



Article

Meta-Transcriptomic Analysis of RNAseq Data Reveals Pacu and Loach Fish with Unusually High Levels of Myoglobin Expression in Skeletal Muscles

Rui-Yi Chen ^{1,2}, Bui Thi Ngoc Hieu ^{3,4}, Gilbert Audira ^{3,4}, Bao Lou ^{5,*}, Ming-Der Lin ^{6,*} 
and Chung-Der Hsiao ^{3,4,7,*} 

¹ Key Lab of Mariculture and Enhancement of Zhejiang Province, Marine Fisheries Research Institute of Zhejiang, Zhoushan 316100, China; xx_cry@163.com

² Marine and Fishery Institute, Zhejiang Ocean University, Zhoushan 316100, China

³ Department of Bioscience Technology, Chung Yuan Christian University, Chung-Li 32023, Taiwan; hieubtn90@gmail.com (B.T.N.H.); gilbertaudira@yahoo.com (G.A.)

⁴ Department of Chemistry, Chung Yuan Christian University, Chung-Li 32023, Taiwan

⁵ Institute of Hydrobiology, Zhejiang Academy of Agricultural Sciences, Shiqiao Road 198, Hangzhou 310021, China

⁶ Department of Molecular Biology and Human Genetics, College of Medicine, Tzu Chi University, 701 Zhongyang Rd, Sec. 3, Hualien 97004, Taiwan

⁷ Center for Nanotechnology, Chung Yuan Christian University, Chung-Li 32023, Taiwan

* Correspondence: Loubao6577@163.com (B.L.); mingder@gms.tcu.edu.tw (M.-D.L.); cdhsiao@cyu.edu.tw (C.-D.H.)

Received: 25 May 2020; Accepted: 27 June 2020; Published: 3 July 2020



Simple Summary: Oxygen-binding proteins that mediate oxygen-binding for storage and consumption, to reduce energy, are very diverse in fish, depending on their habitats. In the present study, oxygen-binding protein gene expression in the skeletal muscle of 25 diverse fish species was examined by a meta-transcriptomic approach. By using RNAseq data, this is the first study to examine the high level of myoglobin, one of the oxygen-binding proteins, transcripts in pacu and loach fish that might be related to their high tolerance for the oxygen-deficient environment. In addition, this study presents the power of the current method to compare the fish oxygen-binding protein expression and its putative gene expansion event.

Abstract: Oxygen-binding proteins, such as myoglobin, hemoglobin, neuroglobin, and cytoglobin, play a role in oxygen binding and delivery to tissues. In icefish, the loss of myoglobin and hemoglobin genes has been reported to be an adaptive evolution event. This interesting finding prompted us to examine oxygen-binding protein expression in diverse fish species. Taking advantage of substantial RNAseq data deposited in the NCBI (National Center for Biotechnology Information) database, we adopted a meta-transcriptomic approach to explore and compare four oxygen-binding protein gene expression levels in the skeletal muscle of 25 diverse fish species for the first time. RNAseq data were downloaded from the NCBI Sequence Read Archive (SRA) database, and de novo assembly was performed to generate transcript contigs. The genes encoding oxygen-binding proteins were then identified by the BLAST search, and the relative expression level of oxygen-binding protein genes was normalized by the RPKM (Reads per Kilobase Million) method. By performing expression profiling, hierarchy clustering, and principal component analysis, pacu and loach fish were noticed by their high myoglobin expression levels in skeletal muscle tissues among 25 diverse fish species. In conclusion, we demonstrated that meta-transcriptomic analysis of RNAseq data is an informative approach to compare the oxygen-binding protein expression and putative gene expansion event in fish.

Keywords: myoglobin; hemoglobin; fish; meta-transcriptomic

1. Introduction

Oxygen homeostasis is an essential mechanism in living organisms that is regulated by four major oxygen-binding proteins, namely hemoglobin, myoglobin, cytoglobin, and neuroglobin [1]. The primary classification of these four oxygen-binding proteins is based on their tissue distribution and oxygen-binding affinity. The major role of these oxygen-binding proteins is on mediating oxygen-binding for storage and consumption, to reduce energy by respiratory progress. Hemoglobin is a tetramer comprising two α and two β globin chains ($\alpha_2\beta_2$), one molecule of heme, and ferric ions. Hemoglobin is essential for aerobic-respiratory fishes, whose living habitats are highly abundant in oxygen [2–4]. Myoglobin is a monomer and predominantly expressed in skeletal and cardiac muscles, and it is an essential intermediary factor to mediate an oxygen delivery from blood to muscle tissues [5,6]. Cytoglobin is a hemoprotein expressed in response to oxidative stress in a variety of tissues. It facilitates the diffusion of oxygen through tissues and helps intracellular oxygen transport [7,8]. The neuroglobin is a monomeric globin with a higher oxygen affinity than hemoglobin. It has been reported that neuroglobin plays a protective role against nitric oxide toxicity in the nervous system [9,10].

Fish are the most widespread and divergent vertebrates that adopted to diverse habitats with different salinity (freshwater, brackish water, or seawater), distinct dissolved oxygen, alkalinity, and temperatures. Some freshwater species, such as crucian carp (*Carassius carassius*) can survive in extremely oxygen-deficient conditions by relying more on subtle neuromodulatory mechanisms, such as ethanol production as their metabolic end-product, caused by duplicated pyruvate dehydrogenase genes, for anoxic metabolic depression [11,12]. In addition, another freshwater species, the pacu fish (*Piaractus mesopotamicus*), which has a high economic value in South America, utilizes aquatic surface respiration (ASR) in order to survive in this condition [13]. Moreover, this species can also survive in conditions with less than 9 mm Hg oxygen concentrations under the intensive aquaculture conditions in fishponds. There is increasing attention being given to the pacu fish, particularly from researchers from Brazil; this interest enhances the farming of this species. Accordingly, some studies on the reproduction [14], feeding, and nutrition [15–17] of pacu fish have been carried out [18]. Several fish species have been evolved in dealing with hypoxia or even anoxia conditions, such as pond loach (*Misgurnus anguillicaudatus*), by aquatic surface respiration [19], and Siamese fighting fish (*Betta splendens*), by labyrinth organ respiration [20]. Instead of the tolerance of hypoxia conditions, the Atlantic cod (*Gadus morhua* L.) displays avoidance behavior from the hypoxia environment, by enhancing oxygen sensing [21].

