Meta Gene 13 (2017) 140–148

Contents lists available at ScienceDirect

Meta Gene

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

In silico characterization of Myogenic Factor 6 transcript of Hilsa, *Tenualosa ilisha* and putative role of its SNPs with differential growth

B.K. Divya, Prabhaker Yadav, Prachi Masih, Rajeev K. Singh, Vindhya Mohindra*

ICAR-National Bureau of Fish Genetic Resources, Canal Ring Road, P. O. Dilkusha, Lucknow 226002, U.P., India

ARTICLE INFO

Keywords:
Myf6
In silico
Open-reading frame
Protein-protein interaction
Single nucleotide polymorphism
Slow/fast growing

ABSTRACT

This study aimed at elucidating the structure of deduced protein Myf6 in an anadromous fish, *T. ilisha* (Hilsa) and polymorphism present in it among different fish species. *Myf6* in Hilsa was identified from muscle Transcriptome sequences through Illumina HiSeq sequencing and designated as TiMyf6, the complete length was 1009 nt, in which ORF (open reading frame) consisted of 720 nt. Its secondary structure of modeled protein consisted of 8 helices, 6 helix-helix interacts, 32 beta turns and 26 gamma turns. The protein-protein interaction prediction of TiMyf6 protein showed functional association with 10 proteins. A total of 3 SNPs were identified in the slow and fast growing fish groups, which had altered the amino acids E/G, Q/G and V/A. These 3 SNPs identified in present study, may function as candidate marker to differentiate slow and fast growing species.

1. Introduction

Skeletal muscle, considered as an organ for the muscular system (Bentzinger et al., 2012) is the consequences of sundry process of myogenesis in vertebrates (Knight and Kothary, 2011). Myogenesis is regulated by complex regulatory mechanisms consisting of different myogenic regulatory factors (MRFs) (Bentzinger et al., 2012). The muscle specific transcription factors belongs to basic helix-loop-helix (bHLH) family proteins (Moncaut et al., 2013) which includes Myoblast Determination Protein (MyoD), Myogenin (MyoG), Myogenic Factor 5 (Myf5) and Myogenic Factor 6 (Myf6/MRF4), (De Almeida et al., 2008). MyoD and Myf5 genes are involved in differentiation and proliferation during embryogenesis, where as MyoG and Myf6 contribute in differentiation of myoblasts (Lin et al., 2015). Myf6 is found to be expressed in many adult muscle fibers (Pin and Konieczny, 2002). The particular role of Myf6 is indistinct due to the phenomenon where the effect of one gene is modified by one or several other genes at the complex myf6/myf5 locus (Hinits and Hughes, 2007). The proteomic studies on fish larvae have elucidated on changes of proteome expression during fish development (Gomez-Requeni et al., 2010). The expression studies on Myf6 showed low expression level in *Trichiurus lepturus* (Zhang et al., 2016) and high in *Schizothorax prenanti* (Lin et al., 2015) depicting the role of Myf6 in growth and development in fish.

Fishes have a wide variety of complex physiological processes involving many genetic and environmental factors (Danzmann et al., 2016). Effect of climate variability is more pronounced in the fast-growing species as populations grow and decline quickly (Pinsky and Byler, 2015), because of the increased hyperplasia in fast than slow growing fish (Kamler, 2008). The tropical fish hilsa (*Tenualosa ilisha*), the Indian anadromous shad belongs to the family Clupeidae and is considered one of the most important species due to its unique taste (Hamilton, 1822; Mohanty et al., 2011). The recorded age (maximum) was 5 to 7 years with the side length of 52.5–61.4 cm and about 4250 g in weight (Pillay et al., 1963; Islam and Talbot, 1968; Ahmed et al., 2008; Bhaumik, 2015). Most fishes continue to grow throughout their lives (Prince et al., 2014) but it is also evident that the growth rate of Hilsa tends to be slow and steady as the fish gets older (Miah, 1997).

Knowledge of the mechanisms of muscle growth and development is one of the most important issue in animal as well as the critical mechanism. Here, the *in silico* investigation was executed for characterization of Ti Myf6 transcript and its deduced protein of Hilsa to understand its mechanism in growth and association of its sequence variation with pattern in growth rate in fish Variations in its sequence in some of fast and slow growing fishes were identified.

