



OPEN

Comparative transcriptome profiles of four sexually size dimorphic fish

DATA DESCRIPTOR

Li-Fei Luo^{1,2}✉, Zi-Sheng Xu^{1,2}, Dan-Yang Li^{1,2}, Zhen Hu³ & Ze-Xia Gao^{1,2,4}✉

Sexual size dimorphism is widespread in fish species. Although sex growth differences in multiple species have been studied successively, the commonalities of regulatory mechanisms across sexually dimorphic species are unknown. In this study, we performed RNA-seq analysis of four representative fish (loach, half-smooth tongue sole, yellow catfish, and Nile tilapia) with significant growth differences between females and males. Clean reads were identified from four fish species, ranging from 45,718,052 to 57,733,120. Following comparison transcriptome analysis, there were 1,132 and 1,108, 1,290 and 1,102, 4,732 and 4,266, 748 and 192 differentially expressed genes (DEGs) in the brain and muscle of loach, half-smooth tongue sole, yellow catfish, and Nile tilapia, respectively. Furthermore, the expression levels were validated by quantitative real-time PCR (qRT-PCR). Comparative transcriptome profiles of four fish described here will provide fundamental information for further studies on the commonalities of sexually size dimorphic fish in regulating growth differences between females and males.

Background & Summary

Many fish species display sexual dimorphism, the most common of which is sexual size dimorphism, wherein one sex is larger than the other, but exhibits species specificity. For example, common carp (*Cyprinus carpio*), rainbow trout (*Oncorhynchus mykiss*), Japanese flyounder (*Paralichthys olivaceus*), and half-smooth tongue sole (*Cynoglossus semilaevis*) show the most extreme sexual size dimorphism with females being larger and growing faster than males^{1–3}. In contrast, in some species, such as Nile tilapia (*Oreochromis niloticus*), yellow catfish (*Pelteobagrus fulvidraco*), and channel catfish (*Ictalurus punctatus*), the growth rate and body size of males is faster and larger than females^{4–6}. The growth of vertebrates is regulated by growth hormones/insulin-like growth factors secreted by the hypothalamic-pituitary-gonad (HPG) axis and other tissues⁷. So many studies on the sexual size dimorphism in fish mainly have focused on the genes and hormones related to the HPG axis^{8–10}. As we all know, most complex traits are controlled by multiple genes, so studies that focus on one or two genes and hormones related to the HPG axis cannot fully reveal the regulatory mechanism of sexual size dimorphism. Therefore, the molecular mechanism of sexual size dimorphism in fish is still unclear.

The difference in gene expression between sexes is thought to be the key contributor to the manifested phenotypic differences^{11,12}. Therefore, in order to understand the causes of sexual size dimorphism, numerous studies have focused on differentially expressed genes (DEGs) between females and males. Comparative transcriptome analysis can help people find DEGs at the entire genome level, and has been widely applied to study sex-biased genes in animals including fish^{13,14}, swimming crab¹⁵, pig¹⁶, cattle¹⁷, and chicken¹⁸. However, the existing transcriptome mostly focus on expression pattern of one or several specific organs at a single specie level. The degree to which sex-biased expression is conserved across the specie lineage and the extent of conservation in different tissues and organ systems are unknown. Assessing the expression of sex-biased genes across species will contribute to a comprehensive understanding of the molecular mechanisms regulating phenotypic sex differences¹⁹. So far, no studies comprehensively and accurately elucidate how gene expression differs between the sexes in a broad range of sexual size dimorphic fish species and tissues, and the universality of the regulation mechanism of growth differences between females and males in sex-dimorphic fish is still unclear.

¹College of Fisheries, Key Lab of Freshwater Animal Breeding, Ministry of Agriculture/Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Engineering Research Center of Green development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt, Ministry of Education, Huazhong Agricultural University, Wuhan, 430070, China. ²Hubei Hongshan Laboratory, Wuhan, 430070, China. ³Hubei Aquatic Products Technology Promotion Station, Wuhan, 430060, China. ⁴Engineering Technology Research Center for Fish Breeding and Culture in Hubei Province, Wuhan, 430070, China. ✉e-mail: luolifei@mail.hzau.edu.cn; gaozx@mail.hzau.edu.cn

