Marine Biotechnology (2021) 23:90–105
https://doi.org/10.1007/s10126-020-10005-4

ORIGINAL ARTICLE

Comparative Genomic and Transcriptomic Analyses Revealed Twenty-Six Candidate Genes Involved in the Air-Breathing Development and Function of the Bighead Catfish *Clarias macrocephalus*

Xiaoli Ma[1,2] · Baofeng Su[1,2] · Max Bangs[1,2,3] · Veronica Alston[1,2] · Nathan J. C. Backenstose[1,2,4] · Rhoda Mae Simora[1,2,5] · Wenwen Wang[1,2] · De Xing[1,2] · Shangjia Li[1,2] · Zhi Ye[1,2,6] · Anthony G. Moss[2,7] · Thuy-Yen Duong[8] · Xu Wang[2,9,10] · Rex A. Dunham[1,2]

Received: 23 June 2020 / Accepted: 18 October 2020 / Published online: 28 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

The bighead catfish (*Clarias macrocephalus*) and channel catfish (*Ictalurus punctatus*) are freshwater species in the Siluriformes order. *C. macrocephalus* has both gills and modified gill structures serving as an air-breathing organ (ABO), while *I. punctatus* does not possess such an organ, and cannot breathe in air, providing an excellent model for studying the molecular basis of ABO development in teleost fish. To investigate the critical time window for the development of air-breathing function, seven development stages were selected based on hypoxia challenge results, and RNA-seq was performed upon *C. macrocephalus* to compare with the non-air-breathing *I. punctatus*. Five-hundred million reads were generated and 25,239 expressed genes were annotated in *C. macrocephalus*. Among those, 8675 genes were differentially expressed across developmental stages. Comparative genomic analysis identified 1458 *C. macrocephalus* specific genes, which were absent in *I. punctatus*. Gene network and protein-protein interaction analyses identified 26 key hub genes involved in the air-breathing function. Three top candidate genes, *mb*, *ngb*, *hbae*, are mainly associated with oxygen carrying, oxygen binding, and heme binding activities. Our study provides a rich data set for exploring the genomic basis of air-breathing function in *C. macrocephalus* and offers insights into the adaption to hypoxic environments.

Keywords Transcriptome analysis · *Clarias macrocephalus* · Air-breathing catfish · Oxygen binding · Oxygen transport

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10126-020-10005-4.

✉ Xu Wang
xzw0070@auburn.edu

✉ Rex A. Dunham
dunhara@auburn.edu

[1] School of Fisheries, Aquaculture and Aquatic Sciences, Auburn University, Auburn, AL 36849, USA

[2] Alabama Agricultural Experiment Station, Auburn, AL 36849, USA

[3] Department of Biological Science, Florida State University, Tallahassee, FL 32304, USA

[4] Department of Biological Sciences, University at Buffalo, Buffalo, NY 14260, USA

[5] College of Fisheries and Ocean Sciences, University of the Philippines Visayas, Miagao, 5023 Iloilo, Philippines

[6] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

[7] Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

[8] College of Aquaculture and Fisheries, Can Tho University, Can Tho City 94000, Vietnam

[9] Department of Pathobiology, Auburn University, Auburn, AL 36849, USA

[10] HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA

Introduction

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 from the ocean to terrestrial living (Brauner et al. 2004). Approximately 350 million years ago, there was a significant change in the environment, higher temperature and the decay of dead organic components consumed oxygen in swamps, rivers, and lakes, resulting in a gradually decline in dissolved oxygen concentration (Johansen 1970). Certain fish ancestors started to develop air-breathing organs (ABO) as a way of adaptation. Some of them left water environments and eventually colonized the land (Hsieh 2010). Air-breathing fish can rise to the surface of water to gulp air or crawl onto land and survive for extended periods of time (Hsieh 2010).

It is estimated that there are more than 370 air-breathing fish species in 49 families (Graham and Lee 2004). The ABOs vary considerably among different fishes. *Clarias macrocephalus* and *Clarias 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 they act as accessory ABOs. Other species use a modified intestine as an ABO, such as *Misgurnus anguillicaudatus* (Mcmahon and Burggren 1987). 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 (Mcmahon and Burggren 1987; Luo et al. 2016). Other fish use their swim bladder as an ABO. These include the Gonorynchiformes, Characiformes. and some species in Siluriformes, such as the suborder Gymnotus and in *Pangasianodon hypophthalmus* (Motta 1984; Graham et al. 1978). The swim bladder extends from the posterior of the head back into the body cavity and serves as the major place for gas exchange for aerial breathing (Liem et al. 1984).

C. macrocephalus, also known as the bighead catfish, belongs to the Siluriformes order. *C. macrocephalus* is native to Thailand and Vietnam and has been introduced to several Southeast Asian countries (Fishbase 2020). Bighead catfish has both gills and a modified gill structure serving as an ABO. Therefore, it is well adapted to hypoxic conditions in muddy marshes (Teugels and Adriaens 2003). It burrows into the mudflats during summer periods, staying alive through direct air-breathing (Chatchaiphan et al. 2017; Islam et al. 2007; Saha and Ratha 2007). After a heavy rainfall, bighead catfish can slither across the land (Li et al. 2018; Bruton 1979). Its "walking" ability allows *C. macrocephalus* to survive in extreme environments, such as hypoxia, short-term desiccation stress and high ammonia, which are not ideal conditions for aquaculture (Belão et al. 2011). This capability makes bighead catfish a perfect model to study the evolution of adaptions to air-breathing, as well as the mechanisms for hypoxia and ammonia tolerance.

Previous studies in *Misgurnus anguillicaudatus* and *Clarias batrachus* have investigated the underlying mechanisms for intestinal ABO development (Saha and Ratha 2007; Li et al. 2018). 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 advances 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 identify the critical time window in the development of the ABO function in *C. macrocephalus*. Comparative genomic analysis and RNA-seq analysis of seven early development stages elucidated the genes that potentially contribute to air-breathing and terrestrial adaptations.

Material and Methods

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 (AU-IACUC, PRN#: 2016-2901 on June 13, 2016 and the Can Tho University Animal Welfare Committee on November 15, 2016).

Experimental Animals and Tissue Collection

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

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 macrocephalus* larvae older than 3 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 macrocephalus larvae were placed in a 2-L container with oxygen supply as a control treatment. A separate group was stocked in a second 2-L container with supplemental oxygen 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 a DO meter). Dissolved oxygen levels, larval behavior, and survival rate were observed and recorded every 10 min; the results at each time point were calculated and responses to hypoxia were represented graphically.

Moist adult *Clarias macrocephalus* can leave the water and walk on land and live indefinitely. Therefore, another oxygen challenge experiment was conducted in a petri dish test to simulate this situation. *Clarias macrocephalus* larvae were exposed directly to air and challenged each day from 3 to 27 dpf. One group of 15–20 larvae was 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. For the petri dish treatment, fish were also stocked into an aerated 2-L water-filled tank as a control treatment. Larval behavior and survival rate were observed and recorded every 10 min.

The bighead catfish candidate air-breathing related genes during development were identified in a stepwise manner. Through annotation with UniProt database and with comparison to the channel catfish database, genes were detected for *Clarias macrocephalus*. Then differentially expressed genes were identified between each adjacent, critical developmental time point based on the anoxia challenges, narrowing the candidate genes. Comparative genomic analysis was conducted between bighead catfish and channel catfish to identify putative *Clarias*-specific genes. The intersection of the genes from the previous two steps was executed to further narrow the list of candidate key genes. Expression profile analysis was conducted, and the genes that are accordant with the development of ABO were retained. Additionally, the narrowed list of genes was inputted to STRING database to detect the putative key hub genes. Bioinformatics analysis was conducted, resulting in 3 key hub genes based on their classifying to oxygen carrier activity as well as oxygen binding ability in UniProt key words and based upon GO enrichment and Reactome pathways. The results were further confirmed by searching for the putative key genes in the literature review. Protein-protein interaction pathway analysis identified the interrelated expression at key developmental time points with the 3 key oxygen-carrying related genes.

RNA Extraction, Library Construction, and Sequencing

Seven time points (3, 5, 13, 14, 16, 17, 24 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 each timepoint and treatment group. Pooled samples of 4 individual fish were homogenized in liquid nitrogen to a fine powder using mortar and pestle. 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 following sequencing with HiSeq 2000 system.

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 2 × 150 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.

