Comparative Biochemistry and Physiology, Part D 20 (2016) 111–117

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

# Comparative transcriptomic analyses of two bighead carp (*Hypophthalmichthys nobilis*) groups with different growth rates

Beide Fu [a], Xinhua Wang [a,b], Xiu Feng [a,b], Xiaomu Yu [a], Jingou Tong [a,*]

[a] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China
[b] University of Chinese Academy of Sciences, Beijing 100039, China

ARTICLE INFO

*Article history:*
Received 22 March 2016
Received in revised form 14 July 2016
Accepted 11 August 2016
Available online 14 September 2016

*Keywords:*
Bighead carp
Growth rate
Comparative transcriptomics
RT-PCR

ABSTRACT

Growth is one of the most important and desired economic traits for aquaculture species, and identification of loci controlling growth is a difficult task without genomic sequences. In this study, the liver transcriptomes of two groups (F and S) of bighead carp (*Hypophthalmichthys nobilis*) within a full-sib family with significant differences in growth rate were sequenced. Following de novo assembly of the combined reads from the two groups, a total of 410 differentially expressed genes were identified. Functional annotation and analysis of these genes indicated that some of these were involved in regulation of glucose levels and lipid metabolism, particularly fatty acid oxidation and transport. In addition to the differences on expression levels between the two groups, we also identified many non-synonymous coding single-nucleotide polymorphisms (SNPs) that were specific to each group, including SNPs from 4 genes involved in the lipid metabolism process (GO: 0006629). These differences in gene expression and DNA sequences may in part comprise the genetic background for the regulation of early growth rate in bighead carp.



## 1. Introduction

Growth is regulated by the integration of many environmental factors, such as season, temperature, and the amount of available food along with the endogenous genetic background. Among all of the abovementioned factors, the genetic background is a vital factor that is considered when breeding new strains with faster growth rates. Rapid growth rates are among the most important and highly desired economic traits for aquaculture species and strongly affect the profitability of fish production. Similar to other vertebrates, fish growth is affected by the hypothalamus-pituitary-liver axis anatomically and by the *growth hormone-insulin growth factor I* (*GH-IGFI*) axis of the neuroendocrine system (Reinecke, 2010). In this axis, *IGFI* is mainly stimulated for synthesis by *GH* and is then expressed in the liver (Norbeck et al., 2007). In the fish body, the liver acts as the main target organ of *GH* and possesses the greatest density of *GH* receptors. Moreover, the liver is a vital organ in the digestive system of teleostei and plays an important role in regulating fish growth rates. It actively participates in the metabolism of fats, carbohydrates and proteins. Together with the skeletal muscle and adipose tissue, the liver is crucial in the regulation of lipid metabolism. At the same time, epigenetic alterations in monozygotic twin pairs were also found, which means that methylation may also play an important role in growth regulation (Pietilainen et al., 2016).

The transcriptome is the readout of the genome of a specific organ or cell at one point in time. In non-model species lacking large-scale genomic sequences, transcriptome sequencing is an effective and quick method for studying gene expression and addressing comparative genomic level questions. Next-generation sequencing (NGS) is much more sensitive and capable of detecting gene expression with a larger dynamic range than microarray-based technologies (Marioni et al., 2008; Mortazavi et al., 2008). The RNA sequencing method (RNA-Seq) has been used in many studies of different fishes, such as zebrafish (Aanes et al., 2011; Vesterlund et al., 2011), silver carp (Fu and He, 2012), and common carp (Zhu et al., 2012). The results of these studies have demonstrated that RNA-Seq is a reliable method for use in comparative transcriptome analyses and provides high-resolution whole-genome expression in fish tissues. But RNA-Seq also had some disadvantages, such as hard to assembly into full-length mRNAs, high price compared with microarray analysis and uneven coverage for transcripts (Wang et al., 2009).

Bighead carp (*Hypophthalmichthys nobilis*) are members of the Cyprinidae family and are widely distributed in Southeast Asia (Nelson, 2006). This species has been cultured for over a thousand years in China, which is the largest producer and consumer of this species in the world. Due to its specific characteristics, such as fast growth, good taste and minimal food requirements, these fish are among the most important aquaculture species in China and around the world (Tong and Sun, 2015). According to a Food and Agriculture Organization (FAO) report, the worldwide production of bighead carp exceeded 2.59 million tons in 2010 (FAO, 2014). Much effort has been spent toward breeding new strains of bighead carp with more characteristics

* Corresponding author.
*E-mail addresses:* fubeide@ihb.ac.cn (B. Fu), xinhuawang123@163.com (X. Wang), fengxiu@ihb.ac.cn (X. Feng), xmyu@ihb.ac.cn (X. Yu), jgtong@ihb.ac.cn (J. Tong).

http://dx.doi.org/10.1016/j.cbd.2016.08.006
1744-117X/

that are suitable for aquaculture (Zhu et al., 2015). However, the lack of genomic and transcriptomic information regarding bighead carp has made studies of this species' genetics and breeding difficult tasks. Moreover, due to their planktonic diet and effective filtering capabilities, bighead carp were introduced to North America and other countries to combat various algal blooms in aquaculture ponds a few decades ago (Kolar et al., 2007). But they caused a lot of troubles for threatening the native fish species. This strategy has proven effective and is used widely in China and worldwide (Zhang et al., 2008).

