博士論文

Molecular biological studies on myomiRs and their host myosin heavy chain genes underlying fish muscle development and growth

（魚類筋肉の発生と成長過程で働く myomiR とその宿主ミオシン重鎖遺伝子に関する分子生物学的研究）

Bhuiyan Sharmin Siddique

ブイヤン　シャーミン　シッディク

Molecular biological studies on myomiRs and their host myosin heavy chain genes underlying fish muscle development and growth

（魚類筋肉の発生と成長過程で働く myomiR とその宿主ミオシン重鎖遺伝子に関する分子生物学的研究）

A

Thesis submitted to

Graduate School of Agricultural and Life Sciences

The University of Tokyo

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in

Department of Aquatic Bioscience

By

Sharmin Siddique Bhuiyan

January, 201

Declaration

I, Bhuiyan Sharmin Siddique, hereby declare that the thesis entitled **"Molecular biological studies on myomiRs and their host myosin heavy chain genes underlying fish muscle development and growth"** is an authentic record of the work done by me and that no part thereof has been presented for the award of any degree, diploma, associateship, fellowship or any other similar title.

15th December, 2014

Bhuiyan Sharmin Siddique

Laboratory of Aquatic Molecular biology and Biotechnology

Department of aquatic Bioscience

Graduate school of Agricultural and Life Sciences

The University of Tokyo.

Acknowledgements

First of all I would like to express my heartfelt gratitude to my honorable supervisor Professor Shuichi Asakawa, laboratory of Aquatic Molecular Biology and Biotechnology, The University of Tokyo. His continuous teaching, monitoring, progress discussion and scholastic direction gave me clear view of understanding my research and finally for correction and critically going through this manuscript. I express my sincere thanks to Professor Shugo Watabe, for his intellectual guidance all through the doctoral research period. He extended his help in every step to perform a good research and guided me in right track. I express my gratitude to Professor Hideki Ushio, Laboratory of Marine Biochemistry, The university of Tokyo, who cared by contributing valuable remarks and necessary direction for my research.

I would like to thank to Associate Professor Shigeharu Kinoshita, laboratory of aquatic molecular biology and biotechnology, the University of Tokyo, without his guidance this research would have been incomplete. He helped in planning and implementation of this research by teaching me necessary techniques and spent his valuable time in correction and changes required for betterment of my research. I thoroughly express my gratitude to him for his endless efforts in my research.

I would like to thank to Assistant Professor Gen Kaneko, laboratory of marine biochemistry, the University of Tokyo who extended his help to teach new techniques as well as putting valuable comments to explore my research. I would like to express gratitude to lab technician Dr. Misako Nakaya who supported me throughout by providing time to time all the necessary chemicals and laboratory equipment. Lastly, I would like to thank to all the members of laboratory of marine biochemistry and laboratory of molecular biology and biotechnology, The university of Tokyo for their help and co-operation during my entire study period.

Contents

Abstract

Skeletal muscle consists of various type of muscle fibers such as slow and fast ones, where muscle fiber-type specification is crucial for the development and growth of skeletal muscle. Fish skeletal muscle is an attractive model to study the mechanisms underlying muscle fiber-type specification because slow and fast muscles are segregated in the trunk myotome. Myosin is the major contractile protein in muscle tissues, which consists of two heavy chains (myosin heavy chains, MYHs) and four light chains. MYH gene (*MYH*) is a multigene family and different expression of *MYHs* leads to the formation of different muscle fiber-types. Among *MYH* family genes, three *MYHs* named *MYH6*, *MYH7*, and *MYH14* has been characterized by existence of microRNA (miRNA) in their introns. These MYH-encoded intronic miRNAs are called as myomiRs. Emerging evidence has demonstrated that the genomic positions and expression patterns of myomiRs and their host *MYHs* are well conserved in mammals and they form an important transcription network involved in muscle fiber-type specification. However, functional analysis of myomiRs and their host *MYHs* as well as their genomic distribution and expression during teleost myogenesis have not been studied in detail. In the present study, distribution of myomiR/MYH loci and their expression patterns were examined with special emphasis on three representative teleosts, torafugu *Takifugu rubripes*, zebrafish *Danio rerio*, and medaka *Oryzias latipes*. Using available genome databases for different vertebrates, the syntenic organization of human *MYH14* and miR-499 with their orthologues was examined (chapter 2). In teleost genome, *MYH14*/miR-499 showed highly diverged structure and the miR-499s phylogenetic relationships were consistent with those of the *MYH14*s. To address expression of *MYH14*/miR-499 *in situ* hybridization performed in the three teleost species. Interestingly, miR-499 expression is exceptionally conserved regardless of the varied expression of their host *MYH14s* (chapter 2). In teleosts, known major cardiac *MYH* isoform, ventricular myosin heavy chain gene (*vmhc*) contains an intronic miRNA, miR-736. Sequence similality and phylogenetic analyses indicates *vmhc*/miR-736 are orthologue of *MYH6*/miR-208a. As well as *MYH14*/miR-499, syntenic and phylogenetic studies revealed that multiple orthologues of *MYH6*/*vmhc*/miR-736 are present in teleost genomes (chapter 3). To address mechanisms of expression regulation of diversified *MYH14* paralogues, *in vivo* reporter assay and their function in muscle fiber-type specification was also examined by knock down analysis was performed (chapter 4). Deletion of the conserved regions significantly reduced the promoter activity of *MYH14-3* but no affect on that of *MYH14-1*, indicating that *cis*-regulatory elements of *MYH14-1* and *MYH14-3* are different in accordance with differential expression between the two *MYHs*. Loss of function experiment of miR-499 was performed in medaka and zebrafish. As expected, knockdowned larvae showed marked reduction of slow muscle fibers in zebrafish and medaka developmental stages (chapter 4). Despite diversification of host *MYHs* in genomic organization and expression patterns, miR-499 expression was exceptionally conserved, indicating pivotal role of the myomiR in teleost muscle formation. Actually, knock down analysis of miR-499 showed perturbation in slow muscle formation during zebrafish/medaka growth, indicating that a myomiR-mediated regulatory network also works in fish muscle formation.

Part of this research is published as follows:

Bhuiyan SS, Kinoshita S, Wongwarangkana C, Asaduzzaman M, Asakawa S. and Watabe S. Evolution of the myosin heavy chain gene *MYH14* and its intronic microRNA miR-499: muscle-specific miR-499 expression persists in the absence of the ancestral host gene. BMC evolutionary biology, **13**:122, 2013. doi:10.1186/1471-2148-13-142.

Abbreviations

ANOVA : Analysis of variance
ATP : Adenosine 5'-triphosphate
bp : Base pair
cDNA : Complementary DNA
CNS : Conserved region
Ct : Cycle threshold
DAPI : Diamidine-20-phenylindole dihydrochloride
DNA : Deoxyribonucleic acid
DIG : Digoxigenin
dpf : Days post fertilization
ED : Erector and depressor
EGFP : Enhanced green florescence Protein
EM : Epaxial muscle
Hh : Hedgehog
hpf : Hours post fertilization
HM : Hypaxial muscl
LS : Lateralis superficialis
miRNA : MicroRNA
MYH : Myosin heavy chain
*MYH*s : Myosin heavy chain genes
NC : Notochord
NADH : Nicotinamide adenine dinucleotide reduced
NBT : Nitro blue tetrazolium chloride
PBSTw : Phosphate-buffered saline with 0.1% tween 20
PCR : Polymerase chain reaction
PFA : Paraformaldehyde
qRT-PCR : Quantitative real-time polymerase chain reaction
RACE : Rapid amplification of cDNA ends
RNA : Ribonucleic acid
RNase : Ribonuclease
RT-PCR : Reverse transcription PCR
SSC : Saline-sodium citrate
SPSS : Statistical package for social science
TBSTw : Tris-buffered saline with 0.1% tween 20
TEEA : The transient embryonic excision assay

TFsearch : Transcription factor search

UTR : Untranslated region

CHAPTER 1

General Introduction

General Introduction

The physiological properties of skeletal muscle are highly flexible in response to internal and external stimuli by switching muscle fiber-types through a process termed muscle fiber-type plasticity (Schiaffino S et al., 2011). Vertebrate skeletal muscle is a heterogeneous tissue which consists of various types of muscle fibers. Muscle fiber type-specification is crucial for the development and growth of skeletal muscle. To study muscle fiber type specification mechanisms, we have focused on teleosts, which is an attractive model to study muscle fiber type specification mechanism. In mammals, most of the skeletal muscle contains a mix of various fibers-types. On the basis of the predominant expression of a particular myosin isoforms and other contractile proteins that determine the speed of contraction (Bassel-Duby and Olson, 2006) muscle have two distinct sub fiber-types, (i.e, type I and type II fibers). Type I fibers display low velocity of shortening and high resistance to fatigue and therefore, termed as slow-twitch fibers. Type II is termed as fast-twitch fibers due to exertion of quick contraction and rapid fatigues (Schiaffino and Reggiani, 1996; Pette and Staron, 2000; Wigmore and Evans, 2002). Contrast to mammals where muscle fibers are organized in mosaic pattern, fiber types are anatomically separated in fish. In fish, the fast twitch fibers are located deep inside the myotome and slow twitch fibers at the lateral end of the myotome and along the horizontal myoseptum region (Sanger and Stoiber, 2001). Myosin is the major contractile protein in muscle tissues, the molecule which contains two heavy chains called myosin heavy chains, (MYHs of 171-244kDa) and two pairs of light chains (Mabuchi and Okino, 1977; Kiehart et al., 1982;

Mooseker and Cheney, 1995). The class II myosin is the major contractile protein in muscle tissues. In vertebrate, on the basis of their motor domain or tail domain, the myosin class IIs can be divided into two types, skeletal/cardiac muscle (sarcomeric) myosins and vertebrate smooth muscle/nonmuscle myosins. *MYHs* form a large gene family that includes sarcomeric MYHs, major contractile proteins of striated muscles that are expressed in a spatio-temporal manner defining the functional properties of different muscle fiber-types (Schiaffino S et al., 2011).

A comprehensive analysis of the human genome revealed two groups of sarcomeric (skeletal and cardiac) MYHs (Weiss et al., 1999a; Weiss et al., 199b). One group contains two tandemly arrayed cardiac MYHs, MYH6 and MYH7, the latter being also expressed in slow skeletal muscle. The other group contains six skeletal muscle MYHs, including adult fast IIa, IIx, IIb, embryonic, neonatal/perinatal, and extra ocular muscle types. Among these *MYHs*, *MYH6*, *MYH7* encoded by MYH-α, MYH-β respectively and *MYH14* (also called *MYH7b*) occupy a unique position because of their specific expression in cardiac and slow muscle fibers, and the existence of miRNAs in their intron. These three muscle-specific myosin genes encoded by the family of intronic miRNAs are called as MyomiRs. In rodents, *α-MYH, a fast ATPase* encodes miR-208a in its 27 intron, which is expressed specifically in the adult heart. *β-MYH, a slow ATPase* encode a closely related miRNA, miR-208b in its 31 intron which is co-expressed with its host gene, showing highest expression in type I myofibers and highly expressed in the developing heart and is down-regulated after birth (Morkin, 2000; van Rooij et al., 2009). In human, *MYH7* is predominantly

expressed in the ventricle, whereas expression of *MYH6* is restricted in the atrium (Everett, 1986). A third member of this myomiRNA gene family, *MYH14*, the recently identified *MYHs*, is encoded by miR-499, (McGuigan et al., 2004). In mammals, *MYH14* contained miR-499 is co-expressed abundantly in the heart and are enriched in the slow muscle (Bell et al., 2010).

Emerging evidence demonstrates a complex network of myomiRs-post-transcriptional regulated gene expression coordinates the overall muscle development. In rodents, miR-499 suppresses the expression of genes involved in muscle fiber-type specification by repressing fast and promoting slow muscle (van Rooij et al., 2009; Bartel DP, 2004; McCarthy JJ et al., 2009; Hagiwara N et al., 2007; von Hofsten J et al., 2008). Experiments have shown, miR-208b or miR-499 plays redundant function and double knockout of these duo miRNAs showed a marked slow-to-fast switch in fiber-type profile and MYHs expression in rodents. Conversely, muscle-specific overexpression of miR-499 leads to a complete fast-to-slow switch in the MYH profile (van Rooij et al., 2008, van Rooij et al., 2009). In mammals, miR-208a regulates the expression of two slow myosins and their intronic miRNAs, *MYH7*/miR-208b and *MYH7b*/miR-499, respectively. Thus, these myo-miRNAs seemingly acts to support normal slow-muscle formation. As schematized in figure 1, the regulatory steps leading to myofiber diversification in rodents is shown.

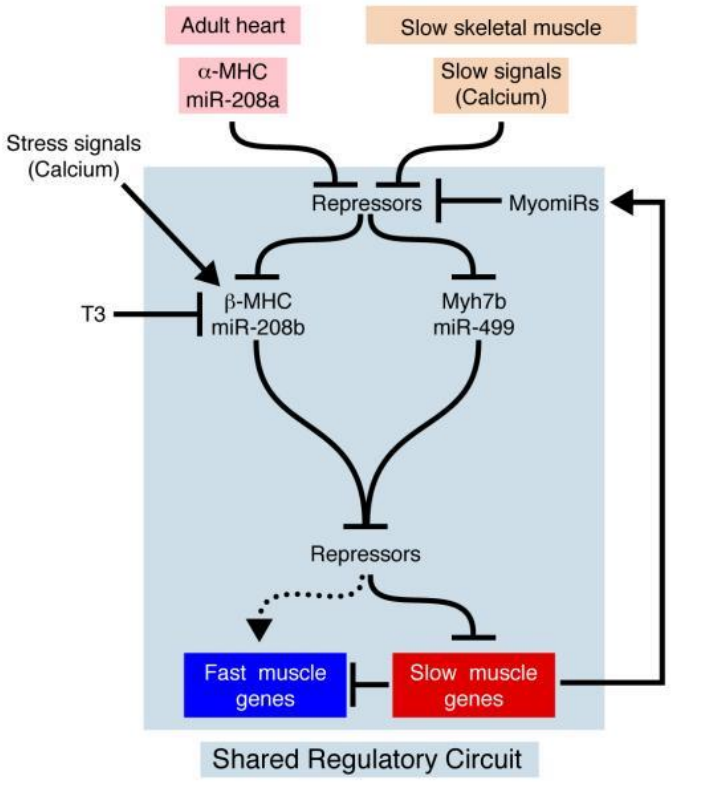

Fig.1 Control of myosin and fast versus slow muscle gene expression by the MyomiR network in rodents (Adopted from van Rooij et al., 2009)

The slow myofiber specific gene activation in skeletal muscle depends on calcium signaling through calcineurin and various kinases (Bassel-Duby and Olson, 2006). It has been already proposed that *MYH7* and *MYH14* are independently activated through such signals (Wu et al.,2000), resulting in the production of their encoded myomiRs, which then reinforce the slow skeletal muscle gene program through the actions of their targets. This model also proposed the existence of a positive feedback circuit whereby the targets of miR-208b and miR-499 in slow skeletal muscle enhance their own expression via their targets, which act on their host myosins. Recent studies have demonstrated that Sox6, which acts as

slow muscle specific transcriptional repressor binds to conserved cis-regulatory elements in slow-twitch fiber genes to represses their transcription in adult fast-twitch muscle (Quiat et al., 2011). miR-499 has been suggested to play a pivotal role in lineage-specific activity of Sox6 in mammals based upon the finding that the 3′UTRs of human, mouse and rat *Sox6* contain consensus recognition sites for the miR-499 and the demonstration that miR-499 reduces *Sox6* mRNA levels both in vitro and in vivo (McCarthy et al., 2009; van Rooij et al., 2009; Bell et al., 2010). It has been revealed that similar to mammals, the functionally conserved zebrafish miR-499 reinforces Sox6/*MYH14*/miR-499 transcriptional network by repressing Sox6 translation in slow muscle progenitor (Wang et al. 2011). Such a mechanism also enables myomiRs to function as conserved binary regulators of slow versus fast muscle gene programs, activating slow and repressing fast through their downstream targets. Thus, these myomiRs and their host myosin genes are well conserved and they form an important transcription network which works in muscle fiber-type specification.

