Characterisation of genes regulating muscle development and growth in two model puffer fish species (*Takifugu rubripes* and *Tetraodon nigroviridis*)

Matthew Graeme MacKenzie

A thesis submitted for the degree of Doctor of Philosophy

University of St Andrews

St Andrews, May 2006

This thesis is dedicated to my Mum and Dad, without their continual belief, love and support this work would never have been possible. Thank you!

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

List of Abbreviations

µg	Micrograms
µl	Microlitres
µm	Micrometres
µM	Micromolar
A1pp	Appr-1-p processing
ABCA3	ATP-binding cassette sub-family A member 3
AcAh	Acetic Anhydride
ACTG	Adenosine, Cytosine, Thymidine, Guanine (The genetic code)
ANOVA	Analyses of Variance
APC	Anaphase Promoting Complex
ATP	Adenosine Triphosphate
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate, P-Toluidine Salt
bHLH	Beta Helix-Loop-Helix
BMP	Bone Morphogenetic Protein
Bp	Base Pairs
BTB_POZ	BR-C, ttk and bab/ Pox virus and Zinc finger
CCNF	G2/ mitotic-specific cyclin F
CD34	Cluster of Differentiation Molecule 34
Cdks	Cyclin Dependent Kinases
cDNA	Complementary Deoxyribonucleic Acid
CDS	Complete Coding Sequence
CENTA1	Centaurin-alpha 1
CGI	Comparative Genomics Identification
CKI	Cyclin Depenent Kinase Inhibitor
cm	Centimetre
cm^3	Centimetres Cubed
c-myc	Cellular- Myelocytomatosis
CNS	Central Nervous System
cRNA	Complementary Ribonucleic Acid
Ct	Cycle Threshold
CYP2W1	Cytochrome P450 2W1
D-box	Destruction Box
DIG	Digoxigenin
DM	Double Muscling
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
DP-1	DNA Binding Protein-1
Ds	Double Stranded
DUF410	Domain of Unknown Function
EDTA	Ethylenediaminetetraacetic Acid
E_{GOI}	PCR Efficiency of Gene of Interest
EGTDC	Environmental Genomics Thematic Programme Data Centre
EST	Expressed Sequence Tag
E_{std}	PCR Efficiency of Gene of Interest
EtBr	Ethidium Bromide
fig	Figure
Flstn	Follistatin
FMstn-1	Fugu Myostatin-1

FMstn-2	Fugu Myostatin-2
FOX01a	Forkhead Box O1A
FoxK1	Forkhead Box K1
G0	Gap0 (Quiescent Stage of Cell Cycle)
G1	Gap1
G2	Gap2
G3PDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GLM	General Linear Model
GO	Gene Ontology
GOI	Gene of Interest
HGMP	The Human Genome Mapping Project
HIGS	Heat inactivated Goat Serum
HLH	Helix-Loop-Helix
HMG	High Mobility Group
Hox	Homeobox
Hpf	Hours Post Fertilisation
Hsp90	Heat Shock Protein 90
ICAT	Isotope-Coded Affinity Tag System
Id	Inhibitor of Differentiation
InsP4	Inositol 1,3,4,5-Tetrakisphosphate
kb	Kilobase
kg	Kilogram
KRAB	Krueppel-Associated Box
LAP	Latency-Associated Peptide
LB	Luria-Bertani
LIM	Lin-11, Isl-1 and Mec-3 genes
M phase	Mitosis Phase of Cell Cycle
MAPK	Mitogen-Activated Protein Kinase
Mb	Megabases
MEF	Myocyte Enhancer Factor
Mesp2	Mesoderm Posterior-2
mg	Milligrams
min	Minute
MIPP	Mouse Intracisterneal A Particle Promoted Placental Gene
ml	Millilitre
MLC	Myosin Light Chain
mM	Nanomolar
MPC	Myogenic Progenitor Cells
MPF	Mitosis Promoting Factor
MRC	Medical Research Council
MRF	Muscle Regulatory Factor
mRNA	Messenger Ribonucleic Acid
mRNP	Messenger Ribonucleoprotein
Mstn	Myostatin
Myf5	Myogenic Factor 5
Myf6	Myogenic Factor 6
MyoD	Myogenic Determination Factor
MYOG	Myogenin
Myr	Million Years Ago
NaCl	Sodium Chloride

NADH	Nicotinamide Adenine Dinucleotide (Reduced)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced)
NBT	Nitro Blue Tetrazolium
NCBI	National Center for Biotechnology Information
NERC	Natural Environment Research Council
NFATc2	Nuclear Factor of Activated T-cells Cytoplasmic Calcineurin-Dependent 2
nm	Nanometre
non-LTR	Non-long Terminal Repeat
OD	Optical Density
ORF	Open Reading Frame
Pax	Paired Box
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
per3	Period 3 (Circadian protein)
PFA	Paraformaldehyde
pg	Picograms
PKC	Protein Kinase C
pRb	Retinoblastoma Protein
psma5	Proteasome Subunit Alpha type 5
PtdInsP3	Phosphatidylinositol 3,4,5-Trisphosphate
qPCR	Quantitative Polymerase Chain Reaction
RASD1	Dexamethasone-induced RAS-related protein
RM	Red Muscle
RNA	Ribonucleic Acid
RNAi	Interfering Ribonucleic Acid
RSV	Rous Sarcoma Virus
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RTR	Retrotransposon
S phase	Synthesis Phase (Cell Cycle)
SART1	*Squamous Cell Carcinoma Antigen Recognized by T cells*
SCAF	Scaffold
Shh	Sonic Hedgehog
SL	Standard Length
Sort1	Sortilin Precursor 1
Sox8	SRY (Y-linked sex-determining gene) related high mobility box protein 8
SPRY	Spla, Ryanodine (Domain)
SRY	Y-linked sex-determining gene
SSC	Sodium Chloride Sodium Citrate
SSH	Suppressive Subtractive Hybridisation
T. nigroviridis	*Tetraodon nigroviridis*
T. rubripes	*Takifugu rubripes*
TAE	Tris Acetate EDTA Buffer
TE	Tris EDTA Buffer
TEA	Tetraethyl Ammonium
TEN	Tris EDTA Sodium Chloride Buffer
Tfeb	Transcription Factor EB
TGF-β	Transforming Growth Factor β
TrisHCl	Tris Hydrochloric Acid

Ts-cdc	Temperature Sensitive- Cell Division Cycle genes
TSP-1	Thrombospondin-1
TSS	Translation Start Site
UBP22	Ubiquitin carboxyl-terminal hydrolase 22
USA	United States of America
UTR	Untranslated Region
UV	Ultraviolet
V	Volts
Vamp3	Vesicle Associated Membrane Protein 3
WGD	Whole Genome Duplication
WM	White Muscle

Acknowledgements

First I wish to thank Professor Ian Johnston for the opportunity to carry out this work and for his guidance and support throughout the last 3 years. In the same light I thank my co-supervisor Dr James Kinghorn for advice and helpful discussions as well as our collaborators at the University of Tokyo, Professor Shugo Watabe and Professor Yuzuru Suzuki for allowing me to carry out some of my research in Japan. All members of the fish muscle research group and Gatty Marine Lab past and present deserve my gratitude for their never ending help, encouragement and keeping me sane. A special thank you to my good friend and colleague Dr Jorge Fernandes for his support, advice and never ending patience with me over the last 3 years without whom I would never have got this far. You deserve a medal Jorge! I would also like to thank Dr Keith Baar for his patience and helpful discussions during the latter stages of writing this thesis. I acknowledge NERC for providing the money to carry out this work. Thanks also to my puffer fish.

Outside the lab I'd like to thank all my friends and family for putting up with me and their support. Too many to name, you know who you are, however a select few who deserve mention are my parents (Graham and Norma MacKenzie), Gran, Peter MacKenzie, Paula Miller, and Lindsay Beattie.

Publications

Page 213: Fernandes J. M. O., MacKenzie M. G., Elgar G., Suzuki Y., Watabe S., Kinghorn J. R. and Johnston I. A. (2005). A genomic approach to reveal novel genes associated with myotube formation in the model teleost, *Takifugu rubripes*. *Physiological Genomics* **22**: 327-338.

Page 225: Fernandes J. M. O., MacKenzie M. G., Wright P. A., Steele S. L., Suzuki Y., Kinghorn J. R. and Johnston I. A. (2006). Myogenin in model pufferfish species: Comparative genomic analysis and thermal plasticity of expression during early development. *Comparative Biochemistry and Physiology, Part D* **1**: 35-45

Declarations

(i) I, Matthew G MacKenzie, hereby certify that this thesis, which is approximately 50,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date..13/10/2006.. Signature of candidate..

(ii) I was admitted as a research student in October 2002 and as a candidate for the degree of Doctor of Philosophy in October 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2002 and 2006.

Date..13/10/2006.. Signature of candidate...

(iii) I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date..13/10/2006.. Signature of supervisor..

(iv) In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date 13/10/2006 Signature of candidate

Thesis Abstract

In the present study the orthologues of a selection of mammalian muscle regulatory factors have been cloned in the puffer fishes *T. rubripes* and *T. nigroviridis*. These factors include the genes coding for the muscle regulatory factors (Myf5, MyoD, myogenin and Myf6), the SRY (Y-linked sex-determining gene) related high mobility box protein Sox8, two isoforms of the TGF-β signalling factor myostatin (FMstn-1 and FMstn-2) and the myostatin inhibitor follistatin.

In situ hybridisation localised *myogenin* transcripts to areas of muscle differentiation in *T. rubripes* embryos with expression in the somites progressing in a rostral/ caudal gradient and later in the muscle of the developing limb bud and face. *Sox8* and *Follistatin* transcripts were shown in the somites and limb buds but with expression also associated with developing neuronal tissue. We have also identified Sox8 as a proposed marker of myogenic progenitor cells (MPCs) in teleosts with transcripts localised to nuclei potentially corresponding to areas of MPCs in adult fast muscle tissue sections. In addition analysing relative expression of *myogenin* and *Sox8* throughout development using qPCR and *in situ* hybrisation revealed a reciprocal pattern of expression and, as previously described in mammalian myogenic cell lines, suggests *Sox8* to interfere with the transcription of *myogenin*.

Investigating relative expression using qPCR with respect to incubation temperature during development identified *myogenin*, *FMstn-1* and *follistatin* as potential plasticity genes. We showed a 3-fold higher expression of *myogenin* at an equivalent stage during early segmentation in embryos reared at 21°C when compared to 15°C. In contrast, at the same developmental stage relative expression of *follistatin* and *FMstn-1* was 2 and 2.5-fold greater

in embryos reared at the lower temperature of 15°C compared to 21°C and 18°C respectively. These changes in gene expression pose a potential transcriptional mechanism to explain the changes in muscle phenotype previously observed in the larvae of embryos developed at different incubation temperatures.

Using suppressive subtractive hybridisation two subtracted cDNA libraries were prepared from the fast muscle of a 180 g (myotube +) *T. rubripes*, still recruiting new muscle fibres and a 1.2 kg (myotube -) *T. rubripes*, no longer recruiting muscle fibres. The myotube (–) library contained 293 putative genes and 24 of these were chosen as possible candidates involved in the cessation of myotube production. Of these 24 11 were identified as consistently showing differential expression in the fast muscle of myotube (-) fish compared to myotube (+). Of the 11 genes 4 strong candidates had the appropriate tissue specificity showing between 5-25-fold increase in expression in myotube (-) fast muscle and little or no up-regulation in other tissues including slow skeletal muscle, heart, liver, skin and brain. We propose these four strong candidates, which are also present in other vertebrate genomes, to play a role in the inhibition of myotube production in vertebrates.

Chapter 1 Introduction

Puffer fish and genomics

1.1 Puffer fish

There are more than 100 species of puffer fish mostly found in the Atlantic, Pacific and Indian Oceans including the Japanese puffer fish (*Takifugu rubripes*). Other species including the green spotted puffer (*Tetraodon nigroviridis*) are found in brackish and fresh water. As a species they vary in size and appearance considerably with *Tetraodon nigroviridis* (*T. nigroviridis*) growing up to 15 cm and *Takifugu rubripes* (*T. rubripes*) up to 70 cm. *T. rubripes* and *T. nigroviridis* are examples of smooth puffer fish and are separated by ~18-30 million years (myr) of evolution having diverged from a common mammalian ancestor ~450 myr (Hedges 2002). The tetraodontiformes order of teleosts include the smooth puffer fish (Tetraodontidae) and their closest relative the spiny puffer fish (Diodontidae), separated by ~50-70 myr. Other species of this order are ocean sunfishes (Molidae), boxfishes and trunk fishes (Ostraciidae), triggerfishes (Balistidae), filefishes (Monacanthidae) and spike fishes and triplespines (Triacanthoidea) (Tyler 1980). Tetraodontiformes are unusual teleosts relying more on paired fin and median swimming over body-caudal fin swimming. They also posses elaborate mechanical defences, for example puffer fishes' ability to inflate if threatened, and fewer skeletal elements than most other fishes, typically with 22 or less vertebrae and the absence of pleural ribs. In addition smooth and spiny puffer fishes and molids are particularly unique with a lack of all caudal bones and the absence of pelvic fins, spines and girdles (Tyler 1980; Wainwright et al., 1995).

Analysis of genome size in a phylogenetically diverse sample of 275 teleost species identified the tetraodontiforms as having below average vertebrate genome size (<2 pg) and in particular the smooth puffer fishes have the smallest vertebrate genome size measured to date ranging from 0.7-1.0 pg (Hinegardner 1968; Brenner et al., 1993; Lamatsch et al., 2000). Members of the balistidae order of tetraodontiforms (including trigger fishes and file fishes) have the next largest genomes, 20-30% larger than those of smooth puffer fishes (Brainerd et al., 2001). The closest relative to smooth puffer fishes, spiny puffer fishes, have a genome 2-fold larger suggesting the compact genome arose within the last 50-70 myr of evolution from the spiny puffers (Neafsey and Palumbi 2003). Other vertebrates have much larger genomes with the *Homo sapien* (human) genome being 8-fold larger (Aparicio et al., 2002) but surprisingly smooth puffer fish and human genomes share a very similar gene repertoire (Brenner et al., 1993). This lead to the proposal that puffer fish genomes may provide an ideal model for comparative genomics and assisting the progression of the human genome project (Brenner et al., 1993). The first draft of the *T. rubripes* (Aparicio et al., 2002) and *T. nigroviridis* (Jaillon et al., 2004) genomes have now been published.

1.2 Genomes and sequencing

The term genome originated around 1930 and was defined as 'the totality of all the genes on all chromosomes in the nucleus of a cell'. This was defined even before DNA was identified. Together with the identification of the genetic code in 1961, the concept that biological organisms contained a blueprint with a finite number of genes was put forward. The human genome mapping project (HGMP) was initiated in 1990 and the draft genome published in 2001 by (Venter et al., 2001). The sequencing of an ever increasing number of other genomes was also initiated as references to assist the progression of the human genome sequencing, in

particular the *T. rubripes* chosen for its small size and lack of repetitive elements (Aparicio et al., 2002). This has been made possible by the huge advances in DNA sequencing over the last 30 years after the initial work from Sanger in 1977 who reported the order of nucleotides of DNA could be determined using chain terminating nucleotide analogues (Sanger et al., 1977). Later that same year the first human gene was isolated and sequenced (Seeburg et al., 1977). The 80s and early 90s showed the sequencing of small genomes starting with the 49 kbp bacteriophage lambda genome using a shot gun restriction digest method in 1982 (Sanger et al., 1982). The shotgun or random sequencing approach is a low cost and less time consuming method of genome sequencing and has been adopted for many sequencing projects including *T. rubripes* and human. Shotgun sequencing involves randomly selected clones containing 1-2 kb inserts being sequenced, assembled using computer algorithms to form a contiguous sequence (contig) and gaps filling in by direct sequencing methods, reviewed in (Sterky and Lundeberg 2000). *T. rubripes* was the second vertebrate to have its genome sequenced to draft level (Aparicio et al., 2002). A number of genome assemblies for a variety of organisms are currently available at http://www.ensembl.org/ including: human, *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Pan troglodytes* (chimpanzee), *Macaca mulatta* (rhesus monkey), *Canis familiaris* (dog), *Bos taurus* (cow), *Monodelphis domestica* (opossum), *Gallus gallus* (chick), *Xenopus tropicalis* (xenopus), *Danio rerio* (zebrafish), *Takifugu rubripes* (fugu), *Tetraodon nigroviridis* (green spotted puffer fish), *Ciona intestinalis* (sea squirt), *Drosophila melanogaster* (fruit fly), *Anopheles gambiae* (Mosquito), *Apis mellifera* (honey bee), *Caenorhabditis elegans* (worm) and *Saccharomyces cerevisiae* (yeast). New species are constantly being added with pre-assemblies available for *Loxodonta africana* (elephant), *Echinops telfairi* (hedgehog), *Dasypus novemcinctus* (armadillo), *Ciona savignyi* (sea squirt) and *Aedes aegypti* (the yellow fever mosquito).

1.3 Comparative genomics

The use of other genomes for research into a particular organism's genome is known as comparative genomics. Many current methods for gene identification are based on predictions using bioinformatics software and despite their increasing accuracy are sometimes unable to correctly identify exon/ intron boundaries and alternatively spliced transcripts (Rogic et al., 2001). Experimental data such as expressed sequence tags (ESTs) are a useful resource enabling expressed sequences to be identified (Kan et al., 2001) however these still lack the ability to identify full length genes. Comparisons between the human genomic DNA and the complete sequence of another less related vertebrate have provided a practical compliment to predictions and experimental data in revealing genomic regions of functional importance (Elgar et al., 1996; Elgar et al., 1999; Kan et al., 2001). The discovery that smooth puffer fish have compact genomes lead to the implication that despite the difference in genome size smooth puffer fish and human have approximately the same number of genes as a result of a lower abundance of non coding DNA (Hinegardner 1968). This hypothesis was confirmed using a random shot gun sequencing approach on the *T. rubripes* genome, revealing the genome to be approximately 400 Mb and with less than 10% repetitive DNA and both organisms have the same amount of coding sequence (Brenner et al., 1993). *T. rubripes* may therefore pose well as a reference genome for gene and functionally relevant sequence identification maintained throughout evolution.

On publication of the draft human genome, Venter and others (Venter et al., 2001) compared the differences in gene families between the human, fruit fly, worm and yeast genomes showing the human genome to contain expansions in proteins involved in acquired immune functions, neural development, structure and function, signalling pathways, hemostasis and

apoptosis. These expansions correlate with the phenotypic complexity observed in vertebrates when compared to the fly and worm. Indeed some striking differences were observed between fish and mammalian systems with the comparison of the *T. nigroviridis*, *T. rubripes*, mouse, human and worm proteomes (Jaillon et al., 2004). Fish have a more diverse collagen family than mammals and a greater number of domains associated with sodium transport possibly a reflection of their adaptation to saline environments. KRAB box transcriptional repressors which are abundant in mammals are completely absent in teleosts. The greatest differences between fish and mammals appear to be among the molecular function of proteins (Jaillon et al., 2004).

1.4 Factors regulating genome size

Diploid genome size varies dramatically among vertebrates and is measured in picograms (pg) of DNA per cell. The small genome of the smooth puffer fishes range from 0.7 to 1 pg (Lamatsch et al., 2000) this is in comparison to the African lungfish (*Protopterus aethiopicus*) genome of ~284 pg (Pedersen 1971). This huge variation in genome size is wide spread throughout vertebrates and is thought to be a result of a combination of genome level mechanisms such as duplication, polyploidy and transposable elements which increase overall genome size and spontaneous deletions and DNA repair mechanisms which tend to decrease genome size (Petrov 2001).

1.4 a Genome duplication and tetraploidisation in teleosts.

Genome duplication and the impacts on the evolution of ray-finned (teleost) fishes has been the topic of vigorous and unresolved debate for several years (Christoffels et al., 2004).

Teleosts are thought to have diverged from a common ancestral vertebrate about 450 Myr (Gu et al., 2002; McLysaght et al., 2002; Panopoulou et al., 2003; Christoffels et al., 2004) and (Holland et al., 1994) suggested that two whole genome duplications (WGD) have occurred in the ray-finned fish (Actinopterygians) lineage since its divergence from tetrapods. Evidence for this was in the identification of seven *Hox* gene clusters in teleost species (Zebrafish, *T. rubripes*, Medaka) in contrast to the four previously identified in mammals, giving rise to the fish specific WGD hypothesis (Amores et al., 1998; Naruse et al., 2000; Aparicio et al., 2002; Christoffels et al., 2004). This hypothesis has been heavily debated (Wolfe 2001) and the opposing view, accepting the presence of multiple gene copies, considers this to be the result of independent gene duplications (Robinson-Rechavi et al., 2001a; Robinson-Rechavi et al., 2001b). Christoffels et al (2004) supports the WGD theory and estimates its occurrence to ~350 Mya. One distinctive method of proving a WGD is to make a comparison with a related genome that did not undergo a WGD. A similar comparison was made to prove the existence of a WGD in *Saccharomyces cerevisiae* based on the comparison with a second yeast species (*Kluyveromyces waltii*) that had diverged before the proposed WGD in that lineage (Dietrich et al., 2004; Kellis et al., 2004). Jaillon et al., (2004) carried out such a comparison with human and *T. nigroviridis* with the recent publication of the draft *T. nigroviridis* genome creating a synteny map associating two regions in *T. nigroviridis* with one region in human. Analysis of the interleaving pattern confirmed a massive gene loss through many small deletions in a balanced fashion between the two *T. nigroviridis* sister chromosomes, a classic characteristic of WGD. This work closed the debate by demonstrating a fish specific WGD revealing the early vertebrate proto-karyotype composed of 12 chromosomes (Jaillon et al., 2004). Consequences resulting from such an event may introduce a variety of changes into the genome including different gene

expression or function (Holland et al., 1994; Sidow 1996) or entirely silenced genes (Lynch and Conery 2000).

1.4b Transposable elements

Transposable elements (TE's) have the ability to change genome size and organisation dramatically by adding to the DNA content and altering genes already present by creating pseudogenes. This method of genome alteration is known as the selfish genome hypothesis (Orgel and Crick 1980). Retrotransposons are a eukaryotic transposon which moves not as DNA but an RNA intermediate. Retrotransposons are thought to be a major influence on the moulding of the eukaryote genome and at least 40% of the human genome has been brought about by retrotransposition (Brosius 1999; Lander et al., 2001). They increase intronic and intergenic sequences and the release of transposable element copy number constraints is a major characteristic of large genomes (Kidwell 2002). Reverse Transcription Retrotransposons (RTR) are retrotransposable elements carrying a reverse transcriptase (pseudo)gene and can be split into 3 major groups: non-long-terminal-repeat (non-LTR) retrotransposons, LTR transposons and Penelope-like elements (Volff et al., 2001). Smooth puffer fish contain ~10% of the number of RTRs compared to mammals however higher diversity exists in puffer fish resulting in many more clades (Volff et al., 2003). This suggests that a reduction in retrotransposon activity may play a partial role in the smaller genome size of smooth puffer fish. Indeed it is not the number of transposons that define a large genome more the rate at which they act. There appears to be a direct correlation between transposistion rate and excision rates and in general TE rates are greater than excision rates (Petrov 2001). Teleosts and in particular smooth puffer fish, have undergone a massive loss of genes (Jaillon et al., 2004) following a complete genome duplication after diversion from

the tetrapods 450 million years ago (Hedges 2002). This mass loss of genes may therefore have been at a greater rate than TE activity and partially explain the smaller genomes amongst tetraodontiformes.

Muscle development in vertebrates

1.5 Muscle structure

Striated skeletal muscle generates movement by contraction. It is a multi-nucleated tissue made up of repeating units called sarcomeres enabling the interaction of two main structural proteins myosin and actin (Huxley 1957; Huxley 1969), which in turn hydrolyse adenosine triphosphate (ATP) and thus provide the energy for muscle contraction (Lymn and Taylor 1971). Myosins (thick filaments) constitute a large superfamily of muscle structural proteins typically constructed of 3 domains: (i) the motor domain for interaction with actin and ATP, (ii) the calmodulin binding/ light chain binding or neck domain, (iii) the tail domain which acts as an anchor when binding to actin (Sellers 2000). Actin (the thin filament) is the more passive partner in the sarcomere with its primary function to provide binding sites for the cross bridges. In skeletal muscle the actin filament contains a number of associated proteins including tropomyosin and troponin that regulate the muscle contraction. Activation of the actin filament is brought about when calcium binds to troponin (Brown and Cohen 2005). A large protein called nebulin is thought to regulate the assembly and final length of actin in skeletal muscle (Wang and Wright 1988). It achieves this in conjunction with Tropomodulin, a pointed-end protein-capping protein that regulates filament lengths in various cell types (Fischer and Fowler 2003). The interaction and efficiency of the actin and myosin filaments, to generate an optimal force, is assisted by a scaffold of other proteins. This scaffold or

sarcomeric cytoskeleton contains three basic structural elements. The Z-band anchors the actin filaments (Goldstein et al., 1977; Goldstein et al., 1988), the M-bands crosslink the thick filaments in the centre (Knappeis and Carlsen 1968) and elastic titan structures connect both transverse structures longitudinally (Wang et al., 1979) (Figure 1.1). A muscle contraction relies on the actin and myosin filaments sliding between one another rather than changing in length. This sliding filament model was proposed in the 1950s (Huxley and Niedergerke 1954; Huxley and Hanson 1954).