Previous studies have identified many differentially expressed genes in the liver transcriptomes of fast- and slow-growing rainbow trout (Tymchuk et al., 2009). However, a similar study has never been conducted in bighead carp or other Cyprinidae fish. In the present study, we sequenced two libraries from the liver of two groups of bighead carp with significantly different growth rates using Illumina RNA-Seq methods. The main goal of this study was to identify the differentially expressed genes and growth-related non-synonymous SNPs in the two groups and, therefore, to decipher the genetic mechanisms that may determine the growth rates of the bighead carp.

## 2. Materials and methods

### 2.1. Ethics statement

All experimental procedures involving the animals in this study were approved by the Committee for Animal Experiments of the Institute of Hydrobiology of the Chinese Academy of Sciences, China, and complied with the Laboratory Animal Management Principles of China. The rearing activities of bighead carp in Jingzhou, Hubei were approved by the owner of the pond.

### 2.2. Fish material and RNA sequencing

The materials used in this study were derived from a full-sib family that was reared in a single pond in Jingzhou of the Hubei Province in China. The fish were one and a half years old when they were sacrificed. Because there is no method for identifying the sex of bighead fish, the sexes of the samples used in this work were unknown. A total of five traits associated with growth were measured and are illustrated in S1 Table. We selected the top six samples with the largest body weight as F (Fast) group and another six samples with the smallest body weight as S (Slow) group samples. The phenotypic mean comparisons between groups were performed using R software. To obtain high-quality transcriptomes, the RNA from the liver tissues from each sample was harvested using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Total RNA samples were extracted in accordance with the protocol of the manufacturer of the Promega Z3100 (Promega, Madison, WI) and were treated with RNase-free DNase I (NEB, UK) for 30 min at 37 °C to remove the residual DNA. After quality checks using electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo, USA), the RNA samples from the different individuals were mixed at equivalent concentrations.

Two paired-end sequencing libraries with insert sizes of approximately 200 bp were constructed and sequenced by MajorBio (Shanghai, China) on an Illumina HiSeq 2000 platform. The short read data were deposited in the NCBI's Short Read Archive at PRJNA305829. Because the read qualities strongly influence the qualities of de novo transcriptome assemblies, we filtered the reads before the next step of the bioinformatics analysis. The reads that contained the adaptor sequence and those with >5% unknown nucleotides were filtered with Perl script. Finally, we removed the reads of <50 bp with FASTX 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/).

### 2.3. Analysis of transcriptome sequencing reads

All of the reads from the two groups that passed the quality checks were pooled together and assembled de novo with Trinity ver.20131110 (Grabherr et al., 2011) with "–seqType fa –JM 20G –CPU 6" and a minimum contig length of 200 bp. Then, we used cd-hit v4.6.1 with "cd-hit-est –t 10 -c 0.95 -n 8" to remove redundancy in the assembly (Li and Godzik, 2006). All of the assembled transcripts were searched against the NCBI Nr database (data 2012.09), and the associated entries for GO with cut-off E-values of 1E − 5 were used for annotation. The Blast results were further parsed with Blast2Go (Conesa et al., 2005) to assign gene ontology to the mapped transcripts. To clarify which pathways the assembled sequences were involved with, we mapped all of the transcriptomes to the KEGG database (http://www.genome.jp/kegg) using BLASTx v2.2.26 (E-value threshold of 1E − 5). For the enrichment analysis of DGEs, we used Fisher's exact test to test whether they were enriched in a specific pathway. Then we corrected the *p*-value with Bonferroni correction. Additionally, we assigned the transcripts to 25 clusters of Orthologous groups (COGs) in the COG database (http://www.ncbi.nlm.nih.gov/COG) after all of the transcripts were mapped to the database.

### 2.4. Gene expression profiles

To identify the genes that were differentially expressed between the two groups of bighead carp, we first used the RSEM package v1.2.15 (Li and Dewey, 2011) to map the reads to the assembled transcripts. RSEM quantified the expression of each gene in terms of transcripts per million (TPMs). Then, the EdgeR package (Robinson et al., 2010) within Trinity was used to identify the differentially expressed genes between the two groups using empirical Bayes methods that permitted the estimations of the gene-specific biological variations. The false discovery rate (FDR) was used to determine the threshold *p*-value for multiple tests, which can offer better results without too many false-positives. We used an FDR < 0.01 as the threshold to define the significantly differentially expressed genes (Benjamini and Yekutieli, 2001). Transcripts with positive or negative log-fold change (log FC) values were considered to be significantly up- and down-regulated genes, respectively. Twelve randomly selected transcripts were validated by RT-PCR in another group of 6-month-old bighead carp that exhibited a significant difference in growth. The slow-growing samples used had body weights of 0.294 kg, 0.315 kg and 0.200 kg. The fast-growing samples used here had body weights of 0.857 kg, 0.759 kg and 0.826 kg. The six bighead samples used here were from a single brood and were fed in the same pond (Supplementary Table 3 contains details regarding the qPCR, including information regarding the primers).

### 2.5. Microsatellite marker discovery

In the present study, we used the MISA microsatellite marker discovery program to identify and localize the microsatellite motifs (http://pgrc.ipkgatersleben.de/misa/). We searched for six types of SSRs, from mononucleotides to hexa-nucleotides, with the following criteria: at least 10 repeats for mononucleotides, 6 repeats for di-nucleotides, and 5 repeats for tri-nucleotides, tetra-nucleotides, penta-nucleotides and hexa-nucleotides. Both perfect SSRs (i.e., SSRs containing a single repeat, such as 'GAT') and compound SSRs (i.e., those composed of two or more repeats separated by 100 bp) were identified.