In contrast to mammalian counterpart, highly diverged MYH multigene family is reported for fish (Gerlach et al., 1990; Kikuchi et al., 1999; Watabe and Ikeda, 2006; and Ikeda et al., 2007). In case of common ancestor of vertebrates, 2 rounds of whole genome duplication (WGD) occurred and a third WGD occurred in fish lineage (Amores et al., 1998; Elgar et al., 1999; Postlethwait et al., 2000; Woods et al., 2000; Smith et al., 2002; McGuigan et al., 2004), resulting in much higher number of *MYHs* and suggesting their function and expression patterns also have diverged in fish musculature. Unexpectedly,

the genome mapping of fish muscle miRNAs revealed that cartilaginous and ray-finned fish genomes do not retain the MYH7/miR-208 gene whereas it exists as single-copy in lobe-finned fishes (Pedro Gabriel Nachtigall, 2014). MYH7 was actually absent into all fish genomes, whereas MYH6 persists but missing the intronic miR-208, except for coelacanth that retain the miR-208 intronic to MYH6 (Pedro Gabriel Nachtigall, 2014). In teleosts, a major cardiac MYHs isoform, ventricular myosin heavy chain (*vmhc*) encodes an intronic miRNA, miR-736. Studies in other vertebrates have shown that miR-736 belongs to the 208 family of miRNA genes specifically expressed in cardiac tissue and can be referred as myomiRNA (Andreassen R et al., 2013). However, the evolutionary history of these ancient MYHs and their myomiRNAs has not been studied in detail as well as their expressional and functional analysis is still limited in case of teleosts.

In this study, multiple *MYH14*/miR-499 and *MYH6*/*vmhc*/miR-736 loci on various teleost fish genomes were identified and their evolutionary history and distribution was examined by synteny and phylogenetic analyses. Divergent distribution of *MYH14*/miR-499 and *MYH6*/*vmhc*/miR-736 locous among teleosts was revealed. Despite varied distribution and expression of myomiRs and their host MYHs, miR-499 expression was exceptionally conserved, indicating its pivotal role in teleost muscle formation. In my findings, miR-499 knockdown caused a marked reduction of slow muscle in larval zebrafish and medaka. Thus, investigation of myomiRs and their host myosin heavy chain genes identified a miRNA-mediated regulatory network in fish muscle fiber-type specification.

Objectives of the study

Different expression of MYH leads to the formation of different muscle types such as fast, slow and cardiac muscle ones. In mammals, the three cardiac and or slow muscle specific myosin genes, *MYH6*, *MYH7*, and *MYH14*, encoded by related myo (myo=muscle) microRNAs (miRs) control muscle myosin content as well as myofiber identity in mammals. The myomiRNAs conserved from fish to human genomes suggests, the ancestral *MYH14* contained the ancestral myomiR in an intron and when that myosin gene duplicated, its intronic myomiRs are also duplicated (van Rooij et al., 2009). Sequence analysis of mammalian *MYHs* also indicates *MYH14* is the most ancient of this trio of myosins and its gene duplications gave rise to *MYH6* and *MYH7 (van Rooij et al, 2009).* Interestingly, teleost fish genomes contain multiple *MYH14* and miR-499 paralogs. However, in teleost, the evolutionary history of the myomiRNAs encoding *MYHs*/has not been studied in detail. In our present study, we focused on the following topics-

- Genomic distribution and expression of *MYH14* /miR-499 in teleosts
- Genomic distribution and expression of *MYH6*/*vmhc*/miR-736 in teleosts
- Expression regulation of *MYH14* paralogues/myomiRs and function analysis of miR-499 in muscle fiber-type specification mechanism in teleosts.

Chapter 2

Genomic organization and expression of *MYH14*/miR-499 in teleosts

The partly content of this paper is published as:

Bhuiyan SS, Kinoshita S, Wongwarangkana C, Asaduzzaman M, Asakawa S. and Watabe S. Evolution of the myosin heavy chain gene *MYH14* and its intronic microRNA miR-499: muscle-specific miR-499 expression persists in the absence of the ancestral host gene. BMC evolutionary biology, **13**:122, 2013. doi:10.1186/1471-2148-13-142.

Abstract

A novel sarcomeric myosin heavy chain gene, *MYH14*, was identified following the completion of the human genome project. *MYH14* contains an intronic microRNA, miR-499, which is expressed in a slow/cardiac muscle specific manner along with its host gene; it plays a key role in muscle fiber-type specification in mammals. Interestingly, teleost fish genomes contain multiple *MYH14* and miR-499 paralogues. However, the evolutionary history of *MYH14* and miR-499 has not been studied in detail. In the present study, we identified *MYH14*/miR-499 loci on various teleost fish genomes and examined their evolutionary history by sequence and expression analyses. Using available genome databases for different vertebrates, the syntenic organization of human *MYH14* and miR-499 with their orthologues was examined. In teleost genome, *MYH14*/miR-499 showed highly diverged structure. Phylogenetic analysis showed that *MYH14* was monophyletic in the amniote lineage, including humans, chickens, and coelacanths, but was duplicated in the ray-finned fish lineage, except for the spotted gar. The miR-499s phylogenetic relationships were consistent with those of the *MYH14*s. *MYH14*/miR-499 locus was duplicated early in teleost evolution by teleost specific whole genome duplication and one of the duplicated miR-499 gene was lost in the common ancestor to cod and the Acanthopterygii, after the split from the zebrafish lineage. Additionally, *MYH14*s have seemingly been lost at independent points of teleost evolution. Interestingly, miR-499 was not located in the *MYH14* introns of certain teleost fish. In medaka, miR-499 is present but *MYH14* is completely absent in the genome. An

MYH14 paralogue, lacking miR-499 exhibited an accelerated rate of evolution compared with those containing miR-499, suggesting a putative functional relationship between *MYH14* and miR-499. In this study, expression of *MYH14-1* and *MYH14-3* was observed whereas transcripts of *MYH14-2* were not detected in both developmental and adult stages. The transcripts of *MYH14-3* were specifically expressed in slow and cardiac muscles at both developmental and adult stages. These results indicate that teleost *MYH14*s are highly diverged in their genomic structure and expression patterns. *In situ* hybridization showed that miR-499 was not expressed in the skeletal muscle at the embryonic stage of zebrafish, torafugu, and medaka. On the other hand, larvae of the three fish showed a clear expression of miR-499 in both cardiac and slow muscles. Similar to the larval stage, adult zebrafish and torafugu showed high expression of miR-499 in cardiac and slow muscles, whereas adult medaka exhibited miR-499 expression only in the cardiac muscle. In torafugu, miR-499 was also detected in the erectors-depressors (ED) muscle which exhibit characteristics of slow muscle fibers. These miR-499 expression patterns were also confirmed by next-generation sequencing of small RNA libraries. It is noted that medaka miR-499 was even expressed in the absense of its host gene *MYH14*. Comparing the flanking sequences of *MYH14*/miR-499 loci between torafugu, zebrafish, and medaka revealed several highly conserved regions including an intronic sequence immediately downstream of miR-499, suggesting *cis*-regulatory elements have been functionally conserved in medaka miR-499 despite the loss of its host gene.

Background

To meet the constantly changing functional demands, the physiological properties of skeletal muscle are highly adjustable and are achieved through a process of switching muscle fiber-types, such as slow and fast muscle fibers, in response to internal and external stimuli, a process termed muscle fiber-type plasticity (Schiaffino and Reggiani, 2011). Myosin heavy chains (MYHs) form a large gene family that includes sarcomeric MYHs, major contractile proteins of striated muscles that are expressed in a spatio-temporal manner defining the functional properties of different muscle fiber subtypes (Schiaffino and Reggiani, 2011). In humans, sarcomeric *MYHs* form two clusters on the genome where skeletal and cardiac *MYHs* are arrayed in tandem on chromosome Chr17 and Chr14, respectively (Mahdavi et al., 1984; Saez et al.,1887; Weiss et al.,1999 and Shrager et al., 2000). Upon completion of the human genome project, a novel *MYH* named *MYH14* (*MYH7b*) was identified on Chr20 (Desjardins et al., 2002), recently, there has been increasing interest in its direct involvement in muscle fiber-type plasticity. Mammalian *MYH14* has a microRNA, miR-499, in its 19th intron that suppresses the expression of genes involved in muscle fiber-type specification (van Rooij et al., 2009; Bartel, 2004; McCarthy et al., 2009; Hagiwara et al., 2007; von Hofsten et al., 2008).); thus, miR-499 seemingly acts to support normal slow-muscle formation in mammals.

Our previous studies revealed that teleost fish also have *MYH14* in their genomes (Watabe and Ikeda, 2006; Ikeda et al., 2007). Expression analysis in torafugu *Takifugu rubripes* Abe 1949 and zebrafish

Danio rerio Hamilton 1822 revealed that *MYH14* is one of the major components of the *MYH* repertoire expressed in the slow and cardiac muscles of teleost fish (Akolkar et al, 2010; Kinoshita et al, 2011), suggesting its role in teleost muscle formation. Consistent with functional conservation with mammals, Wang et al. (2011) showed that the transcriptional network of Sox6/*MYH14*/miR-499 plays an essential role in maintaining slow muscle lineage in larval zebrafish muscle. Our previous study also showed that teleost fish contain a higher number of *MYHs* in their genomes than do their mammalian counterparts (Gerlach et al., 1990; Kikuchi et al., 1999; Watabe and Ikeda, 2006; and Ikeda et al., 2007)). Two *MYH14* paralogues, MYH_{M3383} and MYH_{M5}, were identified in the torafugu genome by phylogenetic and syntenic analyses (Ikeda et al., 2007). Moreover, we have also previously found that medaka *Oryzias latipes* lacks *MYH14* in the syntenic region (Kinoshita et al., 2011). These lines of evidence allowed to speculate on the existence of a highly varied distribution and function of *MYH14* and miR-499 in teleost fish.

The aim of this study was to elucidate the evolutionary history of *MYH14*/miR-499 in fish. *MYH14* and miR-499 genes were screened from available vertebrate genome databases, and their evolutionary history was examined by synteny and phylogenetic analyses. Interestingly, three sets of *MYH14*/miR-499 paralogs located in different loci are present in zebrafish but their expression analysis is still limited.

The tissue distribution of the transcripts of *MYH14/*miR-499 paralogues was examined at embryonic, larval and adult stages. Interestingly, miR-499 expression is exceptionally conserved regardless of the

varied expression of their host *MYH14s*.

Methods

Fish

All procedures in this study were performed according to the Animal Experimental Guidelines for The University of Tokyo. Live adult zebrafish, torafugu and medaka specimens (average body weight of 1.4g, 1kg and 0.78 g) were reared in local tap water with a circulating system at 28.5°C under a 14:10-h light-dark photoperiod, at a fish rearing facility in the Department of Aquatic Bioscience, The University of Tokyo. Tissue for RNA extraction was dissected after instant euthanasia by decapitation and stored in RNAlater (Invitrogen, San Diego, CA, USA). Embryos of zebrafish were obtained by natural spawning raised at The University of Tokyo and reared up to larval stages. The developmental stage was determined by dpf and by morphological criteria. Whole embryos, larvae and adult tissues from zebrafish were used for *in situ* hybridization. For torafugu larvae, artificially fertilized eggs collected from Oshima Fisheries Hatchery, Nagasaki, Japan, were brought to The University of Tokyo, and reared in 500 L tank in seawater at 18°C. Eggs of 6 days post fertilization (dpf) at embryonic stage and 8 dpf at larval stage were collected for further analysis. Whole embryos of 6 dpf, larvae of 8 dpf and adult tissues of torafugu were used for *in situ* hybridization.

Construction of a physical map around MYH14 *and miR-499*

The Ensembl genome browser (http://www.ensembl.org/index.html) was used to determine the syntenic organization in the region surrounding *MYH14* and/or miR-499 in vertebrates. The database versions used were as follows: human (GRCh37), chicken (Galgal4), coelacanth *L. chalumnae* (LatCha1), zebrafish *D. rerio* (Zv9), torafugu *T. rubripes* (FUGU4), green spotted puffer *T. nigroviridis* (TETRAODON8), tilapia *O. niloticus* (Orenil1.0), Atlantic cod *G. morhua* (gadMor1), stickleback *G. aculeatus* (BROADS1), platyfish *X. maculatus* (Xipmac4.4.2), and medaka *O. latipes* (MEDAKA1). The pre Ensembl browser (http://pre.ensembl.org/index.html) was used for analysis of Spotted gar *L. oculatus* (LepOcu1).

Bioinformatics analysis

The *MYH14* and miR-499 sequence data were retrieved from the available genome databases mentioned above (Table 1). NJ and ML trees were constructed on the basis of the *MYH14* coding and miR-499 stem-loop sequences using MEGA5 [39] with 1000 bootstrap replications. The Nei and Gojyobori method [40] (Jukes-Cantor) was employed to consider synonymous and non-synonymous substitutions for the *MYH14* NJ tree. The Tajima-Nei model [41] was employed for the miR-499 NJ tree, whereas the Tamura-Nei model [42] was used for the *MYH14* and miR-499 ML trees. Putative secondary structures of the miR-499 from medaka and torafugu stem-loop sequences and that of the *C. elegans* mirtron miR-62 (miRBase accession number: MI0000033) were predicted using the RNA fold

program CentroidFold (http://www.ncrna.org/centroidfold).

Small RNA library construction and sequencing

Total RNA was extracted from the muscle, intestine, eye, brain, heart, ovary, and testis of adult torafugu and medaka using a mirVana™ miRNA Isolation Kit (Applied Biosystems, Foster City, CA, USA). Small RNAs (less than 40 nucleotides in size) were purified from total RNA using a flashPAGE™ Fractionator (Applied Biosystems), and the small RNA libraries were constructed according to the manufacturer's instructions. Library sequencing was performed with SOLiD™ next-generation sequencer (Applied Biosystems). After elimination of low-quality reads using perl scripts of our own design, 102, 602, 452 reads of 35 nucleotides were obtained. The 18–25 nucleotide reads were subjected to a Blast search against known mature miRNA sequences deposited in miRBase 18.0 (www.mirbase.org/). The sequences with their seed regions (2–8 nucleotides from the 5′-end) showing 100% identity to those of known mature miR-499 sequences were annotated as miR-499. Gene expression was represented as reads per million (RPM), which corresponds to (total reads of a given gene/total reads in the tissue) × 10^6. Sequence data sets used in this study were deposited at the DDBJ Sequence Read Archive under the accession number DRA001039 and DRA001040.

Preparation of total RNA

Total RNAs were extracted from adult zebrafish, torafugu and medaka tissues by using ISOGEN solution (Nippon Gene, Tokyo, Japan). RNA concentrations were determined at 260 nm using a Nanophotometer^{TM} 95 (Implen GmbH, Munich, Germany) and the RNA quality was determined by electrophoresis in a denaturing agarose gel containing 50% formamide. Total RNA pellets were stored at −80°C.

Reverse transcription-PCR

To examine expression of zebrafish *MYH14* paralogs, RT-PCR was performed. First strand cDNAs were prepared by reverse transcription of 5 μg total RNAs with the use of Gene Racer oligonucleotide (dT)-tailed primer (Table 1) and SuperscriptTM III reverse transcriptase (Invitrogen, Carlsbad, USA). For the expression analysis of zebrafish *MYH14*paralogs1 and MYH14paralog2, primers of MYH14paralog1F, MYH14paralog1R and MYH14paralog2F and MYH14paralog2R were designed respectively, for the amplification of the 3' untranslated regions (UTR) (Table 1). First-strand cDNAs from larvae at 2 dpf as well as from heart, brain, intestine, and slow and fast muscles of adult zebrafish were used as the template. PCR amplifications were carried out with the first denaturation at 96°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 1 min, with a final extension step at 72°C for 5 min. A single PCR reaction of 20 μl contained 1 μl each of forward and reverse primers (10 mM), 4 μl of cDNA template (dilution 1:100), 2 μl of 10x

PCR buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 20 mM MgCl2), 1 U *Taq* DNA polymerase (Applied Biosystems, Carlsbad, USA) and 13.8 μl of sterilized water. PCR products were analyzed by 1% agarose gel electrophoresis.

RNA probe synthesis

The PCR products in RT-PCR was subcloned and sequenced by using pGEM-T vector system (Promega, Madison, USA) and an ABI 3100 genetic analyzer (Applied Biosystems) after labeling with ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). After PCR amplification with T7 (5' TAATACGACTCACTATAGGG-3') and SP6 (5' ATTTAGGTGACACTATAGAA-3') primers and the plasmid DNA template, sense and anti-sense RNA probes were synthesized using the PCR product as a template for *in vitro* transcription with T7 and SP6 RNA polymerases (Roche, Basel, Switzerland), respectively.

