Data in Brief

Identification of myogenic regulatory genes in the muscle transcriptome of beltfish (*Trichiurus lepturus*): A major commercial marine fish species with robust swimming ability

Hui Zhang [a,b,1], Chung-Ming Chang [c,1], Kang-Ning Shen [d], Weiwei Xian [a,b,*], Chung-Der Hsiao [e,**]

[a] Key Laboratory of Marine Ecology and Environment Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[b] Laboratory of Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China
[c] Department of Medical Biotechnology and Laboratory Science, College of Medicine, Chang Gung University, Tao-Yuan 33302, Taiwan
[d] Center of Excellence for the Oceans, National Taiwan Ocean University, 20224 Keelung, Taiwan
[e] Department of Bioscience Technology, Chung Yuan Christian University, 32023 Chung-Li, Taiwan

ARTICLE INFO

ABSTRACT

The beltfish (*Trichiurus lepturus*) is considered as one of the most economically important marine fish in East Asia. It is a top predator with a robust swimming ability that is a good model to study muscle physiology in fish. In the present study, we used Illumina sequencing technology (NextSeq500) to sequence, assemble and annotate the muscle transcriptome of juvenile beltfish. A total of 57,509,280 clean reads (deposited in NCBI SRA database with accession number of SRX1674471) were obtained from RNA sequencing and 26,811 unigenes (with N50 of 1033 bp) were obtained after de novo assembling with Trinity software. BLASTX against NR, GO, KEGG and eggNOG databases show 100%, 49%, 31% and 96% annotation rate, respectively. By mining beltfish muscle transcriptome, several key genes which play essential role on regulating myogenesis, including *pax3*, *pax7*, *myf5*, *myoD*, *mrf4/myf6*, *myogenin* and *myostatin* were identified with a low expression level. The muscle transcriptome of beltfish can provide some insight into the understanding of genome-wide transcriptome profile of teleost muscle tissue and give useful information to study myogenesis in juvenile/adult fish.

Specifications [standardized info for the reader]	
Organism/cell line/tissue	*Trichiurus lepturus*/muscle
Sex	N/A
Sequencer or array type	Illumina NextSeq500
Data format	Raw and processed
Experimental factors	Transcriptome profiling of muscle at adult stage
Experimental features	Muscle tissues were dissected from the juvenile beltfish and total RNAs were extracted by using TRIZOL reagent. Prepared cDNA libraries were paired-end sequenced by NextSeq500 platform. The obtained data was subjected for de novo transcriptome assembly using Trinity. The assembled unigene was later functionally annotated by searching NR, GO, KEGG, eggNOR and Swissprot databases.
Consent	N/A
Sample and location	One juvenile fish of beltfish was captured at Yangtze Estuary on Nov 4th, 2015

* Correspondence to: W. Xian, Key Laboratory of Marine Ecology and Environment Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.
** Corresponding author.
E-mail addresses: wwxian@qdio.ac.cn (W. Xian), cdhsiao@cycu.edu.tw (C.-D. Hsiao).
[1] Equal contribution authors.

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRX1674471 for muscle transcriptome.

2. Introduction

The muscle development process is spatially and temporally orchestrated by multiple myogenic regulatory factors. At cell fate determination stage of embryo, the external signals trigger mesodermal muscle progenitors transforming into myoblast. The interplay of *pax3/pax7*, *myoD*, *myf5* and *mrf4/myf6* promotes myoblast differentiating into multinucleate myotube (by a cell fusion process) which highly expressed *myogenin* and *mrf4/myf6*. The *myostatin* (also known as GDF-8), on the contrary, plays an inhibitory role on muscle differentiation. Later, the multinucleate myotube further assembles into muscle fiber and expresses variety of muscle specific proteins (MSPs) to build up the major architecture of muscle fiber [1,10,12,14,17]. Basically, the muscle development process is highly conserved between fish and high vertebrates. However, unlike their high vertebrate counterparts, fish continue their hyperplastic and hypertrophic muscle growth through adult stage [9,13]. Study on how muscle can have continuous cell growth
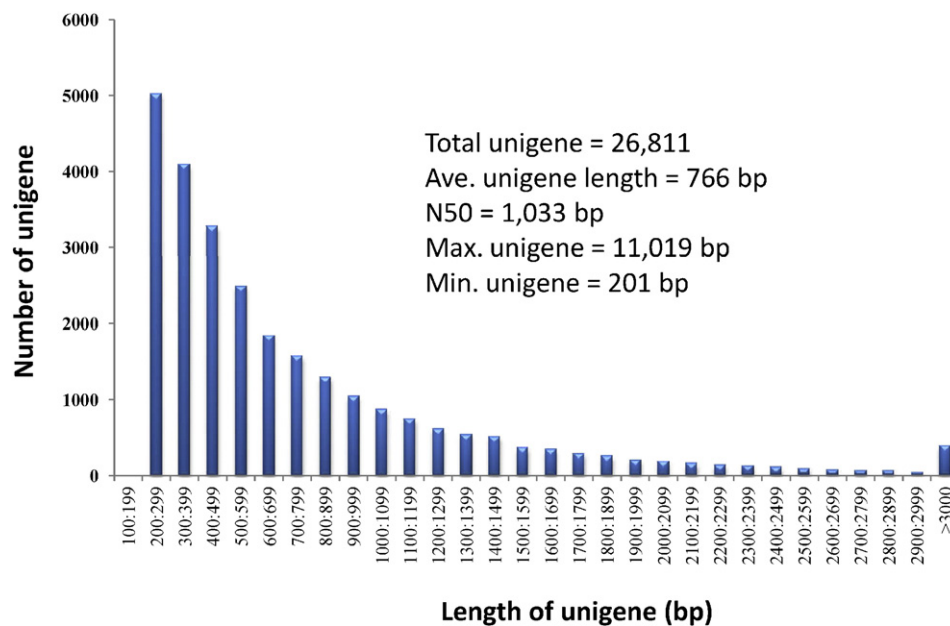
Fig. 1. Length distribution of the assembled unigene of beltfish muscle transcriptome.