Gene expression profiling based on the difference in expression levels, such as microarray, which detects the upregulation or downregulation of target genes in certain physiological or pathological conditions in fish, is widely used in the study of biological processes [22–24]. However, one major disadvantage of such a microarray-based transcript detection method lies in its limitation on detecting transcripts with low expression levels. In addition, due to the limitation of the probe density in a single chip, the gene of interest could be missed if the corresponding probe does not exist on the chip. Compared to the microarray-based method, RNAseq deep sequencing provides a complete transcript content and thus can quantify gene expression according to a given reference genome. Therefore, the next-generation sequencing techniques for RNAseq have become an alternative or even better tool to the researches in the field of transcriptomic analysis due to its high-throughput, unbiased, and relatively cost-effective nature. Taken together, by taking advantage of the abundant RNAseq entries in the NCBI (National Center for Biotechnology Information) Sequence Read Archive (SRA) database, we applied a meta-transcriptomics approach to explore the expression profiling of four oxygen-binding protein genes in skeletal muscle from 25 diverse fish species.

2. Materials and Methods

2.1. RNA Extraction

The muscle tissues dissected from wild juvenile Bombay duck (body length around 2 cm) and beltfish (*Trichiurus lepturus*) were immediately stored in RNAlater (Qiagen, Hilden, Germany) and then stored at $-80\text{ }^{\circ}\text{C}$, prior to RNA extraction. Total RNAs were extracted by using the TRIZOL Kit (Invitrogen, Carlsbad, CA, USA), following the manufacturer's instructions. Total RNA samples were then digested by DNase I (Thermo Scientific, Waltham, MA, USA), to remove potential genomic DNA contamination. Integrity and size distribution were checked with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) (with ethics document approval ID CYCU104024).

2.2. RNA Isolation, Library Construction, and Illumina Sequencing

Initially, about 2.5 μg of starting total RNAs was used to synthesize the cDNA libraries, by following the standard protocols of the Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The final library had an insert size of about 200–300 bp. After qPCR quantitation and dilution (by using KAPA SYBR FAST Master Mix), the sequencing library was sequenced on an Illumina NextSeq500 platform with 150 bp paired-end reads. Adaptor sequences were trimmed, and reads with low quality were further removed by cutadapt software [25]. After the removal of ambiguous nucleotides, duplicates, and low-quality sequences (Phred quality scores < 20), cleaned reads were obtained and deposited in the NCBI SRA database (Table 1).

2.3. De Novo Assembly of RNAseq

Other fish muscles' raw next-generation sequencing (NGS) reads were downloaded from the file transfer protocol (FTP) site of Sequence Read Archive (SRA) in the DNA Data Bank of Japan (DDBJ) database (http://trace.ddbj.nig.ac.jp/dra/index_e.html) by FileZilla software (<https://filezilla-project.org/>) and then converted into fastq format by using NCBI sratoolkit (<http://www.ncbi.nlm.nih.gov/Traces/sra/?view=software>). The raw NGS reads of muscle from each fish species were de novo assembled into contigs by CLCBio software (<http://www.clcbio.com/>) with default parameter settings (k-mer = 23 and bubble size = 50). The number and quality (N50) of assembled contigs for all analyzed fish species are summarized in Table 1.

2.4. Identification of Oxygen-Binding Proteins in Muscle Transcriptome

The assembled contigs were later exported from CLCBio software and imported into Geneious software, to perform in-house basic local alignment search tool (BLAST), to find regions of local similarity between sequences. We used tBLASTN, a mode operation for BLAST that aligns protein sequences to a nucleotide database translated in all six frames, to search the potential contigs coding for either myoglobin, hemoglobin, cytoglobin, or neuroglobin with an e-value cutoff setting at 1×10^{-5} . The potential contig was later BLASTX against the reference sequence (Refseq) database, a comprehensive, integrated, non-redundant, well-annotated set of reference sequences, including genomic, transcript, and protein, to validate the gene annotation with an e-value cutoff of 1×10^{-5} . For gene expression calculation, the raw reads were mapped to the assembled contigs by CLCBio software, and the relative expressional level was calculated as RPKM (Reads per Kilobase Million). The number and quality (N50) of the oxygen-binding-related proteins of 25 analyzed fish species were summarized in Table 1.

2.5. PCA, Heatmap, and Clustering Analysis

In order to compare the relative gene expression patterns of four oxygen-binding proteins in muscle tissues among 25 different fish species, PCA was performed, and heatmap clusters were generated. The relative contig number and RPKM values for each oxygen-binding protein were input

into and saved as an Excel file (.xlsx), by using Microsoft Excel, and the Excel file was converted to a comma-separated values-type file (.csv) to be readable by R software. PCA, heatmap, and clustering analysis were carried out by R software (<https://www.r-project.org/>). In addition, mean centering was carried out, since each variable was treated equally, in the sense that larger relative changes in response will receive the most weighting. In the PCA, variable group classification was determined by using the k-means clustering algorithm, and then the clusters returned by the k-means algorithm were used to color variables [26].

Table 1. Summary of the fish species and muscle transcriptomic data used in this study.