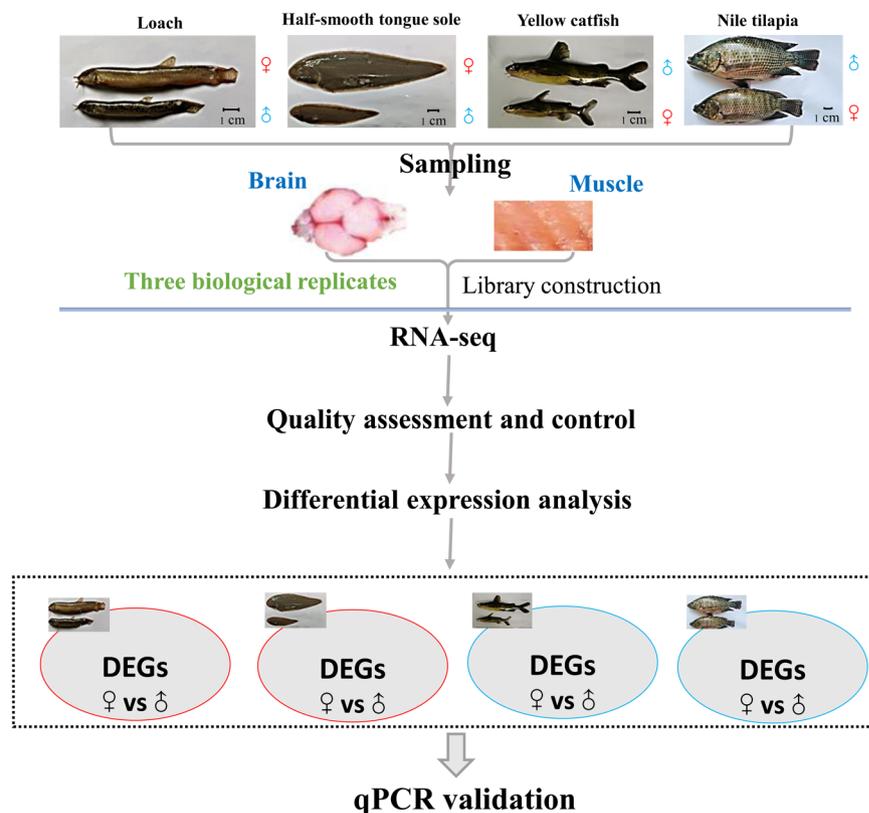


Fig. 1 Flow chart of comparative transcriptome analysis.

Loach (*Misgurnus anguillicaudatus*), half-smooth tongue sole (*Cynoglossus semilaevis*), yellow catfish (*Pelteobagrus fulvidraco*), and Nile tilapia (*Oreochromis niloticus*) are all important economic fish in aquaculture^{20–23}. However, they all have sexually dimorphic growth patterns with significant growth differences between female and male individuals. *M. anguillicaudatus* and *C. semilaevis* have growth advantage in females while *P. fulvidraco* and *O. niloticus* have growth advantage in males, which affects the fish yield and economic value of these cultured species²⁴. It is not clear whether there are some commonalities in the molecular mechanisms regulating the growth differences between females and males in sex-growth dimorphic fish (including fish with both female and male growth dominance). Therefore, exploring the common regulation genes and pathways between females and males of these four representative fish will help us further understand the molecular mechanism of sexual size dimorphism in fish, thus providing an important theoretical basis for breeding fast growing and uniform varieties in aquaculture.

In this study, we performed transcriptomic sequencing on tissues of four representative fish (loach, half-smooth tongue sole, yellow catfish, and Nile tilapia) during significant growth differences between females and males by RNA-seq, including 2 tissues (brain and muscle), and 48 libraries (three biological replicates in each sample). Quality control was conducted to evaluate the quality of our transcriptome data using FastQC, and a high-quality dataset is presented. Additionally, we performed comparative transcriptomic analyses of four sexually size dimorphic fish with the aim of identifying the DEGs between females and males in four species. The schematic overview of the study design or workflow is shown in Fig. 1. Our work represents a valuable resource for re-use, and will provide fundamental information for further studies on the commonalities of sexually size dimorphic fish in regulating growth differences between females and males.

Methods

Ethic statement. All experiments were conducted in accordance with the guidelines of the National Institute of Health Guide for the Care and Use of Laboratory Animals and approved by the Research Ethics Committee, Huazhong Agricultural University, Wuhan, China (approval ID: SYXK2015-0084). All surgery was performed under MS-222 (Sigma, Saint Louis, MO, USA; 100 mg/L) anesthesia, and all efforts were made to minimize suffering.

Experimental fish and sample collection. *M. anguillicaudatus* used in the experiment was collected from the artificial breeding population of our laboratory, and *C. semilaevis* was purchased from Qingdao, Shandong Province. *P. fulvidraco* and *O. niloticus* were provided by Taishan Base of Research Center of Haida Group and Wuxi Fisheries College of Nanjing Agricultural University, respectively. Each fish species was selected from a full sibling family. Body weight and length were measured and compared by student's t-test. The average weight of females and males of *M. anguillicaudatus* was 5.42 ± 0.66 g and 2.77 ± 0.36 g, respectively, which indicated that