and proliferation in fish can benefit better understanding on boosting the growth/repair of muscle tissue at juvenile/adult stage in high vertebrates.

Beltfish (*Trichiurus lepturus*) is a member of the cutlassfish family (Trichiuridae) and is a major commercial marine fish species with robust swimming ability. It is a long, slender fish found throughout the tropical and temperate oceans of the world. Beltfish is a major commercial species in Northwest Pacific, especially in China, South Korea and Japan. Juvenile beltfish participate in the diet vertical migration, rising to feed on krill and small fish during the night and returning to the sea bed in the day. The movement pattern is reversed by large adults, which mainly feed on fish, squid and shrimp. Adults are highly carnivorous and will cannibalize younger individuals regularly [2]. Since Beltfish play as a top predator in the marine ecosystem, they provide a good model to monitor the bioaccumulation of heavy metals (like mercury and selenium) in the muscle tissues for a long time [4,15,16]. In this study, we took Beltfish as a marine fish model to explore muscle gene regulation at a molecular level by RNAseq approach. The establishment of muscle transcriptome provides not only useful information for evaluating the biological impact of heavy metal bioaccumulation on muscle, but also provides fundamental information for myogenic regulatory gene expression at juvenile/adult stages.

3. Experimental design, materials and methods

3.1. RNA extraction

The muscle tissue was dissected from one wild captured juvenile beltfish (body length around 15 cm) and immediately stored in RNAlater (Qiagen, Hilden, Germany) and then stored at −80 °C prior to RNA extraction. Total RNAs were extracted by using the TRIZOL Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Total RNA samples were then digested by DNase I to remove potential genomic DNA contamination. Integrity and size distribution were checked with Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

3.2. RNA isolation, library construction and Illumina sequencing

Initially, about 2.5 μg of starting total RNAs were used to synthesize cDNA libraries by following the standard protocols of the Illumina

TruSeq RNA Sample Preparation Kit (Illumina). The final library had an insert size about 200–300 bp. After qPCR quantitation and dilution, the library was sequenced on an Illumina NextSeq500 with 150 bp paired-end reads. A total of 57,851,358 raw paired-end reads were generated. Adaptor sequences were trimmed; and reads with low quality were removed by cutadapt software [8]. After the removal of ambiguous nucleotides, duplicates and low-quality sequences (Phred quality scores < 20), a total of 57,509,280 cleaned reads (99.4%) were obtained. The raw transcriptome sequences in the present study were deposited in the NCBI SRA database (SRX1674471).
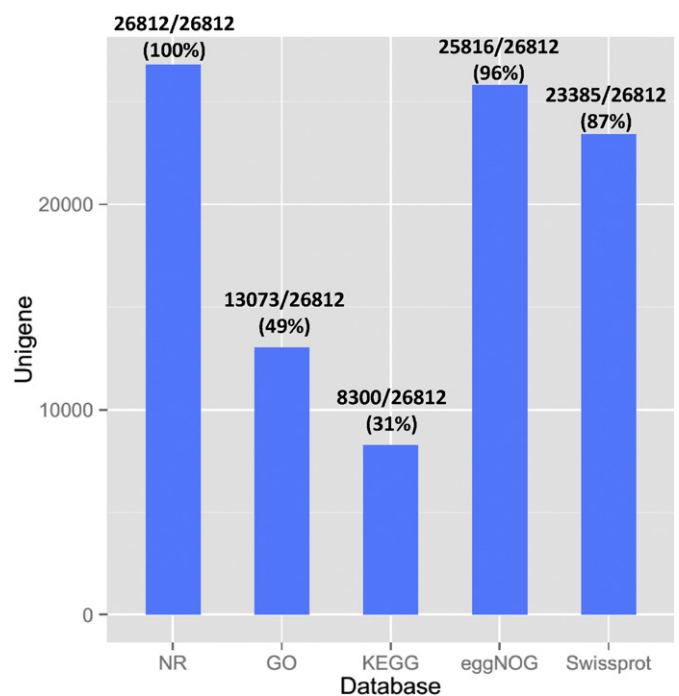

