5.1 Turquoise module (hub genes and node genes)

For the module most associated with the 2 days post-fertilization stage (turquoise module), the hub genes were GDF10, FOXA2, HCEA and SYCE3. GDF10 is a growth differential factor belonging to the TGC-beta (transforming growth factor beta) superfamily, and functions predominantly in bone development [35]. Its pathways are p70S6K signaling and activation of cAMP-dependent PKAGDFs [36, 37]. GDF10 is necessary for head formation, skeletal morphogenesis, and adipogenesis [38-40].

FOXA2 is a transcription factor involved in embryonic development [41], and regulation of gene expression in differentiated tissues and development of multiple endoderm-derived organ systems, such as liver, gland, pancreas and lungs [42-45]. This gene is related to pathways that include heart developmental, Hedgehog signaling events mediated by Gli proteins and cardiac progenitor differentiation [36, 37].

HCEA participates in the breakdown process of the egg envelope, which is derived from the egg extracellular matrix [46, 47]. HCEA has a typical neutral zinc metallopeptidase domain that is involved in the binding of zinc and proteolysis [48].

Another candidate hub gene for the 2 dpf module is SYCE3. It's associated pathways include the cell cycle, mitosis and meiosis [36, 37]. SYCE3 is a significant component of the transverse central element of synaptonemal complexes (SCS), formed between homologous chromosomes during meiotic prophase. This gene is also required for chromosome loading of the central element-specific SCS proteins, and for initiating synapsis between homologous chromosomes as well as required for fertility [49, 50]. The 2-dpf stage is a critical time window related to tissue differentiation, morphogenesis and different organ development, which requires a series of concerted meiosis, mitosis, synapsis activity, and SYCE3 may play an essential role regulating early development through these functions.

The rest of the 2,223 node genes in turquoise module can be found in the supplementary table 3 with gene name and gene ID; the Zebrafish Information Network (ZFIN) network has assembled

and documented function of genes in model species - zebrafish, initial investigation of any of those node genes can be retrieved from ZFIN [51].

2.5.2 Black module

Black module is most associated with 2 dpf, hub genes in this module included VGLL3, CELSR2 and SCARA3. The VGLL proteins are transcriptional co-factors to influence myogenesis in skeletal muscle as well as maturity. Figeac et al. reported that VGLL3 could contribute to muscle fibre composition in mice, knock out of VGLL3 gene in mice supressed myoblast prolifiration, conversely, the overexpression of VGLL3 could highly increase myogenic differentiation. This research stated that VGLL3 as a transcriptional co-factor working with the Hippo signal transduction to control myogenesis [52]. Also, VGLL3 is verified to be linked with age at maturity in Atlantic salmon (*Salmo salar*). Vgll3 and the interrelated Hippo pathway has been reported to be linked to decreased proliferation process in different tissues, and might play a negative role on sertoli cell proliferation in testis and thus compressing the growth of pubertal testis [53].

CELSR 1-3 expression started broadly in the nervous system in early developmental stages, also, these genes were found to be expressed in other organs, such as the cochlea, the kidney, and the whisker. CELSR2 protein was fould to distributed along intercellular boundaries in the whisker and related to neuronal cells [54]. Also, CLESR2 gene was reported to play a role in the

regulation of facial motor neurons migration in neuroepithelium during the developing zebrafish hindbrain [55].

SCARA3 belongs to the class A scavenger receptors (SR-As) family, which is identified to be functioning with the innate immunity in mammals [56], this gene was also identified to potentially contribute to the immunity in rainbow trout (*Oncorhynchus mykiss*) [56]. Further research is needed to see if this gene also functioned as a immunity regulator in all teleost fish.

2.5.3 Blue module

For the blue module, which is more correlated with 2 dpf, hub genes in this module contained ASTN1 and GAD2. The ASTN1 works as a neuron-glial ligand for CNS glial-guided migration [57]. ASTN1 is necessary for normal migration of young postmitotic neuroblasts along glial fibers. It also plays an important role for the migration of granule cells during brain development [57, 58]. In zebrafish, GAD2 has been identified to play role on the dorsal hindbrain development [59]. GAD2 is also involved in the neurotransmitter release cycle and beta -alanine metabolism pathway [36].

2.5.4 Pink module

For the pink module, which is more correlated with 2 dpf, hub genes in this module are: ARF1, NDE1 and RHOA. ARF1 is considered to be coupled with CDC42 to regulate the endocytosis in plasma membrane [60]. The association of ARF1 and membrane is regulated by

the endosomal PH, which controls the PH-dependent association of endosomal COPs. ARF1 could influence COP function through endocytic pathway, which may suggest that ARF1 might act as the cytosolic component for a transmembrane PH-sensing mechanism.

NDE1 is reported to play an essential role for centrosome duplication and mitotic spindle assembly. The function and orientation of the mitotic spindle is critical for normal cerebral cortex development in mammals [61] [62]. RHOA encodes a member of the Rho family of small GTPases, which could promote the reorganization of the actin cytoskeleton as well as regulate the shape and motility [63].

2.5.5 Green module

For the green module, which is more correlated with 5 dpf, hub genes include CK1, DARS1, UBE2V2 and OAZ2. The Casein kinase 1 (CK1) family of protein kinases are serine/threonine-selective enzymes, which function as key regulatory molecules involved in the cell cycle, DNA repair, transcription, translation, the structure of the cytoskeleton, cell-cell adhesion and receptor-coupled signal transduction [64, 65]. CK1 is also involved in the Wnt signaling pathway [36]. Wnt signaling pathway plays an important role in lung organogenesis, the initial formation of the neural plate and many subsequent patterning decisions in the embryonic nervous system [66, 67]. Wnt signaling pathway also works with other signaling systems as primary molecular mechanisms that control embryonic development, and regulate processes such as cell proliferation, survival, or differentiation [68].

DARS1 is critical to the processes of tRNA aminoacylation, selenoamino acid metabolism as well as gene expression [36, 37]. The DARS1 gene encodes a member of a multienzyme complex which catalyzes the attachment of an amino acid (AA) to its connate tRNA in a two-step reaction. The amino acid is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA [69]. Although DARS1 is considered to be expressed in all organs, it has a distinct expression pattern in the brain. Dars - knocked out mice were not viable and displayed early developmental arrest and associated with embryonic lethality [70, 71]. Mutations of DARS1 and its homolog DARS2 have been reported in patients showing hypomyelination in the brainstem, spinal cord and leg spasticity (HBSL), and leukoencephalopathy brain stem and spinal cord involvement and elevated lactate (LBSL), which demonstrates that mutation in tRNA causes a similar disease and shares a common mechanism of neurological pathology [72].

UBE2V2 is thought to be involved in the differentiation of monocytes and enterocytes [73], and it may also play a role in progression through the cell cycle, as well as differentiation [50, 74]. Among its related pathways are DNA double-strand break repair, and class I MHC mediated antigen processing and presentation [36, 37].

OAZ2 plays a role in cell growth and proliferation by regulating intracellular polyamines [50, 75]. Its related pathways are CDK-mediated phosphorylation, and removal of cdc6 and metabolism [36, 37].

2.5.6 Grey module

Another module of interest was grey module, which was most strongly associated with 6 dpf. Hub genes in this module included IFI44L and ZIP10. IFI44L is reported to be associated with virus infection and immune activity, as well as the formation of microtubular structures [76-78]. This gene has not been verified to be closely related to early embryonic development; however, our analysis predicted from the PPI network indicates that IFI44L plays an essential role in this module, illustrating that IFI44L may be related to early immune response and the survival of embryos, and thus, contribute to the early development mechanisms.

The other hub gene in this module is ZIP10, which controls the membrane transport of zinc, calcium, manganese and regulates their intracellular and cytoplasmic concentrations [79].

Functions of most other node genes, i.e. SLC12A8, MTFR1L, Tatdn2, Agxt, MX2, nd4l and nd6, have not been experimentally documented in embryogenesis and somitogenesis of fishes [51]. SLC12A8 (solute carrier family 12 member 8) is an important paralog gene of SLA12A2, Gene Ontology (GO) annotations assign this gene possible function that includes ATPase activity, coupled to transmembrane movement of substances and symporter activity. MTFR1L (Mitochondrial Fission Regulator 1 Like) is a paralog of MTFR2, which may play a role in mitochondrial aerobic respiration in the testis. It also promotes mitochondrial fission. MX2 (MX dynamin like GTPase2) is involved in innate immune system [52].

The related pathways of Tatdn2 (TatD DNase Domain Containing 2) are unfolded protein response (UPR) and metabolism of proteins. The GO annotations related to this gene include deoxyribonuclease activity and endodeoxyribonuclease activity. Agxt (Alanine--Glyoxylate and Serine-Pyruvate Aminotransferase) are predicted to have alanine-glyoxylate transaminase activity and serine-pyruvate transaminase activity. Both nd4l (ND4L-NADH dehydrogenase, subunit 4L) and nd6 are involved in respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins [53].

2.5.7 Purple module

Purple module is correlated with 9 dpf. This module contains TGFB1 and TCK1. TGFB1 can regulate cell proliferation, differentiation of various cell types, and function in other processes such as normal development, immune function and responses to neurodegeneration [50, 80, 81]. Among its related functions are transcription androgen receptor nuclear signaling and p70S6K signaling pathway [36, 37].

TCK1 reversibly catalyzes the transfer of phosphate between ATP and various phosphagens, and also plays a crucial role in tissues with a lot of energy requirements, such as in skeletal muscle, heart, brain and spermatozoa [50, 82, 83].

The other node genes (Supplementary Table 3): AQP7, AQP8, cldn11, MMP9 and clec14a may be involved in the somitogenesis and normal function of organs during early development. AQP7 (Aquaporin 7) in zebrafish is maternally inherited and detected at the 2-4 cell and morula

stages [84]; while other paralogs such as AQP8aa are related to somitogenesis and vascular development [85]. Cldn11a (Claudin 11) expression is detected in vascular endothelium, adjacent to the optic stalk of embryo in normal retinal embryo [86, 87]; cldn11 expression requires zic2 function in the differentiating mammalian cerebellar ganglion cells [88]. MMP-2 (Matrix Metallopeptidase 2), MMP-9 and MMP-13 are necessary for proper zebrafish craniofacial morphogenesis as morpholino knockdown of these genes shows huge alterations in both the mandibular and hyoid arches concurrently [89]. By interacting with Etv2 and Vegf signaling, Clec14a (C-lectin family 14 Member A) in zebrafish is indispensable for function of vascular endothelia cells during vasculogenesis and angiogenesis as knockdown Etv2/Vegf results in an inhibition of intersegmental vessel sprouting [90].

2.5.8 Brown module

The last module is brown module, which is correlated with 10 dpf. The hub genes in this module contain KCNT1 and KCNC. KCNT1 gene belongs to the potassium channel family, which is considered to regulate ion flux. It could also interacts with cytoplasmic peoteins related to developmental signaling pathways [91]. KCNC gene belongs to a potassium voltage-gated channel family, which is critical for the rapid repolarization of fast firing brain neurons [36]. In response to the voltage across the membrane, the channel opens and forming a potassium-selective channel, potassium ions in accordance with their electrochemical gradient could pass the channel [36, 92].

2.6 Conclusions

In conclusion, a comprehensive transcriptome analysis was used to explore the dynamic changes during channel catfish early development. This study provides genomic data that has great promise for improved understanding of channel catfish embryogenesis. A co-expression network was constructed using WGCNA method. As a result of that effort, 12 modules were verified, eight of them were identified to be closely associated with channel catfish early development. Further analysis of these eight selected modules revealed that they were enriched in biological processes related to development, morphogenesis, growth and differentiation. Five and 6 days post fertilization embryos contained the most strongly differentially expressed genes (DEGs). Gene Ontology (GO) analysis revealed that enriched categories at 5 and 6 dpf were most related to embryonic development, morphogenesis, differential and formation of different organs. In addition, these stages display the most striking changes in morphology. Thus, day 5 and 6 are likely to be critical turning point in the progression from fertilized egg to larva in channel catfish. Hub genes identified within each module are likely to direct critical roles during the development and growth processes in channel catfish. Taken together, our results are a first stage in shedding light on the complex biological processes that take place during early development. Our work provides a useful genetic resource for future studies on the metabolism, growth and genetic enhancement programs of channel catfish. Further research should address gene quantification and

genetic behavior. Gene editing technology will be used to confirm the function of these genes in WGCNA network.

Reference

1. Eschmeyer, W.N., et al., *Marine fish diversity: history of knowledge and discovery (Pisces).* Zootaxa, 2010. **2525**(1): p. 19-50.

2. Tan, Y. and S.K. Chang, *Isolation and characterization of collagen extracted from channel catfish (Ictalurus punctatus) skin.* Food chemistry, 2018. **242**: p. 147-155.

3. Vesterlund, L., et al., *The zebrafish transcriptome during early development.* BMC developmental biology, 2011. **11**(1): p. 30.

4. Sharov, A.A., et al., *Transcriptome analysis of mouse stem cells and early embryos.* PLoS biology, 2003. **1**(3).

5. Hartl, D., et al., *Transcriptome and proteome analysis of early embryonic mouse brain development.* Proteomics, 2008. **8**(6): p. 1257-1265.

6. Graveley, B.R., et al., *The developmental transcriptome of Drosophila melanogaster.* Nature, 2011. **471**(7339): p. 473-479.

7. Liu, Z., et al., *The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts.* Nature communications, 2016. **7**: p. 11757.

8. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nature Reviews Genetics, 2011. **12**(10): p. 671.

9. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature reviews genetics, 2009. **10**(1): p. 57.

10. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC bioinformatics, 2008. **9**(1): p. 559.

11. Deng, S.-P., L. Zhu, and D.-S. Huang. *Mining the bladder cancer-associated genes by an integrated strategy for the construction and analysis of differential co-expression networks*. in *BMC genomics*. 2015. BioMed Central.

12. Miller, J.A., S. Horvath, and D.H. Geschwind, *Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways.* Proceedings of the National Academy of Sciences, 2010. **107**(28): p. 12698-12703.

13. Xue, Z., et al., *Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing.* Nature, 2013. **500**(7464): p. 593-597.

14. Smitherman, R.O., R.A. Dunham, and D. Tave, *Review of catfish breeding research 1969–1981 at Auburn University.* Aquaculture, 1983. **33**(1-4): p. 197-205.

15. Backenstose, N., *A Histological Evaluation of the Development of Respiratory Structures in Channel Catfish (Ictalurus punctatus) and Tra (Pangasianodon hypophthalmus).* 2018.

16. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.* The ISME journal, 2012. **6**(8): p. 1621.

17. Wingett, S.W. and S. Andrews, *FastQ Screen: A tool for multi-genome mapping and quality control.* F1000Research, 2018. **7**.

18. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

19. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

20. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.

21. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.

22. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics: a journal of integrative biology, 2012. **16**(5): p. 284-287.

23. Smoot, M.E., et al., *Cytoscape 2.8: new features for data integration and network visualization.* Bioinformatics, 2010. **27**(3): p. 431-432.

24. Zhao, S., et al., *Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.* PloS one, 2014. **9**(1).

25. Clarke, C., et al., *Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis.* Carcinogenesis, 2013. **34**(10): p. 2300-2308.

26. Zhang, B. and S. Horvath, *A general framework for weighted gene co-expression network analysis.* Statistical applications in genetics and molecular biology, 2005. **4**(1).

27. Degli Esposti, D., et al., *Co-expression network analysis identifies gonad-and embryo-associated protein modules in the sentinel species Gammarus fossarum.* Scientific reports, 2019. **9**(1): p. 1-10.

28. Zhang, J., et al., *Weighted gene co-expression network analysis for RNA-sequencing data of the varicose veins transcriptome.* Frontiers in physiology, 2019. **10**: p. 278.

29. Zhao, W., et al., *Weighted gene coexpression network analysis: state of the art.* Journal of biopharmaceutical statistics, 2010. **20**(2): p. 281-300.

30. Kadarmideen, H.N. and N.S. Watson-Haigh, *Building gene co-expression networks using transcriptomics data for systems biology investigations: Comparison of methods using microarray data.* Bioinformation, 2012. **8**(18): p. 855.

31. Wellner, K., *" The Development of the Pronephros during the Embryonic and Early Larval Life of the Catfish (Ictalurus punctatus)"(1932), by Rachel L. Carson.* Embryo Project Encyclopedia, 2014.

32. Carson, R., *The Development of the Pronephros During the Embryonic and Early Larval Life of the Catfish (Ictalurus punctatus).* Master's Thesis (John Hopkins University, Baltimore, Maryland, 1932), 1932.

33. Saksena, V.P., C.D. Riggs, and K. Yamamoto, *Early development of the channel catfish.* The Progressive Fish-Culturist, 1961. **23**(4): p. 156-161.

34. Makeeva, A. and N. Emel'yanova, *Early development of the channel catfish, Ictalurus punctatus.* JOURNAL OF ICHTHYOLOGY C/C OF VOPROSY IKHTIOLOGII, 1993. **33**: p. 87-87.

35. Herpin, A., C. Lelong, and P. Favrel, *Transforming growth factor-β-related proteins: an ancestral and widespread superfamily of cytokines in metazoans.* Developmental & Comparative Immunology, 2004. **28**(5): p. 461-485.

36. Safran, M., et al., *GeneCards Version 3: the human gene integrator.* Database, 2010. **2010**.

37. Belinky, F., et al., *PathCards: multi-source consolidation of human biological pathways.* Database, 2015. **2015**.

38. Hino, J., H. Matsuo, and K. Kangawa, *Bone morphogenetic protein-3b (BMP-3b) gene expression is correlated with differentiation in rat calvarial osteoblasts.* Biochemical and biophysical research communications, 1999. **256**(2): p. 419-424.

39. Matsumoto, Y., et al., *Bone morphogenetic protein-3b (BMP-3b) inhibits osteoblast differentiation via Smad2/3 pathway by counteracting Smad1/5/8 signaling.* Molecular and cellular endocrinology, 2012. **350**(1): p. 78-86.

40. Angioni, M., et al., *Spa therapy induces clinical improvement and protein changes in patients with chronic back pain.* Reumatismo, 2019. **71**(3): p. 119-131.

41. Lavon, N., O. Yanuka, and N. Benvenisty, *The effect of overexpression of Pdx1 and Foxa2 on the differentiation of human embryonic stem cells into pancreatic cells.* Stem Cells, 2006. **24**(8): p. 1923-1930.

42. Wederell, E.D., et al., *Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing.* Nucleic acids research, 2008. **36**(14): p. 4549-4564.

43. Jeong, J.-W., et al., *Foxa2 is essential for mouse endometrial gland development and fertility.* Biology of reproduction, 2010. **83**(3): p. 396-403.

44. Lee, C.S., et al., *Foxa2 controls Pdx1 gene expression in pancreatic β-cells in vivo.* Diabetes, 2002. **51**(8): p. 2546-2551.

45. Wan, H., et al., *Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis.* Journal of Biological Chemistry, 2005. **280**(14): p. 13809-13816.

46. Kawaguchi, M., et al., *Evolution of teleostean hatching enzyme genes and their paralogous genes.* Development genes and evolution, 2006. **216**(12): p. 769.

47. Hishida, R., et al., *hch-1, a gene required for normal hatching and normal migration of a neuroblast in C. elegans, encodes a protein related to TOLLOID and BMP-1.* The EMBO journal, 1996. **15**(16): p. 4111-4122.

48. Rawlings, N.D. and A.J. Barrett, *[13] Evolutionary families of metallopeptidases*, in *Methods in enzymology*. 1995, Elsevier. p. 183-228.

49. Schramm, S., et al., *A novel mouse synaptonemal complex protein is essential for loading of central element proteins, recombination, and fertility.* PLoS genetics, 2011. **7**(5): p. e1002088.

50. Consortium, U., *UniProt: a worldwide hub of protein knowledge.* Nucleic acids research, 2018. **47**(D1): p. D506-D515.

51. Ruzicka, L., et al., *The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources.* Nucleic acids research, 2019. **47**(D1): p. D867-D873.

52. Figeac, N., et al., *Vgll3 operates via Tead1, Tead3 and Tead4 to influence myogenesis in skeletal muscle.* Journal of cell science, 2019. **132**(13): p. jcs225946.

53. Kjærner-Semb, E., et al., *Vgll3 and the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon testis.* Scientific reports, 2018. **8**(1): p. 1-11.

54. Shima, Y., et al., *Differential expression of the seven-pass transmembrane cadherin genes Celsr1-3 and distribution of the Celsr2 protein during mouse development.* Developmental dynamics: an official publication of the American Association of Anatomists, 2002. **223**(3): p. 321-332.

55. Wada, H., et al., *Frizzled3a and Celsr2 function in the neuroepithelium to regulate migration of facial motor neurons in the developing zebrafish hindbrain.* Development, 2006. **133**(23): p. 4749-4759.

56. Poynter, S., A. Monjo, and S. DeWitte-Orr, *Identification of three class A scavenger receptors from rainbow trout (Oncorhynchus mykiss): SCARA3, SCARA4, and SCARA5.* Fish & shellfish immunology, 2018. **76**: p. 121-125.

57. Wilson, P.M., et al., *Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration.* Journal of Neuroscience, 2010. **30**(25): p. 8529-8540.

58. Ni, T., K. Harlos, and R. Gilbert, *Structure of astrotactin-2: a conserved vertebrate-specific and perforin-like membrane protein involved in neuronal development.* Open biology, 2016. **6**(5): p. 160053.

59. Sassa, T., H. Aizawa, and H. Okamoto, *Visualization of two distinct classes of neurons by gad2 and zic1 promoter/enhancer elements in the dorsal hindbrain of developing zebrafish reveals neuronal connectivity related to the auditory and lateral line systems.* Developmental dynamics: an official publication of the American Association of Anatomists, 2007. **236**(3): p. 706-718.

60. Kumari, S. and S. Mayor, *ARF1 is directly involved in dynamin-independent endocytosis.* Nature cell biology, 2008. **10**(1): p. 30-41.

61. Feng, Y. and C.A. Walsh, *Mitotic spindle regulation by Nde1 controls cerebral cortical size.* Neuron, 2004. **44**(2): p. 279-293.

62. Alkuraya, F.S., et al., *Human mutations in NDE1 cause extreme microcephaly with lissencephaly.* The American Journal of Human Genetics, 2011. **88**(5): p. 536-547.

63. Pertz, O., et al., *Spatiotemporal dynamics of RhoA activity in migrating cells.* Nature, 2006. **440**(7087): p. 1069-1072.

64. Eide, E.J. and D.M. Virshup, *Casein kinase I: another cog in the circadian clockworks.* Chronobiology international, 2001. **18**(3): p. 389-398.

65. Schittek, B. and T. Sinnberg, *Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis.* Molecular cancer, 2014. **13**(1): p. 231.

66. Pongracz, J.E. and R.A. Stockley, *Wnt signalling in lung development and diseases.* Respiratory research, 2006. **7**(1): p. 15.

67. Patapoutian, A. and L.F. Reichardt, *Roles of Wnt proteins in neural development and maintenance.* Current opinion in neurobiology, 2000. **10**(3): p. 392-399.

68. Klaus, A. and W. Birchmeier, *Wnt signalling and its impact on development and cancer.* Nature Reviews Cancer, 2008. **8**(5): p. 387.

69. Kim, K.R., et al., *Crystal structure of human cytosolic aspartyl-tRNA synthetase, a component of multi-tRNA synthetase complex.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(10): p. 1840-1846.

70. Fröhlich, D., et al., *In vivocharacterization of the aspartyl-tRNA synthetase DARS: Homing in on the leukodystrophy HBSL.* Neurobiology of disease, 2017. **97**: p. 24-35.

71. Diodato, D., D. Ghezzi, and V. Tiranti, *The mitochondrial aminoacyl tRNA synthetases: genes and syndromes.* International journal of cell biology, 2014. **2014**.

72. Taft, R.J., et al., *Mutations in DARS cause hypomyelination with brain stem and spinal cord involvement and leg spasticity.* The American Journal of Human Genetics, 2013. **92**(5): p. 774-780.

73. Kikuchi, J., et al., *Induction of ubiquitin-conjugating enzyme by aggregated low density lipoprotein in human macrophages and its implications for atherosclerosis.* Arteriosclerosis, thrombosis, and vascular biology, 2000. **20**(1): p. 128-134.

74. David, Y., et al., *The E2 ubiquitin-conjugating enzymes direct polyubiquitination to preferred lysines.* Journal of Biological Chemistry, 2010. **285**(12): p. 8595-8604.

75. Kanerva, K., et al., *Human ornithine decarboxylase paralogue (ODCp) is an antizyme inhibitor but not an arginine decarboxylase.* Biochemical Journal, 2008. **409**(1): p. 187-192.

76. Hallen, L., et al., *Antiproliferative activity of the human IFN-α-inducible protein IFI44.* Journal of interferon & cytokine research, 2007. **27**(8): p. 675-680.

77. Haralambieva, I.H., et al., *Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine.* Human genetics, 2017. **136**(4): p. 421-435.

78. Kong, J., et al., *Potential protein biomarkers for systemic lupus erythematosus determined by bioinformatics analysis.* Computational biology and chemistry, 2019. **83**: p. 107135.

