Genomics Data 12 (2017) 58–61

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Comparative study the expression of calcium cycling genes in Bombay duck (*Harpadon nehereus*) and beltfish (*Trichiurus lepturus*) with different swimming activities

Hui Zhang [a,b], Gilbert Audira [c], Yuan Li [d,e], Weiwei Xian [a,b,*], Muhammed Muhsin Varikkodan [f], Chung-Der Hsiao [c,g,h,**]

[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 Bioscience Technology, Chung Yuan Christian University, 32023 Chung-Li, Taiwan
[d] Third Institute of Oceanography, SOA, Xiamen 361005, China
[e] Open Research Fund Program of MATHAB, SOA, Shanghai 201206, China
[f] Department of Chemistry, Chung Yuan Christian University, 32023 Chung-Li, Taiwan
[g] Center for Biomedical Technology, Chung Yuan Christian University, Chung-Li 32023, Taiwan
[h] Center for Nanotechnology, Chung Yuan Christian University, Chung-Li 32023, Taiwan

ARTICLE INFO

Article history:
Received 25 February 2017
Received in revised form 4 March 2017
Accepted 19 March 2017
Available online 20 March 2017

Keywords:
Muscle
Transcriptome
Beltfish
Bombay duck
Calcium cycling gene

ABSTRACT

The contraction and relaxation events of the muscle is mediated by the coordination of many important calcium cycling proteins of ryanodine receptor (RYR), troponin C (TNNC), parvalbumin (PVALB), sarcoendoplasmic reticulum calcium transport ATPase (SERCA) and calsequestrin (CASQ). In higher vertebrates, the expression level of calcium cycling proteins are positively correlated to the muscle contraction/relaxation ability of the cell. In this study, we used RNAseq to explore the expression profile of calcium cycling genes between two marine fish of Bombay duck (*Harpadon nehereus*) and beltfish (*Trichiurus lepturus*) with poor and robust swimming activities, respectively. We have studied the hypothesis whether the expression level of calcium cycling proteins are also positive correlated to swimming ability in fish. We used Illumina sequencing technology (NextSeq500) to sequence, assemble and annotate the muscle transcriptome of Bombay duck for the first time. A total of 47,752,240 cleaned reads (deposited in NCBI SRA database with accession number of SRX1706379) were obtained from RNA sequencing and 26,288 unigenes (with N50 of 486 bp) were obtained after de novo assembling with Trinity software. BLASTX against NR, GO, KEGG and eggNOG databases show 100%, 65%, 26%, 94% and 88% annotation rate, respectively. Comparison of the dominantly expressed unigenes in fish muscle shows calcium cycling gene expression in beltfish (SRX1674471) is 1.4- to 51.6-fold higher than Bombay duck. Among five calcium cycling genes, the fold change results are very significant in CASQ (51.6 fold) and PVALB (9.1 fold) and both of them are responsive for calcium binding to reduce free calcium concentration in the sarcoendoplasmic reticulum and cytoplasm. In conclusion, we confirmed that the high abundant expression rate of calcium cycling genes in robust swimming fish species. The current muscle transcriptome and identified calcium cycling gene data can provide more insights into the muscle physiology of fish.

Specifications	
Organism/cell line/tissue	*Harpadon nehereus*/muscle
Sex	N/A
Sequencer or array type	Illumina NextSeq500
Data format	Raw and processed
Experimental factors	Transcriptome profiling of muscle at the juvenile stage
Experimental	The muscle tissues collected from juvenile Bombay duck's

* Correspondence to: W. Xian, Key Laboratory of Marine Ecology and Environment Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.
** Correspondence to: C.-D. Hsiao, Department of Bioscience Technology, Chung Yuan Christian University, 32023 Chung-Li, Taiwan.
E-mail addresses: wwxian@qdio.ac.cn (W. Xian), cdhsiao@cycu.edu.tw (C.-D. Hsiao).