Fig. 2. Comparison of the gene annotation rate of unigene against NR, GO, KEGG, eggNOR and Swissprot databases.
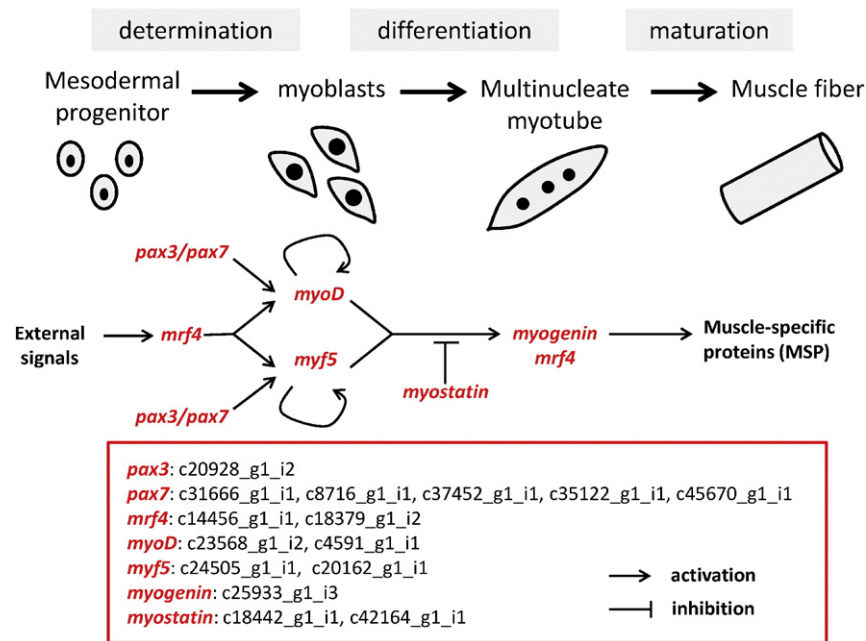
Fig. 3. Schematic picture demonstrating the myogenesis pathway and the myogenic regulatory genes (highlight in red color) identified in beltfish muscle transcriptome.

3.3. De novo transcriptome assembly and functional annotation of muscle expressed genes in beltfish

Cleaned reads were de novo assembled into unigenes by Trinity software [5] with default parameters settings (k mer = 25 bp). The transcriptome was assembled into 26,811 unigenes, ranging from 201 to 11,019 bp in length and that the N50 length was 1033 bp. The unigene length distribution is shown in Fig. 1. The assembled transcriptomic unigenes were subjected to similarity search against non-redundant (NR) protein, Gene ontology (GO), KEGG, eggNOG [11] and Swissprot databases using BLASTX with an e-value cut-off of 1e − 5. Gene names and descriptions were assigned to each unigene based on the BLASTX results. Gene ontology (GO) analysis was then conducted on the assembled unigenes by using Blast2GO [3]. KEGG pathways were assigned to assembled unigenes using online KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/tools/kaas/). The Bi-directional Best Hit (BBH) method was used to obtain KEGG Orthology (KO) assignment. BLASTX against NR, GO, KEGG and eggNOG databases showed 100%, 49%, 31% and 96% annotation rate, respectively. The systematic comparison of gene annotation rate was summarized in Fig. 2.

3.4. Identification of myogenic regulatory genes

The high-quality cleaned reads of RNA-seq library were mapped to the assembled unigenes with Bowtie2 program [6]. The counting of alignments was done using RSEM [7]. Unigenes with RPKM (reads per kilobase of exon per million reads mapped) ≥100 were defined as abundant expressed genes. In juvenile muscle transcriptome of beltfish, we totally identified 15 unigenes with high similarities to either *pax3*, *pax7*, *myoD*, *myf5*, *myogenin*, *mrf4/myf6* or *myostatin*. To validate the identity of those myogenic regulatory gene homologs, we downloaded the complete protein dataset of 27 fish species with complete genome deposited in NCBI ftp sites (ftp://ftp.ncbi.nlm.nih.gov/genomes/), and combined into a single fish protein database to perform in-house BLAST (summarized in Table S1). Later, we constructed gene-specific phylogenetic tree by using Geneious software (http://www.geneious.com/) with 1000 bootstrap Neighbor-Joining calculation (Fig. S1). In total, we identified one *pax3* unigene (c20928_g1_i2), five *pax7* unigenes (c31666_g1_i1, c8716_g1_i1, c37452_g1_i1, c35122_g1_i1, c45670_