### 2.6. SNP identification

SNP identification was performed within and between the two groups with the following stringency criteria: at least 3 reads calling the variant, and >20% of the reads calling the variant, while the nucleotide quality was >20 (Salem et al., 2012). The group-specific SNPs were taken as those with allelic imbalance scores (i.e., the ratio between the allelic frequencies of the F group and those of the S group) >5.0 as an amplification and <0.2 as a loss of heterozygosity (Salem et al., 2012). Finally, we used a Perl script to fetch the SNPs within the coding regions and the non-synonymous SNPs.

### 2.7. Statistical analysis

Data in this study was plotted with SPSS v17 for windows and represented as mean ± standard error. The false discovery rate (FDR) was used to determine the threshold *p*-value for multiple tests, which can offer better results without too many false-positives. We used an FDR < 0.01 as the threshold to define the significantly differentially expressed genes.

## 3. Results and discussion

### 3.1. Phenotypic variations between the extreme groups

A full-sib family was obtained by crossing a wild male bighead and a wild female bighead from the Yangtze River, and all experimental fish were raised in the same pond. After one and a half years of feeding, two hundred fish were harvested and verified to be from the same brood mentioned above. Five traits related to the growth of the fish were measured and found to exhibit strong correlations in terms of body weight (S1 Table). According to the body weight, we selected 6 fish with the greatest body weights as the Fast-growing (F) group and 6 fish with the lowest body weights as the Slow-growing (S) group. The average body weight of the F group was 1.67 times that of the S group (S1 Table). Because all of the fish were raised and fed under the same conditions, the variations in the growth rates between the two groups that resulted from the genetic differences in the phenotypic means between the F and S groups were compared, and significant differences in all five of the analyzed traits were observed (Table 1).

### 3.2. De novo assembly and evaluation of the bighead carp transcriptome

We obtained 17.17 million 81-bp pair-end reads for the F group and 14.80 million for the S group. After cleaning the low-quality reads, we used a modified version of a previously published procedure to assemble the reads for a non-redundant consensus (Surget-Groba and Montoya-Burgos, 2010; Fu and He, 2012). The assembly workflow is depicted in detail in the Materials and methods section. The short read data were deposited in NCBI's Short Read Archive at PRJNA305829. Finally, 38,647 transcripts ranging from 200 to 20,387 bp were collected with N25, N50, N75 was 2479 bp, 1363 bp and 575 bp, respectively. The average length of the assembly was 793 bp, and there were 9001 transcripts larger than 1000 bp. The longest transcript of the bighead carp was the ortholog of the *dst* gene, which was also among the longest transcripts of the zebrafish. The length distribution of all of the transcripts is illustrated in Fig. 1.

Given that no general criterion has yet been accepted as standard for quality evaluations of transcriptome assemblies, we used two substantial factors to judge how well the assembled sequences represented the actual transcriptome: (1) coverage of the known expressed genes and (2) transcript completeness. The coverage of the known expressed genes was judged via comparisons with the sequences that are available for bighead carp. All 13 mitochondrial protein-coding genes and 178 of the 192 (92.7%) mRNAs in NCBI database (Jul 2014) were present in our assembled transcripts. We also searched the UniProt database (Sep 2010) to determine the coverage of the known expressed genes with

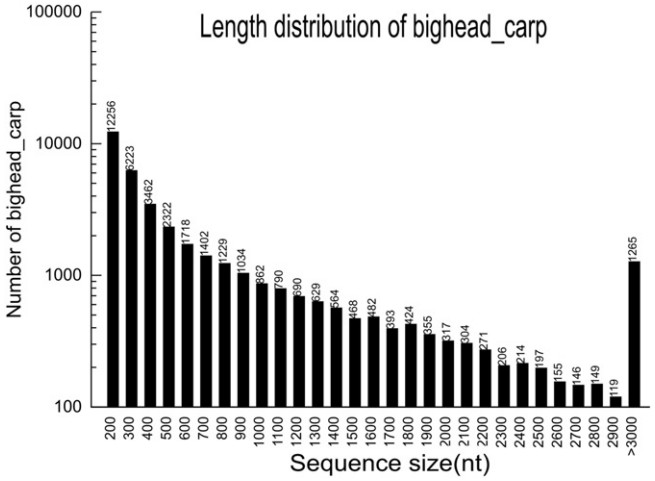


**Fig. 1.** The length distribution of bighead carp liver transcriptome assembled in this study.

E-values of 1E−5. The results indicated that 16,784 transcripts exhibited significant similarities to known proteins in the UniProt database. The completeness levels of the assembled transcripts of the bighead carp were then assessed by comparing the mitochondrial genes assembled from the NGS reads and those based on Sanger sequencing in GenBank (NC_010194). A total of 11,380 bp of 11,423 bp (99.62%) were identical in the two gene sets, which suggested very good transcriptome sequence quality. The remaining 43 bp that differed from each other might have resulted from high levels of intra-specific genetic diversity.

Bighead carp and silver carp (*Hypophthalmichthys molitrix*) are in the same sub-family of Cyprinidae. When compared with the silver carp transcriptome (Fu and He, 2012), the bighead transcriptome in the present study exhibited fewer but longer transcripts. The reason for this difference might be that the silver carp transcriptome was assembled from a mixed library of multiple organs that included more organ-specific transcripts than the liver transcripts in this study. Another reason might be due to the sequencing depth. The sequencing depth in this study was approximately twice that for the silver carp transcriptome, which provided better contiguity for the assembly (Fullwood et al., 2009).