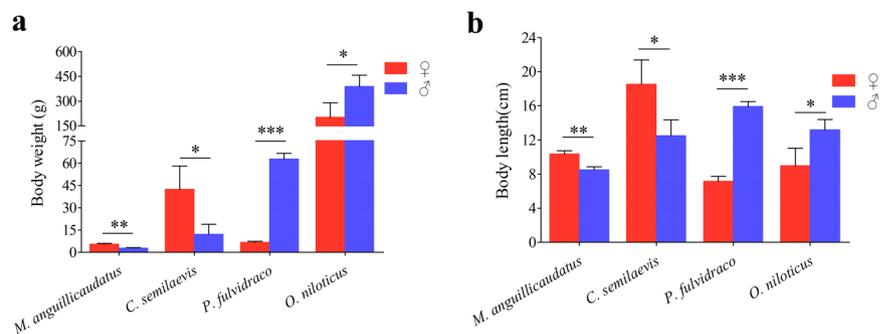


Fig. 2 Comparison of body length and body weight between females and males of four representative fish. T-test was used to calculate significant levels between the indicated groups. * $P < 0.05$, ** $P < 0.005$, and *** $P < 0.0001$.

the weight of females was 95.67% heavier than those of males ($P < 0.005$). In *C. semilaevis*, the average weight of females and males was 42.43 ± 15.65 g and 12.06 ± 6.80 g, respectively, which indicated that the weight of females was 3.5 times heavier than those of males ($P < 0.05$) (Fig. 2a). The comparisons of body length between females and males in these two species also showed that females grew faster than males (Fig. 2b). However, in *P. fulvidraco* and *O. niloticus*, males grew faster than females, especially in *P. fulvidraco*, the difference between females and males was obvious ($P < 0.0001$), with males 9.57 times heavier and 2.23 times longer than females. Nine healthy females and males from each species were selected for transcriptome sampling, and three biological replicates were set up, each containing three fish. After anesthetizing the fish with 100 mg/L MS-222, brain and muscle tissues were quickly dissected, frozen in liquid nitrogen and stored at -80°C for subsequent RNA extraction.

Libraries construction and sequencing. Total RNA was extracted from each sample by trizol with manufacturer's protocol (Ambion). RNA integrity assessment was conducted by Agilent 2100 Bioanalyzer (Agilent Technologies). The procedures of RNA purification and libraries construction were performed by using TruSeq Stranded mRNA LTSample Prep Kit (Illumina) following the manufacturer's instructions. Briefly, high quality mRNA was used to synthesize double stranded cDNA and purified by AMPure XP beads (Beckman Coulter, Beverly, USA). After that, the purified cDNA fragments were repaired on the 3' end and adenylated before being ligated to sequencing adapters. Subsequently, the libraries were purified, enriched by PCR, and purified again to generate the final libraries. At last, the libraries were sequenced on the Illumina sequencing platform (HiSeqTM 2500), generating 150 bp paired-end reads.

Transcriptome assembly and gene expression level analysis. Quality control and reads statistics were determined by Trimmomatic²⁵. At the same time, Q30 and GC contents of the clean reads were calculated, and all the downstream analyses were based on clean reads with high quality. The high-quality clean data of *M. anguillicaudatus* were then de novo assembled using Trinity software with min_kmer_cov set to 2 by default and all other parameters set default²⁶, then the transcriptome assembly was then assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs). And the final clean reads of *C. semilaevis*, *P. fulvidraco*, and *O. niloticus* were mapped to the corresponding reference genome using hisat2²⁷. The expression level of each transcript was calculated by the expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced (FPKM) method²⁸, and the FPKM value and the read counts of each gene was calculated by cufflinks and htseq-count²⁹, respectively.

Differential expression analysis. To identify the differential expression genes (DEGs) in the brain and muscle tissues of each species, DESeq R package with estimate Size Factors and nbinom Test was performed to quantify the expression of two expression profiles³⁰. The unigenes with P -value < 0.05 and $|\log_2(\text{fold-change})| > 1$ were identified as significant DEGs. In the comparisons of brain, there were 1,132, 1,290, 4,732, and 748 DEGs in loach, half-smooth tongue sole, yellow catfish, and Nile tilapia, respectively. In the comparisons of muscle, 1,108, 1,102, 4,266, and 192 DEGs were identified separately in those four species. The information of DEGs of four fish are available on Figshare.