http://dx.doi.org/10.1016/j.gdata.2017.03.003

features	(body length = 2.0 cm) and extracted the total RNAs by TRIZOL method. Well prepared cDNA libraries were paired-end sequenced through the NextSeq500 platform. The obtained data are subjected for the de novo transcriptome assembly by using Trinity. Later, the assembled unigene are functionally annotated by searching NR, GO, KEGG, eggNOR and Swissprot databases.
Consent	N/A
Sample source location	Larval form of Bombay duck captured from Yangtze Estuary on Nov 4th, 2015 by Hui Zhang.

1. Direct link to deposited data

The data related with Bombay duck and beltfish muscle trancriptomes are deposited to NCBI SRA database with accession number of SRX1706379 and SRX1674471, respectively.

2. Introduction

Calcium (Ca^{2+}) is the major element in the body and involved in a number of cellular events, including cell motility, transmission of nerve impulses, excitation-contraction of muscles, release of neurotransmitters, cell secretory, and membrane permeability. The contraction and relaxation events of the muscle are the best investigated mechanisms, which is mediated by the coordination of many important calcium cycling proteins. The previous literature shows that many signals can activate these events. The ryanodine receptor (RYR) can release calcium ions (Ca^{2+}) from sarcoplasmic reticulum (SR) into cytoplasm to induce a calcium spark. Later, this Ca^{2+} can form complex structure with troponin C (TNNC), which can be inducing the muscle contraction in the body. Afterwards, parvalbumin (PVALB), an acidic intracellular Ca^{2+}-binding protein plays an important role on the function of muscle in the body. Afterwards, the retrieval of Ca^{2+} ions from the myofibril can be transfer to the SR. Normally, high concentration of intracellular Ca^{2+} buffer is present in the fast-contracting skeletal muscles across the animals. The Parvalbumin Protein coding gene called PVALB is one of the members of this family for Ca^{2+}-binding molecules, which is always check on Ca^{2+} switching in a cell [1,2]. The loss of function of PVALB can prolong the contraction/relaxation cycle in the fast-twitch muscle of animals [3]. Accordingly the cytoplasmic Ca^{2+} move back to SR through sarcoendoplasmic reticulum calcium transport ATPase (SERCA). Normally, the free Ca^{2+} ions in the SR bound with calsequestrin (CASQ) and acts as dual role in excitation–contraction coupling to the buffer free Ca^{2+} in the cell. It can hold to increase SR capacity and modulate the activity of Ca^{2+} release ryanodine receptor (RYR) channels (Fig. 1). The expression level of those calcium cycling proteins are positively correlated to the muscle contraction/relaxation ability of the cell. For example, the mutations of calcium cycling genes in human can lead to form many health problems like familial ventricular arrhythmias, cardiomyopathy and others [4–6]. The same time, previous study clears that the overexpression of fast-twitch skeletal muscle type of SERCA in transgenic mouse heart can enhance myocardial contractility and increased Ca^{2+} transport function in the body [7].

Bombay duck (*Harpadon nehereus*) is a kind of lizardfish, which can inhabit at the tropical areas of the Indo-Pacific region. It mainly observed and caught from Maharashtra, and Lakshadweep Sea. The little number of this fish can observe at Bay of Bengal and in the East and South China Sea too. Most of the seasons, it can observe at the deep water offshore on sandy mud bottom. The same time, it can also gathers in large shoals at deltas of rivers to feed during monsoons. The Bombay duck can spawn six batches of breeds per year in the life and the adults usually have 25 cm in size [8]. Various observations suggested that the fish can also reach at maximum length of 40 cm in the life (not all fishes).