De Novo Assembly and Gene Annotation

The detailed bioinformatics tools and scripts can be found in GitHub at https://github.com/xzm00171122/Scripts-for-Clarias-ABO. De novo assembly was performed on *C. macrocephalus*-cleaned reads using Trinity (version 2.8.5) (Haas et al. 2013). Prior to assembly, quality of raw reads were evaluated in FastQC (version 0.11.5) (Patel and Jain 2012), and raw reads were filtered by removing adaptor sequences and ambiguous nucleotides using Trimmomatic (version 0.36) (Bolger et al. 2014). 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 the full-length transcripts. The assembled transcriptome was passed to CD-HIT (version 4.7) to reduce redundancy with 95% identity (Li and Godzik 2006).

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). 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.

Identification of Orthologous Protein Groups

The protein sequences of *I. punctatus* were downloaded from NCBI based on sequence orthology (Yu et al. 2012); the protein sequences of *C. macrocephalus* were obtained with transdecoder based on our transcriptome assembly data. Orthologs and orthogroups between *I. punctatus* and *C. macrocephalus* were detected with OrthoFinder software (version 2.2.7) (Barrett et al. 2012; Emms and Kelly 2015; Liu et al. 2016; Li et al. 2018). 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 (Emms and Kelly 2015). Subsequently, a reciprocal BLASTP search was carried out to query genes with no hits from previous steps with a maximum *E* value of 1E-5. These leftover genes were 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 macrocephalus* were aligned with the channel catfish genome using TBLASTN with a maximum *E* value of 1E-10 (Altschul et al. 1990). 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.

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 (Haas et al. 2013). The mapped reads number for each transcript was detected using RSEM software (Li and Dewey 2011). 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) (Love et al. 2014). Differentially expressed genes (DEGs) were detected with a criterion of $|\log2\text{-fold change}| \geq 1$ and adjusted p value < 0.05. The distribution of upregulated and downregulated genes was demonstrated in a volcano plot using the function plotMA, which shows the log2-fold changes attributable to a given variable over the mean of normalized counts for all the samples in the DESeqDataSet.

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) (Yu et al. 2012). 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.

Clustering of Gene Expression and Protein-Protein Interaction (PPI) Network Analysis of the Key Genes As Well as Expression Analysis of Selected Genes

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. GO enrichment analysis of the key genes was performed using ClusterProfiler R package (version 3.6) (Yu et al. 2012), and a q value < 0.05 was 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 (Kumar and Futschik 2007). In addition, the retrieval of interacting genes (STRING) database was used to investigate the network properties of the enzymes encoded by the key genes, and a protein-protein (PPI) network was constructed (Szklarczyk et al. 2016). The STRING database also provided both enrichment and experimental information (Szklarczyk et al. 2016).

The expression patterns of selected genes were detected in *C. macrocephalus* and *I. punctatus* by polymerase chain reaction (PCR). Samples were digested with 600 μl of DNA extraction cocktail buffer [100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 25 mM EDTA (pH 8.0), 0.5% sodium dodecyl sulfate] and 3 μl of 20 mg/ml freshly made proteinase K (Sigma-Aldrich, St. Louis, MO) for 4–5 h at 55 °C (Shang et al. 2018). Then protein precipitation solution (Qiagen) and isopropanol were used to precipitated protein and DNA, respectively. DNA was washed twice in cold 75% ethanol, dried with air, then resuspended in RNase/DNase-free water (Shang et al. 2018). The quantification and integrity of the DNA were measured using Nanodrop 2000 Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware).

The 1-kb plus DNA ladder (Invitrogen) was used as PCR marker; PCR products were resolved in a 1% agarose gel. PCR amplification was in a volume of 10 μL containing 2 μL nuclease-free water; 1 ul of 10 μM each gene-specific primer and 5 μL of EconoTaq®PLUS2X Master Mix (Lucigen, Middleton, WI), and 1 μl (250 ng/ul) DNA. The cycling conditions of PCR were 92 °C/3 min, 40 cycles of 92 °C/30 s, 60 °C/45 s, 72 °C/45 s then 72 °C/5 min, 4 °C/5 min. The 18 s rRNA and β-actin genes were selected as an internal reference gene (Bas et al. 2004). All the specific

primers were designed using Primer3 and listed in Table S1 (Untergasser et al. 2012).

Results

Low Oxygen Challenge

To investigate the critical developmental stages associated with the air-breathing ability, low oxygen challenge experiments were performed to test the survival of *C. macrocephalus* in 0 mg/L dissolved oxygen (DO) in 6–7 early developmental stages. At 3 dpf, *C. macrocephalus* larvae had a survival rate of 93.3% when oxygen was decreased to 0.5 mg/L (14 of 15 fish survived) (Fig. 1a). The aeration was removed from the container at the beginning of this experiment. Oxygen concentration was at 5.3 mg/L, and fish were active and swam on the bottom as normal. After 5 min, DO level dropped below 3.9 mg/L, and fish had a little slower movement and came closer together at the bottom. After 25 min, when DO level reached 1.2 mg/L, fish were swimming slower and some tried to move to the surface. This behavior remained the same until DO concentration decreased to 0.6 mg/L at 75 min; during the challenge, one fish died. The rest moved to the water surface and back to the bottom. No more fish died.

At 5 dpf, when oxygen concentration decreased from 5.5 to 2.5 mg/L 20 min after removing the aeration, all fish swam at the bottom (Fig. 1b). Between 25 and 165 min, the dissolved oxygen level dropped from 2.1 to 0.5 mg/L; the fish were slowly swimming, and some fish went up to the surface. At 195 min, the dissolved oxygen level was 0.5 mg/L and all fish swam slowly on the surface. After 405 min, the dissolved oxygen concentration was of 0.3 mg/L, the fish remained moving slowly at the surface, and no fish died in this experiment. We observed similar fish behavior for challenges at 13 dpf, 14 dpf, 16 dpf, and 17 dpf; and the survival rates were 80%, 90%, 100%, and 100%, respectively (Fig. 1c–f). Detailed information is shown in Table S2.

In addition to the tank experiments with decreasing oxygen levels, we performed additional challenge experiments by directly exposing *Clarias macrocephalus*