The Z-band and titin have been characterised in the most detail. The Z-band is the structure separating each sarcomere and anchoring the actin molecules and titan is a large protein spanning half the sarcomere. The width of the Z-band varies with respect to the various muscle types (cardiac, fast and slow). Fast striated muscle has characteristically thin Z-bands (~50 nm) while slow and cardiac muscle have wider Z-bands (Franzini-Armstrong 1973; Luther 1991). The difference in thickness is due to the density of a zig zag structure formed by α-actinin which is a component of the z-band connecting the actin molecules between the sarcomeres (Takahashi and Hattori 1989). Titin is believed to define the resting sarcomere length and sarcomere extensibility and to keep the myosin filaments in the centre of the sarcomere (Tskhovrebova and Trinick 2003). The M-band is located at the centre of the sarcomere and its location suggests it is needed for the regular packing of the thick filaments (Knappeis and Carlsen 1968). It has also been proposed to reduce the intrinsic instability of myosin filaments (Agarkova et al., 2003) and assist titin in maintaining the stability of the sarcomere (Agarkova and Perriard 2005). The repeated sarcomere patterns throughout striated skeletal and cardiac muscle give it its characteristic striations and the intrinsic interactions of the proteins described above make muscle a highly specialised tissue.

Figure 1.1 Sarcomeric structures

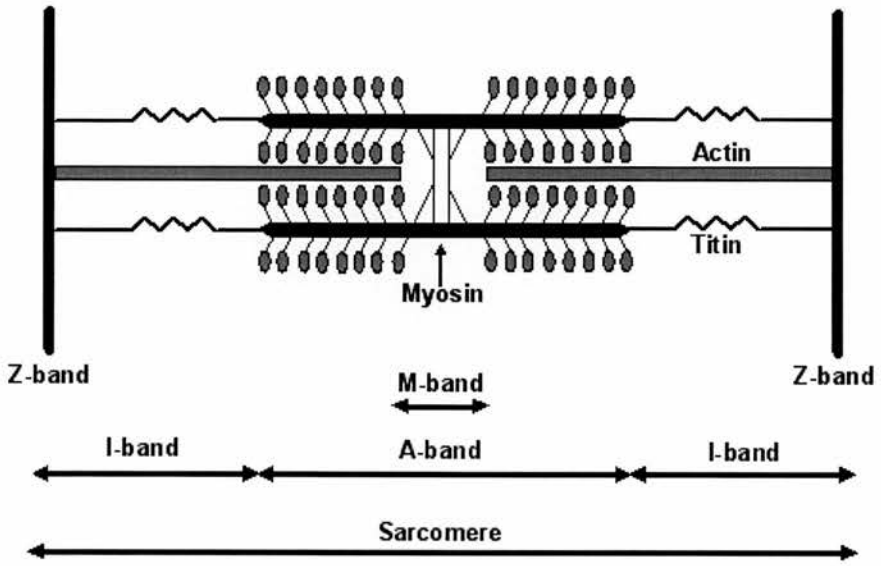


Figure 1.1 Simplified schematic diagram of a sarcomere. Important structures include the Z-band, actin, myosin and titin.

1.6 The somite, origin of muscle

Vertebrate skeletal muscle development begins with mesodermic structures known as somites that develop from the pre-somitic paraxial mesoderm into distinct compartments. The paraxial mesoderm develops from cells around the margin of the early gastrula. These cells converge towards the notochord at the dorsal aspect of the gastrula to develop the paraxial mesoderm adjacent to the axial mesoderm. The notochord regulates the movement of cells during early development through profound patterning influences on the surrounding tissue. Signalling molecules exerted from the notochord during somite formation (segmentation) include sonic hedgehog and a transmembrane receptor protein known as notch. This process has been studied in depth in zebrafish (Mullins 1999). The notch signalling pathway has been shown to have a vital role in vertebrate segmentation with severe somite defects associated with mutant mice such as *Notch1* (Conlon et al., 1995) and lunatic fringe (Zhang and Gridley 1998). A member of the basic helix-loop-helix family of transcription factors Mesp2 has been shown to interact with the Notch pathway in the presomitic mesoderm and to have a key role in vertebrate segmentation (Saga et al., 1997).

Somite formation in vertebrates is very similar between mammals, birds, fish and amphibians (Kimmel et al., 1995) although development of somitic zones varies particularly in fish. The ventral domain of the chick somite gives rise to cells that form the sclerotome which in turn forms the axial skeleton (Dockter and Ordahl 2000). The dorsal region of the avian somite (dermomyotome) gives rise to the muscle precursor cells (MPCs) which forms differentiated somitic muscle cells (myotome). These MPCs in mammals and birds are the pioneers for subsequent limb and trunk muscle development. Face and head muscles are derived from a separate unsegmented rostral mesoderm (Currie and Ingham 1998; Tajbakhsh and

Buckingham 2000; Gros et al., 2005). Proliferating MPCs in avian systems migrate from the dermomyotome (Gros et al., 2005) and must exit the cell cycle and undergo differentiation before fusing to form distinct groups of postmitotic myotubes, the structural component of skeletal muscle (Molkentin and Olson 1996; Yun and Wold 1996b). In contrast to the chick somite, muscle is the major derivative of the zebrafish somite (Kimmel et al., 1995). The sclerotome of zebrafish somites is identified shortly after somite formation as a cluster of cells on the ventromedial surface of the somite expressing genes such as *Pax9* and *twist* (Morin-Kensicki and Eisen 1997). Until recently the dermomyotome has been poorly characterised outside amniotes. A dermomyotome like layer has been described in *Xenopus laevis* (*Xenopus*) expressing *Pax3* (Grimaldi et al., 2004). Devoto and collaborators (Devoto et al., 2006) have identified a layer of cells on the somite surface, external to the developing myotome, showing no signs of terminal myogenic differentiation but expressing *Pax3* and *Pax7* which are known markers of myogenic precursors. In addition some of these cells express the myogenic differentiation factor *myogenin* suggesting the presence of an evolutionary conserved dermomyotome in teleosts (Devoto et al., 2006).

1.7 Organisation of muscle fibre types in adult fish

Skeletal muscle is organised into segmented structures known as myotomes which are categorised into two main groups of striated muscle fibre in vertebrates, described as slow (red) muscle and fast (white) muscle (Fig 1.2). These two muscle fibre types are arranged in anatomically discrete zones in teleosts making them an ideal model organism for analysing different muscle fibre types (Johnston 1977). Slow muscle fibres are parallel to the longitudinal axis of the fish and form either a thin superficial layer of muscle with a thickening at the horizontal septum or an internalised strip. Fast twitch muscle fibres have a

larger diameter and make up ~90% of the cross section of the myomeres (Stickland 1983). A third intermediate (pink) muscle fibre also exists in an intermediate position of some fish species (Bokdawala and George 1967; Johnston 1974; Mosse and Hudson 1977) (Fig 1.2D). The trunk muscles of the fish are arranged in a complex series of myomeres (Fig 1.2A) varying in length, angle of insertion and shape (Alexander 1969). The shape of the myomeres varies along the trunk and they are separated by collagenous sheets called myosepta (Fig 1.2B, C). The anterior cones are relatively larger than the posterior cones (Van Leeuwen 1999). The myotomal muscle acts on the spine producing a travelling wave of curvature along the length of the fish (Hess and Videler 1984; Cheng and Blickhan 1994). This produces a C-shape along the axis of the fish that is then followed by a contralateral contraction bending it the other way (Weihs 1973).

The fast and slow fibres are classified according to their morphology and function. Slow fibres are multi-terminally innervated (Bone 1964) and mitochondrial content varies from 25% (fractional fibre volume) in Coalfish (*Gadus virens*) for example (Patterson and Goldspink 1972) to 56% in *Chaenocephalus aceratus* (Johnston 1987). The characteristic red colour arises from high levels of myoglobin in the slow muscle. The aerobic metabolism and contractile properties suggest utilisation in sustained low intensity swimming patterns (Boddeke 1959; Bone 1975, 1978). Fast fibres contain less mitochondria compared to slow fibres for example 2% in plaice (*Pleuronectes platessa*) (Johnston 1981) and as little as 0.5% in the shark species *Etmopterus spinax* (Kryvi 1977). Up to 95 % of fast muscle fibre volume constitutes myofibrils (Johnston 1980). They also display a low concentration of myoglobin and glycogen utilising anaerobic respiration (Bone 1978) and are associated with burst swimming requiring high energy (Boddeke 1959).

Figure 1.2 Teleost Muscle

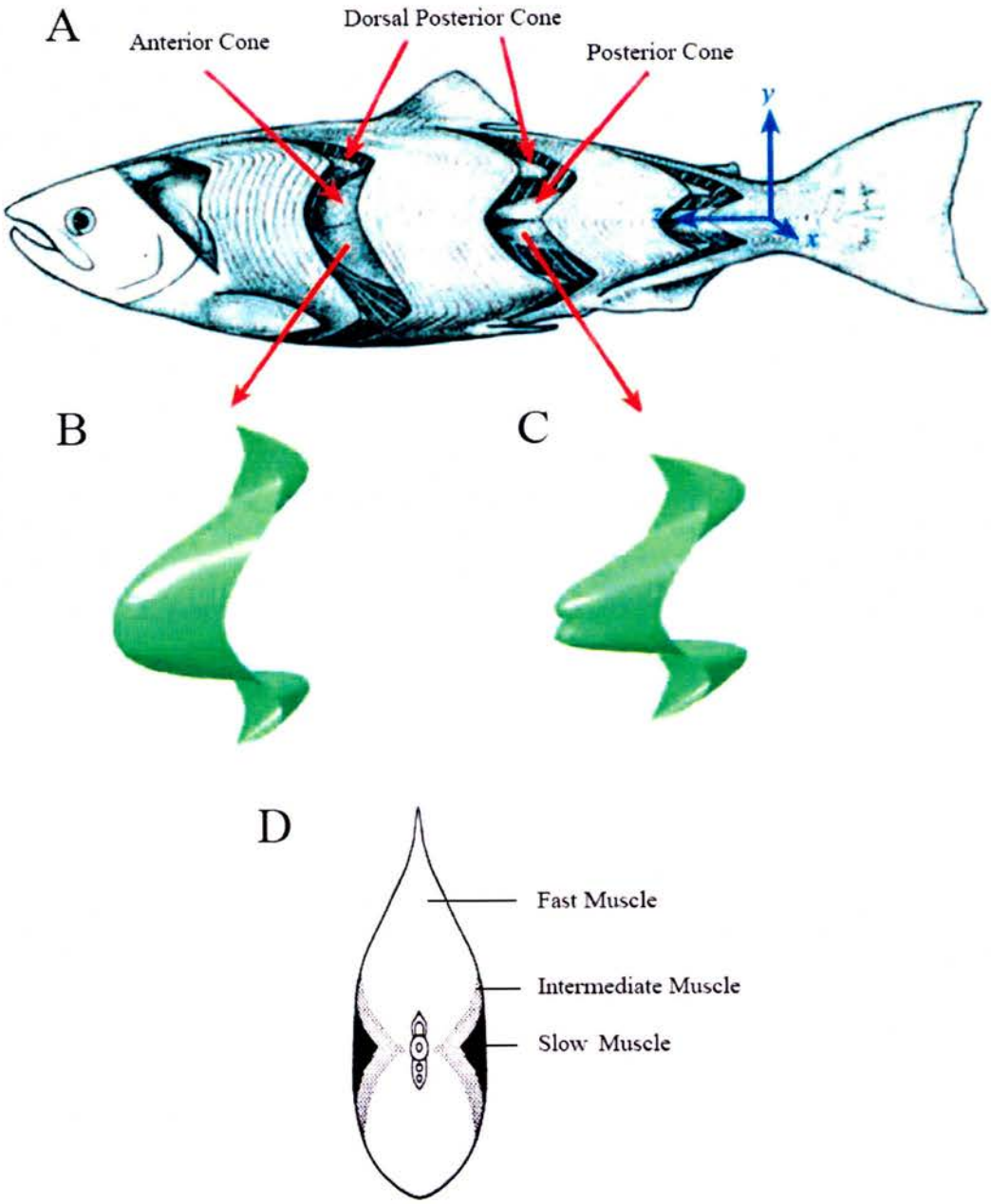


Figure 1.2 Teleost myotome. (A) King salmon myotome. (B, C) Simulated shape of anterior and posterior myosepta. (D) Representation of a cross section of teleost at 0.7 standard length (SL) showing the 3 muscle types. A, B and C are adapted from Leeuwen (1999).

1.8 Early vertebrate muscle specification and differentiation

Early fibre type differentiation varies dramatically between fish and other vertebrates. Using antibodies specific for slow and fast fibre type myosin heavy chains in rats revealed that no differentiation can be made between the fibre types prior to birth (Punkt et al., 2004). This is not the case in teleosts as the two fibre types are distinct very early on in muscle development and arise from separate populations of MPCs known as myoblasts, as demonstrated in zebrafish (Devoto et al., 1996). Fast muscle originates from a lateral population of cells in teleosts found in the segmental plate and is formed later in development than slow fibres (Devoto et al., 1996). The molecular networks directly regulating the differentiation of fast muscle fibres are poorly understood. Slow muscle differentiates prior to fast muscle during teleost embryogenesis (Kelly and Rubinstein 1980). The slow muscle progenitors are initially the most medial myoblasts close to the notochord as large cuboidal (adaxial) cells but shortly after somite development they undergo a lateral shift through the presumptive fast muscle domain to be the most superficial muscle layer (Devoto et al., 1996) (Fig 1.3). A small sub-population of these adaxial cells known as pioneer cells are found in the region of the future horizontal septum and after elongation form the first contractile muscle fibres (Van Raamsdonk et al., 1978) (Fig 1.3). The adaxial cells are the first cells to show expression of the muscle regulatory factor MyoD (Devoto et al., 1996) and the specification of the slow type muscle cell lineage is dependant upon hedgehog signalling. A glyocoprotein Sonic Hedgehog (Shh) is secreted from the notochord initiating the fusion of myoblasts giving rise to myotubes and was thought to be specific for slow muscle formation (Blagden et al., 1997; Du et al., 1997; Currie and Ingham 1998; Coutelle et al., 2001). Functional work using zebrafish systems ectopically expressing *sonic hedgehog* in early embryos leads to the formation of extra slow muscle cells at the expense of fast muscle cells (Du et al., 1997),

likewise mutant zebrafish with defective Sonic Hedgehog signalling lack *MyoD* expression in the adaxial cells and lack superficial slow muscle fibres (Lewis et al., 1999). This work implies the importance of Sonic Hedgehog in the specification of the migrating adaxial cells. More recently however Hedgehog signalling has been revealed to impact not only the slow fibre differentiation but also a small sub-population of fast twitch fibres expressing Engrailed protein (Wolff et al., 2003). In addition, Henry and Amacher (Henry and Amacher 2004) also revealed Hedgehog signalling is required for normal fast fibre elongation. They also showed that it is not a direct influence rather migrating wild-type slow muscle cells that induce fast muscle cell elongation. Expression of muscle specific genes, including myosins and tropomyosins, during the expansion of the myotome of hatching trout larvae reveals a distinct separation of fibre type specific genes to the appropriate areas of fast or slow muscle. In contrast at later stages an externally added group of slow fibres expressed not only slow twitch muscle isoforms but also a subset of fast twitch muscle isoforms adding further complexity to the gene networks regulating fibre type identity (Chauvigne et al., 2006).

Figure 1. 3 Adaxial cell migration

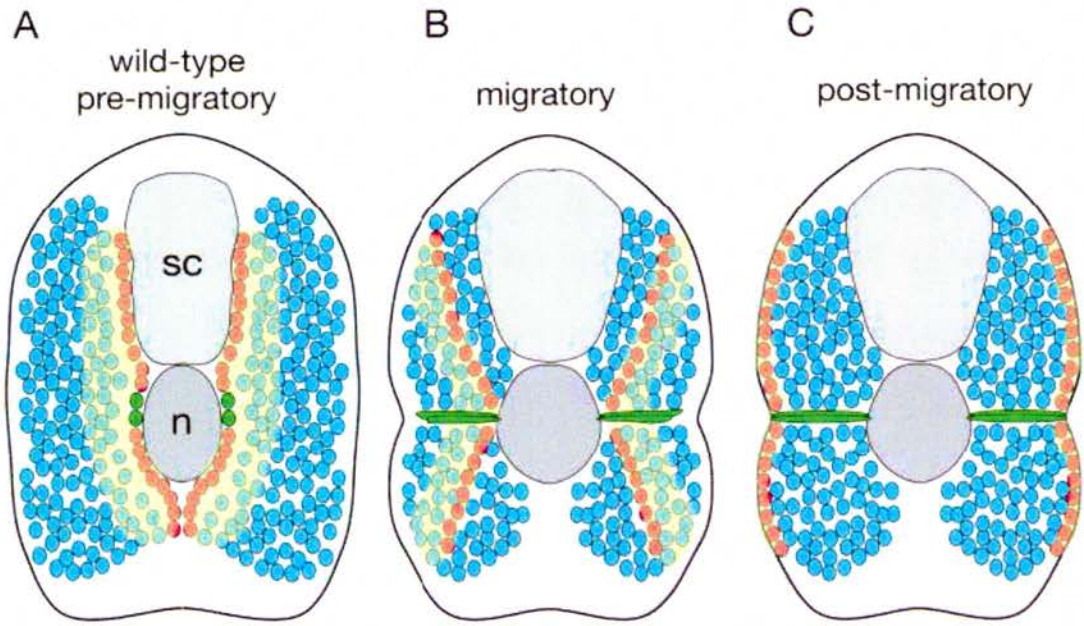


Figure 1.3 Schematic representation of adaxial cell migration in teleost embryo cross-section. The fast muscle progenitor cells are blue and the migrating adaxial cells are orange. Pioneer cells are highlighted green. SC (spinal chord), N (notochord).

1.9 Skeletal muscle growth

Muscle tissue is post mitotic and highly differentiated, however in teleosts it must be able to respond to the requirements of growth as the fish matures (Weatherley and Gill 1985). Both skeletal fibre types (fast and slow) therefore grow through fibre recruitment (hyperplasia) and fibre enlargement (hypertrophy) (Koumans and Akster 1995). Muscle development and growth in vertebrates is managed by the opposing activity of a group of beta helix-loop-helix transcription factors known as the muscle regulatory factors (MRFs) and other factors including Id (inhibitors of differentiation) and myostatin (Yun and Wold 1996b). These factors will be discussed in more detail (section 1.12). The community effect is a phenomenon in all developing tissues whereby cells must contact a sufficient number of like neighbours if they are to undergo coordinate differentiation within a developing tissue (Gurdon et al., 1993a). This was first demonstrated in *Xenopus* where large groups of 100+ cells were able to interact thus enabling muscle specific gene expression and terminal differentiation but smaller groups would not (Gurdon 1988; Gurdon et al., 1993b). There are two forms of muscle growth in vertebrates: (i) Hyperplasia whereby developing muscle grows with the addition of new myotubes and, (ii) hypertrophy where muscle growth occurs via the expansion of fibres already present.

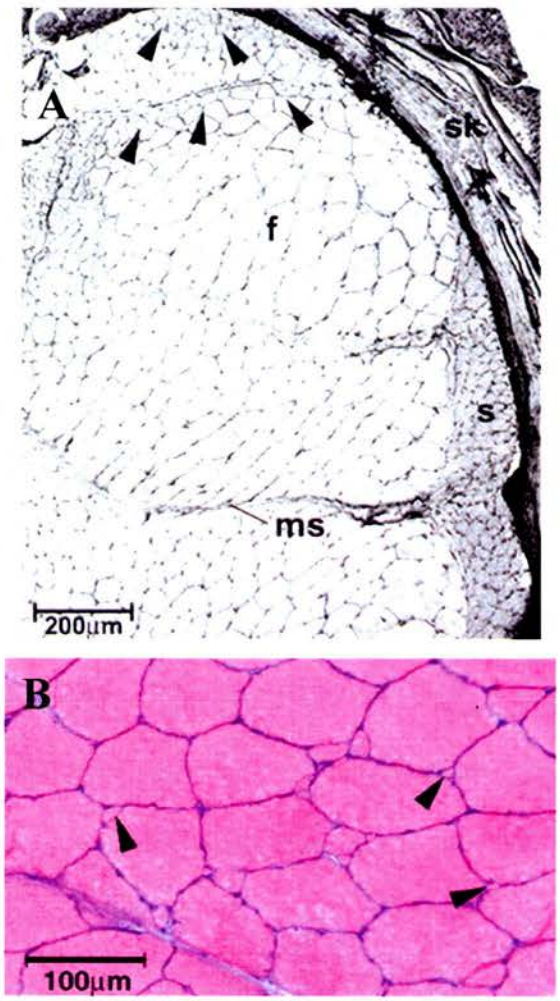
Myotube formation (hyperplasic growth) in mammals occurs only during foetal development and subsequent growth after birth occurs solely through the elongation and hypertrophy of muscle fibres already present (Rowe and Goldspink 1969). Myofibers generally develop in two waves in small mammalian species such as mouse. The existence of a third wave of myotube formation has been suggested in large mammals such as pigs (Mascarello et al., 1992; Lefaucheur et al., 1995) which may enhance growth rate of the developing muscle

tissue, however it generally arises after the total fiber number is considered fixed (Lefaucheur et al., 1995). The first wave of myotube formation arises from embryonic myoblasts and forms primary muscle fibres. These fibres are present early on in foetal development and a definitive number is reached rapidly (McKoy et al., 1998). Secondary fibres are formed from the fetal myoblasts and primary fibres act as a scaffold on which secondary fibres can form. The cross-sectional area of primary myofibers is characteristically larger than that of the secondary fibres (Christensen et al., 2000) and therefore many more secondary myotubes than primary ones are formed in all mammals (Zhang et al., 1998). The ratio between final numbers of primary and secondary fibres varies between mammals depending on the maximum size of the animal (Handel and Stickland 1987a; Zhang et al., 1998).

Muscle growth in teleosts has some unique properties particularly in post-embryonic growth. One such property unique to teleosts is their ability to recruit new fibres (hyperplasia) throughout growth after hatch until ~40% of their maximum body length (Weatherley et al., 1988; Johnston et al., 2000). This is to allow for the massive growth of fish from larvae to adult (Weatherley et al., 1979). Atlantic salmon for example has an estimated 5000 muscle fibres (per trunk cross-sectional area) at hatching increasing to 150,000 during smoltification with excess of 1,000,000 at 4 kg (Johnston 1999, 2001). Hyperplasic growth of teleosts occurs in two sequential phases; stratified and mosaic (Rowlerson and Veggetti 2001) (Fig 1.4). The stratified hyperplastic growth phase takes place in proliferation/ growth zones during the post hatch larval stages completing the embryonic fast and slow muscle layers (Van Raamsdonk et al., 1974; Rowlerson et al., 1995; Johnston and McLay 1997). The resulting new fibres that are added to the germinal zones give rise to a characteristic gradient of muscle fibre diameter from the superficial to the deep layer of the myotome (Rowlerson and Veggetti 2001). This stage of muscle growth gives rise to both fast and slow muscle

lineages however disrupted *sonic hedgehog* expression in zebrafish mutants show no phenotypic change in the stratified hyperpastic growth of slow muscle (Barresi et al., 2001). This opens up the possibility of a different mechanism in specifying new slow fibres during this stage of growth compared to the specification of migrating adaxial cells in the somites. Mosaic hyperplasia is observed during the final phase of myogenesis and is so called because the myotubes form on the surface of fast muscle fibres throughout the myotome showing a mosaic pattern of fibre diameters (Weatherley and Gill 1987). The small diameter fibres formed during hyperplastic growth express a unique myosin heavy chain which can be a useful marker for muscle undergoing muscle fibre recruitment (Ennion et al., 1995). This phase of hyperplastic growth occurs across the entire myotome and in some cases overlaps with stratified hyperplasia (Stoiber and Sanger 1996). Fast muscle fibre recruitment in fish stops at 25-50% of the maximum body length, depending on the species (Weatherley et al., 1988), this is in contrast to slow muscle in which fibre number continues to increase with body length (Van Raamsdonk et al., 1983; Johnston et al., 2004). Despite the physiological and phenotypic knowledge with regard to hyperplastic muscle growth little is known about the transcriptional networks that regulate it.