| SRR ID | Species Name | Data Size (Gb) | Sequencing Platform | Sequencing Method | Assembled Contigs | N50 |
|------------------------|------------------------------------|----------------|-----------------------------|-------------------|-------------------|-----|
| SRR1613325 | <i>Megalobrama amblycephala</i> | 2.3 | Illumina HiSeq 2000 | 2X101 bp | 35052 | 557 |
| SRR1652342 | <i>Paramisgurnus dabryanus</i> | 3.6 | Illumina HiSeq 2000 | 2X102 bp | 49125 | 972 |
| SRR1533666 | <i>Thymallus thymallus</i> | 3 | Illumina HiSeq 2000 | 2X101 bp | 102377 | 489 |
| SRR1533644 | <i>Pangasianodon hypophthalmus</i> | 3.6 | Illumina HiSeq 2000 | 2X101 bp | 86400 | 454 |
| SRR1056356 | <i>Piaractus mesopotamicus</i> | 8.1 | Illumina HiSeq 2000 | 2X101 bp | 131747 | 483 |
| SRR514104 | <i>Callorhynchus milii</i> | 6.4 | Illumina Genome Analyzer II | 2X77 bp | 44417 | 338 |
| SRR1524253 | <i>Lepisosteus oculatus</i> | 3.7 | Illumina HiSeq 2000 | 2X101 bp | 82408 | 517 |
| SRR1524264 | <i>Amia calva</i> | 4.1 | Illumina HiSeq 2000 | 2X101 bp | 94702 | 554 |
| SRR1524274 | <i>Oryzias latipes</i> | 4.6 | Illumina HiSeq 2000 | 2X101 bp | 100420 | 560 |
| SRR1532750 | <i>Apteronotus albifrons</i> | 3.6 | Illumina HiSeq 2000 | 2X101 bp | 124794 | 551 |
| SRR1532760 | <i>Anguilla anguilla</i> | 2.7 | Illumina HiSeq 2000 | 2X101 bp | 86397 | 429 |
| SRR1532781 | <i>Salmo trutta</i> | 5.5 | Illumina HiSeq 2000 | 2X101 bp | 121092 | 486 |
| SRR1532792 | <i>Osteoglossum bicirrhosum</i> | 3.2 | Illumina HiSeq 2000 | 2X101 bp | 107973 | 469 |
| SRR1532803 | <i>Alosa alosa</i> | 4.8 | Illumina HiSeq 2000 | 2X101 bp | 115049 | 502 |
| SRR1532771 | <i>Pantodon buchholzi</i> | 2.6 | Illumina HiSeq 2000 | 2X101 bp | 93563 | 567 |
| SRR1533633 | <i>Umbra pygmaea</i> | 3.2 | Illumina HiSeq 2000 | 2X101 bp | 107513 | 583 |
| SRR1533655 | <i>Esox lucius</i> | 4 | Illumina HiSeq 2000 | 2X101 bp | 73132 | 550 |
| SRR1533677 | <i>Coregonus lavaretus</i> | 3.9 | Illumina HiSeq 2000 | 2X101 bp | 93899 | 481 |
| SRR1533689 | <i>Perca fluviatilis</i> | 5.1 | Illumina HiSeq 2000 | 2X101 bp | 75135 | 563 |
| SRR1533701 | <i>Gnathonemus petersii</i> | 4.3 | Illumina HiSeq 2000 | 2X101 bp | 136066 | 569 |
| SRR1533711 | <i>Plecoglossus altivelis</i> | 3.1 | Illumina HiSeq 2000 | 2X101 bp | 115114 | 537 |
| SRR3029231 | <i>Larimichthys polyactis</i> | 1.9 | Illumina HiSeq 2000 | 2X101 bp | 60237 | 715 |
| SRR1524241 | <i>Danio rerio</i> | 4.3 | Illumina HiSeq 2000 | 2X101 bp | 118459 | 509 |
| SRR332188 ¹ | <i>Trichiuus lepturus</i> | 5.7 | Illumina NextSeq 500 | 2X151 bp | 36214 | 860 |
| SRR338795 ¹ | <i>Harpadon nehereus</i> | 2.3 | Illumina NextSeq 500 | 2X151 bp | 21599 | 615 |

¹ Those two RNAseq data were performed in the current study and were partly reported in our previous publications [27,28]. The other 23 RNAseq data were downloaded from the public Sequence Read Archive (SRA) database.

3. Results and Discussion

3.1. Meta-Transcriptomic Analysis of Oxygen-Binding Protein Expression

Fish can survive in a diverse range of water oxygen levels. By radiation adaptation, the expression of hemoglobin and myoglobin genes is absent in some Antarctic icefishes [29,30]. This interesting phenomenon led us to ask whether the expression levels of oxygen-binding protein genes display a unique pattern among various fish species that live in a diverse habitat with different water oxygen levels. To answer this question, we used a bioinformatics approach to compare the relative expression levels of two major oxygen-binding protein genes, *hemoglobin* and *myoglobin*, in the skeletal muscle of fish. In addition, other two oxygen-binding protein genes, *cytoglobin* and *neuroglobin* were also analyzed, in order to serve as control groups. Taking the advantage of the rich information in the Sequence Read Archive (SRA) database, we sequenced or collected muscle RNAseq data from 25 fish species, to perform comparative studies systematically (species names are listed in Table 1). We used CLCBio software to assemble raw RNAseq reads into contigs by using the default setting (analysis pipeline is summarized in Figure 1). The detailed information for fish species, raw data size, sequencing method, assembly contig number, and assembly quality (N50) were summarized in Table 1. The size of raw SRA data for different fish species ranged from 1.9 to 8.1 Gb. The size of the assembled contig number for different fish species ranged from 21,259 to 136,066, with N50 ranging from 338 to 972. The oxygen-binding protein homologs of hemoglobin, myoglobin, cytoglobin, and neuroglobin were identified by using BLAST search, and the relative number of assembled

contigs matched with oxygen-binding proteins for different fish species is summarized in Appendix A, Table A1. In RNAseq datasets of fish skeletal muscle, based on relative gene expression level, we found the predominant oxygen-binding proteins expressed are hemoglobin and myoglobin. On the contrary, the number of annotated contigs encoding cytoglobin or neuroglobin is very low or even absent in most fish species (Appendix A, Table A1). Among 25 fish species analyzed, we found the pacu fish (*Piaractus mesopotamicus*) displays a unique oxygen-binding protein expression due to its outstanding high number for assembled contigs annotated as either *hemoglobin* (65 transcripts) and *myoglobin* (31 transcripts) (Appendix A, Table A1). By contrast, the contig numbers of *hemoglobin* and *myoglobin* in other fish species are usually less than 16 and 6, respectively (Appendix A, Table A1).