79. Levy, M., et al., *Zinc transporter 10 (ZnT10)-dependent extrusion of cellular Mn2+ is driven by an active Ca2+-coupled exchange.* Journal of Biological Chemistry, 2019. **294**(15): p. 5879-5889.

80. Hull, M.L., et al., *Host-derived TGFB1 deficiency suppresses lesion development in a mouse model of endometriosis.* The American journal of pathology, 2012. **180**(3): p. 880-887.

81. Ingman, W.V. and S.A. Robertson, *The essential roles of TGFB1 in reproduction.* Cytokine & growth factor reviews, 2009. **20**(3): p. 233-239.

82. Manos, P., G.K. Bryan, and J. Edmond, *Creatine kinase activity in postnatal rat brain development and in cultured neurons, astrocytes, and oligodendrocytes.* Journal of neurochemistry, 1991. **56**(6): p. 2101-2107.

83. Ventura-Clapier, R., H. Mekhfi, and G. Vassort, *Role of creatine kinase in force development in chemically skinned rat cardiac muscle.* The Journal of general physiology, 1987. **89**(5): p. 815-837.

84. Chauvigné, F., C. Zapater, and J. Cerdà, *Role of aquaporins during teleost gametogenesis and early embryogenesis.* Frontiers in physiology, 2011. **2**: p. 66.

85. Sumanas, S., T. Jorniak, and S. Lin, *Identification of novel vascular endothelial–specific genes by the microarray analysis of the zebrafish cloche mutants.* Blood, 2005. **106**(2): p. 534-541.

86. Cannon, J., et al., *Global analysis of the haematopoietic and endothelial transcriptome during zebrafish development.* Mechanisms of development, 2013. **130**(2-3): p. 122-131.

87. Sedykh, I., et al., *Zebrafish zic2 controls formation of periocular neural crest and choroid fissure morphogenesis.* Developmental biology, 2017. **429**(1): p. 92-104.

88. Frank, C.L., et al., *Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum.* Nature neuroscience, 2015. **18**(5): p. 647-656.

89. Hillegass, J.M., *The role of matrix metalloproteinases in zebrafish (danio rerio) embryogenesis and their regulation by glucocorticoids*. 2008, Rutgers University-Graduate School-New Brunswick.

90. Pociute, K., J.A. Schumacher, and S. Sumanas, *Clec14a genetically interacts with Etv2 and Vegf signaling during vasculogenesis and angiogenesis in zebrafish.* BMC developmental biology, 2019. **19**(1): p. 6.

91. Barcia, G., et al., *De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy.* Nature genetics, 2012. **44**(11): p. 1255-1259.

92. Oliver, K.L., et al., *Myoclonus epilepsy and ataxia due to KCNC 1 mutation: Analysis of 20 cases and K+ channel properties.* Annals of neurology, 2017. **81**(5): p. 677-689.

Chapter 3 Comparative transcriptome analysis during the seven developmental stages of Channel Catfish (*Ictalurus punctatus*) and Tra Catfish (*Pangasianodon hypophthalmus*) provides novel insights for the terrestrial adaptation

3.1 Abstract

Channel catfish (*Ictalurus punctatus*) and tra catfish (*Pangasianodon hypophthalmus*) both belong to the order Siluriformes. Channel catfish does not possess an air-breathing organ (ABO), and thus cannot breathe in the air, while tra catfish is a facultative air-breather and use the swim bladder as its air-breathing organ, which provides for aerial breathing in low oxygen conditions. Tra and channel catfish serve as a great comparative model for studying the transition of life from water to terrestrial living, as well as for understanding genes that are crucial for development of the swim bladder and the function of air breathing in tra catfish. In this study, seven developmental stages in tra catfish were selected for RNA-Seq analysis based on their transition to a stage that could live at 0 ppm oxygen. More than 587 million sequencing clean reads were generated in tra catfish, and a total of 21, 448 unique genes were detected. A comparative genomic analysis between channel catfish and tra catfish revealed 109 genes were present in tra catfish, but absent from channel catfish. In order to further narrow down the list of these candidate genes, gene expression analysis was performed for these tra catfish specific genes. Hypoxia challenge and microtomy experiments collectively suggested that there are

critical timepoints for the development of the air breathing function and swim bladder development stages in tra catfish. Fourteen genes were inferred to be important for air breathing. Of these, *hry, grp and cx3cl1* genes were identified to be the best candidates of genes related to the air breathing ability in tra catfish. This study provides a large data resource for functional genomic studies in air breathing function in tra catfish, and sheds light on the adaption of aquatic organisms to the terrestrial environment.

3.2 Background

Oxygen is indispensable for all aerobic creatures. For animals, breathing is a critical biophysical and voluntary process that involves inhaling oxygen and transferring it to the cells. The transport and consumption of oxygen involves multiple physiological and biochemical processes. For water breathing fish, gills serve as the primary site for gas exchange with the environment [1]. In addition to their respiratory function, gills also serve as a pathway for the exchange of non-volatile molecules between the blood and environment, and most fish exchange gases through gills that are protected under an operculum on both sides of the pharynx [2]. In addition to gill-based breathing, there are many fish that can perform aerial respiration [3]. Primitive fish were the first vertebrates to breathe air through atmosphere, in addition to breathing gases dissolved in the aquatic environment [4]. Air breathing fish are fish that can undergo gas exchange respiration directly with the aerial environment, rather than in water. Air breathing fish may rise to the surface of water, gulp air, and even crawl onto land and survive for

a long time [5]. Although a number of different morphological adaptations for air breathing exists, all gas exchangers share basic features such as the barrier between gas and blood is thin, pleated structure with a large interface [6]. This structure creates small diffusion distances of gases and decreases branchial vascular resistance, and also increases the respiratory surface area. It eventually results in the low consumption respiration in air-breathing fish and improves the respiratory efficiency, allowing them to survive even being exposed to air deficient with oxygen [6].

It is estimated that there are more than 370 extant air breathing fish species in 49 families and their air-breathing organs vary considerably [7]. For the species that use modified gills as air-breathing organ (ABO), such as *Clarias macrocephalus* and *C. batrachus,* the efferent branchial arteries of anterior (first and second) gill arches serve as the accessory air-breathing organ and are also the site for gas exchange. The ventral aorta originates from the heart and splits into a ventral and dorsal branch. The ventral branch supplies blood to anterior gill arches, which then flows through the accessory ABO and, back to the heart. The dorsal branch distributes blood to the posterior (third and fourth) gill arches and proceeds to the circulatory system, which transports oxygen-rich blood to other tissues [8]. In some species, such as *Misgurnus anguillicaudatus* and *Corydoras aeneus,* a modified intestine serves as the ABO, the air is taken into the mouth with unidirectional ventilation of the posterior region of the intestine and continuous exhaust of gas from the vent [9]. In the posterior region of the intestine, the mucosa has very smooth surfaces and is lined with respiratory epithelium and capillary networks, and a

thin air-blood barrier (0.24-3.00 μm) is produced for air exchange [10]. Intestinal gas exchange and digestion in fish are not mutually exclusive processes. In such animals, eating and breathing alternates over very short time in fish [9].

The modified swim bladder is another air-breathing organ. The teleost swim bladder is a large, trabeculated, well vascularized organ, and is widely considered as homologous lung of immemorial Osteichthyes [11, 12]. The swim bladder can serve as a gas (usually oxygen) container, and can function as a hydrostatic organ, allowing the fish to maintain itself in the water column without floating up down with the stream. It also serves as a resonating chamber to produce or receive sound. Orders of fish using the swim bladder for aerial respiration includes *Gonorynchiformes* [13], *Characiformes* and some species in *Siluriformes*, such as suborder *Gymnotus* and the *Pangasianodon hypophthalmus* (tra catfish) [14, 15]. Tra catfish are facultative air breathers. Their swim bladder extends from the posterior of the head to the tail beyond the anus. The collagen rich fibrous walls form subdivisions that are support the swim bladder. There are two types of epithelial cells on the surface of the collagen-like fibrous wall, one is a thin respiration type, which covers the majority of the surface and is highly vascularized, making it a major place for gas exchange between air and blood in the swim bladder. The second type is thicker with a brush border, which also works to support the structure of swim bladder [16]. Phuong et al (2018) also found that the volume and the respiratory surface area of the swim bladder in tra catfish is strongly and positively correlated with the body mass [16].

Catfish (Order Siluriformes) is one of the most commonly raise food fish in the United States. The channel catfish (*Ictalurus punctatus*) and its hybrid from the mating with blue catfish (*I. furcatus*) are recognized as the most extensively cultured type of catfish in the USA. The culture of tra catfish *(Pangasianodon hypophthalmus)* accounts for 2/3 of Vietnam's overall aquaculture yields [17].

The Mekong Delta in Vietnam has become one of the world's largest aquaculture producers, with an annual output of tra catfish reaching 1.14 million tonnes, with an export income of about 1.4 billion dollars [18]. The air-breathing capability of tra catfish allows them to live under conditions of low aquatic oxygen concentrations and so tra have a substantial advantage over channel catfish, which cannot breathe air and so which are confined to high oxygen, aerobic environments [19][20].

In practical catfish production, hypoxia is a frequent and significant problem, resulting in enormous economic losses. Aerators are extensively used in the US catfish farming industry, but are subject to mechanical failure or human error, and are energy intensive. A better understanding of the mechanisms for tolerating hypoxia is critical for continued successful US catfish aquaculture productivity, expecially in face of competition from SE Asia. In addition, the study of hypoxia tolerance, and the cellular and physiological bases for understanding how this has occurred, will better guide our understanding of the evolution of life from the ocean to the terrestrial environment.

In this study, hypoxia challenge and histology experiments were conducted to reveal the development of the swim bladder in tra catfish, and to better understand its air breathing function. RNA-Seq analysis of seven critical stages during early development of the tra catfish associated with different levels of air breathing ability was conducted to identify the genes and pathways leading to the development of swim bladder and functioning of air breathing in tra catfish larvae.

3.3 Materials and Methods

3.3.1 Ethics statement

All experimental procedures involving animal care and tissue collection were approved by the Auburn University and Can Tho University Institutional Animal Care and Use Committee. All animal related procedures were performed following the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States and in Vietnam.

3.3.2 Experimental animals and tissue collection

Tra catfish embryos were produced at Can Tho University, Vietnam. Tra catfish embryos were hatched within 1-day post fertilization and samples were collected every 24 hours over a 30-day period since hatching. A total of 20-50 eggs/embryo/fry were collected at each sampling. Two sets of samples were taken at each time point and later used for histological and gene expression analysis [21]. At each sampling point, one set of fry was euthanized with 200 ppm

buffered MS-222, and then stored in RNA-Later solution (Thermo Fisher Scientific). Samples were shipped to the US on dry ice and then stored at -80°C until RNA isolation was carried out. The second set of fry was sampled in the same way, placed in 10% neutral buffered formalin and sealed with screw top lids [21]. All samples for histology were kept at room temperature.

3.3.3 Low oxygen (Anoxia) challenge

Anoxia challenge was conducted at Can Tho University, Vietnam, to determine the tolerance of tra catfish to anoxia conditions at a temperature of 27 ºC. Tra catfish larvae were challenged each day from 2-12 days post fertilization. One group of 20 larvae were placed in a two-liter container with oxygen supply as a control treatment. The other group was stocked in a second two-liter container at first with aeration and then the dissolved oxygen level was lowered by bubbling nitrogen gas into the water until a 0 mg/L dissolved oxygen concentration was obtained (measured by DO meter). Dissolved oxygen levels, larvae behavior and survival rate were measured and recorded every 15 minutes. The experiment was repeated daily until all 20 fish survived in the eliminated oxygen container (the fish demonstrated facultative air breathing), after 12 days post-fertilization.

3.3.4 RNA isolation, library construction, and sequencing

Seven time points, 2, 4, 6, 8, 9, 10, 11 days post fertilization (dpf), were selected based on the low oxygen challenge results. At each time point, two replicates were taken for RNA

isolation and Illumina sequencing. Samples were moved from the -80°C freezer, and four whole fry per replicate were pooled for each time point in order to obtain sufficient nucleic acid for analysis. Pooled samples were ground to a fine powder using mortar and pestle in liquid nitrogen. RNA extraction was performed using Qiagen RNeasy Plus Kit (Qiagen, Valencia, California), following the manufacturer's directions. RNA concentration and integrity of each sample was measured on a NanoDrop 2000 Spectrophotometer (NanoDrop Technologies). For each life stage, equal amounts of RNA from the two pooled replicates was used for RNA-seq library construction and RNA sequencing.

Library preparations and sequencing reactions were conducted at GENEWIZ, LLC. (South Plainfield, NJ, USA). Ribosomal RNA, which occurs in vast abundance in all samples, was reduced so that it would not interfere with analysis of mRNA. The rRNA depletion method chosen was via the Ribozero rRNA Removal Kit (Illumina, San Diego, CA, USA). The NEBNext Ultra RNA Library Prep Kit was utilized for Illumina RNA sequencing library preparation by following standard protocols (NEB, Ipswich, MA, USA). Amplified library yields were validated using the Agilent Tapestation 4200 (Agilent Technologies, Palo Alto, CA, USA) and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). Then the sequencing libraries were multiplexed with eight samples clustered per lane of a flowcell and loaded on the Illumina HiSeq instrument with a 2x150 Paired End (PE) configuration.

3.3.5 Reads mapping and differential expression analysis.

Raw reads quality was evaluated in FastQC (version 0.11.5) [22], and low-quality bases, adapter sequences and ambiguous nucleotides were trimmed from the raw sequences using Trimmomatic (version 0.36) [23]. The reads were removed if an average slidingwindow Phred score over four bases was less than 25, and the reads with length shorter than 36 bases after trimming were removed. The remaining high-quality reads were used for subsequent analysis. The recently assembled tra catfish genome was used as a reference for reads mapping. The tra catfish genome was approximately 700 Mb, assembled into 568 scaffolds, with a scaffold N50 of 14.29 Mbp [24].

To profile tra catfish gene expression, tra catfish filtered reads were mapped to their genome using STAR alignment software (version 2.7.0) with a max 4% mismatching rate of the mapped length allowed and a minimum 90% of the bases mapped to the genome [25]. HTSeq-count [26] was conducted to extract and count the read from the mapping files. After counting the number of clean reads mapped to each gene, the FPKM (fragments per kilobase of exon model per million reads mapped) method was performed for normalization and the genes with a FPKM smaller than 0.1 were filtered out of each sample. To account for differences with the development in the two each species, differential expression analysis was performed with the R package DESeq2 [27]. Differentially expressed genes (DEGs) were defined as having a P-value < 0.05 and an $|\log2 \text{ fold change}| > 1$.

3.3.6 Gene ontology and enrichment analysis

The GO terms for each differential expression comparison were obtained by using zebrafish database annotations for the unigene set as well as using ClusterProfiler R software (version 3.6) [28]. The annotation result was then sorted with respect to biological process, cellular component and molecular functions. ClusterProfiler R was also used for a GO functional enrichment analysis of certain genes. A criterion of p-value and q-value cutoff of 0.05 was chosen as the threshold of significance.

3.3.7 Comparative genomic analysis of channel catfish and tra catfish

The protein sequences of channel catfish and tra catfish were obtained from the NCBI website to determine chromosome orthology [29, 30]. The orthologs and orthogroups between channel catfish and tra catfish were identified using OrthoFinder version 2.2.7 [31]. In order to obtain the tra catfish-specific genes, an additional round of protein BLAST (BLASTP) was conducted for genes that are not in the orthologue groups. These genes were queried against the genes in the orthologue groups within the same species with a maximum e-value threshold of 1e-10. In the end, reciprocal BLASTP searches were performed to query genes with no hits from last steps with a maximal e-value threshold of 1e -5. The remaining genes with no hits to any orthologs were identified as species-specific genes for subsequent analysis.

To further confirm the tra-catfish specific genes from previous steps, the species-specific genes in tra catfish were queried with the channel catfish genome using TBLASTN with a

maximum e-value threshold of 1e-10. The recognized tra-catfish specific genes were then screened based on the percentage of identical matches (pident) and query coverage per subject (qcovs). The genes without a TBLASTN hits with channel catfish genome were finally recognized as tra-catfish specific genes present in tra catfish, but absent in the channel catfish genome.

3.3.8 Clustering of time series gene expression data

Clustering is universally used for gene expression data analysis. Mfuzz [32] is one of the most commonly implemented soft clustering software. A minimization of weighted square error function based on fuzzy c-means algorithm was performed to reveal structures underlying large gene expression data sets. Hard clustering approaches are preferable to identify well separated clusters, at a cost of excluding biologically relevant genes. To solve this issue, Mfuzz soft clustering provides an overall relation between clusters, and it is more robust to noise since gene/protein clusters frequently overlap in biological data.

3.3.9 Histological analysis

Tra catfish samples for each life stages were removed from 10% formalin and used for paraffin processing for embedment and subsequent sectioning. Following dehydration by graded ethanol, hyalinization and infiltration by dimethylbenzene with standard protocols, samples were embedded in paraffin. Then we conducted transverse sectioning of 7 μm thickness. Sections goes through a water-bath at 40°C and then placed on glass slides. The slides were kept in an

incubator at 37°C overnight. Following deparaffination by xylene substitution (HEMO-DE), the slides were then stained with hematoxylin and eosin (HE) according to standard procedures. After HE staining, slides were covered with a cover slip and left for 48 hours at room temperature to dry. Slides were observed and imaged with an Olympus® BHS fluorescence binocular microscope equipped with a 3.4-megapixel color digital camera (Qimaging® model Micropublisher 3.3 RTV). Image Pro Plus 7 software (Media Cybernetics, Bethesda, MD) was used to capture the image.

3.4 Results

3.4.1 Sequencing and global analysis of tra catfish transcriptome

A total of 1,303 million raw reads were generated for tra catfish through RNA-Seq analysis. As shown in Table 4, after removing low-quality reads (quality score < 25) and trimmed reads that are less than 36 bases, approximately 587 million clean reads retained in tra catfish. All clean reads were aligned to the tra catfish reference genome using STAR software (v 2.7.0) with the mapping rate ranging from 62.44% to 73.77% (Supplementary Figure 4).

Reads were assigned to transcripts by their overlaps with tra catfish reference gene models. FPKM (fragments per kilobase of exon model per million reads mapped) method was performed for normalization and an FPKM cutoff for expressed genes was set at >= 0.1. As shown in Figure 11a, a total of 21,448 gene transcripts were detected in the RNA-Seq data set, the highest number of expressed genes (21,004) were detected at 9 dpf and the lowest number of expressed genes

(20,516) were verified at 8 dpf. A total of 19,728 genes were discovered to be expressed at every sample point (Supplementary Table 6).

Principal component analysis (PCA) was used to identify the outliers tra catfish transcriptomes during different development time points (Figure 12). The clustering results were in agreement with the developmental stage groups. In Figure 2, the expression profiles of different time points were divided into 7 clusters from left to right, with biological replicates within group clustered together. We could observe considerable expression variability across different developmental points, accounting for 53% of expression variation.

Table 4. Summary of the transcriptome sequencing in tra catfish (*Pangasianodon hypophthalmus*). Pan2-1, Pan2-2…Pan11-2 are abbreviations for sampling of tra catfish embryos/fry replicate 1 at 2 days post-fertilization (dpf), replicate 2 at 2 dpf, …, replicate 2 at 10 dpf, respectively.

Days post fertilization	No. of raw reads	No. of Clean reads	%GC	No. of mapped reads	mapped reads (%)
Pan2-1	91,258,588	34,263,036	48	23,315,024	68.04%
Pan2-2	99,428,290	45,246,230	48	32,999,524	72.93%
Pan4-1	84,303,060	35,220,266	49	24,687,004	70.09%
Pan4-2	94,169,112	35,220,266	49	32,125,804	73.77%
Pan6-1	101,986,182	50,214,702	48	34,695,330	69.10%

Pan6-2	83,463,360	39,952,806	48	27,393,394	68.57%
Pan8-1	79,577,474	41,618,368	50	26628028	63.98%
Pan8-2	122,305,060	63,011,586	49	435,083,86	69.05%
Pan9-1	95,165,314	44,452,736	48	28,410,570	63.91%
Pan9-2	87,731,496	41,556,154	48	26,434,938	63.61%
Pan10-1	90,931,716	37,469,612	50	23,738,636	63.35%
Pan10-2	91,213,504	36,673,982	50	24,093,544	65.7%
Pan11-1	97,272,818	42,492,728	47	26,532,068	62.44%
Pan11-2	83,885,030	39,744,530	47	25,736,390	64.75%

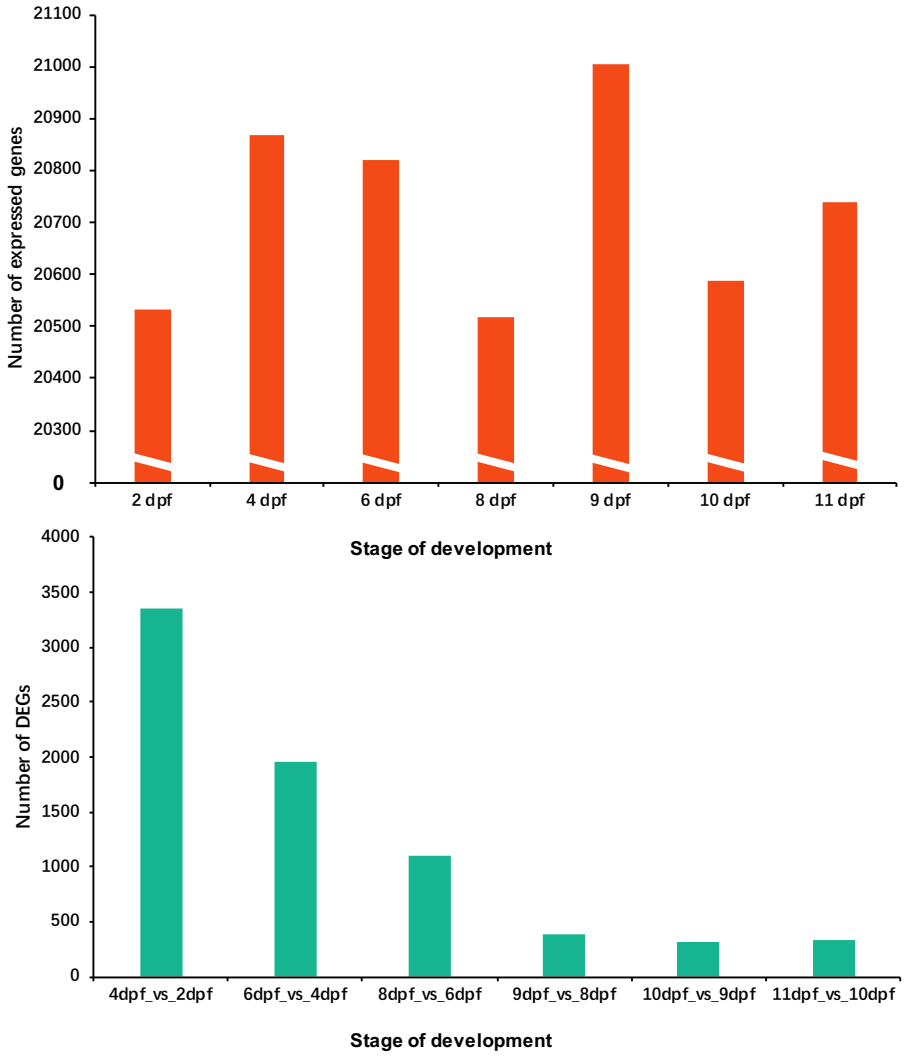


Figure 11. Gene expression during early embryonic development in tra catfish (*Pangasianodon hypophthalmus*). (**A**) Number of expressed genes at each development stage averaged for two replicates; (**B**) The number of DEGs (Differentially Expressed Genes) for comparison of each stage with the previous stage.

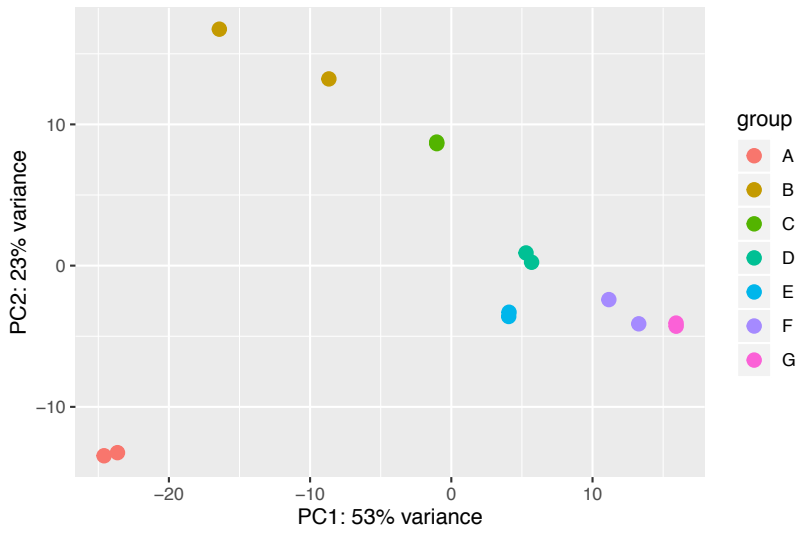


Figure 12. Principal component analysis (PCA) of tra catfish (*Pangasianodon hypophthalmus*) transcriptome. A: 2 dpf; B: 4 dpf; C: 6 dpf; D: 8 dpf; E: 9 dpf; F: 10 dpf; G: 11 dpf.