Figure 1.4 Two stages of hyperplastic muscle growth



Stratified Hyperplasia

Mosaic hyperplasia

Figure 1.4 The two stages of hyperplastic muscle growth in teleosts. (A) Stratified hyperplasia, the arrow heads indicate areas of stratified hyperplasia containing fibres of smaller diameter than the surrounding tissue. (B) Classic active mosaic hyperplasia, the arrowheads indicate areas of mosaic hyperplasia. Tissue sections are fast muscle from *Notothenia coriiceps* and *Eleginops maclovinus* respectively. Slow muscle (s); fast muscle (f); skin (sk); myosepta (ms). Scale bars indicate 200 µm (A), 100 µm (B). Figure adapted from (Johnston et al., 2003a)

1.10 Developmental plasticity

Teleosts are ectotherms and as such are susceptible to environmental influences during development, with embryonic temperature (Johnston et al., 2003b) and photoperiod (Johnston et al., 2003c) profoundly affecting the rate and outcome of organogenesis. There are numerous examples in plants and animals whereby a given genotype can result in multiple phenotypes in response to environmental factors during embryogenesis (Stearns 1991). Developmental processes can be uncoupled in response to temperature changes prior to the completion of development resulting in various intermediate phenotypes (Johnston and Hall 2004). Embryonic temperature can influence the number and diameter of fast and slow muscle fibres in the myotomes of larvae and juveniles in a range of teleost species (Johnston and Hall 2004). For example subjecting Atlantic herring to different temperatures with the range 5°C to 12°C greatly affects the rostral caudal development of axial muscles and the outgrowth of primary motor neurones. The onset of myogenic stages, including the synthesis of contractile filaments and myofibril assembly, become greatly accelerated at higher temperatures with respect to somite stage (Johnston et al., 1995). Similar heterochronic shifts were observed in larval stages with the growth rate and development of the median fins retarded with respect to body length at 5°C in comparison to embryos reared at 8°C. In addition muscle fibre recruitment at 37 mm standard length (SL) was increased at higher temperatures with an almost 2-fold increase in muscle fibres recruited per day at 8°C compared to 5°C (Johnston and Cole 1998). Indeed developmental plasticity is also apparent in mammals, including humans and rats, with studies suggesting environmental factors experienced by the mother can have a profound effect on the phenotypic expression of her offspring and termed "maternal environmental effects" (Barker 1992, 1994; Barker and Clark 1997; Mousseau 1998; Wolf et al., 1998; Champagne et al., 2003). The phenotypic changes

as a result of developmental plasticity are apparent but little is known regarding the genetic networks that regulate these phenotypes and how they are affected. Changes in gene expression have also been investigated with respect to embryonic temperature. Up to the 45-somite stage in rainbow trout *MyoD* and *myogenin* transcripts were expressed in more somites of embryos reared at 12°C compared to 4°C (Xie et al., 2001). Conversely during the eyed stage of development *myogenin* and *MyoD* mRNA and protein levels were stronger in the caudal somites at 4°C compared to 12°C. Northern analysis revealed myosin heavy chain expression was higher at 12°C compared to 4°C consistant with a more advanced stage of myogenic differentiation (Xie et al., 2001). In contrast temperature was shown to have no significant effect on *myogenin* expression with respect to somite stage in Atlantic herring (Temple et al., 2001). It been proposed that there is an optimal temperature for both myogenic regulatory factors and myosin heavy chain expression in teleosts (Wilkes et al., 2001).

Myogenic determination and differentiation

1.11 The cell cycle and exit

The cell cycle is the essential process whereby cells duplicate their genetic information (mitosis) and proliferate, creating new cells from those already present, unicellular organisms produce a completely new organism whereas multicellular organisms develop and repair. Initially the proliferating cells are undifferentiated and transcription factors enable them to differentiate into the various cell types and exit the cell cycle to terminal differentiation (Alberts et al., 1998).

The cell cycle is split into 4 distinct phases (Fig 1.5). The two active phases are the S phase (DNA synthesis) and the M phase (mitosis and cytokinesis) and are separated by the Gap phases 1 and 2 where the majority of cell growth occurs (Johnson and Rao 1971). Gap 1 (G1) is the most expandable period of the cycle and gives the cell the opportunity to prepare for DNA synthesis or importantly exit the cell cycle for terminal tissue differentiation. Gap2 (G2) allows the cell to prepare for mitosis by growth where the cell will segregate its sister chromosomes and undergo cytokinesis (cell division) (Nasmyth 1996). G0 is a quiescent state the cells can enter on leaving the cell cycle. The G0 cells are able to re-enter the cell cycle when needed in response to increased growth factor concentration (Iyer et al., 1999), for example myosatellite cells in muscle regeneration (Grounds and Yablonka-Reuveni 1993; Charge and Rudnicki 2004). The collection of G1, S and G2 make up what is known as interphase. Originally this was thought to be an intermediate phase of mitosis where little happened except cell growth until the discovery of chromosome duplication (Howard and Pelc 1951).

Figure 1.5 The cell cycle

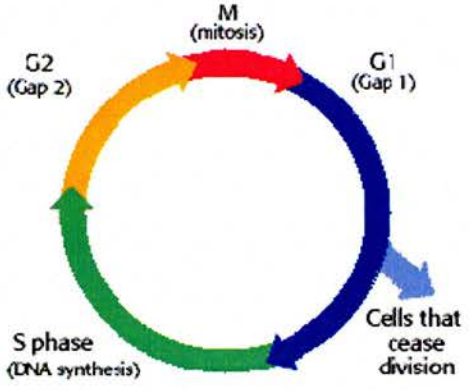


Figure 1.5 The 4 phases of the cell cycle.

Cell cycle research identifying the factors that regulate the progression into these phases identified the concept of checkpoints. Checkpoints were originally defined as cell cycle surveillance mechanisms but were re-defined as any control that ensures the correct order of cell cycle events. P53 for example is the classic blocker of G1 – S phase by arresting cells in the G1 state and ultimately leading to the programmed cell death (apoptosis) of the dividing cell (Nasmyth 1996; Vousden 2000; Taylor and Stark 2001). Entry into M-phase is only permitted if the dividing cell has reached an appropriate size and its DNA has replicated completely (Tapon et al., 2001).

The master regulators of the cell cycle are a group of kinases known as the cdks (cyclin dependent kinases) together with their regulatory cyclins. Their joint action allows progression through the cell cycle. Cdks were initially discovered during cell cycle research on the yeast strains, *Saccharomyces cerevisia* (budding yeast) mostly but also *Schizosaccaromyces pombe* (fission yeast) (Schwob et al., 1994; Dirik et al., 1995). In yeast there is only one Cdk called Cdk1 (cdc28) and works in conjunction with one of nine cyclins depending on the stage of the cell cycle (Tapon et al., 2001). With the use of temperature sensitive cell division cycle (Ts-cdc) mutants it was demonstrated that the cell cycle would not function without cdks. Moving a ts-mutant for cdc28 to a non-permissive temperature blocks the cell in G1 (Kitazono and Kron 2002). In mammals there are various Cdks each being a representative for a particular stage of the cell cycle and has its specific regulatory cyclin for activity. For example mammalian Cdk1 works along with cyclin B as a mitosis promoting factor, MPF (Ohi and Gould 1999). In general the Cdk is present throughout the cell cycle and is activated at the appropriate time by the regulated translation of its respective cyclin (see figure 1.6). Its activity ceases on queue by the proteolytic breakdown of the cyclin (Ekholm and Reed 2000). The Cdk/ cyclin complexes are inhibited by a group of proteins

known as the CKIs. They are made up of two families, the INKs and the cip/kips (Tanner et al., 2000). Both act very differently, the INKs (e.g. $p16^{INK4a}$) bind to the cdk preventing the cyclin from interacting and the cip/kips (e.g. $p21^{Cip1}$ and $p27^{kip1}$) bind to the Cdk/cyclin complex blocking substrate recognition and ATP binding . p21 for example is markedly upregulated upon myogenic differentiation and cell cycle exit (Guo et al., 1995; Halevy et al., 1995; Hawke et al., 2003). The process of myogenic cell cycle exit and differentiation is regulated by many factors including a group of muscle regulatory factors (MRFs) and their interactions with the Cdks and inhibitors (twist, Id) (Yun and Wold 1996a; Walsh and Perlman 1997).

Figure 1.6 MPF activity

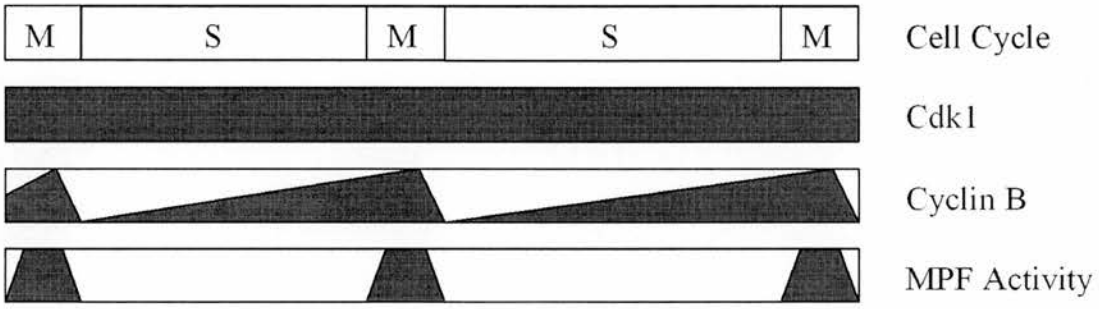


Figure 1.6 MPF activity during the cell cycle with respect to cyclin B degredation. M (mitosis) and S (DNA synthesis) phases of the cell cycle are indicated.

1.12 The MRFs

Myogenic regulatory factors (MRFs) guide myogenic identity from determination to differentiation in a perfected synchronised manner. A summary of their interactions during myogenesis is shown in figure 1.7. The MRFs are a group of four transcription factors (Myf5, MyoD, myogenin and Myf6) collectively known as the MyoD family and a subset of a super family based on their conserved regulatory motifs, such as the basic helix-loop-helix domain.

This family includes proteins such as the proto-oncogene c-myc (Stone et al., 1987) and the protein Twist involved in mesoderm formation (Thisse et al., 1988). The basic region is the DNA binding motif and binds to the E box consensus sequence CANNTG present in the promoter regions of many muscle-specific genes such as myosin heavy chain, troponin I and muscle creatine kinase. This is in contrast to the helix-loop-helix motif which forms heterodimers with ubiquitously expressed E proteins encoded by the E2A gene transcripts (Lin et al., 1991; Edmondson and Olson 1993). Two other less well characterised domains, histidine/ cysteine-rich (H/C) and helix III, have been implicated in the activation of muscle specific genes and an involvement in the chromatin remodelling of muscle specific gene enhancers enabling transcription of previously silent loci by MyoD (Gerber et al., 1997). The activities of the MRFs are controlled by both positive (CBP/p300, MEF2 proteins) and negative (Id, LIM protein, Twist, I-mf proteins) regulatory factors including interactions with each other (Buckingham 2001; Brand-Saberi 2005). Remarkably, when ectopically expressed in a variety of cell types from different germ layer origins, each of the MRFs has the ability to convert non-muscle cells into muscle (Lin et al., 1989; Weintraub et al., 1989; Choi et al., 1990)

Myf5 is the earliest known muscle specific factor expressed as shown in vivo with the myogenic cell line C2. Its expression is restricted to cycling cells and the translated protein is regulated by proteolytic breakdown prior to mitosis (Lindon 1998). A motif within the bHLH domain resembles a 'destruction box' that is characteristic to substrates of mitotic proteolysis, this domain is also recognised by a mitosis initiating complex called APC (Anaphase Promoting Complex). Myf5 is thought to be the only MRF regulated in this manner (Lindon et al., 2000a). MyoD interacts with Cdks specifically in the G1/S transition of the cell cycle encouraging proliferating cells into the myogenic lineage (Skapek et al., 1995; Zhang 1999;

Zhang et al., 1999; Wei and Paterson 2001). The Myogenic conversion of 10T1/2 cells by MyoD has been shown to be dependent on the level of MyoD transcripts. Over expression of *MyoD* using the Rous Sarcoma Virus (RSV) promoter results in the differentiation of single myoblasts whereas expression under the control of the weaker SV40 promoter results in a myogenic cell line capable of proliferating while still expressing low levels of *MyoD* (Rudnicki et al., 1993). MyoD$^-$/Myf5$^-$ knockout mice lack myofibres indicating that one or other of these two factors is essential for myoblast establishment (Rudnicki et al., 1993). Myf5 induction of *myogenin* transcription has been shown to be cell density dependent, which suggests a regulatory mechanism ensuring a sufficient number of myoblasts differentiate to interact and form a multi-nucleated myotube (Lindon et al., 2001). *Myogenin* is a direct transcriptional target of both Myf5 and MyoD (Hollenberg et al., 1993; Buchberger et al., 1994) and is irreversibly activated upon growth factor removal and cell differentiation (Olson and Klein 1994). It is a marker of terminal differentiation whereby myoblasts are commited to a myogenic fate (Hasty et al., 1993; Nabeshima et al., 1993). Myf6 is the last of the four MRFs to be expressed and has an involvement in the maturation of the myotubes, terminal differentiation of the muscle fibres (Edmondson and Olson 1993). More recently Myf6 has been shown to have a further role the initial determination of the myogenic lineage along with Myf5 and MyoD (Kassar-Duchossoy et al., 2004).

Expression of the MRFs during teleost development reveals a strong association with the progression of somite development (Weinberg et al., 1996; Chen et al., 2001; Coutelle et al., 2001; Temple et al., 2001; Cole et al., 2004). *Myf5* and *MyoD* are initially expressed in the mesoderm prior to segmentation in common carp followed by *myogenin* as the somites develop (Cole et al., 2004). More specifically *Myf5* and *MyoD* transcripts are detected in the medial adaxial cells with *Myf5* expression migrating laterally into the pre-somitic mesoderm

before fading as the adaxial cells become incorporated into the developing somites. MyoD expression progresses laterally, encompassing the entire somite closely followed by myogenin expression with myogenin expression persisting beyond that of MyoD (Cole et al., 2004). This was consistent with previous work in zebrafish (Weinberg et al., 1996; Chen et al., 2001; Coutelle et al., 2001) however some striking differences have been observed in other teleosts including the rainbow trout which shows *Myf5* expression only in the adaxial cells and medial region of the somite (Delalande and Rescan 1999) and herring where MyoD expression persists beyond that of myogenin (Temple et al., 2001).

Orthologues of the mammalian MyoD family have been characterised in a variety of teleosts; *myf5* (Coutelle *et al.*, 2001 [zebrafish], Tan *et al.*, 2002 [striped bass]), *myoD* (Rescan *et al.*, 1994 [rainbow trout], Kobiyama *et al*, 1998 [carp], Tan &Du, 2002 [gilthead sea bream], Gotensparre *et al.* 2004 [atlantic salmon]), *myogenin* (Rescan *et al.*, 1995 [rainbow trout], Chen *et al.*, 2000 [zebrafish], Tan *et al.*, 2002 [striped bass]) and *myf6* (Hinits *et al.*, 2003 [zebrafish], Coutelle *et al.*, 2001 [Takifugu]. MyoD has been shown to have multiple paralogues in salmonid species with differential mRNA expression at different stages of development and in different muscle fibre types (Rescan and Gauvry 1996; Delalande and Rescan 1999).

Figure 1.7 Summary of myogenesis

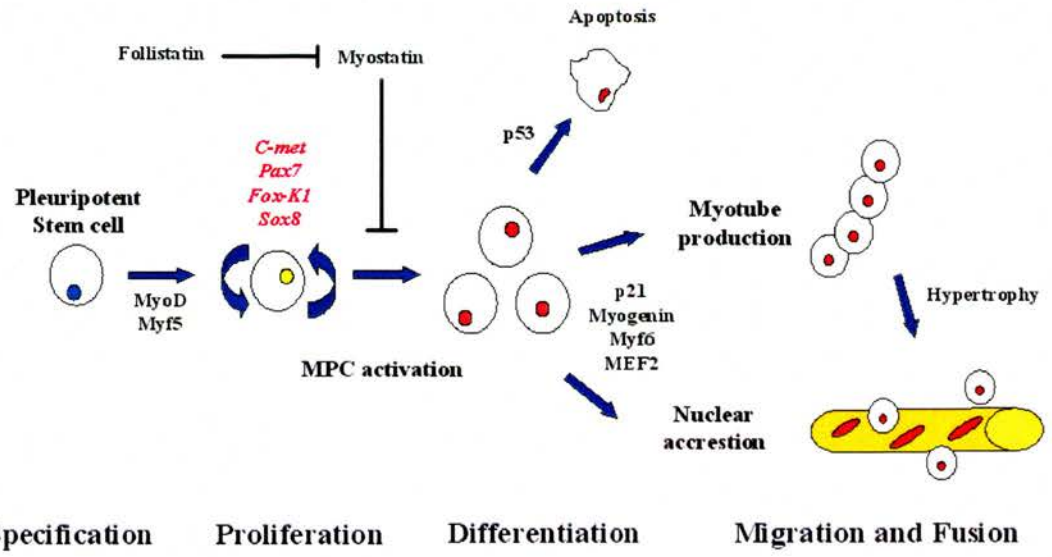


Figure 1.7 A model describing the main events of myogenesis in teleost skeletal muscle. Some myogenic factors including the MRFs are represented at their appropriate stage of specification, proliferation, differentaion migration and fusion. Myostatin inhibits muscle growth and markers of MPCs are highlighted red. Many of these factors and others will be discussed in the following sections. On exit from the cell cycle myogenic cells have 3 potential fates: (i) apoptosis (cell death). (ii) cell differentiation, fusion and production of new myotubes. (iii) cell differentiation and fusion with growing muscle fibres.

1.13 Regulating the regulators

1.13a p300

A member of a family of co-activators (p300) has been shown to interact with MyoD and together they have an involvement in *myogenin* transcription and cell cycle arrest. The association occurs through the carboxy terminal cystine/ histidine-rich domain of p300 (Yuan et al., 1996). The p300/ MyoD complexes are recruited specifically at the MyoD DNA binding sites of differentiating myoblasts initiating the transcription of the down stream muscle specific factors myogenin and muscle creatine kinase (Puri et al., 1997). It has therefore been proposed by Puri and collaborators (1997) that p300 is required for MyoD specific cell cycle arrest in both differentiating myoblasts and MyoD converted fibroblasts.

1.13b Sox proteins

The Sox family of transcription factors have wide spread functions throughout embryogenesis including certain aspects of muscle growth and development (Wegner 1999). Both Sox8 (Sock et al., 2001; Schmidt et al., 2003) and Sox15 (Beranger et al., 2000; Lee et al., 2004) have been shown to have a major role in muscle growth and interactions with the MRFs. Beranger and collaborators (2000) used murine C2C12 myogenic cell lines and revealed Sox15 transcripts were at the highest level in myoblasts prior to their differentiation into myotubes and specifically inhibited the activation of muscle specific genes including myogenin and MyoD, resulting in a failure of these cells to terminally differentiate. Cell lines with C-terminal truncated Sox15 did not show this behaviour. More recently Lee et al (2004) suggested Sox15 to be involved in the specification of myogenic cell lineages or interestingly

in the regulation of the fusion of myoblasts to form myotubes during muscle development and regeneration. Analysis of $Sox15^{-/-}$ cell lines revealed a downregulation of *MyoD* transcripts and an upregulation of *Myf5* transcripts suggesting Sox15 role in cell cycle exit by regulating early myogenic regulatory factors. Sox8 is one of the more recently characterised Sox proteins but has already been shown to have a major role in both myogenesis (Schmidt et al., 2003) and osteoblast differentiation (Schmidt et al., 2005). In 2003 the first suggestions of the role of Sox8 in muscle growth became apparent with *Sox8* transcripts shown to be confined to the myosatellite cells in adult skeletal muscle with a downregulation during myogenic differentiation (Schmidt et al., 2003). The study also revealed that overexpression of *Sox8* disrupted myoblasts in their ability to form myotubes alongside a decrease in transcript number of both *MyoD* and *myogenin*. A similar observation was shown with the over expression of *Sox9*, a close relative of Sox8. This led to the proposal that Sox8 acts as a negative regulator of muscle differentiation by interfering with the function or transcription of myogenic basic helix-loop-helix proteins.

1.13c Id proteins

An inhibitor of DNA binding (Id) is a member of the helix-loop-helix family however differs from the MRFs in that it does not possess the basic DNA binding motif. They inhibit the action of bHLH transcription factors including the MyoD family by forming heterodimers with them, preventing them from binding to the E-box motif, and consequentially inhibiting the differentiation pathway (Kadesch 1993; Littlewood 1995). Included in this apparent inhibitory role, they also possess a positive regulating property with regard to the cell cycle. The negative effect on cell differentiation holds the cells in a proliferative state with a specific role in the G0 to S phase (Norton et al., 1998). Four Id paralogues have been found

in higher vertebrates: Id1 (Benezra et al., 1990), Id2 (Sun et al., 1991), Id3 (Christy et al., 1991) and Id4 (Riechmann et al., 1994). Expression patterns of all four have been well characterised in murine systems separating them into two subclasses with Id1, 2 and 3 showing overlapping expression in multiple tissues and Id4 transcripts restricted to neuronal tissues and the ventral portion of the epithelium of the developing stomach. Their expression in most tissues correlates with cells actively undergoing morphogenesis (Jen et al., 1996). Orthologues of mammalian Id 1 and 2 have been have been identified and characterised from the slow muscle of rainbow trout (Rescan 1997). A further family member Id6 first identified in zebrafish (Sawai and Campos-Ortega 1997) has 2 non allelic orthologues (Id6a, Id6b) in rainbow trout (Ralliere et al., 2004). All four Id family members have been shown to have distinct expression patterns in the developing somites of rainbow trout demonstrating their role in myogenesis (Ralliere et al., 2004).

1.13d LIM protein and MEF2

The LIM protein (CRP3) is a muscle specific protein which interacts with the heterodimer MyoD/E47 through two zinc finger structures (Kong et al., 1997). Myocyte enhancer factor-2 (MEF2) is part of the MADS (MCM1, agamous, deficiens, serum response factor) box protein family of transcription factors. MEF2 contains two motifs (MADS box and MEF2) characteristic of its DNA-binding properties and ability to form homo- and heterodimers (Molkentin et al., 1995; Black and Olson 1998). Three CRP proteins (1-3) have been identified in higher vertebrates (Arber et al., 1994) and CRP2 has been characterised in rainbow trout (Delalande and Rescan 1998). Both are putative positive regulators of myogenesis. There are four MEF2 genes (A-D) identified to date in higher vertebrates (Pollock and Treisman 1991; Yu et al., 1992; Breitbart et al., 1993; Martin et al., 1994) and

three of the four (MEF2a, MEF2c and MEF2d) have been identified in teleosts (Ticho et al., 1996; Kobiyama et al., 1998). They associate with certain co-factors which interact with and activate (MyoD, GATA4, NFAT, p300) or inhibit (MITR, cabin) MEF2 function (McKinsey 2002). The MEF2 binding site is A/T rich and present alongside E-boxes in the regulatory regions of many muscle specific genes (Black and Olson 1998).

1.13e Myostatin and Follistatin

Myostatin is a member of the transforming growth factor- β (TGF- β) family and is a negative regulator of muscle growth (McPherron et al., 1997; McPherron and Lee 1997). Mutant *myostatin* null (*myostatin* -/-) mice show a massive increase in both hyperplasia and hypertrophy (McPherron et al., 1997), this is also evident naturally in certain breeds of cattle (Belgian blue and Piedmontese) with misense mutations in Myostatin showing a large increase in muscle mass relative to conventional cattle (McPherron and Lee 1997). Myostatin is thought to function by controlling the proliferation of MPCs. It prevents the progression of cells from the G1 phase of the cell cycle into DNA synthesis (S phase) by upregulating the cdk inhibitor p21 and decreasing the activity of cdk2. This leads to an accumulation of retinoblastoma (Rb) protein in its active hypophophorylated form and subsequent arrest of myoblasts in the G1 phase of the cell cycle (Thomas et al., 2000b). Analysis of the promoter region of bovine *myostatin* revealed several E-boxes and a MEF2 site suggesting regulation by the MRFs and MEF2. Indeed one E-box was experimentally shown to bind to MyoD, indicating myostatin as a downstream target of MyoD (Spiller 2002).

Myostatin has been identified in a numerous teleosts to date including zebrafish (Xu et al., 2003), catfish (Gregory et al., 2004), rainbow trout (Rescan et al., 2001) and Atlantic salmon

(Ostbye et al., 2001). Two paralogues of myostatin have been identified in salmonids (Ostbye et al., 2001; Rescan et al., 2001) and zebrafish (Kerr et al., 2005). Expression of myostatin in mammals is primarily restricted to skeletal muscle (McPherron et al., 1997) however expression of *myostatin* othologues in teleosts appears to be ubiquitous (Ostbye et al., 2001; Roberts and Goetz 2001; Vianello et al., 2003). Despite this, the salmonid paralogue *Myostatin 2* appears to be more closely related to muscle growth and development (Rescan et al., 2001; Kerr et al., 2005). Kerr and collaborators (2005) have conducted a phylogenetic analysis of the entire *myostatin* gene family in telesost and proposed all salmonid *myostatin* paralogues described to date are in fact orthologues of mammalian *myostatin* and a second *myostatin* gene not related to mammalian mysostatin is also present in most, if not all, teleosts.

Follistatin is a secreted glycoprotein with many functional roles throughout mammalian embryogenesis and growth (Patel 1998) including antagonising members of the TGF- β family such as Myostatin (Fainsod et al., 1997; Amthor et al., 2002; Zimmers et al., 2002; Amthor et al., 2004). Over expression of *follistatin* induces a similar phenotypic response to myostatin knock out mouse whereas a *follistatin* -/- mouse displays muscle deficiency (Matzuk et al., 1995; Lee and McPherron 2001). Amthor and collaborators (2004) were the first group to demonstrate a direct interaction of Myostatin and Follistatin. Despite its obvious functional importance in embryogenesis and growth very little work on follistatin has been carried out in teleosts, particularly with respect to muscle growth.

Project Aims

1.14

Muscle fibre number and size is influenced by environmental conditions during development in teleosts. Little is know about the transcriptional networks regulating myotube production in teleosts and how these and other muscle specific networks are affected as a result of developmental plasticity. Research into how gene expression is altered as a result of developmental plasticity opens up the potential for the discovery of previously undescribed gene function and transcriptional regulation. The recent publication of the compact puffer fish genomes (*T. rubripes* and *T. nigroviridis*) to draft level offers a vast resource and relatively straight forward method of identifying orthologues of known mammalian muscle regulatory factors.