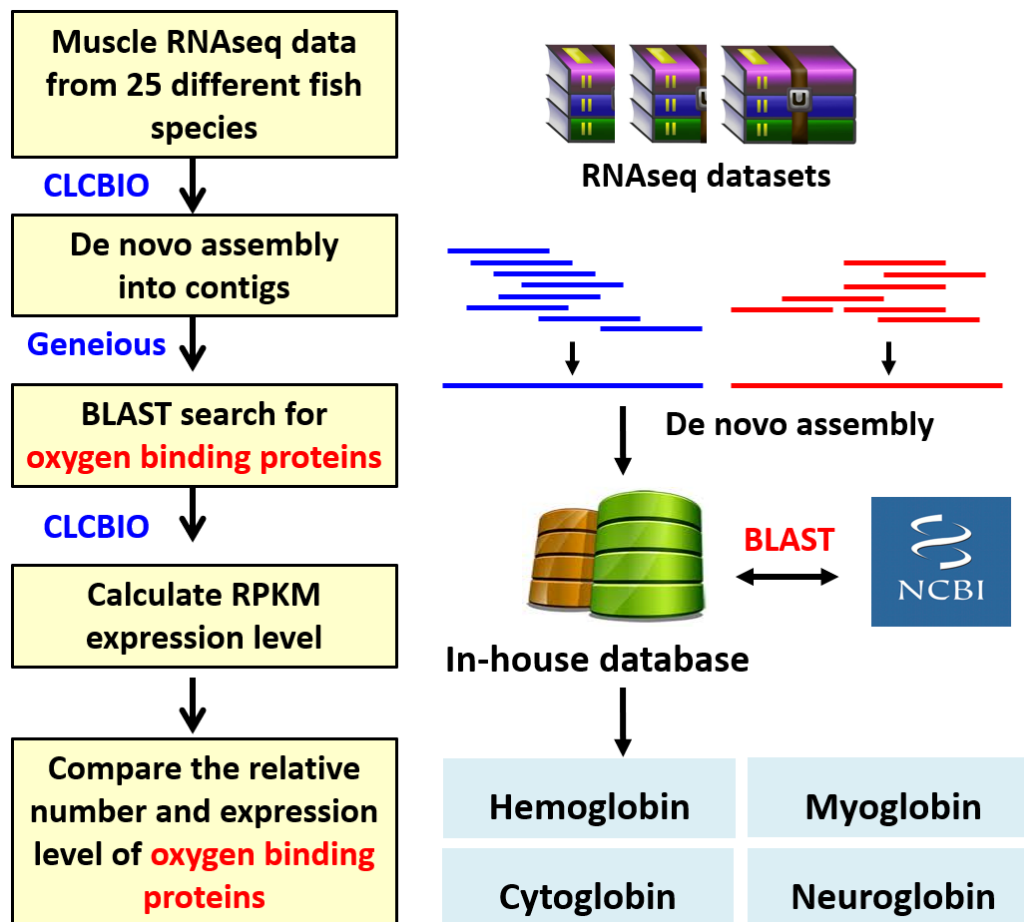


Figure 1. Summary of the experimental workflow from bioinformatic analysis of the expression of oxygen-binding proteins in fish. We performed de novo gene assembly, gene mining by BLAST search, and gene mapping to compare the expressional level of four major oxygen-binding proteins (hemoglobin, myoglobin, cytoglobin, and neuroglobin) in 25 different fish species.

3.2. Comparison of the Expression Levels of Oxygen-Binding Protein Genes among Fish Species

Next, we performed gene mapping to compare the relative expression level of *hemoglobin* and *myoglobin* in skeletal muscle tissues among diverse fish species. The relative expression level of oxygen-binding protein homologs was calculated and presented as RPKM (Reads per Kilobase Million) for expressional normalization [31]. As a result, we identified that the pacu fish has a high *myoglobin* expression level of 1327.8, based on RPKM (Figure 2A, Appendix A, Table A2). Although the number of contigs annotated as *hemoglobin* is high in pacu fish, their relative expression level is low (Figure 2B). Another interesting species is the loach fish (*Paramisgurnus dabryanus*), showing one *myoglobin*-associated contig, but with an extremely high RPKM expression level, at 3136.4 (Figure 2A, Appendix A, Table A2). Based on hemoglobin transcript abundance, we found the Blunt snout bream

(*Megalobrama amblycephala*), an oxygen-sensitive fish species, displays a relatively high hemoglobin expression level in the skeletal muscle (Figure 2B). The other two oxygen-binding protein genes, *cytoglobin* and *neuroglobin* can be considered as low-expressed genes in skeletal muscle since their RPKM level is less than 2.5 in all fish species examined. The phylogenetic tree based on complete mitochondrial DNA sequences from these fish species can be found in Figure 2C.

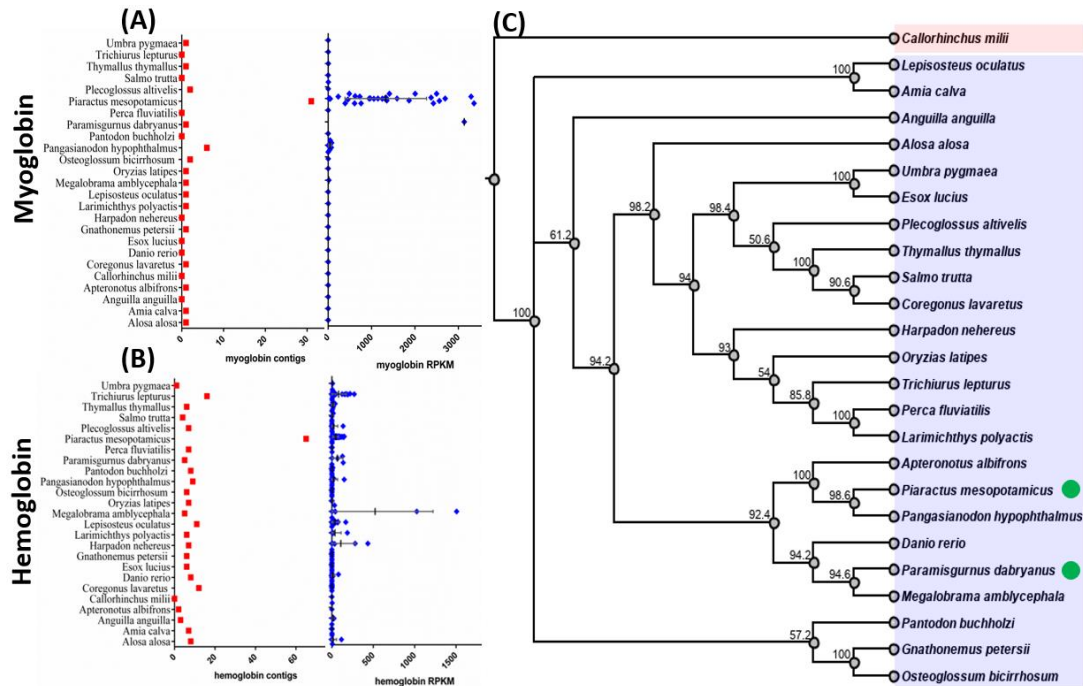


Figure 2. The assembled contig number and relative expression levels of myoglobin and hemoglobin of 25 diverse fish species. (A) The gene contig number (red color) and expression level (blue color) comparison for myoglobin in fish. (B) The gene contig number (red color) and expression level (blue color) comparison for hemoglobin in fish. The gene expression level was normalized and compared by Reads per Kilobase Million (RPKM) values. (C) The phylogenetic tree based on complete mitochondrial DNA sequences from 25 fish species. The multiple sequence alignment and phylogenetic tree construction were performed by Geneious software and bootstrap value, to support the tree topology, and are indicated in the root position of each branch. The red- and blue-shaded color column represents cartilaginous and teleost lineages, respectively. The green dots represent two fish species that were identified with high myoglobin expression in the skeletal muscle.