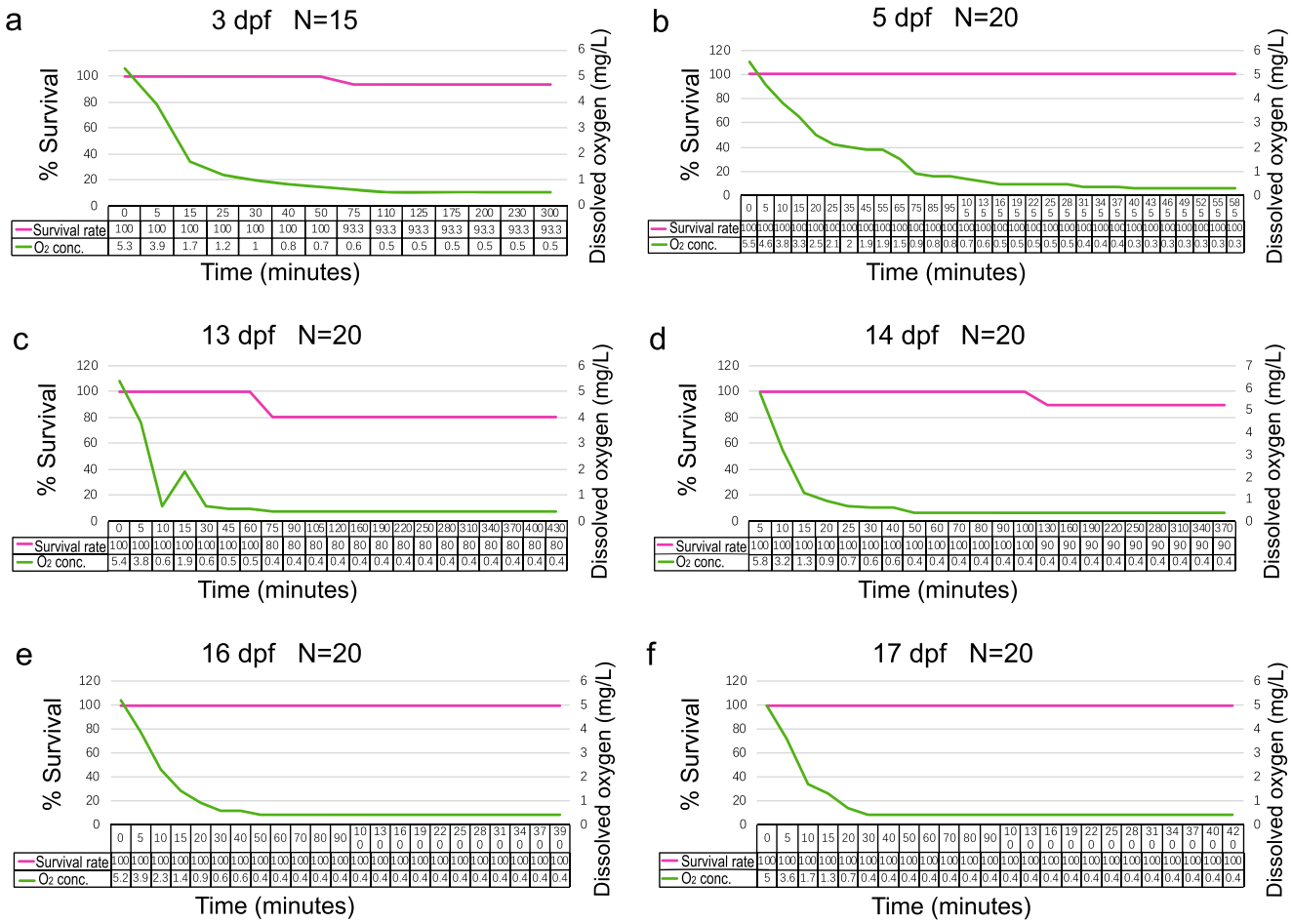

Fig. 1 Oxygen depletion challenge for *Clarias macrocephalus* at 6 early developmental stages. Dissolved oxygen concentrations (in blue, right *y*-axis) and survival curve (in red, left *y*-axis) for *Clarias macrocephalus* at 3, 5, 13, 14, 16, 17 days post-fertilization (dpf) was plotted at the low dissolved oxygen challenge. Dissolved oxygen level was reduced by replacing with the oxygen stripping. Fish were determined moribund when the opercular movement ceased

larvae to the air in a petri dish from 3 to 24 dpf. At 3 dpf, two fish (out of 15) died within 25 min, and two more died after 30 min (Table 1, Fig. 2a). The rest of the fish stopped moving and all were dead by 80 min. At 5 dpf, all *Clarias macrocephalus* larvae survived in petri dish without water until the first two fish were found dead at 62 min. The survival rate within the first hour was much higher compared to 3 dpf (90% vs. 40%). No fish survived after 108 min (Fig. 2b). At 13-dpf larvae, all fish survived during the first 93 min, which is half an hour longer than at 5 dpf. The first dead fish was found at 93 min into challenge and no fish survived after 382 min (Fig. 2c). Similar results were observed at 14 dpf and 16 dpf, and the ability to survive in the air was increasing gradually, but no fish survived the entire 7-h experiment (Fig. 2d–e). At 17 dpf and 24 dpf, a large proportion of fish remained alive after 420 min, with a 45% and 60% survival rate, respectively (Fig. 2f–g and Table S3).

In summary, from 3 dpf, the first day after *C. macrocephalus* completely absorbed all of the yolk sac, the fish already had a high tolerance of extremely low oxygen levels. At this stage, some fish could move to the surface and had air-gulping behavior. From 3 to 16 dpf, the air-breathing ability gradually increased during the development of *C. macrocephalus.* By 16 dpf, all fish larvae could live in extremely low oxygen water environment. However, at this stage, no larvae could survive in a water-free aerial environment for an extend period (more than 2.5–6 h). Forty-five percent of the fish could survive by air-breathing at 17 dpf, and the proportion increased to 60% at 24 dpf in the environment. *C. macrocephalus* were close to having full air-breathing ability at 24 dpf as 60% of the population survived.

Transcriptome Sequencing and de Novo Assembly

A total of 1331 million raw reads was generated through RNA sequencing. Low-quality bases (quality score < 25) and adapter sequences were trimmed; reads less than 36 bases after trimming were excluded from subsequent analysis (Table S4). Five-hundred-four million high-quality reads from the seven different transcriptome timepoints were assembled using the Trinity software. After removing the redundant transcripts using CD-HIT tool of 95% identity, approximately 1,841,162 transcripts remained with an average contig length of 829 bp and N50 = 1390 bp (Table 2). The average ORF length was 592 bp, at a GC content of 42.3%. The correlation of the two biological replicates IS illustrated in Fig. S1.

Table 1 Time of first death, last death, and survival of larval bighead catfish (*Clarias macrocephalus*) at various times day post-fertilization (dpf) in an aerial environment

dpf	First death (min)	Last death (min)	Total survival (%)
3	1	80	0
5	42	110	0
13	110	390	0
14	110	290	0
16	155	340	0
17	280	410	42
24	280	370	60

Gene Content Differences Between *Clarias macrocephalus* and Channel Catfish

C. macrocephalus and *I. punctatus* both belong to the same Siluriformes order, but air-breathing ability is a major difference between the two species. To identify genes involved in the air-breathing function in *Clarias macrocephalus*, we compared the gene contents between *C. macrocephalus* and *I. punctatus* genomes, and found 1458 genes that are present in *C. macrocephalus*, but absent from *I. punctatus* (Table S5). Some of these *Clarias*-specific genes are expected to contribute to air-breathing organ structure and function in *C. macrocephalus*. Air-breathing is not the only phenotypical and physiological difference between *C. macrocephalus* and *I. punctatus*. The other major difference between the two species is the lack of cold tolerance in *Clarias macrocephalus*. Genes related to ABO development and air-breathing should be differentially expressed during *C. macrocephalus* development.

Transcriptome Annotation and Identification of Differentially Expressed Genes (DEGs)

The assembled transcripts were BLASTed against three databases: UniProt protein database (SwissProt), the channel catfish transcript database, and zebrafish transcript database, using an *E* value cutoff of 1e-5. The best hit contigs were retained, resulting a total of 25,239 annotated unigene transcript for expression analysis (Table 2).

Differential gene expression analysis was performed to identify the DEGs by comparing of each stage with its previous stage. In *Clarias macrocephalus*, comparisons between the first three developmental timepoints (5 dpf vs. 3 dpf and 13 dpf vs. 5 dpf) discovered more than 5000 DEGs (Fig. 3 and Table S6), which is consistent with the dramatic gene expression changes during early development. There were 1906, 2703, and 1718 DEGs detected in the next three comparisons (14 dpf vs. 13 dpf, 16 dpf vs. 14 dpf, and 17 dpf vs. 16 dpf). The comparison between the last developmental point (24 dpf) and its previous stage (17 dpf) identified the least number of DEGs (449). The significance and fold-changes