In situ *hybridization with RNA probe*

Whole embryo, larva, adult zebrafish, torafugu and medaka trunk muscles and cardiac samples were fixed in 4% paraformaldehyde (PFA) and *in situ* hybridization was performed on transverse sections (14 μm thickness) by the method of Darby et al. (2006). The samples were re-hydrated with 75% methanol/phosphate buffered-saline with 0.1% tween 20 (PBSTw), 50% methanol/PBSTw and 25%

methanol/PBSTw consecutively at room temperature for 5 min each. Subsequently samples were washed for 5 min twice in PBSTw. The samples were permeabilized with 5 μg/ml proteinase K (Takara, Japan) in PBSTw for a time period depending on the stage of tissues at room temperature. The samples were fixed in 4% PFA and then washed for 5 min each for five times in PBSTw at room temperature. The samples were pre-hybridized in hybridization mix solution at 60ºC for 2-5 hours and hybridization was performed in a humidified incubator at 60ºC for 16-20 hours followed by high-stringency washes at the same temperature for 30 min in 2xsaline sodium citrate (SSC) and two rinses in 1xSSC. Unhybridized probes were removed by the incubation with RNase A in a solution containing 0.01M Tris-HCl (pH 8.0) and 0.5 M NaCl at 60ºC followed by two washes in 0.1xSSC for 30 min each at room temperature. The samples were blocked with 2% blocking solution for 60 min at room temperature. Blocking buffer was replaced with a solution containing anti-digoxigenin (DIG) Fab-alkaline phosphatese conjugate (1:2000 in blocking solution) for 1 hour at room temperature. Unbound DIG antibodies were removed after six washes in PBSTw for 15 min each followed by incubation with staining buffer (0.1M NaCl, 1.1 M Tris-HCl, 0.05 M $MgCl_2$, pH 9.5) for 5 min each and for three times. Finally the samples were stained with nitroblue tetrazolium chloride (NBT) (Roche) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (Roche) for detection of riboprobes. Serial sections were examined with a stereomicroscope MVX10 (Olympus, Tokyo, Japan).

In situ *hybridization with LNA probe*

The digoxigenin (DIG)-labeled MiRCURY detection probe (Exiqon, Copenhagen, Denmark), an LNA-modified oligo DNA probe containing the miR-499 mature sequence (5′-AAACATCACTGCAAGTCTTAA-3′) was used to detect miR-499 transcripts. *In situ* hybridizations were performed according to Kloosterman et al. [43]. The embryos and larvae of zebrafish, torafugu and medaka and adult trunk skeletal and cardiac muscles of these fish were fixed in 4% PFA at 4°C overnight. Transverse sections of the tissues were cut at 16-µm thickness. All hybridizations were performed at 66°C, which was 20°C below the predicted melting temperature (Tm) of the LNA probe. Alkaline phosphatase-conjugated anti-DIG antibody (Roche Diagnostics, Penzberg, Germany) and nitroblue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate were used for signal detection with an MVX10 stereomicroscope (Olympus, Tokyo, Japan).

Results

Distribution of MYH14 *and miR-499 in teleost fish genomes*

Using the genomic databases available for different vertebrates, the syntenic organization of human *MYH14* and miR-499 with their orthologues was examined. The locations and IDs of *MYH14* and miR-499 used in this study are shown in Supplementary Table 1 and Figure 1. The outcome show that the tandem arrayed location of the ER degradation enhancer, mannosidase alpha-like 2 gene (*EDEM2*),

transient receptor potential cation channel subfamily C member 4 associated protein gene (*TRPC4AP*), and *MYH14* containing miR-499 were conserved in humans, chickens, and coelacanths *Latimeria chalumnae*. The synteny was also found LG18 in spotted gar *Lepisosteus oculatus*. In zebrafish Chr11, *MYH14* containing miR-499 was located next to *TRPC4AP*. In addition, two *MYH14s* were also found on Chr23 located near a putative *TRPC4AP* paralogues. Both zebrafish *MYH14* contained miR-499, totaling three *MYH14*/miR-499 pairs in this species. Ikeda et al. (2007) reported two *MYH14* paralogues, MYH_{M5} and MYH_{M3383}, in the torafugu genome. The former was located on scaffold79 and the latter on scaffold398. MYH_{M5} was located next to *TRPC4AP* and contained miR-499, whereas MYH_{M3383} was located next to sulfatase 2 gene (*SULF2*) and did not contain miR-499 in its intron. In tetrapods, however, *SULF2* is located in the same chromosome as *MYH14*/miR-499, but far from the locus. Based on the synteny, two putative *MYH14s*, one containing miR-499 and the other lacking it, were also found in green spotted puffer *Tetraodon nigroviridis* and tilapia *Oreochromis niloticus*. Interestingly, in Atlantic cod *Gadus morhua*, stickleback *Gasterosteus aculeatus*, platyfish *Xiphophorus maculatus*, and medaka, miR-499 was present within the expected syntenic region that contained *TRPC4AP*, *NDRG3*, *SULF2*. However, *MYH14* was absent in each case. Cod and stickleback retained a single *MYH14* paralogueue lacking miR-499 in the other syntenic region that contained *SULF2*. *SULF2* seems to be consistently located next to *MYH14* in most teleost fish species. Interestingly, the medaka genome was lacking *MYH14*. Although the *MYH14* sequence was screened

from the Ensembl medaka genome and medaka EST data sets deposited to DDBJ/EMBL/GenBank using tBLASTn and the torafugu MYH14-1 (MYH_{M5}) protein sequence as a query, no *MYH14* sequence was retrieved.

Phylogenetic analysis of MYH14 *and miR-499*

Phylogenetic analyses based on the *MYH14* coding and miR-499 stem-loop sequences were performed to clarify the evolutionary history of the *MYH14*/miR-499 locus in teleost fish. Figure 2A and supplementary Figure 1A show neighbor-joining (NJ) and maximum-likelihood (ML) trees of the *MYH14*s. Both trees show almost the same phylogenetic relationship, indicating the reliability of the phylogenetic relationships observed in this study. *MYH14* was monophyletic in the amniote lineage, including humans, chickens, and coelacanths, but was duplicated in the ray-finned fish lineage, except for the spotted gar (Figure 2A). Therefore, both *MYH14*s in teleost fish are paralogous genes that diverged at the base of neoteleostei lineage. *MYH14* paralogues were separated, except for zebrafish, according to the presence or absence of miR-499 in their introns. Note that accelerated evolution was clearly observed in *MYH14*s lacking miR-499 by their large genetic distance from *MYH14* possessing miR-499, suggesting a functional relationship between *MYH14* and miR-499.

The miR-499s phylogenetic relationships (Figure 2B and Supplementary Figure 1B) were consistent with those of the *MYH14*s. Although the bootstrap value in each node was quite low, three zebrafish

miR-499 paralogues, miR-499-1, -2, and -3, were divided into two clades. Zebrafish miR-499-1 formed a single cluster with other teleost fish miR-499s.

The combined phylogenetic and synteny analyses suggest that the *MYH14*/miR-499 locus was duplicated early in teleost evolution and one of the duplicated miR-499 genes was lost in the common ancestor to cod and the Acanthopterygii, after the split from the zebrafish lineage. Additionally, *MYH14*s have seemingly been lost at independent points of teleost evolution.

Divergent expression of zebrafish MYH14-1, MYH14-2 *and* MYH14-3 *paralogueues*

Among the three *MYH14*/miR-499 loci on zebrafish genome, a *MYH14*/miR-499 locus on chromosome 11 shows highly conserved synteny with its mammalian counterpart. This locus was termed as *MYH14-1*/miR-499_1. Additional two *MYH14s, MYH14-2* and *MYH14-3* were tandemly arrayed on Chr23 of zebrafish located near a putative *TRPC4AP* paralogue. The latter two *MYH14* paralogues contained putative intronic miRNA-499 sequences which were termed as *MYH14*-2/miR-499_2 and *MYH14*-3/miR-499_3, respectively. The expression of *MYH14-2* and *MYH14-3* in adult zebrafish was examined by RT-PCR with the *β-actin* gene as an internal control. The transcripts of *MYH14-1* were expressed in both slow and fast skeletal muscle tissue along with cardiac muscle of adult and larval zebrafish (Kinoshita et al., 2011). On the contrary, *MYH14-2* transcripts were not amplified by RT-PCR (Fig. 3A) and no probe could be generated for *MYH14-2 in situ* hybridization. The transcripts of

MYH14-3 were expressed in slow muscle tissues and in cardiac muscle in developmental and adults stage by RT-PCR (Fig. 3A). *In situ* data showed, zebrafish larvae at 2 dpf expressed *MYH14-3* transcripts along the skeletal and cardiac muscle region (Fig. 3B). Figure 3C showed larval *in situ* hybridization of 2dpf zebrafish with sense probe of *MYH14-3*. Trunk skeletal muscle of zebrafish larvae expressed *MYH14-3* transcripts in lateral surface of the myotome beneath the skin which is specifically slow muscle fiber region whereas the fast muscle fibers located at the deep portion of the myotome, remained unstained (Fig. 3D). Similar to the the developmental stage, *MYH14-3* transcripts were localized in superficial slow muscle whereas no transcripts could be detected in the intermediate or in fast muscle fibers region of adult zebrafish (Fig. 3E.).

miR-499 expression in medaka, torafugu and zebrafish developmental stage

To find out miRNA-499 developmental expression in zebrafish, torafugu and medaka, *in situ* hybridization was performed. In zebrafish, *in situ* hybridization showed miR-499 was not expressed in the skeletal muscle of embryos at 1 dpf (Fig. 4A,B). Note that miR-499 expression was observed in the heart at the same stage (Fig. 4A). On the other hand, larvae at 2 dpf and 5 dpf showed a clear expression of miR-499 in both cardiac and skeletal muscle (Fig. 4C, E). Transverse sections of the larvae at the trunk region revealed expression of miR-499 restricted in slow muscle fibers located at the lateral surface of the myotome, whereas fast muscle fibers situated deep inside of the myotome remained unstained (Fig. 4D,

F). In torafugu, the fast twitch fibers are located deep in the trunk myotomal muscle, where the slow twitch fibers are at the surface part of the myotome called lateralis superficialis (LS) and at the erector-depressor (ED) of the median fins (dorsal and anal fins). The organization of different muscle types examined in torafugu musculature is schematized in Figure 5A. In our study, miR-499 expression in torafugu examined by whole mount *in situ* hybridization did not reveal any expression in 6 dpf torafugu embryos (Fig. 5B). However, hatched larvae at 8 dpf showed a skeletal muscle specific expression of miR-499 (Fig. 5C,D). The transverse sections of larval trunk muscle showed, miR-499 was specifically expressed along the lateral surface of the myotome and in the horizontal myoseptum region where the slow muscle fibers are located (Fig. 5E). Meanwhile, the expression of miR-499 in heart is first detected at 8 dpf larval stage (Fig. 5C). We observed that medaka miR-499 was expressed at the embryonic stage in the notochord (Fig. 6A) even lacking its host myosin gene. miR-499 expression in the notochord has not been previously reported in other animals. At the hatching stage, miR-499 was expressed in cardiac and trunk skeletal muscles (Fig. 6B,C). The transverse sections of the medaka larva clearly showed miR-499 expression in the heart (Fig. 6D) and the lateral surface of the myotomal muscle (Fig. 6E) where slow muscle fibers are present.

Distribution of miR-499 transcripts in zebrafish and torafugu and medaka adults

Similar to the larval stage, *in situ* hybridization of adult zebrafish showed high expression of miR-499 in

cardiac muscle (Fig.7A). The adult trunk muscle showed superficial slow muscle specific expression of miR-499. The slow muscle fibers of small diameters were stained and the fast muscle with large diameters as well as intermediate muscle fibers remained unstained for miR-499 (Fig. 7B). In adult stage, torafugu trunk muscle cross sections showed, miR-499 is localized in the lateralis superficialis muscle (LS) (Fig. 7C,D) and in erectors-depressors muscle (ED) (Fig. 7E). No expression of miR-499 is found in deep fast trunk muscle (Fig. 7F). Cardiac muscle cryosections clearly revealed expression of miR-499 by *in situ* hybridization (Fig. 7G,H). Thus, torafugu miR-499 is slow/cardiac specific at larval stages of muscle development. At medaka adult stage, miR-499 expression was detected only in cardiac muscle by *in situ* hybridization (Fig. 7F-H). These miR-499 expression pattern in the adult stage of torafugu and medaka was also confirmed by next-generation sequencing (Table 3, Figure 8). After sequencing, miR-499 is found in muscle, intestine, eye, brain, cardiac, liver, ovary from adult torafugu. In torafugu, the highest number of miR-499 in cardiac muscle (22707) with a significant read number in trunk skeletal muscle (10353) was retrieved, however minor amount of expression is detected in eye (603), brain (84), intestine (22), liver (82), ovary (24) and testis (0) (Table 1, Fig.8A). Although miR-499 was detected in the adult medaka tissues examined, much higher miR-499 reads were obtained from the cardiac muscle (reads per million [RPM] = 20,624) when compared with skeletal muscle (544), eye (256), brain (40), intestine (22), testis (11), and ovary tissues (0) (Figure Figure 8B).

Secondary structure of the miR-499 stem-loop sequence

Intronic miRNA is transcribed as pre-mRNA from a part of an intron in the host gene (Kim et al., 2009). miRNA endowed by an intron folds to form a local double-stranded stem-loop structure called the primary miRNA (pri-miRNA). In animals, RNase III drosha crops pri-miRNA at the stem-loop during splicing and produces a precursor miRNA (pre-miRNA), which is then processed by dicer to form mature miRNA. From these canonical intronic miRNAs, a new type of intronic miRNA called mirtron has been discovered. Mirtrons are embedded in short introns, and their biogenesis does not require drosha cropping. The pre-miRNA of mirtron is produced directly by splicing (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007). Figure 9 shows miR-499 predicted stem-loop structures from medaka, torafugu, and the representative mirtron, miR-62, from *Caenorhabditis elegans*. miR-499s have longer stem-loop regions than those of mirtrons and are processed by drosha to produce pre-miRNAs. The torafugu *MYH14* intron containing miR-499 is 247 bp in length (see supplementary Figure 2), which is long enough to produce canonical miRNA hairpins to be cut by drosha. These results combined suggest that miR-499 is not a mirtron but a canonical intronic miRNA. However, experimental proof is required to confirm whether miR-499 requires drosha processing.

Discussion

Figure 10 shows the putative evolutionary history of the *MYH14*/miR-499 locus in teleost fish. It has been proven that after two rounds of whole genome duplication (WGD) in a common ancestor of vertebrates, a

third WGD occurred in the fish lineage (Amores et al., 1998; Elgar et al., 1999; Postlethwait et al., 2000; Woods et al., 2000; Smith et al., 2002). This fish-specific WGD occurred at the base of the Teleostei lineage, after diverging from ancient fish groups such as Polypteriformes, Acipenseriformes, and Lepisosteidae (Hoegg et al., 2004). Our phylogenetic analysis clearly shows duplication of the *MYH14*/miR-499 locus after the divergence of spotted gar, indicating that the teleostei-specific WGD provided present-day *MYH14*/miR-499 paralogue in teleost fish. *TRPC4AP* and *SULF2* genes located next to *MYH14*, were also duplicated in the fish-specific WGD. However, information on Osteoglossomorpha, Elopomorpha, Clupeomorpha, and Protacanthopterygii, which are important fish groups comprised of neoteleostei, was not reviewed in this study. Therefore, further analysis is required to fully reveal *MYH14*/miR-499 evolution in fish.