The dried *H. nehereus* is the regional food in India can produce extremely odor. Normally fresh fish can usually fried and served as a starter in regional shops and homes. In Mumbai, Konkan, and the western coastal areas in India, this dish is popularly known as “Bombil fry”. The previous studies say that, 90% of this fish includes moisture content, whereas the fewer amounts of protein and fat in muscle comparing with other species [9]. In addition, the swimming ability of Bombay duck is very poor and move only with the tidal oscillations [9]. The slow swimming activity mainly helps the muscle fibers to perform basic aerobic metabolic functions including, circulatory and respiratory systems to supply needful substrates and oxygen. As swimming speed increases, the contracting muscle fibers become faster in the tissues. The maximum performance can achieved throughout fast-starts associated with predation, escape responses and involves the mobilization of the entire white muscle mass. These white muscle masses are typically arranged by single fiber type that expresses fast isotypes of the myofibrillar proteins and containing high concentrations of the cytoplasmic Ca^{2+}-binding protein parvalbumin. It also has larger average diameters than red fibers and contains higher volume densities of myofibrils. And a more sarcoplasmic reticulum for faster Ca^{2+} cycling [10].

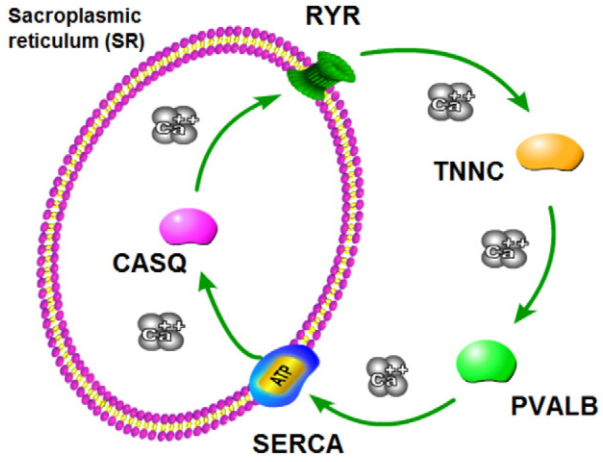

Fig. 1. Schematic picture demonstrate the function of five calcium cycling genes (CASQ, PVALB, RYR, SERCA and TNNC) identified in Bombay duck muscle transcriptome.

In the previous work, we performed RNAseq to explore the muscle transcriptome in beltfish (*Trichiurus lepturus*) with a robust swimming ability (RNAseq data are deposited as SRX1674471) [11]. This study, we performed the RNAseq to explore the muscle transcriptome in Bombay duck with a poor swimming ability. We have studied the hypothesis whether the expression level of calcium cycling proteins are positive correlated to fish swimming ability.

3. Experimental design, materials and methods

3.1. RNA extraction

The muscle tissue dissected from the wild juvenile Bombay duck (body length around 2.0 cm) and stored in RNAlater (Qiagen, Hilden, Germany) at −80 °C prior to RNA extraction. The total RNAs were extracted by TRIZOL Kit (Invitrogen, Carlsbad, CA, USA) and samples digested by DNase I to prevent the genomic DNA contamination. The integrity and size distribution of RNA was checked with Bioanalyzer 2100 (Agilent technologies, Santa Clara, CA, USA).

3.2. RNA isolation, library construction and Illumina sequencing

2.5 μg of RNAs were used to synthesize the cDNA libraries by Illumina TruSeq RNA Sample Preparation Kit. The final library had an insert size about 200–300 bp. After qPCR quantitation and dilution, the library was sequenced with Illumina NextSeq500 through 150 bp paired-end reads. The total of 45,944,846 raw paired-end reads were generated