3.4.2 Differentially expressed genes during developmental stages of tra catfish

RNA sequencing reads were aligned with the tra catfish genome and gene read counts were calculated using HTSeq. DESeq2 software was used to identify the significantly and differentially expressed genes (DEGs) with comparison of each stage to the previous stage. In tra catfish, the number of DEGs ranged from 3,360 (2,324 up-regulated, 1,036 down-regulated), between 4 dpf and 2 dpf, to 338 (213 up-regulated, 125 down-regulated) between 11 dpf and 10 dpf (Supplementary Table 7). In general, the number of identified DEGs decreased during the development stages in tra catfish (Figure 11b). 1,956 (861 up-regulated, 1,095 down-regulated), 1,108 (661 up-regulated, 447 down-regulated), 380 (209 up-regulated, 171 down-regulated) and 324 (209 up-regulated, 115 down-regulated). DEGs were detected between 6 dpf and 4 dpf, 8 dpf

and 6 dpf, 9 dpf and 8 dpf, 10 dpf and 9 dpf, respectively. Taken together, 5,419 DEGs were detected at different developmental stages. Notably, the greatest number of DEGs was identified between the first two timepoints (4 dpf and 2 dpf), which is not surprising since the transition from fertilized egg to hatchling is dramatic, and various biological processes would be expected to occur between these stages. A Venn diagram showed the intersection between DEGs at different stages (Supplementary Figure 5). Only one gene was detected to be differentially expressed in all stage comparisons.

3.4.3 Gene ontology enrichment analysis of differentially expressed genes (DEGs) at different stages

To classify if there was a gene ontology (GO) enrichment category with related functions, the GO term enrichment analysis of DEGs was performed for each developmental stage. The GO categories that were significantly enriched for different stages of tra catfish are listed in Supplementary Table 7; The GO enriched categories at each stage are shown in Figure 13.

At 4 dpf compared to 2 dpf, the significantly enriched categories for the DEGs were mainly related to blood vessel formation, respiratory chain complex assembly, reactive oxygen species (ROS) and Wnt signaling pathway, that included vasculogenesis (GO:0001570). Day 4 was also associated with respiratory chain complex IV assembly (GO:0008535), regulation of reactive oxygen species metabolic process (GO:2000377) and canonical Wnt signaling pathway (GO:0060070). Several mitochondrial activity and respiratory chain functions were enriched at 6 dpf compared to 4 dpf, including respiratory chain complex IV assembly (GO:0008535),

mitochondrial inner membrane (GO:0005743), mitochondrial ribosome (GO:0005761) and mitochondrial protein complex (GO:0098798).

At 8 dpf compared to 6 dpf, enriched categories for the DEGs were mainly related to morphogenesis of heart and muscle organ, transport and binding of oxygen, some other ion transmembrane transport as well as some complex assembly and development, such as heart morphogenesis (GO:0003007), muscle organ morphogenesis (GO:0048644), oxygen transport (GO:0015671), oxygen carrier activity (GO:0005344), oxygen binding (GO:0019825), calcium ion transmembrane transporter activity (GO:0015085), hemoglobin complex (GO:0005833) and cardiac myofibril assembly (GO:0055003). At 9 dpf compared to 8 dpf, DEGs were mainly enriched in transmembrane transport activity, dendrite and some channel activity. The activities included regulation of ion transmembrane transport (GO:0034765), regulation of cation transmembrane transport (GO:1904062), transmembrane transporter complex (GO:1902495), dendrite membrane (GO:0032590), dendritic tree (GO:0097447), calcium channel complex (GO:0034704), potassium channel complex (GO:0034705) and ion gated channel activity (GO:0022839).

At the 10 dpf compared to 9 dpf stage, the DEGs were mainly enriched in metabolic process and endopeptidase activity, including glutamine family amino acid metabolic process (GO:0009064), aminoglycan metabolic process (GO:0006022), serine-type endopeptidase activity (GO:0004252), peptidase regulator activity (GO:0061134) and carboxylic ester hydrolase activity (GO:0052689). Additionally, DEGs from 11 dpf compared to 10 dpf were

enriched in oxygen and ion transport and ATP activity, including oxygen carrier activity (GO:0005344), oxygen binding (GO:0019825), bicarbonate transport (GO:0015701), sulfur compound transport (GO:0072348), sodium ion transmembrane transporter activity (GO:0015081), anion:anion antiporter activity (GO:0015301), chloride transmembrane transporter activity (GO:0015108), ATPase activator activity (GO:0001671) and sodium:potassium-exchanging ATPase activity (GO:0005391).

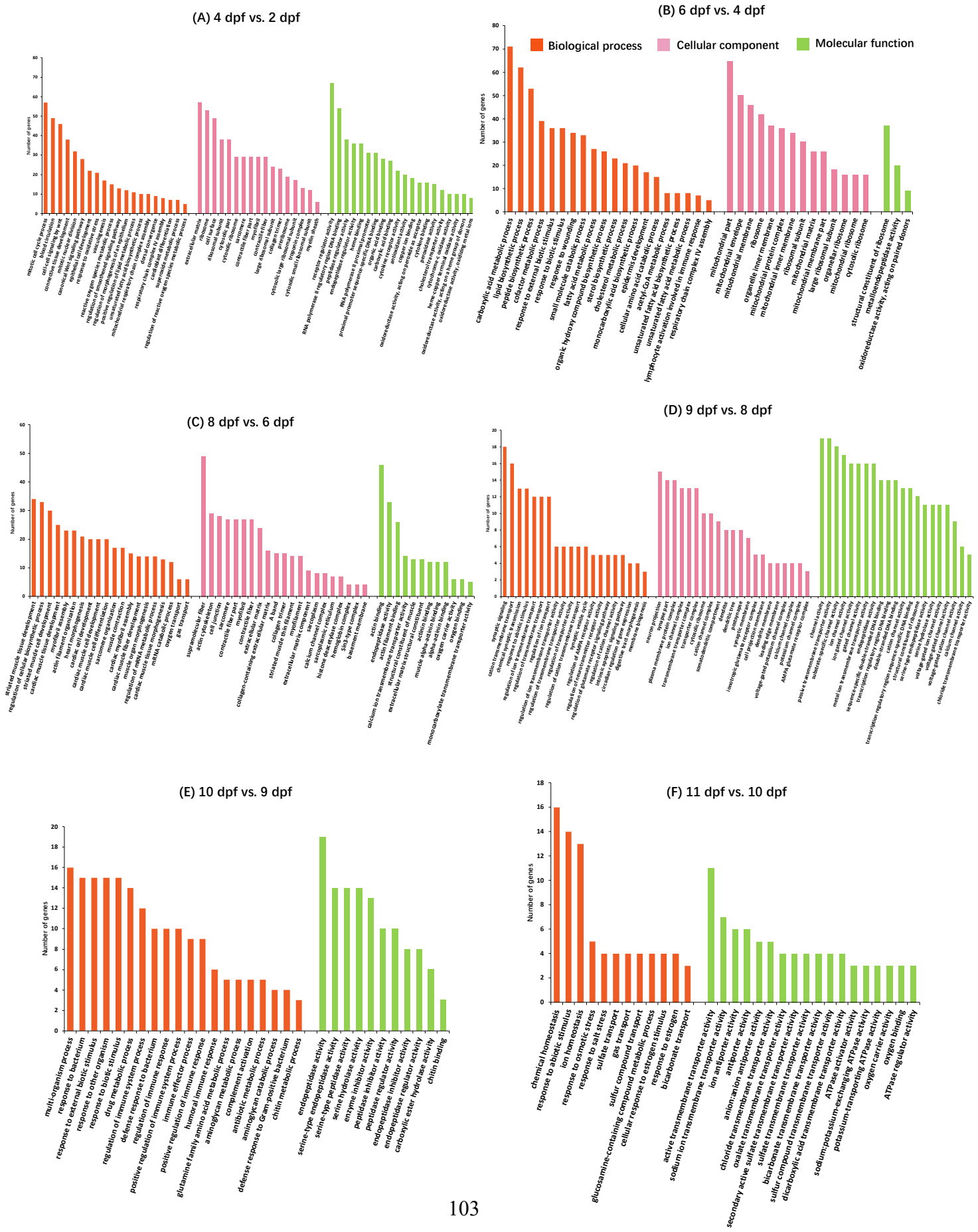

(A) 4 dpf vs. 2 dpf
(B) 6 dpf vs. 4 dpf
Biological process
Cellular component
Molecular function
Number of genes
(C) 8 dpf vs. 6 dpf
(D) 9 dpf vs. 8 dpf
(E) 10 dpf vs. 9 dpf
(F) 11 dpf vs. 10 dpf

Figure 13. Gene ontology enrichment of DEGs at different development stages in tra catfish (*Pangasianodon hypophthalmus*). (A): 4 dpf and 2 dpf; (B) 6 dpf and 4 dpf; (C) 8 dpf and 6 dpf; (D) 9 dpf and 8 dpf; (E) 10 dpf and 9 dpf; (F) 11 dpf and 10 dpf. The vertical axis indicates the number of DEGs between two adjacent sampling datasets, and the horizontal axis represents the GO terms significantly enriched in the DEGs.

3.4.4 Comparison of gene contents of tra catfish and channel catfish

Tra catfish and channel catfish are both Siluriformes and so, are evolutionarily closely related to each other, but air breathing ability varies greatly between these two species. The tra catfish is a facultative air-breather and utilizes the swim bladder as its air-breathing organ in low oxygen environment; while channel catfish does not possess an air-breathing organ and thus cannot breathe in the air. At first, we compared the gene contents between tra catfish genome and channel catfish genome, and a total of 109 genes were identified to be present in tra catfish that were absent from channel catfish (Supplementary Table 8). Many traits differ between tra catfish and channel catfish, not only the ability to breathe in the air. But many genes should be involved in the formation of swim bladder in tra catfish and the air breathing ability. In this regard, genes which contribute to swim bladder development and aerial breathing ability should be differentially expressed during tra catfish development, so the genes, which are (1) present in tra catfish but absent from channel catfish, and (2) differentially expressed during tra catfish development would be considered to be key genes involved in the morphogenesis of swim bladder and differences in aerial breathing ability. Taken together, 109 tra catfish-specific genes

were compared with the 5,419 genes that were differentially expressed during tra catfish development, and 27 genes were present in tra catfish that were absent from channel catfish, that were also differentially expressed during tra catfish development (Figure 14, Table 5). All these 27 genes may be closely related with the formation of air breathing ability in tra catfish.

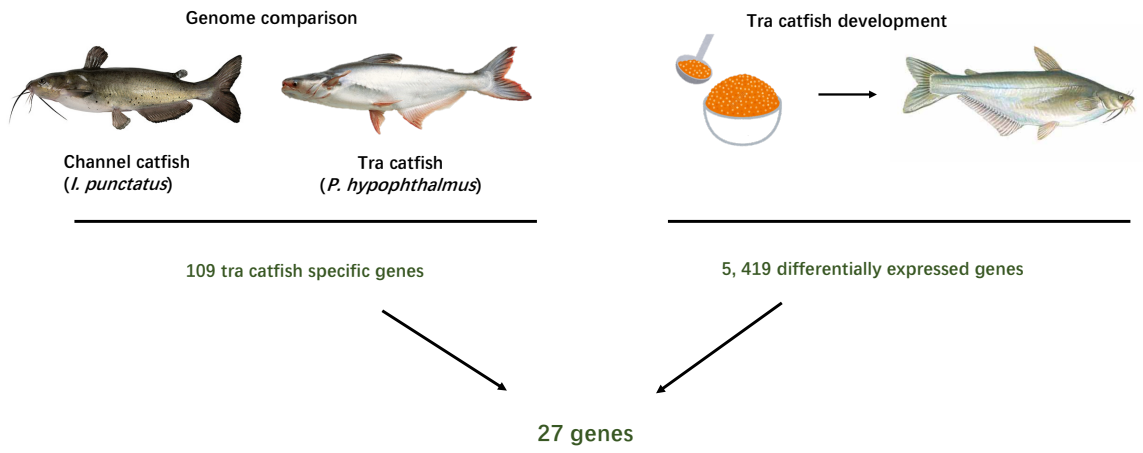


Figure 14. Twenty-seven candidate key genes were identified to be present in tra catfish (*Pangasianodon hypophthalmus*) but absent in channel catfish (*Ictalurus punctatus*), also, it was differentially expressed during tra catfish (*Pangasianodon hypophthalmus*) development.

Table 5. Genes specific to tra catfish (*Pangasianodon hypophthalmus*) that were differentially expressed during tra catfish development.

Gene ID	Name	Description

113524458	LOC113524458	uncharacterized protein
113524522	LOC113524522	uncharacterized protein YBL113C-like
113524613	HRG	histidine-rich glycoprotein-like, partial
113531560	CUNH3ORF85	uncharacterized protein C3orf85 homolog
113533778	SHKD	dual specificity protein kinase shkD-like isoform X2
113534207	LOC113534207	uncharacterized protein LOC113534207
113535709	FAM216A	protein FAM216A isoform X1
113536891	COX7C	cytochrome c oxidase subunit 7C, mitochondrial
113538378	LOC113538378	uncharacterized protein LOC113538378
113539350	LY6D	lymphocyte antigen 6D
113539654	ZNF862	zinc finger protein 862-like
113541042	VXN	vexin-like
113535796	GRP	gastrin-releasing peptide-like isoform X2
113541665	LOC113541665	uncharacterized protein LOC113541665
113542106	LOC113542106	uncharacterized protein LOC113542106 isoform X1
113542136	LOC113542136	uncharacterized protein LOC113542136
113543977	LOC113543977	uncharacterized protein LOC113543977
113544002	LOC113544002	uncharacterized protein YBL113C-like

113544546	CLCA3A1	calcium-activated chloride channel regulator 3A-1-like
113544571	CLCA1	calcium-activated chloride channel regulator 1-like
113544929	LAGE3	L antigen family member 3
113545006	NCMAP	noncompact myelin-associated protein
113545009	LOC113545009	uncharacterized protein DDB_G0287625-like
113546726	LOC113546726	uncharacterized protein LOC113546726
113547415	PET117	PET117 Cytochrome C Oxidase Chaperone
113547475	SPIDROIN	spidroin-2-like isoform X1
113533685	CX3CL1	chromo domain-containing protein cec-1-like isoform X2

3.4.5 Anoxia Challenge

A low oxygen challenge experiment were conducted to test the survival ability of tra catfish in anoxic conditions (0 ppm dissolved oxygen (DO)) as a function of age (Supplementary Table 9) [21]. At 2 dpf, tra larval survival dropped to 0% when the oxygen was lowered to 0 ppm (i.e. anoxia) (Figure 15). Initially when the aeration was removed from the container, the larvae swam as normal. After 15 minutes, when the dissolved oxygen level dropped below 2.1 mg/L, the fish again swam rapidly at the surface, a behavior of oxygen stress. After 30 minutes, when the

dissolved oxygen level in water dropped to 1.3 mg/L, locomotion was dramatically reduced. After 45 minutes, when the dissolved oxygen level reached 0.7 mg/L, fish ceased to swim and sank to the bottom of container; all of the fish were dead.

At 4 dpf, the oxygen concentration fell rapidly from 5 mg/L to 1.5 mg/L 15 minutes after removing the aeration, and fish swam slowly. After 30 minutes, the dissolved oxygen level dropped to 1.1 mg/L, and the fish swam rapidly at the surface. At 45 minutes, the dissolved oxygen levels decreased to 0 mg/L and 50% of the fish dropped to the bottom of the tank, although no fish died at this point. After 55 minutes, with the dissolved oxygen concentration remaining at 0 mg/L the tra catfish larvae were verified to be dead.

When the fish were 6 dpf, the challenge began with a dissolved oxygen level of 4.7 mg/L. Oxygen concentration dropped to 1.2 mg/L then 0.7 mg/L at 15 and 30 minutes, respectively. When the dissolved oxygen level decreased below 1.2 mg/L, the fish were actively swimming on the surface. At 45 minutes, the dissolved oxygen concentration reached 0 mg/L, 50% of the fish dropped to the bottom. At 60 minutes, all fish were dead.

At 8 dpf, the dissolved oxygen level started at 4.6 mg/L, and the fish were actively swimming within the container. Fifteen minutes and 30 minutes into the challenge, the dissolved oxygen level fell to 0.9 mg/L and then 0.6 mg/L, respectively. At 45 minutes the dissolved oxygen level decreased to 0 mg/L, and thirty percent of the fish responded by dropping to the bottom. At 60 minutes, 70% of the fish were dead and ten minutes later all fish had died.

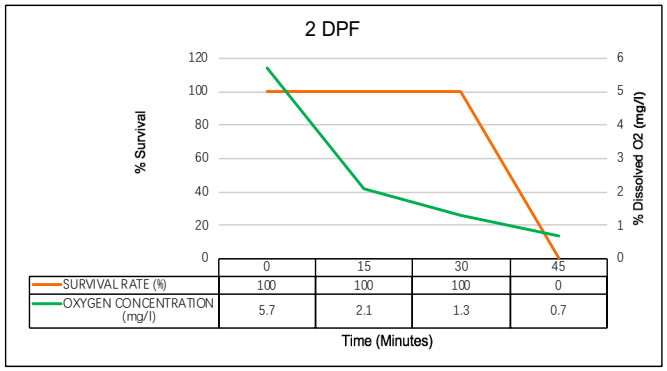
At 9 dpf, the challenge started at a dissolved oxygen reading of 5.4 mg/L. Oxygen level was lowered to 1.6 mg/L, 0.8 mg/L and then 0.4 mg/L at 15, 30 and 45 minutes, respectively, and fish were actively swimming on the surface during these stages. At 50 minutes, the dissolved oxygen concentration decreased to 0 mg/L. After 75 minutes, 27% of the fish died and sank to the bottom. The death rate remained the same until 50 minutes later when an additional 13% fish died. The challenge concluded at 120 minutes, and there was a 60 % survival rate for tra catfish larvae at a dissolved oxygen level of 0 mg/L.

When the fish were 10 dpf, the challenge began at 4.6 mg/L. At fifteen minutes after initiation of the challenge, the dissolved oxygen reading was 0.9 mg/L, and fish were actively swimming in the container. Thirty minutes in to the challenge, the dissolved oxygen reading decreased to 0.5 mg/L, 50% of the fish were on the surface. At 45 minutes the dissolved oxygen level fell to 0 mg/L, and all the fish were at the surface to gulp air. At 55 minutes, 13% of the fish died followed ten minutes later with an additional 7% fish dead, equating to 20% mortality. The challenge concluded after 120 minutes with the tra catfish larvae having 80% survival in 0 ppm DO.

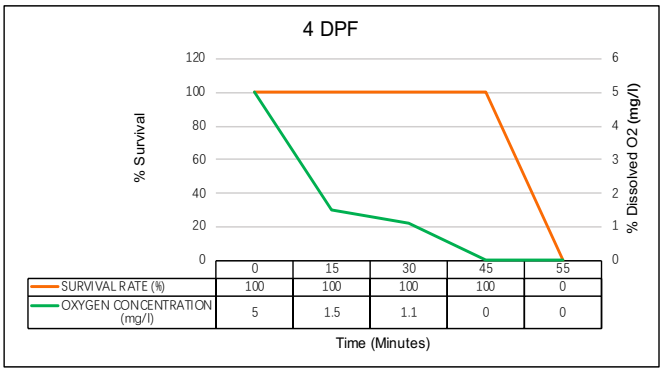
At 11dpf the challenge began at 4.6 mg/L. The dissolved oxygen concentration decreased to 1 and then 0.5 mg/L at 15 and 40 minutes, respectively. Fish were actively swimming at the bottom of the container. After 50 minutes, the dissolved oxygen level fell to 0 mg/L. At 0 mg/L dissolved oxygen, fish were swimming actively in the midwater column, and occasionally fish

swam to the surface gulping air. The challenge was concluded after 120 minutes with 100% survival of the tra catfish larvae.

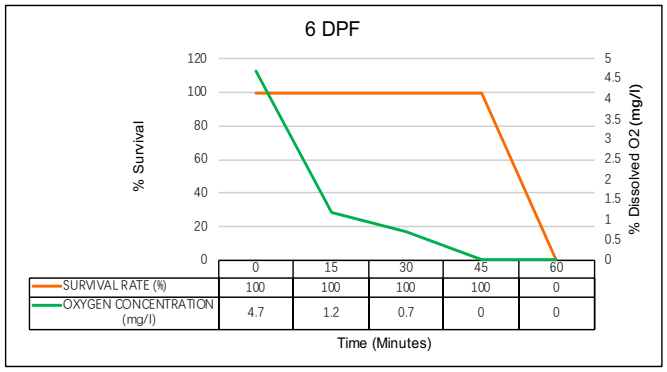
In conclusions, at a temperature of 27 ºC , within the first 6 dpf tra catfish larvae showed some levels of anoxia behavior when the DO reached below 2 ppm after 15 minutes being exposed to the decrease of DO. Their survival ability was 0% when oxygen was reduced to 0 ppm and maintained at 0 ppm for a while. However, at 4 dpf and 6 dpf, the tra catfish larvae already possessed some ability to survive at low dissolved oxygen conditions (0.7- 1.1 mg/L) , and sometimes exhibited air gulping behavior, which indicated a portion of the tra catfish larvae had reached a point to have initial gas exchange activity through air breathing. At 8 dpf, tra catfish larvae had 100% survival rate at 0.6 mg/L water conditions, and could survive for a while at a dissolved oxygen level of 0 mg/L (anoxia). The survival rate of tra catfish larvae was 60%, 80%, 100% in 9, 10, 11 dpf when challenge concluded after 120 minutes, indicating that at 9 and 10 dpf, the air breathing ability was relatively complete. At 11 dpf, the swim bladder in tra catfish was fully developed and tra catfish possessed complete air breathing capability.



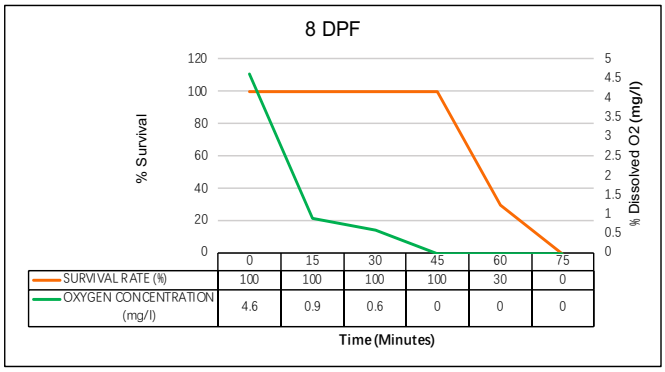
	0	15	30	45
SURVIVAL RATE (%)	100	100	100	0
OXYGEN CONCENTRATION (mg/l)	5.7	2.1	1.3	0.7



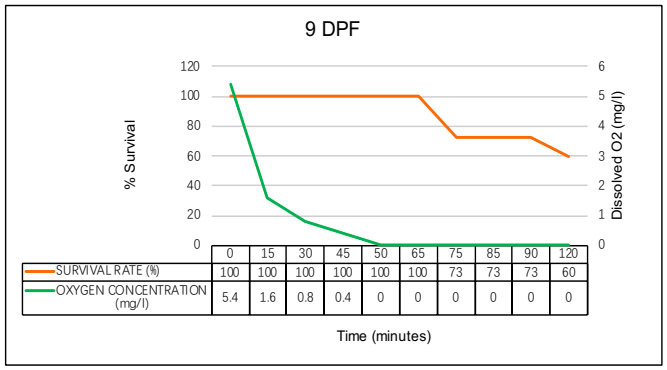
	0	15	30	45	55
SURVIVAL RATE (%)	100	100	100	100	0
OXYGEN CONCENTRATION (mg/l)	5	1.5	1.1	0	0



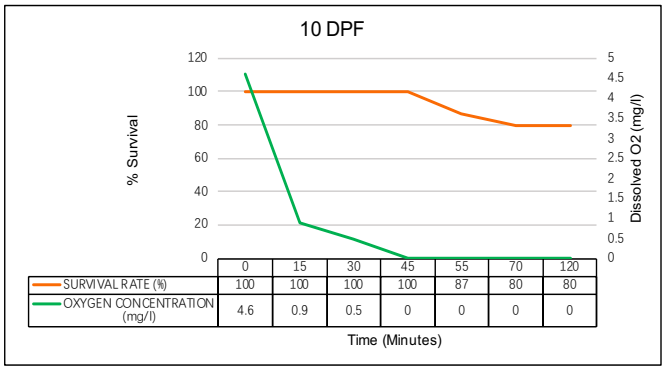
	0	15	30	45	60
SURVIVAL RATE (%)	100	100	100	100	0
OXYGEN CONCENTRATION (mg/l)	4.7	1.2	0.7	0	0



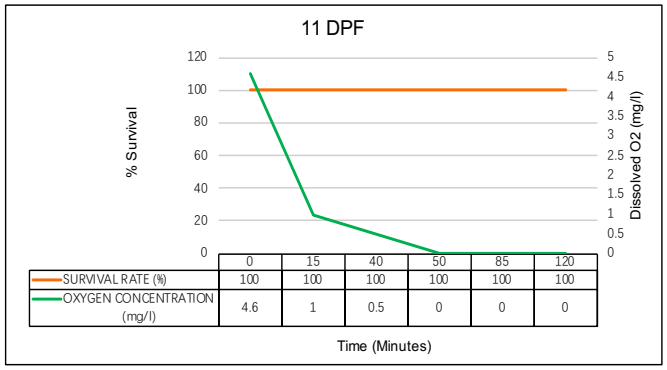
	0	15	30	45	60	75
SURVIVAL RATE (%)	100	100	100	100	30	0
OXYGEN CONCENTRATION (mg/l)	4.6	0.9	0.6	0	0	0



	0	15	30	45	50	65	75	85	90	120
SURVIVAL RATE (%)	100	100	100	100	100	100	73	73	73	60
OXYGEN CONCENTRATION (mg/l)	5.4	1.6	0.8	0.4	0	0	0	0	0	0



	0	15	30	45	55	70	120
SURVIVAL RATE (%)	100	100	100	100	87	80	80
OXYGEN CONCENTRATION (mg/l)	4.6	0.9	0.5	0	0	0	0



	0	15	40	50	85	120
SURVIVAL RATE (%)	100	100	100	100	100	100
OXYGEN CONCENTRATION (mg/l)	4.6	1	0.5	0	0	0

Figure 15. Dissolved oxygen concentrations (mg/L) and survival curve for tra catfish (*Pangasianodon hypophthalmus*) at 2, 4, 6, 8, 9, 10, 11 days post fertilization (dpf), during a low dissolved oxygen challenge. Dissolved oxygen level was reduced by replacing with the oxygen stripping. Catfish were determined moribund when the opercular movement ceased.