The existence of multiple *MYH14* and miR-499 genes in various teleost fish suggests their expressional and functional versatilities. Torafugu *MYH14-1* (MYH_{M5}) expression was observed in both slow and cardiac muscles in the developmental and adult stages, whereas *MYH14-2* (MYH_{M3383}) expression was restricted to adult slow muscle (Akolkar et al., 2010; Ikeda et al., 2007). Zebrafish *MYH14-1* was expressed in both slow and cardiac muscles in the early developmental stages and in slow and intermediate muscles in the adult stage (Kinoshita et al., 2011). Our previous analysis revealed, zebrafish *MYH14-1* was expressed in most myotomal regions at embryonic stage and in superficial slow muscle fibers in the adult stage and a low level of *MYH14* transcripts were also detected in intermediate muscle

fibers located between superficial slow and inner fast muscle fibers in the adults (Kinoshita et al., 2011). The present investigation also revealed a transient expression of *MYH14-1* in the cardiac/myotomal muscle of developmental zebrafish. However, *MYH14-2* transcripts were not detected in examined tissues including cardiac and myotomal skeletal muscle and *MYH14-3* was expressed specifically in slow/cardiac muscle at larval and adult stage (see figure 3). Thus, the existence of multiple *MYH14/*miR-499 paralogs in zebrafish genome therefore, resulted expression difference among them. Unlike vertebrate, an additional round of whole genome duplication occurred in teleost lineage, resulting in duplication of most genes (Christoffels et al., 2004; Vandepoele et al., 2004). This is also the case with *MYH14/*miR-499s as well as other sarcomeic *MYH*s (Kikuchi et al., 1999; Ikeda et al., 2004; Muramatsu-Uno et al., 2005; Watabe and Ikeda, 2006; Ikeda et al., 2007) and Liang et al., 2007). The redundancy of duplicated gene often causes differentiation in their expression patterns (Scemama et al., 2006 and Carlos et al., 2008**)** which may explains the divergent expression observed between *MYH14* paralogs as well. *MYH14-3* transcripts expressed in adult cardiac muscle in addition to slow skeletal muscle is well coincided with our previous observations of torafugu *MYH14* (Ikeda et al., 2007 and Akolkar et al., 2010) as well as *MYH14* and miR-499 for mammals (Rossi et al., 2010). Furthermore, this present study demonstrates that miR-499 expression differed from the above-mentioned divergent *MYH14* expression patterns. *In situ* hybridization showed miR-499 was not expressed in the skeletal muscle at the embryonic stage of zebrafish, torafugu and in medaka. However, miR-499 expression was observed in the heart at the same

stage. On the other hand, zebrafish, torafugu and larvae showed a clear expression of miR-499 in both cardiac and skeletal muscle. Adult zebrafish and torafugu showed high expression of miR-499 in cardiac/skeletal muscle. In torafugu, miR-499 is also localized in the erectors-depressors (ED) muscle which exhibit characteristics of slow muscle fibers. miRNA transcripts detected using next generation sequencing platform showed, miR-499 is most abundantly found in cardiac muscle with a significant read number in trunk skeletal muscle in medaka and torafugu. It would be interesting to determine whether such differences in *MYH14* and miR-499 are related to physiological and ecological variations among teleost fish species. Fish are the most diverse vertebrate group consisting of over 22,000 species. In response to the wide range of environmental and physiological conditions they encounter, the characteristics of fish musculature, including muscle fiber-type composition, are also highly diverse. Medaka makes a particularly interesting subject because of the complete elimination of *MYH14* from its genome. Although muscle fiber-type composition has not been well characterized in medaka, Ono et al. (2010) reported an *MYH* gene specifically expressed in slow muscle fibers at the horizontal myoseptum. Such *MYH* expression has never been reported in other teleost fish species. In contrast, medaka fast muscle exhibits high plasticity to adapt to temperature fluctuations by changing *MYH* expression (Liang et al., 2007 and Liang et al., 2008). Further comparative analyses of *MYH14* and miR-499 may shed light on the mechanisms involved in the formation of species-specific musculature evolution.

The loss of the intronic miRNA in the ancestor of cod and the Acanthopterygii might be explained by

functional redundancy. The loss of intronic miRNA from the host gene is possible if mutations are introduced into an intron without any effect on the function and expression of the host gene. Stickleback, medaka, and Atlantic cod display the opposite pattern with the intronic miRNA lacking its host gene. Intronic miRNAs are transcribed with their host genes, and thus, coordinated expression between an intronic miRNA and its host gene is frequently observed (Baskerville and Bartel, 2005). In the present study, however, medaka miR-499 was actually expressed in various tissues despite the absence of *MYH14* (see Figure 7). How does intronic miRNA remain after the loss of its host gene? We speculate that miR-499 is a canonical intronic miRNA produced by drosha cropping (see Figure 9). Recent studies have revealed that splicing and pre-miRNA cropping by drosha are independent processes, indicating that splicing is not essential for intronic miRNA production (Ref). In other words, severe mutations of the host gene may not affect the production of intronic miRNAs in the presence of the host gene transcriptional system. To our knowledge, this is the first report that describes the conversion of intronic into non-intronic miRNA during evolution. Comparative analysis of transcriptional regulation between intronic and intergenic miR-499s will provide new insights into miRNA evolution.

Chapter 3

Expression and distribution of *MYH6*/*vmhc*/miR-736 in teleots

Abstract

Myosin heavy chain (MYH), actin-based motor protein which converts chemical energy to mechanical force by hydrolyzing ATP, is the major component of contractility of striated muscles. MYH gene (*MYH*) is a multi gene family and spatio-temporally regulated expression of each family member determines physiological and contractile muscle property. In mammals, 8 well characterized *MYHs* and three ancient *MYHs* have been characterized in cardiac and skeletal muscles. Among them, *MYH6* (MYH-α), *MYH7* (MYH-β), and *MYH14* (also called *MYH7b*) occupy a unique position because of their specific expression in cardiac and slow muscle fibers, and existence of microRNAs (miRNAs). *MYH6* encodes miR-208a and *MYH7* and *MYH14* encode miR-208b and miR-499, respectively. miR-208 and miR-499 are also called as myo(muscle) miRNAs for their muscle specific expression. In mammals, genomic positions and expression patterns of myomiRs and their host genes are well conserved and they form an important transcription network which works in muscle fiber-type specification. Highly diverged distribution and expression of *MYH14* and its myomiR, miR-499 has been reported suggesting unique evolution of myomiRs in teleost lineage. However, the evolutionary history of MYH6/vmhc/miR-736 has not been reported yet. We examined expression and distribution of these myomiRs in three representative teleost genome, torafugu *Takifugu rubipes*, zebrafish *Danio rerio* and medaka *Oryzias latipes*. Our result suggests, as well as in *MYH14*/miR-499 locus, teleost genome contains multiple MYH orthologues at *MYH6*/*vmhc*/miR-736 locus. However, only one *MYH*, *vmhc*, contains intronic miRNA, miR-736.

Expression of miR-736 in adult torafugu and medaka was examined in the present study using next generation sequencing platform, showing trace expression in various tissues including muscle. These results indicate uncoordinated expression of miR-736 with its host gene, suggesting miR-736 is not a functional orthologue of mammalian miR-208s. Taken together with highly conserved expression of miR-499, these results suggest that the complete deletion of miR-208s function in teleost is counterbalanced by the role played by miR-499.

Introduction

Myosin is the major contractile protein in muscle tissues, the molecule which consists of two myosin heavy chains (MYHs) and four light chains. While MYHs are encoded by the multigene family in which vertebrate muscles contain a variety of myosin heavy chain genes (*MYHs*) with different properties and their spatio-temporal expressions specify muscle fiber types such as slow and fast ones. Myosin heavy chain genes (*MYHs*) are expressed in an interesting switch on-off manner throughout growth and development of vertebrates. A comprehensive analysis of the human genome revealed two groups of sarcomeric *MYHs* (Weiss et al., 1999a; Weiss et al., 199b). One group contains two tandemly arrayed cardiac *MYHs*, and, the latter being also expressed in slow skeletal muscle. The other group contains six skeletal muscle *MYHs*, including adult fast IIa, IIx, IIb, embryonic, neonatal/perinatal, and extra ocular muscle types. Among them, *MYH6*, *MYH7* and *MYH7b* contained intronic miRNAs, miR-208a,

miR-208b and miR-499 respectively. These miRNAs are co-expressed in a slow/cardiac specific manner with their host gene. Hence, these miRs expressed exclusively in the muscle are termed as myo(muscle) miRNA. However, the details of the molecular mechanisms involved in a complex expression pattern of *MYHs/*miRNAs in the muscle development are still unknown. In rodents, *MYH6* is the predominant isoform expressed in the adult heart and 27th intron of this gene encodes miR-208a, which is expressed specifically in the heart as well as its host gene (van Rooij et al., 2009). *MYH7* encode miR-208b on its 31th intron. *MYH7*/miR-208a is co-expressed predominantly in slow skeletal muscle fibers and a lower level of expression in heart (Morkin et al, 2000 and van Rooij et al., 2009). In rodents, miR-208a regulates the function of 208b and miR-499. In human, *MYH7* is predominantly expressed in the ventricle, whereas expression of *MYH6* is restricted in the atrium (Everett, 1986). The most recently identified sarcomeric *MYH*, *MYH14 (*also called *MYH7b)* is known for the existence of miRNA, miR-499, in its 19th intron (McGuigan et al., 2004), showed coupled expression in a slow/cardiac specific manner in mammalian muscles (Bell et al., 2010). These *MYHs* and myomiRs are expressed specifically in cardiac and slow muscles in mammals. In mammals, thus, the genomic positions and expression patterns of myomiRs and their host myosin genes are well conserved and they form an important transcription network which works in muscle fiber-type specification. In the previous investgation, highly diverged distribution of *MYH14* and its myomiRNA, miR-499 is revealed, suggesting myomiRNAs unique evolution of in teleost lineage. Unexpectedly, the genome mapping of fish muscle miRNAs revealed that

cartilaginous and ray-finned fish genomes do not retain the MYH7/ miR-208 whereas *MYH6* persists but missing the intronic miR-208 (Pedro Gabriel Nachtigall, 2014). It has been revealed that similar to mammals, the functionally conserved zebrafish miR-499 reinforces Sox6/*MYH14*/miR-499 transcriptional network by repressing Sox6 translation in slow muscle progenitor (Wang et al. 2011). Such a mechanism also enables myomiRs to function as conserved binary regulators of slow versus fast muscle gene programs, activating slow and repressing fast through their downstream targets. In teleosts, a major cardiac MYHs isoform, ventricular myosin heavy chain (*vmhc*) encodes an intronic miRNA, miR-736. Studies in other vertebrates have shown that miR-736 belongs to the 208 family of miRNA genes specifically expressed in cardiac tissue and can be referred as myomiRNA (Andreassen R et al., 2013). However, the evolutionary history of these MYHs and their myomiRNAs has not been studied in detail. In the present study, the distribution of *MYH6*, *vmhc*/miR-736 and *MYH14*/miR-499 loci on various teleost fish genomes was examined their evolutionary history was revealed by sequence and expression analyses.

Materials and methods

Fish

All procedures in this study were performed according to the Animal Experimental Guidelines for The University of Tokyo. Live specimens of adult torafugu (body mass about 1 kg) were reared in local tap

water with a circulating system at 28.5°C under a 14:10-h light-dark photoperiod raised at The University of Tokyo. Various tissues including brain, fast, slow and cardiac muscles were dissected after instant killing by cutting head of the fishes. These tissues were immediately frozen in liquid nitrogen and stored at –80°C until use for total RNA extraction.

Construction of a physical map around MYH14 *and miR-499*

The Ensembl genome browser (http://www.ensembl.org/index.html) was used to determine the syntenic organization in the region surrounding *MYH6/vmh/MYH14* and/or miR-499/miR-736 in vertebrates. The database versions used were as follows: human (GRCh37), chicken (Galgal4), coelacanth *L. chalumnae* (LatCha1), zebrafish *D. rerio* (Zv9), torafugu *T. rubripes* (FUGU4), green spotted puffer *T. nigroviridis* (TETRAODON8), tilapia *O. niloticus* (Orenil1.0), Atlantic cod *G. morhua* (gadMor1), stickleback *G. aculeatus* (BROADS1), platyfish *X. maculatus* (Xipmac4.4.2), and medaka *O. latipes* (MEDAKA1). The pre Ensembl browser (http://pre.ensembl.org/index.html) was used for analysis of Spotted gar *L. oculatus* (LepOcu1).

Bioinformatics analysis

The *MYH6*, *vmhc* and miR-736 sequence data were retrieved from the available genome databases mentioned above (Table 1). NJ and ML trees were constructed on the basis of the *MYH6*, *vmhc* coding and miR-736 stem-loop sequences using MEGA5 (Tamura et al., 2011) with 1000 bootstrap replications.

The Nei and Gojyobori method (Nei and Gojyobori, 1984) (Jukes-Cantor) was employed to consider synonymous and non-synonymous substitutions for the *MYH6*, *vmhc* NJ tree. The Tajima-Nei model (Tajima and Nei, 1986) was employed for the miR-736 NJ tree, whereas the Tamura-Nei model (Tamura nad Nei, 1993) was used for the *MYH6*, *vmhc* and miR-736 ML trees. Putative secondary structures of the miR-499 from medaka and torafugu stem-loop sequences and that of the *C. elegans* mirtron miR-62 (miRBase accession number: MI0000033) were predicted using the RNA fold program CentroidFold (http://www.ncrna.org/centroidfold).

miR-736 library construction and sequencing

Small RNA libraries were prepared from various tissues of adult torafugu skeletal muscle, intestine, eye, brain, heart, both mature and immature ovary and testis. For this purpose, total RNA from each sample was extracted using mirVana^{TM}miRNA Isolation Kit (Ambion). Small RNAs less than 40 nucleotides in size were isolated using flashPAGE™ Fractionator (Ambion). First strand cDNA synthesis was then carried out and PCR amplified using barcode specific primers (Table 1). PCR products of about 100 nucleotides in length were then purified from nondenaturing polyacrylamide gels. The 7 PCR products were mixed and sequenced using Solid3 next generation sequencer (Applied Biosystem). The obtained 172,656,341 of 35 nucleotides reads in total with exact barcode sequences. The 2-base encoding sequence data were decoded into nucleotide sequence data and divided into sub-dataset of each tissue according to the barcode sequence. Low quality reads were eliminated using our own designed perl scripts. From this,

102,602,452 high quality reads were obtained. After trimming of adaptor, 97,186,950 reads were retrieved. The reads of 18-25 nucleotides were subjected to Blast search against known mature miRNA sequences deposited in miRBase 18.0 (www.mirbase.org/). Sequences with their seed regions 5'-end 2-8 nucleotides shown 100% identity to seed regions of known mature miR-736 sequences were annotated.

Results

Distribution of myomiRs in teleost genome

Our evolutionary analysis on MYH-6, *vmhc*/miR-736 and MYH7b/miR-499 revealed, multiple orthologues are present on the locus of these genes and miRNAs in teleost lineage. In human, MYH6, MYH7, MYH7b are present but *vmhc* is absent. However, in zebrafish, the najor cardiac MYH, vmhc, contains intronic miRNA, named miR-736. Sequence similality indicates miR-736 is orthologue of miR-208. In the same synytenic region with human, zebrafish contains multiple *vmhc* loci, one containing miR-736 and the others lacking it and three slow muscle specific gene (*smyhc1*, *smyhc2*, *smyhc3*). In torafugu genome, *vmhc* ($MYH_{M2126\text{-}1}$) contains intronic miR-736 and the other lacking it ($MYH_{M2126\text{-}2}$). However, torafugu genome contains two MYH_{M5} isoform, however one contains miR-499, *MYH14-1*(MYH_{M5}) and the other, *MYH14-2* ($MYHM_{3383}$) lacking it. Torafugu and medaka genomes contain 3 and 5 *MYH6*/*vmhc* paralogues respectively, and only one paralogue in each fish contains miR-736. Thus, torafugu, zebrafish and medaka revealed that they contain multiple orthogues of *vmhc*/miR-736 in their genome. Thus, MYH6/MYH7/ miR-208 showed a highly varied distribution

among species.

Phylogenetic analysis MYH6/vmhc/miR-736

Phylogenetic studies revealed that multiple orthologues of *MYH6*/*vmhc*/miR-736 are present in teleost genomes. Zebrafish genome contains a total of 10 *MYH6*/*vmhc* paralogues of which one of them possess miR-736 and the others lack the intronic miRNA.

miR-736 expression by next generation sequencing

To confirm the miR-499 and miR-736 expression in zebrafish and torafugu, miRNAs library construction and sequencing using next generation sequencing platform was consecutively performed. After sequencing, miR-499 is found in muscle, intestine, eye, brain, cardiac, liver, ovary from adult torafugu. The highest number of miR-499 in cardiac muscle (22707) with a significant read number in trunk skeletal muscle (10353) was retrieved, however minor amount of expression is detected in eye (603), brain (84), intestine (22), liver (82), ovary (24) and testis (0) (Table 1, Fig. 3). Expression of torafugu and zebrafish miR-736 was quietly low in examined tissues although their host gene, *vmhc*, is a major isoform among cardiac muscle MYHs. These expression patterns indicate that miR-736 does not contribute to muscle formation in teleost.