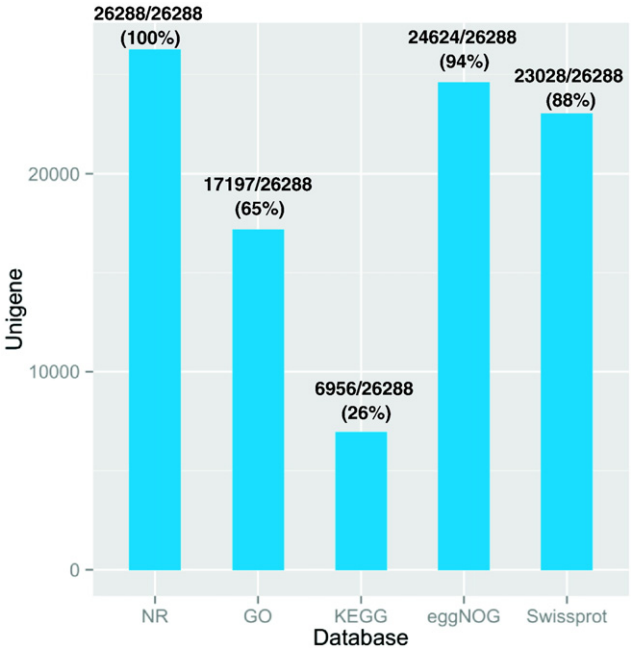

Fig. 2. Comparison of the gene annotation rate of unigene against NR, GO, KEGG, eggNOR and Swissprot databases of Bombay duck muscle transcriptome.

from the sample and the adaptor sequences were trimmed. Further, the low quality reads removed by cutadapt software [12]. Finally, the removal of ambiguous nucleotides, duplicates and low-quality sequences (Phred quality scores < 20) from the data can get total of 47,752,240 cleaned reads (99.6%) from the sample. The raw transcriptome sequences in the present study were deposited in the NCBI SRA database (SRX1706379).

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

The cleaned reads were de novo assembled into contigs by Trinity software [13] with default parameters settings. The transcriptome was assembled into 26,288 unigenes with the N50 length of 486 bp. The assembled transcriptomic unigenes subjected to the similarity search against non-redundant (NR) protein, Gene ontology (GO), KEGG, eggNOG [14] 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 transcriptome by using Blast2GO [15]. KEGG pathways were assigned to assembled unigenes using the 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, eggNOG and Swissprot databases show 100%, 65%, 26%, 94% and 88% annotation rate, respectively. The systematic comparison of gene annotation rate was summarized in Fig. 2.

3.4. Identification of calcium cycling genes

The high-quality cleaned reads of each RNAseq library were mapped to the assembled transcripts with Bowtie2 program [16]. The counting of alignments was done using RSEM [17]. The unigene with RPKM (reads per kilobase of exon per million reads mapped) ≥ 100 was defined as abundant expressed genes. For the validation of protein identity of calcium cycling homologs, we downloaded the data of fish species includes zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*), tilapia (*Oreochromis niloticus*) and large yellow croaker (*Larimichthys crocea*) from NCBI ftp sites (ftp://ftp.ncbi.nlm.nih.gov/genomes/). We made a single protein database to perform in-house BLAST and constructed the gene-specific phylogenetic tree by using Geneious software (http://www.geneious.com/) with 1000 bootstrap Neighbor-Joining calculation (Fig. S1). Phylogenic tree topology analysis