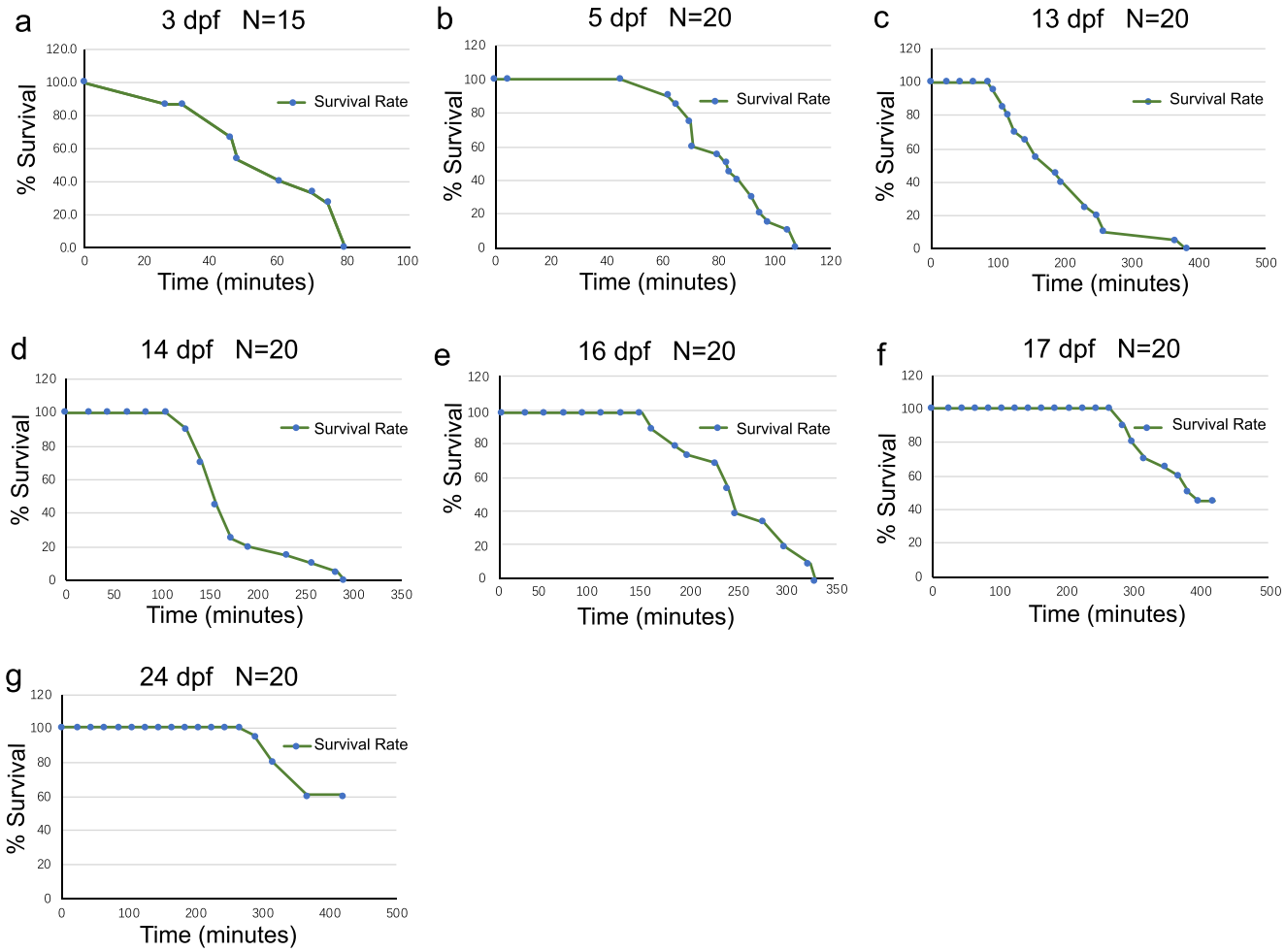

Fig. 2 Survival curve was plotted for *Clarias macrocephalus* exposed to the aerial environment at 3, 5, 13, 14, 16, 17, 24 days post-fertilization (dpf). Fish were determined moribund when the opercular movement ceased

of DEGs were shown in volcano plots (Fig. 4). As we already detected 1458 *C. macrocephalus*-specific genes, in which 291 genes were also identified to be differentially expressed across *C. macrocephalus* development (Fig. S2). These genes served as candidate key genes for air-breathing function development.

Table 2 Statistics and quality assessment transcript contig assembly of bighead catfish (*Clarias macrocephalus*)

Assembly and annotations	*Clarias macrocephalus*
Number of transcripts	1,841,162
Number of nonredundant transcripts	1,004,981
GC%	42.3%
Average contig length (bp)	829
Average ORF length	592
Transcript contig N50	1390
N50 longest isoform	799
Number of unigenes	25,239

Gene Ontology Enrichment Analysis of DEGs 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 macrocephalus* development are listed in Table S6. The most significantly enriched GO terms for each stage are shown in Fig. S3.

Between days 3 and 5, the DEGs were mainly related to hemopoiesis (GO:0030097), response to aerobic respiration (GO:0009060), response to decreased oxygen levels (GO:0036293), response to hypoxia (GO:0001666), blood vessel development (GO:0001568), and angiogenesis (GO:0001525). The enriched GO terms in response to hypoxia, blood vessel development, and angiogenesis is evidenced with the fact that the air-breathing ability started to develop as early as 5 dpf.

Between days 5 and 13, DEGs were mainly enriched for aerobic respiration (GO:0009060), response to oxidative stress

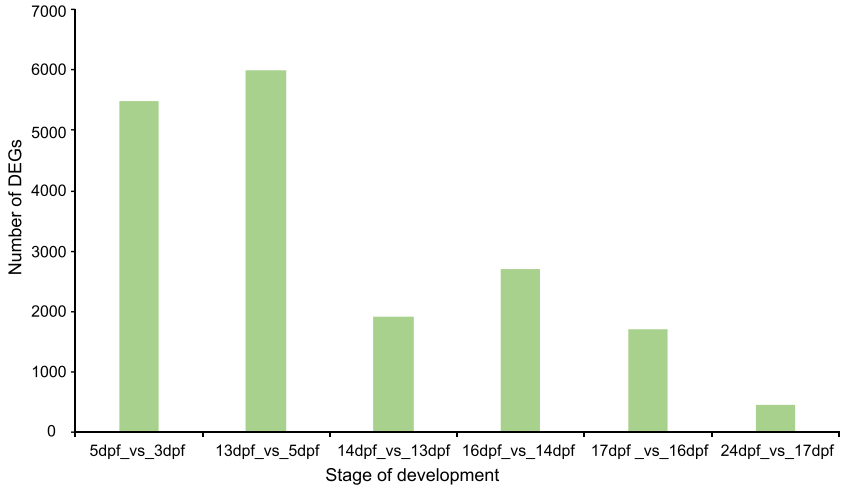

Fig. 3 Gene expression during early embryonic development in *Clarias macrocephalus.* The number of differentially expressed genes (DEGs) was detected for comparison of each stage with the previous stage. Differential expression of genes peaked between 13 and 5 dpf stage. Over time, the number of DEGs generally decreased

(GO:0006979), response to hypoxia (GO:0001666), respiratory electron transport chain (GO:0022904), and respiratory chain complex (GO:0098803). 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), and myosin filament assembly (GO:0031034).

For 16 dpf compared 14 dpf stage, the DEGs were mainly enriched for erythrocyte homeostasis (GO:0034101), ATP metabolic process (GO:0046034), aerobic electron transport chain (GO:0019646), myosin filament organization (GO:0031033),

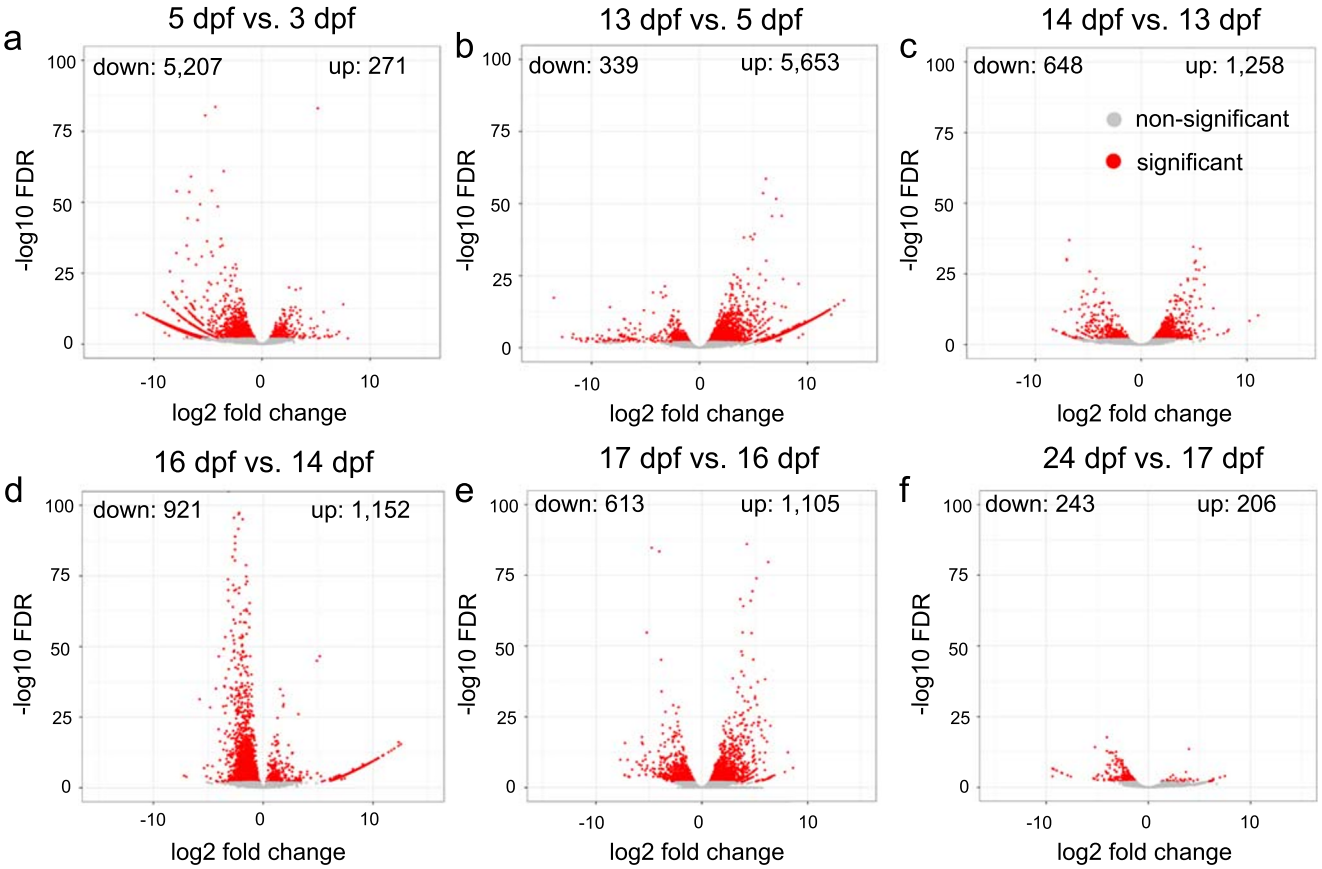

Fig. 4 Volcano plot of the differentially expressed genes (DEGs) during *Clarias macrocephalus* air-breathing organ development. Red dots indicate downregulated (negative value) and upregulated (positive value) genes at each stage. Gray dots represent non-DEGs. **a** 5 days post-fertilization (dpf) vs. 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

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), oxygen binding (GO:0019825), oxygen carrier activity (GO:0005344), hemoglobin complex (GO:0005833), response to oxidative stress (GO:0006979), and morphogenesis of an epithelial sheet (GO:0002011). 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), heart contraction (GO:0060047), hemoglobin metabolic process (GO:0020027), myofibril (GO:0030016), and heme binding (GO:0020037). These most enriched GO categories were consistent with the observed behaviors during air-breathing development.