Discussion

Recent study by our group revealed that teleost fish genomes contain multiple *MYH14* and miR-499

paralogs (Sharmin et al., 2013). Using the genomic databases available for different vertebrates, we examined the syntenic organization of human *MYH6/MYH7* and miR-208 with their orthologs. In this study, synteny analysis revealed, as well as in *MYH14*/miR-499 locus, teleost genome contains multiple MYH orthologues at *MYH6*/*MYH7*/miR-208 locus. However, miR-208 is completely absent and the only one *MYH*, *vmhc*, contains intronic miRNA, named miR-736 is present in this syntenic region (see Figure 1). Our phylogenetic analysis shows duplication of the *vmhc*/miR-736 locus in zebrafish, torafugu and medaka genome resulting from fish specific additional round of whole genome duplication (WGD). Sequence similality indicates miR-736 is an orthologue of miR-208. Recently it has been reported that artilaginous and ray-finned fish genomes do not retain the miR-208/MYH7 gene whereas MYH6 persists but missing the intronic miR-208, except for coelacanth that retain *MYH7* as well as miR-208 intronic to *MYH6* (Pedro Gabriel Nachtigall, 2014). Together with our syntenic analysis it can be concluded that myomiRs and their host *MYHs* are highly diverged in teleost lineage.

The existence of multiple MYH6/*MYH14/vmhc* genes and their intronic miRNAs miR-499 and miR-736 in various teleost fish suggests their expressional and functional versatilities. Torafugu *MYH14-1* expression was observed in slow and cardiac muscles in the developmental and adult stages, whereas *MYH14-2* expression was restricted to adult slow muscle (Ikeda et al., 2007; Akolkar et al., 2010). Zebrafish *MYH14-1* was expressed in both slow and cardiac muscles in the early developmental stages and in slow and intermediate muscles in the adult stage (Kinoshita et al., 2011). Our sequencing data

suggests, miR-726 expression did not coordinate to its host gene, *vmhc* in torafugu adults. This result is also supported by previous observations by northern blot analysis as revealed miR-499 is expressed highly in heart and skeletal muscle with miR-736 expressed throughout developmental stage of zebrafish but almost absent in adult stage except some expression in gut and eye (Kloosterman et al., 2006). These expression patterns indicate, miR-736 is not a functional orthologue of miR-208. However, our data analysis referred to the expression pattern of miR-499 indicates that the co-transcription within their host gene was conserved.

Thus, despite varied distribution and expression of myomiRs and their host MYHs, a slow/cardiac specific miR-499 expression was exceptionally conserved from fish to human. Recent study reported that three myosin genes, *MYH6 and MYH7* and *MYH14* encoded by related myo(myo=muscle)miRNAs within their introns control muscle myosin content, myofiber identity and muscle performance in rodents (van Rooij et al., 2009). The miR-208 is considered essential for the differentiation and maintenance of slow-twitch muscle fiber type as well as for heart development. Our data suggests that the complete deletion of *MYH7*/208 in teleost is counterbalanced by the role played by miR-499.

From this study it has been confirmed that MYH7/miR-208 or miR-736 was absent in teleost cardiac muscle. These results imply, putative pivotal role of miR-499 in muscle fiber-type specification in teleots. Future comparative analysis on the physiological functions of miR-499 will hopefully shed light on the mechanisms involved in the formation of unique musculature observed in divergent fish species.

Chapter 5

Expression regulation of MYH14 paralogues/myomiRs and functional analysis of miR-499 in teleosts

Abstract

Recently, after the completion of the human genome project, a new sarcomeric *MYHs*, *MYH14 (MYH7b)*, which contains an intronic microRNA, miR-499, has been identified on chromosome 20. This new sarcomeric *MYHs* differs significantly from the other sarcomeric *MYH*s in sequence, size and exon–intron organization. Comprehensive genome-wide screenings of *MYHs* revealed that fish also contain *MYH14* orthologs in their genome. Our previous study, however, demonstrated that *MYH14*/miR-499 syntenic region was highly diverged in teleost lineage and their functional and expressional analyses during muscle formation are still limited. Detail promoter analysis of these transgene suggested, 5710b 5'-flanking region of *MYH14-1* and 5641b of *MYH-3* contained a necessary regulatory region to recapitulate their expression during embryonic and larval muscle development of zebrafish. The embryos of stable transgenic p*MYH14-1:5710b-Tol$_2$EGFP* zebrafish generated expression in both slow and fast muscle fibers consistent with the endogenous expression. Among the paralogs, the 5'-flanking region of zebrafish *MYH14-1* with its torafugu ortholog revealed the existence of two distal and a single proximal conserved regions. However, deletion study of these conserved regions suggests, they had no effect on *MYH14-1* expression in muscle fibers in case of zebrafish. The 5'-flanking region of *MYH14-1* also revealed existence of two conserved regions (CSR-1, CSR-2) with *MYH14-2* and *MYH14-3* paralogs. Deletion of these conserved regions significantly reduced the promoter activity of *MYH14-3* conversely to *MYH14-1* which remained unchanged. Interestingly, deletion of both the conserved regions from

p*MYH14-3:5641b-Tol*$_2$*-EGFP* dramatically changed fiber type-specific expression of the reporter gene and the number of EGFP-expressing slow muscle fibers per embryo was also reduced. Comparing the flanking sequences of *MYH14*/miR-499 loci between torafugu, zebrafish, and medaka revealed some highly conserved regions and an intronic sequence immediately downstream of miR-499 is conserved among them suggesting *cis*-regulatory elements have been functionally conserved in medaka miR-499 despite the loss of its host gene. Interestingly, medaka miR-499 is transcribed lacking its host gene *MYH14*, which suggests the presence of its own promoter for transcription. Detail promoter analysis revealed, deleting the regions containing several highly conserved sequences of *MYH14*/miR-499 loci between torafugu and medaka upstream of miR-499 (4769/-1897) in medaka generated reporter gene expression specifically in cardiac and fast muscle fibers. Loss of function experiment of miR-499 was performed by injection of antisense miR-499 oligonucleotide into medaka and zebrafish eggs. As expected, knockdowned larvae showed marked reduction of slow muscle fibers in zebrafish and medaka developmental stages. In zebrafish, such reductions were also validated by real-time PCR analysis using a transgenic line which expresses GFP in slow muscle fibers. Taken together with the fact of trace expression of miR-736, teleost miR-208s orthologue, in muscles, miR-499 is a sole functional myomiR and thus have indispensable role in muscle fiber-type specification in teleost muscle formation.

Background

Vertebrate muscles contain a variety of myosin heavy chain genes (*MYHs*) with different properties and their spatio-temporal expressions specify muscle fiber types such as slow and fast ones. A comprehensive analysis of the human genome revealed two groups of sarcomeric (skeletal and cardiac) *MYHs* (Weiss et al., 1999a; Weiss et al., 1999b). One group contains two tandemly arrayed cardiac *MYHs*, alpha and beta, the latter being also expressed in slow skeletal muscle. The other group contains six skeletal muscle *MYHs*, including adult fast IIa, IIx, IIb, embryonic, neonatal/perinatal, and extra ocular muscle types.

The completion of the human genome project led to the identification of three additional sarcomeric *MYH*s ([Berg et al., 2001] and [Desjardins et al., 2002]). Among them, the new sarcomeric *MYH, MYH14 (*also called *MYH7b)* is known for the existence of an intronic miRNA, miR-499. In mammalian muscles, *MYH14* and miR-499, showed a coupled expression in a slow/cardiac specific manner in developmental and adult stages (Bell et al., 2010) although their functions are partly known.

Interestingly, *MYH14* expression is controlled post-transcriptionally by miR-499 as a regulatory feedback circuit in mammals. One of the putative targets of miR-499 is a transcription factor, Sox6, which is involved in fast skeletal muscle development by repression of slow muscle formation (van Rooij et al., 2009 and Bell et al., 2010). In zebrafish myogenesis, hedgehog signaling induces Prdm1a transcriptional repressor which represses Sox6 expression in adaxial cells, resulting in formation of slow muscle fibers (Roy et al., 2001; Baxendale et al., 2004 and von Hofsten et al., 2008). Wang et al. (2011) reported that in

zebrafish, miR-499 reinforces Sox6/*MYH14*/miR-499 transcriptional network by repressing Sox6 translation in slow muscle progenitor unveiling an interesting question whether miRNAs can control myosin switching and contractile protein gene programs in fast versus slow skeletal muscle in case of fish.

In contrast to mammalian counterpart, highly diverged MYH multigene family is reported for fish (Gerlach et al., 1990; Kikuchi et al., 1999; Watabe and Ikeda, 2006 and Ikeda et al., 2007). In case of common ancestor of vertebrates, 2 rounds of whole genome duplication (WGD) occurred and a third WGD occurred in fish lineage (Amores et al., 1998; Elgar et al., 1999; Postlethwait et al., 2000; Woods et al., 2000; Smith et al., 2002, McGuigan et al., 2004), resulting in much higher number of *MYHs* and suggesting their function and expression patterns might also have diverged in fish musculature. Phylogenetic analysis clearly showed duplication of the *MYH14*/miR-499 locus after the divergence of spotted gar, indicating that the teleostei-specific WGD provided present-day *MYH14*/miR-499 paralogs in teleost fish (Bhuiyan et al., 2013). The existence of multiple *MYH14* and miR-499 genes in various teleost fish suggests their expressional and functional versatilities. Torafugu *MYH14-1* (MYH_{M5}) expression was observed in both slow and cardiac muscles in the developmental and adult stages, whereas *MYH14-2* (MYH_{M3383}) expression was restricted to adult slow muscle (Ikeda et al., 2007 and Akolkar et al., 2010). Zebrafish *MYH14-1* was expressed in whole myotomal and cardiac muscles in the early developmental stages and in slow and intermediate muscles in the adult stage (Kinoshita et al., 2011). Furthermore, our

recent study demonstrated that medaka miR-499 expression differed from the above-mentioned *MYH14* expression (Sharmin et al., 2013). Meanwhile, the transcriptional regulatory regions have been mapped for certain number of fish muscle-specific genes which identified unique and shared *cis* acting elements (Du et al., 2003; Kobiyama et al., 2006; Chen et al., 2007; Liang et al., 2008). In case of fish, the *cis*-acting elements located at various distances upstream or downstream of the skeletal muscle-specific genes responsible for development as well as tissue-specific expression are partially known. The expression regulation of MYH14 paralogues/myomiRs by *in vivo* reporter assay and their function in muscle fiber-type specification was examined by knockdown analysis in this current chapter.

Methods

Construction of physical map around MYH14

The ensemble genome browser was used to determine the syntenic organization of human *MYH14* and its intronic microRNA (miR-499) with its orthologous genes for zebrafish and torafugu. Gene names are written in italics and protein names in roman characters throughout the text, according to a widely accepted use. Database versions are as follows: human (GRCh37, release 60), zebrafish (Zv9), torafugu (FUGU4).

Fish

Live specimens of adult zebrafish and medaka (average body weight 1.4g and 0.78 g respectively) were reared in local tap water with a circulating system at 28.5°C under a 14:10-h light-dark photoperiod raised at The University of Tokyo. Various tissues including brain, fast, slow and cardiac muscles were dissected after instant killing by cutting head of the fishes. These tissues were immediately frozen in liquid nitrogen and stored at –80°C until use for total RNA extraction. Slow muscles were dissected beneath the lateral surface of the myotome. Embryos of zebrafish and medaka were obtained by natural spawning raised at The University of Tokyo and reared up to larval stages. The developmental stage was determined by dpf and by morphological criteria. 2 dpf zebrafish larvae and adult tissues were used for reverse transcription-polymerase chain reaction (RT-PCR). Whole embryos, larvae and adult tissues from zebrafish and were used for *in situ* hybridization.

Preparation of total RNA

Total RNAs were extracted from adult zebrafish tissues by using ISOGEN solution (Nippon Gene, Tokyo, Japan). RNA concentrations were determined at 260 nm using a Nanophotometer™ 95 (Implen GmbH, Munich, Germany) and the RNA quality was determined by electrophoresis in a denaturing agarose gel containing 50% formamide. Total RNA pellets were stored at −80°C.

Generation of MYH14-1: EGFP Stable Transgenic Zebrafish

To confirm the expression patterns of *MYH14-1*:P5710 during the early muscle development stage, stable transgenic lines were established. We adapted Tol_2 transposon vector system which was originally identified in medaka (Koga et al., 1996) later used as a vehicle for efficiently integrate ectopic DNA into the zebrafish genome (Kawakami et al., 2000, 2004). Transient injection of Tol2 based plasmids was previously shown as a valuable method to analyze tissue specific promoters in zebrafish (Fisher et al., 2006; Korzh, 2007). In this method, the *MYH14-1*:P5710 construct was co-injected with transposase mRNA, which was transcribed from pCS-TP vector. The injected embryos were subjected to the transient embryonic excision assay (TEEA) at 8-10 h after microinjection for confirmation of the excision. The excision product was detected from all embryos injected with the transposase mRNA and a transposon-donor plasmid containing T2AL200R150G. Muscle-specific *EGFP* expressing embryos were selected and raised to adulthood. After 3 months, the surviving founders were mated with wild-type zebrafish to find successful germline-transmitted transgenic zebrafish. Stable muscle-specific *EGFP*-expressing zebrafish lines (F1) were established from 10 founders tested. About 50% of F2 offspring embryos carried muscle-specific *EGFP* expression by outcrossing the F1 fish with wild-type fish. All of the F1 transgenic line fish were mated and produced normal offspring. This generation of F_2 transgenic line is referred as stable transgenic *MYH14-1* zebrafish line.

Bioinformatic analysis

The nucleotide sequences of 5'-flanking sequences of torafugu *MYH14-1* (*MYH*$_{M5}$), zebrafish *MYH14-1*, *MYH-2*, *MYH-3* and 3′-flanking sequences, and the medaka miR-499 stem-loop sequences, which contain *Snai1* and *TRPC4AP* genes, were retrieved from the Ensembl genome browser. The homology search on the flanking sequences was carried out using the mVISTA alignment program through the vista server (http://genome.lbl.gov/vista/index.shtml).

Knockdown of miR-499

For the knockdown analysis, miRCURY LNA inhibitors (Exiqon) were used, which were antisense oligonucleotides complementary to miR-499. When inhibitors are introduced into the cells, they sequester miR-499 in highly stable hereto-duplexes thereby effectively prevent miR-499 hybridizing with its normal cellular interaction partners. Introduction of the inhibitor into the cell was done by microinjection. To perform microinjection, fertilized eggs of wild zebrafish and medaka *MYHL1-2.6k-DsRed* stable line fish were collected at 15-30 min after fertilization from fish rearing facility of the Department of Aquatic Bioscience, The University of Tokyo. The inhibitor was used at 100 nM after dilution into distilled water and introduced by using microinjection into fertilized eggs at one to two-cell stages. Embryos were reared at 28°C and were collected at different stages embryos up to larval stages for further analysis.

Real-time PCR analysis

Total RNA was obtained from miR-499 knockdown 4dpf zebrafish *MYH14-1:5710* stable transgenic line and from the control. For relative quantification, reaction was performed in a total volume of 20 ml, containing10 ml 2_SYBR premix Ex Taq II kit (Takara), 10 mm each of primers,1 ml diluted template cDNA (about 10ng) and 0.4 ml ROX reference dye. Real-time quantitative PCR was performed with ABI Prism7300 Sequence Detection System (Applied Biosystems). PCR consisted of pre- incubation at 50°C for 2min and denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15s and annealing and extension at 60°C for 1min. The samples were analyzed in triplicates. A housekeeping gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAP-DH) was selected as a reference for the relative expression levels of EGFP, which were calculated using the comparative C_T difference method [34]. The GAPDH primer sequence was forward 5'-TGTTTCCTCGTCCCGTAG-3' and reverse 5'-CAATCTCCACTTTGCCACT-3'. Primers used in real- time PCR for EGFP are listed in Supplementary Table S1.

Statistical analysis

For various constructs comparison of the percentage of embryos with EGFP expression in muscle fibers and relative quantification of target gene, statistical analyses were carried out using one-way analysis of variance (ANOVA) followed by Tukey's test in Statistical Package for Social Science(SPSS) version14 (SPSS Inc.,IL,USA). Data were represented as the mean SE and the differences were considered

significant at $P<0.05$.