The main focus of this project was primarily to identify and characterise orthologues of known mammalian muscle regulatory factors representing various stages of myogenesis in puffer fish (*T. rubripes* and *T. nigroviridis*) with the use of the draft genomes and investigate their expression throughout development with respect to embryonic temperature using *in situ* hybridisation and quantitative real-time PCR. A second major aim of this work was to identify novel and indeed known transcripts whose putative proteins may have a role in myotube formation. We aimed to achieve this with the use of subtracted cDNA libraries representing differentially expressed genes from fast muscle still recruiting new fibres and fast muscle that had ceased myotube production.

Chapter 2: Materials and Methods

2.1 Sample Collection

2.1 *i* Animals and sample collection

Adult and juvenile Tiger Puffer Fish (*T. rubripes*) were bred in captivity and maintained at 18°C in the Fisheries Marine Laboratory (Maisaka, Shizuoka prefecture, University of Tokyo, Japan). Adult *T. rubripes* (3-4kg) were wild caught and obtained from Maisaka City fish market, Japan. *T. rubripes* eggs from a single female (3.8 kg) were fertilised at Nisshin Marinetech Co. (Yokohama, Japan) by 2 males (~6 kg) at 18°C and transferred to the Fisheries Laboratory within 4 h. They were separated into 3 groups and incubated at temperatures of (15°C, 18°C and 21°C ± 0.5°C). After hatch the larvae were all transferred to 18°C. Adult *Tetraodon nigroviridis* (*T. nigroviridis*) were obtained from a commercial supplier (Ultimate Discount Aquatics, Cupar, Fife) and maintained at the Gatty Marine Laboratory (St Andrews, Fife) at 26°C in brackish water (20 ppm NaCl) under a light regime of 12 hours light and 12 hours dark. Adult and juvenile fish were killed humanely by over-anaesthesia in a solution of 0.2 mM 3-aminobenzoic acid ethyl ester (Sigma, Dorset, UK) buffered with sodium bicarbonate (Sigma) and tissue samples were stored in RNA later (Ambion, Cambridgeshire, UK) or snap frozen in liquid nitrogen for future nucleic acid extraction. A 0.5 cm steak was cut at 0.7 standard length (SL) anterior to the dorsal and anal fins and a series of blocks were cut to sample various regions of the myotome. Blocks were frozen in Cryomatrix embedding medium (ThermoShandon, Pittsburgh, USA) using isopentane cooled to its freezing point with liquid nitrogen (-159°C). *T. rubripes* eggs for *in situ* hybridisation were punctured and fixed in 4% PFA/PBS (Sigma) before being

dehydrated in increasing concentrations of methanol to (25%, 50%, 75% and 100%) and stored at -20°C.

2.2 Nucleic acids

2.2 *i* Nucleic acid extraction

Total RNA was extracted from 100 mg of tissue using the Tri Reagent protocol from Sigma (Sigma 1999). The FastPrep Instrument (Qbiogene, Cambridge, UK) was used in conjunction with FastRNA Pro Green beads (Qbiogene) to homogenise the tissue in 1ml of Tri Reagent (Sigma, Dorset, UK) for 40 seconds at a speed setting of 6.0. RNA was checked for quality and denaturation on a 1.2 % Agarose (BDH, Dorset, UK) containing 10 µg/ml EtBr gel in Tris (Sigma)/ Acetate (BDH, Dorset, UK)/ EDTA (Sigma) buffer. Genomic DNA was extracted from 25 mg of each sample using the DNeasy Tissue kit (Qiagen, West Sussex, UK), following the suggested protocol.

2.2 *ii* Nucleic acid quantification

Plasmid DNA (2.3 *ii*), PCR products (2.3 *i*) and RNA for use in RT-PCR were quantified using a UV-1601 spectrophotometer (Shimadzu-Deutschland GmbH, Duisburg, Germany). Readings were taken at 260 nm and 280 nm. The OD reading at 260 nm was used to calculate the concentration of nucleic acids in a sample and the 280 nm reading permitted the ratio of absorbance (260/280) indicating the purity of the sample. Ratios were ~1.8 and values <1.7 indicated a high phenol or protein contamination. Concentration (c) was calculated from the

OD value using the folowing rule: where c = ε x OD x dilution. ε is a constant (ε = 40 for RNA and ε = 50 for DNA). Nucleic acids were generally diluted 1:200.

Total RNA for use in qPCR (2.3 *ii*) was quantified with the fluorescent nucleic acid stain RiboGreen (Molecular Probes/ Invitrogen, Paisley, UK) according to the manufacturer's protocol. DNA standards (Molecular probes) were quantified with the RNA on 96 well plates (Nunc) to form a straight line. The straight line equation (y = mx + c) was then used to calculate the nucleic acid concentration. This provided a more accurate measure of the concentration which was essential for qPCR.

2.2 *iii* Agarose gels

Nucleic acid preparations were checked for integrity using gel electrophoresis. 6x gel loading buffer (30% glycerol, 0.25% Xylene cyanol, 0.25% bromophenol blue) was applied to each sample in a 1:6 ratio and samples were fractionated through an agarose gel. Agarose gels were prepared by dissolving 1.2% agarose in 1xTris (Sigma)/ Acetate (BDH, Dorset, UK)/ EDTA (Sigma) buffer (1xTAE) with 10 μg/ml EtBr. Agarose was dissolved by heating the agarose and 1xTAE and EtBr was added once the solution had cooled to ~50°C. The set gel was submerged in 1xTAE and a 100V current was applied for ~45 minutes. A 1 kb marker (Promega) was run with samples to determine the size of bands. The gels were visualised under ultraviolet light and images captured with a Versadoc 3000 (Bio-Rad Laboratories, Hemel Hempstead, UK)

2.2 *iv* Reverse Transcription

1 μg of total RNA (0.75 μg for use with real time PCR, see 2.3 *ii*) was used to synthesise 1st strand cDNA with a retroscript kit (Ambion) using the manufacturers recommendations. RNA was denatured with oligo (DT) primers at 85°C for 3 minutes and the reverse transcription reaction was performed at 50°C for 1 hour. The enzyme was denatured for 10 minutes at 92°C. cDNA was then stored at -20°C for subsequent use.

2.3 Sequence analysis

2.3 *i* Genomic database mining

The genoscope *T. nigroviridis* and *T. rubripes* genomic sequence assembly (http://www.ensembl.org/Tetraodon_nigroviridis/ and http://www.ensembl.org/Fugu_rubripes/ respectively) were screened using the default parameters for orthologues of mammalian muscle regulatory factors (including *Myf5*, *MyoD*, *myogenin*, *Myf6*, *Pax7*, *P300*, *Mstn-1*, *Mstn-2*, *follistatin*, *Sox8*, *MLC2*) using the mammalian protein sequences. A TBLASTN was performed using a BLOSUM80 matrix, a word size of 4 and a maximum expected value cut-off equal to $1 \cdot e^{-5}$.

2.3 *ii* Primer Design

Gene specific primers were designed generally to amplify the entire coding sequence of predicted transcripts or alignments with mammalian sequences. In all cases primers were chosen by "eye" using the following guidelines: (i) primers were between 18- 24 bases in length, (ii) a Guanine (G)/ Cytosine (C) content of between 50- 65%, (iii) The last base at the

3' end on a every primer was a G or a C to facilitate initiation of polymerisation. Primers were checked for secondary structures such as dimers and hairpins using Netprimer (Premier Biosoft International, CA) and obtained from MWG biotech (Ebersberg bei München, Germany). In many cases the same primers designed for *T. rubripes* sequences were also used for amplifying *T. nigroviridis* genes and vice versa.

Table 2. 1 Primers for gene amplification using RT-PCR

Gene name	*Fwd primer (5'→ 3')*	*Rev Primer (5'→ 3')*
Myf5	GCCCCTGTTCTTTGGTTCTG	CATCAGGAGTCATGTGACCC
MyoD	ATGGAGCTGTCGGAGATCT	CAGGACCTGATAAATCAGGTT
myogenin	TGGAGCTTTTTGAGACCAACC	TGGGAATGTCCATAGAAAAGG
Myf6	GAGAGGGGCGCAACAATATG	CTCTTCTCTGGGAAGCTCAG
Pax7	ATGACTGCGTTGGCGGGGT	CGTAACAAAGTTTGCCATCTG
P300	ACTCGTACATGTCGCCCTCA	ATGGCCGATAATGTCCTGGA
FMstn- 1	ATGCTGGTCTTAGCTGTGCTG	GGATTACTTTTCCTCGCTGG
FMstn- 2	ATGCAAATGTCTCCGAGC	GACTTAAAGACATCCACAACG
follistatin	TTCAACACTGCCACCGTAAG	GAGGAGGGCGTTTTCTTTCAT
Sox8	CGTAAAAGACGCACCGAGC	CGAACGAGTCTCTGGGAAAGC
Sox8 probe for in situ	ACCACCACGCTGAACACGC	TAAGGCCGGGACAGGGTG
MLC2	GTCAGTCACATCGGCTTTGG	GGTGAGGATTGTGAAAGAGG

2.4 Gene amplification and bioinformatic analysis

2.4 *i* Polymerase Chain Reaction (PCR)

T. rubripes and *T. nigroviridis* transcripts were amplified using sequence specific sense and antisense primers designed from the genomic sequence (see table 2. 1). PCRs were performed using a 20 µl reaction mix comprising 1.5 µl DNA template, 50 nmol of each primer (Fwd and Rev), 0.1 µM dNTPs, 1x PCR buffer (Amersham, Buckinghamshire, UK) and 1 U Taq DNA polymerase (Amersham). A Techne Genius (Techne, Glasgow, UK) PCR machine was used with the following parameters; initial denaturation at 95°C for 3 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 56°C for 30 seconds and

extension at 72°C for 90 seconds. The extension was increased to 2 minutes for amplifying genomic DNA and a final extension at 72°C was applied for 10 minutes. 6x gel loading buffer was added (1x final concentration) and the PCR products were visualised on a 1.2 % Agarose (BDH, Dorset, UK) containing 10 μg/ml EtBr gel in Tris (Sigma)/ Acetate (BDH, Dorset, UK)/ EDTA (Sigma) buffer and excised from the gel using the Qiagen gel extraction kit (Qiagen).

2.4 *ii* Gene Cloning

After gel extraction, 10-20 ng of PCR product (maximum 4 μl) were ligated to a pCR4-TOPO T/A vector (Invitrogen, Paisley, UK), which was then used to chemically transform TOP10 *Escherichia coli* cells (Invitrogen). Colonies were grown on LB plates containing Ampicilin (50 mg/ml) generally 5 colonies were cultured in LB broth containing ampicillin (50 mg/ml). Miniprep plasmid purifications were performed using the QIAprep Miniprep kit (Qiagen) following the manufacturers recommended protocol.

Restriction enzyme digestion was used as an insert check using restriction endonucleases supplied by New England Biolabs (Ipswich, MA, USA). Digest reaction conditions followed those suggested by the manufacturer with regard to enzyme specific buffer and optimum reaction temperature. The vector contained *EcoR 1* splice sites flanking the insert and *EcoR* 1 was used most frequently for digestion at 37°C for 2 hours in a 20μl reaction with 1μl 10x buffer H and 1μl cDNA. Occasionally EcoR1 splice sites were present in the insert, in such cases digestion was repeated using *Not 1* (splice site 3' to insert) and *Spe 1* (splice site 5' to insert). Digest reactions were stopped with gel loading buffer (1x final concentration) and run on a 1.2% agarose gel (2.2 ii*i*).

2.4 *iii* Sequencing and analysis

Plasmid DNA was sequenced at the University of Dundee Sequencing Service (University of Dundee, Dundee, UK) with universal T3 and T7 primers. The DNA was sequenced in both directions using an ABI PRISM 377 DNA Sequencer (Applied Biosystems, Warrington, UK).

Genomic organisation was determined by comparing the genomic sequence with the cDNA using Spidey . Synteny data between *T. rubripes*, *T. nigroviridis* and human orthologues were obtained with MartView (http://www.ensembl.org/Fugu_rubripes/martview). cDNA was conceptually translated using DNAman (Lynnon Biosoft, Quebec, Canada) and putative peptides were aligned with orthologues obtained from Swiss-Prot Protein knowledgebase (http://us.expasy.org/sprot/) using ClustalW .

2.5 Gene expression analysis

2.5 *i In situ* hybridisation

a. Whole mount

Primers were designed from the cDNA sequence to amplify either the entire coding region or a >750 bp region of the *T. rubripes* sequence outwith the conserved domain for use in probe synthesis for *in situ* hybridisation (Table 2. 1). The region of the *T. rubripes* cDNA for probe synthesis was amplified from its corresponding pcr4-TOPO vector using standard fwd and rev M13 primers and the PCR parameters described above. The product was used as a template for DIG RNA labelling by *in vitro* transcription using T7 and T3 polymerases

(Roche, East Sussex, UK) for sense and antisense probe synthesis according to the manufacturer's recommended instructions. *In situ* hybridisation was performed essentially as described by (Hall et al., 2003). At room temperature the optimised incubation times for permeabilisation of *T. rubripes* embryos with 20 μg/ml proteinase K (Roche) were 5, 10 and 15 min for pre-somitogenesis, segmentation and post-somitogenesis stages, respectively. Images were photographed using a Leica MZ7.5 (Milton Keynes, UK) binocular microscope with a Nikon (Surrey, UK) coolpix 4500 digital camera.

b. Tissue sections

Fast muscle blocks (~1 cm^3) were embedded in cryomatrix (ThermoShandon, Pittsburgh, USA) under RNase free conditions and frozen in Isopentane cooled to its melting point (-160°C) in liquid nitrogen. 15 μm frozen sections were cut at -20°C in a cryostat (Leica CM 1850, Leica Microsystems, Milton Keynes, UK) and air dried. RNA was fixed on the sections by heating the sections at 50°C for 5 minutes. Sections were stored at -80°C and rehydrated in 2 x PBS (Sigma) before use. Localisation of transcripts was performed using a method modified from Weltzein and collaborators (Weltzien et al., 2003). Sections under went a pre-fixation step by one 5 minute wash in TrisHCl (0.05M, pH 7.5), followed by 2 minutes in 4% paraformaldehyde/ PBS (pH7.5) at room temperature (RT) and a further 5 minute wash in PBS. Tissue was permeabilised in 5 μg Proteinase K/ ml TrisHCl (0.05M, pH7.5) for 5 minutes at room temperature after equilibration for 5 minutes in TrisHCl (0.05M, pH7.5). A post-fixation step was then administered as described for pre-fixation. Sections were then subjected to acetic anhydride (Sigma) treatment; 10 minutes in TEA (0.1M, pH 8) (Sigma), 10 minutes in 0.25% AcAh in TEA at RT and rinsed for 5 minutes, 3 times in PBS. For pre-hybridisation 250 μl of hybridisation buffer (50% formamide, 5xSSC, 5xDenhardts, 250 μg/ml tRNA, 1 mg/ml salmon sperm DNA, 10% Dextran Sulfate) was added to each slide

and incubated in a moisture chamber (wipes soaked in 5xSSC) at room temperature for 2 hours. Hybridisation was performed over night at 65°C in a sealed moisture chamber (wipes soaked in 5 x SSC) with 500 µl of hybridisation buffer/ 300-800 ng/ml probe. Post-hybridisation washes were carried out: 30 minutes 5xSSC at RT, 15 minutes 30% formamide/ 5xSSC at 65°C, 2 washes in 0.2xSSC for 15 minutes at 65°C, slides were then allowed to cool to room temperature followed by a further 5 minutes wash in 0.2xSSC. Sections were RNase treated (~10 µg/ ml RNase A at 37°C) at this point as a control to ensure any signal was true hybridisation. Sections were then washed in antibody buffer (0.1M Tris-HCl, 0.15 M NaCl, pH 7.5) for 5 minutes at RT followed by 1 hour in antibody buffer/ 1% heat inactivated goat serum (HIGS). Anti-DIG antibody (Roche) 1:2000 was made up in antibody buffer/ 1% HIGS and slides were incubated over night at 4°C. Samples were washed 3 x 5 minutes at RT in antibody buffer before a 5 minute incubation in colour buffer (0.1M Tris-HCl, 0.1M NaCl, 0.005M $MgCl_2$, pH 9.5). Sections were stained in the dark using colour buffer with 3.4 µl/ml NBT, 3.5 µl BCIP, 1mM Levamisole. The reaction was stopped in TEN (10mM Tris-HCl, 1mM EDTA, 0.9% NaCl pH 8.0). Stained sections were photographed using a Leica MZ7.5 (Milton Keynes, UK) binocular microscope with a Nikon (Surrey, UK) coolpix 4500 digital camera.

2.5 *ii* Quantitative real time PCR (qPCR)

0.75 µg of total RNA was used to synthesise 1st strand cDNA as described in 2.2 *iv*. Oligo DTs were used in conjunction with random decamers in a 1:1 ratio and a negative control containing no template was included ensuring no contamination. cDNA was then diluted 5 times before use due to the sensitivity of the assay. qPCR was performed using an ABI prism 7000 (Applied Biosystems, Foster City, CA, USA) real time Thermocycler along with a

SYBR green kit (QuantiTect SYBR Green PCR, Qiagen) and 96 well plates were obtained from (Applied Biosystems). The qPCR mix contained 12.5 µl of SYBR green master mix, 1 µl of template cDNA, 0.5 µM of each primer and made up to a 25 µl reaction with nuclease free water. Specific primers were designed as described in 2.3 *ii* (see table 2. 2) avoiding the conserved domain and crossing at least one exon/exon boundary to avoid amplifying genomic content. PCR conditions were as follows: initial denaturation at 95°C for 15 minutes followed by 40 cycles of denaturation at 94°C for 15 seconds, annealing at 56°C for 30 seconds and extension at 72°C for 30 seconds. A dissociation protocol with a gradient from 60°C to 90°C was performed to investigate the specificity of the primers and the presence of primer dimers. As an internal reference for normalisation of SYBR Green fluorescence the fluorescent dye ROX was used. 3 internal standards were used (18s, G3PDH and RNA polymerase II), 18S primers (Ambion) were chosen preferentially when using cDNA from adult tissues and RNA polymerase II primers (see table 2.2) when investigating expression in developmental stages. A no template control was included on the plate for both gene specific primers and the respective standard. PCR efficiencies of each amplicon were estimated by linear regression analysis of the logarithm of SYBR Green fluorescence versus cycle number, using the LinRegPCR software (Ramakers et al., 2003). The ratio of expression of the gene of interest (GOI) between each sample and the data point with the highest expression level (calibrator) normalised against the standard is computed according to the following equation, based on a mathematical model (Pfaffl 2001):

$$Ratio = \frac{E_{GOI}^{\Delta C_{T,GOI}(calibrator-sample)}}{E_{std}^{\Delta C_{T,std}(calibrator-sample)}}$$

Where E_{GOI} and E_{std} represent real-time PCR amplification efficiencies of the GOI and standard (18S rRNA or RNA polymerase II), respectively, and ΔC_T is the difference between threshold cycle values between sample and calibrator.

Table 2. 2 Primers crossing exon boundaries for qPCR

Gene name	Fwd primer (5'→ 3')	Rev Primer (5'→ 3')
myogenin	ACTGCACTTCCGGACCAGC	GCTCAGAAGATCCTCGCTGC
Sox8	ACTGTGGCGTTTGCTCTCTCGG	GCCCGGCTCAGCTTTATACATG
FMstn-1	AGCACAGCAAGCAGATGAGGC	TGGGCGATGGGATTAGGC
FMstn-2	CGGAGACTAACTGGGGCATCG	GCGGGTAGCGGCAGCAAC
Follistatin	CAAAGAAACCTGCGACAATG	CGGGGCACAAGACATCG
18s rRNA (Ambion)	Unknown	Unknown
RNA Polymerase II	CAGCCCAGATGAACTTAAACGG	CCAGGACACTCTGTCATGTTGC
Elongation factor-1α	TGGAGGCGGCATCGGAACTG	CTCGGTCAACGCCTCGTGG

2.5 *iii* Statistical analysis of qPCR data

Relative expression from qPCR analysis was plotted against hours post fertilisation (hpf) using Sigma Plot (version 8.02). All data were represented as means (n=4 for developmental stage, n=7 for fast/slow skeletal muscle) ± standard error. For all other tissues n=2 and data are represented as a mean, values did not vary by > 25%. For statistical analysis of relative expression during development a general linear model was carried out with hpf as the fixed factor. Further post hoc comparisons were carried out with the Tukey HSD and Benferroni tests. Significant differences between two individual points were tested using the Mann-Whitney U non-parametric test. A value of $p=<0.005$ was considered significant.

2.6 Subtracted Library Construction and Analysis.

2.6 *i* cDNA synthesis

First strand cDNA from the fast muscle of one large adult male *Takifugu* (myotube (-) 3.4 kg, 50 cm SL) and one juvenile specimen (myotube (+) 180 g, 18cm SL) were synthesised from RNA samples (2.1 *ii*) using the SMART PCR cDNA synthesis kit (Clonetech, CA, USA). The optimum number of cycles of amplification (ensuring that the double stranded (ds) cDNA remained in the exponential phase of amplification) was 17 from both samples. This was taken to be 1 cycle fewer than the number of cycles to reach the plateau of amplification. The two ds cDNA samples were then digested with *Rsa 1* (BD biosciences, Oxford, UK) which generates shorter, blunt-ended ds cDNA fragments which are essential for adapter ligation and optimal for subsequent subtraction.

2.6 *ii* Construction of subtracted cDNA libraries.

Two independent subtractions were performed using the cDNA Subtraction Kit (Clontech): a forward (fwd) subtraction (where the cDNA from the myotube (-) *T. rubripes* was used as the tester and the cDNA from the myotube (+) *T. rubripes* was used as the driver) and reverse subtraction (where the cDNA from the myotube (+) *T. rubripes* was used as the tester and the cDNA from the myotube (-) *T. rubripes* was the driver). Myotube (-) fish had stopped recruiting new fibres and myotube (+) fish were still recruiting new muscle fibres. Adapter ligation was performed on two tester cDNAs for each library prior to subtraction, adapter 1 to the first tester and adapter 2R to the second tester. Adapter ligation was also carried out on a control tester skeletal muscle sample. The samples then underwent two hybridisations. For

the first hybridisation an excess of driver cDNA was added to each tester cDNA along with 1x hybridisation buffer and the samples were heat denatured at 98°C for 90 seconds and then incubated at 68°C for o/n (not exceeding 12 hours). The remaining single stranded (ss) cDNAs (used for the second hybridisation) coding for differentially expressed genes were significantly enriched at this stage as non target sequences present in the tester cDNA and driver hybridise. For the second hybridisation both samples from each subtraction were added together without a second denaturing along with fresh denatured driver cDNA to further enrich differentially expressed transcripts. Hybridisation was performed at 68°C o/n before addition of 200 µl of dilution buffer followed by a further incubation at 68°C for 7 minutes. The result of the second hybridisation was newly formed hybrid molecules consisting of differentially expressed ds cDNAs with different adapters at the 5' ends.

The subtracted libraries then underwent two PCR amplifications to amplify the differentially expressed sequences following a brief incubation at 75°C to fill in the missing strands of adapters creating a binding site for PCR primers. The first PCR only amplified ds cDNAs with different adapters on each end and the second PCR was a nested PCR further reducing background. The nested PCR uses 2 sets of primers: The first amplifies products similar to conventional PCR (2.3 *i*), and the second (nested) set binds within the first PCR product amplifying a shorter product, this reduces the probability of amplifying a non specific locus.

2.6 *iii* EST sequencing

The subtracted PCR mixes were ligated into pCR4-TOPO vector as described in 2.4 *ii*. Approximately 1150 random colonies were picked from each library and used as the template in a PCR (2.3 *i*), using short T3 (5'-ATTAACCCTCACTAAAG-3') and short T7 (5'-

AATACGACTCACTATAG-3') primers. Colony PCRs were carried out on 96 well plates (Applied Biosystems) and the products run on a 1.2% agarose gel to check for inserts which ranged from 250 to 3000 bp. Amplified PCR products were sequenced at the MRC Rosalind Franklin Institute for Genomics Research, Cambridge using a T3 primer (5'-AATTAACCCTCACTAAAGGG-3') and Big Dye terminator sequencing mix (Applied Biosystems). The sequencing reaction cycle had the following parameters: 95°C for 20 s and 60°C for 140 s, and was repeated 25 times. An Applied Biosystems 3700 capillary sequencer was used to attain the DNA sequences.

All EST data has been submitted to EST database, dbEST, and can be accessed with the GenBank accession numbers CK829215- CK830666.

2.6 *iv* Library Bioinformatics

The EST analysis pipeline developed by the NERC (Natural Environmental Research Council) funded Environmental Genomics Thematic Programme Data Centre (EGTDC), University of Edinburgh, was used to process the raw sequence trace data. The pipeline consists of the following programmes:

(1) **Trace2dbEST** (accessible through **http://envgen.nox.ac.uk/est.html**), analyses the electropherograms integrating three programmes, Phred (Ewing et al., 1998), Cross_match (P. Green, unpublished, available from **http://www.phrap.org/phredphrapconsed.html)**, and Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1997). These programmes trim the sequences by removing vector sequence, poor quality base calls and any sequence 3' to the poly (A) tail. BLAST then searches for similarity matches, using BLASTX against the

National Centre for Biotechnology Information non redundant protein sequence database (available for download from **ftp://ftp.ncbi.nlm.nih.gov/blast/db/nr.tar.gz**). 1541 good quality sequences from both libraries were submitted to dbEST.