Dynamic Genes 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 form 4 co-expression clusters (Fig. 5). The most abundant group was cluster 1, with 169 genes showing a peak of expression at 5 dpf, which may correspond to the developmental fate determination of the air-breathing organ. Cluster 2 (35 genes) showed an unaltered expression profile at the first four developmental timepoints, and peaked at 17 dpf, which are correlated to the key transition between 16 and 17 dpf. Our hypoxia challenge experiments suggest the ability of surviving in air environment was obtained during this critical window. The 29 genes assigned to cluster 3 displayed an expression profile that peaked at 14 dpf. The 29 genes of cluster 4 had variable expression across seven developmental stages, which are less likely 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 for more than half the population. 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.*

Twenty-Six Hub Genes Were Identified to Be Critical for Air-Breathing Function in *Clarias macrocephalus*

To further understand these 291 genes in *C. macrocephalus* and identify their interactions, we searched through Retrieval of Interacting Genes (STRING) (Table S7). A protein-protein

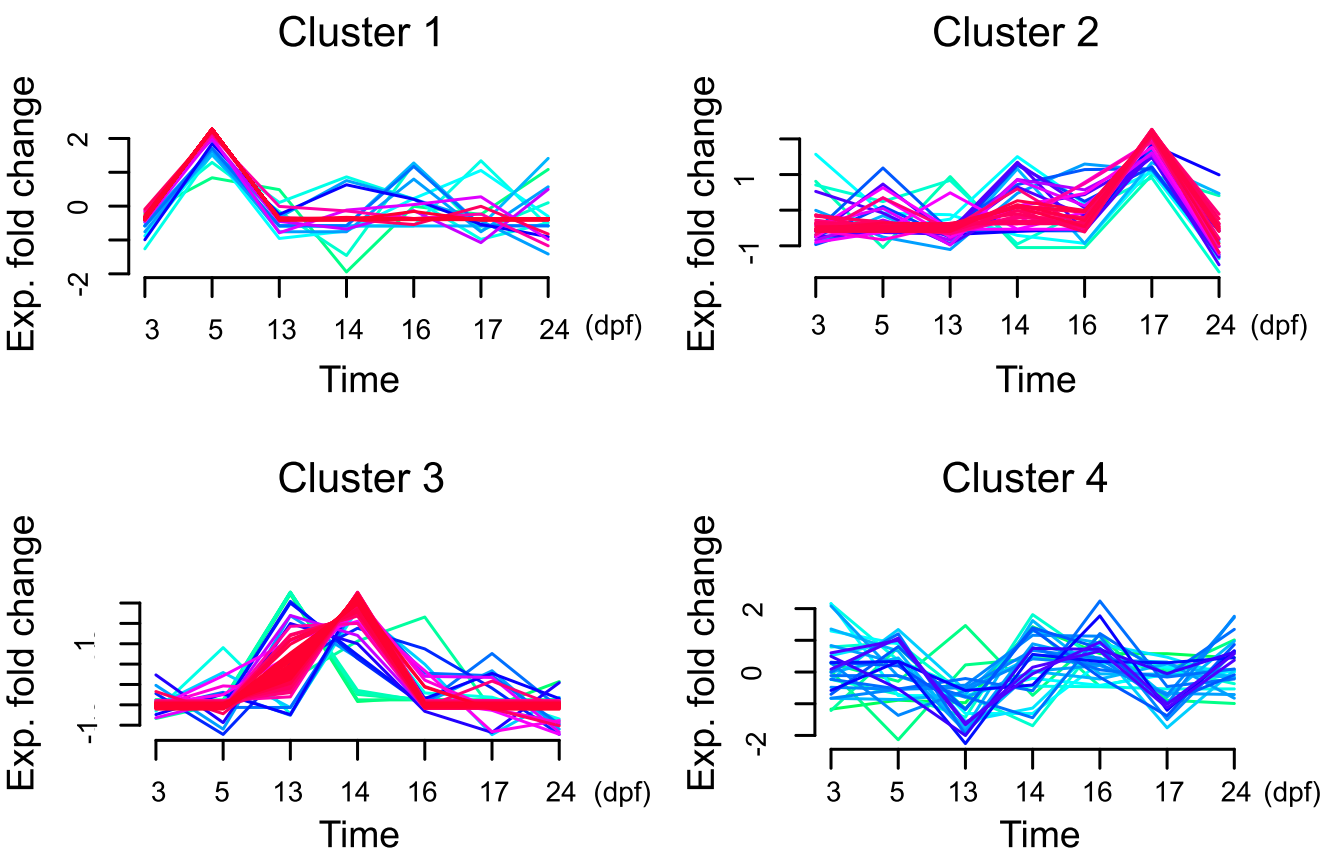

Fig. 5 Expression clusters of 291 key genes across *Clarias macrocephalus* developmental stages

interaction (PPI) network was constructed (Fig. 6). A subset of the genes was closely associated with hypoxia-related pathways (Table 3). Oxygen transport was associated with the formation of air-breathing organ in *C. macrocephalu*s, 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