3.3. PCA and Hierarchical Clustering Analyses of the Oxygen-Binding Protein Genes Expression among Fish Species

Next, we performed principal component analysis (PCA) and hierarchical clustering, to reduce the data dimension and complexity, and to generate a heatmap for inter-species comparison. PCA is a mathematical method able to reduce data dimension and complexity. From the PCA analysis, all of the 25 fishes were grouped into four major clusters, based on their gene expression patterns for *hemoglobin*, *myoglobin*, *cytoglobin*, and *neuroglobin*, with the first cluster (red color) as the biggest cluster among all the clusters (Figure 3B). The majority of the fishes, including zebrafish and medaka fish, were grouped in the first cluster (red color) (Figure 3B), which has higher *hemoglobin* expression than the fish species in other clusters (Figure 3A). Four fishes, namely *Amia calva*, *Pangasianodon hypophthalmus*, *Paramisgurnus dabryanus*, and *Piaractus mesopotamicus*, were grouped into the second cluster (blue color). Based on the heatmap results, it can be concluded that this grouping is reasonable because these fishes have unique and similar patterns regarding their *myoglobin* expression level, which is quite high compared to the rest of the fishes (Figure 3A). The third cluster (pink color) consists of two species, *Gnathonemus petersii* and *Callorhynchus milii*. This categorization is plausible since both of the fishes showed a high

level of *cytoglobin* expression in the skeletal muscle, while this phenomenon was not observed in other fishes (Figure 3A). Furthermore, the fourth cluster (green color) consists of *Apteronotus albifrons*, *Coregonus lavaretus*, and *Osteoglossum bicirrhosum*, showing a high-level expression of both *myoglobin* and *hemoglobin* in their skeletal muscle (Figure 3A). Taken together, the different expression levels of oxygen-binding protein genes in each fish species examined resulted in four major clusters in the PCA and heatmap clustering analyses.

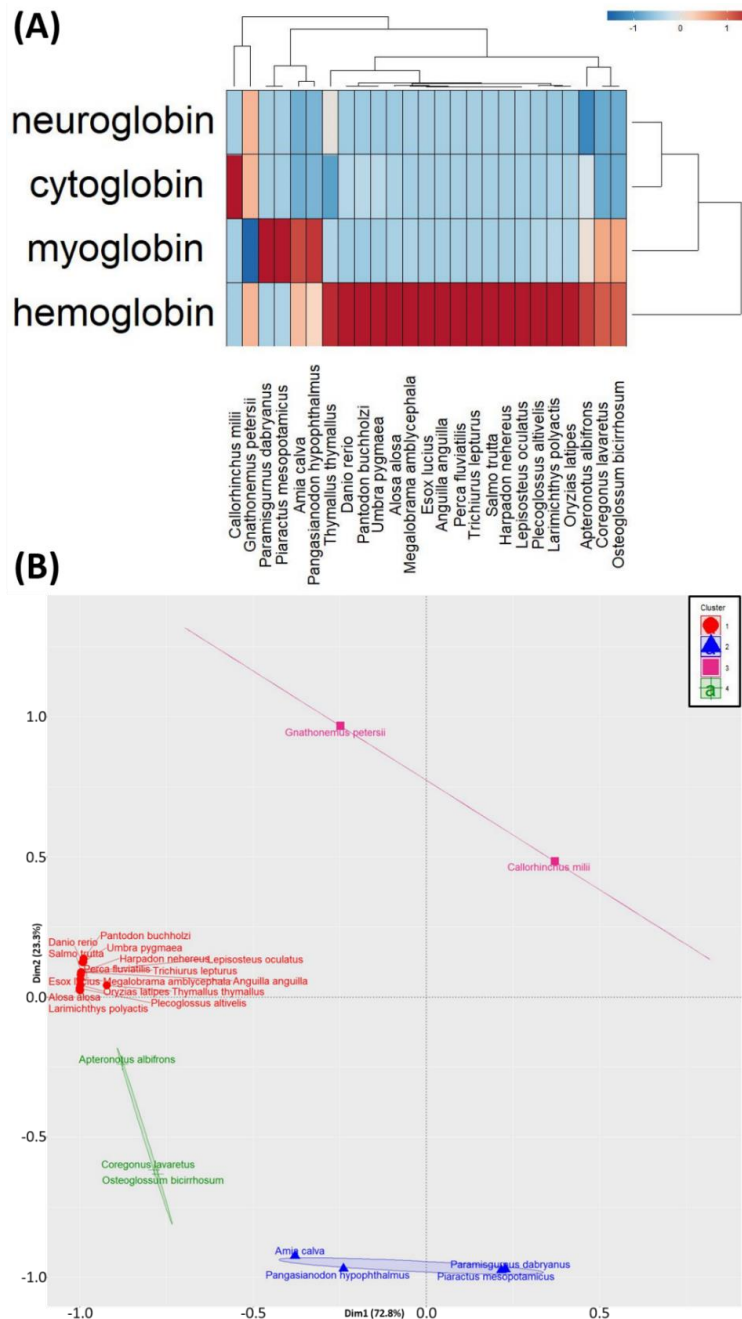


Figure 3. Comparison of the relative oxygen-binding protein gene expression among 25 diverse fish species. **(A)** Hierarchical clustering analysis of the relative oxygen-binding protein gene expression levels among 25 diverse fish species. **(B)** Principal component analysis (PCA) of the relative oxygen-binding protein gene expression levels among 25 diverse fish species. Totally, 25 fish species can be categorized into four major groups (shown by different color labeling). Pacu (*Piaractus mesopotamicus*) and loach (*Paramisgurnus dabryanus*) fish were identified with high myoglobin expression in skeletal muscle and were grouped in the second group (blue color).