3.4.6 Histological analysis

In Figure 16, at 4 dpf, the vertebrae column, the notochord and vertebral column were observed at central location of the front ventral surrounded by skeletal muscle. A small cavity formed within the yolk sac that was surrounded with simple gastrointestinal structures.

At 6 dpf, moderate increase in the dimensions was found in the developing internal organs. The yolk sac was absorbed entirely at this point, and tra catfish larvae already possessed the ability of free swimming and could regulate their orientation under water. The newly observed swim bladder was situated retroperitoneally; i.e. and located ventral to the notochord and dorsal to the peritoneal cavity, encompassing a significant portion of the body cavity. At this point, the tra larvae already had improved survival ability in hypoxic conditions (0.7 mg/L).

The swim bladder and other internal organs were further developed at eight days post fertilization. The swim bladder increased slightly in size, and t two distinct lobes were present, divided by a longitudinal, central septum. The ventral surface of the gas bladder was located outside of the visceral cavity and was attached to the parietal peritoneum. During this stage,

survival in hypoxic conditions was slightly increased, as they could live in 0.6 mg/L, and could survive for a short period of time under anoxia.

At 9 dpf, the swim bladder was further expanded, the inner layer of the swim bladder consisted of cuboidal epithelium and the smooth outer layer consisting of an elastic collagenic fibrosa. The tra larvae possessed a long-term, survival rate of 60% at 0 mg/L oxygen conditions.

At 10 dpf, there were distinct changes to swim bladder morphology. The musculature was slightly thicker. The swim bladder had expanded significantly, the inner layers pushed to the outer perimeter of the organ. It was extremely dilated and presumably engorged with air. Survival rate improved to 80% at 0 mg/L O_2.

Pangasianodon larvae at 11 dpf were similar in morphology to fish at 10 dpf. However, the swim bladder expanded significantly in size and reached the outer limits of the body cavity. The bi-lobed structure was demarcated by a protuberated central obstruct. Larvae at this stagehad 100% survival rate during anoxia challenge. From 6 dpf, the tra catfish swim bladder served as an air- breathing organ, and aerial breathing ability was fully functional only five days later, at 11 dof.

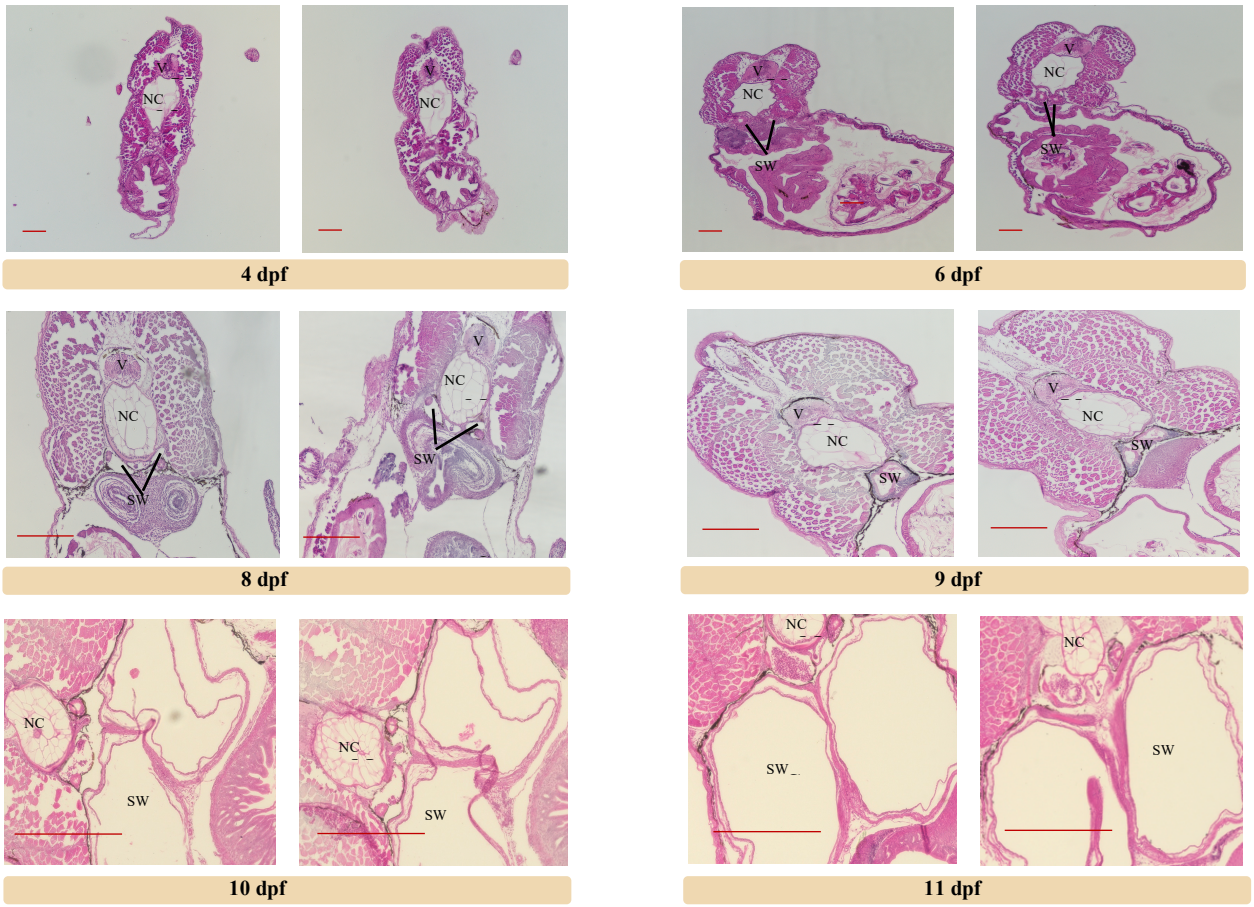


Figure 16. Transverse section, 7 um thick, from tra catfish (*Pangasianodon hypophthalmus*) 4 days post fertilization (dpf) to 11 dpf showing internal structures, including the vertebrae (V), notochord (NC) and swim bladder (SB). The vertebral column (V) was located just dorsal to the notochord (NC) and was surrounded by musculature. The developing swim bladder (SB) was distinctly bi-lobed and was located retroperitoneally.

3.4.7 Gene expression profiling of tra catfish-specific genes

To further narrow down the list of the candidate key genes that, may have a key role in the developmental of the swim bladder and on the function of aerial breathing, the expression model of the 109 tra catfish-specific genes were drawn according to the total FPKM value at each development stage, and clear variation was observed (Figure 17). The 109 tra catfish-specific genes were categorized into 6 different clusters (Supplementary Table 10). Cluster 1, 2, 5, including 19, 12 and 20 genes, respectively, in these clusters showed a peak of expression at 2 or 4 days-post fertilization, and then decreased. Cluster 3 including 16 genes showed a peak of expression values at 9 dpf. Also, there were 20 genes in cluster 4 showing an increasing expression profile and 20 genes in cluster 6 for which the expression profile remained unchanged during the development of tra catfish. Since the Mfuzz software implemented a soft clustering algorithm, each gene could be assigned to more than one cluster, which is a great advantage to reduce noise. Therefore, the total number of genes in all clusters was 129, more than our input number of 109. Furthermore, from the 0-ppm oxygen challenge and histology experiments, we concluded that the swim bladder can be observed in tra catfish larvae from 6 dpf, and the tra fish larvae possessed partial aerial breathing ability even before that, however, the fish did not possess full air breathing ability until 11 dpf. Thus, for gene expression profiles, the candidate key genes, which contributed to swim bladder development and air breathing function would be turned on before 6 dpf (maybe at 4 dpf or as early as 2 dpf) or had an increasing expression value with the formation of air breathing function over time. In addition, the genes which had an

expression value reach peak at 9 dpf may also contribute to the acceleration of the formation of ABO and the air breathing ability in late stages. In this regard, we presumed that the genes in clusters 1, 3, 4, 5 and 6 have the greatest possibility to play an important role in the swim bladder development and function of air breathing. Also, the key genes should be differentially expressed across tra catfish development. Based on these criteria, we already identified 27 genes to be tra catfish specific genes as well as differentially expressed during tra catfish development (Table 5). These genes were summarized in groups based on their expression profiles (Table 6). Except genes with unknown function, we are left with 14 genes in cluster 1, 3, 4, 5, 6 as the best candidates of air breathing function related genes in tra catfish: PET117 Cytochrome C Oxidase Chaperone (*pet117*), spidroin-2-like isoform X1 (*spidroin*), gastrin-releasing peptide-like isoform X2 (*grp*), chromo domain-containing protein cec-1-like isoform X2 (*cx3cl1*), lymphocyte antigen 6D (*ly6d*), dual specificity protein kinase shkD-like isoform X2 (*shkd*), noncompact myelin-associated protein (*ncmap*), calcium-activated chloride channel regulator 3A-1-like (*clca3a1*), calcium-activated chloride channel regulator 1-like (*clca1*), histidine-rich glycoprotein-like (*hrg*), cytochrome c oxidase subunit 7C, mitochondrial (*cox7c*), zinc finger protein 862-like (*znf862*), L antigen family member 3 (*lage3*) and protein FAM216A isoform X1 (*fam216a*).

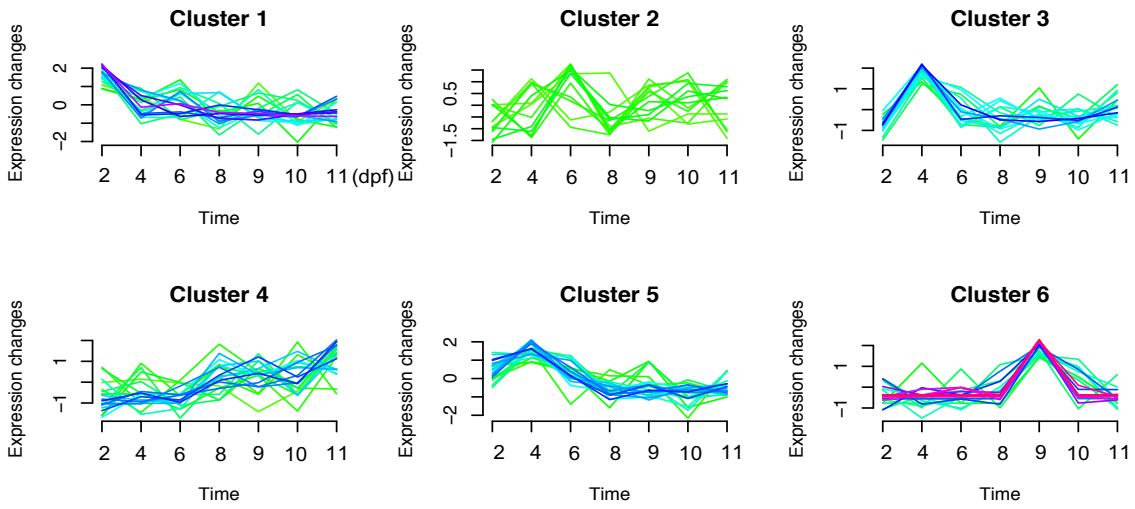


Figure 17. Time series expression profiles for tra catfish (*Pangasianodon hypophthalmus*) specific genes. These tra catfish specific genes grouped into 6 clusters. dpf- days post fertilization.

Table 6. Twenty-seven candidate key genes were grouped into 6 clusters based on their expression profiles with the development in tra catfish (*Pangasianodon hypophthalmus*).

Cluster	Gene Name	Gene ID	Gene Description
Cluster 1	PET117	113547415	PET117 Cytochrome C Oxidase Chaperone
	SPIDROIN	113547475	spidroin-2-like isoform X1
Cluster 2	CUNH3ORF85	113531560	uncharacterized C3orf85 homolog
	VXN	113541042	vexin-like
	LOC113542136	113542136	uncharacterized protein LOC113542136

Cluster 3	GRP	113535796	gastrin-releasing peptide-like isoform X2
	CL3CL1	113533685	chromo domain-containing protein cec-1-like isoform X2
	LY6D	113539350	lymphocyte antigen 6D
	SHKD	113533778	dual specificity protein kinase shkD-like isoform X2
	NCMAP	113545006	noncompact myelin-associated protein
	LOC113534207	113534207	uncharacterized protein LOC113534207
	LOC113524522	113524522	uncharacterized protein YBL113C-like
Cluster 4	CLCA3A1	113544546	calcium-activated chloride channel regulator 3A-1-like
	CLCA1	113544571	calcium-activated chloride channel regulator 1-like
	LOC113541665	113541665	uncharacterized protein LOC113541665
	LOC113542106	113542106	uncharacterized protein LOC113542106 isoform X1
	LOC113546726	113546726	uncharacterized protein LOC113546726
Cluster 5	HRG	113524613	histidine-rich glycoprotein-like

	COX7C	113536891	cytochrome c oxidase subunit 7C, mitochondrial
	ZNF862	113539654	zinc finger protein 862-like
	LAGE3	113544929	L antigen family member 3
	FAM216A	113535709	protein FAM216A isoform X1
	LOC113538378	113538378	uncharacterized protein LOC113538378
Cluster 6	LOC113524458	113524458	uncharacterized protein LOC113524458
	LOC113543977	113543977	uncharacterized protein LOC113543977
	LOC113544002	113544002	uncharacterized protein YBL113C-like
	LOC113545009	113545009	uncharacterized protein DDB_G0287625-like

3.5 Discussion

Life evolved on an anoxic Earth [33], but aerobic respiration is critical for efficient energy metabolism, which is a precondition for the beginning of complex creatures [34]. Various respiratory systems have evolved for obtaining oxygen from the environment. For water breathing fish, gills are used as the chief gas exchangers, while the decreases of dissolved oxygen in water promoted the migration of life from water to land and compelled the evolution of water breathers to air breathers [35]. Amphibians, some turtles and mammals have been

theorized to have undergone the key evolutionary processes [36-38]. In the intermediary step of aquatic to terrestrial breathing, the air-breathing organs have evolved into many different forms in different fish, including modified gills, skins, trachea, intestine and swim bladder. The swim bladder has long been postulated as a homolog of the lung in terrestrial vertebrates [39]. The swim bladder of zebrafish arose from branches of the foregut endoderm, close to liver and pancreas, which conforms with mammalian lung [40, 41]. The Wnt signaling pathways were found to play a critical role in the development of zebrafish swim bladder as well as in that of vertebrates' lung [42-44]. Moreover, Zheng et al (2011) reported a strong resemblance between the zebrafish swim bladder and mammalian lungs by transcriptome comparison [45]. These studies suggest that mammalian lungs may have originated from the teleost swim bladder, and that the genes contributed to the function of lungs may also be critical for the formation of air breathing ability in fish.

In this study, for the first time, we sequenced and analyzed the transcriptome from seven developmental stages in tra catfish, providing a comprehensive understanding of this species and their unique ability in air breathing. A total of 5,419 DEGs were identified during the early development of tra catfish. Gene ontology enrichment analysis revealed these genes were mainly enriched in blood vessel formation, respiratory chain complex, transportation and binding of oxygen, transmembrane activities, and others, which were considered to be air breathing-related categories. We then compared the genomic contents of channel catfish and tra catfish, which belong to the same order (Siluriformes). One hundred and nine unique genes were identified that

were active in tra, but absent from channel catfish, and our studies suggest that they might play key roles in the development of the swim bladder and air breathing functions in tra. Low to zero oxygen challenge experiments and histology were combined to reveal the development of air breathing capabilities in concert with the formation of the swim bladder in tra catfish. Gene expression analysis was also performed for tra catfish specific genes, and 14 genes were selected and believed to be implicated in the air breathing ability in tra catfish. Luo et. al (2016) conducted a similar analysis in 2016 [46]. They conducted Illumina RNA sequencing for the six developmental stages (4 dph (days post-hatch), 8 dph, 12 dph, 20 dph, 40 dph and one-year-old) for the posterior intestine, the ABO, of dojo loach (*M. anguillicaudatus*). According to the differential expression analysis among different developmental stages and gene expression analysis, 25 key genes were detected to be potential target genes involved in the formation of intestinal air-breathing function in *M. anguillicaudatus*. These included: GSN, YES1, CISH, RHOA, PRKCE, RAF1, BAD, SOCS3, PIK3CA, AKT1, KDR, EGFR, TP53, JUN, SMAD4, PKT2, MAPK14, GRB2, VEGFA, MYC, TNF, DVL2, ROS1, ETV5 and FZD10 [46]. None of these genes match the key genes that were identified as key ABO genes in tra, which use the swim bladder instead of the intestine for air breathing, in our study. Luo et. al (2016) also reported that these genes were seldom mentioned with air-breathing function in fish [46]. They are reported to be mainly involved in development, angiogenesis and cytoskeleton, and thus considered to contribute to the intestinal aerial breathing function formed process during posterior intestine development in dojo loach. The lack of intersection among these 25 genes

with the 14 genes we identified to be candidate key genes for the air-breathing function in tra catfish, may emphasize that using radically different organs for air breathing, also leads to the use of very different genes to accomplish this task. Further studies on the 2 species using methods such as in situ hybridization may confirm these apparent differences. The putative functional roles and related pathways of the key genes are discussed below.

PET117 Cytochrome C Oxidase Chaperone (PET117) is a protein coding gene. Diseases associated with *pet117*include Mitochondrial Complex Iv Deficiency and Charcot-Marie-Tooth Disease, Type 4K [47]. Human (*Homo sapiens*) PET117 functions the same way in cytochrome c oxidase (*cox*) biogenesis as that in yeast (*Saccharomyces cerevisiae*), although further experimentation needs to provide conclusive evidence in mammalian systems [48]. Vidoni et. al (2017) demonstrated that in the presence of PET100, PET117 interacts with myofibrillogenesis regulator 1 (MR-1S) and with some COX subunits. Such interaction proved to assist COX biogenesis in higher eukaryotes [49].

SPIDROIN-2 is also named Dragline silk fibroin 2 and is known as silk protein. Sequence analysis indicates that it belongs to the silk fibroin family, and this is a highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units. In *Golden silk orbweaver (Nephila clavipes),* spiders' major ampullate silk possesses unique characteristics of strength and elasticity. Until recently, there has been little evidence of this gene reported in teleost fishes, but its specific function in aquatic species needs further investigation [50, 51].

Gastrin-releasing peptide (GRP) is a bombesin-like peptide, generated by pulmonary neuroendocrine cells (PNEC), Many studies reported this gene to play a key role in pulmonary neuroendocrine cells grown and lembryonic lungs branching [45, 52]. As reported by Spindel et al (1987), GRP expression was elevated in human embryonic pulmonary during the canalicular phase of lung development (16 to 30 wk post fertilization). Through RNA blot and in situ hybridization analyses, GRP expression was first detected in fetal lung at 9-10 wk, reaching a plateau from 16 to approximately 30 wk, which was 25-fold higher than in mature lungs, and then declined to adult level at 34 wk post-partum. By contrast, GRP peptide levels remained elevated until several months after birth. The ephemeral high level expression of GRP during nearly 12-wk phase of embryonic lung development suggested that the secretion of GRP or its COOH-terminal peptides from pulmonary neuroendocrine cells might be closely associated with normal lung development in humans [53]. In human early pregnancy, in situ hybridization has indicated that GRP had the greatest expression in the proximal lung, and that as the lung continued to develop, the intensity of expression increased in the distal lung, suggesting that GRP gene was activated in the progression from a proximal to distal development, which was consistent with the differentiation and development of respiratory bronchioles. These related observations suggest that GRP plays a predominant role in human fetal lung development. The activation of the GRP gene is critical for normal lung growth. In addition, GRP gene is necessary for inducing the formation of primitive air saccules along respiratory bronchioles as well as the continuing extension of airway epithelium [53]. In 1984, Uddman reported that GRP may also

possess the function of regulation of local blood flow, glandular secretion and the activity of smooth muscle [54]. Martínez et al. (2005) found that the addition of GRP in human cells could increase endothelial cell migration, cord formation as well as induce angiogenesis *in vitro* [55].

GRP was reported to have similar distribution and function in mouse (*Mus musculus*) as humans. There is evidence that GRP is associated with fetal mouse lung branching during morphogenesis [56]. In mice, GRP and GRP receptor genes are expressed in embryonic lung as early as embryonic day 12, when the lung begins to branch [56]. In situ hybridization in mouse revealed that the GRP receptor reaches its highest level of expression in mesenchymal cells at cleft regions of branching airways and blood vessels [56, 57].

Fractalkine-like isoform X1 (CX3CL1) is another gene we identified in our developmental time series analysis, which is potentially to be important in the development of swim bladder and air breathing function. In human pulmonary circulation, the high blood flow and low pressure was preserved by distal arterioles with limited smooth muscles [58]. In hypoxia, human lung usually increases the air exchange area and capillary length to compensate pulmonary alveolar. Also, lung microvascular endothelium produces excessive CX3CL1, in response to hypoxia, which could stimulate phenotypic switching, proliferation, and muscle expansion in SMC [58]. Though probably related to hypoxia tolerance, little is known about this gene in fish.

The lymphocyte antigen 6 (LY6) gene family belongs to the superfamily of lymphocyte antigen-6 (LY6)/urokinase-type plasminogen activator receptor (uPAR) proteins [59]. This superfamily is characterized by a LU domain (60-80 amino acid), which is composed of 6-10

cysteines arranged in a specific spacing pattern that allows distinct disulfide bridges which create the three-fingered (3F) structural motif [59]. The three-fingered structural motif is an ancient motif. Lymphocyte antigen 6D (LY6D or Ly-6D or E48) in humans is involved in cell adhesion, lymphocyte differentiation, and response to stilbenoids [60]. It can be used as molecular marker to distinguish between B- and T-cell lymphocytes types at the earliest stage. This gene is expressed at the outer cell surface of translational epithelia and keratinocytes of stratified squamous epithelia, indicating tissue specificity [59]. A few of human orthologs in mouse of LY6 include LY6E, LY6K, Lynx1, Slurp1, and Gpihbp1. Knockout of these genes results in embryonic lethality, infertility in both sexes, visual cortex, palmoplantar keratoderma with metabolic and neuromuscular abnormal phenotype, and hypertriglyceridemia phenotypes. Most of the LY6 homologues (i.e. *Sca1, Ly6B, Ly6C, Ly6G, Ly6I/Ly6M*, and *Ly6F*) in mice are expressed in various immune cells, such as B-cells, T-cells, NK cells, monocytes, and dendritic cells, indicating involvement in immune system [61]. Interestingly, LY6K from clinical data demonstrated that overexpression of LY6K leads to a few organ cancers, including breast cancer, esophageal squamous cancer, gingivobuccal cancers, bladder cancer, and lung cancer [61]. Therefore, we speculate that overexpression of LY6 in fish may cause dysfunction of swim bladder, gill and other organs. However, in the current study, this transcript contig was too short to predict any biological structure.

Non-compact myelin associated protein (ncmap) has other alternative names, myelin protein 11 kDa, or C1orf130, or short for MP11, and plays a role in myelin formation [62]. Diseases

associated with NCMAP include Dieulafoy Lesion and Gene Ontology (GO) annotations related to this gene include structural constituents of the myelin sheath [62]. It is a membrane protein that inhibits myelination when either over- or underexpressed. MP11 expression is restricted to the placenta and peripheral nervous system, where it is expressed by Schwann cells and localized to paranodes and Schmidt-Lanterman incisures (SLIs) of non-compact myelin [63].