Confirmation by quantitative real-time PCR (qRT-PCR). In order to validate the results of RNA-seq, several DEGs in brain and muscle tissues of four fishes were selected for quantitative real-time polymerase chain reaction (qRT-PCR) analysis. The total RNA of the brain and muscle tissues from both sexes of *M. anguillicaudatus*, *C. semilaevis*, *P. fulvidraco*, and *O. niloticus* was reverse transcribed by using the PrimeScriptRT reagent Kit (Takara) following the manufacturer's protocol. Primers were designed using Primer Premier 5.0 software and were listed in Supplementary Table 1. Each 20 μL reaction volume contained 10 μL $2 \times$ Hieff[®] qPCR SYBR Green Master Mix (Yeasen, Shanghai, China), 1.0 μL diluted cDNA template, 0.8 μL each of sense and reverse primers, and 7.4 μL ddH₂O. The qRT-PCR reaction was performed using the Applied Biosystems QuantStudio 6 Flex Real-time PCR System (Applied Biosystems, Foster City, CA, USA) with the following program: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s and 72°C for 30 s. Five biological replicates were performed in

Species	Sample name	Number of uniquely mapped reads	Percentage of uniquely mapped reads%	Number of multiply mapped reads	Percentage of multiply mapped reads%
<i>C. semilaevis</i>	F_brain1	45,080,563	91.28%	1,275,758	2.58%
	F_brain2	47,268,898	90.83%	1,399,890	2.69%
	F_brain3	43,645,440	90.99%	1,289,586	2.69%
	F_muscle1	41,421,038	86.18%	3,677,788	7.65%
	F_muscle2	47,897,954	86.65%	4,300,193	7.78%
	F_muscle3	45,414,577	87.15%	3,866,463	7.42%
	M_brain1	45,505,320	91.43%	1,254,380	2.52%
	M_brain2	47,031,880	91.50%	1,246,549	2.43%
	M_brain3	47,045,429	91.65%	1,315,726	2.56%
	M_muscle1	49,216,376	87.87%	3,659,373	6.53%
	M_muscle2	39,491,270	87.32%	3,022,780	6.68%
	M_muscle3	48,628,744	87.02%	3,807,886	6.81%
<i>P. fulvidraco</i>	F_brain1	45,394,150	92.63%	1,299,059	2.65%
	F_brain2	47,837,486	92.47%	1,268,419	2.45%
	F_brain3	46,165,503	92.26%	1,268,917	2.54%
	F_muscle1	47,219,749	86.65%	5,460,284	10.02%
	F_muscle2	44,567,138	88.57%	3,987,067	7.92%
	F_muscle3	45,011,306	86.83%	5,024,160	9.69%
	M_brain1	51,683,137	92.64%	1,363,262	2.44%
	M_brain2	50,720,918	92.38%	1,381,930	2.52%
	M_brain3	47,488,856	92.62%	1,272,313	2.48%
	M_muscle1	49,184,986	87.24%	4,942,037	8.77%
	M_muscle2	46,620,994	85.13%	6,241,636	11.40%
	M_muscle3	46,092,038	87.05%	5,237,187	9.89%
<i>O. niloticus</i>	F_brain1	46,561,189	91.50%	1,123,021	2.21%
	F_brain2	44,809,941	92.44%	1,045,914	2.16%
	F_brain3	47,442,736	90.26%	1,281,658	2.44%
	F_muscle1	43,722,771	87.46%	4,793,153	9.59%
	F_muscle2	43,776,082	89.96%	3,877,328	7.97%
	F_muscle3	39,746,755	87.13%	4,296,799	9.42%
	M_brain1	47,446,250	91.92%	1,157,992	2.24%
	M_brain2	46,747,697	90.57%	1,174,645	2.28%
	M_brain3	46,200,304	93.66%	1,145,974	2.32%
	M_muscle1	43,873,255	87.36%	4,070,023	8.10%
	M_muscle2	46,863,229	88.37%	4,204,486	7.93%
	M_muscle3	48,019,110	90.35%	4,005,195	7.54%

Table 1. Statistics analysis of clean reads mapping onto reference genome.

each reaction. The expression of β -actin was served as the reference for internal standardization to normalize the Ct values to conduct the $2^{-\Delta\Delta Ct}$ method³¹.

Data Records

All sequencing data of *M. anguillicaudatus*, *C. semilaevis*, *P. fulvidraco*, and *O. niloticus* were uploaded to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information under accession number SRP313711, SRP313744, SRP314481, and SRP313936^{32–35}, respectively. The information of annotations for uni-genes of *M. anguillicaudatus*, and the information of DEGs of four fish can be found on the Figshare³⁶.

Technical Validation

A similar number of raw reads was obtained for four fish species, ranging from 46,481,320 to 60,661,760 (Supplementary Table 2). After trimming, a total of 45,718,052 to 57,733,120 clean reads remained and the overall mapping efficiency of these reads against the reference genome of the corresponding species ranged from 88.43 to 97.89%, the above results indicated that the quality of the sequencing data was high enough for subsequent analysis. BUSCO analysis of *M. anguillicaudatus* de novo assembled data revealed that a total of 86.32% genes that were completely matched in BUSCO library (3,957 out of 4,584 genes) (Fig. 3a). Among them, 80.27% were those of complete and single-copy BUSCOs (3,677 out of 4,584) and 6.11% were complete