2. Materials and methods

2.1. Transcriptome sequencing and identification of Myf6 transcript

The muscle tissue sample of *T. ilisha* from Padma river at Farakka

Abbreviations: MRFs, myogenic regulatory factors; MyoD, Myoblast Determination Protein; MyoG, Myogenin; Myf5, Myogenic Factor 5; Myf6, Myogenic Factor 6; ORF, open reading frame; CREB, cAMP response-element binding protein; MEF2C, Myocyte Enhancing Factor 2C; MEF2A, myocyte enhancer factor 2A; TCF3, transcription factor 3

* Corresponding author.
E-mail address: vmohindra@nbfgr.res.in (V. Mohindra).

http://dx.doi.org/10.1016/j.mgene.2017.06.003
Received 15 February 2017; Received in revised form 26 April 2017; Accepted 14 June 2017
Available online 15 June 2017

Table 1
Nucleotide and protein sequences of Myf6, MyoD and Myf5 sequences from different organisms used for multiple sequence alignment and phylogenetic analysis respectively. Sequence No.1 in both categories is from present study and others downloaded from NCBI.

Sl. no.	Protein/ nucleotide sequences	Organism	Accession number/ source
Myf6 nucleotide sequences used for multiple alignment and other bioinformatic analyses			
1		*Tenualosa ilisha*	Present study
2		*Ictalurus punctatus*	XM_017493745.1 (NCBI)
3		*Oreochromis niloticus*	JQ246950.1 (NCBI)
4		*Salmo salar*	BT057306.1 (NCBI)
Protein sequences used for phylogeny			
1	Myf6	*Tenualosa ilisha*	Present study
2		*Siniperca chuatsi*	AFW89953.1 (NCBI)
3		*Larimichthys crocea*	KKF14207.1 (NCBI)
4		*Monopterus albus*	AIS22055.1 (NCBI)
5		*Sparus aurata*	AEV53630.1 (NCBI)
6		*Oreochromis niloticus*	AFN20594.1 (NCBI)
7		*Tetraodon nigroviridis*	AAS88438.1 (NCBI)
8		*Takifugu rubripes*	NP_001027943.1 (NCBI)
9		*Salmo salar*	ABB02375.1 (NCBI)
10		*Megalobrama amblycephala*	AHW49180.1 (NCBI)
11		*Cyprinus carpio*	ADC38865.1 (NCBI)
12		*Danio rerio*	NP_001003982.1 (NCBI)
13		*Homo sapiens*	NP_002460.1 (NCBI)
14	MyoD	*Oreochromis aureus*	ADA84041.1 (NCBI)
15		*Oreochromis niloticus*	NP_001266649.1 (NCBI)
16		*Sparus aurata*	AAL85337.1 (NCBI)
17		*Monopterus albus*	AIS22059.1 (NCBI)
18		*Oreochromis niloticus*	ADA84042.1 (NCBI)
19	Myf5	*Megalobrama amblycephala*	AHW49179.1 (NCBI)
20		*Ctenopharyngodon idella*	ADB56965.1 (NCBI)
21		*Danio rerio*	AF253470_1 (NCBI)
22		*Oncorhynchus mykiss*	NP_001118001.1 (NCBI)
23		*Salmo salar*	NP_001117116.1 (NCBI)
24		*Takifugu rubripes*	NP_001027942.1 (NCBI)
25		*Tetraodon nigroviridis*	ABE11537.1 (NCBI)
26		*Paralichthys olivaceus*	ABI96686.1 (NCBI)
27		*Monopterus albus*	AIS22057.1 (NCBI)
28		*Siniperca chuatsi*	AHB18042.1 (NCBI)
29		*Larimichthys crocea*	KKF14210.1 (NCBI)
30		*Morone saxatilis*	AF463525_1 (NCBI)

(24° 49′ 0″ N, 87° 54′ 0″ E) India, stored in LN_2 was used for total RNA extraction, by guanidinium thiocyanate-phenol-chloroform extraction (Trizol) method. Paired end library was prepared and sequenced through Illumina HiSeq platform and the sequences (unpublished) were processed and denovo assembled with Assemblerv1.3 (Zheng et al., 2011). The high quality assembled sequences were annotated by Blast2GO tool (www.blast2go.com) against non redundant, SwissProt and UniProt databases with blastX algorithm and e value of e^{-5}. The annotated sequences were mapped against Gene Ontology (GO) database to assign GO terms. The annotations were manually screened based on literature for the presence of muscle growth related genes. The identified gene was further analysed for their molecular characteristics.