3.4. The Skeletal Muscle of Pacu Fish Has Multiple Myoglobin Transcripts with High Expression Level

Our meta-transcriptomic study provides molecular evidence that there is a high level of *myoglobin* expression in the skeletal muscle of pacu fish. This finding may help to explain their ability to cope with oxygen-deficiency habitats. Moreover, we identified 31 transcripts annotated as *myoglobin*. The numerous forms of *myoglobin* transcripts identified in the pacu fish muscle could be originated by a large number of duplications of *myoglobin* genes in its genome. In common carp, duplication and differentiation of *myoglobin* genes have been identified by bacterial artificial chromosome (BAC) analysis [32]. In West African lungfish, *Protopterus annectens*, at least seven distinct *myoglobin* genes (PanMb1–7) have been identified, and each of them displays differential expression patterns in different tissues [33]. Consistent with the hypothesis that *myoglobin* genes are highly duplicated in the pacu genome, we observed several non-redundant *myoglobin* transcripts encoding myoglobin with amino acid substitutions by multiple sequence alignment (Appendix A, Figure A1). Alternatively, we hypothesize that multiple *myoglobin* transcript contigs in pacu fish may be generated by alternative splicing. This hypothesis is based on a previous study that detected sixteen alternatively-spliced *myoglobin* transcripts in human cancer [34]. Future studies on whole-genome sequencing for pacu fish are required to validate the above hypotheses. In addition to *myoglobin*, this study also identified relative high-level *hemoglobin* expression in the skeletal muscle of blunt snout bream (*Megalobrama amblycephala*). Blunt snout bream is an oxygen-sensitive fish species showing massive mortality once the dissolved oxygen level declines in the pond. Therefore, the high-level hemoglobin expression in the skeletal muscle in blunt snout bream can be considered to be a physiological adaptation. More studies are needed to address the underlying mechanism for this interesting phenotype.

4. Conclusions

In conclusion, in this study, we provided good evidence to show the power of meta-transcriptomic analysis on mining and comparing oxygen-binding protein expression in fish muscle RNAseq data for the first time. This approach led us to identify pacu and loach fish with exceptional levels of *myoglobin* transcripts in their skeletal muscle that might be functionally associated with their high-tolerance potential for the oxygen-deficient environments. However, one has to keep in mind that deposited datasets in the SRA that were used in the present study often represent a relatively small number of biological samples, and there is a possibility that the different animals included were maintained under identical or comparable conditions. Thus, validation with RT-qPCR for the fish species is needed in the future. In addition, this approach opens up a new avenue to perform data mining on more species or tissue types, to validate *myoglobin* and *hemoglobin* expressional adaptation among diverse fish species.

Author Contributions: Conceptualization, M.-D.L. and C.-D.H.; Data curation and Formal analysis, R.-Y.C., B.T.N.H. and G.A.; Funding acquisition, C.-D.H.; Investigation and Methodology, B.T.N.H. and G.A.; Project administration, C.-D.H.; Resources, B.L.; Software, G.A.; Supervision, B.L. and C.-D.H.; Validation, B.T.N.H. and G.A.; Visualization, R.-Y.C.; Writing—original draft, R.-Y.C., M.-D.L. and C.-D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the grants sponsored by the Ministry of Science and Technology MOST 108-2313-B-033-001-MY3 and MOST 107-2622-B-033-001-CC2 to C.-D.H.

Acknowledgments: We appreciate the two anonymous reviewers and editors for their professional comments, which improved the quality of this paper. We also thank Jung-Ren Chen at the Department of Biological Science and Technology, College of Medicine, I-Shou University, for doing English editing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Summary of the assembled contigs annotated as transcripts of oxygen-binding protein genes in 25 fish species.

| Species Name | Hemoglobin | Myoglobin | Cytoglobin | Neuroglobin |
|------------------------------------|------------|-----------|------------|-------------|
| <i>Alosa alosa</i> | 8 | 1 | 2 | 0 |
| <i>Amia calva</i> | 7 | 1 | 0 | 0 |
| <i>Anguilla anguilla</i> | 3 | 0 | 0 | 0 |
| <i>Apteronotus albifrons</i> | 2 | 1 | 1 | 0 |
| <i>Callorhinchus milii</i> | 0 | 0 | 1 | 0 |
| <i>Coregonus lavaretus</i> | 12 | 1 | 0 | 0 |
| <i>Danio rerio</i> | 8 | 0 | 2 | 0 |
| <i>Esox lucius</i> | 6 | 0 | 0 | 0 |
| <i>Gnathonemus petersii</i> | 6 | 1 | 1 | 1 |
| <i>Harpadon nehereus</i> | 7 | 0 | 0 | 0 |
| <i>Larimichthys polyactis</i> | 6 | 1 | 1 | 0 |
| <i>Lepisosteus oculatus</i> | 11 | 1 | 0 | 0 |
| <i>Megalobrama amblycephala</i> | 5 | 1 | 0 | 0 |
| <i>Oryzias latipes</i> | 7 | 1 | 2 | 0 |
| <i>Osteoglossum bicirrhosum</i> | 6 | 2 | 0 | 0 |
| <i>Pangasianodon hypophthalmus</i> | 9 | 6 | 0 | 0 |
| <i>Pantodon buchholzi</i> | 8 | 0 | 2 | 0 |
| <i>Paramisgurnus dabryanus</i> | 5 | 1 | 0 | 0 |
| <i>Perca fluviatilis</i> | 7 | 0 | 0 | 0 |
| <i>Piaractus mesopotamicus</i> | 65 | 31 | 0 | 0 |
| <i>Plecoglossus altivelis</i> | 7 | 2 | 1 | 0 |
| <i>Salmo trutta</i> | 4 | 0 | 0 | 0 |
| <i>Thymallus thymallus</i> | 6 | 1 | 2 | 1 |
| <i>Trichiurus lepturus</i> | 16 | 0 | 0 | 0 |
| <i>Umbra pygmaea</i> | 1 | 1 | 1 | 0 |