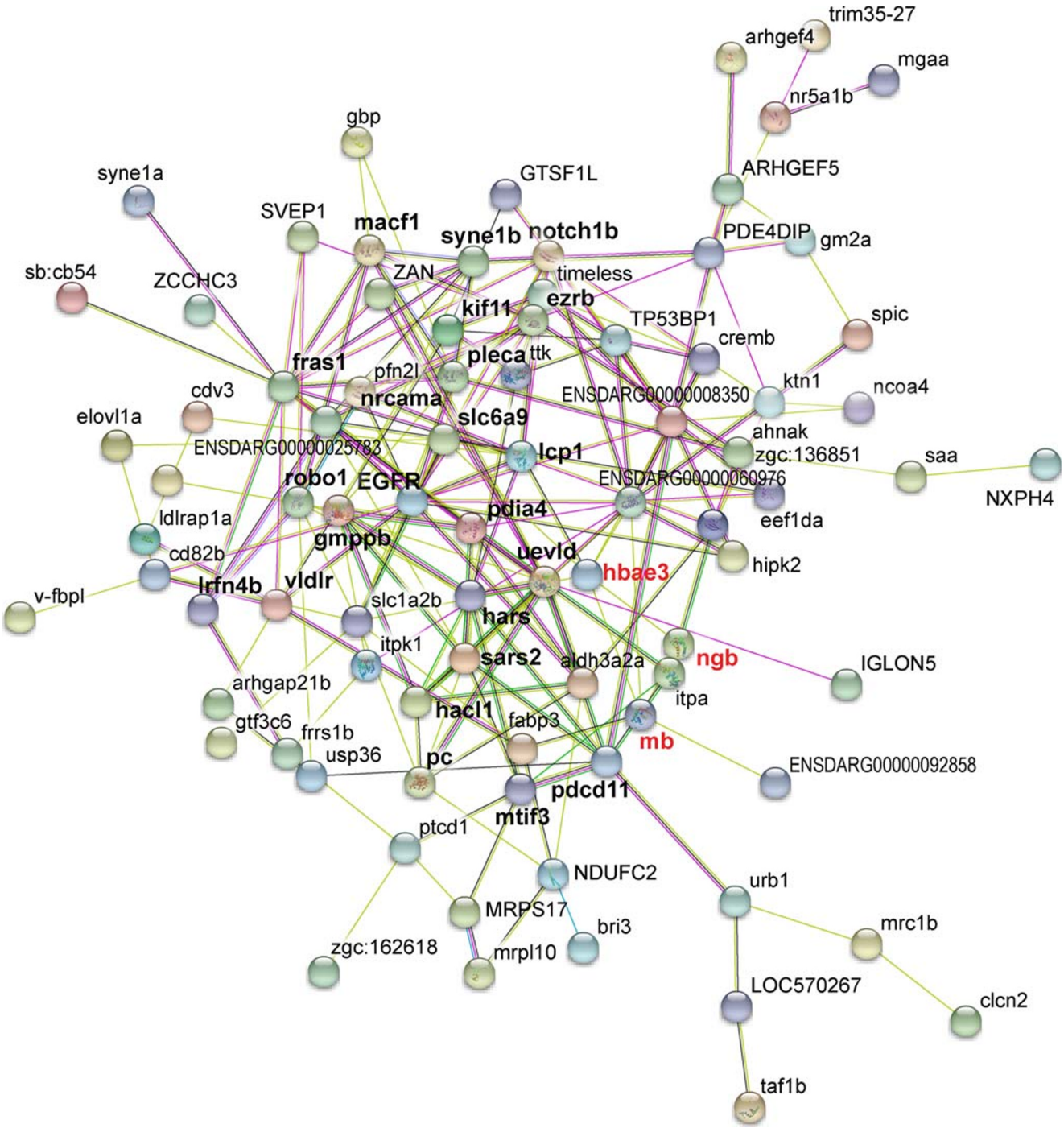

Fig. 6 Gene network analyses identified 26 key regulatory hubs genes associated with *Clarias macrocephalus* air-breathing function. Genes with no interaction were excluded in this diagram. Only genes with high degree of connectivity were selected. Each node represents one gene, and the interactions among these genes were represented with different lines

Table 3 Functional enrichment and pathway analysis for 291 key genes in *Clarias macrocephalus*

UniProt key words					
No. 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	*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*

function. Taken together, *mb*, *ngb*, and *hbae* were highly associated with the formation of air-breathing in *C. macrocephalus*.

Subsequent PPI analysis identified a total of 26 genes with a high degree of connectivity (threshold of 0.200) with each node. The three genes, *mb*, *ngb*, and *hbae* are among them (Fig. 6), suggesting a critical role in the occurrence of air-breathing organ function and growth in the development of *C. macrocephalus*. The other hub genes are mainly related to the function of calcium ion binding and chromatin binding, maintenance of myofiber integrity, the regulation of epidermal-basement membrane adhesion and organogenesis during development, maintenance of cell and tissue integrity, and orchestrating dynamic changes in cytoarchitecture and cell shape, serving as scaffolding platforms for the assembly, positioning, and regulation of signaling complexes. These 26 key genes were expressed at low levels at 3 dpf, dramatically increased at 5 dpf and then steadily decreased through 24 dpf (Fig. 7 and Table S8).

Validation of Species-Specific Genes

The 18S rDNA and β-actin reference genes were expressed high levels in both *C. macrocephalus* and *I. punctatus*. All the selected key genes, including *pdcd*, *mb*, *fras*, *vldlr*, *pdia4*, *uevld*, *slc6a9*, and *hars*, exhibited species-specific expression patterns in *C. macrocephalus* as well as absent in *I. punctatus* (Fig. S4). This is consistent with our hypothesis that these genes are potentially critical for the air-breathing ability in *C. macrocephalus*.

Discussion

For water breathing fish, gills served as the primary gas exchange organ for millions of years. Certain fish ancestors developed air-breathing organs to adapt to low dissolved oxygen level environments (Szklarczyk et al. 2016). *C. macrocephalus* is a descendent of those air-breathing fishes, which can use an accessory air-breathing organ to breathe in air, allowing them to survive in extreme environments, such as hypoxia and high ammonia (Hsia et al. 2013). Therefore, *C. macrocephalus* could be a perfect model to explore the mechanisms of air-breathing function. In this study, low oxygen challenge experiments were conducted under anoxic conditions. We also tested larval fish with exposure to atmospheric air. These latter tests revealed critical timepoints for the development of air-breathing capability as well as the formation of ABO in *C. macrocephalus*.

The water-filled tank experiments did not elucidate the air breathing development of *C. macrocephalus* as the system was not completely effective and dropped the oxygen level to 0.3 mg/L instead of the water becoming completely hypoxic. At all developmental stages, the bighead catfish had 80–100% survival in this system. This indicates that at even at a very young age, the bighead catfish is highly tolerant of low dissolve oxygen, which is likely correlated to

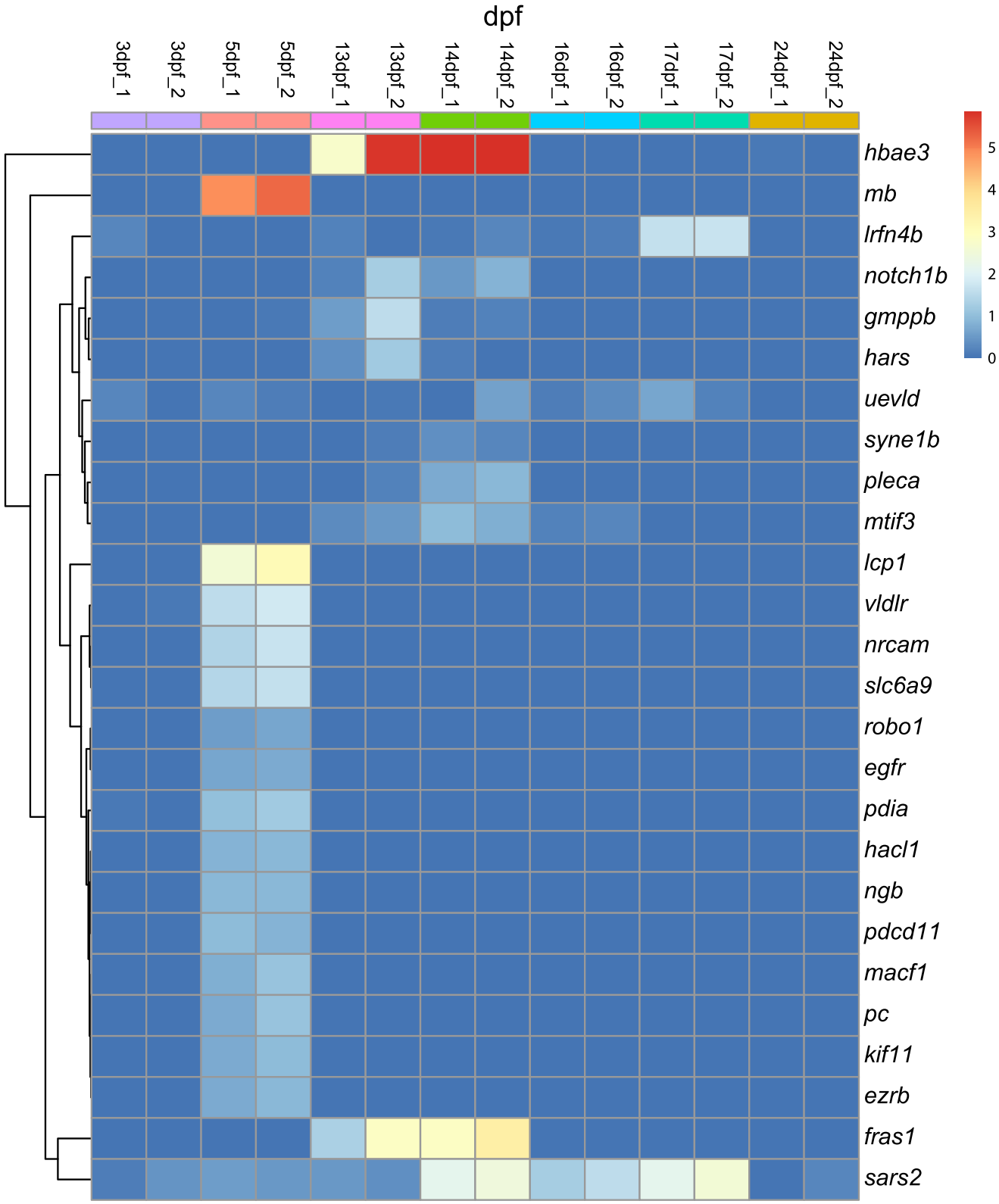

Fig. 7 Heatmap of gene expression of 26 key regulatory hubs genes associated with *Clarias macrocephalus* air-breathing function from day 3 post-fertilization through day 24 post-fertilization

its air breathing ability, and may be an important step towards terrestrial living.

The petri dish experiment, which forces the fish to breathe air while keeping their skin moist by misting, was more

revealing. No fish survived from 3 to 16 dpf, but time to first death and last death progressively increased during this time period, indicating that none of the fish had complete air breathing ability, but it was developing over time. At day 17 dpf, 42% of the population had air breathing ability based on 42% of the population surviving out of water for 7 h. By 24 dpf, this number had increased to 60% of the population. At these 2 dates, the survivors were the largest individuals. Apparently, like other developmental traits in fish like sex determination (Dunham 2011; Gabriel et al. 2017), size is likely more important than age for achieving complete development. This is another reason that asynchronous development is expected in the population. Some gene expressions might have stronger signals if the experiment were to be extended a larger period of time.