g1_i1), two *mrf4/myf6* unigenes (c14456_g1_i1 and c18379_g1_i2), two *myoD* unigenes (c23568_g1_i2 and c4591_g1_i1), two *myf5* unigenes (c24505_g1_i1 and c20162_g1_i1), one *myogenin* unigene (c25933_g1_i3) and two *myostatin* unigenes (c18442_g1_i1 and c42164_g1_i1) (Fig. 3 and Fig. S2). The relative expression level of those myogenic regulatory genes in the juvenile beltfish muscle is relatively low (RPKM < 20, summarized in Table S2) when compared to those highly expressed muscle specific proteins like *parvalbumin 3* (c47515_g1_i1, RPKM = 65.269) and *myosin heavy chain* (c28260_g1_i1, RPKM = 57.107) etc. Based on monophylogenic tree topology, we provided strong evidences to support the molecular identity of those myogenic regulatory genes identified in beltfish muscle transcriptome. We believe this beltfish muscle transcriptome and those identified myogenic regulatory genes will provide insights into the muscle development in beltfish.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.04.005.

References

[1] C.A. Berkes, S.J. Tapscott, MyoD and the transcriptional control of myogenesis. Paper Presented at: Seminars in Cell & Developmental Biology, Elsevier, 2005.
[2] V.T. Bittar, D.R. Awabdi, W.C.T. Tonini, M.V. Vidal Junior, D. Beneditto, A.P. Madeira, Feeding preference of adult females of ribbonfish *Trichiurus lepturus* through prey proximate-composition and caloric values. Neotrop. Ichthyol. 10 (2012) 197–203.
[3] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676.
[4] M.F. Costa, S.C. Barbosa, M. Barletta, D.V. Dantas, H.A. Kehrig, T.G. Seixas, O. Malm, Seasonal differences in mercury accumulation in *Trichiurus lepturus* (Cutlassfish) in relation to length and weight in a Northeast Brazilian estuary. Environ. Sci. Pollut. Res. 16 (2009) 423–430.
[5] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (2013) 1494–1512.

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

[7] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12 (2011) 323.

[8] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17 (2011) 10–12.

[9] F. Mascarello, A. Rowlerson, G. Radaelli, P.-A. Scapolo, A. Veggetti, Differentiation and growth of muscle in the fish *Sparus aurata* (L): I. Myosin expression and organization of fibre types in lateral muscle from hatching to adult. J. Muscle Res. Cell Motil. 16 (1995) 213–222.

[10] E.N. Olson, W.H. Klein, bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev. 8 (1994) 1–8.

[11] S. Powell, K. Forslund, D. Szklarczyk, K. Trachana, A. Roth, J. Huerta-Cepas, T. Gabaldón, T. Rattei, C. Creevey, M. Kuhn, eggNOG v4. 0: nested orthology inference across 3686 organisms. Nucleic Acids Res. (2013) gkt1253.

[12] R. Ríos, I. Carneiro, V.M. Arce, J. Devesa, Myostatin is an inhibitor of myogenic differentiation. Am. J. Phys. Cell Phys. 282 (2002) C993–C999.

[13] A. Rowlerson, F. Mascarello, G. Radaelli, A. Veggetti, Differentiation and growth of muscle in the fish *Sparus aurata* (L): II. Hyperplastic and hypertrophic growth of lateral muscle from hatching to adult. J. Muscle Res. Cell Motil. 16 (1995) 223–236.

[14] P. Seale, L.A. Sabourin, A. Girgis-Gabardo, A. Mansouri, P. Gruss, M.A. Rudnicki, Pax7 is required for the specification of myogenic satellite cells. Cell 102 (2000) 777–786.

[15] T.G. Seixas, I. Moreira, O. Malm, H.A. Kehrig, Bioaccumulation of mercury and selenium in *Trichiurus lepturus*. J. Braz. Chem. Soc. 23 (2012) 1280–1288.

[16] T.G. Seixas, I. Moreira, O. Malm, H.A. Kehrig, Mercury and selenium in a top-predator fish, *Trichiurus lepturus* (Linnaeus, 1758), from the Tropical Brazilian Coast, Rio de Janeiro. Bull. Environ. Contam. Toxicol. 89 (2012) 434–438.

[17] K. Yun, B. Wold, Skeletal muscle determination and differentiation: story of a core regulatory network and its context. Curr. Opin. Cell Biol. 8 (1996) 877–889.