Calcium-activated chloride channel regulator 1 (CLCA1) and calcium-activated chloride channel regulator 3A-1 (CLCA4A1) both belong to part of larger family of CLCA proteins that has conserved domain architectures, such as, CLCA domain, VWA domain and transmembrane domain. CLCA1 factor has widely been found in many epithelial cells, endothelial cells and smooth muscle cells. Its main roles include chloride transport and mucin expression [64]. The VWA name comes from von Willebrand factor (vWF) type A domain. The von Willebrand factor is a large multimeric glycoprotein found in blood plasma, and mutation of VWA causes bleeding disorders [65]. CLCA family members have been reported in different species, including human (4 genes from hCLCA1 to hCLCA4), mouse and rat (8 homologues from mCLCA1 to mCLCA8), and cow (bCLCA1, bCLCA2 (Lu-ECAM-1), bCLCA3, and bCLCA4) [66]. There are five orthologue genes documented in zebrafish, CLCA1, CLCA5.1, CLCA5.2, CLCA1-201, CLCA1-203 and CLCA5.1-201 [67]. In our current hypoxia study, two orthologues are found, CLCA1 and CLCA3. Recent studies demonstrated that CLCA1 forms non-covalent oligomers in colonic mucus and has Mucin 2-processing properties, playing an important role in regulating the structural arrangement of the mucus and thereby partly mediating mucus

processing [68, 69]. Few reports are available on the functional analysis of CLCA3. However, CLCA3 has conserved domains with CLCA1, and share 82% identity to each other using MatGATv program (data not shown). In addition, CLCA3 has one specific domain, FN3, fibronectin type 3 domain, which is involved in cell adhesion, cell morphology, thrombosis, cell migration, and embryonic differentiation [70].

Another candidate gene for air breathing in tra was histidine-rich glycoprotein (HRG). Li et al (2018) reported that genes related to angiogenesis maybe one of the adaptations for the air-breathing organ to retain the high efficiency of gas exchange, and thus are one of the critical components for air breathing fish to adapt the low oxygen terrestrial conditions [29]. Histidine-rich glycoprotein-like (HRG) is mainly present in plasma fluid and is thought to play various roles in the human blood, such as angiogenesis, vascularization, coagulation contained and immunity [71]. For the function in angiogenesis, HRG can bind to thrombospondin (TSP) and TSP-1, which is a powerful inhibitor in angiogenesis. HRG was reported to inhibit the antiangiogenic effect of TSP-1 [72, 73].

Cytochrome c oxidase (COX) is composed of 13 subunits, three encoded by mitochondrial (mt)DNA and 10 encoded by nuclear genes. Cytochrome c oxidase subunit 7C (COX7C) is one of the last enzymes in the mitochondrial electron transport chain that drives oxidative phosphorylation. This respiratory chain catalyzes the reduction of oxygen to water. COX7C, NRF1 and PGC1α itself, in the putative PGC1α axis, showed no increase in mRNA in response to AMPK activation, while cold acclimation induced 4.1 fold increase in COX activity relative to

warm acclimated goldfish [74]. Low temperature induced mitochondrial biogenesis in many fish species. However, Duggan et al (2011) showed that not all of the subunits of COX. COX4-1, COX5A1, COX6B1, COX6C and COX7C are cold responsive in dace (*Chrosomus eos*), goldfish (*Carassius auratus*) and zebrafish (*Danio rerio*), suggesting coordination of cytochrome c oxidase gene expression in the remodeling of fish skeletal muscle [75].

Zinc Finger Protein 862 (ZNF862) in human functions in transcriptional regulation by binding metal ions and nucleic acid, and has protein dimerization activity [76]. Few reports from fish could be found. This gene only has a partial cDNA sequence in our RNA-seq data.

L antigen family member 3 (LAGE3) has a typical transmembrane domain and Pcc1 domain. Pcc1 family is conserved and can be found in yeast (*Saccharomyces cerevisiae*), such as, EKC/KEOPS complex subunit Pcc1, and also in mammals, EKC/KEOPS complex subunit LAGE3, and human cancer/testis antigen (CTAG) 1/2 [77]. Human (*Homo sapiens*) lage3 is homologue to ECK, which in both yeast and human is essential for the universal tRNA modification [78, 79].

Family with sequence similarity 216 member A (FAM216B) is predicted to have a domain FAM216B. Its family members are approximately 150 - 270 amino acids in length. In humans, the gene encoding FAM216B protein is located in the position, C13orf30. In *Pangasianodon hypophthalmus*, FAM216A has 180 amino acids in length and shows 43.75% identity to zebrafish (data now shown) and 39% to humans (NCBI blast). Function of this gene is not well identified.

VXN (vexin-like) gene belongs to cluster 2, which was a cluster differentially expressed and assumed to be involved in air breathing. However, VXN gene expression does not change over time, and likely does not appear to be associated with air-breathing ability.

The uncharacterized genes are potentially of great importance. Since the functional air breathing using the swim bladder is a unique trait and structure, perhaps that is why these genes are not characterized in other organisms.

All of the 14 characterized genes, excluding VXN, are closely related to the formation of air breathing ability in tra catfish, in which, HRG, GRP and CX3CL1 are the most important candidate genes, as they were reported to be critical for the formation and function of human lung and angiogenesis. Further transgenic, overexpression, knock out and in situ hybridization experiments will be needed to verify the molecular mechanisms and special distribution underlying air breathing functions in vertebrates and tra catfish.

3.6 Conclusion

Tra catfish are aquatic, but can use its swim bladder to breathe in air, while channel catfish cannot perform air breathing. As such, these two species provide remarkable models to study the transition from aquatic to terrestrial living, and the genes that are critical for the development of swim bladder, as well as the function of air breathing in tra catfish. Through comparative gene contents analysis between tra catfish and channel catfish, 109 genes were initially and uniquely identified to be in tra catfish, but absent from channel catfish. Hypoxia challenge and histology

experiments revealed the timepoints for the air breathing ability and swim bladder development in tra catfish. Further analysis was performed to narrow the list of candidate key genes for air breathing. Fourteen genes were ultimately selected to play an important role in the formation of air breathing ability in tra catfish. In which, HRG, GRP and CX3CL1 were confirmed to be critical in human lung growth, maintenance of function, angiogenesis as well as improving respiratory efficiency, suggesting these genes may play an important role in functioning of air breathing ability in tra catfish. Further research should include use of in situ hybridization, overexpression and gene knock-out methods to confirm the special distribution and functional expression of these genes.

Reference

1. Palzenberger, M. and H. Pohla, *Gill surface area of water-breathing freshwater fish.* Reviews in Fish Biology and Fisheries, 1992. **2**(3): p. 187-216.

2. Olson, K.R., *Vasculature of the fish gill: anatomical correlates of physiological functions.* Journal of Electron Microscopy Technique, 1991. **19**(4): p. 389-405.

3. Barrell, J., *Influence of Silurian-Devonian climates on the rise of air-breathing vertebrates.* Bulletin of the Geological Society of America, 1916. **27**(1): p. 387-436.

4. Hedrick, M.S. and S.L. Katz, *Control of breathing in primitive fishes.* Phylogeny, anatomy and physiology of ancient fishes, 2015: p. 179-200.

5. Graham, J.B., *Air-breathing fishes: evolution, diversity, and adaptation*. 1997: Elsevier.

6. Maina, J.N., *Structure, function and evolution of the gas exchangers: comparative perspectives.* Journal of anatomy, 2002. **201**(4): p. 281-304.

7. Graham, J.B. and H.J. Lee, *Breathing air in air: in what ways might extant amphibious fish biology relate to prevailing concepts about early tetrapods, the evolution of vertebrate air breathing, and the vertebrate land transition?* Physiological and Biochemical Zoology, 2004. **77**(5): p. 720-731.

8. Olson, K., et al., *Gill microcirculation of the air-breathing climbing perch, Anabas testudineus (Bloch): Relationships with the accessory respiratory organs and systemic circulation.* American journal of anatomy, 1986. **176**(3): p. 305-320.

9. McMAHON, B.R. and W.W. BURGGREN, *Respiratory physiology of intestinal air breathing in the teleost fish Misgurnus anguillicaudatus.* Journal of Experimental Biology, 1987. **133**(1): p. 371-393.

10. Podkowa, D. and L. Goniakowska-Witalińska, *Adaptations to the air breathing in the posterior intestine of the catfish (Corydoras aeneus, Callichthyidae). A histological and ultrastructural study.* Folia biologica, 2002. **50**(1-2): p. 69-82.

11. Alexander, R.M., *Physical aspects of swimbladder function.* Biological Reviews, 1966. **41**(1): p. 141-176.

12. Nguyen, T.H.P., *Effects of temperature and salinity on growth performance in cultured Tra catfish (Pangasianodon hypophthalmus) in Vietnam*. 2015, Queensland University of Technology.

13. Motta, P.J., *Mechanics and functions of jaw protrusion in teleost fishes: a review.* Copeia, 1984: p. 1-18.

14. Graham, J.B., D.L. Kramer, and E. Pineda, *Comparative respiration of an air-breathing and a non-air-breathing characoid fish and the evolution of aerial respiration in characins.* Physiological Zoology, 1978. **51**(3): p. 279-288.

15. Liem, K.F., B. Eclancher, and W.L. Fink, *Aerial respiration in the banded knife fish Gymnotus carapo (Teleostei: Gymnotoidei).* Physiological Zoology, 1984. **57**(1): p. 185-195.

16. Phuong, L.M., et al., *Ontogeny and morphometrics of the gills and swim bladder of air-breathing striped catfish Pangasianodon hypophthalmus.* Journal of Experimental Biology, 2018. **221**(3): p. jeb168658.

17. Nguyen, A.L., et al., *Simulated impacts of climate change on current farming locations of striped catfish (Pangasianodon hypophthalmus; Sauvage) in the Mekong Delta, Vietnam.* Ambio, 2014. **43**(8): p. 1059-1068.

18. De Silva, S.S. and N.T. Phuong, *Striped catfish farming in the Mekong Delta, Vietnam: a tumultuous path to a global success.* Reviews in Aquaculture, 2011. **3**(2): p. 45-73.

19. Burleson, M.L. and N.J. Smatresk, *Branchial chemoreceptors mediate ventilatory responses to hypercapnic acidosis in channel catfish.* Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 2000. **125**(3): p. 403-414.

20. Lefevre, S., et al., *Effects of nitrite exposure on functional haemoglobin levels, bimodal respiration, and swimming performance in the facultative air-breathing fish Pangasianodon hypophthalmus.* Aquatic toxicology, 2011. **104**(1-2): p. 86-93.

21. Backenstose, N., *A Histological Evaluation of the Development of Respiratory Structures in Channel Catfish (Ictalurus punctatus) and Tra (Pangasianodon hypophthalmus).* 2018.

22. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

23. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

24. Kim, O.T., et al., *A draft genome of the striped catfish, Pangasianodon hypophthalmus, for comparative analysis of genes relevant to development and a resource for aquaculture improvement.* BMC genomics, 2018. **19**(1): p. 733.

25. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

26. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.

27. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.

28. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics: a journal of integrative biology, 2012. **16**(5): p. 284-287.

29. Li, N., et al., *Genome sequence of walking catfish (Clarias batrachus) provides insights into terrestrial adaptation.* BMC genomics, 2018. **19**(1): p. 952.

30. Zhou, T., et al., *Chemokine CC motif ligand 33 is a key regulator of teleost fish barbel development.* Proceedings of the National Academy of Sciences, 2018. **115**(22): p. E5018-E5027.

31. Emms, D.M. and S. Kelly, *OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.* Genome biology, 2015. **16**(1): p. 157.

32. Kumar, L. and M.E. Futschik, *Mfuzz: a software package for soft clustering of microarray data.* Bioinformation, 2007. **2**(1): p. 5.

33. Föllmi, K.B., *Early Cretaceous life, climate and anoxia.* Cretaceous Research, 2012. **35**: p. 230-257.

34. Catling, D.C., et al., *Why O2 Is Required by Complex Life on Habitable Planets and the Concept of Planetary" Oxygenation Time".* Astrobiology, 2005. **5**(3): p. 415-438.

35. Hsia, C.C., et al., *Evolution of air breathing: oxygen homeostasis and the transitions from water to land and sky.* Comprehensive Physiology, 2013. **3**(2): p. 849-915.

36. Owen, R., *Lectures on the comparative anatomy and physiology of the invertebrate animals*. 1855: Longman.

37. Radinsky, L.B., *Evolution of vertebrate design*. 2015: University of Chicago Press.

38. Tokita, M., W. Chaeychomsri, and J. Siruntawineti, *Developmental basis of toothlessness in turtles: insight into convergent evolution of vertebrate morphology.* Evolution: International Journal of Organic Evolution, 2013. **67**(1): p. 260-273.

39. Owen, R., *Lectures on the Comparative Anatomy and Physiology of the Vertebrate Animals: Delivered at the Royal College of Surgeons of England, in 1844 and 1846*. Vol. 2. 1846: Longman, Brown, Green, and Longmans.

40. Field, H.A., et al., *Formation of the digestive system in zebrafish. ii. pancreas morphogenesis☆.* Developmental biology, 2003. **261**(1): p. 197-208.

41. Spooner, B.S. and N.K. Wessells, *Mammalian lung development: interactions in primordium formation and bronchial morphogenesis.* Journal of Experimental Zoology, 1970. **175**(4): p. 445-454.

42. Yin, A., et al., *Wnt signaling is required for early development of zebrafish swimbladder.* PloS one, 2011. **6**(3).

43. Yin, A., et al., *Expression of components of Wnt and Hedgehog pathways in different tissue layers during lung development in Xenopus laevis.* Gene expression patterns, 2010. **10**(7-8): p. 338-344.

44. Yang, Y., et al., *Comparative transcriptome analysis reveals conserved branching morphogenesis related genes involved in chamber formation of catfish swimbladder.* Physiological genomics, 2018. **50**(1): p. 67-76.

45. Zheng, W., et al., *Comparative transcriptome analyses indicate molecular homology of zebrafish swimbladder and mammalian lung.* PloS one, 2011. **6**(8).

46. Luo, W., et al., *Developmental transcriptome analysis and identification of genes involved*

in formation of intestinal air-breathing function of Dojo loach, Misgurnus anguillicaudatus. Scientific reports, 2016. **6**: p. 31845.

47. Renkema, G.H., et al., *Mutated PET117 causes complex IV deficiency and is associated with neurodevelopmental regression and medulla oblongata lesions.* Human genetics, 2017. **136**(6): p. 759-769.

48. Szklarczyk, R., et al., *Iterative orthology prediction uncovers new mitochondrial proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in the assembly of cytochrome coxidase.* Genome biology, 2012. **13**(2): p. R12.

49. Vidoni, S., et al., *MR-1S interacts with PET100 and PET117 in module-based assembly of human cytochrome c oxidase.* Cell reports, 2017. **18**(7): p. 1727-1738.

50. Gaines IV, W.A. and W.R. Marcotte Jr, *Identification and characterization of multiple Spidroin 1 genes encoding major ampullate silk proteins in Nephila clavipes.* Insect molecular biology, 2008. **17**(5): p. 465-474.

51. Correa-Garhwal, S.M., et al., *Spidroins and Silk Fibers of Aquatic Spiders.* Scientific reports, 2019. **9**(1): p. 1-12.

52. Spurzem, J.R., S.I. Rennard, and D.J. Romberger, *Bombesin-like peptides and airway repair: a recapitulation of lung development?* American journal of respiratory cell and molecular biology, 1997. **16**(3): p. 209-211.

53. Spindel, E., et al., *Transient elevation of messenger RNA encoding gastrin-releasing peptide, a putative pulmonary growth factor in human fetal lung.* The Journal of clinical investigation, 1987. **80**(4): p. 1172-1179.

54. Uddman, R., E. Moghimzadeh, and F. Sundler, *Occurrence and distribution of GRP-immunoreactive nerve fibres in the respiratory tract.* Archives of oto-rhino-laryngology, 1984. **239**(2): p. 145-151.

55. Martínez, A., et al., *Gastrin-releasing peptide (GRP) induces angiogenesis and the specific GRP blocker 77427 inhibits tumor growth in vitro and in vivo.* Oncogene, 2005. **24**(25): p. 4106-4113.

56. Aguayo, S.M., et al., *Regulation of lung branching morphogenesis by bombesin-like peptides and neutral endopeptidase.* American journal of respiratory cell and molecular

biology, 1994. **10**(6): p. 635-642.

57. King, K.A., J.S. Torday, and M.E. Sunday, *Bombesin and [Leu8] phyllolitorin promote fetal mouse lung branching morphogenesis via a receptor-mediated mechanism.* Proceedings of the National Academy of Sciences, 1995. **92**(10): p. 4357-4361.

58. Zhang, J., et al., *Hypoxia-induced endothelial CX3CL1 triggers lung smooth muscle cell phenotypic switching and proliferative expansion.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2012. **303**(10): p. L912-L922.

59. Consortium, U., *UniProt: a hub for protein information.* Nucleic acids research, 2015. **43**(D1): p. D204-D212.

60. Bateman, A., et al., *The Pfam protein families database.* Nucleic acids research, 2004. **32**(suppl_1): p. D138-D141.

61. Upadhyay, G., *Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells.* Frontiers in immunology, 2019. **10**: p. 819.

62. Safran, M., et al., *GeneCards Version 3: the human gene integrator.* Database, 2010. **2010**.

63. Ryu, E.J., et al., *Analysis of peripheral nerve expression profiles identifies a novel myelin glycoprotein, MP11.* Journal of Neuroscience, 2008. **28**(30): p. 7563-7573.

64. Evans, S.R., W.B. Thoreson, and C.L. Beck, *Molecular and functional analyses of two new calcium-activated chloride channel family members from mouse eye and intestine.* Journal of Biological Chemistry, 2004. **279**(40): p. 41792-41800.

65. Ruggeri, Z.M. and J. Ware, *von Willebrand factor.* The FASEB journal, 1993. **7**(2): p. 308-316.

66. Liu, C.-L. and G.-P. Shi, *Calcium-activated chloride channel regulator 1 (CLCA1): More than a regulator of chloride transport and mucus production.* World Allergy Organization Journal, 2019. **12**(11): p. 100077.

67. Howe, D.G., et al., *ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics.* Nucleic acids research, 2012. **41**(D1): p. D854-D860.

68. Nyström, E.E., et al., *Calcium-activated chloride channel regulator 1 (CLCA1) forms non-*

covalent oligomers in colonic mucus and has mucin 2–processing properties. Journal of Biological Chemistry, 2019. **294**(45): p. 17075-17089.

69. Dayal, A., S.F.J. Ng, and M. Grabner, *Ca 2+-activated Cl− channel TMEM16A/ANO1 identified in zebrafish skeletal muscle is crucial for action potential acceleration.* Nature communications, 2019. **10**(1): p. 1-15.

70. Petersen, T.E., et al., *Partial primary structure of bovine plasma fibronectin: three types of internal homology.* Proceedings of the National Academy of Sciences, 1983. **80**(1): p. 137-141.

71. Leung, L., et al., *Histidine-rich glycoprotein is present in human platelets and is released following thrombin stimulation.* 1983.

72. Blank, M. and Y. Shoenfeld, *Histidine-rich glycoprotein modulation of immune/autoimmune, vascular, and coagulation systems.* Clinical reviews in allergy & immunology, 2008. **34**(3): p. 307-312.

73. Wakabayashi, S., *New insights into the functions of histidine-rich glycoprotein*, in *International review of cell and molecular biology*. 2013, Elsevier. p. 467-493.

74. Bremer, K., et al., *Sensing and responding to energetic stress: the role of the AMPK-PGC1α-NRF1 axis in control of mitochondrial biogenesis in fish.* Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 2016. **199**: p. 4-12.

75. Duggan, A.T., et al., *Coordination of cytochrome c oxidase gene expression in the remodelling of skeletal muscle.* Journal of Experimental Biology, 2011. **214**(11): p. 1880-1887.

76. Zhang, L., et al., *Synthetic Zinc Finger Transcription Factor Action at an Endogenous Chromosomal Site ACTIVATION OF THE HUMAN ERYTHROPOIETIN GENE.* Journal of Biological Chemistry, 2000. **275**(43): p. 33850-33860.

77. Bateman, A., et al., *Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins.* Nucleic acids research, 1999. **27**(1): p. 260-262.

78. Kisseleva-Romanova, E., et al., *Yeast homolog of a cancer-testis antigen defines a new transcription complex.* The EMBO journal, 2006. **25**(15): p. 3576-3585.

79. Costessi, A., et al., *The human EKC/KEOPS complex is recruited to Cullin2 ubiquitin ligases by the human tumour antigen PRAME.* PloS one, 2012. **7**(8).

Chapter 4 *De novo* transcriptome assembly and gene expression analysis in seven developmental stages of *Clarias macrocephalus* to investigate the evolution innovation of air-breathing

4.1 Abstract

Walking catfish (*Clarias macrocephalus*) and channel catfish (*Ictalurus punctatus*) are freshwater fish species of the Siluriformes order. *C. macrocephalus* has both gills and modified gill structures serving as an air breathing organ (ABO) which allows them aerial breathing (AB), while *I. punctatus* does not possess an air-breathing organ (ABO), and thus cannot breathe in air. These two species provide an excellent model for studying the molecular basis of accessory air-breathing organ development in teleost fish. In this study, seven development stages in *C. macrocephalus* were selected for RNA-seq analysis to compare with channel catfish as the time when air breathing developed and became functional. More than 504 million clean reads were ultimately generated in *C. macrocephalus*, and a total of 25,239 expressed genes were detected and annotated. Subsequently, 8,675 differentially expressed genes (DEGs) were identified among different developmental stages. Through comparative genomic analysis between *C. macrocephalus* and *I. punctatus*, 1,458 genes were identified to be present in *C. macrocephalus,* but absent from *I. punctatus*. Gene expression analysis and protein-protein intersection (PPI) analysis were performed to select the top candidate genes involved in the formation of air-

breathing function. Finally, 26 genes were selected in *C. macrocephalus*, including *mb*, *ngb*, *hbae* genes, which are mainly associated with oxygen carrier activity, oxygen binding and heme binding activities. Hypoxia challenge suggested the timepoints for the functioning and development of air breathing ability in *C. macrocephalus*.

This study provides a large data resource for exploring the genomic basis of air breathing function in *C. macrocephalus* and offers an insight into the adaption of aquatic organisms to hypoxia and high ammonia environment.

4.2 Background

Scientists have been long focused on the transition from aquatic to aerial gas exchange in vertebrates. Fish that directly breath air, i.e., conduct aerial respiration, provide critical evidence for the evolution of life in the ocean to terrestrial living [1]. Approximately 350 million years ago, there is a change in the environment, the higher temperature and the decay of dead organic components that used up dissolved oxygen in swamps, rivers and lakes, which resulting in a gradually decline in dissolved water concentration [2]. Certain fish ancestors start to develop air-breathing organs as a result of lack of dissolved oxygen in water. Some of them left water environments and colonized the land; the progeny of others are air-breathing fish [3]. Air breathing fish can perform gas exchange directly with the atmosphere. Some may rise to the surface of water to gulp air while others crawl onto land and survive for extended periods of time [4].

It is estimated that there are more than 370 air breathing fish species in 49 families [5]. The air-breathing organs (ABO) vary considerably among different fishes. *C. macrocephalus* and *C. batrachus* have the air-breathing organs evolved from the gill. In these fish, the efferent branchial arteries of the anterior (first and second) gill arches are sites for gas exchange and so act as accessory ABOs [6-8]. Some other species use a modified intestine as an ABO, such as *Misgurnus anguillicaudatus.* In such cases*,* the posterior region of intestine is highly modified: it is well vascularized with intraepithelial capillaries, which provide a suitable place for gas exchange [9, 10]. Other fish use their swim bladder as an ABO. These include the *Gonorynchiformes* [11], *Characiformes* and some species in *Siluriformes*, such as suborders *Gymnotus* and *Pangasianodon hypophthalmus* [12, 13]. The swim bladder extends from the posterior of the head to the tail beyond the anus and serves as the major place for gas exchange in aerial breathing [14].

Walking catfish *(Clarias macrocephalus)* belong to the Siluriformes, and are widely distributed in Asia, the Indian subcontinent and Africa and has great economic value as a food fish [15]. *Clarias* has both gills and a modified gill structure, which serves as an ABO. Therefore *Clarias* is well adapted to hypoxic conditions in muddy marshes [16]. It burrows into the mudflats during summer periods, staying alive through direct air-breathing [17-19]. After a heavy rainfall, *Clarias* can make slither, snake-like, in order to travel across the land; thus it is also known as the walking catfish [20, 21]. Its “walking” ability allow *Clarias* to survive in extreme environments, such as hypoxia, desiccation stress and high ammonia, which are not

ideal conditions for aquaculture [18]. This capability makes *Clarias* a perfect model to study the evolution of adaptions to air breathing, as well as the mechanisms for hypoxia and ammonia tolerance.

Previous studies have demonstrated the genetic basis for development transcriptome analysis of intestinal air-breathing in *M. anguillicaudatus*, and in adult *C. batrachus* through transcriptome analysis [10, 19]. However, little is known about the underlying mechanisms in the formation and function of the ABO in *C. macrocephalus* during early developmental stages. Our research filled this knowledge gap and advance the understanding of the molecular basis of accessory ABO organ development in fish, as well as adaptions to terrestrial life. In this study, two hypoxia challenges were performed to reveal the development and functioning of the ABO in *C. macrocephalus*, along with RNA-seq analysis of seven early development stages of *C. macrocephalus* to reveal the genomic features that potentially contribute to air breathing and terrestrial adaptations.