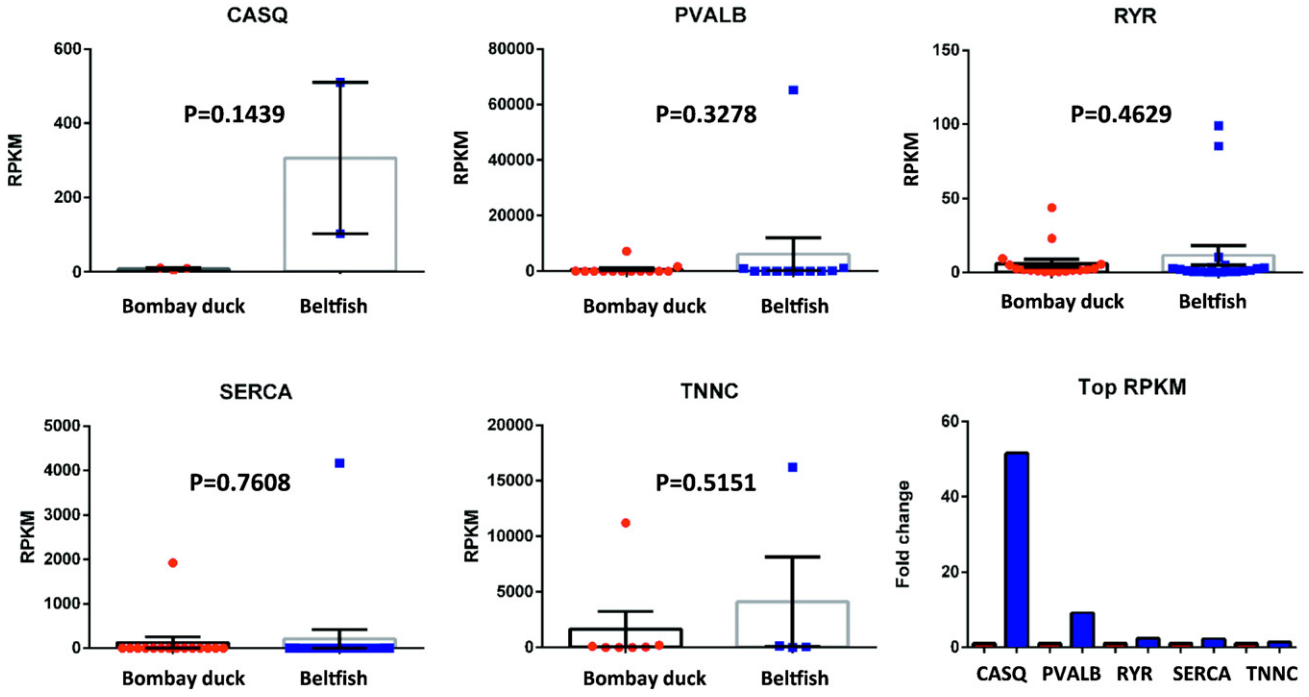

Fig. 3. Comparison of the gene expression level of calcium cycling genes between Bombay duck (red) and Beltfish (blue). The significance is statistically compared by *t*-test.

revealed the strong evidences to support the molecular identity of those calcium cycling genes identified in Bombay duck muscle transcriptome.

The comparison of gene expression level in fishes shows that 3 CASQ, 13 PVALB, 17 RYR, 15 SERCA and 7 TNNC unigenes in Bombay duck. The same time, 2 CASQ, 11 PVALB, 19 RYR, 20 SERCA and 4 TNNC unigenes in beltfish muscle transcriptomes. The relative expression by RPKM method found that the 2 PVALB unigenes (Hne_c22292_g1_i1 and Hne_c52631_g1_i1), 1 SERCA unigene (Hne_c4397_g1_i1) and 1 TNNC unigene (Hne_c52572_g1_i1) have high relative expression level in Bombay duck. In the beltfish, 3 PVALB unigenes (Tle_c47515_g1_i1, Tle_c18933_g1_i1 and Tle_c18968_g1_i1), 1 SERCA unigene (Tle_c13632_g1_i1) and 1 TNNC unigene (Tle_c22302_g1_i1) shows high relative expression level (RPKM > 1000, summarized in Table 1). The statistical comparison of all expressed calcium cycling unigenes in muscle has no significant difference between Bombay duck and beltfish. However, the comparison of dominantly expressed unigenes in fish muscle shows that the calcium cycling gene expression in beltfish is 1.4- to 51.6-fold higher than Bombay duck. Among five calcium cycling genes, the fold change results are very significant in CASQ (51.6 fold) and PVALB (9.1 fold) and both of them are responsive for calcium binding to reduce free calcium concentration in the SR and cytoplasm (see Fig. 3). By the similar approach, literature reported that the robust swimming fish species like (Pacific bluefin tuna) and Pacific cod have abundant expression of glycolytic enzyme genes [19]. In this study, the results confirmed that the high abundant expression rate of calcium cycling genes in robust swimming fish species. The current muscle transcriptome and identified calcium cycling gene data can provide more insights into the muscle physiology of fish.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2017.03.003.