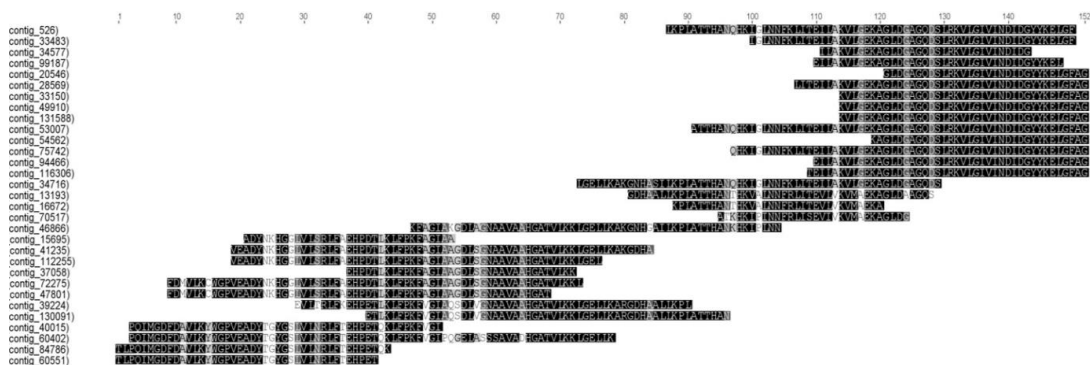


Figure A1. Sequence alignment of the translated contigs annotated as myoglobin in pacu fish. Thirty-one polypeptide translated from *myoglobin* contigs were used for multiple sequence alignment by using Geneious software. The conserved residues are marked as shaded black color, and the non-conserved residues are marked as open or gray-shaded colors.

Table A2. Summary of the transcript expression level based on the RPKM method for oxygen-binding protein genes in 25 fish species.

| Species Name | Hemoglobin | Myoglobin | Cytoglobin | Neuroglobin |
|------------------------------------|------------|-----------|------------|-------------|
| <i>Alosa alosa</i> | 16.1 | 0.6 | 0.8 | 0 |
| <i>Amia calva</i> | 0.5 | 0.8 | 0 | 0 |
| <i>Anguilla anguilla</i> | 14.7 | 0 | 0 | 0 |
| <i>Apteronotus albifrons</i> | 2.2 | 1.1 | 0.9 | 0 |
| <i>Callorhinchus milii</i> | 0 | 0 | 2.4 | 0 |
| <i>Coregonus lavaretus</i> | 0.4 | 0.3 | 0 | 0 |
| <i>Danio rerio</i> | 10.4 | 0 | 0.7 | 0 |
| <i>Esox lucius</i> | 0.4 | 0 | 0 | 0 |
| <i>Gnathonemus petersii</i> | 0.3 | 0.2 | 0.3 | 0.3 |
| <i>Harpadon nehereus</i> | 108.5 | 0 | 0 | 0 |
| <i>Larimichthys polyactis</i> | 38.6 | 3.9 | 2.5 | 0 |
| <i>Lepisosteus oculatus</i> | 37.2 | 0 | 0 | 0 |
| <i>Megalobrama amblycephala</i> | 523.3 | 14.0 | 0 | 0 |
| <i>Oryzias latipes</i> | 5 | 0.4 | 0.3 | 0 |
| <i>Osteoglossum bicirrhosum</i> | 1.3 | 1 | 0 | 0 |
| <i>Pangasianodon hypophthalmus</i> | 20.8 | 43.5 | 0 | 0 |
| <i>Pantodon buchholzi</i> | 2.2 | 0 | 0.2 | 0 |
| <i>Paramisgurnus dabryanus</i> | 67.6 | 3136.4 | 0 | 0 |
| <i>Perca fluviatilis</i> | 1.4 | 0 | 0 | 0 |
| <i>Piaractus mesopotamicus</i> | 46 | 1327.8 | 0 | 0 |
| <i>Plecoglossus altivelis</i> | 20.8 | 1.4 | 0.3 | 0 |
| <i>Salmo trutta</i> | 4.5 | 0 | 0 | 0 |
| <i>Thymallus thymallus</i> | 13.6 | 3.2 | 0.3 | 5.6 |
| <i>Trichiurus lepturus</i> | 81.9 | 0 | 0 | 0 |
| <i>Umbra pygmaea</i> | 8.4 | 0.12 | 0.8 | 0 |

References

1. Wilson, M.T.; Reeder, B.J. Oxygen-binding haem proteins. *Exp. Physiol.* **2008**, *93*, 128–132. [[CrossRef](#)] [[PubMed](#)]
2. Nikinmaa, M.; Rees, B.B. Oxygen-dependent gene expression in fishes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2005**, *288*, R1079–R1090. [[CrossRef](#)]
3. Terwilliger, N.B. Functional adaptations of oxygen-transport proteins. *J. Exp. Biol.* **1998**, *201*, 1085–1098.
4. Lekang, O.I. *Aquaculture Engineering*; Wiley: Hoboken, NJ, USA, 2013.
5. Wittenberg, B.A.; Wittenberg, J.B. Myoglobin-mediated oxygen delivery to mitochondria of isolated cardiac myocytes. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7503–7507. [[CrossRef](#)] [[PubMed](#)]
6. Ordway, G.A.; Garry, D.J. Myoglobin: An essential hemoprotein in striated muscle. *J. Exp. Biol.* **2004**, *207*, 3441–3446. [[CrossRef](#)]
7. Singh, S.; Canseco, D.C.; Manda, S.M.; Shelton, J.M.; Chirumamilla, R.R.; Goetsch, S.C.; Ye, Q.; Gerard, R.D.; Schneider, J.W.; Richardson, J.A. Cytoglobin modulates myogenic progenitor cell viability and muscle regeneration. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E129–E138. [[CrossRef](#)] [[PubMed](#)]
8. Burmester, T.; Ebner, B.; Weich, B.; Hankeln, T. Cytoglobin: A novel globin type ubiquitously expressed invertebrate tissues. *Mol. Biol. Evol.* **2002**, *19*, 416–421. [[CrossRef](#)] [[PubMed](#)]