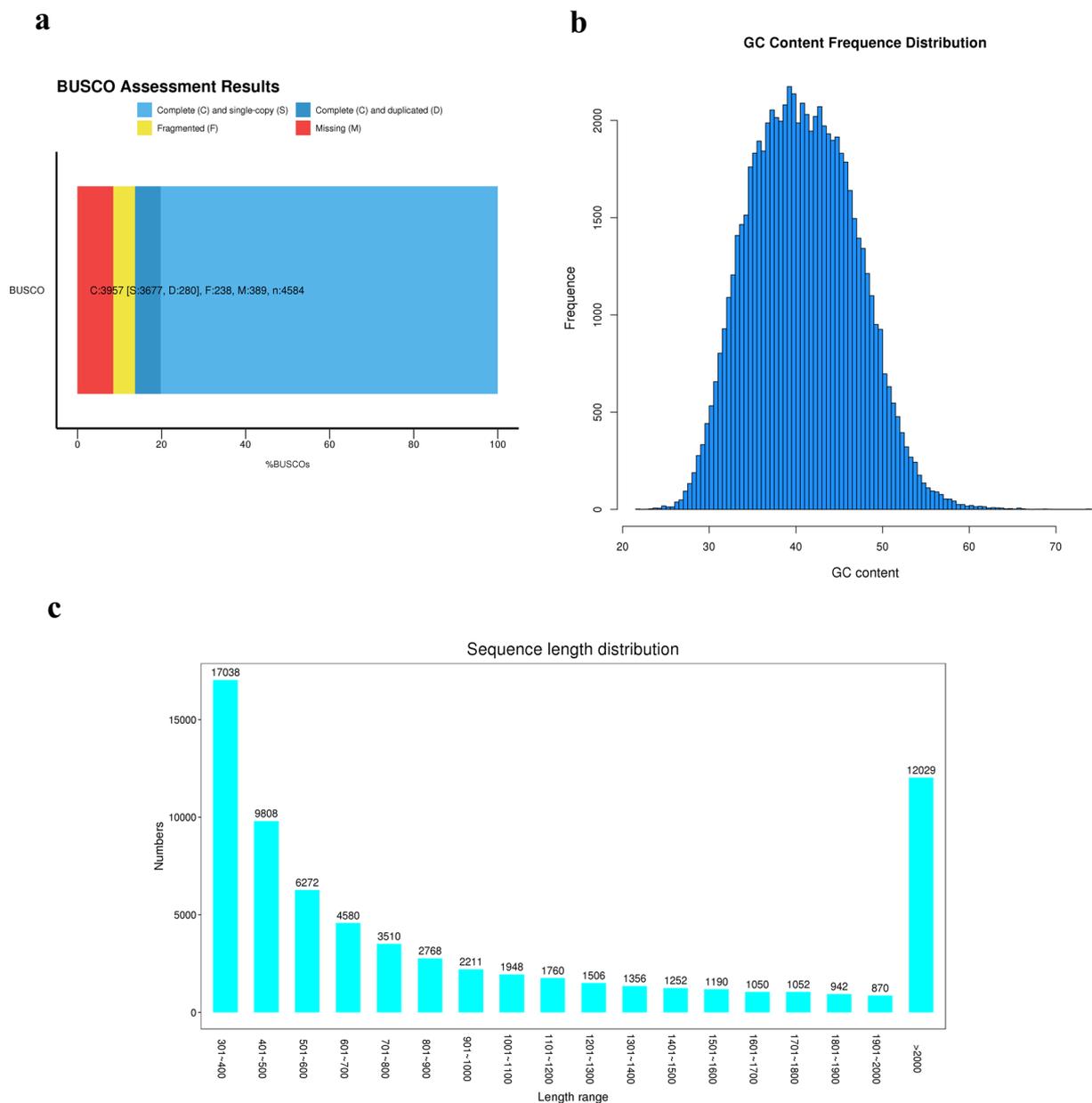


Fig. 3 Quality assessment of *M. anguillicaudatus* de novo assembled data. **(a)** BUSCO assessment results of *M. anguillicaudatus* assembled data. Complete and single-copy BUSCOs (S): the number of genes that are completely matched and match the previous one in BUSCO library; Complete and duplicated BUSCOs (D): number of duplicated and duplicated genes in the BUSCO library; Fragmented BUSCOs (F): number of genes in partially matched BUSCO library; Missing BUSCOs (M): number of unmatched genes in the BUSCO library; Total BUSCO groups searched (n): the chosen BUSCO library of all the number of genes. **(b)** GC content distribution map of unigenes. **(c)** Length distribution of unigenes.