### 3.3. Sequence annotation

We used several complementary methods to annotate the assembled transcripts in this study. First, the transcripts were searched against the NCBI Nr (non-redundant) protein database using BLASTx with an E-value of 1E−5. Of the 38,647 assembled transcripts, 19,582 exhibited significant results. If we searched against zebrafish proteome, we found 19,109 transcripts got significant results, and they covered 74.53% of all genes in zebrafish genome (Zv10). Subsequently, the transcripts with significant matches were assigned into three main Gene Ontology (GO) categories, i.e., biological processes, cellular components, and molecular functions with 13,636, 10,851 and 6646 transcripts, respectively (Fig. 2). In the biological processes category, cellular and metabolic process terms were dominant. We also observed that 58 transcripts were assigned to the growth class. These transcripts were searched against the zebrafish proteome and mapped to 43 zebrafish genes that included four insulin-like growth factor-binding proteins (i.e., *Igfbp1a*, *Igfbp2a*, *Igfbp5b* and *Igfbp7*) and two insulin receptor genes (i.e., *Insra* and *Insrb*). In the molecular function category, binding and catalytic activities were dominant. We also observed that 141 transcripts were assigned to the enzyme regulator activity class and could be mapped to 98 zebrafish genes.

**Table 1**
Mean comparison ± standard deviation between F and S groups for all the 5 traits.

| Trait | Correlation with WT | Mean high | Mean low | Significance |
|---|---|---|---|---|
| TL (cm) | 0.98 | 54.38 ± 1.39 | 47.23 ± 1.08 | 1.74E−06 |
| BL (cm) | 0.98 | 46.91 ± 1.45 | 40.46 ± 0.63 | 1.64E−06 |
| BH (cm) | 0.91 | 13.38 ± 0.43 | 11.08 ± 0.86 | 1.69E−04 |
| HL (cm) | 0.97 | 15.42 ± 0.27 | 12.91 ± 0.47 | 5.61E−07 |
| WT (kg) | 1.00 | 2.17 ± 0.13 | 1.29 ± 0.09 | 1.09E−07 |

TL: total length; BL; body length; BH body height; HL; head length; WT; weight.

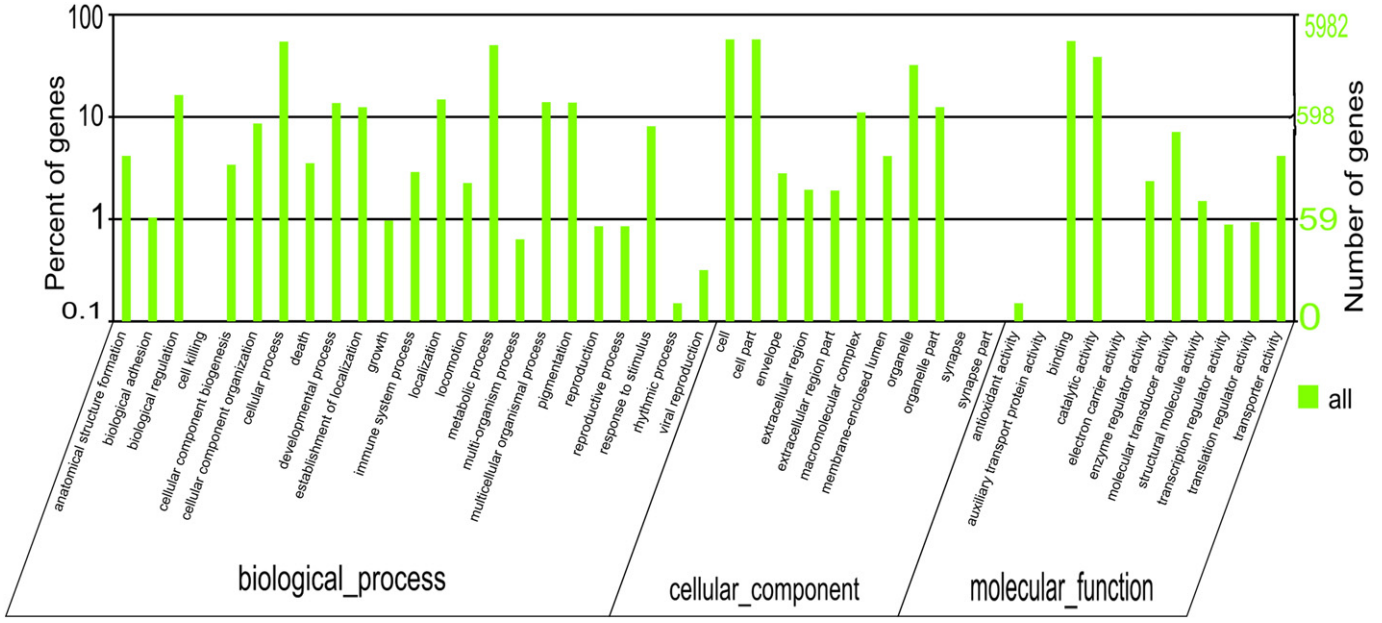


**Fig. 2.** Distribution of Gene Ontology terms for the bighead carp transcriptome, supported by Blast2GO. Transcripts were functionally mapped to GO terms and annotated by setting the following parameters: E-value: 1E−5; Annotation cut-off: 55; Hsp-Hit Coverage cut-off: 0.6.