Acknowledgments

This work was supported by grants from National Natural Science Foundation of China (Nos. 41406136 and U1406403), Key Laboratory of Marine Ecology and Environmental Science and Engineering, SOA (MESE-2015-02), Nantong Municipal Science and Technology Project (MS12015118) and Open Research Fund Program of MATHAB, SOA (MATHAB201603).

References

[1] S.H. Arif, A Ca^{2+}-binding protein with numerous roles and uses: parvalbumin in molecular biology and physiology. BioEssays 31 (4) (2009) 410–421.
[2] J. Rall, Role of parvalbumin in skeletal muscle relaxation. Physiology 11 (6) (1996) 249–255.
[3] B. Schwaller, J. Dick, G. Dhoot, S. Carroll, G. Vrbova, P. Nicotera, et al., Prolonged contraction-relaxation cycle of fast-twitch muscles in parvalbumin knockout mice. Am. J. Phys. Cell Phys. 276 (2) (1999) C395–C403.
[4] A. Marjamaa, P. Laitinen-Forsblom, A.M. Lahtinen, M. Viitasalo, L. Toivonen, K. Kontula, et al., Search for cardiac calcium cycling gene mutations in familial ventricular arrhythmias resembling catecholaminergic polymorphic ventricular tachycardia. BMC Med. Genet. 10 (1) (2009) 12.
[5] S.R. Houser, V. Piacentino III, J. Weisser, Abnormalities of calcium cycling in the hypertrophied and failing heart. J. Mol. Cell. Cardiol. 32 (9) (2000) 1595–1607.
[6] S. Minamisawa, M. Hoshijima, G. Chu, C.A. Ward, K. Frank, Y. Gu, et al., Chronic phospholamban–sarcoplasmic reticulum calcium ATPase interaction is the critical calcium cycling defect in dilated cardiomyopathy. Cell 99 (3) (1999) 313–322.
[7] E. Loukianov, Y. Ji, I.L. Grupp, D.L. Kirkpatrick, D.L. Baker, T. Loukianova, et al., Enhanced myocardial contractility and increased Ca^{2+} transport function in transgenic hearts expressing the fast-twitch skeletal muscle sarcoplasmic reticulum Ca^{2+}-ATPase. Circ. Res. 83 (9) (1998) 889–897.
[8] I. Fernandez, M. Devaraj, Dynamics of the Bombay duck (*Harpodon nehereus*) stock along the northwest coast of India. Indian J. Fish. 43 (1) (1996) 1–11.
[9] V. Deshmukh, Responsible Marine Fisheries: Reflections from Maharashtra. 2013.
[10] C.M. Wood, D.G. McDonald, Global Warming: Implications for Freshwater and Marine Fish. Cambridge University Press, 1997.
[11] H. Zhang, C.-M. Chang, K.-N. Shen, W. Xian, C.-D. Hsiao, 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 8 (2016) 81–84.
[12] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17 (1) (2011) 10–12.
[13] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, et al., De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (8) (2013) 1494–1512.
[14] S. Powell, K. Forslund, D. Szklarczyk, K. Trachana, A. Roth, J. Huerta-Cepas, et al., eggNOG v4. 0: nested orthology inference across 3686 organisms. Nucleic Acids Res. (2013), gkt1253.
[15] 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 (18) (2005) 3674–3676.
[16] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4) (2012) 357–359.
[17] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12 (2011) 323 Epub 2011/08/06 10.1186/1471-2105-12-323 (PubMed PMID: 21816040; PubMed Central PMCID: PMC3163565).