4.3 Material and Methods

4.3.1 Ethics statement

This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States and in Vietnam. All experimental protocols involving animal care and tissue collection were approved

by the Auburn University and Can Tho University Institutional Animal Care and Use Committee.

4.3.2 Experimental animals and tissue collection

Walking catfish (*Clarias macrocephalus*) samples were collected from Can Tho University, Vietnam in June of 2017. *C. macrocephalus* fertilized eggs were cultured in a flowing water at 26.5 ± 2 ºC. Twenty to fifty *C. macrocephalus* embryos or larvae were harvested every 24 hours over a 30-day period after hatching. Samples were euthanized with 200 ppm buffered MS-222 (Finquel, Argent Chemicals) and stored in RNA Later solution (Thermo Fisher Scientific). Samples were shipped to the US with dry ice and stored at -80°C until to RNA extraction.

4.3.3 Oxygen challenge for *Clarias macrocephalus*

Two oxygen challenge experiments were conducted at Can Tho University, Vietnam, to determine the air breathing ability of *C. macrocephalus*. *Clarias* larvae older than three days post-fertilization (dpf), which had completely absorbed the yolk sac, were challenged with low dissolved oxygen each day from 3 days to 20 dpf. Fifteen to twenty *Clarias* larvae were placed in a two-liter container with oxygen supply as a control treatment. A separate group was stocked in a second two-liter container, supplemental oxygen was removed, and the dissolved oxygen level was decreased by bubbling nitrogen gas into the water until a 0 mg/L dissolved oxygen concentration was obtained (measured by DO meter). Dissolved oxygen levels, larval behavior

and survival rate were observed and recorded every 10 minutes, and the results at each time point calculated and response to hypoxia represented graphically.

Moist adult *Clarias* can leave the water and walk on land and live indefinitely. Therefore, another oxygen challenge experiment was conducted in a petri dish test simulate this situation. *Clarias* larvae were exposed directly to air and challenged each day from 3 dpf to 27 dpf. To do this, one group of 15-20 larvae were placed in a Petri dish without water and misted to keep them wet. This treatment forced the fish to obtain O_2 from the air. In a second treatment, fish were identically loaded into a aerated two-liter water-filled tank with as a control treatment for each experiment. Larval behavior and survival rate were observed and recorded every 10 minutes.

4.3.4 RNA extraction, library construction and sequencing

Seven time points: 3, 5, 13, 14, 16, 17, 24 days post-fertilization (dpf), were selected based on the low oxygen challenge results. Samples were collected at each time point and stored in the -80°C freezer. Two biological replicates were included for timepoint and treatment group. Pooled samples of 4 individual fish were homogenized in liquid nitrogen and ground to a fine powder using mortar. RNA was extracted using a RNeasy Plus Kit (Qiagen, Valencia, California) following the manufacturer's directions. For each time point, equal amounts of RNA from the two pooled replicates were used for library construction and Illumina RNA sequencing.

Library construction and sequencing were performed by GENEWIZ, LLC. (South Plainfield, NJ, USA). RNA integrity of each sample was measured by 4200 TapeStation (Agilent

Technologies, Palo Alto, CA, USA). Ribosomal RNA depletion method was carried out using a Ribozero rRNA Removal Kit (Illumina, San Diego, CA, USA). Sequencing libraries were checked with the Agilent Tapestation 4200 (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (Applied Biosystems, Carlsbad, CA, USA). The sequencing libraries were multiplexed, clustered on three lanes of a flowcell and loaded on the Illumina HiSeq instrument with a 2x150 Paired End (PE) configuration. Raw sequence data (.bcl files) generated from the HiSeq were converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software.

4.3.5 *De novo* assembly and gene annotation

De novo assembly was performed on *C. macrocephalus* cleaned reads using Trinity (version 2.8.5) [22]. Prior to assembly, raw reads quality was evaluated in FastQC (version 0.11.5) [23], and raw reads were filtered by removing adaptor sequences and ambiguous nucleotides using Trimmomatic (version 0.36) [24]. Reads with quality scores less than 25 and length shorter than 36 bp were all trimmed. The remaining high-quality reads were used in subsequent assembly. In brief, the clean reads of the seven libraries were jointly assembled into unique sequences of transcripts in Inchworm via greedy K-mer extension (K-mer 25). After mapping of reads to Inchworm contigs, Chrysalis incorporated reads into de Bruijn graphs. Butterfly ultimately generated full-length transcripts. The assembled transcriptome was passed to CD-HIT (version 4.7) to reduce redundancy with 95% identity [25].

The non-redundant contigs were annotated against the UniProt database (Swissprot), channel catfish (*Ictalurus punctatus*) database and zebrafish (*Danio rerio*) database using BLASTX program (version 2.6.0) [26]. The cutoff Expected value (E-value) was set at 1e-5, and only the top hit result was allocated as the annotation for each contig.

4.3.6 Identification and analysis of differentially expressed genes

The merged transcriptome assembly served as the reference transcriptome. The high-quality reads from each sample were mapped back to it using the align_and_estimate_abundance.pl package in Trinity [22]. The mapped reads number for each transcript was detected using RSEM software [27]. Differential expression analysis was performed between two adjacent time points using an R package DESeq2 (3 dpf vs. 5 dpf, 5 dpf vs. 13 dpf, 13 dpf vs. 14 dpf, 14 dpf vs. 16 dpf, 16 dpf vs. 17 dpf, 17 dpf vs. 24 dpf) [28]. Differentially expressed genes (DEGs) were detected with a criterion of $|\log2 \text{ fold change}| \geq 1$ (log2 fold change) and adjusted P-value < 0.05. The distribution of upregulated and downregulated genes was demonstrated in a volcano plot.

4.3.7 Gene ontology and functional enrichment analysis

To identify the overrepresented GO terms with in the DEGs, GO terms for each gene were assigned by using zebrafish annotations for the Unigenes set. Gene Ontology (GO) analysis and enrichment analysis were performed using the ClusterProfiler R package (version 3.6) [29]. The final annotation result was categorized in terms of biological process, cellular component and

molecular functions. ClusterProfiler was also used for a GO functional enrichment analysis of certain genes. P-value and q-value cutoff criteria of 0.05 were used as thresholds of significance.

4.3.8 Identification of orthologous protein groups

The protein sequences of *I. punctatus* and *C. macrocephalus* were obtained from NCBI based on sequence orthology [30], according methods in the literature [19, 31]. Orthologs and orthogroups between *I. punctatus* and *C. macrocephalus* were detected with OrthoFinder software (version 2.2.7) [32]. To identify the *Clarias*-specific genes, a second round of Protein BLAST (BLASTP) was performed for genes with no match in the orthologue groups with a criterion of maximum e-value of 1e-10 [33]. Subsequently, a reciprocal BLASTP search was carried out to query genes with no hits from previous steps with a maximal e-value of 1e -5. These leftover genes are recognized as *Clarias*-specific gene candidates for subsequent analysis.

To further define the *Clarias*-specific genes from the previous step, the species-specific genes in *Clarias* were aligned with the channel catfish genome using TBLASTN with a maximum e-value of 1e-10 [33]. The identified *Clarias*-catfish specific genes were then filtered with the percentage of identical matches (pident) and query coverage per subject (qcovs). The remaining genes without a TBLASTN hits in channel catfish genome were finally classified as the *Clarias*-specific genes absent from the channel catfish genome.

Ultimately, the genes that are present in *C. macrocephalus* but absent from *I. punctatus*, and differentially expressed for at least in one time point over *C. macrocephalus* development, were

recognized to be critical genes involved in the morphogenesis of the ABO and for differences in aerial breathing ability.

4.3.9 Clustering of gene expression and protein-protein (PPI) network analysis of the key genes

GO enrichment analysis of the key genes was performed using ClusterProfiler R package (version 3.6) [29], and a q-value < 0.05 were used as the threshold of significance. A soft clustering software, Mfuzz, revealed expression patterns of key genes through development stage and assigned a general classification based on expression profiles [34]. In addition, the STRING database was used to investigate the network properties of the enzymes encoded by the key genes, as well as providing both enrichment and experimental information [35].

4.4 Results

4.4.1 Oxygen challenge

Low oxygen challenge experiments were performed to test the survival of *C. macrocephalus* in 0 ppm dissolved oxygen (DO) (Supplementary Table 11) and in an aerial environment (Supplementary Table 12). The survival rate of *Clarias* larvae at each time point was calculated, and statistical graphics were drawn in Figure 18 to show the changes in survival rate over time. At 3 dpf, *C. macrocephalus* larvae had a survival rate of 93.3% when oxygen was decreased to 0.5 ppm. Initially, when the aeration was removed from the container, the oxygen concentration

was at 5.3 mg/L, and the fish were active and swam on the bottom as normal. Then after 5 minutes, when the dissolved oxygen level dropped below 3.9 mg/L, they had a little slower movement and came closer together at the bottom. After 15 minutes, when the dissolved oxygen level in water dropped to 1.7 mg/L, the fish moved slower, and were still at the bottom. After 25 minutes, when the dissolved oxygen level reached 1.2 mg/L, fish were swimming slower and some tried to move to the surface. This behavior remained the same until the dissolved oxygen concentration decreased to 0.6 mg/L at 75 minutes into the challenge, one fish died at this time, and some fish moved to the water surface and then back to the bottom. Then, from 110 minutes after the beginning of the challenge, at each time about 7-8 fish tried to move to the water surface and then went back to the bottom, but no more fish died.

For the challenge at 5 dpf, the oxygen concentration fell from 5.5 mg/L to 2.5 mg/L at 20 minutes after removing the aeration, and fish swam at the bottom. At 25 to 165 minutes, the dissolved oxygen level dropped from 2.1 mg/L to 0.5 mg/L, the fish were slowly swimming, and some fish up to the surface. At 195 minutes, the dissolved oxygen levels was 0.5 mg/L and all fish swam slowly at the surface. After 405 minutes, the dissolved oxygen concentration was of 0.3 mg/L, the fish remained moving slowly at the surface, and no fish died at this point.

The dissolved oxygen level started at 5.4 mg/L for the challenge at 13 dpf, and the fish were actively swimming within the container. Thirty minutes into the challenge, the dissolved oxygen level fell to 0.6 mg/L, and some fish started swimming to the surface and moved slowly. At 75 minutes the dissolved oxygen level decreased to 0.4 mg/L, and all fish were at the surface

moving slowly. There was an 80% survival rate for *C. macrocephalus* larvae at a dissolved oxygen concentration of 0.4 mg/L.

The challenge at 14 dpf started with a dissolved oxygen reading of 5.8 mg/L. Oxygen level was lowered to 1.3 mg/L, then 0.6 mg/L at 15 and 40 minutes into the challenge, respectively, and fish slowly swam at the bottom at this time. At 100 minutes, the dissolved oxygen concentration decreased to 0.4 mg/L, some fish came to the surface and moved slowly. At 130 minutes, the dissolved level was 0.4 mg/L, and two fish died. After 250 minutes, all fish swam to the surface and moved slowly. The death rate remained the same from 130 minutes into challenge, and there was a 90 % survival rate for *C. macrocephalus* larvae at a dissolved oxygen level of 0.4 mg/L.

When the fish were 16 dpf, the challenge began with a dissolved oxygen level of 5.2 mg/L, and fish were swimming normally. Oxygen concentration dropped to 0.9 mg/L at 20 minutes, and some fish swam to the surface and moved slowly. Then the dissolved oxygen level decreased and stabilized at 0.4 mg/L from 20 to 390 minutes into the challenge, the fish were moving slowly, and swam to the surface. No fish died. The survival rate was 100% during this stage.

At 17 dpf, the challenge began at 5 mg/L. The dissolved oxygen concentration decreased to 0.4 mg/L at 20 minutes. Some fish were moving slowly and swam to the surface. The dissolved oxygen remained at 0.4 mg/L The challenge concluded after 420 minutes with 100% survival rate of the *C. macrocephalus* larvae.

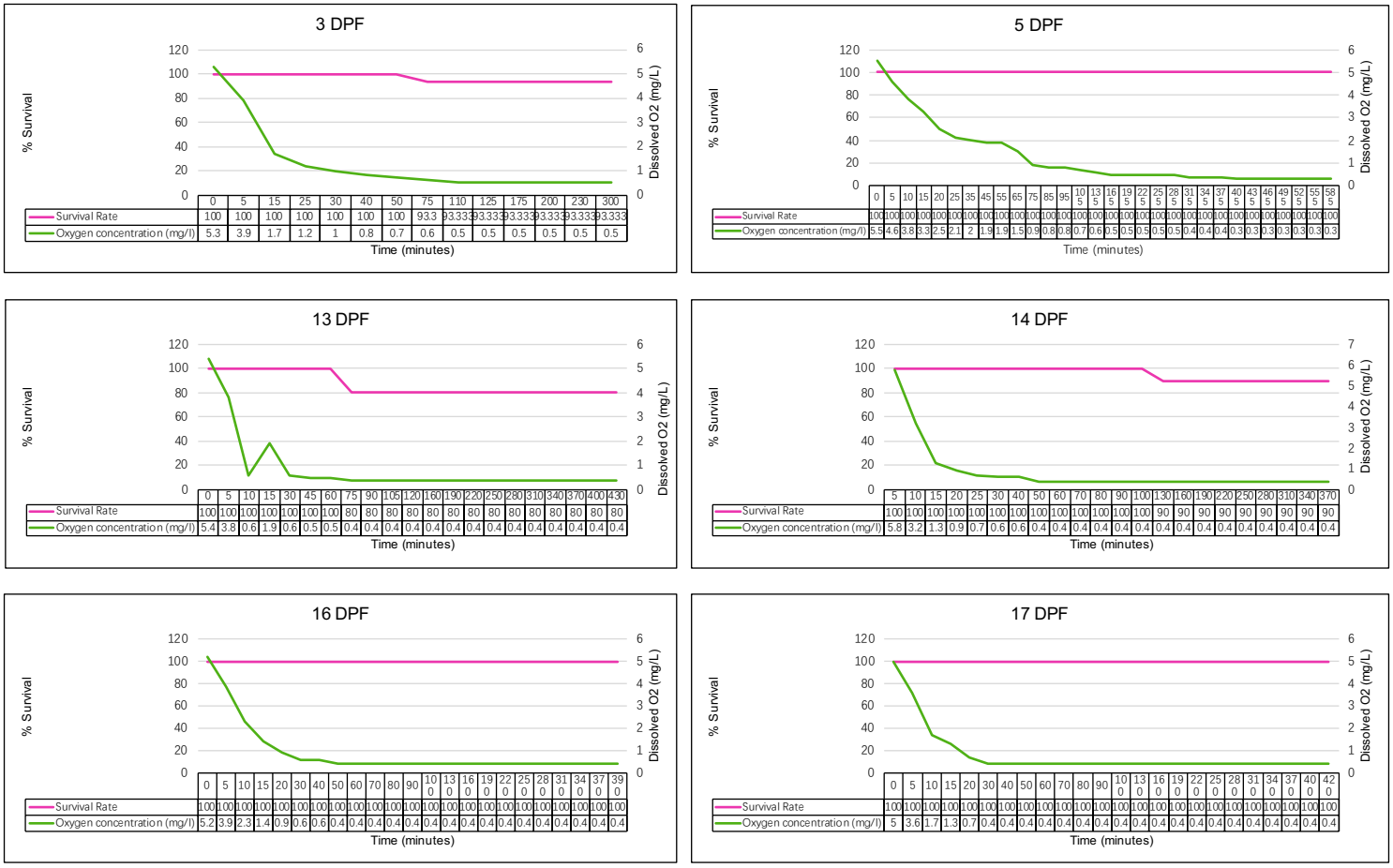


Figure 18. Dissolved oxygen concentrations (mg/L) and survival curve for *Clasias macrocephalus* at 3, 5, 13, 14, 16, 17, 24 days post-fertilization (dpf), during a low dissolved oxygen challenge. Dissolved oxygen level was reduced by replacing with the oxygen stripping. Fish were determined moribund when the opercular movement ceased.

In the Petri dish experiments, *Clarias* larvae were exposed to the air directly and challenged each day from 3 dpf to 27 dpf. At 3 dpf, two fish died after 25 minutes, and the survival rate was 86.7% (Figure 19). At 30 minutes, two additional fish died, the fish were not active and then stopped moving. Survival rate was 73.3%. Survival rate decreased to 53.3%, 40% and 26.7% at

47, 60 and 75 minutes, respectively, and fish stopped moving during these stages. At 80 minutes, all fish died. There was a 0 % survival rate for *C. macrocephalus* in the aerial environment.

At 5 dpf, *Clarias* larvae had a survival rate of 100% at 45 minutes into the challenge, but fish were not active and stopped moving at this period. Then at 62 minutes, two fish died, survival rate was reduced to 90%, and fish stopped moving. At 70, 80, 95 and 105 minutes, the survival rated decreased to 75%, 55%, 20% and 10%, respectively. After 108 minutes, all fish had died. For 13 dpf larvae, fish were not active at 25 minutes and stopped moving. At 65 minutes, fish stopped moving and a bubble appeared around fish body. Fish had a survival rate of 100% until exposed to the air for 93 minutes. At 93 minutes into challenge, one fish died, and survival rate was 95%. Then the survival rate decreased to 70%, 45%, 20% and 5% at 125 minutes, 186 minutes, 248 minutes and 365 minutes, respectively. After 382 minutes, all fish had died.

At 14 dpf, all fish survived through air breathing for 105 minutes. From 25 minutes into the challenge onward, fish stopped moving, and a bubble appeared around the fish body. At 126 minutes, two fish died, and survival rate decreased to 90%. Then the survival rated decreased to 70%, 45% 20% and 5% at 142 minutes, 157 minutes, 191 minutes and 283 minutes, respectively. At 291 minutes, all of the fish were dead.

When the fish were 16 dpf, fish stopped moving at 25 minutes after exposure to the air, and a bubble started appearing around the body from respiration. No fish died until 145 minutes into

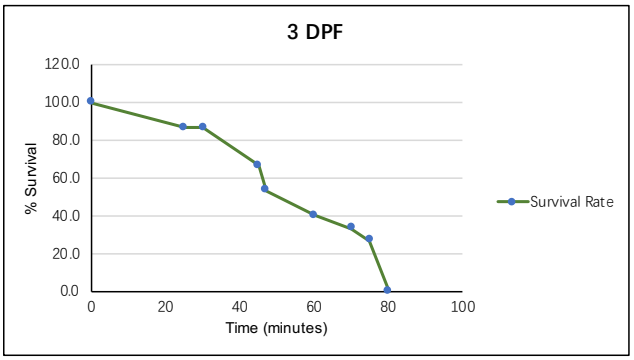
the challenge. The survival rate continuously decreased from 90% at 157 minutes to 10% at 321 minutes. At 327 minutes, all fish had died.

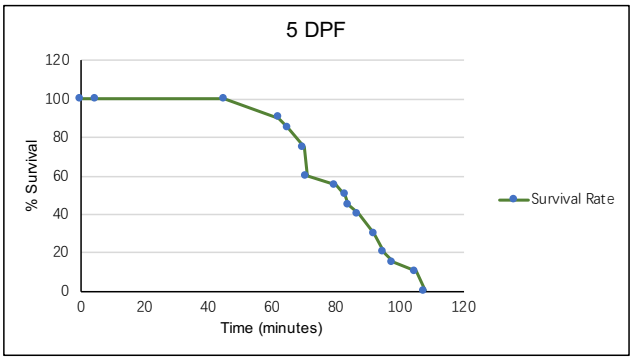
At 17 dpf, fish stopped moving 25 minutes into the challenge, and a bubble started appearing around the fish. All fish survived in the aerial environment for 265 minutes through air breathing. At 285 minutes, two fish died, dropping survival rate to 90%. Then the survival rate decreased to 70%, 50% and 45% at 317 minutes, 382 minutes and 397 minutes, respectively. Then no additional fish died, the survival rate remained the same from 397 minutes onward. Fish possessed a final survival rate of 45%.

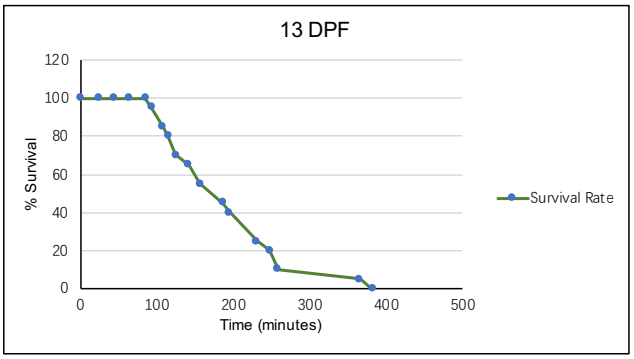
Again, fish stopped moving at 25 minutes into the air challenge at 24 dpf, and no fish died until 289 minutes into the challenge. Then four fish died after exposure to the air after 316 minutes, resulting in 95% survival. The survival rate remained the same until 51 minutes later, when four additional fish died (80% survival). The challenge concluded at 420 minutes, and there was a 60% survival rate for *C. macrocephalus* in the aerial environment.

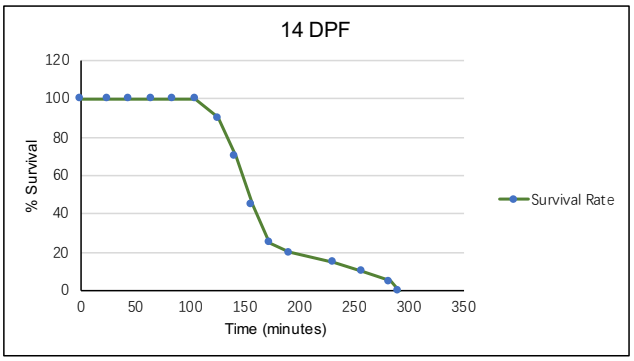
In summary, from 3 dpf, the first day after *C. macrocephalus* completely absorbed all of the yolk sac, the fish already possessed partial aerial breathing ability. At that age, some fish could move to the surface and had air gulping behavior. From 3 dpf to 16 dpf, the air breathing ability gradually increased with the development of *C. macrocephalus.* However, no larvae could survive in air environment for an indefinite period during these stages. Forty-five percent fish could survive through air breathing at 17 dpf, sixty percent at 24 dpf in the aerial environment, as

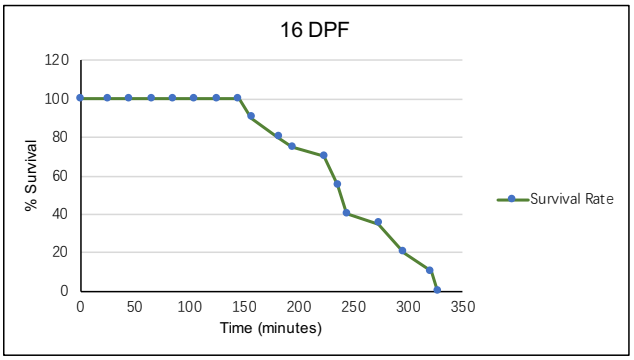
this proportion of these fish could breathe air and possessed fully air breathing ability during this stage.

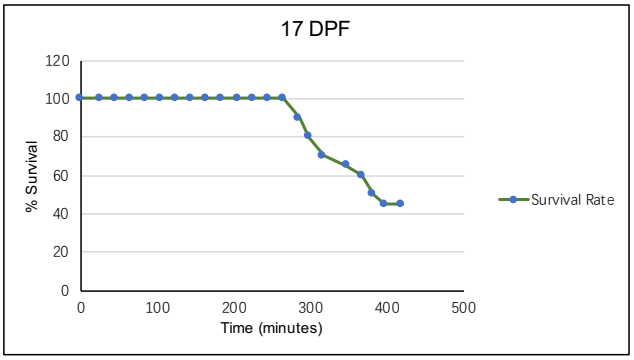












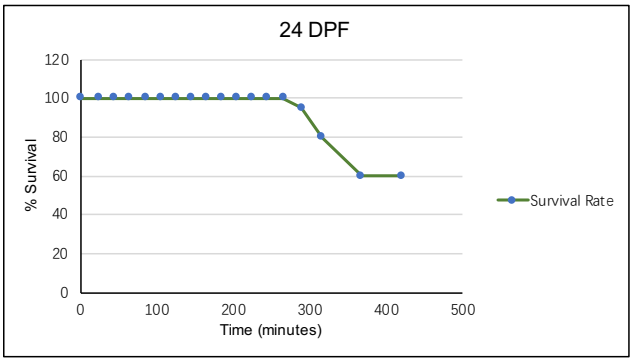


Figure 19. Survival curve for *Clarias macrocephalus* exposed to the aerial environment at 3, 5, 13, 14, 16, 17, 24 days post-fertilization (dpf). Fish were determined moribund when the opercular movement ceased.