(2) **Partigene** (Harcus et al., 2004) groups or clusters the ESTs on the basis of their similarities using CLOBB (Parkinson et al., 2002) and clusters were then assembled into contigs using Phrap (see above). This partial transcriptome database of non redundant sequences was annotated on the basis of their similarity using BLAST searches. Clusters were compared with two data sets: the NCBI non redundant (nr) protein sequence database and the Ensembl fugu assembly v.2 cDNA database (**ftp://ftp.ensembl.org/pub/current_fugu/data/fasta/cDNA/**). Partigene was also used to compare the *T. rubripes* partial transcriptome database against the 27.1b.1 release of the *T. nigroviridis* genome and cDNA sets (**ftp://ftp.ensembl.org/pub/current_tetraodon/data/fasta/cdna/; ftp://ftp.ensembl.org/pub/current_fugu/data/fasta/dna/)**

(3) **Prot4EST** was used as an EST translation tool to identify possible open reading frames (ORFs) and predict the corresponding polypeptide translations. These putative polypeptide protein sequences were classified according to their Gene Ontology (GO) as described by the GO consortium (Harris et al., 2004) using a further programme Annot8er_blast2GO (available from **http://www.nematodes.org/Partigene**). Annotation was based on BLAST searches against a GO slim subset of GO annotated SwissProt/TrEMBL database (downloadable from **ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/**). The consistently differentially expressed candidate genes were subjected to BLASTN searches against the *T. rubripes* genome database at ensembl using the default parameters.

Chapter 3: Cloning and structural characterization of orthologues of mammalian muscle regulatory factors in *Tetraodon nigroviridis.*

3. 1 Abstract

The myogenic regulatory factors (MRFs) are a group of four master transcription factors (Myf5, MyoD, myogenin and Myf6) that function by controlling muscle lineage determination and differentiation. The aim of this study was to clone and characterise all four members of the MRF family in the puffer fish species *T. nigroviridis* for structural comparison with MRFs in other species. This allows the possibility of further analysis for use as markers of myogenic cells or to improve the current understanding of the MRFs role in puffer fish muscle development. All four transcription factors were cloned and bioinformatic analysis was carried out identifying the functional motifs and other conserved regions including a 14-residue motif highly conserved among the MRFs and different species. All members show a high identity with orthologues from other species with MyoD for example sharing a 90% and 58% identity with *T. rubripes* and human MyoD respectively. The genes coding for the MRFs are all composed of 3 exons and have a relatively similar genomic structure in *T. nigroviridis*, *T. rubripes* and human. Introns are up to 8-fold larger in human, as shown with the second intron in *myogenin*. Structurally and functionally *T. rubripes* MRFs can be separated into two sub groups Myf5/ MyoD and myogenin/ Myf6.

3. 2 Introduction

Myogenic cells proliferate and differentiate under the synchronised control of the myogenic regulatory factors (MRFs) Myf5, MyoD, myogenin and Myf6 (Edmondson and Olson 1993). Their specific importance to muscle determination and differentiation has been demonstrated by their ability to transform a variety of cell types into myoblasts (Weintraub et al., 1989). The analysis of null mutations of MRFs in murine systems has shown these genes to function in a hierarchical and redundant fashion (Rudnicki et al., 1993). Myf5- /MyoD- double knock out mice lack myofibres (Rudnicki et al., 1993), indicating a potential role in myoblast formation for either factor while *myogenin* knockouts possess a disrupted differential step resulting in less skeletal muscle tissue (Hasty et al., 1993; Nabeshima et al., 1993). Myf5 and MyoD primarily function in myogenic determination of proliferating myoblasts while myogenin functions as a terminal differentiation factor of myogenesis. Myf6 has more recently been shown to serve both roles in part (Edmondson and Olson 1993; Kassar-Duchossoy et al., 2004). They are characterised by the presence of the homologous basic and helix-loop-helix motifs which bind to the E-box consensus sequence CANNTG and assist the formation of heterodimers with E-proteins respectively. E-proteins, such as E12 and E47 are found in many tissues and are involved in protein-protein interactions (Murre et al., 1989). E-boxes are found in the promoter region of many muscle specific genes including myosin heavy chains and creatine kinase (Lin et al., 1991; Edmondson and Olson 1993).

Several studies have characterised the MRF genes from teleost skeletal muscle and revealed their expression is restricted to somites and early muscle growth (Rescan et al., 1994; Rescan et al., 1995; Weinberg et al., 1996; Kobiyama et al., 1998; Chen et al., 2001; Coutelle et al., 2001; Temple et al., 2001; Tan et al., 2002; Hall et al., 2003). Taking advantage of the

current draft genome assemblies from puffer fish species, the MRFs were cloned enabling further sequence comparisons with other teleost and mammalian orthologues and the possibility of identifying novel functional motifs and transcriptional regulators. These clones are also then available to make cRNA probes designed for *in situ* hybridisation and other expressional analysis including qPCR.

3. 3 Results

3. 3 *i* Identification of Puffer fish MRFs

The coding sequence for three of the four MRFs, *MyoD*, *myogenin* and *Myf6,* were identified by *in silico* cloning in *T. nigroviridis*, amplified from fast myotomal muscle by RT-PCR, cloned and sequenced. MyoD, myogenin and Myf6 are located on Scaffold SCAF7217 of the genome assembly v36 (5th December 2005), chromosome 11 and chromosome 19 respectively. The genomic *T. nigroviridis Myf5* sequence was amplified from fast muscle and a predicted cDNA sequence was used for comparison. *Myf5* is located on chromosome 19 in close proximity to *Myf6*.

3. 3 *ii* Puffer fish MyoD

The *MyoD* cDNA sequence has been submitted to the Genbank database under the accession number AY616520. The majority of the *MyoD* transcript was amplified (818 bp) accounting for 97% of the complete coding sequence. Interestingly 3 splice variants of *MyoD* (Fernandes et al., unpublished) have been found in *T. rubripes* (accession numbers AY445315, AY4453156 and AY4453157) and the *T. nigroviridis* sequence has the highest identity with AY445315, the shortest splice variant with 843 bp. No splice variants were found in *T. nigroviridis MyoD*. The sequence codes for a putative protein of 272 amino acids (Q6IWL7) and contains both the basic and helix-loop-helix domains characteristic of myogenic regulatory factors. Multi-sequence alignment of *T. nigroviridis* MyoD with other vertebrate orthologues revealed a high conservation between species, confirming its identity as MyoD. It shares a 90%, 69%, 65%, 64%, 64%, 60% and 58% identity with *T. rubripes* (Q6SYW0),

zebrafish (Q90477), trout 2 (Q91206), trout 1 (Q91205), chick (P16075), mouse (P10085) and human (P15172) respectively (Fig 3. 1a). In contrast the helix-loop-helix domain showed a very high identity between species with *T. nigroviridis* sharing 92% identity with the corresponding human sequence. Single conservative amino acid changes were noted throughout the peptide sequences between species and some larger differences between teleosts and mammals and indeed between puffer fish and other teleosts were apparent. The most striking difference appears in the basic domain where both puffer fish species have a 27-residue serine and histidine rich region not found in the mammalian or other teleost sequences studied. Also C-terminus to the helix-loop-helix domain two sequences (4 and 8 residues) separated by 5 residues are unique to human and mouse but not found in teleosts or chick. This is followed by two highly conserved regions spanning 20 (HYSGDSDASSPRSNCSDGMT) and 14 (SSLDCLSSIVERIS) residues respectively. Towards the C-terminus there is a 25-residue sequence semi-conserved in mammalian MyoD not found in puffer fish.

3. 3 *iii* Puffer fish Myogenin

The cDNA sequences for *T. nigroviridis and T. rubripes myogenin* were submitted to GenBank under the accession numbers AY566282 and AY445318_respectively. The transcripts encoded for a putative protein composed of 248 residues (Q6PUV6), which contained the expected basic and helix-loop-helix domains. Multi-sequence alignment showed that myogenin from puffer fishes has a high degree of global conservation with homologous proteins from other vertebrates, sharing 61% identity with its human orthologue. In particular, within the helix-loop-helix domain the identity between *T. nigroviridis* myogenin and its homologues in human (P15173), chicken (P17920), *Xenopus* (Q8UUX8)

and *T. rubripes* (Q6Q2A7) was 91%, 88%, 86% and 98%, respectively (Fig 3.1b). Amino acid residues 178-197 of puffer fish myogenin corresponded to a serine-rich region (VSSSSEPSSGSTCCSSPEWS) that is highly conserved amongst teleosts.

3. 3 *iv* Puffer fish Myf5 and Myf6

Both genes coding for *Myf5* and *Myf6* lie on the negative strand of chromosome 19 in *T. nigroviridis* with *Myf5* being the first gene downstream of *Myf6*. RT-PCR, using gene specific primers, was attempted using cDNA from fast skeletal muscle but Myf5 could not be amplified. *Myf5* transcripts are most likely to be restricted to developing muscle of juvenile and embryonic stage fish as *Myf5* has been shown to be restricted to proliferating cells in developing muscle (Lindon 1998). The coding sequence for *T. nigroviridis Myf5* was predicted by aligning the *T. rubripes* coding region (NM_001032770) with the *Myf5* genomic region (DQ453127) of *T. nigroviridis*. The predicted transcript is 741 bp in length and codes for a 246-residue peptide (Q6SYV6) containing the characteristic basic and helix-loop-helix functional motifs. The putative *T. nigroviridis* Myf5 protein shares an 88% identity with its *T. rubripes* orthologue. Multiple sequence alignment with Myf5 peptides from other species including human (P13349), chick (Q08856) and rainbow trout (Q5UEM3) revealed a 59%, 58% and 77% identity respectively (Fig 3.1c). The 9-residue destruction box (D-box) (RRRLKKVNH, residues 103-111), previously shown in mammals to be the site of cell cycle proteolytic breakdown (Lindon et al., 2000b), is conserved in the basic helix-loop-helix motif of *T. nigroviridis* Myf5 and all other orthologues with the only difference being the last residue which is a glutamine in mammals, birds and amphibians. Towards the N-terminus there is a 10-residue sequence (EFGDEFVPRV) found in human Myf5 and other mammals but not present in the teleost species investigated. A 12-residue domain (SSLECLSSIVDR)

towards the C-terminal of the protein is highly conserved among all species. In close proximity to this 12-residue domain lies a run of nine glycine residues that only appears to be present in *T. nigroviridis* Myf5.

The *T. nigroviridis Myf6* mRNA sequence was submitted to the genbank database under the accession number AY576806. The sequence contained some 5' and 3' UTR as well as the complete coding sequence (CDS). The 714 bp CDS translates to a putative 238 amino acid peptide containing the expected basic and helix-loop-helix domain. Multi-alignment with orthologues in other species confirmed its identity as Myf6. It has a 96% and 57% identity with *T. rubripes* (Q90ZL0) and human (P23409) peptide sequences respectively and as expected a high identity (82%) in the conserved helix-loop-helix domain with human (Fig 3. 1d). Very few major differences were found between the peptides in teleosts and mammals except for a 4-residue sequence (NDLR) found at the N-terminal of the protein. A 14-residue sequence (SSLECLSSIVDR(L/I)S) towards the C-terminus of the protein is highly conserved among all species in both Myf5 and Myf6.

3. 3 *v* Comparison of MRF peptides in puffer fish.

Conceptual translation of all the MRF transcripts revealed the conserved 12-residue sub-domain in the basic region, containing the muscle recognition motif 'AT' (Davis et al., 1990; Brennan et al., 1991). Overall identity between the four proteins averaged between 35 and 40% with the lowest being 33% between myogenin and Myf5. The more conserved region between the 4 peptides is the 41-residue helix-loop-helix domain sharing between 70 and 80% identity, with MyoD and Myf5 having the highest at 80% (Fig 3.2). Little or no conservation out with these two regions was observed. A 14-residue sequence towards the C-

terminal of MyoD (SSLDCLSSIVERIS) appears to be conserved in all family members with only 5 of the residues conserved in myogenin. This region appears to be universally conserved among Myf5, MyoD and Myf6 in all species considered and poorly conserved in myogenin. Despite the conservation of the basic domain of the same gene between species there is very little conservation between MRF family members in this region. The C-terminus of the MRFs varies considerably after the 14-residue conserved region, with the region of Myf5 and MyoD being almost 3-fold longer than the corresponding region in Myf6 and myogenin.

3. 3 *vi* Genomic organisation of the MRFs in puffer fish.

Alignment of the cDNA with the corresponding genomic sequence revealed the *T. nigroviridis* MRFs to be composed of 3 exons (Fig 3.3). *T. nigroviridis* genomic *myogenin* (1184 bp) and *Myf5* (1418 bp) were amplified by RT-PCR, cloned and the sequences submitted to genbank under the accession numbers AY822074 and DQ453127. The genomic sequences for *MyoD* and *Myf6* (1366 bp and 1324 bp respectively) were taken from the *T. nigroviridis* ensembl genomic database. The first exon codes for the basic helix-loop-helix motif in all MRFs and the exon splice sites for MyoD, myogenin and Myf6 are conserved between the puffer fish and their human orthologues. Myf5 exon 1/ exon2 splice site is conserved between puffer fish and human, however the exon 2/ exon 3 splice site is not conserved.

Exon 1 varies in size marginally between the four genes: 588 bp, 534 bp and 523 bp in *MyoD*, *myogenin* and *Myf6* respectively, with *Myf5* being the smallest at 454 bp. The greatest differences were observed in the intronic regions and the second and third exons. Exon 3 is

proportionally similar with respect to size between *myogenin* and *Myf6* and likewise between *MyoD* and *Myf5* with the latter pair being almost 2-fold bigger than the corresponding exon in *myogenin* and *Myf6*. The opposite is true when comparing intron size where, particularly in the second intron, *Myf5* and *Myf6* are more similar in size when compared to the corresponding intron in *MyoD* and *myogenin*. Interestingly, intron 1 in human *Myf6* is approximately 2-fold smaller than the corresponding puffer fish intron.

The overall identity of the exons between *T. nigroviridis Myf5*, *MyoD*, *myogenin* and *Myf6* and their orthologues in human was 63%, 68%, 62%, 66% respectively. However no evolutionary conserved intronic regions (identity >60% over 100bp) were observed. Despite the similarities between puffer fish and human MRFs some striking structural differences were apparent. In general the human MRF genes had larger introns. Taking *myogenin* as an example, its exons are similar in size between both puffer fish and human but notable differences are observed when comparing intron sizes. The first and second human introns are approximately 2-fold and 8-fold longer than the corresponding introns in puffer fish. Intron 1 is 20% longer in *T. rubripes* than the corresponding intron in *T. nigroviridis*.

3. 4 Discussion

The orthologues of three of the four mammalian myogenic regulatory factors (MRFs) *MyoD*, *myogenin* and *Myf 6* have been cloned and characterised in the green spotted puffer fish *T. nigroviridis*. *Myf5* predicted cDNA sequence was identified by aligning the *T. rubripes* coding sequence with *T. nigroviridis* genomic sequence to identify the exon structure. Structurally the genes are comparable, being made up of 3 exons and 2 introns all of similar size. The similarities between *T. nigroviridis* and *T. rubripes* at the peptide and indeed nucleotide level are apparent. The most striking comparison is between the introns in puffer fish and human where human intronic regions in all genes are substantially larger. For instance intron 2 of *myogenin* is 8-fold larger than the corresponding intron in puffer fish. This is in agreement with the fact that puffer fish genomes are highly compact when compared to mammalian genomes due to small intergenic and intronic regions and a low percentage of repetitive elements (Aparicio et al., 2002; Jaillon et al., 2004). For this reason the puffer fishes are ideal model species for comparative genomics work (Venkatesh and Gilligan 2000).

Three splice variants of *T. rubripes MyoD* have been cloned (Fernandes et al, unpublished) and only one was identified in *T. nigroviridis*. It is possible these splice variants are only found in *T. rubripes*. A more likely explanation may be due to the stage of development from which the cDNA used for transcript amplification came. *T. rubripes MyoD* was amplified from whole larvae whereas the *T. nigroviridis* orthologue was amplified from adult fast skeletal muscle. It can therefore be suggested these splice variants are only transcribed in larval/ embryonic muscle. Indeed they may not code for functional proteins and further functional work would be required to investigate this.

Myf5, MyoD, myogenin and MRF4 peptides in *T. nigroviridis* have more than 55% identity with their human orthologues and more specifically the conserved basic helix-loop-helix (HLH) domain of human myogenin has extremely high identity with the corresponding domain of *T. nigroviridis* myogenin sharing 91% identity. The highly conserved 12 amino acid region (RR(K/R)A**AT**LRE(R/K)RRL) of the basic domain, responsible for DNA binding and heterodimerisation with the HLH protein E12 (Brennan et al., 1991), is present in all four MRFs as well as the two adjacent amino acids Ala and Thr of this motif characteristic of their muscle specificity (Davis et al., 1990; Brennan et al., 1991). A serine rich region of the basic domain was identified in MyoD which only appears to be present among puffer fish. Another serine rich domain in the C-terminal of myogenin is universally conserved among amongst zebrafish (*Danio rerio*) (Chen et al., 2000), striped sea bass (*Morone saxatilis*) (Tan et al., 2002), rainbow trout (*Oncorhynchus mykiss*) (Rescan et al., 1995), channel catfish (*Ictalurus punctatus*) (Gregory et al., 2004), common carp (*Cyprinus carpio*) (Kobiyama et al., 1998), *T. rubripes* and *T. nigroviridis*. These regions contain possible serine phosphorylation sites and therefore may have a role in the transcriptional activity of the proteins. The serine rich domain of MyoD unique to puffer fish suggests a possible transcriptional role of MyoD only required in tetraodontiforms. A further highly conserved 14-residue region near the C-terminus was identified in three of the MRFs (Myf5, MyoD and Myf6) and forms part of the previously characterised 50-residue α helical structure referred to as helix III involved in the activation of transcriptionally repressed genes during myogenesis (Gerber et al., 1997; Bergstrom and Tapscott 2001). This region is less well characterised in myogenin and has previously been shown to not function as the corresponding MyoD sequence (Bergstrom and Tapscott 2001). The remainder of this motif is less well characterised between mammalian and teleost MRFs suggesting these 14-residues are an important region for the function of the domain. It is feasible to propose this domain

has a function common to Myf5, MyoD and Myf6 that is not required by myogenin. Taking the function of the domain and the MRFs individual function, Myf5, MyoD and Myf6 all have functions in muscle determination whilst myogenin functions solely as a differentiation factor, suggesting this domain common to Myf5, MyoD and Myf6 may be related to the initial determination of MPCs and initiation of *myogenin* transcription and genes transcriptionally repressed during myogenesis.

Analysis of the MRFs as a group revealed a separation of the four transcription factors into two more similar groups of two. Genomic structure and peptide size are more comparable between Myf5/ MyoD and myogenin/ Myf6. Myogenin shares a 40% identity with Myf6 and a 33% identity with Myf5. Exon 3 which translates out with the conserved bHLH domain is also almost 2-fold bigger in MyoD/ Myf 5 than myogenin/ Myf6 accounting for the larger C terminus found in MyoD and Myf5. This is consistent with previous work proposing the four genes were derived from a single ancestral gene before being split into 2 lineages (*MyoD/ Myf5*; *myogenin/ Myf6*) during an event early in vertebrate evolution (Atchley et al., 1994).

3. 5 Figures

Figure 3. 1a MyoD Protein Alignment

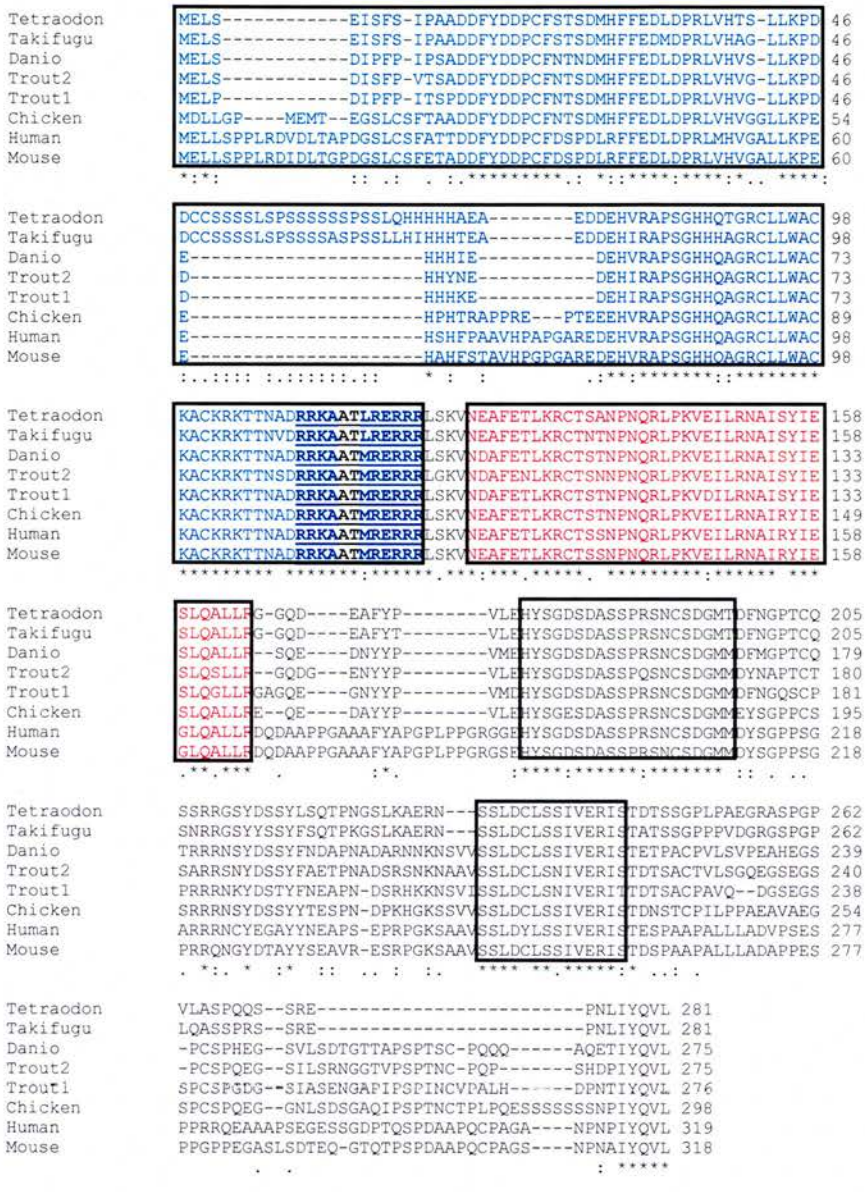


Figure 3. 1a. Multiple sequence alignment of MyoD protein sequences. The inferred *T. nigroviridis* sequence (Q6IWL7) is compared with those of *T. rubripes* (Q6SYW0), zebrafish (Q90477), trout (Q91205), chick (P16075), mouse (P10085) and human (P15172). Conserved substitutions are marked with a colon, semi conserved substitutions are marked with a dot and global conservation is represented with an asterisk. The basic (blue) and HLH (red) motifs are highlighted and 2 other highly conserved regions are represented with boxes downstream of the HLH region. The 12-residue subdomain of the basic region is underlined with the muscle recognition motif 'AT' (black).

Figure 3. 1b Myogenin Protein Alignment

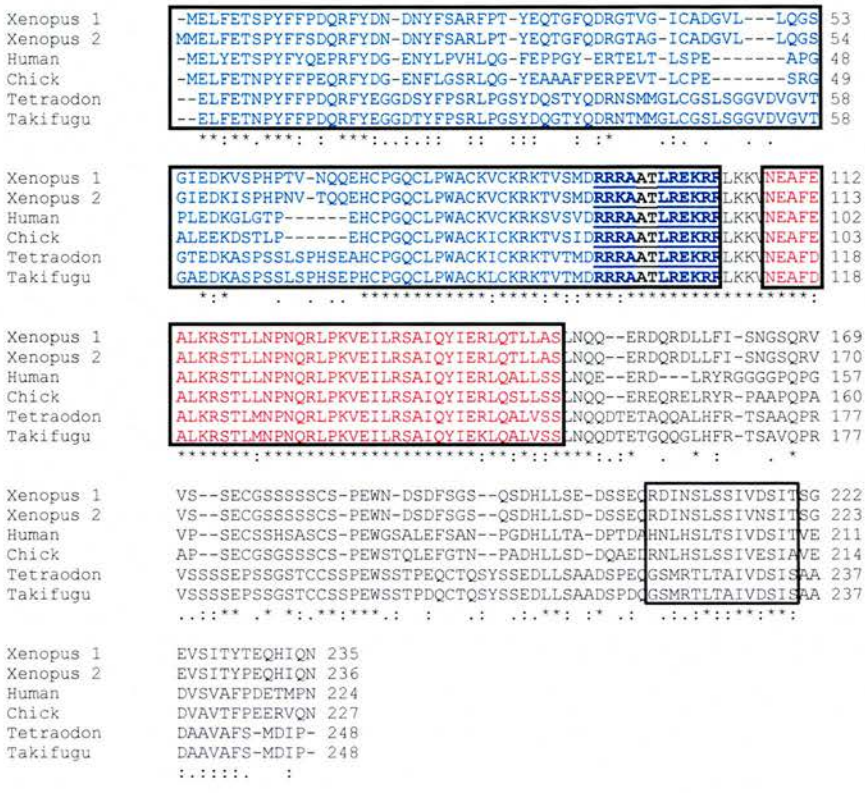


Figure 3. 1b. Multiple sequence alignment of myogenin protein sequences. The inferred *T. nigroviridis* sequence (Q6IWL7) is compared with those of human (P15173), chicken (P17920), xenopus 1 (Q8UUX8), Xenopus 2 (Q8UUX7) and *T. rubripes* (Q6Q2A7). Conserved substitutions are marked with a colon, semi conserved substitutions with a dot and global conservation is represented with an asterisk. The basic (blue) and HLH (red) motifs are highlighted with boxes and the region corresponding to a 12-residue domain, downstream of the HLH domain, conserved in the other MRFs is also highlighted with a box. The 12-residue subdomain of the basic region is underlined with the muscle recognition motif 'AT' (black).