Immunohistochemical analysis

Immunohistological observations were conducted to clarify types of muscle fibers that express *EGFP*. *EGFP*-expressing embryos were stained with F310 (Crow and Stockdale, 1986) and F59 (Crow and Stockdale, 1986; Devoto et al., 1996) monoclonal antibodies specific to fast-type myosin light chain and slow-type MYHs, respectively. For immunohistochemistry, after microinjection mentioned above, miR-499 knockdown zebrafish embryos and larvae were fixed with 4% PFA (Westerfield, 1993) in Tris-buffered saline (25 mM Tris, 137 mM NaCl, 2.7 mM KCl, pH 7.4) containing 0.1% Tween 20 (TBSTw) overnight at 4°C. Fixed embryos were washed with TBSTw and blocking was performed with 1.5% blocking solution (Roche) in TBSTw. Transverse sections were prepared at a thickness of 16 μm using a cryostat Tissue-Tek Cryo3 (Sakura Finetech, Tokyo, Japan) at -20°C before the first immunoreactions. Both whole mount and cryosection samples were used for immunohistochemistry. F59 antibody (Developmental Studies Hybridoma Bank, Iowa, USA) with a dilution of 1:20 and Sox6 monoclonal antibody (Abnova GmbH, Germany) with a dilution of 1:100 in 1.5% blocking solution were used as primary antibodies. These are specific to slow-type MYH and Sox6, respectively. Immuno-reaction with the first antibody was performed overnight at 4°C. After incubation, embryos were washed with TBSTw and labeled with the second antibodies, together with anti-mouse IgG Alexa Fluor

555 (Invitrogen) and anti-rabbit IgG Alexa Fluor 488 (Invitrogen) at a dilution of 1:250 for overnight at 4°C. The localization of cell nuclei was observed after staining for 10 min with 4', 6-diamidino-2-phenylindole (DAPI) (Roche) diluted at 1 μg/ml in TBSTw. The signals in the whole mount and cryosection samples were visualized using an Olympus FluoView 1000 confocal laser scanning microscope (Olympus).

Results

Promoter analysis of MYH-1, MYH-2 and MYH-3 paralogs

To identify the minimal promoter that can recapitulate *MYH14* paralogs expression in the myotomal compartment, a series of constructs containing progressive deletions from the 5'-end were fused to the enhanced green fluorescent protein gene (*EGFP*) in pT2AL200R150G vector were generated for *MYH14-1*, *MYH14-2* and *MYH14-3*. In *MYH14-1*, the resulting constructs contained fragments of 7528, 6205, 5710, and 2941bp, respectively. When *MYH14-1*:P7528, *MYH14-1*:P6205, *MYH14-1*:P5710 construct was microinjected into one to two-cell embryos, approximately 90% of the injected embryos had *EGFP* expression in the myotomal compartments and in heart (Fig.1A-I). Embryos microinjected with P2941 displayed a gradual reduction in percentages of *EGFP* expression per embryo in the myotomal compartments and in heart as well (Fig.2J-L). The EGFP expression of *MYH14-1* started in the migratory slow muscle fibers originating from adaxial cells at 14 hours post fertilization (hpf; 10 somite

stage) (Fig.3M-P). For *MYH14-2*, 5630b, 4960b, 4860b gene constructs were generated. However, none of the *MYH14-2* constructs generate any EGFP expression and *MYH-2:P5630-Tol$_2$-EGFP* reporter construct injected embryo and larva is shown in Fig.3A-C. *MYH-3:P5641-Tol$_2$-EGFP* reporter construct generated a strong expression of its reporter gene along the slow myotomal muscle fibers and in heart (Fig.3D-F). These data suggests that the 5710bp and 5641bp of the 5'-flanking region of zebrafish *MYH14-1* and *MYH14-3* respectively contained the essential regulatory sequences for cardiac/skeletal muscle-specific expression. For *MYH-1*, *EGFP*-expressing muscle fibers of P5710 injected larvae consisted both fast and slow type *MYHs* which reacted with both slow muscle specific F59 antibody (Fig.3G) and fast muscle specific F310 antibody (Fig. 3H). On the contrary, *MYH14-3* reacted exclusively with slow type muscle specific antibody F59 (Fig.3I,J).

MYH14-1:EGFP stable transgenic zebrafish showed both slow and fast muscle specific expression

To confirm the expression patterns of *MYH14* paralogs during the early muscle development stage, stable transgenic lines were established. The temporal and spatial *EGFP* expression was analyzed in these transgenic lines. In the stable transgenic line, the expression was detected at 10 somite stage and continued at 24 hpf (Fig. 4A). After hatching, *EGFP* continued to be expressed in the whole myotomal region of larva and in the cardiac muscle (Fig. 4B,C). The observed *EGFP* expression in the stable transgenic line was consistent with the transient expression pattern of *MYH14-1*:P5710.

Immunohistochemistry confirmed expression of *MYH14-1* in both the fast/slow muscle fibers (Fig.4D,E). Interestingly for *MYH14-1* expression, EGFP was concentrated especially along the horizontal myoseptum regions of the myotome.

Conserved regions are dispensable for MYH14-1 expression

Our previous studies showed, a small 5'-flanking region is enough to regulate spatio-temporal expressions of various fish *MYH*s ([Liang et al., 2007], [Ono et al., 2010], [Yasmin et al., 2011] and [Asaduzzaman et al., 2011]). Therefore, we compared 5'-flanking sequences of *MYH14* between torafugu and zebrafish. Within the genes of 10kb of 5'-upstream region, the comparison between torafugu and zebrafish *MYH14-1* revealed 3 conserved regions from 415bp to 485bp, 6665bp to 6712bp and 7339bp to 7383bp with (CR-1,CR-2 and CR-3) and 5 putative Sox6 binding sites (Sox6-1, Sox6-2, Sox6-3, Sox6-4, Sox6-5 and Sox6-6) (Fig.5A). In contrast, comparison between torafugu *MYH14-1* with that of zebrafish *MYH14-2* and *MYH14-3* revealed no significant homology (Fig.Supplementary Fig.). To precisely map the promoter region of zebrafish *MYH14-1*, three deletion constructs (⊿CR-1, ⊿CR-2 and ⊿CR-3) of P5710 were generated and microinjected into one to two-cell embryos. Constructs of P5710 recapitulated a cardiac/skeletal expression which actually excludes the distal two conserved regions and suggests, the distal conserved regions (CR-2 and CR-3) had no effect on *MYH14-1* expression (Fig.). Although *MYH14-1:5710bTol*$_2$-EGFP contained the proximal conserved region recapitulated its expression

completely and deletion of that single proximal region (CR-1:-415b~-549b) from this transgene did not generate any effect on cardiac/skeletal muscle specific EGFP expression (Figure 3A-E). This suggests, all of the three conserved regions are dispensable for *MYH14-1* expression. We also performed comparison of the upstream sequences between *MYH14-1* with that of *MYH14-2* and *MYH14-3*. Sequence comparison between *MYH14-1* with *MYH14-2* and *MYH14-3* revealed two conserved regions (CNR) at more than 70% sequence homology (Fig.5C). The sequence comparison between *MYH14-2, MYH14-3* revealed 100% conserved sequence starting from -5518b upstream up to -10000b (Supplementary Figure.) suggesting strong possibility of *MYH14-2* as pseudogene. Deletion of each conserved regions spanning from 1350-1474=124b, 1704 to 1931=227b (⊿CNR-1 and ⊿CNR-2) from p*MYH14-1:5710b-Tol$_2$EGFP* did not have significant impact on the transgene expression (Supplementary Figure). Compared to the p*MYH14-1:5710-Tol$_2$EGFP* reporter gene expression (Fig. 5A,C,G) deletion of ⊿CNR-1 and ⊿CNR-2 together, resulted significant reduction of *MYH14-3* expression in muscle fibers (Fig.5A,D,H). Deletion of each conserved regions (⊿CNR-1 and ⊿CNR-2) spanning from 3187 to 3336=149b and 3411 to 3637=226bp of *pMYH14-3:5641b-Tol$_2$EGFP* did not have significant effect on *MYH14-3* promoter activity (Suppl.Fig.1). Interestingly, compared to the non-deletion control (Fig.5 B,D,I), deletion of⊿HCNR-1 and ⊿HCNR-2 together (*pMYH14-3:⊿HCNR1&2:Tol$_2$EGFP*) resulted in significant reduction of *MYH14-3* expression in slow muscle fibers (Fig.B,F,J) and dramatically changed fiber type-specific expression of the reporter gene. Embryos at 2dpf injected with *pMYH14-3:5641b-Tol$_2$EGFP*

reporter gene expressed EGFP exclusively in slow muscle but *pMYH14-3:5641b⊿HCNR1&2:Tol$_2$EGFP* expressed EGFP not only in slow but also in fast muscle fibers (Fig.5K,L).

MYH14 promoter activity depends on the hedgehog signaling

Hedgehog signaling is important for slow muscle development in vertebrates. We examined the involvement of hedgehog signaling pathway, a positive regulator of adaxial cell differentiation using cyclopamine, a well-known hedgehog signaling inhibitor [37,38]. The stable transgenic embryos harboring *MYH14-1:5710b-Tol$_2$EGFP* and *MYH14-3:5641b-Tol$_2$EGFP* transgenic fish were treated with 10 μg/ml of cyclopamine. No significant changes observed on the survival rate among the cyclopamine treated and control eggs. Cyclopamine treated embryos showed partial and/or complete elimination of EGFP expression in slow muscle fibers (Ono, et al., 2010; Asaduzzaman et al., 2013). The *MYH14-1:5710b-Tol$_2$EGFP* stable transgenic zebrafish (Fig.8A,C) treated with cyclopamine showed only a partial inhibition of EGFP expression as the transgene in fast muscle fibers continued to express after cyclopamine treatment (Fig.6B,D). Meanwhile, the rate of zebrafish embryos at 2dpf harboring *MYH14-3:5641-Tol$_2$EGFP* transgene showed a significant reduction of EGFP expression after cyclopamine treatment (Fig.6E,F).

Sequence analysis of MYH14*/miR-499 locus flanking regions*

Intronic miRNAs can be independently transcribed from their host gene by using their own promoter positioned immediately upstream of miRNAs [19]. For medaka, miR-499 is transcribed lacking its host gene *MYH14*, which suggests the presence of its own promoter for transcription. Figure 4A shows comparisons of torafugu *MYH14-1* (MYH_{M5}) flanking regions with corresponding regions in zebrafish *MYH14-1* and medaka miR-499. In the case of medaka, *MYH14* was completely absent, with the exception of miR-499 (Figure 4A and supplementary Figure 2) and an intron immediately downstream of miR-499 (intronic conserved region in Figure 4A, supplementary Figure 3). Interestingly, the torafugu and zebrafish *MYH14s* 5′-flanking sequences showed clear similarity with those of medaka miR-499 (5′-upstream conserved regions in Figure 4A, supplementary Figure 4). Although the conservation in the zebrafish *MYH14-1* 5′-flanking region was not so obvious, it still contained several highly conserved regions (supplementary Figure 4).

Conserved regions analysis of miR-499

To validitate the conserved regions that can recapitulate miR-499 expression in the myotomal compartment, a 2000 bp sequences from the 5'-end of miR-499 were fused to the enhanced green fluorescent protein gene (*EGFP*) in pT2AL200R150G vector. When miR-499:P7528 construct was microinjected into one to two-cell zebrafish embryos, approximately 90% of the injected embryos had *EGFP* expression in the myotomal compartments specifically in fast muscle fibers and in heart at 1 dpf

and 2 dpf (Fig.5A-B). The EGFP-expressing cells tested positive for staining with F59 in the cardiac muscle (the anti-slow type MYH antibody; Fig. 5C) but none of the EGFP expressing cells in the myotomal muscle were tested positive with the same antibody (Fig. 5D,E). However, the EGFP expressing cells were stained successfully with F310 (anti-fast type MYH antibody; Fig. 5F). These data indicated that these hrGFP-expressing cells are components of superficial slow muscle fibers.

Knockdown of miR-499 in zebrafish and medaka developmental stage

miR-499 inhibited zebrafish larvae showed a loss of a part of slow muscle fibers at the lateral surface of the myotome both in the hypaxial muscle region and epaxial muscle region (Fig. 7A). The control zebrafish embryo showed a clear slow muscle layer formation (Fig. 7B). One of the targets of miR-499 is transcriptional repressor, Sox6 (Bell et al., 2010). Inhibition of miR-499 caused strong expression of Sox6 ectopically in some slow muscle fibers (Fig.7C). In control larvae, expression of Sox6 was restricted in fast muscle fibers (Fig.9D). Further confirmation on the effect of miR-499 knockdown on zebrafish muscle fibers using slow muscle specific p*M86-2614-Tol2EGFP* stable transgenic line is carried out. miR-499 knocked down transgenic larvae at 4 dpf showed a loss of slow muscle fibers in contrast to the control (Fig.7 E,F).

We validated down regulated expression of *MYH14-1* for miR-499 knockdown using p*MYH14-1:5710b-Tol*$_2$*EGFP* stable transgenic line by quantitative real-time PCR (qRT-PCR) analysis.

Quantitative real-time PCR consistently showed that the normalized mRNA expression of EGFP to GAPDH was significantly decreased in p*MYH14-1:5.7Tol$_2$EGFP* stable transgenic line than to control groups (Fig.7.G). Due to unavailability of slow muscle specific antibody for medaka, we used sarcomeric myosin heavy chain specific MF20 antibody to observe the effect of miR-499 knockdown on medaka slow muscle. However, the slow and fast muscle differentiation was not really evident using MF20 antibody. Therefore, a medaka transgenic fish *mMYHL1-2.6kbDs-Red* expressed in a fast muscle specific manner in medaka (Ono et al., 2010) was used. Applying MF20 and *mMYHL1-2.6kbDs-Red* together, slow muscle in green stain from fast muscle in purple stain can be differentiated. A clear slow muscle formation by green staining is observed for control medaka larvae (Fig.9 A,C,D) where the fast muscle is evident in purple colorization (Fig.9 B) and slow muscle by green stain (Fig.9D) .The results of knockdown of miR-499 in medaka larvae at 10 dpf showed marked loss of slow muscle as slow muscle can be evident by green stain at the outer end of myotome and at the horizontal myoseptum region (Fig. 9 F,G) whereas the fast muscle is separated from green slow muscle by purple colorization (Fig. 9E).

Discussion

In this present study, the 5710b and 5641b of the 5'-flanking region of zebrafish *MYH14-1* and *MYH14-3* respectively contained the essential regulatory sequences for cardiac/skeletal muscle-specific expression. These results supports that the promoter activity is conserved across different fish species. The transient

myotomal muscle-specific expression of *MYH14-1* was confirmed by generating stable transgenic zebrafish lines. However, stable transgenic zebrafish showed *MYH14-1* expression in both slow and fast muscle fibers and along the horizontal myoseptum region which is actually a connective tissue partition developing at the apex of the chevron-shaped myotome and separating dorsal (epaxial) and ventral (hypaxial) body wall muscle masses (Kimmel et al., 1995). In zebrafish, the decision for a cell to generate fast or slow muscle is made early in development while the somites are still forming (Devoto et al., 1996). Fate-mapping studies showed that the adaxial cells, positioned adjacent to the notochord, give rise to slow muscle cell types. As somites form, most adaxial cells migrate laterally to become superficial slow fibers (SSF), while a few adaxial cells remain adjacent to the notochord and become muscle pioneer cells (MPCs) and are found in slow-muscle cells located at the horizontal myoseptum (Hatta et al., 1991; Ochi & Westerfield, 2007). Expression of *MYH-1* in horizontal myoseptum therefore, suggests its essential organizational roles for somite structure which subsequently imparts guidance cues to pattern other developing systems. Apart from the slow muscle-specific expression, both the endogenous *MYH14-1* and *MYH14-2* the reporter gene were also found to be expressed in cardiac muscles (see Fig.). However, our present study could not conceive the promoter regulation of *MYH14-1 and MYH-2* in cardiac muscle. Among the paralogs, the 5'-flanking region of zebrafish *MYH14-1* with its torafugu ortholog revealed the existence of two distal and a single proximal conserved regions. However, deletion study of these conserved regions had no effect on *MYH14-1* expression in myotomal muscle fibers in case of zebrafish.

The 5'-flanking sequences comparison among *MYH14-1* with that of *MYH14-2 and MYH14-3* showed two conserved regions among them. Deletion of each of these two conserved regions or deletion of both the regions together from *MYH14-1* revealed no significant changes of reporter gene expression in the myotomal and cardiac muscle. However, deletion of each of the conserved regions from *MYH14-3* reduced the promoter activity and the number of EGFP-expressing slow muscle fibers per embryo was also reduced (see Fig.). The promoter activity of *pMYH14-3:5641b⊿CSR1and-2-Tol$_2$EGFP*, which lacks both the conserved regions was reduced and results in ectopic expression of the reporter gene in fast muscle fibers. This finding is suggestive of the presence of unidentified *cis*-acting elements in the *MYH14-1* and *MYH14-2* promoter.