To the best of our knowledge, our work is the first to perform RNA sequencing during the development of *C. macrocephalus*, including 3 dpf, 5 dpf, 13 dpf, 14 dpf, 16 dpf, 17 dpf, and 24 dpf. In total, 8675 DEGs were detected during the early development of *C. macrocephalus.* 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. These enriched GO terms were illustrated to be mainly related to the oxygen binding and transport, which is believed to be benefits for improving the ability and efficiency of air-breathing ability. The genomic contents of *C. macrocephalus* and *I. punctatus* were compared, and 1458 unique genes were identified in *C. macrocephalus* that were absent from *I. punctatus*. Two hundred ninety-one genes were detected to be both *Clarias*-specific genes and DEGs across the development of *C. macrocephalus*, indicating that these genes might be critical to the formation of air-breathing function in *C. macrocephalus*. Twenty-six hub genes were identified through gene expression and STRING database analyses. There genes have no orthologs between *C. macrocephalus* and *I. punctatus*, with both percentage of identical matches (pident) and query coverage per subject (qcovs) less than 70%. Of these 26 candidate genes, *hbae3*, *pc*, *pleca*, *ezrb*, and *lrfn4b* were found strictly in *Clarias macrocephalus* compared to channel catfish. The remaining 21 genes have at least one paralog in *I. punctatus*, but the sequence similarity is less than 70%. We speculate that all these 26 genes are potentially important for the function of ABO in *Clarias macrocephalus*.

Luo et al. (2016) examined the development of air breathing in larva of Dojo loach, *Misgurnus anguillicaudatus*, which is an intestinal air breather. This strategy is much different than in *Clarias*, and apparently, a totally different evolutionary and genomic path was taken to accomplish this type of air breathing as none of the 26 *Clarias macrocephalus* genes in the current study matched the 25 key Dojo loach genes. Li et al. (2018) conducted comparative genomics for *I. punctatus* and *C. batrachus* using adult fish. They identified copy number expansion of *mb* (myoglobin), *ora1* (olfactory receptor class A related 1), and *sult6b1* (sulfotransferase 6b1) genes in the air-breathing *C. batrachus* genome, with 15, 15, and 12 copies, respectively. 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* (Li et al. 2018). The authors pointed out that ABOs are highly committed to oxygen transport and cellular respiration (Li et al. 2018). Duong et al. (2020) also conducted a genome sequencing and assembly for adult bighead catfish, and also reported an obvious expansion of myoglobin (*mb*), olfactory receptor class A related 1 (*ora1*) genes in *C. macrocephalus* compared to non-air-breathing fishes. The findings of Li et al. and Duong et al. are, in general, in accordance with the current study for air-breathing ability in developing larval *C. macrocephalus*. The *mb* gene, which was also identified in the 2 species of adult *Clarias*, was a key gene as well as *ngb* and *hbae3* that were not reported in the previous studies. All three genes found to be critical in the bighead catfish larva belong to the globin family, which is directly related to oxygen binding and transport (Li et al. 2018; Vlecken et al. 2009; Duong et al. 2020). The two other copy number expansion genes, *sult6b1* and *ora1*, were not identified as key larval genes in our study, which may not play a role in air breathing, or not until later on in development.

Myoglobin (*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 (Tian et al. 2017; Tiedke et al. 2011). In mammals, *mb* is expressed in muscle and cardiac muscles in early embryonic stages and the expression dramatically increases in the later developmental stages (Wittenberg and Wittenberg 2003). In Japanese medaka (*Oryzias latipes*), a fish species that is well adapted to tolerate hypoxia, exposure of adult fish to low oxygen resulted in strong upregulation of myoglobin (Weller et al. 1986). 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 h post-fertilization (hpf). A 50-fold increase in expression was observed between 18 and 31 hpf. Later, at 4–5 dpf, this expression further increased by 10 to 20-fold (Wawrowski et al. 2011, Tiedke et al. 2011). The *mb* expression upregulation during embryonic stages is consistent with the onset of the blood circulation (Vlecken et al. 2009; Tiedke et al. 2011). In our study, the expression profile of *mb* genes belonged to cluster 1 with an increasing expression value through 3–5 dpf (Fig. 5), which is

consistent with what was reported in zebrafish (Vlecken et al. 2009; Tiedke et al. 2011).

Neuroglobin (*ngb*) is a recently discovered vertebrate globin preferentially expressed in the brain and nervous system. It reversibly binds oxygen (Vlecken et al. 2009) and is related to oxygen-dependent oxidative metabolism (Hankeln et al. 2005; Bentmann et al. 2005). In zebrafish, *ngb* expression level is low during early stages, when respiration function is achieved by simple oxygen diffusion. At 4–5 dpf when circulation starts, *ngb* expression starts to increase along with *mb* (Roesner et al. 2006; Grillitsch et al. 2005). Our expression profile analysis found that *ngb* gene also belongs to cluster 1, which in perfectly correlation 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 (Tiedke et al. 2011). Hemoglobin (Hb) consists of two α and two β chains located in the erythrocytes. It dramatically increases the oxygen carrying capacity of the blood (Tian et al. 2017). *Hb* can enhance the efficiency of oxygen transport from the respiratory surfaces to the interior of the body in the lungs, gills, and skin (Roesner et al. 2006). In zebrafish, activation of hemoglobin biosynthesis was reported to be associated with hypoxia-protection (Tian et al. 2017). All of these genes play 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.

These 3 genes and the other 23 hub genes showed variable waves of expression during the critical developmental time points. At 3 dpf, 6 (*hbae3*, *pdia*, *uevld*, *lrfn4b*, *vldlr*, *and sars2*) of these genes had detectable, but low levels of expression. At 5 dpf, when we see evidence of the beginning of air breathing ability, there is an explosion of gene expression with 18 (*ngb*, *mb*, *macf1*, *kif11*, *ezrb*, *nrcam*, *slc6a9*, *lcb1*, *robo1*, *egfr*, *pdia*, *uevld*, *vldlr*, *gmppb*, *sars2*, *hacl1*, *pc*, *and pdcd11*) of the 26 key genes having detectable or high expression. Two stand out with *lcb1* having relatively high and *mb* having extremely high expression. *mb* must have been particularly important at this stage as its expression was not detected before or after this timepoint. From 5 to 13 dpf, survival times in the hypoxic environment greatly increased. By day 13 dpf, the expression of these key genes was steadily decreasing as 10 genes are expressing at detectable levels, but expression of *syne1b*, *notch1b*, *pleca*, *fras1*, *lrnf4b*, *hars*, and *mtif3* was observed for the first time. In particular, expression of *hbae3* was extremely high; *syne1b*, *pleca*, *fras1*, and *notch1b* were only active at this time; *fras1* expression was high and *mtif3* and *sars2* were peaking. Expression levels continued to drop at day 16 dpf with only 7 genes having detectable expression. At day 17 dpf, the first individuals had attained complete air breathing ability, the overall gene expression continued to drop with only 5 of these genes having detectable gene expression, *lrnf4b peaked* and *sars2* had a second peak. It appears most of the groundwork for air breathing had been accomplished by this time. More than half of the population had reached complete air breathing ability by 24 dpf, and only 3 key genes had detectable expression.

These 26 key genes appear to be expressed in waves over a 21-day period for air breathing function and form to come into existence. *sars2* was the only gene detectable the entire 21 days, and experienced 3 separate peaks in expression. *sars2* (*seryl-tRNA synthetase 2, mitochondrial*) encodes the mitochondrial seryl-tRNA synthethase precursor (https://ghr.nlm.nih.gov/gene). Its molecular function involves protein biosynthesis in contents of ATP binding, serine-tRNA ligase activity, and tRNA binding. It is responsible for mitochondrial seryl-tRNA aminoacylation and seryl-tRNA aminoacylation. The survival data indicates that complete air breathing ability is reached between 11 and 15 mm in total length. Future research should examine gene expression based on size as well as age to better define key moments in time and key genes for air-breathing development.

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 function. These two species provide an excellent model to reveal the critical genes that contribute to the development of ABO. Our study is the first report using high-throughput sequencing for transcriptome analysis during the early developmental stages in *C. macrocephalus*. A total of 25,239 unique genes were detected, and 8,675 DEGs were identified during development in *C. macrocephalus*. Through comparative genomic contents analysis between *C. macrocephalus* and *I. punctatus*, 1458 genes were identified to be unique in *C. macrocephalus.* Through gene expression profile and PPI network analysis, and 26 hub genes were identified to be candidate key genes involved in the formation of air-breathing function in *C. macrocephalus.* Future overexpression, knock-out, and in situ hybridization could be performed to further investigate the candidate key genes for air-breathing.