Figure 3. 1c Myf5 Protein Alignment

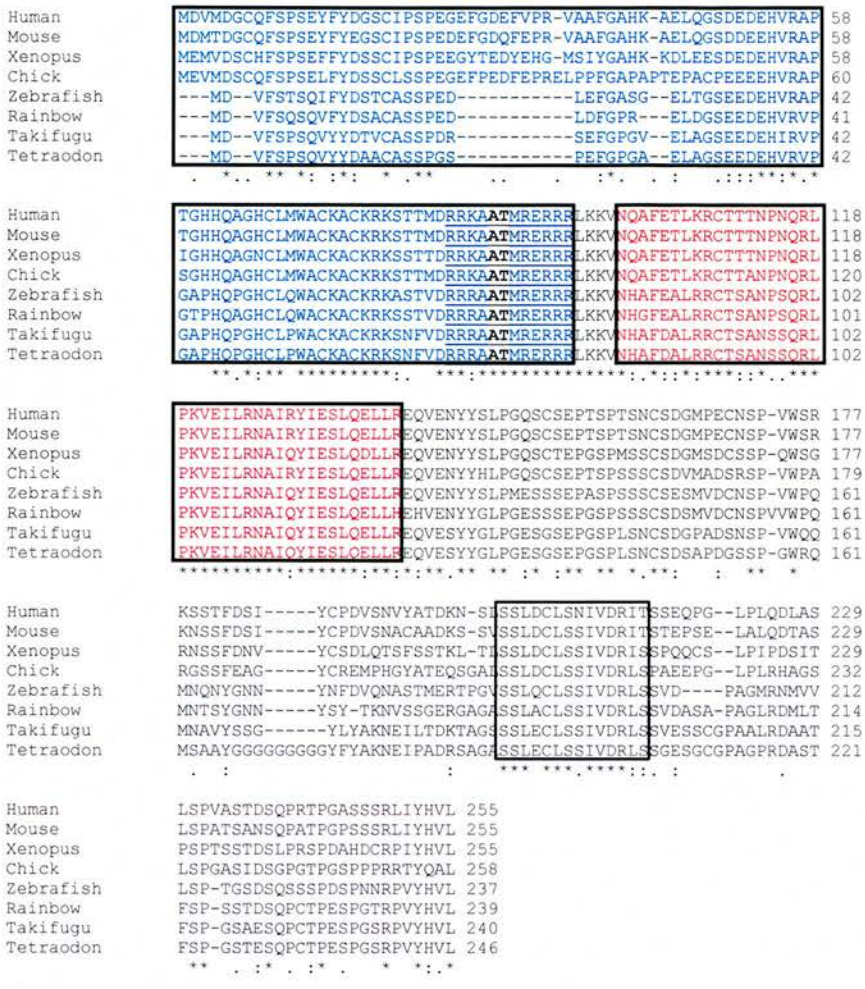


Figure 3. 1c. Multiple sequence alignment of Myf5 protein sequences. The inferred *T. nigroviridis* sequence (Q6IWL7) is compared with those of *T. rubripes* (Q6SYV6), rainbow trout (Q5UEM3), zebrafish (Q5CZP8), chick (Q08856), *Xenopus* (P24700), mouse (P24699) and human (Q5UEM3). Conserved substitutions are marked with a colon and global conservation is represented with an asterisk. The basic (blue) and HLH (red) motifs are highlighted and a highly conserved 12 residue region is represented with a box C-terminal to the HLH motif. The 12-residue subdomain of the basic region is underlined with the muscle recognition motif 'AT' (black).

Figure 3. 1d Myf6 Protein Alignment

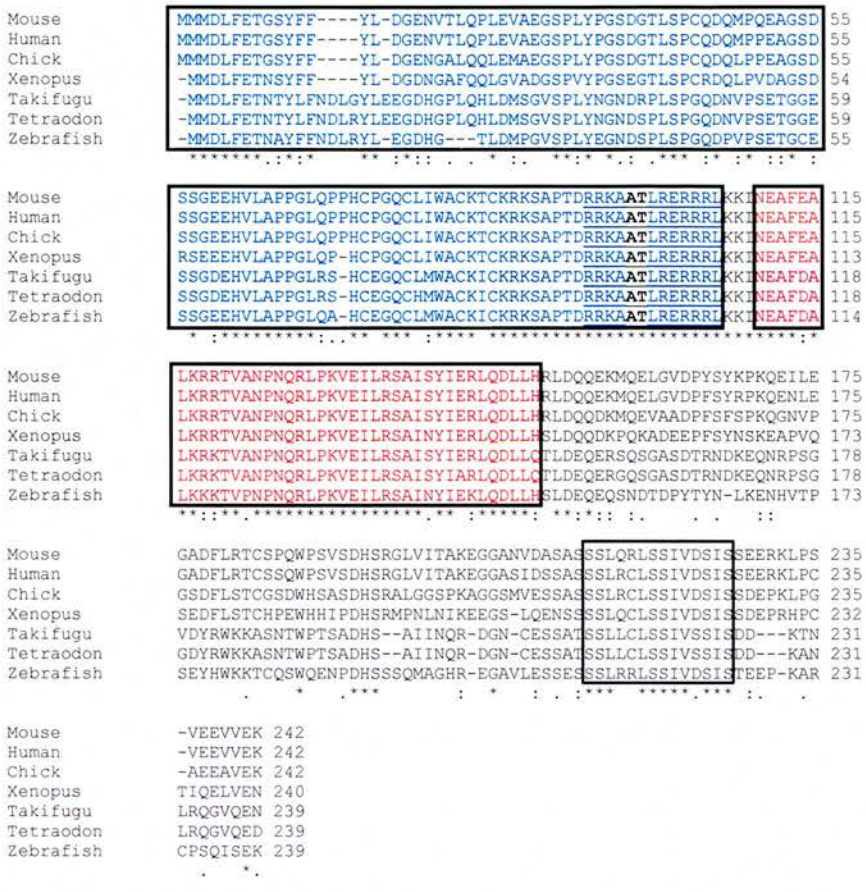


Figure 3. 1d. Multiple sequence alignment of Myf6 protein sequences. The inferred *T. nigroviridis* sequence (Q6PUV) is compared with those of *T. rubripes* (Q90ZL0), zebrafish (Q6VNZ9), chick (Q01795), *Xenopus* (Q92020), mouse (P15375) and human (P23409). Conserved substitutions are marked with a colon, semiconserved substitutions with a dot and global conservation is represented with an asterisk. The basic (blue) and HLH (red) motifs are highlighted and a highly conserved 12-residue region is represented with a box, downstream of the HLH motif. The 12-residue subdomain of the basic region is underlined with the muscle recognition motif 'AT' (black).

Figure 3. 2a *T. nigroviridis* MRF Protein Alignment

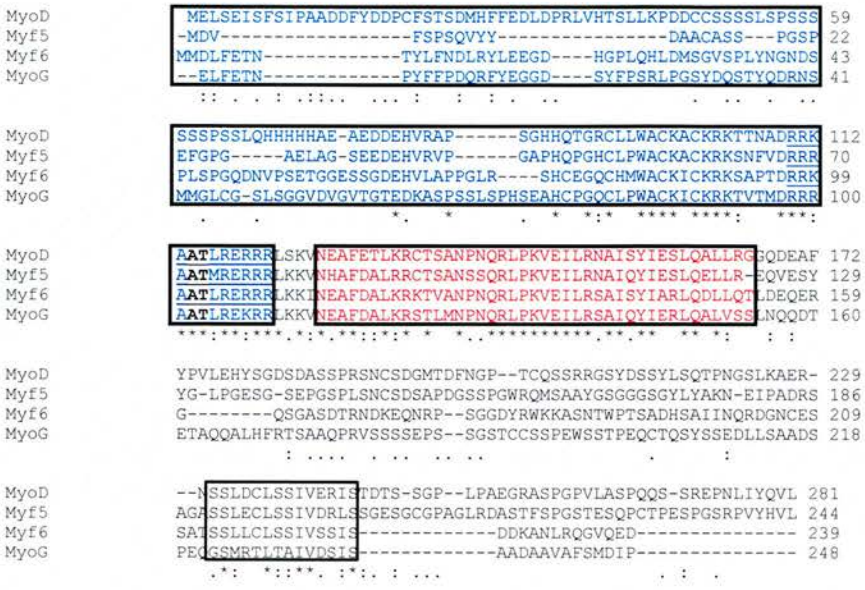


Figure 3. 2a. Multiple sequence alignment of *T. nigroviridis* MRF protein sequences. Conserved substitutions are marked with a colon, semi conserved substitutions with a dot and global conservation is represented with an asterisk. The basic (blue) and HLH (red) motifs are highlighted and a highly conserved 12-residue region is represented with a box, downstream of the HLH motif. The 12-residue subdomain of the basic region is underlined with the muscle recognition motif 'AT' (black).

Figure 3. 2b *T. nigroviridis* MRF exon structure

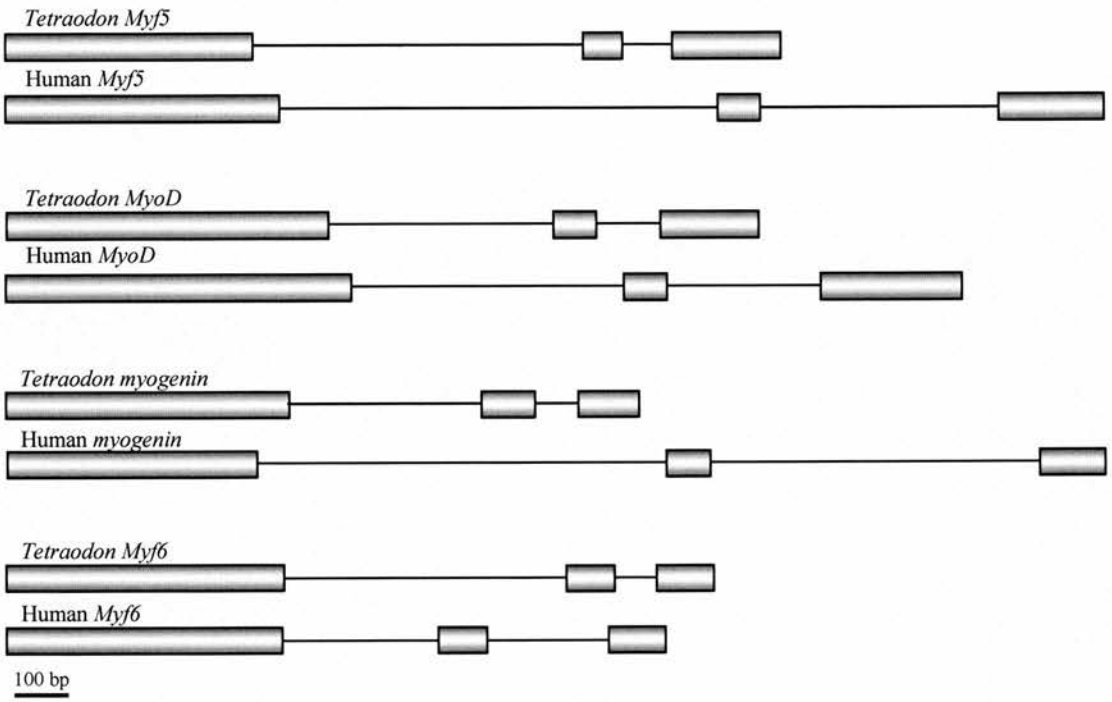


Figure 3. 2b Transcript structure of the MRFs from *T. nigroviridis* and human. Bars represent exons and lines represent introns. The scale bar indicates 100 bp.

Chapter 4: Further characterization of *myogenin* and the thermal plasticity of *myogenin* expression during early development of *T. rubripes*

4. 1 Abstract

Myogenin is a member of the MyoD family of muscle specific transcription factors, characterised by their functional basic helix-loop-helix motifs. The expression of *myogenin* in *T. rubripes* was characterised and its genomic structure compared to *T. nigroviridis* and human. Analysis of the genomic neighbourhood of the *myogenin* gene in the two puffer fish and human revealed a large region (>100 kb) of conserved synteny between the three species and promoter analysis revealed putative binding sites for the muscle specific transcription factors MEF2 and MEF3. Expression analysis of *myogenin* with an *in situ* hybridisation assay revealed a similar expression pattern to that previously described in zebrafish with transcripts appearing in a rostral-caudal gradient during somite development followed by a second wave of expression in the pectoral fin bud primordia, jaw and extraocular muscle precursors. *Myogenin* transcripts were also shown to be localised to nuclei surrounding small diameter fibres in adult *T. rubripes* skeletal muscle sections. Relative expression of *myogenin* was investigated throughout development in embryos incubated at 15, 18 and 21°C. The time taken to develop one somite pair was more than 3-fold slower at 15°C compared to 21°C. Normalising expression between the temperatures relative to 18°C revealed a 3-fold higher peak of *myogenin* expression at 21°C compared to 15°C during early somitogenesis.

4. 2 Introduction

Myogenin is a muscle-specific basic helix-loop-helix transcription factor that plays an essential role in the specification and differentiation of myoblasts (Edmondson and Olson 1993). Members of the same family of muscle regulatory factors (Myf5, MyoD, Myf6) characteristically bind to E-boxes in the promoter region of muscle-specific genes (Lin et al., 1991). Myogenin is a direct transcriptional target of both Myf5 and MyoD (Hollenberg et al., 1993; Buchberger et al., 1994) and is irreversibly activated upon growth factor removal and cell differentiation (Olson and Klein 1994). MRF genes show unique spacial and temporal expression patterns during somitogenesis and fin muscle development in teleosts (Rescan et al., 1994; Rescan et al., 1995; Chen et al., 2001; Temple et al., 2001).

Embryonic temperature has a profound effect on both the rate and phenotypic outcome of myogenesis in ectotherms such as teleosts (Johnston and Hall 2004). Phenotypic changes in fibre number and size in larval and juvenile fish are apparent. However little is known about the effect temperature has on transcript levels and the genetic networks regulating myogenesis. Temple et al. (2001) found that incubation temperature had no effect on the timing of *myogenin* expression with respect to somite stage in Atlantic Herring. In contrast, *myogenin* and *MyoD* transcripts were found in more somites of rainbow trout up to the 45 somite stage reared at 12°C compared to 4°C (Xie et al., 2001). In the present study, using a combination of *in situ* hybridisation and quantitative real time PCR (qPCR), *myogenin* expression was investigated testing the hypothesis that expression shows developmental-stage specific changes with respect to embryonic incubation temperature. In addition a comparative genomic study of puffer fish *myogenin* was carried out investigating its synteny with respect to human to identify genes with potential co-regulation.

4. 3 Results

4. 3 *i* Promoter analysis of *myogenin*

Analysis of the 5 kb region (from ensembl *T. rubripes* assembly version 35) upstream of the *myogenin* translation start site (TSS) revealed putative binding sites for a number of transcription factors in *T. rubripes* and *T. nigroviridis*. A muscle specific TATA box and binding sites for the muscle specific transcription factors MEF2 (GCTAAATTTAACCCTA) and MEF3 (TGTCGGGTTTC) were found within the first 250 bp upstream of the translation start site (TSS) (Fig 4.1). A number of possible E-boxes were also found further upstream suggesting binding sites for a variety of transcription factors including E2F, E47 and MyoD (Murre et al., 1989).

4. 3 *ii Myogenin* Synteny analysis

Comparative mapping of the genes surrounding *myogenin* in the two puffer fish species (*T. rubripes* and *T. nigroviridis*) and human revealed a large region of conserved synteny (>100 kb). The *myogenin* gene is located on chromosome 11 and 1 in *T. nigroviridis* and human respectively and in *T. rubripes* is found on scaffold 208 (Fig 4.2). Upstream of *myogenin* in puffer fish lies the gene coding for an important P53 binding protein, *Mdm4*. Further upstream lie the genes *sort1* (sortilin precursor), *psma5* (proteasome subunit alpha type 5), *per3* (period circadian protein 3) and *vamp3* ( ). Downstream of *myogenin* lies an unknown gene with no apparent conserved domains and the transcription factor EB (*tfeb*).

The synteny is remarkably well conserved between *T. rubripes* and *T. nigroviridis*. Small differences can be noted in the size of genes and intergenic regions but gene order appears to have remained intact over the 18-30 million years of evolutionary distance between the two puffer fishes (Hedges 2002). Despite the size of intergenic regions and some chromosomal rearrangements the genes all lie in the same region of conserved synteny between puffer fish and human, with the exception of *tfeb* which lies on chromosome 6 in human.

4. 3 *iii* Developmental expression patterns and effect of developmental temperature on *myogenin* in *T. rubripes*

Expression of *myogenin* and timing of developmental stages in *T. rubripes* was investigated throughout embryogenesis at 3 temperatures (15°C, 18°C and 21°C) and also in adult fast muscle tissue sections. Environmental temperature had a profound effect on the timing of embryonic events. Somitogenesis in *T. rubripes* begins ~55 hours post fertilisation (hpf) at 18°C; in contrast at 15°C and 21°C the onset of segmentation was at 87 hpf and 45 hpf respectively. The times for the development of one somite pair (somite interval) at the 3 temperatures were 120 min at 15°C, 96 min at 18°C and 38 min at 21°C respectively (Fig 4. 3). Somite interval was then used as a time point to normalise development between the temperatures from the onset of segmentation.

Expression of *myogenin* was analysed throughout development using quantitative real time PCR (qPCR) and *in situ* hybridisation. qPCR results revealed a peak of relative expression which correlated with early somitogenesis at all temperatures and a second less severe rise in expression before hatch (Fig 4.4a). This pattern of expression was consistent between temperatures however some striking differences in relative expression were evident.

Developmental time (hpf) was normalised between temperatures using somite interval as described above (4.3 *iii*). An ANOVA using hpf as the factor revealed an overall significant difference in relative expression levels throughout development at 15°C ($p=0.009$), 18°C ($p=0.001$) and 21°C ($p=0.002$). The expression of *myogenin* during early somitogenesis was highest at 21°C with the corresponding peaks at 18°C and 15°C being ~2-fold ($p=0.029$) and ~3-fold ($p=0.029$) less respectively (Fig 4.4b). The subsequent decrease in expression appeared most severe at 21°C and the second rise in expression was most prominent during hatch at 18°C compared to 15°C ($p<0.005$).

For correlation with the qPCR results expression was investigated throughout development at 18°C by *in situ* hybridisation and 3 stages were chosen to analyse potential changes in expression with respect to the 3 temperatures: 10-somite stage, end of segmentation and just prior to hatch. The first signs of *myogenin* expression were detected at the onset of segmentation (~55 hpf), in the adaxial region of the somites adjacent to the developing notochord (fig 4. 5B). *Myogenin* expression was shown as bands that extended laterally from the medial region of the developing somites. Throughout somitogenesis *myogenin* was expressed in a rostral-caudal gradient and was progressively down-regulated in mature somites. By the end of segmentation (~100 hpf) it was confined to the most posterior tail somites (fig 4.5B-F). At this stage the number of somites ranged from 27 to 29. At the high-pec stage of the pharyngula (~160 hpf) period *myogenin* transcripts could be detected as two pairs of symmetrical stripes in the pectoral fin buds (fig 4.5H) and in the extraocular muscles and jaw muscle progenitors (Fig 4.5G). *Myogenin* was expressed in these cells during late embryonic development until hatching. No obvious differences in expression could be seen with respect to temperature using *in situ* hybridisation (results not shown). In adult fast muscle *myogenin* expression was confined to potential differentiating myoblasts surrounding

the smaller myofibres (Fig 4.6). Samples used with sense probes in both *in situ* assays were unstained.

4. 4 Discussion

In the present study the developmental expression of the myogenic regulatory factor *myogenin* has been investigated in *T. rubripes* in response to changes in embryonic temperature. Promoter and synteny analyses were also carried out.

Analysis of the genomic neighbourhood surrounding *myogenin* revealed a large region of conserved synteny (>100kb) containing genes of varied function and interest. The gene coding for the transcription factor EB (Tfeb) lies downstream of *myogenin* and is from the same family of bHLH proteins that recognise E box binding sites (Carr and A. 1990; Le Blanc et al., 1998). *Mdm4* is the first neighbouring gene upstream of *myogenin* and codes for an important protein which inhibits p53-mediated cell cycle arrest and apoptosis (Shvarts et al., 1997). As these genes are in close proximity to one another it is acceptable to propose they may share transcriptional regulators. Another interesting gene found in the vicinity of *myogenin* is *per3*. This circadian clock gene is one of 3 mammalian homologues of the *Drosophila* period gene known to be widely expressed with prominent synchronous circadian oscillations in most mouse tissues, including skeletal muscle (Zylka et al., 1998). In light of this previous work by Johnston et al. (2003) revealed manipulating photoperiod to have a profound effect on muscle growth in Atlantic salmon, *Salmo salar* (Johnston et al., 2003c). A continuous light treatment had the effect of increasing the number of muscle progenitor cells (MPCs) by either increasing the speed of the cell cycle and/or increasing the number of cell divisions (Johnston et al., 2003c). Further to this the discovery of a cell cycle regulating protein synchronized by circadian rhythms in zebrafish (Dekens et al., 2003) implicated the cell-autonomous circadian clock in the regulation of the vertebrate cell cycle by light.

Myogenin has been shown to be specifically expressed in the developing somites and skeletal muscle of mouse (Wright et al., 1989) and teleosts including zebrafish (Weinberg et al., 1996) and herring (Temple et al., 2001). We have analysed the developmental expression pattern of *myogenin* in *T. rubripes* embryos incubated at three temperatures (15, 18 and 21°C) and our results show that *myogenin* expression in *T. rubripes* follows the same pattern as previously observed in other teleosts including zebrafish and herring. It follows a similar pattern to *MyoD* expression which, in zebrafish, starts during mid gastrulation and persists until just prior to segmentation (Weinberg et al., 1996). The expression progresses in a rostral caudal progression following somite development and indeed fades following the same pattern (fig 4. 5B-F). *Myogenin* expression in the somites increases laterally away from the adaxial cells but progressively fades medially resulting in expression always being highest in the medial central portion of the somites. After 100 hpf expression fades to a minimum with expression being confined to the most caudal somites before transcripts were no longer detected post-somitogenesis. In contrast to *MyoD* (Temple et al., 2001) transcripts were not present in the adaxial cells of the pre-somitic mesoderm. This is consistent with the specific role of *myogenin* in terminal differentiation unlike the other MRF family members who all have a role in myogenic determination. Expression appears again in the fin bud primordia and extraocular and jaw muscle progenitors close to hatching (fig 4. 5G, H). We have also shown *myogenin* expression beyond embryogenesis in adult muscle tissue sections with transcripts appearing in areas of potential differentiating myoblasts surrounding small muscle fibres undergoing hypertrophic growth (fig 4.6). The high degree of structural conservation and similarity in expression patterns between *T. rubripes myogenin* and its homologues in other teleosts and mammals indicates there may be a similarity in function between fish and mammals.

Myogenin expression is first evident in the somites beginning at the onset of segmentation (10.5 hpf) during normal zebrafish development (Weinberg et al., 1996). Our results also show expression initiating at the commencement of segmentation with transcripts first appearing in the somites at ~55 hpf in *T. rubripes* reared at 18°C. Using embryos of known age (hpf) stained for *myogenin* expression we were able to ascertain rates of somitogenesis as well as a final somite number for 3 different incubation temperatures. Higher developmental temperature increased rate of somitogenesis significantly with a somite forming approximately every 2 hours at 15°C compared to every 0.8 hours at 21°C. The total somite number varied from 27 to 29 between individuals during development at 18°C.

Environmental temperature has been shown to have a substantial effect on muscle growth in teleosts (Cossins and Bowler 1987; Johnston and Temple 2002). This is thought to affect the expression of genes that are crucial for muscle growth. *In situ* hybridisation results did not show any significant change in *myogenin* expression with respect to developmental temperature; however, our qPCR data revealed substantial differences in *myogenin* transcript number. Xie and collaborators (2001) working on trout discovered *MyoD* and *myogenin* mRNAs in a greater number of somites in embryos of the same developmental stage but reared at 12°C compared to 4°C. Our qPCR results paralleled this observation in *T. rubripes* (fig 4.4A, B) showing that embryonic temperature has a marked effect on *myogenin* expression levels and suggest a possible genetic basis for temperature-induced developmental plasticity. Higher embryonic temperature not only accelerates development but increases the relative up-regulation of *myogenin* during segmentation. The second rise in *myogenin* expression suggests a secondary wave of myogenic differentiation post-segmentation. This correlates with *myogenin* expression in the fin bud primordia and extraocular and jaw muscle progenitors around the time of hatching. Indeed, mammalian muscle progenitor cells and

satellite cells originate and migrate from the medial wall and dorsal portion of the somite (dermomyotome) as pioneers for subsequent limb and trunk muscle development. Face and head muscles are derived from a separate unsegmented rostral mesoderm (Currie and Ingham 1998; Gros et al., 2005). Although little is know regarding the process of cell migration to form limbs and face muscle of teleosts, recently a dermomyotomal layer has been identified in developing somites of zebrafish. This poses a potential origin for the myogenic cells that migrate and subsequently differentiate into fin and face muscles (Devoto et al., 2006). It was therefore concluded that this second wave of *myogenin* expression is a result of the subsequent differention of the muscle progenitors into myotubes during limb and face skeletal muscle development. The changes in *myogenin* expression with respect to developmental temperature identified in this study have the potential to alter the outcome of the myogenic programme through changes in myoblast differentiation. In order to couple these findings with phenotypic developmental plasticity of muscle functional analysis in which *myogenin* expression was manipulated experimentally would be required.