Second, the annotation of the assembled transcripts using the Clusters of Orthologous Groups (COG) database yielded good results for the 6554 putative proteins (Fig. 3). These COG-annotated putative proteins were functionally assigned to 25 molecular families that included cellular structure, signal transduction and biochemical metabolism, in accordance with the categories observed in the GO annotations.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that aims to increase the understanding of the high-level functions and utilities of biological processes. We observed a total of 14,793 assembled

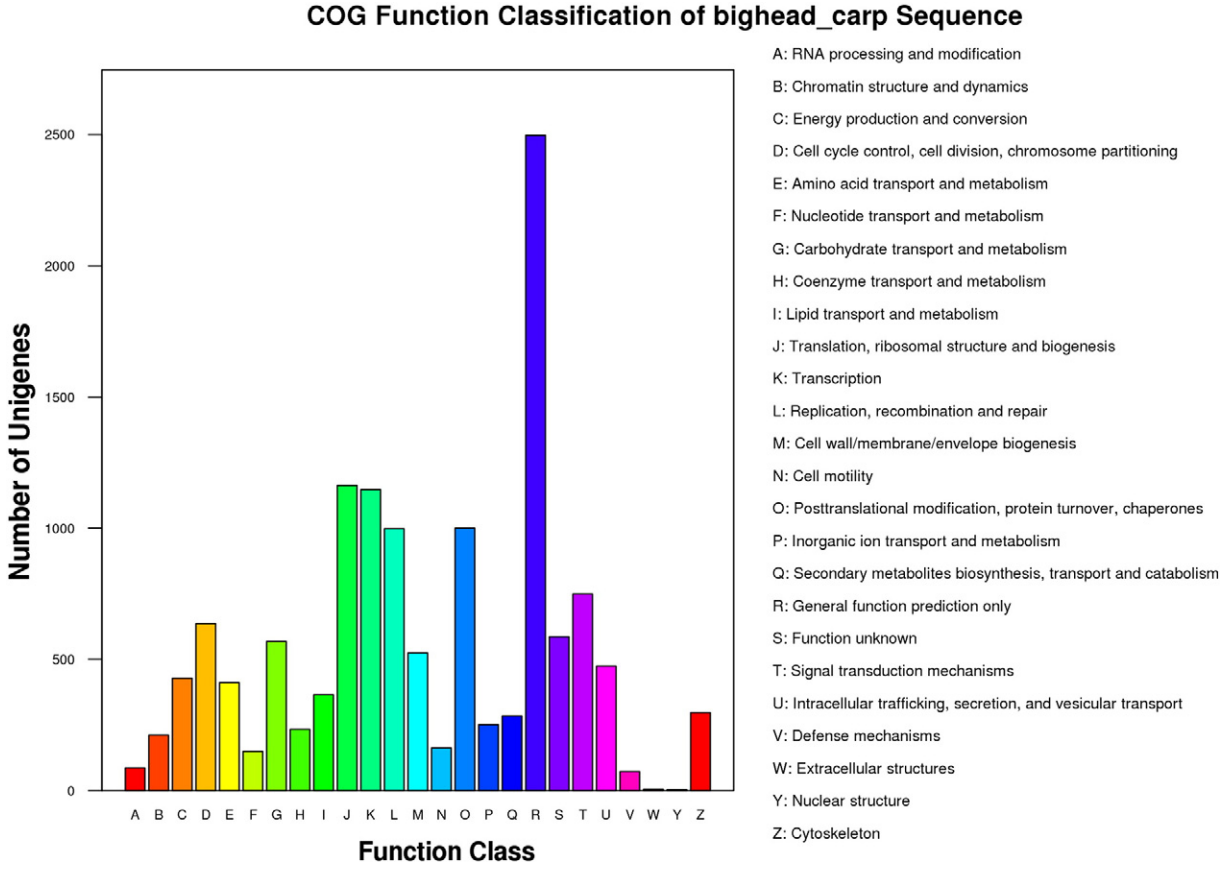


**Fig. 3.** COG annotations of putative bighead carp proteins. All putative proteins translated from the liver transcriptome were aligned to COG database and classified functionally into at least 25 molecular families.

sequences that were involved in 218 predicted KEGG metabolic pathways. The 20 pathways with the greatest numbers of transcripts are shown in S2 Table. The best of which was metabolism.

### 3.4. Differential gene expression in the two groups

After assembling the transcriptomes of the bighead carp liver, we mapped the sequencing reads of the transcriptomes to evaluate the differential gene expression between the two groups. The gene expression profiles of the two groups were extracted using the RSEM package (Li and Dewey, 2011) and Bowtie2 (Langmead and Salzberg, 2012), and the results are provided as the normalized numbers of fragments per kilobase of exon per million reads (FPKMs). We found that 82.20% and 81.59% of filtered reads were mapped to the assembled transcripts for F group and S group, respectively. Accounting only for those genes with FPKMs greater than zero, we observed that 29,682 and 28,719 genes were expressed in the F and S groups, respectively, and there were 25,994 genes that were common to both groups. The correlation of the gene expression levels between the two groups was very high ($r = 0.96$), which suggests that the main portion of the liver transcriptome was conserved. To determine which of the 38,647 genes were differentially expressed between the two groups, we filtered the results with EdgeR (Robinson et al., 2010) and found that 410 genes matched our criterion (FDR < 0.01). We compared the F and S groups and observed that 166 genes were up-regulated and 244 genes were down-regulated. KEGG pathway analyses of the 410 differentially expressed genes indicated that many of the genes were enriched in the peroxisome proliferator-activated receptor (PPAR) signaling pathway, which is involved in lipid metabolism (Zardi et al., 2013) (Table 2). Genes involved in fatty acid oxidation (i.e., *Glpk, Ubc, Cpt1,* and *Pgar*) were up-regulated in the F group compared with the S group. In contrast, the genes that play important roles in fatty acid transport (*Fatp1*) and transcription factor regulation of the adipocytokine signal pathway (*PPARA*) were down-regulated in the F group compared with the S group. Our result is different to the results found in faster growing domesticated and slow growing wild rainbow trout, in which the genes involved in transport in domesticated fish were up-regulated compared with wild ones (Tymchuk et al., 2009). We thought that samples used in that study might be responsible for these phenomena. For example, the samples used in this study were at the same time after hatch, but rainbow trout mentioned above was in almost the same size while wild rainbow trout is older than domesticated one.