4.4.2 Transcriptome sequencing and *de novo* assembly

A total of 1,331 million raw reads was generated for *Clarias* through RNA sequencing. After trimming the low-quality reads (quality score < 25), very short reads (< 36 bases) and adapter sequences, approximately 504 million clean reads were ultimately retained for *Clarias* (Supplementary Table 13). Trinity software was used to assemble the clean reads obtained from the seven different transcriptome timepoints and an assembled full-length transcriptome was generated. After removing the redundant transcripts using CD-HIT tool at 90% identity, approximately 1,841,162 transcripts remained with an average contig length of 829 bp and N50 size of 1,390 in its assembly (Table 7). The average ORF length was 592 bp, at a GC content of 42.3%.

Table 7. The sequencing data output and quality assessment on the RNA-seq results of walking catfish (*Clarias macrocephalus*).

	Clarias macrocephalus
Assembly	
Transcript number	1,841,162
Genes	1,004,981

Total assembled bases	1,525,751,722
GC%	42.3%
Average contig length	829
Average ORF length	592
N50	1,390
N50 longest isoform	799
Annotation	
Unigenes	25,239

4.4.3 Gene identification and annotation

To maximize the information of assembled transcripts, a BLAST-based gene identification was carried out to annotate the *Clarias* transcriptome and inform following differential expression analysis. All of the transcript sequences were searched against three databases: UniProt protein database (SwissProt), the channel catfish database and zebrafish database, based on a criterion of E-value of 1e-5. The best hit contigs were retained, with a total of 25,239 unigenes identified among these three databases (Table 7).

4.4.4 Identification and analysis of differentially expressed genes

Differentially expressed gene analysis was performed to identify the DEGs with comparison of each stage with the previous stage. In *Clarias*, the number of DEGs varied from 5,992 (5,653

up-regulated, 339 down-regulated), between 13 dpf to 5dpf, to 449 (206 up-regulated, 243 down-regulated) between 24 dpf and 17 dpf (Supplementary Table 14). In detail, 5,478 (271 up-regulated, 5,027 down-regulated),1,906 (1,258 up-regulated, 648 down-regulated), 2,703 (1,152 up-regulated, 921 down-regulated), 1,718 (1,105 up-regulated, 613 down-regulated). DEGs were detected between 5 dpf and 3 dpf, 14 dpf and 13 dpf, 16 dpf and 14 dpf, 17 dpf and 16 dpf, respectively. Eight thousand, six hundred fifty seven DEGs were ultimately identified over different developmental stages in *Clarias*. Generally, the number of identified DEGs decreased during the developmental stages in *Clarias* (Figure 20), except for a slight fluctuation between 13 dpf and 5 dpf, which is not surprising since it is a relatively long interval from 13 to 5 dpf. Multiple biological processes would be expected to 'turn on' during this time, which would result in a large amount of DEGs. A Venn diagram was drawn to show the intersection between DEGs in different stages (Supplementary Figure 6). Moreover, a Volcano plot was constructed to detect the transcripts significantly changed during the early development in *Clarias* (Figure 21).

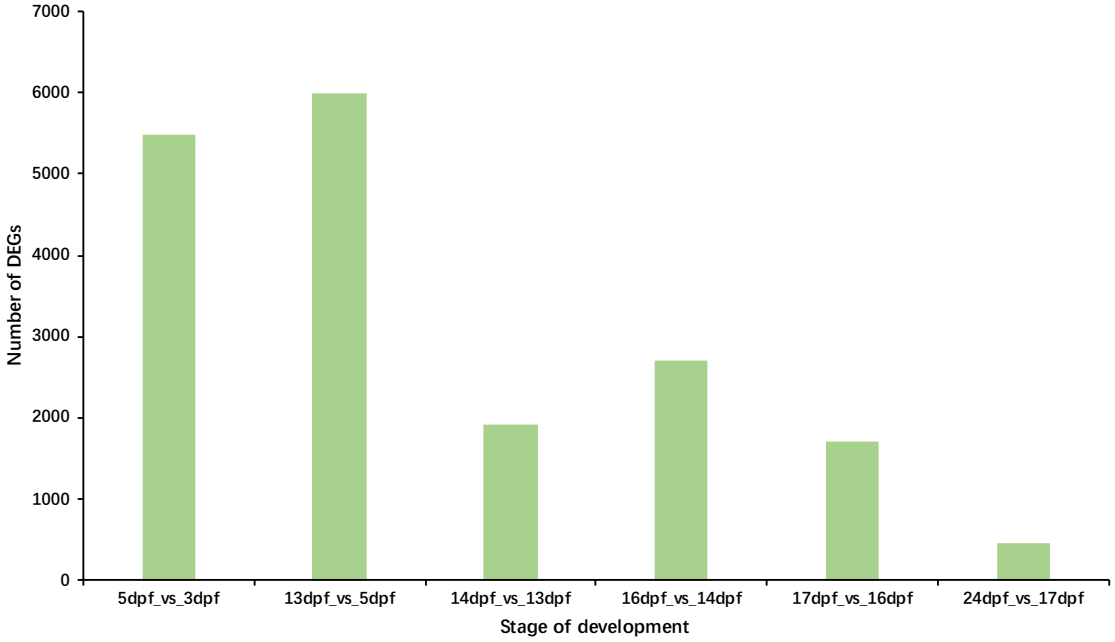


Figure 20. Gene expression during early embryonic development in *Clarias macrocephalus.* The number of differentially expressed genes (DEGs) were detected for comparison of each stage with the previous stage. Differential expression of genes peaked between 13 dpf and 5 dpf stage. Over time, the number of DEGs generally decreased.

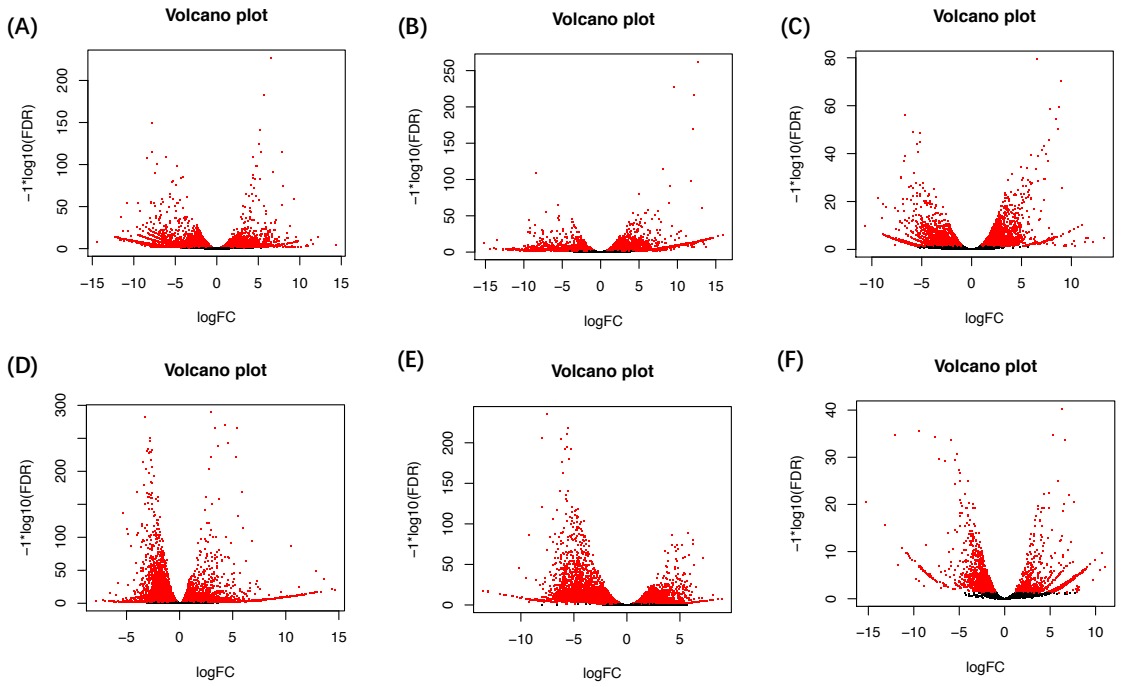


Figure 21. Volcano plot of the differentially expressed genes (DEGs) in different comparisons during development in *Clarias macrocephelus.* Red dots indicate downregulation (negative value) and upregulation (positive value). Black dots represent non-DEGs. (A) 5 days post-fertilization (dpf) vs. 3 dpf; (B) 13 dpf vs. 5 dpf; (C) 14 dpf vs. 13 dpf; (D) 16 dpf vs. 14 dpf; (E) 17 dpf vs. 16 dpf; (F) 24 dpf vs. 17 dpf.

4.4.5 Gene ontology enrichment analysis of differentially expressed genes at different stages

To classify the gene ontology (GO) enrichment category with related functions, the enrichment analysis of DEGs at different development stages was conducted, and GO categories significantly enriched during *Clarias* development are listed in Supplementary Table 14. The enriched GO terms for each stage are showed in Figure 22.

Between days 3 and 5, the DEGs were mainly related to hemopoiesis (GO:0030097), response to oxidative stress (GO:0006979), aerobic respiration (GO:0009060), response to decreased oxygen levels (GO:0036293), response to hypoxia (GO:0001666), blood vessel development (GO:0001568), angiogenesis (GO:0001525), activation of MAPK activity (GO:0000187), Ras protein signal transduction (GO:0007265), Wnt signaling (GO:0016055), ATP generation from ADP (GO:0006757) and ATP metabolic process (GO:0046034).

Between days 5 and 13, DEGs were mainly enriched for aerobic respiration (GO:0009060), response to oxidative stress (GO:0006979), response to hypoxia (GO:0001666), ATP metabolic process (GO:0046034), respiratory electron transport chain (GO:0022904), respiratory chain complex (GO:0098803), ATP synthesis coupled electron transport (GO:0042773), ADP metabolic process (GO:0046031) and NADH dehydrogenase complex (GO:0030964).

Between days 13 and 14, enriched categories for the DEGs were mainly related to cardiac muscle cell development (GO:0055013), blood coagulation (GO:0007596), myosin complex (GO:0016459), myosin filament assembly (GO:0031034), ligase activity, forming carbon-oxygen bonds (GO:0016875), lipid transporter activity (GO:0005319), steroid binding (GO:0005496) and ATPase activity, coupled to movement of substances (GO:0043492).

For 16 dpf compared 14 dpf stage, the DEGs were mainly enriched for erythrocyte homeostasis (GO:0034101), ATP metabolic process (GO:0046034), ATP generation from ADP (GO:0006757), ATP hydrolysis coupled transmembrane transport (GO:0090662), ATP hydrolysis coupled ion transmembrane transport (GO:0099131), electron transport chain

(GO:0022900), energy derivation by oxidation of organic compounds (GO:0015980), aerobic electron transport chain (GO:0019646), myosin filament organization (GO:0031033) and NADH dehydrogenase complex (GO:0030964).

Between days 16 and 17, the DEGs were mainly related to blood coagulation (GO:0007596), gas transport (GO:0015669), oxygen transport (GO:0015671), mitochondrial respiratory chain complex IV (GO:0005751), oxygen binding (GO:0019825), oxygen carrier activity (GO:0005344), hemoglobin complex (GO:0005833), response to oxidative stress (GO:0006979), morphogenesis of an epithelial sheet (GO:0002011), respiratory chain (GO:0070469), ATP metabolic process (GO:0046034), NADP metabolic process (GO:0006739) and NAD metabolic process (GO:0019674).

Additionally, DEGs between 17 and 24 were enriched in gas transport (GO:0015669), oxygen transport (GO:0015671), oxygen binding (GO:0019825), oxygen carrier activity (GO:0005344), respiratory electron transport chain (GO:0022904), ATP synthesis coupled electron transport (GO:0042773), striated muscle contraction (GO:0006941), heart contraction (GO:0060047), hemoglobin metabolic process (GO:0020027), hemoglobin complex (GO:0005833), myofibril (GO:0030016), myosin complex (GO:0016459), voltage-gated sodium channel complex (GO:0001518) and heme binding (GO:0020037).

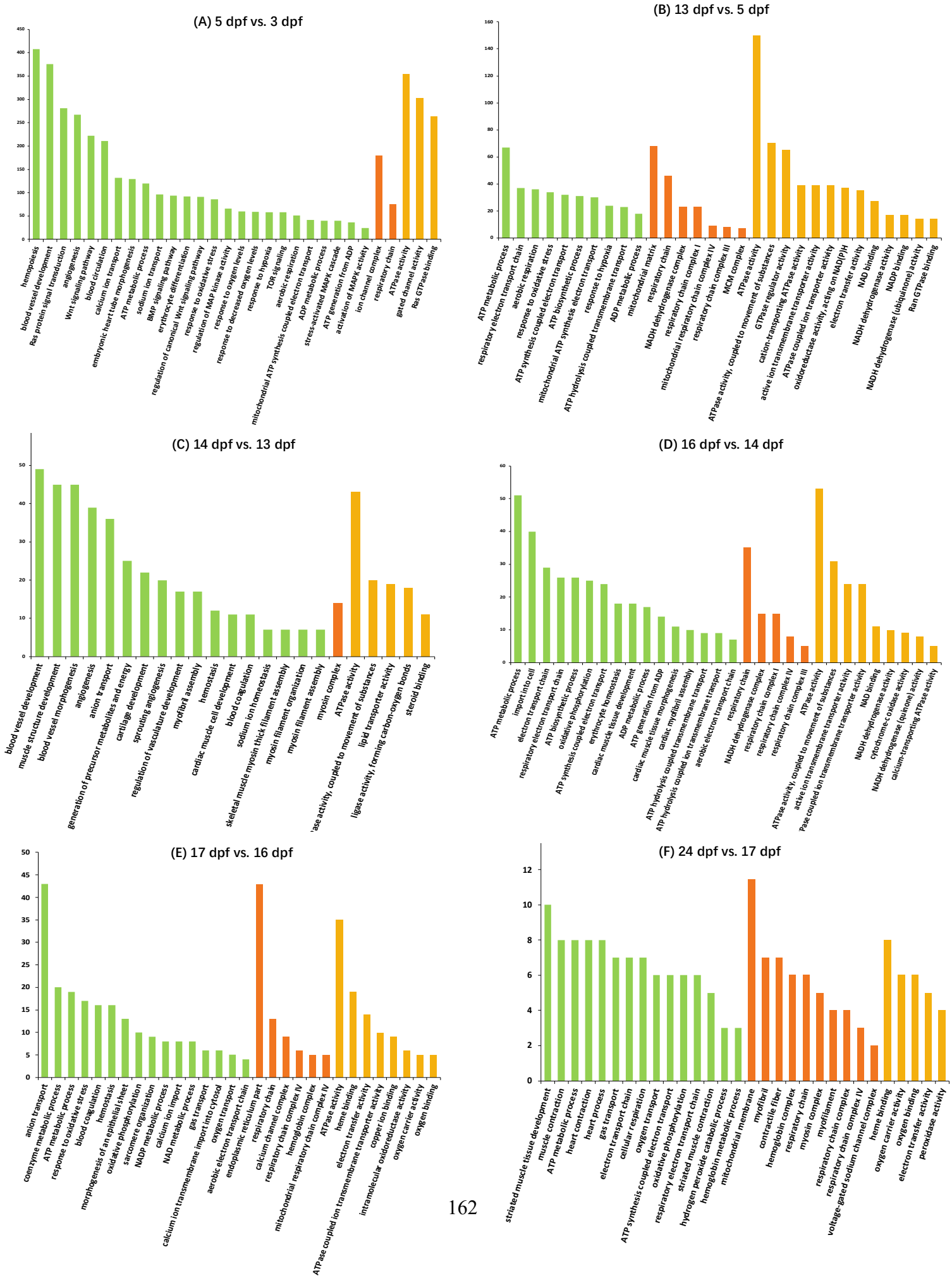

(A) 5 dpf vs. 3 dpf
hemopoiesis
blood vessel development
Ras protein signal transduction
angiogenesis
Wnt signaling pathway
blood circulation
calcium ion transport
embryonic heart tube morphogenesis
ATP metabolic process
sodium ion transport
BMP signaling pathway
erythrocyte differentiation
regulation of canonical Wnt signaling pathway
response to oxidative stress
regulation of MAP kinase activity
response to oxygen levels
response to decreased oxygen levels
response to hypoxia
TOR signaling
aerobic respiration
mitochondrial ATP synthesis coupled electron transport
ADP metabolic process
stress-activated MAPK cascade
ATP generation from ADP
activation of MAPK activity
ion channel complex
respiratory chain
ATPase activity
gated channel activity
Ras GTPase binding
(B) 13 dpf vs. 5 dpf
ATP metabolic process
respiratory electron transport chain
aerobic respiration
response to oxidative stress
ATP synthesis coupled electron transport
ATP biosynthetic process
mitochondrial ATP synthesis electron transport
response to hypoxia
ATP hydrolysis coupled transmembrane transport
ADP metabolic process
mitochondrial matrix
respiratory chain
NADH dehydrogenase complex
respiratory chain complex I
mitochondrial respiratory chain complex I
respiratory chain complex IV
respiratory chain complex III
MCM complex
ATPase activity
ATPase activity, coupled to movement of substances
GTPase regulator activity
cation-transporting ATPase activity
active ion transmembrane transporter activity
ATPase coupled ion transporter activity
oxidoreductase activity, acting on NAD(P)H
electron transfer activity
NAD binding
NADH dehydrogenase activity
NADP binding
NADH dehydrogenase (ubiquinone) activity
Ran GTPase binding
(C) 14 dpf vs. 13 dpf
blood vessel development
muscle structure development
blood vessel morphogenesis
angiogenesis
anion transport
generation of precursor metabolites and energy
cartilage development
sprouting angiogenesis
regulation of vasculature development
myofibril assembly
hemostasis
cardiac muscle cell development
blood coagulation
sodium ion homeostasis
skeletal muscle myosin thick filament assembly
myosin filament organization
myosin filament assembly
myosin complex
ATPase activity
ase activity, coupled to movement of substances
lipid transporter activity
ligase activity, forming carbon-oxygen bonds
steroid binding
(D) 16 dpf vs. 14 dpf
ATP metabolic process
import into cell
electron transport chain
respiratory electron transport chain
ATP biosynthetic process
oxidative phosphorylation
ATP synthesis coupled electron transport
erythrocyte homeostasis
cardiac muscle tissue development
ADP metabolic process
ATP generation from ADP
cardiac muscle tissue morphogenesis
cardiac myofibril assembly
ATP hydrolysis coupled transmembrane transport
ATP hydrolysis coupled ion transmembrane transport
aerobic electron transport chain
respiratory chain
NADH dehydrogenase complex
respiratory chain complex I
respiratory chain complex IV
respiratory chain complex III
ATPase activity
ATPase activity, coupled to movement of substances
active ion transmembrane transporter activity
ase coupled ion transmembrane transporter activity
NAD binding
NADH dehydrogenase activity
cytochrome-c oxidase activity
NADH dehydrogenase (quinone) activity
calcium-transporting ATPase activity
(E) 17 dpf vs. 16 dpf
anion transport
coenzyme metabolic process
ATP metabolic process
response to oxidative stress
blood coagulation
hemostasis
morphogenesis of an epithelial sheet
oxidative phosphorylation
sarcomere organization
NADP metabolic process
calcium ion import
NAD metabolic process
gas transport
calcium ion transmembrane import into cytosol
oxygen transport
aerobic electron transport chain
endoplasmic reticulum part
respiratory chain
calcium channel complex
respiratory chain complex IV
hemoglobin complex
mitochondrial respiratory chain complex IV
ATPase activity
heme binding
electron transfer activity
ATPase coupled ion transmembrane transporter activity
copper ion binding
intramolecular oxidoreductase activity
oxygen carrier activity
oxygen binding
(F) 24 dpf vs. 17 dpf
striated muscle tissue development
muscle contraction
ATP metabolic process
heart contraction
heart process
gas transport
electron transport chain
cellular respiration
oxygen transport
ATP synthesis coupled electron transport
oxidative phosphorylation
respiratory electron transport chain
striated muscle contraction
hydrogen peroxide catabolic process
hemoglobin metabolic process
mitochondrial membrane
myofibril
contractile fiber
hemoglobin complex
respiratory chain
myosin complex
myofilament
respiratory chain complex
respiratory chain complex IV
voltage-gated sodium channel complex
heme binding
oxygen carrier activity
oxygen binding
electron transfer activity
peroxidase activity

Figure 22. Gene ontology enrichment of differentially expressed genes (DEGs) at different development stages in *Clarias macrocephalus*. (A): 5 dpf vs. 3 dpf; (B) 13 dpf vs. 5 dpf; (C) 14 dpf vs. 13 dpf; (D) 16 dpf vs. 14 dpf; (E) 17 dpf vs. 16 dpf; (F) 24 dpf vs. 17 dpf. The vertical axis indicates the number of DEGs between two adjacent sampling times, and the horizontal axis represents the GO terms significantly enriched in the DEGs.

4.4.6 Comparison of gene contents of *Clarias* and channel catfish

C. macrocephalus and *I. punctatus* both belong to the Siluriformes order and thus are genetically close with each other, but air breathing is a major difference between the two species. Initially, we compared the gene contents between *C. macrocephalus* genome and *I. punctatus* genome, and 1,458 genes were identified to be present in *C. macrocephalus,* but absent from *I. punctatus* (Supplementary Table 15). *C. macrocephalus* and *I. punctatus* phenotypically vary in many ways, not only in the ability to breathe in the air. One major difference is the lack of cold tolerance in *Clarias*. Many genes would be expected to contribute to air breathing organ structure and function in *C. macrocephalus*. Genes related to ABO development and air breathing ability would be expected to be differentially expressed during *C. macrocephalus* development. In total, 1,458 *C. microcephalus* - specific genes aligned with the 8,675 DEGs, and 291 genes were verified to be present in *C. macrocephalus* while absent from *I. punctatus.* The same genes were also differentially expressed throught the development of *C. macrocephalus*

(Figure 23). These 291 genes may play important roles in ABO morphogenesis and air breathing ability.

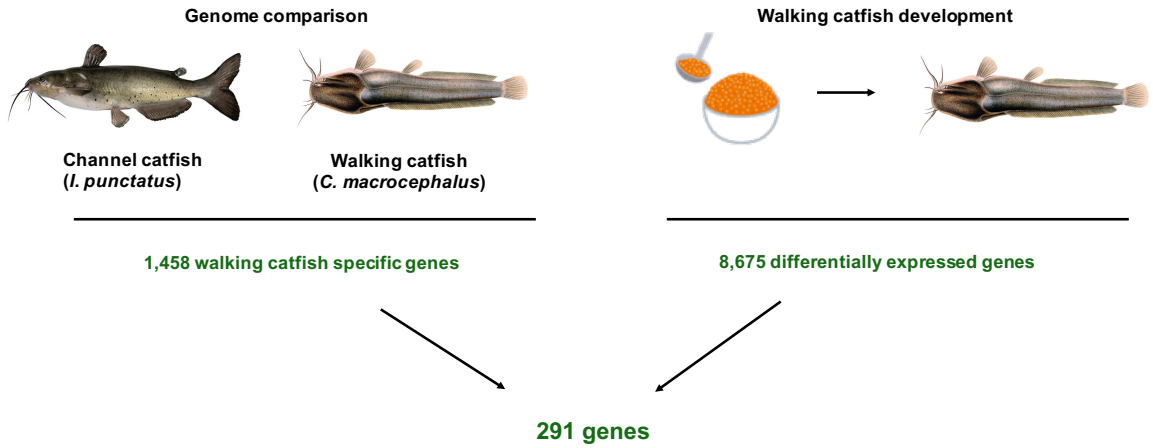


Figure 23. Two hundred and ninety one candidate key genes were identified to be present in *Clarias macrocephalus* but absent in *Ictalurus punctatus*; they were also differentially expressed during development.

4.4.7 Dynamic expression profiles of key genes in *Clarias macrocephalus*

To obtain the expression patterns of the 291 key genes over different development stages, Mfuzz was used to classify genes based on changes in their expression. The 291 genes were gathered into 4 clusters. Clear variation was observed (Figure 24). Multiple different genes were detected in distinct clusters we named Cluster 1, 2, 3 and 4. The most abundant group was Cluster 1, with 169 genes showing a peak of expression at 5 dpf (Supplementary Table 16); after that, the expression values decreased. Cluster 2, including 35 genes, showed an unaltered expression profile at the first four developmental time points, and peaked at 17 dpf. The 29 genes assigned to Cluster 3 showed an expression profile that peaked at 14 dpf. The 29 genes of Cluster

4 had flat expression over the seven developmental stages. We do not expect, Cluster 4 genes to be involved in the formation of air-breathing function in *C. macrocephalus.* In addition, the two oxygen challenge experiments, indicate that the air-breathing organ in *C. macrocephalus* can partially function even as early as the stage of complete yolk sac absorption (3 dpf); and that full air-breathing ability is present at 24 dpf. In this regard, we considered that the genes in clusters 1-3 have the greatest potential to play a critical role in ABO development and function in *C. macrocephalus.*

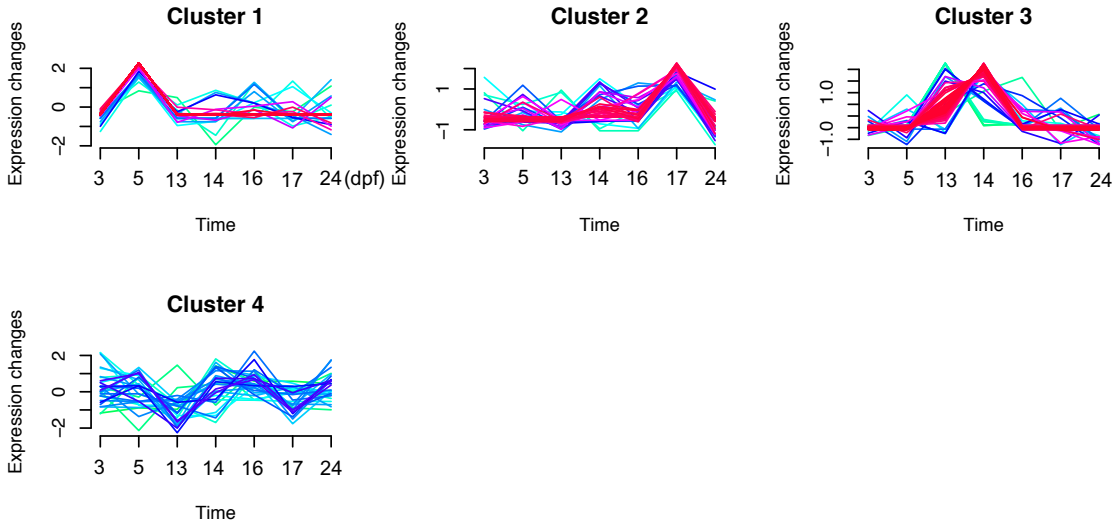


Figure 24. Expression profiles and clusters for the 291 key genes in *Clarias macrocephalus*. These genes gathered into 6 significantly different profiles.