4. 5 Figures

Figure 4. 1 *Myogenin* promoter analysis.

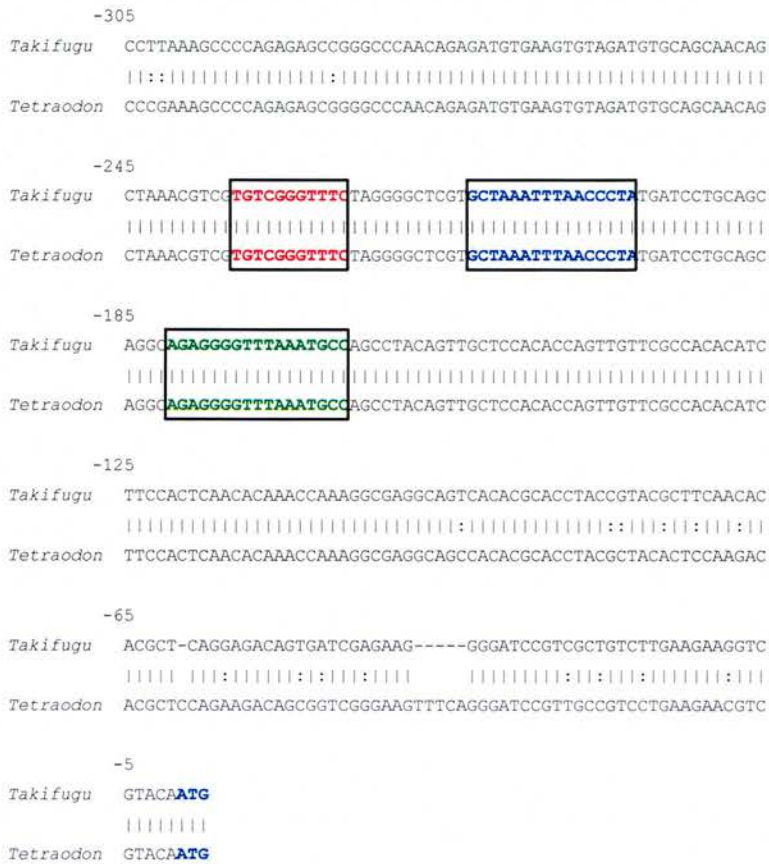


Figure 4. 1. Alignment of the 300 bp region upstream of the translation start site (ATG) of *myogenin* in *T. rubripes* (*Takifugu*) and *T. nigroviridis* (*Tetraodon*). The MEF3 (red) and MEF2 (blue) transcription factor binding sites are boxed. A muscle specific TATA box is also marked in green text. Lines represent conservation and discrepancies are highlighted with colons. Gaps in the sequence are marked with a dash.

Figure 4. 2 Partial *myogenin* synteny map

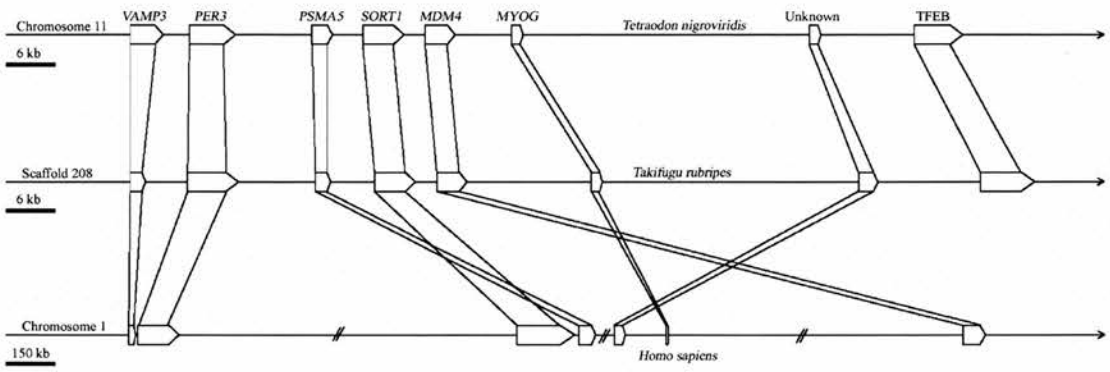


Figure 4. 2. Partial synteny map surrounding the *myogenin* gene (MYOG) in *T. rubripes*, *T. nigroviridis* and human. Genes are represented by block arrows relating to gene direction and the scale is shown for each organism.

Figure 4. 3 Somitogenesis at 15°C, 18°C and 21°C.

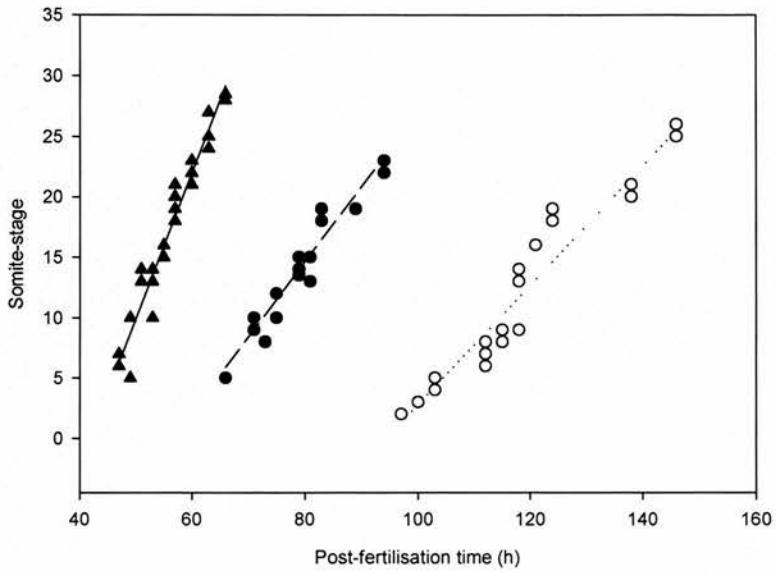


Figure 4. 3. Relationship between somite number and time elapsed post-fertilisation for *T. rubripes* embryos. Embryos reared at 15°C are represented with open circles, closed circles and closed triangles correspond to embryos reared at 18°C and 21°C respectively. Lines refer to the fitted first order linear regression obtained using the least squares method (n = 23 for 21°C, n = 16 for 18°C, n = 19 for 15°C).

Figure 4. 4 Relative expression of *myogenin* at 15°C, 18°C and 21°C.

Figure 4. 4a

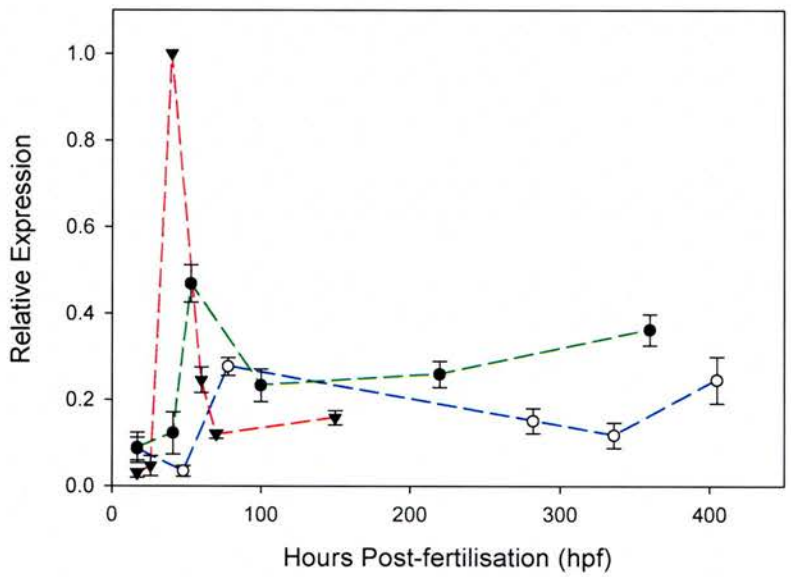


Figure 4. 4b

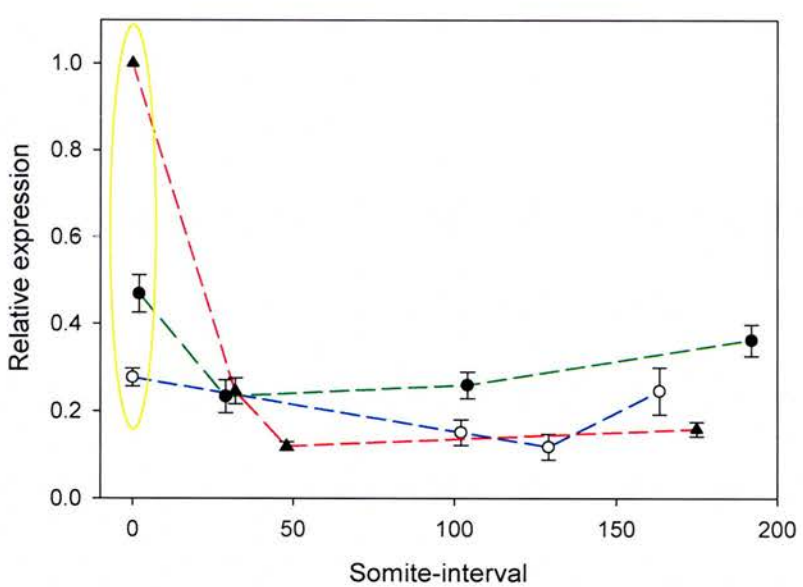


Figure 4. 4c

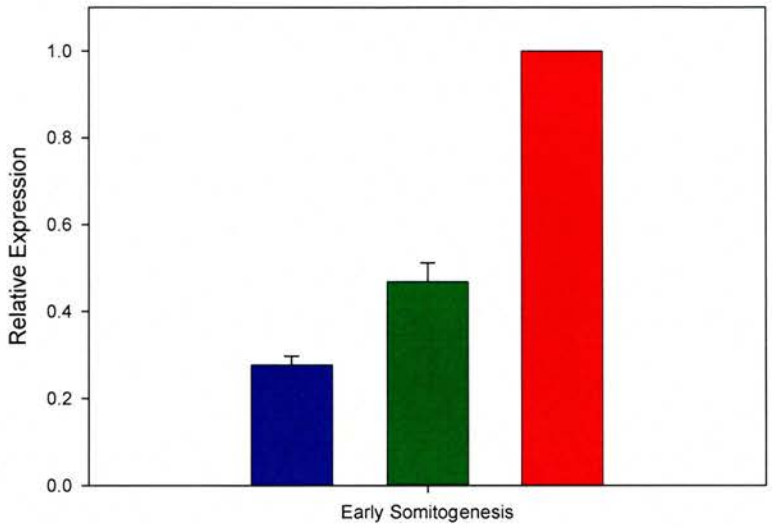


Figure 4. 4. Relative *myogenin* expression as determined using qPCR illustrating changes in relative expression during *T. rubripes* development at 15°C (blue), 18°C (green), 21°C (red) with respect to developmental time (4a) and somite interval (4b). 4c shows the relative expression at the same developmental stage during early somitogensis between temperatures, highlighted (yellow) on 4b. *Myogenin* expression was shown to differ significantly throughout development (15°C p=0.009, 18°C p=0.001, 21°C p=0.002). Corresponding points during early somitogenesis (circled yellow in 4b) were analysed for significant difference using a Mann-Whitney U test revealing a 2-fold (p=0.029) and 3-fold (p=0.029) greater expression during somitogenesis at 21°C compared to 18°C and 15°C respectively (4b). 18s Ribosomal RNA was used as the endogenous internal control and all points (n=4) were normalised to the highest expression value (40 hpf, 21°C).

Figure 4. 5 Whole mount *in situ* hybridisation

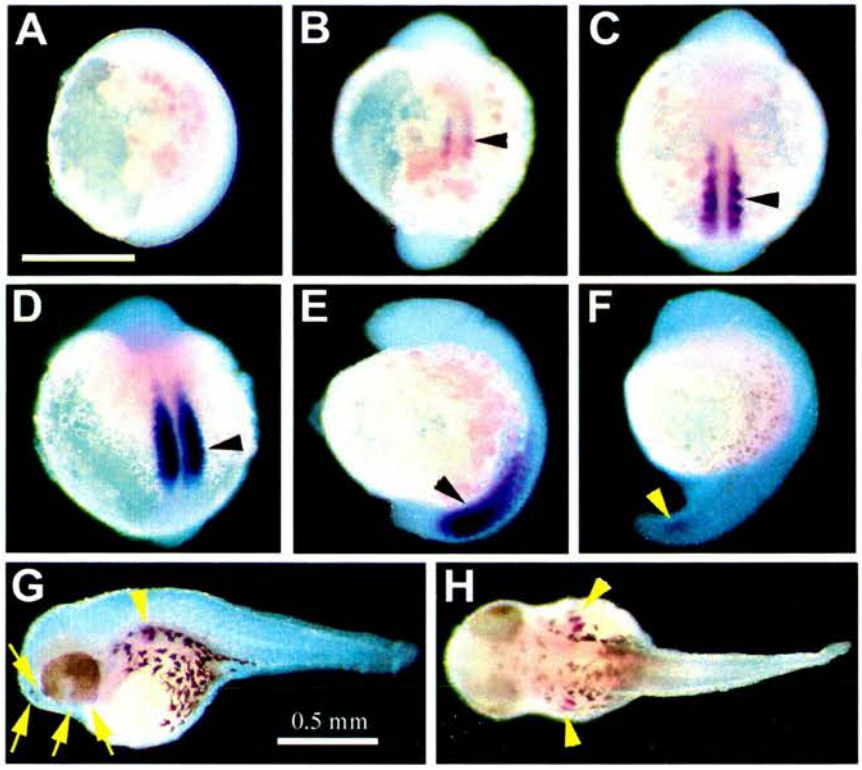


Figure 4. 5. Localisation of *myogenin* transcripts in *T. rubripes* embryos throughout development at 18°C by whole mount *in situ* hybridisation with a cRNA probe: (A) 42 hpf, bud stage of gastrulation, (B) 57 hpf, onset of segmentation, (C) 63 hpf, 5-somite stage, (D) 80 hpf, 14-somite stage, (E) 85 hpf, 17-somite stage, (F) 100 hpf, 25-somite stage, (G, H) 160 hpf, high-pec stage of pharyngula period. The anterior of the embryos is top except (G, H) where anterior is left. (B-F) Cells of the developing somites expressing *myogenin* are indicated with arrowheads. (G, H) cells of the fin buds are indicated with arrow heads, jaw and extraocular muscle are indicated with full arrows. Scale bar in (G) indicates 0.5 mm. Adapted from Fernandes et al, 2005.

Figure 4. 6. *In situ* hybridisation on fast muscle tissue section

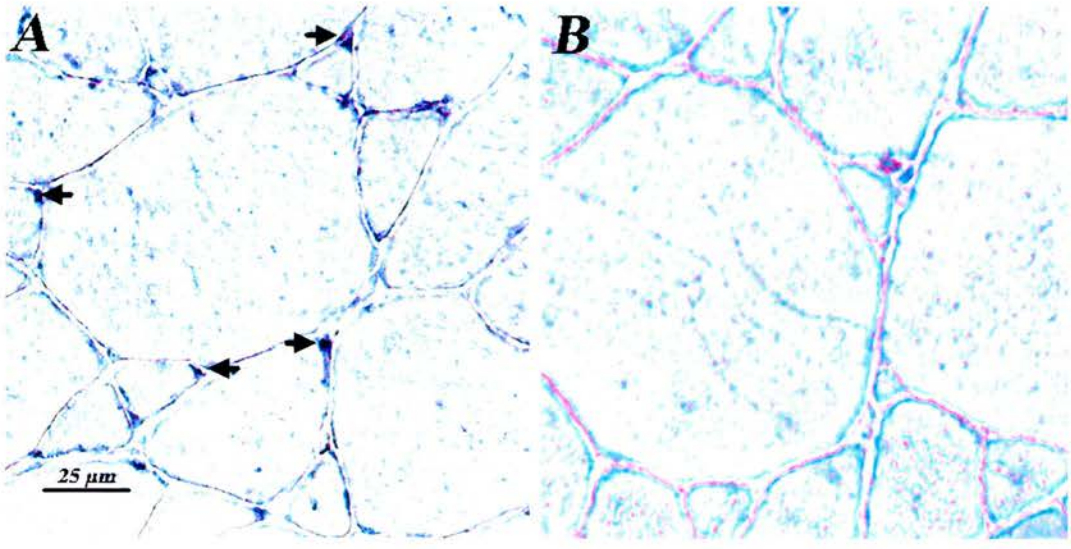


Figure 4.6. Localisation of *myogenin* transcripts in fast muscle cryosection from *T. rubripes* using antisense (A) and sense (B) riboprobes. Arrows indicate selected areas of putative differentiating myoblasts. Scale bar indicate 25 µm.

Chapter 5: Characterization of Sox8 and expressional analysis with respect to developmental temperature in *T. rubripes.*

5. 1 Abstract

The SRY (Y-linked sex-determining gene) related high mobility box protein Sox8 is a member of a large family of transcription factors involved in development and has been implicated in myogenic differentiation. *Sox8* genes from the puffer fishes *Takifugu rubripes* and *Tetraodon nigroviridis* are composed of 3 exons and have two introns, approximately 5- and 2-fold shorter than human *Sox8*. Analysis using qPCR and *in situ* hybridisation in adult and embryonic *T. rubripes*, respectively, localised *Sox8* expression in the developing mid and hind brain, somites and the pectoral fin bud primordia of embryonic stages and the brain, skeletal muscle and testes of adult *T. rubripes*. In adult fast muscle *Sox8* transcripts were localised to mononuclear cells corresponding to the myogenic progenitor cells (MPCs). We tested the hypothesis that the relative expression of *Sox8* would be influenced by egg incubation temperature. However relative expression throughout normalised development at three temperatures (15°C, 18°C and 21°C) using a qPCR assay identified no significant change in *Sox8* expression with respect to temperature. In addition we identified a large region of conserved synteny surrounding *Sox8* between *T. rubripes* and *T. nigroviridis* not conserved with its human counterpart. Our results pose Sox8 as a promising marker for identifying MPCs in teleosts and to have a major role in the development of muscle and brain tissues in teleosts.

5.2 Introduction

The Sox family of transcription factors have wide spread functions throughout embryogenesis and have been shown to be associated with a number of human congenital diseases (Jager et al., 1990; Foster et al., 1994; Southard-Smith et al., 1998). They were first discovered in the early nineties as a group of genes related to the Y-linked sex-determining gene SRY (Gubbay et al., 1990) characterised by the presence of the High Mobility Group (HMG) box domain. The HMG-box domain is found in a large superfamily of proteins including HMG-1, HMG-2, TCF1 and LEF1. It is a highly conserved motif and is involved in DNA-binding and bending (Laudet et al., 1993). A family of 20 pairs of *Sox* genes have been identified based on sequence homology in human and mouse, although new Sox genes are constantly being identified, (Schepers et al., 2002) and have been categorised into 8 groups (A-H) according to their functional and structural similarities (Bowles et al., 2000). Most Sox groups are represented by a single gene in invertebrates such as *Caenorhabditis elegans* suggesting an increase in gene number during evolution and genome expansion. Recently the complete complement of *Sox* genes has been cloned and characterised in *T. rubripes* (Koopman et al., 2004) revealing some unexpected findings including unique isoforms compared to mammals and the absence of orthologues to several mammalian isoforms. For example, group E, which contains *Sox8*, *Sox9* and *Sox10*, is made up of 6 genes in *T. rubripes* compared to the 3 previously found in mammals. A possible explanation for this may be the whole genome duplication in teleosts occurring since the divergence from a common ancestor to mammals or a lineage specific duplication within puffer fishes (Jaillon et al., 2004).

Sox8 has important roles in sexual development (Takada and Koopman 2003), osteoblast differentiation (Schmidt et al., 2005) and has been shown to be a marker of myogenic progenitor cells (MPCs) in mammals (Schmidt et al., 2003). MPCs in vertebrate adult muscle express a group of myogenic regulatory factors, the MRFs, and differentiate into muscle nuclei for the growth of muscle fibres (Edmondson and Olson 1993). Generally they are considered to arise from a population of pleuripotent stem cells that, under the right conditions, have the ability to form cell types other than muscle including adipocytes and chondrocytes (Wada et al., 2002). Mammalian stem cells express genes such as CD34, *FoxK1* and *Pax7* all of which can be used as markers for MPCs (Tajbakhsh 2003). Pax7 is thought to be important in the maintenance of mouse myogenic cells with *Pax7* knockout mice lacking myosatellite cells (Seale et al., 2000). *FoxK1* is highly expressed in mouse MPCs and *FoxK1* knock out mice have a severely impaired myogenic cell function (Garry et al., 2000). The paired box protein (Pax7) and the forkhead box protein (FoxK1) have previously been shown to be markers for MPCs in teleost muscle using immunohistochemistry (Johnston et al., 2004). Sox8 also has a potential role in delaying muscle differentiation by holding MPCs in a proliferative state. It is thought to achieve this through an inhibitory effect on the transcription of the MRFs *MyoD* and *myogenin* (Schmidt et al., 2003). In the chicken, *Sox8* is expressed throughout development in many tissues including the dermomyotome of differentiated somites (Bell et al., 2000), which is a precursor of the musculature and specifically the muscle precursors that give rise to the hypaxial musculature of the limbs and the epaxial muscles that attach to the spinal cord (Gros et al., 2005). Despite the importance of Sox8 during many aspects of development the *Sox8* knockout mouse model revealed no major phenotypic abnormalities other than idiopathic weight loss (Sock et al., 2001).

Sox8 has been cloned and characterised in mammals (mouse and human), birds (chicken) (Bell et al., 2000; Schepers et al., 2000) and fish *Oncorhynchus mykiss* (rainbow trout) (Ito et al., 1995) and more recently *T. rubripes* (Koopman et al., 2004). However its characterisation and expression pattern, particularly in relation to muscle development, is yet to be investigated in teleosts. Teleosts, as ectotherms, pose an ideal model to investigate the influence environmental factors, such as temperature, have on embryogenesis. Temperature profoundly affects the rate and outcome of embryogenesis (Johnston and Hall 2004). In Atlantic herring for example increasing incubation temperature accelerated the onset of the rostral to caudal progression of myofibril assembly (Johnston et al., 1995) and similar heterochronic shifts were observed in larval stages with growth rate and development of median fins retarded (Cole et al., 2004). It is unknown whether developmental plasticity is an emergent property of many individual effects of temperature at the transcriptional and translational level or as a result of so called plasticity genes that are subject to selection such as heat shock protein 90 (Hsp90) in *Arabidopsis thaliana* (Sangster and Queitsch 2005) and goby fish (*Gillichthys mirabilis* and *Gillichthys seta*) (Dietz and Somero 1992). Summer-acclimatized goby fish had higher levels of *Hsp90* in brain tissues than fish acclimatized to winter conditions. For winter-acclimatized fish, increased synthesis of *Hsp90* was observed when the control temperature was raised from 18°C to 28°C (Dietz and Somero 1992). We have shown *myogenin* (chapter 4) to be a potential candidate as a plasticity gene in *T. rubripes* with relative expression levels higher at elevated embryonic incubation temperatures. We hypothesised *Sox8* to be a further plasticity gene candidate as Sox8 protein interferes with the transcription of *MyoD* and *myogenin*, thus inhibiting myogenic differentiation (Schmidt et al., 2003). The objectives of this study were to clone the puffer fish orthologue of *Sox8* and characterise its expression in response to changes in incubation

temperature during *T. rubripes* embryogenesis. In addition, *Sox8* was investigated as a potential marker for MPCs in teleosts as previously shown in mammals.