and duplicated BUSCOs (280 out of 4,584). Although the proportion of complete alignments is not as high as that of the previous loach transcript assembly, the proportion of complete and single-copy BUSCOs is higher than the previous 70.6%³⁷. Distribution map of GC content and length of unigenes in *M. anguillicaudatu* were shown in Fig. 3b,c. Collectly, it indicated high quality of loach transcriptome assembly. As for species with reference genomes (half-smooth tongue sole, yellow catfish, and Nile tilapia), after all the clean reads were mapped onto the corresponding reference genome, the number and percentage of uniquely mapped reads and multiply mapped reads was calculated and presented in Table 1. The correlation of gene expression levels between samples is an important index to verify the reliability of an experiment, and the Pearson correlation coefficient (r) with a square value greater than 0.85 was a prerequisite for differential expression analysis (Supplementary Fig. 1 and Fig. 2). Additionally, although the values of the $\log_2(\text{fold-change})$ from the transcriptomic analysis and qRT-PCR analysis were different, the differential expression levels of these selected genes by qRT-PCR were highly consistent with those observed by RNA-seq (Fig. 4).

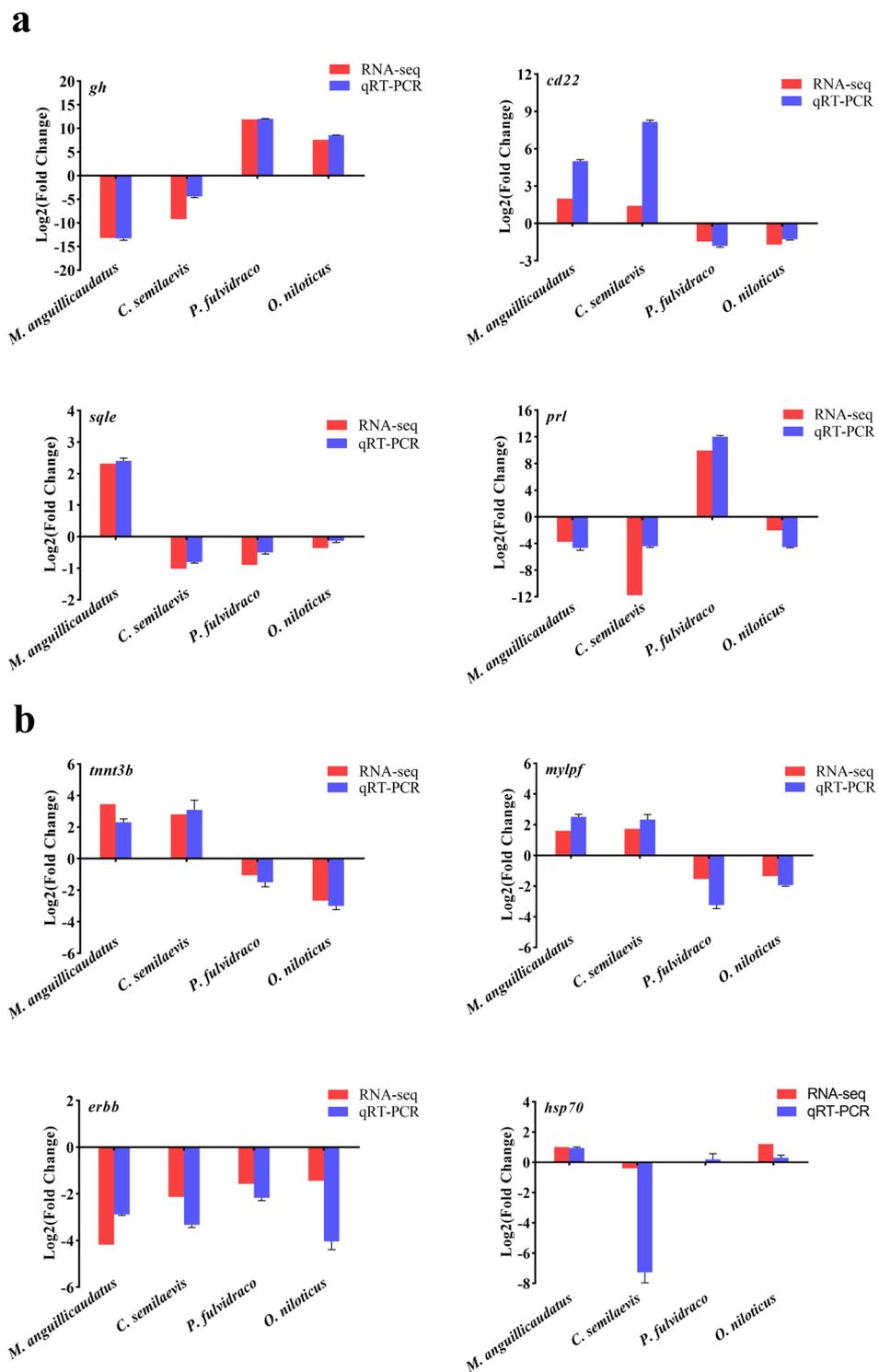


Fig. 4 Illustrating of qRT-PCR confirmation for RNA-seq. Each bar represents the expression fold change in a gene compared to that in the males. **(a)** Expression patterns of DEGs in four fish brains. **(b)** Expression patterns of DEGs in four fish muscles.