2.2. Phylogenetic analysis

For phylogenetic analysis of TiMyf6 in *T. ilisha* and the protein sequences of Myf6, MyoD and Myf5 from different fishes and Myf6 of *Homo sapiens* were obtained from NCBI (http://www.ncbi.nlm.nih.gov/Entrez/) (Table 1) and neighbour-joining method (Saitou and Nei, 1987) was applied for construction of phylogenetic tree, with boot strap 1000 replicates and evolutionary distance by Poisson correction method (Zuckerkand and Pauling, 1965) in MEGA 6 (Tamura et al., 2013). All missing data and alignment gaps were deleted pair-wise.

2.3. Characterization of TiMyf6 and its ORF

Complete ORF of TiMyf6 was identified with genescan (http://genes.mit.edu/GENSCAN.html) and expasy tool (http://www.expasy.org). Physicochemical properties of Myf6 transcript including molecular weight, amino acid composition, theoretical isoelectric point (pI), extinction coefficient (EC), instability index (II) and grand average hydropathy (GRAVY) were computed using the Expasy's ProtParam server (http://www.expasy.org). Phosphorylation sites for serine, threonine and tyrosine were determined with NetPhos 3.1 (www.cbs.dtu.dk/services/NetPhos/). Presence of disulfide bonds in TiMyf6 were identified using DiANNA 1.1 web server (http://clavius.bc.edu/~clotelab/DiANNA/), Superfamily and functional domains by Superfamily 1.75 (http://supfam.org/SUPERFAMILY) and SMART tool (http://smart.embl-heidelberg.de) respectively. Subcellular localization and potential signal peptide of TiMyf6 was predicted through PSORT (http://psort.hgc.jp/) and SignalP 4.1 (http://www.Cms.dtu.dk/services/SignalP/), respectively.

2.4. Secondary structure prediction and homology modeling

Secondary structure predictions were conducted using the PDBsum at EMBL server (Laskowski et al., 1993) and solvent accessibility by the Predict Protein Server (http://www.predictprotein.org) (Rost et al., 2004). Homology modeling was performed for TiMyf6 of *T. ilisha*, and compared to reference models of TiMyf6 proteins of *I. punctatus*, and *S. salar* and *O. niloticus* by SWISS-MODEL server (Arnold et al., 2006). MYOD bHLH domain was used as template for generating predicted models for Myf6 of all four fishes. Quality of protein model was estimated by PROCHECK tools (Laskowski et al., 1993). Protein-Protein interaction prediction of myf6 was performed on STRING 9.1 server (http://string-db.org/). Interactions with score < 0.3 were considered as low confidence, scores ranging from 0.3 to 0.7 were classified as medium confidence and scores > 0.7 yield high confidence (Szklarczyk et al., 2011).

2.5. Docking

Hex server software (http://hex.loria.fr/) was used for docking of deduced Myf6 Protein of 4 fishes, *T. ilisha*, *I. punctatus*, *and S. salar* and *O. niloticus* with default parameters. As Copper is a main element affecting blood vessel growth and development in muscle (Mroczek-Sosnowska et al., 2015), the Copper ion (Cu^{2+}) was used as Ligand for docking (obtained from NCBI PubChem) Ref No. 1A2V (Li et al., 1997). LIgplot 1.4.5 (Wallace et al., 1995) was used for visualization of molecules involved in protein Myf6 ligand (Cu^{2+}) interaction.

2.6. Identification of single nucleotide polymorphism (SNP)

To find out SNPs in Myf6 protein between slow and fast growing fish species (Table 1) the Myf6 sequences were used. *I. punctatus* (NCBI acc. No. XM_017493745.1) (Dunham and Brummett, 1999) and *T. ilisha* of present study (Miah, 1997) of slow growth; *O. niloticus* (NCBI acc. No. JQ246950.1) (Damle and Chari 2011) and *S. salar* (NCBI acc. No. BT057306) (Valente et al., 1999) of faster growth were compared and multiple sequence alignment was performed through Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).