In zebrafish embryos, superficial slow muscle fibers are subdivided into 3 components, according to their *MYH*-expressing repertoire and hedgehog signal dependence (Elworthy et al., 2008). One component consists of primary slow muscle fibers that differentiated from adaxial cells. The primary slow muscle fibers require hedgehog signaling during their development, expressing *MYH* named *smyhc1*. After formation of the primary slow muscle, 2 types of secondary slow muscle fibers participate in formation of the superficial slow muscle. One group expresses *smyhc1* just like primary slow muscle fibers do, but this group's differentiation does not require a hedgehog signal. The other group expresses *smyhc2*, and this group's differentiation depends on hedgehog signaling. Evidently, the *MYH14-1* and *MYH-2* promoter works in primary slow muscle fibers as revealed by expression of the reporter gene in adaxial cells (Fig.).

On the other hand, it is unclear whether *MYH-1* promoter works in secondary slow muscle fibers. Because the inhibition of hedgehog signaling partially repressed the reporter gene expression (Fig.), it is likely that some percentage of hrGFP-expressing cells require no hedgehog signal for their differentiation. Interestingly, sequence comparison analysis showed highly conserved 5′-flanking regions between torafugu MYH_{M5} and medaka miR-499 (see Figure 5A). The spatio-temporal expression of the major skeletal *MYHs* in teleost fish is regulated by small regions scattered throughout the 5′-flanking sequence (Liang et al., 2007; Ono et al., 2010; Yasmin et al., 2011; Asaduzzaman et al., 2011). Recently, Yeung et al. (2012) reported promoter activity in a 6.2-kb upstream sequence of mouse *MYH14* that mimics endogenous *MYH14* and miR-499 expression. Therefore, these conserved regions in the 5′-flanking sequence may act as a promoter for the spatio-temporal expression of *MYH14*, and the regulatory sequences are conserved in medaka miR-499 despite the loss of the *MYH14* gene. We could also speculate that miR-499 has its own promoter as do some intronic miRNAs. In fact, Matthew et al. [37] reported uncoupled *MYH14* and miRNA-499 expression in mice, suggesting the independent transcriptional regulation of miR-499 from *MYH14*. Isik et al. [38] found a conserved region immediately upstream of some intronic miRNAs in *C. elegans* and demonstrated in promoter activity the conserved region. An intronic sequence immediately downstream of miR-499 is conserved among zebrafish, torafugu, and medaka, as shown in Figure 4A, which could be the miR-499 promoter. To validitate such possibilities, we generated miR-499 upstram conserved sequence containing constructs with the reporter

vector which resulted in expression of the reporter construct exclusively in cardiac and fast muscle fibres. This result suggests, the small intronic conserved regions might have an important contribution for miR-499 expression in a slow/cardiac specific manner. These findings can potentially explain why miR-499 has remained despite the loss of *MYH14* in some teleost fish genomes.

Recent studies have demonstrated that Sox6 binds to conserved cis-regulatory elements in slow-twitch fiber genes to represses their transcription in adult fast-twitch muscle (Quiat et al., 2011). miR-499 has been suggested to play a pivotal role in lineage-specific activity of Sox6 in mammals based upon the finding that the 3′UTRs of human, mouse and rat *Sox6* contain consensus recognition sites for the miR-499 and the demonstration that miR-499 reduces *Sox6* mRNA levels both in vitro and in vivo (McCarthy et al., 2009; van Rooij et al., 2009; Bell et al., 2010). It has been revealed that similar to mammals, the functionally conserved zebrafish miR-499 reinforces Sox6/*MYH14*/miR-499 transcriptional network by repressing Sox6 translation in slow muscle progenitor (Wang et al. 2011). Such a mechanism also enables myomiRs to function as binary regulators of slow versus fast muscle gene programs activating slow and repressing fast through their downstream targets. It was reported that miR-499 knockout mice showed a substantial loss of type I muscle fibers in the soleus by the reduced expression of slow muscle specific *β-MYH* (McCarthy et al., 2009). This muscle atrophy caused by miR-499 knockout is consistant with our case where miR-499 inhibition in zebrafish resulted in loss of slow muscle in larvae. Actually, strong ectopical Sox6 expression in slow muscle was obtained in miR-499 knockdown zebrafish

larvae (see Fig. 9C). This line of information clearly indicates that the slow/cardiac muscle-specific expression of miR-499 is important for fish and mammalian muscle formation. The fact that medaka genome also contains miR-499 despite lack of *MYH14*, and medaka miR-499 is still expressed in slow and cardiac muscles (Sharmin et al., 2013) well indicates pivotal role of miR-499. The present study thus, confirmed the transcriptional network of *MYH14* and miR-499. Recent findings have proved that some other *MYH*s also contained intronic miRNA and such *MYH*-encoded miRNAs regulate expression of various muscle specific genes including *MYH*s themselves (van Rooij et al., 2009). In mouse, *α-MYH* encodes miR-208a which is expressed specifically in the heart and *β-MHC* gene encode a closely related miRNA, miR-208b which is co-expressed with its host gene, showing highest expression in type I myofibers and a lower expression in heart (van Rooij et al, 2009). In mammals, miR-208a regulates the expression of two slow myosins and their intronic miRNAs, *MYH7*/miR-208b and *MYH7b*/miR-499, respectively. Thus, these myo-miRNAs seemingly acts to support normal slow-muscle formation. These lines of information, together with the present findings, provide new aspects where a network of miRNAs affects not only the expression of the major contractile proteins of muscle, but also act more broadly to control muscle gene expression. In this regards, expression regulation of *MYH14*/miR-499 paralogues and functional analysis of miR-499 in zebrafish and medaka are thought to be an interesting piece of finding. Future comparative analyses on physiological functions of miR-499 will hopefully shed light on the mechanisms involved in the formation of species-specific musculature observed in fish.

Chapter 5

General Discussion and Conclusion

General Discussion

The skeletal and cardiac muscle contractility and bioenergetic requirements vary among vertebrates as well as spatiotemporally within species. Multiple myosin proteins serve to meet these variable demands (reviewed by Schiaffino and Reggiani 1996). Apart from playing the key role as primary determinant of the efficiency of muscle contraction, the functions of myosin genes extends far beyond the mere expression of proteins to the control of myriad functions of striated muscles through a network of miRNAs. Thus, *MYHs* enoded by their intronic myomiRNAs plays vital role in switching muscle properties. However, our knowledge of the MYHs/myomiRNA gene family in teleosts is limited to other vertebrates. Thus, in our present study we examined detailed genomic distribution and expression of myomiRNAs and their host myosin genes as well as transcription regulation of myomiRs and their function in muscle fiber-type specification (chapter 2- 4).

*Divergent distribution and expression of MYH14/*miR-499 *genes in teleosts*

In this study, the genomic distribution of myosin genes and their intronic myomiRs in three representative teleost genomes of torafugu, zebrafish and medaka was examined. Previously it has been reported that fish genomes contain *MYH14* orthologues (Ikeda et al, 2007; Watabe and Ikeda, 2006; Ikeda et al, 2004 and McGuigan et al.,2004) and miR-499 has also been found in the intron of *MYH14* on the zebrafish genome (Kloosterman et al.,2006). In our study, we elucidated the evolutionary history of *MYH14*/miR-499 in teleost genome. Interestingly, multiple *MYH14*/miR-499 loci are found on various teleost fish genomes (Chapter 2). The synteny and phylogenetic of analyses of *MYH14*/miR-499 depicts the unique evolutionary history of *MYH14* and its myomiR, miR-499. The teleost specific duplication and several subsequent rounds of species-specific gene loss events took place. Thus, the evolutionary history of the *MYH14*/miRNA-499 suggests divergent distribution of *MYH14* and miR-499 genes in different teleost lineage. Our previous investigation on comprehensive genome-wide screenings of MYHs revealed,

fish contain a higher number of *MYHs* in their genomes than their mammalian counterparts (Watabe and Ikeda, 2006; Ikeda et al., 2007; Ikeda, et al 2004; Liang et al., 2007). Together with present findings, the existence of a highly varied distribution of *MYH14* and miR-499 in teleost fish has been demonstrated. An *MYH14* paralog, lacking miR-499, exhibited an accelerated rate of evolution compared with those containing miR-499, suggesting a putative functional relationship between *MYH14* and miR-499 (Chapter 2). Again, teleost genome contains multiple MYHs orthologues at *MYH6*/*vmhc*/miR-736 locus (Chapter 3). In contrast to mammals, *MYH6* is present in teleots but it lacks any miRNA within its intron (Nachtigall et al., 2014). However, our evolutionary analysis depicts, teleost contain a single isoform among the cardiac muslc specific MYHs, *vmhc*, encoded by an intronic miRNA, miR-736. Our sequence analysis indicates, miR-736 is an orthologue of miR-208. Recent findings suggested, *MYH7*/miR-208b is completely absent in cartilaginous and ray finned fish genome (Nachtigall et al., 2014). Our synteny and phylogenetic analysis corroborates absence of *MYH7*/miR-208 in zebrafish/torafugu/medaka genome. In zebrafish, total three *MYH14*/miR-499 pairs are present in its genome. Previous study revealed that *MYH14-1* was expressed in whole myotomal region including fast muscle area at embryonic stage but in superficial slow, intermediate, and cardiac muscles at adult stage (Kinoshita et al., 2011). In this study, expression of *MYH14-1* and *MYH14-3* was observed whereas transcripts of *MYH14-2* were not detected in both developmental and adult stages. The transcripts of *MYH14-3* were specifically expressed in slow and cardiac muscles at both developmental and adult stages. Meanwhile, torafugu genome contains two *MYH14*s. One paralogue named MYH_{M5} contains miR-499, whereas the other one named MYH_{M3383} lacks the miRNA. These two *MYH14* paralogues shows different expression pattern where MYH_{M5} expression was observed in both slow and cardiac muscles in the developmental and adult stages, and MYH_{M3383} expression was restricted to adult slow muscle (Ikeda et al., 2007 and Akolkar, 2010). In medaka, miR-499 is present but *MYH14* is completely absent in the genome. These results indicate that teleost *MYH14*s are highly diverged in their genomic structure and expression patterns.

Expression of miR-499 was also examined in the three teleost species. Interestingly, miR-499 expression is exceptionally conserved regardless of the varied expression of their host *MYH14s*. *In situ* hybridization showed that miR-499 was not expressed in the skeletal muscle at the embryonic stage of zebrafish, torafugu, and medaka. On the other hand, larvae of the three fish showed a clear expression of miR-499 in both cardiac and slow muscles. Similar to the larval stage, adult zebrafish and torafugu showed high expression of miR-499 in cardiac and slow muscles, whereas adult medaka exhibited miR-499 expression only in the cardiac muscle. In torafugu, miR-499 was also detected in the erectors-depressors (ED) muscle which exhibit characteristics of slow muscle fibers. These miR-499 expression patterns were also confirmed by next-generation sequencing of small RNA libraries. It is noted that medaka miR-499 was even expressed in the absense of its host gene *MYH14*. Comparing the flanking sequences of *MYH14*/miR-499 loci between torafugu, zebrafish, and medaka revealed several highly conserved regions including an intronic sequence immediately downstream of miR-499, suggesting *cis*-regulatory elements have been functionally conserved in medaka miR-499 despite the loss of its host gene.

*Divergent distribution and expression of MYH6/vmhc/*miR-736 g*enes in teleosts*

MYH6 and *MYH7* are well-known cardiac *MYHs in mammals and contain* myomiRs, miR-208a and miR-208b, respectively within their introns. In adult mouse, *MYH6* is expressed specifically in the heart whereas *MYH7* was expressed in slow muscle fibers (van Rooij et al., 2009). *MYH7* also showed high expression in the developing heart but is down-regulated after birth. miR-208a is co-expressed with *MYH6* and regulates the expression of *MYH7*/miR-208b and *MYH14*/miR-499. Recent studies have revealed that miR-208b and miR-499 play redundant functions in muscle fiber-type specification by activating slow and repressing fast muscle fiber gene programs in mammals. In available teleost genome databases, *MYH7* was not detected. *MYH6* persists in cartilaginous and ray-finned fish genomes but missing any intronic miRNAs. In teleosts, known major cardiac *MYH* isoform is ventricular myosin

heavy chain gene (*vmhc*) which contains an intronic miRNA, miR-736. Sequence similality and phylogenetic analyses indicates *vmhc*/miR-736 are orthologue of *MYH6*/miR-208a. As well as *MYH14*/miR-499, syntenic and phylogenetic studies revealed that multiple orthologues of *MYH6*/*vmhc*/miR-736 are present in teleost genomes. Zebrafish genome contains a total of 10 *MYH6*/*vmhc* paralogues of which one of them possess miR-736 and the others lack the intronic miRNA. Torafugu and medaka genomes contain 3 and 5 *MYH6*/*vmhc* paralogues respectively, and only one paralogue in each fish contains miR-736. Kloosterman et al. (2006) reported that zebrafish miR-736 is expressed in eye and gut by northern blot analysis. Expression of miR-736 in adult torafugu and medaka was examined in the present study using next generation sequencing platform, showing trace expression in various tissues including muscle. These results indicate uncoordinated expression of miR-736 with its host gene, suggesting miR-736 is not a functional orthologue of mammalian miR-208s. Taken together with highly conserved expression of miR-499, these results suggest that the complete deletion of miR-208s function in teleost is counterbalanced by the role played by miR-499.

Functional analysis of miR-499 in teleost

To address mechanisms of expression regulation of diversified *MYH14* paralogues, *in vivo* reporter assay was performed by injecting reporter gene EGFP conjugating with various length of 5'-flanking sequence of zebrafish *MYH14* paralogues into zebrafish eggs. As a result, a 5710bp 5'-flanking region of *MYH14-1* and 5641bp of *MYH14-3* contained a necessary regulatory region to recapitulate their endogenous expression during embryonic and larval muscle development. *MYH14-3* promoter activated reporter gene expression specifically in slow muscle fibers, whereas the embryos of stable zebrafish transgenic line Tg:MYH14-1:5710bp-EGFP expressed EGFP in both slow and fast muscle fibers, well consistent with their endogenous expression in zebrafish. Immunohistochemistry and cyclopamine (an inhibitor of slow muscle development) treatment of the Tg:MYH14-1:5710bp-EGFP confirmed the reporter gene

expression in both fast and slow muscle fibers where EGFP was concentrated especially along with the horizontal myoseptum regions. 5'-flanking region of zebrafish *MYH14-1* and its torafugu orthologue shared two distal and a single proximal conserved region. However, deletion study of these conserved regions had no effect on reporter gene expression in case of zebrafish. On the other hand, additional two conserved regions were detected by comparing 5'-flanking regions of zebrafish three *MYH14* paralogues. Deletion of the two conserved regions significantly reduced the promoter activity of *MYH14-3* but no affect on that of *MYH14-1*, indicating that *cis*-regulatory elements of *MYH14-1* and *MYH14-3* are different in accordance with differential expression between the two *MYHs*.

As observed in previous section, medaka miR-499 is transcribed despite lacking its host gene *MYH14*, suggesting the presence of its own promoter for transcription. The present study also examined promoter activity of 5'-flanking sequence of medaka miR-499. As well as in zebrafish paralogues, various length of 5'-flanking sequence of medaka miR-499 was conjugated with EGFP reporter gene and injected into zebrafish eggs. Detail promoter analysis revealed, deleting the regions containing several highly conserved sequences of *MYH14*/miR-499 loci between torafugu and medaka upstream of miR-499 (4769/-1897) in medaka generated reporter gene expression specifically in cardiac and fast muscle fibers.

Loss of function experiment of miR-499 was performed by injection of antisense miR-499 oligonucleotide into medaka and zebrafish eggs. As expected, knockdowned larvae showed marked reduction of slow muscle fibers in zebrafish and medaka developmental stages. In zebrafish, such reductions were also validated by real-time PCR analysis using a transgenic line which expresses GFP in slow muscle fibers. As explained, mammalian miR-499 and miR-208b have redundant role in muscle fiber-type specification by activating slow muscle gene program. van Rooij et al. (2009) reported that knock down of each of miR-499 and miR-208a shows no affect on mouse muscle formation, but double knock down of the two miRNAs caused reduction of slow muscle fibers. On the other hand, in teleost, miR-499 suppression is enough to reduce the slow muscle fiber formation. Taken together with the fact of

trace expression of miR-736, teleost miR-208s orthologue, in muscles, miR-499 is a sole functional myomiR and thus have indispensable role in muscle fiber-type specification in teleost muscle formation.

Conclusion

Despite diversification of host *MYHs* in genomic organization and expression patterns, miR-499 expression was exceptionally conserved, indicating pivotal role of the myomiR in teleost muscle formation. Actually, knock dowm analysis of miR-499 showed perturbation in slow muscle formation during zebrafish/medaka growth, indicating that a myomiR-mediated regulatory network also works in fish muscle formation although the network composition will be different from that of mammals. On the other hand, diversification of host *MYHs* suggests their functional versatilities involved in diverged musculature in teleost. Fish is the most diverse vertebrate group. In response to the wide range of environmental and physiological conditions, the characteristics of fish musculature, including muscle fiber-type composition, are also highly diverse. It would be interesting to determine whether diversification in myomiRs and their host genes is related to physiological and ecological variations among teleost fish species.