Taken together, our findings present a high-quality transcriptomic dataset characterizing differences in transcription levels between females and males in sexually size dimorphic fish, benefiting the study of exploring the common genes regulating the growth differences between females and males in fish with sexually size dimorphism.

Code availability

All software used in this study were executed according to the manual and protocols of the published bioinformatic tools. FastQC, version 0.11.3, was used for the quality check of the raw FASTQ sequencing files. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. The versions and parameters of the transcriptome assembly and expression analysis software described in the methods section are as follows:

Trimmomatic, version 0.36, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50.
 Trinity, version 2.4.0, --seqType fq--SS_lib_type RF.
 hisat2, version 2.2.1.0, --rna-strandness rf -fr.
 cufflinks, version 2.2.1, --library-type fr-firststrand.
 htseq-count, version 0.9.1, -s reverse
 DESeq, version 1.18.0, pvalue < 0.05, |log2FoldChange| > 1

Received: 26 October 2022; Accepted: 6 December 2022;

Published online: 17 December 2022

References

- Bye, V. J. & Lincoln, R. F. Commercial methods for the control of sexual maturation in rainbow trout (*Salmo gairdneri* R.). *Aquaculture* **57**, 299–309, [https://doi.org/10.1016/0044-8486\(86\)90208-5](https://doi.org/10.1016/0044-8486(86)90208-5) (1986).
- Chen, S. L. *et al.* Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar. Biotechnol.* **9**, 273–280, <https://doi.org/10.1007/s10126-006-6081-x> (2007).
- Yoneda, M. *et al.* Age validation and growth variability of Japanese flounder *Paralichthys olivaceus* off the Pacific coast of northern Japan. *Fisheries Sci.* **73**, 585–592, <https://doi.org/10.1111/j.1444-2906.2007.01371.x> (2007).
- Jing, J. *et al.* Sex-biased miRNAs in gonad and their potential roles for testis development in yellow catfish. *PLoS One* **9**, e107946, <https://doi.org/10.1371/journal.pone.0107946> (2014).
- Beardmore, J. A., Mair, G. C. & Lewis, R. I. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* **197**, 283–301, [https://doi.org/10.1016/S0044-8486\(01\)00590-7](https://doi.org/10.1016/S0044-8486(01)00590-7) (2001).
- Wang, D., Mao, H. L., Chen, H. X., Liu, H. Q. & Gui, J. F. Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. *Anim. Genet.* **40**, 978–981, <https://doi.org/10.1111/j.1365-2052.2009.01941.x> (2009).
- Hyun, S. Body size regulation and insulin-like growth factor signaling. *Cell. Mol. Life Sci.* **70**, 2351–2365, <https://doi.org/10.1007/s00018-013-1313-5> (2013).
- Xiong, S. T. *et al.* Characterization and sexual dimorphic expression of Cytochrome P450 genes in the hypothalamic-pituitary-gonad axis of yellow catfish. *Gen. Comp. Endocrinol.* **216**, 90–97, <https://doi.org/10.1016/j.ygcen.2015.04.015> (2015).
- Ma, W. G. *et al.* Sex differences in the expression of GH/IGF axis genes underlie sexual size dimorphism in the yellow catfish (*Pelteobagrus fulvidraco*). *Sci. China Life Sci.* **59**, 431–433, <https://doi.org/10.1007/s11427-015-4957-6> (2016).
- Kang, H., Kenealy, T. M. & Cohen, R. E. The hypothalamic-pituitary-gonadal axis and thyroid hormone regulation interact to influence seasonal breeding in green anole lizards (*Anolis carolinensis*). *Gen. Comp. Endocrinol.* **292**, 113446, <https://doi.org/10.1016/j.ygcen.2020.113446> (2020).
- Connallon, T. & Knowles, L. L. Intergenic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* **21**, 495–499, <https://doi.org/10.1016/j.tig.2005.07.006> (2005).
- Assis, R., Zhou, Q. & Bachtrog, D. Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol. Evol.* **4**, 1189–1200, <https://doi.org/10.1093/gbe/evs093> (2012).
- Zhang, J. L., Zhang, C. N., Liu, M., Fan, M. Z. & Huang, M. X. Transcriptome analyses of sex differential gene expression in brains of rare minnow (*Gobiocypris rarus*) and effects of tributyltin exposure. *Data in Brief* **18**, 1193–1195, <https://doi.org/10.1016/j.dib.2018.03.119> (2018).
- Wan, Z. Y., Lin, G. & Yue, G. H. Genes for sexual body size dimorphism in hybrid tilapia (*Oreochromis sp. x Oreochromis mossambicus*). *Aquaculture and Fisheries* **4**, 231–238, <https://doi.org/10.1016/j.aaf.2019.05.003> (2019).
- Ly, J. J., Zhang, D. N., Gao, B. Q., Liu, P. & Li, J. Transcriptome and MassARRAY analysis for identification of transcripts and SNPs for growth traits of the swimming crab *Portunus trituberculatus*. *Gene* **566**, 229–235, <https://doi.org/10.1016/j.gene.2015.04.050> (2015).
- Perez-Montarelo, D. *et al.* Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. *Physiol. Genomics* **46**, 195–206, <https://doi.org/10.1152/physiolgenomics.00151.2013> (2014).
- Moisa, S. J. *et al.* Bioinformatics analysis of transcriptome dynamics during growth in angus cattle longissimus muscle. *Bioinform. Biol. Insights* **7**, 253–270, <https://doi.org/10.4137/BBI.S12328> (2013).
- Davis, R. V. N. *et al.* Transcriptome analysis of post-hatch breast muscle in legacy and modern broiler chickens reveals enrichment of several regulators of myogenic growth. *PLoS One* **10**, e0122525, <https://doi.org/10.1371/journal.pone.0122525> (2015).
- Naqvi, S. *et al.* Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, eaaw7317, <https://doi.org/10.1126/science.aaw7317> (2019).
- Huang, P. P. *et al.* Genome-wide association study reveals the genetic basis of growth trait in catfish with sexual size. *Genomics* **114**, 110380, <https://doi.org/10.1016/j.ygeno.2022.110380> (2022).
- Kuang, G. Q., Tao, W. J., Zheng, S. Q., Wang, X. S. & Wang, D. S. Genome-wide identification, evolution and expression of the complete set of cytoplasmic ribosomal protein genes in Nile tilapia. *Int. J. Mol. Sci.* **21**, 1230, <https://doi.org/10.3390/ijms21041230> (2020).
- Zhao, M. L., Chi, H. & Sun, L. Neutrophil extracellular traps of *Cynoglossus semilaevis*: production characteristics and antibacterial effect. *Front. Immunol.* **8**, 290, <https://doi.org/10.3389/fimmu.2017.00290> (2017).
- Zhou, H. *et al.* Gonadal transcriptome analysis of hybrid triploid loaches (*Misgurnus anguillicaudatus*) and their diploid and tetraploid parents. *PLoS One* **13**, e0198179, <https://doi.org/10.1371/journal.pone.0198179> (2018).
- Mei, J. & Gui, J. F. Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. *Sci. China Life Sci.* **58**, 124–136, <https://doi.org/10.1007/s11427-014-4797-9> (2015).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
- Grabherr, M. *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652, <https://doi.org/10.1038/nbt.1883> (2011).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360, <https://doi.org/10.1038/nmeth.3317> (2015).
- Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515, <https://doi.org/10.1038/nbt.1621> (2010).

29. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, <https://doi.org/10.1093/bioinformatics/btu638> (2015).
30. Anders, S. & Huber, W. Differential expression of RNA-seq data at the gene level—the DESeq package. *EMBL* (2012).
31. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**, 1559–1582, <https://doi.org/10.1038/nprot.2006.236> (2006).
32. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP313711> (2021).
33. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP313744> (2021).
34. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP314481> (2021).
35. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP313936> (2021).
36. Luo, L. F., Xu, Z. S., Li, D. Y., Hu, Z. & Gao, Z. X. Comparative transcriptome profiles of four sexually size dimorphic fish. *Figshare* <https://doi.org/10.6084/m9.figshare.c.6244080.v2> (2022).
37. Kim, C. H., Kim, E. J., Seo, C. & Nam, Y. K. Transcriptome analysis of maternal gene transcripts in unfertilized eggs of *Misgurnus anguillicaudatus* and identification of immune-related maternal genes. *Int. J. Mol. Sci.* **21**, 3872, <https://doi.org/10.3390/ijms21113872> (2020).

Acknowledgements

This work was financially supported by the National Key Research and Development Program (No. 2018YFD0900205), National Natural Science Foundation of China (No. 31872559 and U21A20263) and China Agriculture Research System of MOF and MARA (No. CARS-45-08), Hubei Hongshan Laboratory (No. 2021hszd001).

Author contributions

Z.G. designed and supervised the study. L.L., Z.X., D.L. and Z.H. performed the sample collection, RNA isolation, and qRT-PCR validation experiments. L.L. performed the data curation and wrote the original draft. Z.G. and L.L. revised the manuscript. All authors have read and agreed to publish the version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01887-1>.

Correspondence and requests for materials should be addressed to L.-F.L. or Z.-X.G.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022