Among the 166 up-regulated genes, we observed that many were involved in the regulation of growth rate. For example, the insulin-like growth factor-binding protein 1 (*Igfbp1*) gene binds both *IGFI* and *II* in the plasma, prolongs the half-lives of IGFs and alters their interactions with cell surface receptors (Wood et al., 2005). The protein product thyroid hormone-inducible hepatic protein (*THRSP*) is similar to the gene product of S14, which is expressed in the liver and adipocytes, particularly in lipomatous modules (Grillasca et al., 1997). These genes have roles in fatty acid metabolism processes and may be considered candidate genes for marker-assisted selection (MAS) in bighead carp.

The liver plays a major role in glucose metabolism and glycogen storage. In this study, we observed many differentially expressed genes that are involved in the manipulation of the glucose level in the blood, such as *Gcgra*, which is a member of the class B G protein-coupled receptor (GPCR) family (Jelinek et al., 1993) that mediates the activity of glucagon (Unger and Cherrington, 2012). We observed that the expression level of *Gcgra* was lower in the F group than in the S group, which might indicate that the glucose level in the F group was greater than that in the S group.

To validate these differentially expressed genes, we used the qPCR method to test another group of bighead carp with a different growth rate. We randomly selected twelve genes out of the 410 genes and verified their expression levels with three samples from each group (Fig. 4 and S3 and S4 Table). We observed that more than half of the selected genes exhibited similar expression patterns between the RNA-Seq and qPCR methods. These findings provide direct evidence supporting our findings that the genes were truly differentially expressed between the F and S groups of the bighead carp. The discordant cases might have occurred because different samples were used in the validation. An additional reason for the discordance might be because the gene expression atlas of all of the genes in the liver was altered in the different growth periods.

### 3.5. Identification of the group-specific non-synonymous SNPs

Reads from the two groups were mapped to the previously assembled transcripts and used to identify the SNPs in each group. According to the Trinity results, the nucleic acids with more supported reads were chosen as the references for the final assembly (Grabherr et al., 2011). After filtering with a stringent criterion, 74,663 and 110,058 SNPs were identified in F and S groups, respectively. In combination with the protein-coding annotation, the SNPs were divided into coding SNPs (cSNPs) and non-coding SNPs (ncSNPs). We identified 29,716 cSNPs from the 5412 transcripts in the F group and 31,330 cSNPs from the 5282 transcripts in the S group. As non-synonymous mutations can alter the protein sequence and influence the function of the protein, we further filtered the cSNPs based on the genetic code information regarding the locations of the SNPs within the transcripts. Finally, we

**Table 2**
Enriched pathways in 410 DEGs.

| Num[a] | Gene num in 410 DEG[b] | Gene num in zebrafish[c] | Pathway number | Pathway | q-Value |
|---|---|---|---|---|---|
| 1 | 9 | 31 | 00860 | Porphyrin and chlorophyll metabolism | 5.46E−7 |
| 2 | 15 | 130 | 04630 | Jak-STAT signaling pathway | 2.02E−6 |
| 3 | 8 | 33 | 00020 | Citrate cycle (TCA cycle) | 6.23E−6 |
| 4 | 3 | 4 | 00740 | Riboflavin metabolism | 7.35E−5 |
| 5 | 8 | 68 | 03320 | PPAR signaling pathway | 1.09E−3 |
| 6 | 5 | 27 | 00601 | Glycosphingolipid biosynthesis — lacto and neolacto series | 2.68E−3 |
| 7 | 44 | 1274 | 01100 | Metabolic pathways | 2.37E−3 |
| 8 | 8 | 86 | 04920 | Adipocytokine signaling pathway | 3.61E−3 |
| 9 | 5 | 43 | 00620 | Pyruvate metabolism | 1.64E−2 |
| 10 | 6 | 64 | 00310 | Lysine degradation | 1.58E−2 |
| 11 | 4 | 29 | 04140 | Regulation of autophagy | 2.13E−2 |
| 12 | 3 | 15 | 00450 | Selenocompound metabolism | 2.37E−2 |
| 13 | 3 | 18 | 00053 | Ascorbate and aldarate metabolism | 3.75E−2 |
| 14 | 4 | 36 | 02010 | ABC transporters | 3.73E−2 |
| 15 | 4 | 39 | 00980 | Metabolism of xenobiotics by cytochrome P450 | 4.65E−2 |

[a] The number ranked in sorted order about the enriched pathways of DGEs.
[b] Gene number of 410 DEGs is within this KEGG pathway.
[c] Gene number of zebrafish genes is within this pathway.