9. Jin, K.; Mao, X.O.; Xie, L.; Khan, A.A.; Greenberg, D.A. Neuroglobin protects against nitric oxide toxicity. *Neurosci. Lett.* **2008**, *430*, 135–137. [[CrossRef](#)] [[PubMed](#)]
10. Burmester, T.; Weich, B.; Reinhardt, S.; Hankeln, T. A vertebrate globin expressed in the brain. *Nature* **2000**, *407*, 520–523. [[CrossRef](#)] [[PubMed](#)]
11. Nilsson, G.E.; Renshaw, G.M. Hypoxic survival strategies in two fishes: Extreme anoxia tolerance in the North European crucian carp and natural hypoxic preconditioning in a coral-reef shark. *J. Exp. Biol.* **2004**, *207*, 3131–3139. [[CrossRef](#)] [[PubMed](#)]
12. Fagernes, C.E.; Stensløkken, K.-O.; Røhr, Å.K.; Berenbrink, M.; Ellefsen, S.; Nilsson, G.E. Extreme anoxia tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new ethanol-producing pyruvate decarboxylase pathway. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)]
13. Saint-Paul, U.; Bernardino, G. Behavioural and ecomorphological responses of the neotropical pacu *Piaractus mesopotamicus* (Teleostei, Serrasalminae) to oxygen-deficient waters. *Exp. Biol.* **1988**, *48*, 19–26.
14. Romagosa, E.; De Paiva, P.; Godinho, H.M. Pattern of oocyte diameter frequency distribution in females of the pacu, *Piaractus mesopotamicus* (Holmberg 1887)(= *Colossoma mitrei* Berg 1895), induced to spawn. *Aquaculture* **1990**, *86*, 105–110. [[CrossRef](#)]
15. Souza, V.L.; Oliveira, E.; Urbinati, E. Effects of food restriction and refeeding on energy stores and growth of pacu, *Piaractus mesopotamicus* (Characidae). *J. Aquac. Trop.* **2000**, *15*, 371–379.
16. Bechara, J.A.; Roux, J.P.; Ruiz Diaz, F.J.; Flores Quintana, C.I.; Longoni de Meabe, C.A. The effect of dietary protein level on pond water quality and feed utilization efficiency of pacu *Piaractus mesopotamicus* (Holmberg, 1887). *Aquac. Res.* **2005**, *36*, 546–553. [[CrossRef](#)]
17. Takahashi, L.S.; de Abreu, J.S.; Biller, J.D.; Urbinati, E.C. Efeito do ambiente pós-transporte na recuperação dos indicadores de estresse de pacus juvenis, *Piaractus mesopotamicus*-doi:10.4025/actascianimsci. v28i4. 610. *Acta Sci. Anim. Sci.* **2008**, *28*, 469–475.
18. Rantin, F.T.; Guerra, C.D.R.; Kalinin, A.L.; Glass, M.L. The influence of aquatic surface respiration (ASR) on cardio-respiratory function of the serrasalminid fish *Piaractus mesopotamicus*. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **1998**, *119*, 991–997. [[CrossRef](#)]
19. Jasiński, A. Air-blood barrier in the respiratory intestine of the pond-loach, *Misgurnus fossilis* L. *Cells Tissues Organs* **1973**, *86*, 376–393. [[CrossRef](#)]
20. Mendez-Sanchez, J.F.; Burggren, W.W. Hypoxia-induced developmental plasticity of larval growth, gill and labyrinth organ morphometrics in two anabantoid fish: The facultative air-breather Siamese fighting fish (*Betta splendens*) and the obligate air-breather the blue gourami (*Trichopodus trichopterus*). *J. Morphol.* **2019**, *280*, 193–204.
21. Herbert, N.A.; Skjæraasen, J.E.; Nilsen, T.; Salvanes, A.G.; Steffensen, J.F. The hypoxia avoidance behaviour of juvenile Atlantic cod (*Gadus morhua* L.) depends on the provision and pressure level of an O₂ refuge. *Mar. Biol.* **2011**, *158*, 737–746. [[CrossRef](#)]
22. Salem, M.; Kenney, P.B.; Rexroad, C.E., 3rd; Yao, J. Microarray gene expression analysis in atrophying rainbow trout muscle: A unique nonmammalian muscle degradation model. *Physiol. Genom.* **2006**, *28*, 33–45. [[CrossRef](#)]
23. Denslow, N.D.; Garcia-Reyero, N.; Barber, D.S. Fish ‘n’ chips: The use of microarrays for aquatic toxicology. *Mol. Biosyst.* **2007**, *3*, 172–177. [[CrossRef](#)] [[PubMed](#)]
24. Sarropoulou, E.; Kotoulas, G.; Power, D.M.; Geisler, R. Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. *Physiol. Genom.* **2005**, *23*, 182–191. [[CrossRef](#)]
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
26. Kassambara, A. *Practical Guide To Principal Component Methods in R: PCA, M(CA), FAMD, MFA, HCPC, Factoextra*; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2017.
27. Zhang, H.; Audira, G.; Li, Y.; Xian, W.; Varikkodan, M.M.; Hsiao, C.-D. Comparative study the expression of calcium cycling genes in Bombay duck (*Harpadon nehereus*) and beltfish (*Trichiurus lepturus*) with different swimming activities. *Genom. Data* **2017**, *12*, 58–61. [[CrossRef](#)]
28. Zhang, H.; Chang, C.-M.; Shen, K.-N.; Xian, W.; Hsiao, C.-D. Identification of myogenic regulatory genes in the muscle transcriptome of beltfish (*Trichiurus lepturus*): A major commercial marine fish species with robust swimming ability. *Genom. Data* **2016**, *8*, 81–84. [[CrossRef](#)]

29. Sidell, B.D.; O'Brien, K.M. When bad things happen to good fish: The loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J. Exp. Biol.* **2006**, *209*, 1791–1802. [[CrossRef](#)]
30. Grove, T.J.; Hendrickson, J.W.; Sidell, B.D. Two species of Antarctic icefishes (genus *Champscephalus*) share a common genetic lesion leading to the loss of myoglobin expression. *Polar Biol.* **2004**, *27*, 579–585. [[CrossRef](#)]
31. Wagner, G.P.; Kin, K.; Lynch, V.J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **2012**, *131*, 281–285. [[CrossRef](#)]
32. Zhao, Z.-X.; Xu, P.; Cao, D.-C.; Kuang, Y.-Y.; Deng, H.-X.; Zhang, Y.; Xu, L.-M.; Li, J.-T.; Xu, J.; Sun, X.-W. Duplication and differentiation of common carp (*Cyprinus carpio*) myoglobin genes revealed by BAC analysis. *Gene* **2014**, *548*, 210–216. [[CrossRef](#)]
33. Koch, J.; Lüdemann, J.; Spies, R.; Last, M.; Amemiya, C.T.; Burmester, T. Unusual diversity of myoglobin genes in the lungfish. *Mol. Biol.* **2016**, *33*, 3033–3041. [[CrossRef](#)] [[PubMed](#)]
34. Bicker, A.; Dietrich, D.; Gleixner, E.; Kristiansen, G.; Gorr, T.A.; Hankeln, T. Extensive transcriptional complexity during hypoxia-regulated expression of the myoglobin gene in cancer. *Hum. Mol. Genet.* **2014**, *23*, 479–490. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).