Acknowledgments Ma, Xiaoli, a graduate student at School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, was supported by the China Scholarship Council. X.W. is supported by Alabama Agricultural Experiment Station Enabling Grant, National Science Foundation EPSCoR RII Track-4 Research Fellowship (NSF OIA 1928770), the USDA National Institute of Food and Agriculture (Hatch

project 1018100), and a generous laboratory start-up fund from Auburn University College of Veterinary Medicine. The authors would like to thank the Auburn Hopper supercomputer clusters for computational support.

Funding This work was supported by an IGP grant (no. 180302) from the Auburn University Office of Vice President for Research.

Data Availability The raw RNA-seq data is available at NCBI GEO (Gene Expression Omnibus) databases under the accession number GSE151993.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M (2012) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41:D991–D995

Bas A, Forsberg G, Hammarström S, Hammarström ML (2004) Utility of the housekeeping genes 18S rRNA, β-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. Scand J Immunol 59:566–573

Belão TC, Leite CAC, Florindo LH, Kalinin AL, Rantin FT (2011) Cardiorespiratory responses to hypoxia in the African catfish, *Clarias gariepinus* (Burchell 1822), an air-breathing fish. J Comp Physiol B 181:905–916

Bentmann A, Schmidt M, Reuss S, Wolfrum U, Hankeln T, Burmester T (2005) Divergent distribution in vascular and avascular mammalian retinae links neuroglobin to cellular respiration. J Biol Chem 280: 20660–20665

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Brauner C, Matey V, Wilson J, Bernier N, Val A (2004) Transition in organ function during the evolution of air-breathing; insights from *Arapaima gigas*, an obligate air-breathing teleost from the Amazon. J Exp Biol 207:1433–1438

Bruton MN (1979) The survival of habitat desiccation by air breathing clariid catfishes. Environ Biol Fish 4:273–280

Chatchaiphan S, Srisapoome P, Kim J-H, Devlin RH, Na-Nakorn U (2017) *De novo* transcriptome characterization and growth-related gene expression profiling of diploid and triploid bighead catfish (*Clarias macrocephalus* Günther, 1864). Mar Biotechnol 19:36–48

Dunham RA (2011) Aquaculture and fisheries biotechnology: genetic approaches 2nd edition. CABI Publishing, Wallingford 504p

Duong T-Y, Tan MH, Lee YP, Croft L, Austin CM (2020) Dataset for genome sequencing and *de novo* assembly of the Vietnamese bighead catfish (*Clarias macrocephalus* Günther, 1864). Data Brief 31: 105861

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157

Fishbase (2020) URL: https://www.fishbase.se/summary/Clarias-macrocephalus.html. Accessed 10 June 2020

Gabriel NN, Qiang J, Ma XY, He J, Xu P, Omoregie E (2017) Sex-reversal effect of dietary Aloe vera (Liliaceae) on genetically improved farmed Nile tilapia fry. N Am J Aquac 79:100–105

Graham JB, Lee HJ (2004) 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? Physiol Biochem Zool 77:720–731

Graham JB, Kramer DL, Pineda E (1978) Comparative respiration of an air-breathing and a non-air-breathing characoid fish and the evolution of aerial respiration in characins. Physiol Zool 51:279–288

Grillitsch S, Medgyesy N, Schwerte T, Pelster B (2005) The influence of environmental PO2 on hemoglobin oxygen saturation in developing zebrafish *Danio rerio*. J Exp Biol 208:309–316

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M (2013) *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

Hankeln T, Ebner B, Fuchs C, Gerlach F, Haberkamp M, Laufs TL, Roesner A, Schmidt M, Weich B, Wystub S (2005) Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. J Inorg Biochem 99:110–119

Hsia CC, Schmitz A, Lambertz M, Perry SF, Maina JN (2013) Evolution of air breathing: oxygen homeostasis and the transitions from water to land and sky. Compr Physiol 3:849–915

Hsieh S-TT (2010) A locomotor innovation enables water-land transition in a marine fish. PLoS One 5:e11197

Islam MN, Islam MS, Alam MS (2007) Genetic structure of different populations of walking catfish (*Clarias batrachus* L.) in Bangladesh. Biochem Genet 45:647–662

Johansen K (1970) Air breathing in fishes. In: Hoar WS, Randall DJ (eds) Fish Physiology. Academic Press, New York, pp 361–411

Kumar L, Futschik ME (2007) Mfuzz: a software package for soft clustering of microarray data. Bioinformation 2:5–7

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323

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

Li N, Bao L, Zhou T, Yuan Z, Liu S, Dunham R, Li Y, Wang K, Xu X, Jin Y, Zeng Q, Gao S, Fu Q, Liu Y, Yang Y, Li Q, Meyer A, Gao D, Liu Z (2018) Genome sequence of walking catfish (*Clarias batrachus*) provides insights into terrestrial adaptation. BMC Genomics 19:952

Liem KF, Eclancher B, Fink WL (1984) Aerial respiration in the banded knife fish *Gymnotus carapo* (Teleostei: Gymnotoidei). Physiol Zool 57:185–195

Liu ZJ, Liu SK, Yao J, Bao LS, Zhang JR, Li Y, Jiang C, Sun LY, Wang RJ, Zhang Y, Zhou T, Zeng QF, Fu Q, Gao S, Li N, Koren S, Jiang YL, Zimin A, Xu P, Phillippy AM, Geng X, Song L, Sun FY, Li C, Wang XZ, Chen A, Jin YL, Yuan ZH, Yang YJ, Tan SX, Peatman E, Lu JQ, Qin ZK, Dunham RA, Li ZX, Sonstegard T, Feng JB, Danzmann RG, Schroeder S, Scheffler B, Duke MV, Ballard L, Kucuktas H, Kaltenboeck L, Liu HX, Armbruster J, Xie YJ, Kirby ML, Tian Y, Flanagan ME, Mu WJ, Waldbieser GC (2016) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nat Commun 7:1–13

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Luo W, Cao X, Xu X, Huang S, Liu C, Tomljanovic T (2016) Developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function of Dojo loach. *Misgurnus anguillicaudatus*. Sci Rep 6:31845

Mcmahon BR, Burggren WW (1987) Respiratory physiology of intestinal air breathing in the teleost fish *Misgurnus anguillicaudatus*. J Exp Biol 133:371–393

Motta PJ (1984) Mechanics and functions of jaw protrusion in teleost fishes: a review. Copeia 1984:1–18

Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7(2):230619

Roesner A, Hankeln T, Burmester T (2006) Hypoxia induces a complex response of globin expression in zebrafish (*Danio rerio*). J Exp Biol 209:2129–2137

Saha N, Ratha BK (2007) Functional ureogenesis and adaptation to ammonia metabolism in Indian freshwater air-breathing catfishes. Fish Physiol Biochem 33:283–295

Shang M, Su B, Perera DA, Alsaqufi A, Lipke EA, Cek S, Dunn DA, Qin Z, Peatman E, Dunham RA (2018) Testicular germ line cell identification, isolation, and transplantation in two North American catfish species. Fish Physiol Biochem 44:717–733

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res 45:D362–D368

Teugels GG, Adriaens D (2003) Taxonomy and phylogeny of Clariidae: an overview. Catfishes 1:465–487

Tian R, Losilla M, Lu Y, Yang G, Zakon H (2017) Molecular evolution of globin genes in Gymnotiform electric fishes: relation to hypoxia tolerance. BMC Evol Biol 17:51

Tiedke J, Gerlach F, Mitz SA, Hankeln T, Burmester T (2011) Ontogeny of globin expression in zebrafish (*Danio rerio*). J Comp Physiol B 181:1011–1021

Untergasser A, Cutcutache L, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115

Vlecken DH, Testerink J, Ott EB, Sakalis PA, Jaspers RT, Bagowski CP (2009) A critical role for myoglobin in zebrafish development. Int J Dev Biol 53:517–524

Wawrowski A, Gerlach F, Hankeln T, Burmester T (2011) Changes of globin expression in the Japanese medaka (*Oryzias latipes*) in response to acute and chronic hypoxia. J Comp Physiol B 181:199–208

Weller P, Price M, Isenberg H, Edwards Y, Jeffreys A (1986) Myoglobin expression: early induction and subsequent modulation of myoglobin and myoglobin mRNA during myogenesis. Mol Cell Biol 6: 4539–4547

Wittenberg JB, Wittenberg BA (2003) Myoglobin function reassessed. J Exp Biol 206:2011–2020

Yu G, Wang L-G, Han Y, He Q-Y (2012) ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16: 284–287

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