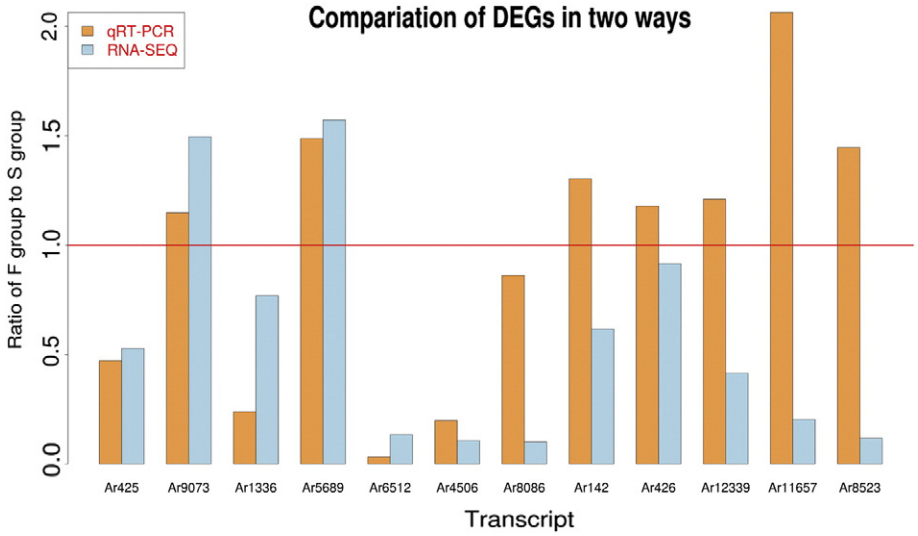


**Fig. 4.** The expression validation of 12 randomly selected genes with qRT-PCR methods. The X axis is the name of each transcript. The Y axis is the ratio of expression level comparison of F group to S group. DEGs: differentially expressed genes.

identified 13,705 and 15,920 non-synonymous SNPs in the F and S groups, respectively. Among these SNPs, 5240 SNPs of the 1836 transcripts were specific to the F group, and 7455 SNPs within the 1650 transcripts were specific to the S group. GO enrichment analyses were performed for the 824 transcripts with fewer than 2 non-synonymous SNPs in the F group, and the results indicated that biological processes (GO: 0008150), regulation of transcription (GO: 0006355) and signal transduction (GO: 0007165) represented the three largest categories (S5 Table). We also found 4 genes that were involved in the lipid metabolism process (GO: 0006629), i.e., *asah1a*, *ptgdsa*, *gba2*, and *srebf2*.

As co-dominantly inherited and easy to scale up automated genotyping markers, SNPs are suitable for mapping quantitative trait loci (QTLs) associated with important traits (Wang et al., 2008). NGS has enabled SNP discovery with unprecedented accuracy and low cost (Canovas et al., 2010). In the present study, we observed 74,663 and 110,058 SNPs in the F and S groups, respectively. Identifying SNPs from pooled samples with different traits is a convenient method of connecting phenotypic and genotypic information. The SNPs that were unevenly distributed in the two groups might have resulted from the allelic imbalances between the groups (Salem et al., 2012). Therefore, these SNPs could potentially be used as candidate markers in the MAS of bighead carp.

### *3.6. Identification of microsatellites*

Microsatellite markers have been widely used in genetic map construction and QTL analyses in aquaculture (Zhang et al., 2010; Laghari et al., 2013; Laine et al., 2013; Zhu et al., 2015). In bighead carp, few genetic studies have employed microsatellites in genetic map construction (Liao et al., 2007). We obtained 8342 simple sequence repeat (SSR) markers in 6416 transcripts with MISA (Thiel et al., 2003). Mononucleotide repeats were the most abundant (5194, 62.26%), followed by di-nucleotide repeats (2078, 24.91%) and tri-nucleotide repeats (936, 11.22%). Repeats of more than three nucleotides occurred at rates below 1.5%. The SSR markers were divided into perfect markers that had only a single repeat, such as 'GTT' and compound SSR markers that were composed of two or more SSR markers that were separated by fewer than 100 bp. We found a total of 637 (7.6%) compound SSR markers. After excluding the mono-nucleotide SSR markers, the SSR frequency was calculated. Among the di-nucleotide repeat motifs, 64.8% were AC/GT repeats. Among the tri-nucleotide repeat motifs, ATC/GAT was the most abundant at 26.9%. To validate these SSRs, we mapped all the filtered reads to the assembled transcripts, and got the depth for all SSRs. The quantile depths for all sites were 0, 29, 75 (median value), 187 and 302,600. At the same time, the mean depth for these SSRs is 495, which we thought was a high-confidence evidence for this study.

It has been reported that the largest proportion of SSRs in the human genome consists of di-nucleotide repeats, which account for approximately 16% of all SSRs in the human genome (Venter et al., 2001). In the present study, 24.91% of the 8342 bighead carp SSRs were di-nucleotide repeats; this rate is much higher than that of the human genome but lower than that of the silver carp genome (Fu and He, 2012). The most common di-nucleotide repeats in the bighead carp genome were AC/GT repeats, which differed from the silver carp genome, in which AC and AT were the most common.

## 4. Conclusion

The present transcriptome analyses of the F and S groups of bighead carp liver samples provided the first large-scale sequencing data for this important aquaculture species. In this study, we want to uncover the growth physiology differences between the two groups in terms of the gene expression level and the level of the gene sequences themselves. First, we observed 410 differentially expressed genes between the F and S groups. Second, we identified 13,705 and 15,920 non-synonymous SNPs in the F and S groups. The differences were mainly enriched in the metabolism of lipid in liver and they provided the genetic background regarding the regulation of the growth rates of bighead carp and can be further used in investigating the impact of lipid synthesis and transportation to the growth physiology.

## Disclosure statement

Conflicts of interest: none.

## Acknowledgments

This work was supported by the Knowledge Innovation Program of the Chinese Academy of Science and the National Natural Science Foundation of China (grant no. 31472268), and Open Funding of State Key Laboratory of Freshwater Ecology and Biotechnology of Institute of Hydrobiology, Chinese Academy of Sciences (2011FBZ20). We also thank Wuhan Branch, Supercomputing Center of the Chinese Academy of Sciences for support of high performance cluster.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cbd.2016.08.006.