4.4.8 Genes related to the function of air breathing organ in *Clarias macrocephalus*

To further understand these 262 gens in *C. macrocephalus,* and identify their interaction informations, we searched through Retrieval of Interacting Genes (STRING). A protein-protein

(PPI) network was constructed (Figure 25). A subset of the genes was closely associated with hypoxia-related pathways (Table 8). As shown in Table 2, oxygen transport was associated with the formation of air-breathing organ in *C. macrocephalus* and Uniprot analysis detected three genes from this category: hemoglobin alpha embryonic-3 (*hbae*), myoglobin (*mb*), and neuroglobin isoform X1 (*ngb*). Also, from GO enrichment analysis, *mb, ngb* genes were considered to be related to oxygen carrier activity, oxygen binding and heme binding categories, which likely contribute to the air-breathing function. In addition, Reactome pathway analysis revealed that *mb* and *ngb* were closely related with the intracellular oxygen transport function. Taken together, *mb*, *ngb* and *hbae* were highly associated with the formation of air-breathing in *C. macrocephalus*.

Subsequent PPI analysis also indicated that *mb, ngb* and *hbae* had a high degree of connectivity (Figure 25). The genes which are disconnected from other genes were excluded in Figure 8. Finally twenty-three more genes were counted as having a high degree of connectivity with each node, and thus, considered to be related to the occurrence of air-breathing organ function and growth in the development of *C. macrocephalus.* These enzymes included translation initiation factor IF-3 (*mtif3*), RRP5 homolog isoform X1 (*pdcd11*), 2-hydroxyacyl-CoA lyase 1 (*hacl1*), serine--tRNA ligase (*sars2*), ubiquitin-conjugating enzyme E2 variant 3 isoform X3 (*uevld*), leucine rich repeat and fibronectin type III domain containing 4b (*lrfn4b*), polycomb group protein (*pc*), protein disulfide-isomerase A4 precursor (*pdia4*), roundabout homolog 1 isoform X4 (*robo1*), very low-density lipoprotein receptor (*vldlr*), extracellular matrix protein FRAS1 isoform X3 (*fras1*), microtubule-actin cross-linking factor 1 isoform X19

(*macf1*), mannose-1-phosphate guanyltransferase beta (*gmppb*), Larval cuticle protein 1 (*lcp1*), plectin isoform X5 (*pleca*), Epidermal growth factor receptor (*egfr*), sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 isoform X2 (*svep1*), neurogenic locus notch homolog protein 1 isoform X1 (*notch1b*), nesprin-1 isoform X1 (*syne1b*), homeodomain-interacting protein kinase 2 isoform X3 (*hipk2*), kinesin-like protein KIF11 isoform X2 (*kif11*), elongation factor-1, delta, a isoform X2 (*eef1da*) and myomegalin isoform X4 (*pde4dip*). Further knockout, over expression and in situ hybridization experiments will be needed to clarify the associations between these genes and the formation of air-breathing organ function in *C. macrocephalus.*

Table 8. Enrichment and pathway analysis for 262 key genes in *Clarias macrocephalus.* Genes are closely related to the air breathing functions (oxygen-binding, oxygen transport) are identified.

Uniprot key words

#term ID	term description	observed gene count	background gene count	false discovery rate (FDR)	matching genes in network (labels)
KW-0009	Actin-binding	6	143	0.0036	*lcp, macf1, pleca*
KW-0561	Oxygen transport	3	19	0.0047	*hbae, mb, ngb*

GO enrichment					
GO:0005344	oxygen carrier activity	2	6	0.0415	*mb, ngb*
GO:0008144	drug binding	5	214	0.0415	*mb, ngb*
GO:0016772	transferase activity, transferring phosphorus-containing groups	4	129	0.0415	*gmppb*
GO:0019825	oxygen binding	2	6	0.0415	*mb, ngb*
GO:0036094	small molecule binding	7	322	0.0415	*gmppb, mb, ngb*
GO:0043167	ion binding	10	832	0.0415	*gmppb, lcp1, mb, ngb*
GO:0046872	metal ion binding	9	573	0.0415	*lcp1, mb, ngb*
GO:002003	heme binding	2	23	0.0425	*mb, ngb*
GO:0000166	nucleotide binding	5	286	0.0475	*gmppb*
Reactome pathways					
DRE-8981607	Intracellular oxygen transport	2	3	0.0361	*mb, ngb*

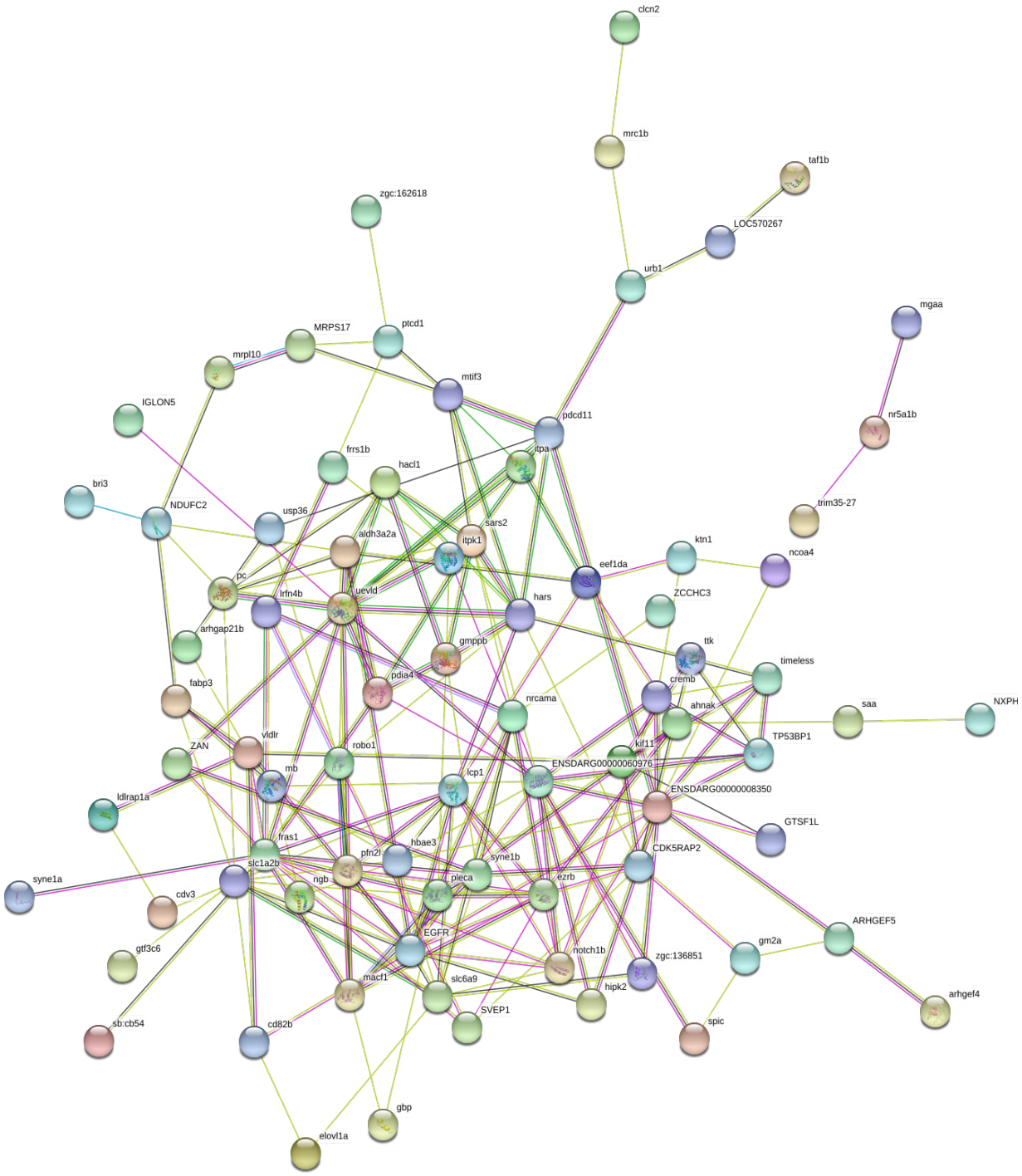


Figure 25. Protein-protein interaction (PPI) network for the key genes in *Clarias macrocephalus.* Genes with no interaction were excluded in this diagram. Only genes with high degree of connectivity were selected. Each node represents one gene, and the interactions among these genes were represented with different lines.

4.5 Discussion

For water breathing fish, gills served as the primary gas exchangers for millions of years, and then decreasing dissolved oxygen in water forced certain fish ancestors to develop air-breathing organs [36]. *C. macrocephalus* is a descendent of those air-breathing fishes, which can use an accessary air-breathing organ to breathe in air, allowing them to survive extreme environments, such as hypoxia and high ammonia [37]. Therefore, *C. macrocephalus* could be a perfect model to explore the formative mechanisms of air-breathing function. Experiments were conducted on anoxia. We tested larval fish with varying exposure to atmospheric air. These latter tests were used to detect the development of air breathing capability as well as the formation of ABO in *C. macrocephalus*. To the best of our knowledge, our work is the first to perform RNA sequencing development in *C. macrocephalus*, including 3 dpf, 5 dpf, 13 dpf, 14 dpf, 16 dpf, 17 dpf, 24 dpf. In total, 8, 675 DEGs were detected during the early development of *C. microcephalus.* The most significantly enriched GO categories revealed that these DEGs were mainly enriched in response to oxidative stress, decreased oxygen levels, hypoxia, angiogenesis, hemopoiesis, ATP metabolic process, gas transport, epithelial sheet morphogenesis and the hemoglobin complex. The genomic contents of *C. macrocephalus* and *I. punctatus* were compared, and 1,458 unique genes were identified in *C. microcephalus* that are absent from *I. punctatus*. Two-hundred-nighty-one genes were detected to be both *Clarias*-specific genes and DEGs throught the development of *C. macrocephalus*, indicating that these genes might be closely related to the formation of air-breathing function in *C.*

macrocephalus. Gene expression analysis also conducted for these 291 genes, and STRING software was carried out to construct PPT network for further analysis for these genes. Twenty-six genes were identified: *mb, ngb, hbae3, mtif3, pdcd11, hacl1, sars2, uevld, lrfn4b, pc, pdia4, robo1, vldlr, fras1, macf1, gmppb, lcp1, pleca, egfr, svep1, notch1b, syne1b, hipk2, kif11, eef1da* and *cdk5rap2* that have no orthologs between *C. macrocephalus* and *I. punctatus,* and the percentage of identical matches (pident) and query coverage per subject (qcovs) were less than 70% over these two species, indicting these genes to be *Clarias*-specific. Of these 26 candidate genes, hbae3, pc, pleca, ezrb and lrfn4b were found strictly in *Clarias*. The remaining 21 genes were detected to have different isoforms in *I. punctatus*, however, their similarity and coverage between these two species was less than 70%. We expect that all these 26 genes are potentially important for the function of ABO in *Clarias*.

Ning et. al (2018) also conducted comparative genomics for *Ictalurus punctatus* and *Clarias batrachus,* but for adult fish. They identified the expansion of *mb* (myoglobin), *ora1* (olfactory receptor related to the class A G protein-coupled receptor 1), and *sult6b1* (sulfotransferase 6b1) genes in the air-breathing *C. batrachus* genome, with 15, 15, and 12 copies, respectively [19]. In addition, a comparative transcriptomic analysis of the gills and air breathing organ revealed that there were eight “elastic fiber formation” genes, eight “hemoglobin” genes, and eighteen “angiogenesis” genes related to air breathing in *C. batrachus* [19]. Ning pointed out that ABOs are highly committed to oxygen transport and cellular respiration [19]. Their finding is in general accordance with ours, as we identified 26 genes potentially contributing to air breathing ability in

C. macrocephalus, of which, *mb, ngb and hbae3* were strongly associated with the formation of ABO function. All belong to the globin family, a family related to the classic respiratory pathway. All play an important role in oxygen binding and transport [38, 39].

MB has a critical function in the supply of oxygen to muscle tissue, and sustains the oxidative metabolism of heart and muscle through facilitating oxygen diffusion to the respiratory chain and storing of oxygen [40, 41]. In mammals, *mb* was observed to be expressed in muscle and cardiac muscles in early embryonic stages and the expression value dramatically increases in the later developmental stages [42]. In Japanese medaka (*Oryzias latipes*), a fish which is relatively well adapted to tolerate hypoxia, exposure of adult medaka to low oxygen result in strong upregulation of myoglobin [43]. In zebrafish (*Danio rerio*), *mb* is associated with circulatory oxygen transport and supply to muscles. Maternal *mb* mRNA is present in the early stages of development, as early as 22 hours post-fertilization (hpf). In fact it displays approximately a 50-fold increase in expression between 18 and 31 hpf. These expression values further increase by 10- to 20-fold in later stages such as like 4-5 dpf [38, 40]. The higher *mb* expression level during those stages is consistent with the onset of the blood circulation [38, 40]. In addition, our gene expression analysis also revealed that the expression profile of *mb* genes belonged to Cluster 1, and the genes in this cluster had an increasing expression value through 3 dpf and reached peak expression at 5 dpf, which agrees with the report in zebrafish [38, 40].

Neuroglobin (*ngb*) is a recently detected vertebrate globin preferentially expressed in the brain and nervous system. It reversibly binds oxygen [44]. In mammals and zebrafish, *ngb* is reported

to be related to oxygen-dependent oxidative metabolism [45, 46]. In zebrafish, *ngb* expression levels are low during early stages. When simple oxygen diffusion is sufficient, in the early stages of embryology, *ngb* expression is low. Its expression increase parallel along with the expression of *mb* at 4-5 dpf when circulation starts [40, 47]. Our expression profile analysis found that *ngb* gene also belonged to Cluster 1, which again peaked at 5 dpf, consistent with *mb* expression.

Hemoglobin alpha embryonic-3 (*hbae3*) and *mb* are best known for respiratory functions, which enable the cellular oxygen supply in support of aerobic metabolism [39]. Hemoglobin consists of two α and two β chains, located in the erythrocytes. It dramatically increases the oxygen carrying capacity of the blood [46]. *Hb* can enhance the efficiency of oxygen transport from the respiratory surfaces to the interior of the body in lungs, gills and skin [39, 46, 48]. In zebrafish, activation of hemoglobin biosynthesis was reported to be associated with hypoxia-protection [49]. All of these genes played different roles in the hypoxia response and alternative metabolic processes of several fish species coping with O_2 deprivation. Further experiments will consider using over expression, knock out and in situ hybridization method to characterize their specific functions and to provide a comprehensive understanding of their molecular roles in air breathing.

4.6 Conclusion

In conclusion, *C. macrocephalus* is an aquatic species that can utilize its accessory ABO to obtain oxygen directly from air, while *I. punctatus* cannot perform air breathing. Therefore, these two species provide an excellent model to reveal the critical genes that contribute to the

development of ABO function. Our study is the first report concerning the high throughout sequencing for transcriptome studies during the early developmental stages in *C. macrocephalus*. A total of 25, 239 unique genes were generated, and 8,675 DEGs were detected during development in *C. macrocephalus*. Through comparative genomic contents analysis between *C. macrocephalus* and *I. punctatus*, 1, 458 genes were initially and uniquely identified in *C. macrocephalus* but shown to be absent from *I. punctatus*. Gene expression profile analysis and PPI network was explored, and 26 genes were identified to be potential candidate genes for the formation of air-breathing function in *C. macrocephalus.* Hypoxia challenge experiments revealed the timing of expression related to air breathing ability in *C. macrocephalus.* Future over expression, knock-out and in situ hybridization should be performed to further verify the candidate key genes for air breathing.

Reference

1. Brauner, C., et al., *Transition in organ function during the evolution of air-breathing; insights from Arapaima gigas, an obligate air-breathing teleost from the Amazon.* Journal of Experimental Biology, 2004. **207**(9): p. 1433-1438.

2. Johansen, K., *9 Air Breathing in Fishes*, in *Fish physiology*. 1970, Elsevier. p. 361-411.

3. Hsieh, S.-T.T., *A locomotor innovation enables water-land transition in a marine fish.* PloS one, 2010. **5**(6): p. e11197.

4. Graham, J.B., *Air-breathing fishes: evolution, diversity, and adaptation*. 1997: Elsevier.

5. Graham, J.B. and H.J. Lee, *Breathing air in air: in what ways might extant amphibious fish biology relate to prevailing concepts about early tetrapods, the evolution of vertebrate air breathing, and the vertebrate land transition?* Physiological and Biochemical Zoology, 2004. **77**(5): p. 720-731.

6. Olson, K., et al., *Gill microcirculation of the air-breathing climbing perch, Anabas testudineus (Bloch): Relationships with the accessory respiratory organs and systemic circulation.* American journal of anatomy, 1986. **176**(3): p. 305-320.

7. Pantung, N., et al., *Histopathological alterations of hybrid walking catfish (Clarias macrocephalus x Clarias gariepinus) in acute and subacute cadmium exposure.* Environment Asia, 2008. **1**: p. 22-27.

8. Olson, K., et al., *Microcirculation of gills and accessory respiratory organs of the walking catfish Clarias batrachus.* The Anatomical Record, 1995. **242**(3): p. 383-399.

9. McMAHON, B.R. and W.W. BURGGREN, *Respiratory physiology of intestinal air breathing in the teleost fish Misgurnus anguillicaudatus.* Journal of Experimental Biology, 1987. **133**(1): p. 371-393.

10. Luo, W., et al., *Developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function of Dojo loach, Misgurnus anguillicaudatus.* Scientific reports, 2016. **6**: p. 31845.

11. Motta, P.J., *Mechanics and functions of jaw protrusion in teleost fishes: a review.* Copeia, 1984: p. 1-18.

12. Graham, J.B., D.L. Kramer, and E. Pineda, *Comparative respiration of an air-breathing and a non-air-breathing characoid fish and the evolution of aerial respiration in characins.* Physiological Zoology, 1978. **51**(3): p. 279-288.

13. Liem, K.F., B. Eclancher, and W.L. Fink, *Aerial respiration in the banded knife fish Gymnotus carapo (Teleostei: Gymnotoidei).* Physiological Zoology, 1984. **57**(1): p. 185-195.

14. Phuong, L.M., et al., *Ontogeny and morphometrics of the gills and swim bladder of air-breathing striped catfish Pangasianodon hypophthalmus.* Journal of Experimental Biology, 2018. **221**(3): p. jeb168658.

15. Teugels, G.G. and D. Adriaens, *Taxonomy and phylogeny of Clariidae: an overview.* Catfishes, 2003. **1**: p. 465-487.

16. Chatchaiphan, S., et al., *De novo transcriptome characterization and growth-related gene expression profiling of diploid and triploid bighead catfish (Clarias macrocephalus Günther, 1864).* Marine biotechnology, 2017. **19**(1): p. 36-48.

17. Islam, M.N., M.S. Islam, and M.S. Alam, *Genetic structure of different populations of walking catfish (Clarias batrachus L.) in Bangladesh.* Biochemical genetics, 2007. **45**(9-10): p. 647-662.

18. Saha, N. and B. Ratha, *Functional ureogenesis and adaptation to ammonia metabolism in Indian freshwater air-breathing catfishes.* Fish physiology and biochemistry, 2007. **33**(4): p. 283-295.

19. Li, N., et al., *Genome sequence of walking catfish (Clarias batrachus) provides insights into terrestrial adaptation.* BMC genomics, 2018. **19**(1): p. 952.

20. Bruton, M.N., *The survival of habitat desiccation by air breathing clariid catfishes.* Environmental Biology of Fishes, 1979. **4**(3): p. 273-280.

21. Belão, T., et al., *Cardiorespiratory responses to hypoxia in the African catfish, Clarias gariepinus (Burchell 1822), an air-breathing fish.* Journal of Comparative Physiology B, 2011. **181**(7): p. 905.

22. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the*

Trinity platform for reference generation and analysis. Nature protocols, 2013. **8**(8): p. 1494.

23. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

24. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

25. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.* Bioinformatics, 2006. **22**(13): p. 1658-1659.

26. Mount, D.W., *Using the basic local alignment search tool (BLAST).* Cold Spring Harbor Protocols, 2007. **2007**(7): p. pdb. top17.

27. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC bioinformatics, 2011. **12**(1): p. 323.

28. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.

29. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics: a journal of integrative biology, 2012. **16**(5): p. 284-287.

30. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets—update.* Nucleic acids research, 2012. **41**(D1): p. D991-D995.

31. Liu, Z., et al., *The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts.* Nature communications, 2016. **7**(1): p. 1-13.

32. Emms, D.M. and S. Kelly, *OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.* Genome biology, 2015. **16**(1): p. 157.

33. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.

34. Kumar, L. and M.E. Futschik, *Mfuzz: a software package for soft clustering of microarray data.* Bioinformation, 2007. **2**(1): p. 5.

35. Szklarczyk, D., et al., *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.* Nucleic acids research, 2010. **39**(suppl_1): p. D561-D568.

36. Hsia, C.C., et al., *Evolution of air breathing: oxygen homeostasis and the transitions from water to land and sky.* Comprehensive Physiology, 2013. **3**(2): p. 849-915.

37. Olshanskiy, V., et al. *Ejection of specialized electric discharges during prey pursuit and unspecialized electric activity related to respiratory behavior of the clariid catfish Clarias macrocephalus (Clariidae, Siluriformes)*. in *Doklady Biological Sciences*. 2011. Springer Science & Business Media.

38. Vlecken, D.H., et al., *A critical role for myoglobin in zebrafish development.* International Journal of Developmental Biology, 2009. **53**(4): p. 517-524.

39. Tian, R., et al., *Molecular evolution of globin genes in Gymnotiform electric fishes: relation to hypoxia tolerance.* BMC evolutionary biology, 2017. **17**(1): p. 51.

40. Tiedke, J., et al., *Ontogeny of globin expression in zebrafish (Danio rerio).* Journal of Comparative Physiology B, 2011. **181**(8): p. 1011-1021.

41. Wittenberg, J.B. and B.A. Wittenberg, *Myoglobin function reassessed.* Journal of Experimental Biology, 2003. **206**(12): p. 2011-2020.

42. Weller, P., et al., *Myoglobin expression: early induction and subsequent modulation of myoglobin and myoglobin mRNA during myogenesis.* Molecular and Cellular Biology, 1986. **6**(12): p. 4539-4547.

43. Wawrowski, A., et al., *Changes of globin expression in the Japanese medaka (Oryzias latipes) in response to acute and chronic hypoxia.* Journal of Comparative Physiology B, 2011. **181**(2): p. 199-208.

44. Hankeln, T., et al., *Neuroglobin and cytoglobin in search of their role in the vertebrate globin family.* Journal of inorganic biochemistry, 2005. **99**(1): p. 110-119.

45. Bentmann, A., et al., *Divergent distribution in vascular and avascular mammalian retinae links neuroglobin to cellular respiration.* Journal of Biological Chemistry, 2005. **280**(21): p. 20660-20665.

46. Roesner, A., T. Hankeln, and T. Burmester, *Hypoxia induces a complex response of globin expression in zebrafish (Danio rerio).* Journal of Experimental Biology, 2006. **209**(11): p. 2129-2137.

47. Grillitsch, S., et al., *The influence of environmental PO2 on hemoglobin oxygen saturation in developing zebrafish Danio rerio.* Journal of experimental biology, 2005. **208**(2): p. 309-316.

48. Rahbar, S., *Hemoglobin: Structure, function, evolution, and pathology.* American journal of human genetics, 1983. **35**(4): p. 781.

49. Long, Y., et al., *Transcriptional events co-regulated by hypoxia and cold stresses in Zebrafish larvae.* BMC genomics, 2015. **16**(1): p. 385.