References

1. Akolkar DB., Kinoshita S, Yasmin L, Ono Y, Ikeda D, Yamaguchi H, Nakaya M, Erdogan O, Watabe S. 2010. Fibre type-specific expression patterns of myosin heavy chain genes in adult torafugu *Takifugu rubripes* muscles. *J. Exp. Biol.* 213, 137-145.

2. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH.1998. Zebrafish hox clusters and vertebrate genome evolution. *Science*, 282:1711-1714.

3. Asaduzzaman M, Kinoshita S, Bhuiyan SS, Asakawa S, Watabe S.2011. Multiple *cis*-elements in the 5′-flanking region of embryonic/larval fast-type of the myosin heavy chain gene of torafugu, $MYH_{M743\text{-}2}$, function in the transcriptional regulation of its expression. *Gene*, 489:41-54.

4. Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*.116:281-297.

5. Baskerville S. and Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, 11:241-247.

6. Baxendale S, Davison C, Muxworthy C, Wolff C, Ingham P W, Roy S. 2004. The B-cell maturation factor Blimp-1 specifies vertebrate slow-twitch muscle fiber identity in response to Hedgehog signaling. *Nat. Genet.* 36, 88-93.

7. Bell ML, Buvoli M, Leinwand LA, 2010. Uncoupling of expression of an intronic microRNA and its myosin host gene by exon skipping. *Mol. Cell Biol.* 30, 1937-1945.

8. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. 2007. Mammalian mirtron genes. *Mol Cell*, 28:328-336.

9. Berg JS, Powell BC, Cheney RE, 2001. A millennial myosin census. *Mol. Biol. Cell* 12,780–794.

10. Bhuiyan SS, Kinoshita S, Wongwarangkana C, Asaduzzaman M, Asakawa S, Watabe S. 2013. Evolution of the myosin heavy chain gene MYH14 and its intronic microRNA miR-499: muscle-specific miR-499 expression persists in the absence of the ancestral host gene. *BMC evolutionary Biol*, 13, 122. doi:10.1186/1471-2148-13-142.

11. Bryson-Richardson RJ, Daggett DF, Cortes F, Neyt C, Keenan DG, Currie PD. 2005. Myosin heavy chain expression in zebrafish and slow muscle composition. *Dev. Dyn.* 233, 1018-1022.

12. Carlos I, Makoto PM, Esther A, Jose PC, Michael R, Manuel M. 2008. Selection of housekeeping genes for gene expression studies in larvae from flatfish using real-time PCR. *BMC Mol. Biol.* 9, 28.

13. Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* 21, 1146–1151

14. Darby IA, Bisucci T, Desmouliere A, Hewitson TD. 2006. In situ hybridization using cRNA probes: isotopic and nonisotopic detection methods. *Methods Mol. Biol.* 326, 17-31.

15. Desjardins PR, Burkman JM, Shrager JB, Allmond LA, Stedman HH: Evolutionary implications of three novel members of the human sarcomeric myosin heavy chain gene family. *Mol Biol Evol* 2002, 19:375-393.

16. Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJ, Bouchireb N, Cottage A, Yeo GS, Umrania Y, Williams G, Brenner S. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. *Genome Res,* 9:960-971.

17. Everett AW. 1986. Isomyosin expression in human heart in early pre- and post-natal life. J Mol Cell Cardiol 18*:* 607–615.

18. Garriock RJ, Meadows SM, Krieg PA. 2005. Developmental expression and comparative genomic analysis of Xenopus cardiac myosin heavy chain genes. *Dev. Dyn.* 233, 1287–1293.

19. Gerlach G, Turay L, Mailik K, Lida J, Scutt A, Goldspink G. 1990. The mechanisms of seasonal temperature acclimation in carp: a combined physiological and molecular biology approach. *Am. J. Physiol*. 259: 237–244.

20. Griffiths-Jones, S., 2004. The microRNA registry. Nucleic Acids Res. 32 (suppl 1):D109-D111.

21. Haddad F, Qin AX, Bodell PW, Jiang W, Giger JM, Baldwin KM.Intergenic transcription and developmental regulation of cardiac myosin heavy chain genes. Am J Physiol Heart Circ

Physiol 294: H29–H40, 2008.Hagiwara N, Yeh M, Liu A. 2007. Sox6 is required for normal fiber type differentiation of fetal skeletal muscle in mice. *Dev Dyn*, 236:2062-2076.

22. Hagiwara N, Ma B, Ly A. Slow and fast fiber isoform gene expression is systematically altered in skeletal muscle of the Sox6 mutant, p100H. Dev Dyn 234: 301–311, 2005

23. Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of the teleost fish. *J Mol Evol*, 59:190-203.

24. Ikeda D, Clark MS, Liang CS, Snell P, Edwards YJK, Elgar G, Watabe S. 2007. Genomic structural analysis of the pufferfish (*Takifugu rubripes*) skeletal muscle myosin heavy chain genes. *Mar Biotechnol*, 6:S462-S467.

25. Ikeda D, Ono Y, Snell P, Edwards YJ, Elgar G, Watabe S. 2007. Divergent evolution of the myosin heavy chain gene family in fish and tetrapods: evidence from comparative genomic analysis. *Physiol Genomics*, 32:1-15.

26. Ikeda D, Ono Y, Snell P, Edwards YJ, Elgar G, Watabe S. 2007. Divergent evolution of the myosin heavy chain gene family in fish and tetrapods: evidence from comparative genomic analysis. *Physiol. Genomics*, 32, 1-15.

27. Ikeda D, Toramoto T, Ochiai Y, Suetake H, Suzuki Y, Minoshima S, Shimizu N, Watabe S, 2003. Identification of novel tropomyosin 1 gene of pufferfish (Fugu rubripes) on genomic sequences and

tissue distribution of their transcripts. *Mol. Biol. Rep.* 30, 83-90.

28. Isik M, Hendrik CK, Berezikov. 2010. Expression patterns of intronic microRNAs in*Caenorhabditis elegans. Silence*, 1: 1-5.

29. Johnston IA. 2006. Environmental and plasticity of myogenesis in teleost fish. *J. Exp. Biol.* 209, 2249-2264.

30. Kikuchi K, Muramatsu M, Hirayama Y, Watabe S. 1999. Characterization of the carp myosin heavy chain multigene family. *Gene*, 228, 189–196.

31. Kim VN, Han J, Siomi MC: Biogenesis of small RNAs in animals. 2009. *Nat Rev Mol Cell Biol*, 10:126-139.

32. Kim YK, Kim VN.2007. Processing of intronic microRNAs. *EMBO J*, 26:775-783.

33. Kinoshita S, Bhuiyan SS, Ceyhun SB, Asaduzzaman M, Asakawa S, Watabe S. 2011. Species-specific expression variation of fish *MYH14*, an ancient vertebrate myosin heavy chain gene ortholog. *Fish. Sci.* 77, 847-853.

34. Kloosterman WP, Wienholds E, Bruijn E de, Kauppinen S, Plasterk RH. 2006. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat Methods*, 3:27-29.

35. Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, Verheul M, Cuppen E, Plasterk RH, 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acid Res.* 34, 2558-2569.

36. Liang CS, Ikeda D, Kinoshita S, Shimizu A, Sasaki T, Asakawa S, Shimizu N, Watabe S. 2008, Myocyte enhancer factor 2 regulates expression of medaka *Oryzias latipes* fast skeletal myosin heavy chain genes in a temperature-dependent manner. *Gene*, 407:42-53.

37. Liang CS, Kobiyama A, Shimizu A, Sasaki T, Asakawa S, Shimizu N, Watabe S. 2007. Fast skeletal muscle myosin heavy chain gene cluster of medaka *Oryzias latipes* enrolled in temperature adaptation. *Physiol Genomics*, 29:201–214.

38. Lompre AM, Nadal-Ginard B, Mahdavi V. 1984. Expression of the cardiac alpha- and beta-myosin heavy chain genes is developmentally and hormonally regulated. J. Biol. Chem. 259, 6437-6446.

39. Mahdavi V, Chambers AP, Nadal-Ginard B.1984. Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem. *Proc Natl Acad Sci USA*, 81:2626-2630.

40. Matthew LB, Massimo B, Leslie AL.2010. Uncoupling of expression of an intronic microRNA and its myosin host gene by exon skipping. *Mol Cell Biol*, 30:1937–1945.

41. McCarthy JJ, Esser AK, Peterson AC, Dupont-Versteegden EE. 2009. Evidence of MyomiR network regulation of β-myosin heavy chain gene expression during skeletal muscle atrophy. *Physiol Genomics*, 39:219-226.

42.McCullagh KJ, Calabria E, Pallafacchina G, Ciciliot S, Serrano AL, Argentini C, Kalhovde JM, L omo T,Schiaffino S NFAT is a nerve activity sensor in skeletal muscle and controls

activity-dependent myosin switching. Proc Natl Acad Sci USA 101: 10590–10595, 2004.

43. McGuigan K, Phillips PC, Postlethwait JH. 2004. Evolution of sarcomeric myosin heavy chain genes: evidence from fish. *Mol. Biol. Evol.* 21, 1042–1056.

44. Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox, KA, Xing Y, Davidson BL. 2010. Structure and activity of putative intronic miRNA promoters. *RNA*, 16:495–505.

45. Muramatsu-Uno M, Kikuchi K, Suetake H, Ikeda D, Watabe S, 2005. The complete genomic sequence of the carp fast skeletal myosin heavy chain gene. *Gene*, 349, 143–151.

46. Nei M and Gojobori T. 2010. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol and Evol* ,3:418-426.

47. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC.2007. The mirtron pathway generates microRNA-class regulatory RNAs in drosophila. *Cell*, 130:89-100.

48. Ono Y, Liang C, Ikeda D, Watabe S, 2006. cDNA cloning of myosin heavy chain genes from medaka Oryzias latipes embryos and larvae and their expression patterns during development. *Dev. Dyn.* 235, 3092-3101.

49. Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* , 10:1890-1902.

50. Qin H, Hsu MK, Morris BJ, Hoh JF. 2002. A distinct subclass of mammalian striated myosins:

structure and molecular evolution of ‘superfast' or masticatory myosin heavy chain. *J. Mol. Evol.*, 55, 544–552.

51. Rinaldi C, Haddad F, Bodell PW, Qin AX, Jiang W, Baldwin KM. 2008. Intergenic bidirectional promoter and cooperative regulation of the IIx and IIb MHC genes in fast skeletal muscle. *Am J Physiol Regul Integr Comp Physiol.*, 295: 208–218.

52. Rossi AC, Mammucari C, Argentini C, Reggiani C, Schiaffino S. 2010. Two novel/ancient myosins in mammalian skeletal muscles: MYH14/7b and MYH15 are expressed in extraocular muscles and muscle spindles. *J. Physiol.* 588, 353–364.

53. Roy S, Wolff C, Ingham PW. 2001. The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo. *Genes Dev.* 15, 1563-1576.

54. Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass drosha processing. *Nature*, 448:83-86.

55. Saez LJ, Gianola KM, McNally EM, Feghali R, Eddy R, Shows TB, Leinwand LA. 1987. Human cardiac myosin heavy chain genes and their linkage in the genome. *Nucleic Acids Res*, 15:5443-5459.

56. Scemama JL, Vernon JL, Stellwag EJ. 2006. Differential expression of hoxa2a and hoxa2b genes during striped bass embryonic development. *Gene Expr. Patterns*, 6, 843-848.

57. Schachat F, Briggs MM. 1999. Identification of two patterns of exon organization in the human

striated muscle myosin heavy chain genes. *Mol. Cell Biol.* 10, 34a.

58. Schiaffino S, Reggiani C: Fiber types in mammalian skeletal muscles. *Physiol Rev* 2011, 91:1447-1531.

59. Schiaffino, S., Reggiani, C., 1996. Molecular diversity of myofibrillar proteins: gene regulation and functional significance. Physiological Reviews. 76, 371-423.

60. Shrager JB, Desjardins PR, Burkman JM, Konig SK, Stewart SK, Su L, Shah MC, Bricklin E, Tewari M, Hoffman R, Rickels MR, Jullian EH, Rubinstein NA, Stedman HH: Human skeletal myosin heavy chain genes are tightly linked in the order

61. embryonic-IIa-IId/x-ILb-perinatal-extraocular. *J Muscle Res Cell Motil* 2000, 21:345-355.

62. Smith SF, Snell P, Gruetzner F, Bench AJ, Haaf T, Metcalfe JA, Green AR, Elgar G. 2002. Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and human 20q. *Genome Res*., 12:776-784.

63. Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage, *Nature*, 428, 415–418.

64. Tajima F and Nei M.1984. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol and Evol*., 1:269-285.

65. Tamura K, Nei M.1993 Estimation of the number of nucleotide substitutions in the control region

of mitochondrial DNA in humans and chimpanzees. *Mol Biol and Evol.*, 10:512–526.

66. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.*, 28:2731-2739.

67. Ton C, Hwang DM, Dempsey AA, Tang HC, Yoon J, Lim M, Mably JD, Fishman MC., Liew CC. 2000. Identification, characterization, and mapping of expressed sequence tags from an embryonic zebrafish heart cDNA library. *J. Genome Res.* 10, 1915-1927.

68. Toramoto T, Ikeda, D, Ochiai Y, Minoshima S, Shimizu N, Watabe S. 2004. Multiple gene organization of pufferfish Fugu rubripes tropomyosin isoforms and tissue distribution of their transcripts. *Gene* 331, 41-51.

69. Vandepoele K, De Vos, W Taylor, JS Meyer A, Van de Peer Y, 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci.* USA 101, 1638–1643.

70. Wang X, Ono Y, Tan CS, Chai RJ, Philip C, Ingham PW. 2011. Prdm1a and miR-499 act sequentially to restrict Sox6 activity to the fast-twitch muscle lineage in the zebrafish embryo. *Development*, 138:4399-4404.

71. Watabe S, Ikeda D.2006. Diversity of the pufferfish *Takifugu rubripes* fast skeletal myosin heavy chain genes. *Comp Biochem Physiol.*, 1:28-34.

72. Watabe, S., 2002. Temperature plasticity of contractile proteins in fish muscle. *J. Exp. Biol.* 205, 2231-2236.

73. Watabe S, Ikeda D. 2006. Diversity of the pufferfish Takifugu rubripes fast skeletal myosin heavy chain genes. *Comp. Biochem. Physiol.*, 1: 28-34.

74. Weiss A, McDonough D, Wertman B, Acakpo-Satchivi L, Montgomery K, Kucherlapati R Leinwand L, Krauter K. 1999. Organization of human and mouse skeletal myosin heavy chain gene clusters is highly conserved. *Proc Natl Acad Sci USA*, 96:2958-2963.

75. Westerfield M. 1993. The Zebrafish book. A Guide for the Laboratory Use of Zebrafish (Danio rerio). University of Oregon Press, Eugene.

76. Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, Bruijn d. E, Horvitz HR, Kauppinen S, Plasterk HAR. 2005. MicroRNA expression in zebrafish embryonic development. *Science*, 309: 310-311.

77. Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS. 2000. A comparative map of the zebrafish genome. *Genome Res*, 10:1903-1914.

78.Yasmin L, Kinoshita S, Akolkar DB, Asaduzzaman M, Ikeda D, Ono Y, Watabe S. 2010. A 5′-flanking region of embryonic-type myosin heavy chain gene, $MYH_{M743\text{-}2}$, from torafugu (*Takifugu rubripes*) regulates developmental muscle-specific expression. *Comp Biochem Physiol.*, 6:76-81.

79. Yeung F, Chung E, Guess MG, Bell ML, Leinwand LA. 2012.Myh7b/miR-499 gene expression is transcriptionally regulated by MRFs and EOS. *Nucleic Acids Res.*, 40:7303-7318.

80. van Rooij E, Quiat D, Johnson BA, Sutherland LB, Qi X, Richardson JA, Kelm RJ Jr, Olson EN.2009. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev Cell*, 17:662-673.

81. van Rooij, E Liu N, Olson EN, 2008. MicroRNAs flex their muscles. Trends Genet. 24, 159–166.

82. von Hofsten J, Elworthy S, Gilchrist MJ, Smith JC, Wardle FC, Ingham PW. 2008. Prdm1-and Sox6-mediated transcriptional repression specifies muscle fiber type in the zebrafish embryo. *EMBO Rep.*, 9:683-689.