## References


Aanes, H., Winata, C.L., et al., 2011. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res. 21 (8), 1328–1338.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 1165–1188.

Canovas, A., Rincon, G., et al., 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm. Genome.

Conesa, A., Gotz, S., et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18), 3674–3676.

FAO, 2014. Yearbook of Fishery and Aquaculture Statistics 2013. Fao Inter-Departmental Working Group, Rome, Italy.

Fu, B., He, S., 2012. Transcriptome analysis of silver carp (*Hypophthalmichthys molitrix*) by paired-end RNA sequencing. DNA Res. 19, 131–142.

Fullwood, M.J., Wei, C.L., et al., 2009. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res. 19 (4), 521–532.

Grabherr, M.G., Haas, B.J., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652.

Grillasca, J.P., Gastaldi, M., et al., 1997. Cloning and initial characterization of human and mouse Spot 14 genes. FEBS Lett. 401 (1), 38–42.

Jelinek, L.J., Lok, S., et al., 1993. Expression cloning and signaling properties of the rat glucagon receptor. Science 259 (5101), 1614–1616.

Kolar, C.S.C.D., Courtenay Jr., W.R., Housel, C.M., Williams, J.D., Jennings, D.P., 2007. Bighead Carps: A Biological Risk Assessment. American Fisheries Society, Bethesda, Maryland.

Laghari, M., Zhang, Y., et al., 2013. Quantitative trait loci (QTL) associated with growth rate trait in common carp (*Cyprinus carpio*). Aquac. Int. 21 (6), 1373–1379.

Laine, V.N., Shikano, T., et al., 2013. Quantitative trait loci for growth and body size in the nine-spined stickleback *Pungitius pungitius* L. Mol. Ecol. 22 (23), 5861–5876.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 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 (13), 1658–1659.

Liao, M., Zhang, L., et al., 2007. Development of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) genetic maps using microsatellite and AFLP markers and a pseudo-testcross strategy. Anim. Genet. 38 (4), 364–370.

Marioni, J.C., Mason, C.E., et al., 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18 (9), 1509–1517.

Mortazavi, A., Williams, B.A., et al., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5 (7), 621–628.

Nelson, J.S., 2006. Fishes of the World. John Wiley, Hoboken, N.J.

Norbeck, L.A., Kittilson, J.D., et al., 2007. Resolving the growth-promoting and metabolic effects of growth hormone: differential regulation of GH-IGF-I system components. Gen. Comp. Endocrinol. 151 (3), 332–341.

Pietilainen, K.H., Ismail, K., et al., 2016. DNA methylation and gene expression patterns in adipose tissue differ significantly within young adult monozygotic BMI-discordant twin pairs. Int. J. Obes. 40 (4), 654–661.

Reinecke, M., 2010. Influences of the environment on the endocrine and paracrine fish growth hormone-insulin-like growth factor-I system. J. Fish Biol. 76 (6), 1233–1254.

Robinson, M.D., McCarthy, D.J., et al., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1), 139–140.

Salem, M., Vallejo, R.L., et al., 2012. RNA-Seq identifies SNP markers for growth traits in rainbow trout. PLoS One 7 (5), e36264.

Surget-Groba, Y., Montoya-Burgos, J.I., 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. Genome Res. 20 (10), 1432–1440.

Thiel, T., Michalek, W., et al., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106 (3), 411–422.

Tong, J., Sun, X., 2015. Genetic and genomic analyses for economically important traits and their applications in molecular breeding of cultured fish. Sci. China Life Sci. 58 (2), 178–186.

Tymchuk, W., Sakhrani, D., et al., 2009. Domestication causes large-scale effects on gene expression in rainbow trout: analysis of muscle, liver and brain transcriptomes. Gen. Comp. Endocrinol. 164 (2–3), 175–183.

Unger, R.H., Cherrington, A.D., 2012. Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. J. Clin. Invest. 122 (1), 4–12.

Venter, J.C., Adams, M.D., et al., 2001. The sequence of the human genome. Science 291 (5507), 1304–1351.

Vesterlund, L., Jiao, H., et al., 2011. The zebrafish transcriptome during early development. BMC Dev. Biol. 11 (1), 30.

Wang, S., Sha, Z., et al., 2008. Quality assessment parameters for EST-derived SNPs from catfish. BMC Genomics 9, 450.

Wang, Z., Gerstein, M., et al., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Wood, A.W., Duan, C., et al., 2005. Insulin-like growth factor signaling in fish. Int. Rev. Cytol. 243, 215–285.

Zardi, E.M., Navarini, L., et al., 2013. Hepatic PPARs: their role in liver physiology, fibrosis and treatment. Curr. Med. Chem. 20 (27), 3370–3396.

Zhang, X., Xie, P., et al., 2008. A review of nontraditional biomanipulation. ScientificWorldJournal 8, 1184–1196.

Zhang, L., Yang, G., et al., 2010. Construction of a genetic linkage map for silver carp (*Hypophthalmichthys molitrix*). Anim. Genet. 41 (5), 523–530.

Zhu, Y.P., Xue, W., et al., 2012. Identification of common carp (*Cyprinus carpio*) microRNAs and microRNA-related SNPs. BMC Genomics 13, 413.

Zhu, C., Tong, J., et al., 2015. Comparative mapping for bighead carp (*Aristichthys nobilis*) against model and non-model fishes provides insights into the genomic evolution of cyprinids. Mol. Gen. Genomics. 290 (4), 1313–1326.