5. 3 Results

5. 3 *i* Cloning *Sox8* in two puffer fish species

The *Sox8* gene was amplified by RT-PCR and cloned in *T. rubripes* and *T. nigroviridis.* It was located on scaffold 875 and chromosome 3 of the respective genome assemblies (Ensembl). The cDNA sequences have been submitted to the nucleotide genbank database under the accession numbers AY935980 and AY612092 respectively. Putative translation of the cDNA gave a 462-residue peptide in both *T. rubripes* (Q5BM60) and *T. nigroviridis* (Q61248) containing the characteristic SOX-TCF_HMG-box (79-residues). Multiple sequence alignment of the proteins confirmed their identity as Sox8 (Fig 5.1a) on the basis of their homology with the human, mouse, chick and trout proteins particularly in the conserved SOX-TCF_HMG-box motif. The *T. rubripes* Sox8 protein shares 64% identity with its human orthologue (P5703) compared to 95% identity with *T. nigroviridis.* More specifically the 79-residue *T. rubripes* conserved domain shares a 99%, 96%, 96%, 95% and 95% identity with homologous proteins in *T. nigroviridis*, human, mouse (Q04886) chick (P5704) and rainbow trout (Q91216) respectively (Fig 5.1a). There are single conservative amino acid changes in the HMG-box domains between puffer fish and mammals and this size of this domain remains conserved. Sox8 is very closely related to Sox9 and Sox10 and the *T. rubripes* paralogues of Sox9a (Q8UWL6) and 10a (Q6WNS7) have a high degree of conservation with Sox8 particularly in the conserved region, 50% and 48% respectively (Fig 5.1b). Sox8 is shorter than both Sox9 and Sox10. A 34-residue region N-terminal to the HMG box (EDDERFPACIRDAVSQVLKGYDWSLVPMPVRGNG) is conserved between all species studied starting at position 75. C-terminal to the HMG-box motif lies a second conserved 20-residue hydrophilic region (HAGQPHGPPTPPTTPKTDLH). These two

regions flanking the HMG-box are also conserved among the other Sox group E proteins (Fig 5.1b). A 50-residue sequence C-terminal to the 20-residue hydrophilic region of Sox8 (Fig 5.1a) is highly conserved and a 36 amino acid portion of this is conserved in other group E proteins (Fig 5.1b). Outwith the conserved region of Sox8 there are some differences specific to birds, mammals or teleosts. For example, a 9-residue region (PAPRPPGAA) of chicken Sox8 starting at position 51 is not present in teleosts or mammals (Fig 5.1a).

5. 3 *ii* Genomic organisation of *Sox8* in puffer fish

The genomic region of *Sox8* was amplified and cloned in *T. rubripes* and found *in silico* in *T. nigroviridis* using the *T. nigroviridis* genome assembly (Ensembl). Genomic DNA was aligned with the cDNA sequence revealing the exon intron boundaries. The *T. rubripes* and *T. nigroviridis Sox8* genes are 1760 bp and 1794 bp, respectively, containing three exons of identical size between the two species (416 bp, 236 bp and 737 bp) and 2 introns that are slightly larger in *T. nigroviridis* (Fig 5.2). The exon sizes between human and *T. rubripes* are very similar although the third exon is more than 50 bp longer in *T. rubripes.* The exon/ intron splice sites remain conserved between species and also between all group E genes in *T. rubripes* (indicated by black arrow heads in Fig 5.1a, 5.1b). The most striking difference is in the size of introns one and two, which are 5-fold and 2-fold larger respectively in human when compared to the puffer fish *Sox8* genes.

5. 3 *iii* Sox8 Synteny

Comparative mapping of the genes adjacent to *Sox8* in *T. rubripes*, *T. nigroviridis* and human revealed a large region of conserved synteny (130 kb) between the two puffer fishes (Fig

5.3). This region is found on chromosome 3 in *T. nigroviridis* and corresponding genes are found on chromosomes 7, 16 and 17 in human. Upstream of *Sox8* lie the genes CENTA1 (Centaurin-alpha 1), CYP2W1 (Cytochrome P450 2W1), ABCA3 (ATP-binding cassette sub-family A member 3) and CCNF (G2/ mitotic-specific cyclin F). Downstream of *Sox8* lie the genes RASD1 (*Dexamethasone-induced RAS-related protein*) and UBP22 (*Ubiquitin carboxyl-terminal hydrolase 22*). The gene order is conserved between the puffer fishes, but two upstream (ABCA3 and CYP2W1) and two downstream (RASD1 and USP22) genes are located on chromosomes 7 and 17 in human respectively.

5.3 *iv Sox8* expression in *T. rubripes* adult tissues

A comparative real-time PCR assay was performed to assess the relative expression of *Sox8* in adult tissues. Transcript levels were 10-fold greater in brain than in fast muscle, and *Sox8* was also found to be expressed in both fast and slow skeletal muscle, skin and testes (Fig 5.4a). Minimal expression levels were observed in other tissues including heart, liver, kidney and spleen. The cycle threshold (Ct) values obtained using primers specific for the internal standards were consistent between all samples.

Sox8 expression was localised in fast muscle tissue sections using *in situ* hybridisation. cRNA probes were ~750 bp long and designed out with the conserved domain to prevent the possibility of identifying cross expression from Sox9 and Sox10. Conservation between Sox group E transcripts at the nucleotide level is low. *Sox8* transcripts were localised to mononuclear cells corresponding to potential MPCs. The sense control sections were unstained (Fig 5.4b).

5.3 *v* Developmental expression of *Sox8* in *T. rubripes*

Sox8 expression was investigated throughout *T. rubripes* development using *in situ* hybridisation on embryos reared at 18°C. *Sox8* transcripts were first detected in the developing notochord at 60 hpf (Fig 5.5B). Strong staining was also evident anterior to the most rostral somites possibly corresponding to the rhombomeres at most stages during segmentation (Fig 5.5C, H), which was not in the sense control (Fig 5.5G). Expression in the eye ceased and became stronger in the mid and hind-brain by 79 hpf (Fig 5.5H), returning faintly in the lateral portion of the more rostral somites at 100 hpf (Fig 5.5C) by which time somitogenesis was nearing completion. Also at 100 hpf expression was observed in the medial periphery of the eye (Fig 5.5C). Prior to hatch (162 hpf) strong expression was observed as 2 parallel stripes in the pectoral fin bud (Fig 5.5D). Expression in the brain had faded considerably by 162 hpf and appeared as faint ubiquitous expression in the head region. Post-hatch (195 hpf) the expression was most evident uniformly throughout the developing muscle of the pectoral fin buds (Fig 5.5F) and in the hind brain and otoliths (Fig 5.5E).

5.3 *vi* Effects of Temperature on *Sox8* expression

Relative *Sox8* expression during development was investigated using qPCR at 3 different incubation temperatures (15°C, 18°C and 21°C) (Fig 5.6). *Sox8* expression varied significantly throughout development at 15°C ($p<0.001$), 18°C ($p<0.001$) and 21°C ($p<0.001$) (Fig 5.6). The initial up-regulation of *Sox8* transcripts occurred during early somitogenesis (>50 hpf at 18°C). This correlates with the *in situ* hybridisation data showing expression in the notochord at 60 hpf in embryos incubated at 18°C. *Sox8* expression then follows a sharp increase in relative expression until late somitogenesis after which it begins to plateau.

During this sharp increase in expression *Sox8* transcripts are located in the brain, eyes, rhombomeres and rostral somites. Normalising the data points using somite interval (see 2.3 *iii*) revealed small differences in relative expression (Fig 5.7), in particular during early somitogenesis relative expression appeared highest at 15°C (Fig 5.7), however no significant differences were identified between temperature groups.

5.3 *vii* Comparing *Sox8* and *myogenin* exression during *T. rubripes* development.

Since Sox8 has been identified as a potential inhibitor of *myogenin* promoter activity (Schmidt et al., 2003) we investigated the relative expression of *Sox8* and *myogenin* at the same developmental temperature. 21°C was chosen because the peak of *myogenin* expression during somitogenesis (40 hpf) was most prominent at this temperature (see chapter 4). The expression of the two transcripts appeared to be somewhat reciprocal with *myogenin* expression decreasing dramatically at the onset of *Sox8* up-regulation (Fig 5.8). This correlates with the *in situ* hybridisation results. *Sox8* expression appears in the rostral somites during late somitogenesis (Fig 5.5C) by which time *myogenin* transcripts are no longer present in the rostral somites but confined to the caudal somites (chapter 4). Contrary to this however both transcripts are expressed simultaneously in the fin buds at 162 hpf (Fig 4.5, 5.5).

5.4 Discussion

In this study the gene corresponding to the Sox group E protein *Sox8*, has been described in two puffer fish species, *T. rubripes* and *T. nigroviridis*. The genomic organisation, including exon splice sites, is conserved between the 2 species. This is also true when comparing *Sox8* with the other group E genes in *T. rubripes* (Koopman et al., 2004) and has previously been shown in mouse and human suggesting a possible common evolutionary ancestor to the group E Sox genes (Schepers et al., 2000). Two regions flanking the HMG domain are conserved in Sox8 between vertebrate species and also Sox group E proteins and may be responsible for a number of additional DNA binding and protein folding and interacting properties. This high conservation in multiple functional regions of Sox group E proteins highlights the fact they possess some functional redundancy and may be able to substitute for one another (Wegner 1999; Sock et al., 2001; Stolt et al., 2004b). *Sox8* studies in mouse (Schepers et al., 2000) and chick (Bell et al., 2000) have shown expression in many different cell types that support our results. The expression profile of *Sox8* during *T. rubripes* development suggests important roles in the CNS (central nervous system), early muscle differentiation and fin development. In adult *T. rubripes Sox8* transcripts are highest in the brain, testes and skeletal muscle. This widespread expression and functional redundancy with other Sox proteins has made it very difficult to determine Sox8 function in development and growth. For this study we have focused on the role of Sox8 in muscle growth and development.

The *Sox8* transcript is located in a large segment of conserved synteny (~130 kb) between puffer fish species. This region contains a number of genes with different functions. Upstream of *Sox8* in puffer fish and human lies the gene coding for cyclin F (CCNF) which is

involved in control of the cell cycle during the S phase and G2 (Bai et al., 1994). *Sox8* has been shown to have an inhibitory effect on myoblast differentiation by holding cultured C2C12 cells in a proliferative state (Schmidt et al., 2003). *Sox8* and CCNF may have similar transcriptional regulators being located in the same region of conserved synteny and both being involved in the progression of the cell cycle. The gene with closest proximity to *Sox8* in all species studied is CENA1 and codes for a putative MAPK-activating protein (Centaurin-alpha-1) (Tanaka et al., 1997; Venkateswarlu et al., 2004) that binds phosphatidylinositol 3,4,5-trisphosphate (PtdInsP3) and inositol 1,3,4,5-tetrakisphosphate (InsP4) (Hammonds-Odie et al., 1996). It is highly expressed in brain and at lower levels in peripheral blood leukocytes and is phosphorylated by protein kinase C (PKC) (Hammonds-Odie et al., 1996; Zemlickova et al., 2003). Other genes located in the vicinity of *Sox8* in puffer fish species include; CYP2W1 whose expression or possible activities are still currently unknown (Guengerich et al., 2005), ABCA3 which has only been identified in mammals and is expressed in mammalian brain, skeletal muscle and heart tissue (Shulenin et al., 2004), RASPD1 which codes for Dexamethasone-induced Ras-related protein 1 and is expressed most strongly in heart tissue but also in skeletal muscle. Interestingly it is proposed to have a role in the alterations of cell morphology, growth and cell-extracellular matrix interactions (Tu and Wu 1999) and UBP22 which codes for the protein Ubiquitin carboxyl-terminal hydrolase 22 which was cloned from human brain tissue but its function is as yet unknown (Kikuno et al., 1999). Most of these genes are yet to be identified in teleosts and as their function and tissue expression in mammalian systems suggests they may have important functions within teleost skeletal muscle and brain. The lack of conservation between puffer fish and human indicates it is unlikely the genes described share function or transcriptional regulation with Sox8.

Previously we identified *myogenin* transcripts in developing somites following a rostral to caudal gradient of expression (chapter 4). *Myogenin* expression was no longer present in the rostral somites of the embryos showing *Sox8* expression. *Myogenin*, like *Sox8*, transcripts showed a second wave of expression in the muscle of the fin buds during the high pec stage of the pharyngula stage of development. In addition we have compared the relative expression of both transcripts and identified a reciprocal pattern of expression with *myogenin* transcripts decreasing on queue with the up-regulation of *Sox8* during late somitogenesis. Schmidt et al (2003) identified Sox8 as an inhibitor of muscle differentiation by interfering with the transcription of the MRFs *MyoD* and *myogenin*. In particular, it reduced their expression and was proposed to interact with the promoter region of *myogenin*. Although *Sox8* expression is most prominent in the brain of *T. rubripes* the reciprocal expression pattern with *myogenin* suggests a similar mechanism controlling myogenic differentiation in teleosts. There is however simultaneous expression of *Sox8* and *myogenin* in the fin buds at 162 hpf. For further confirmation of this observation it would be beneficial to perform double *in situ* hybridisation using cRNA probes for both *myogenin* and *Sox8*. Analysis of tissue sections from double stained embryos would identify whether they are expressed in overlapping regions of the fin bud or in separate compartments.

We identified relative expression of *myogenin* during somitogenesis to increase in response to elevated incubation temperatures during early somitogenesis. In light of a possible interaction and opposing effect of Sox8 and myogenin during muscle differentiation it seemed feasible to suggest *Sox8* expression may possess a similar response to temperature. Despite the evident acceleration of development at higher temperatures, shown by the initial up-regulation of *Sox8* at 21°C occurring ~20 hours prior to that observed at 18°C, we failed to detect any significant differences in expression of *Sox8* at equivalent stages of

development between three temperature groups, 15°C, 18°C and 21°C. It was therefore concluded *Sox8* is not a candidate as a plasticity gene.

In situ hybridisation on *T. rubripes* fast muscle tissue sections localised *Sox8* transcripts in discrete zones along the periphery of potential hypertrophic muscle fibres. These regions correspond to putative areas of MPCs. Sox8 has been identified as a possible marker for myosatellite cells in mice with expression in adult muscle fibres confined to muscle satellite cells (Schmidt et al., 2003). Further to this *Sox8* is expressed in the dermomyotome of chick somites (Bell et al., 2000) which gives rise to the muscle progenitor cells of limb and facial muscles in mammals (Gros et al., 2005). A similar structure with the same properties as the dermomyotome has recently been identified in zebrafish (Devoto et al., 2006) and we have shown Sox8 expression in *T. rubripes* somites and developing limb muscle. This suggests Sox8 may be of use as a marker for MPCs in other vertebrates including teleosts. Further functional work to define the specific role of Sox8 in MPC function may be hindered due to the functional redundancy of group E sox proteins (Wegner 1999; Sock et al., 2001; Koopman et al., 2004; Stolt et al., 2004a). The *Sox8* knock out mouse revealed no major phenotypic change which was thought to be due to compensation by *Sox9* and *Sox10* (Sock et al., 2001). This redundancy may dilute the affect any over expression, knock down or knock out models, for example in cell lines, used for functional analysis may have. Identification of genes such as *Sox8* as specific MPC markers in teleosts however opens up the possibility to identify changes in MPC number and location in the myotome of fish subjected to different factors such as environmental temperature, exercise and disease.

5. 5 Figures

Figure 5. 1a Sox8 Protein Alignment

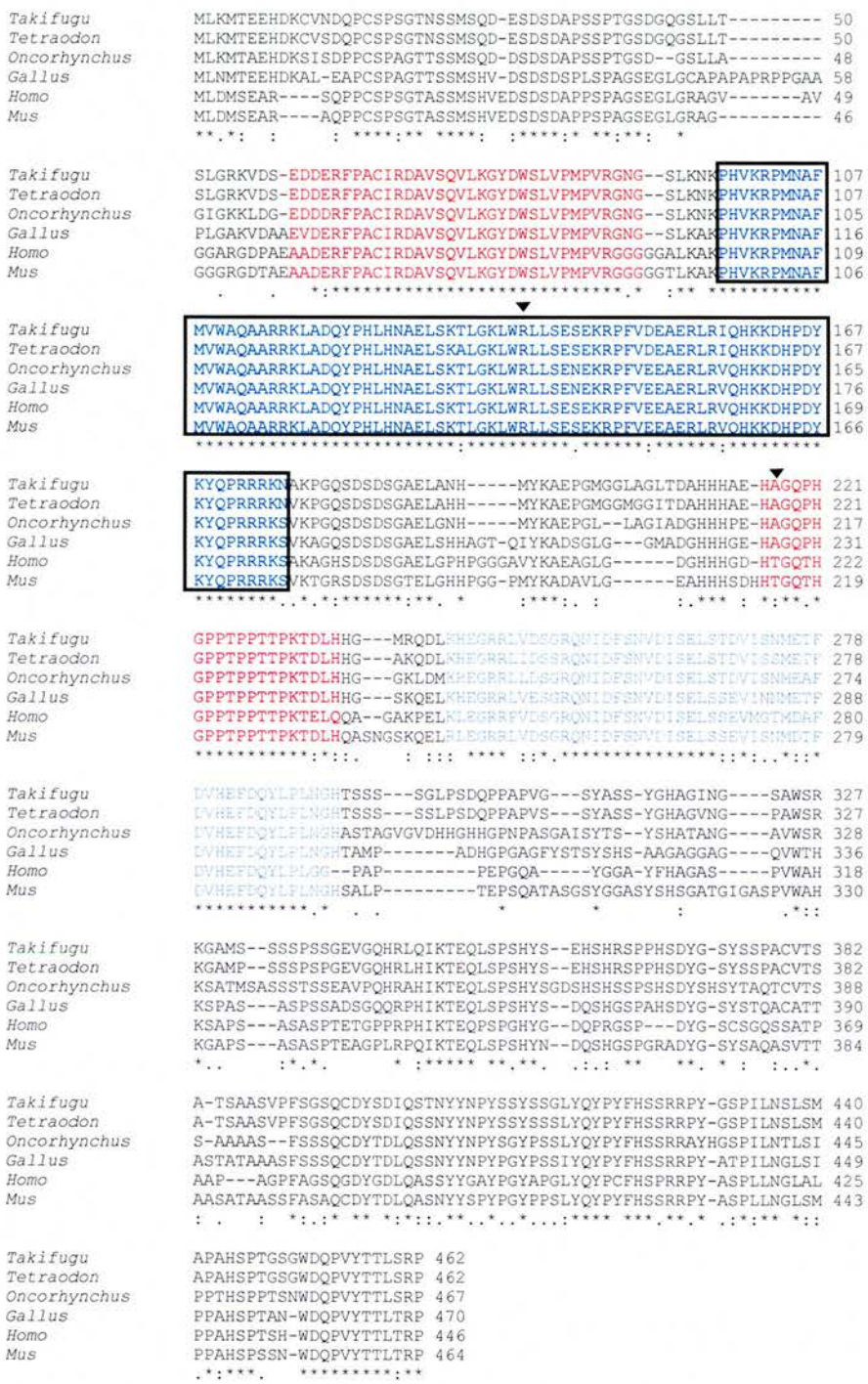


Figure 5. 1a Multiple sequence alignment of Sox8 protein sequences. The inferred *T. rubripes* (*Takifugu*) sequence (Q5BM60) is compared with those of *T. nigroviridis* (*Tetraodon*) (Q61248), trout (Q91216), chick (P5704), human (P5703) and mouse (Q04886). Conserved substitutions are marked with a colon, semi-conserved substitutions with a dot and global conservation is represented with an asterisk. The conserved 79-residue SOX-TCF_HMG-box motif (blue) is highlighted with a box. Two regions flanking the functional domain with a high degree of conservation are highlighted in red. A 50-residue region conserved among all Sox8 orthologues studied and less conserved in other Sox group E proteins has been highlighted in grey. Exon/ intron boundaries are marked with black arrow heads.

Figure 5. 1b Sox group E Protein Alignment

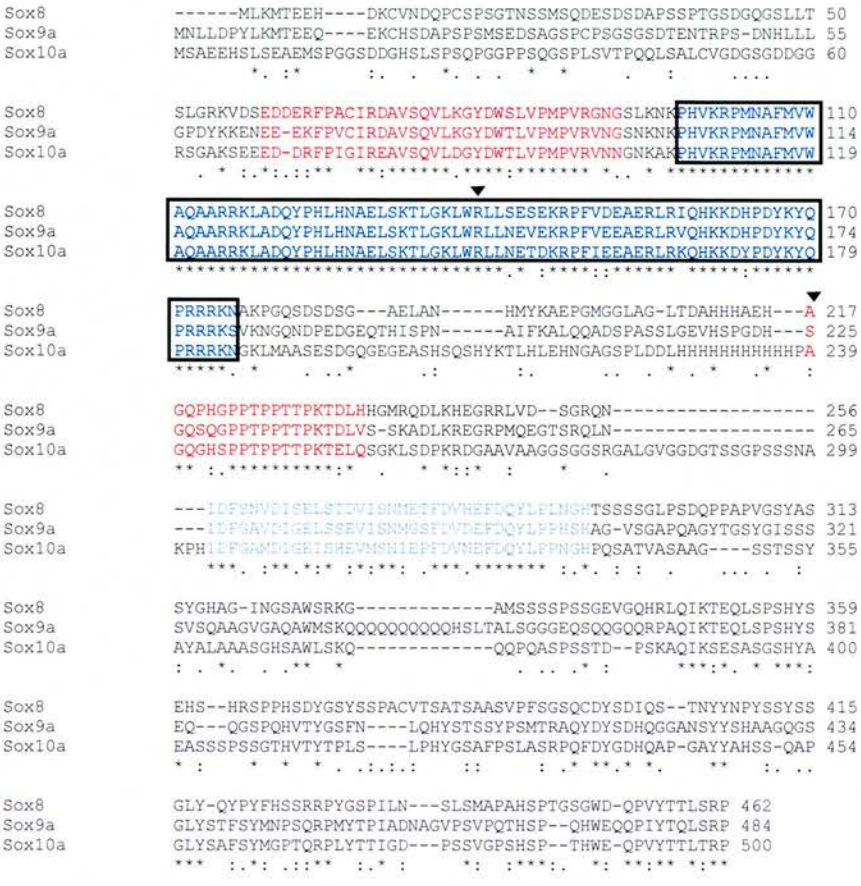


Figure 5. 1b Multiple sequence alignment of Sox group E protein sequences. The inferred *T. rubripes* Sox8 sequence (Q5BM60) is compared with those of *T. rubripes* Sox9a (Q8UWL6) and Sox10a (Q6WNS7). Conserved substitutions are marked with a colon, semi-conserved substitutions with a dot and global conservation is represented with an asterisk. The conserved 79-residue SOX-TCF_HMG-box motif (blue) is highlighted with a box. Two regions flanking the functional motif with a high degree of conservation are highlighted in red. A further conserved sequence is highlighted in grey. Exon/ intron boundaries are marked with black arrow heads.

Figure 5. 2 *Sox8* exon/ intron structure

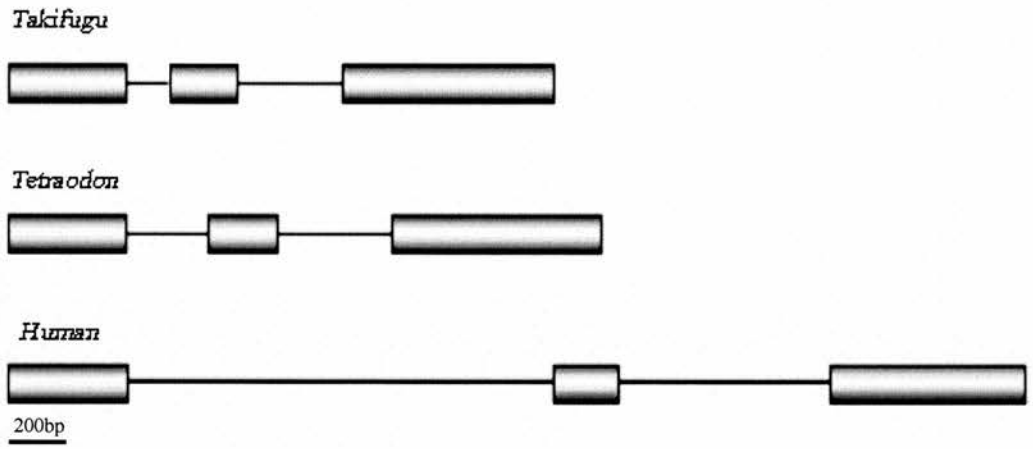


Figure 5. 2 Transcript structure of the *Sox8* gene from *T. rubripes*, *T. nigroviridis* and human. Bars represent exons and lines represent introns. The scale bar indicates 200 bp. Exon sizes are comparable between species and the introns are 5 and 2-fold larger in human *Sox8*.

Figure 5. 3 *Sox8* Synteny

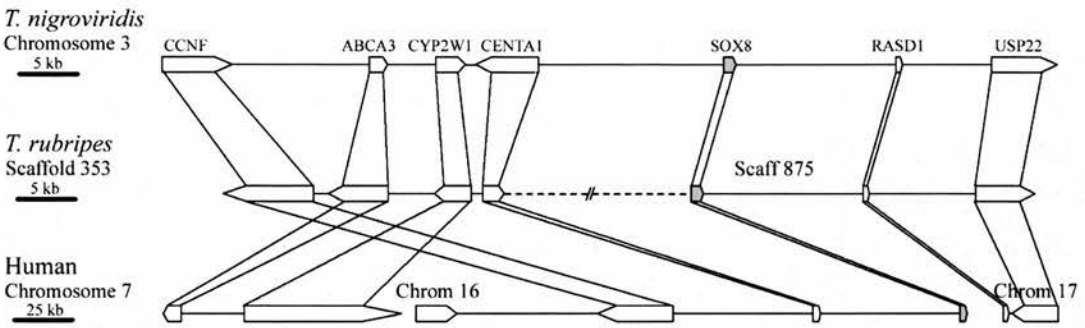


Figure 5 .3 Partial synteny map of the genomic neighbourhood surrounding *Sox8* in *T. rubripes*, *T. nigroviridis* and human. Genes are represented with block arrows indicating the direction of the genes in the respective genome. *Sox8* is highlighted yellow and scale bars are indicated beneath the species name.

Figure 5. 4 Expression analysis, qPCR

Figure 5. 4a Relative tissue expression of *Sox8* in adult *T. rubripes*

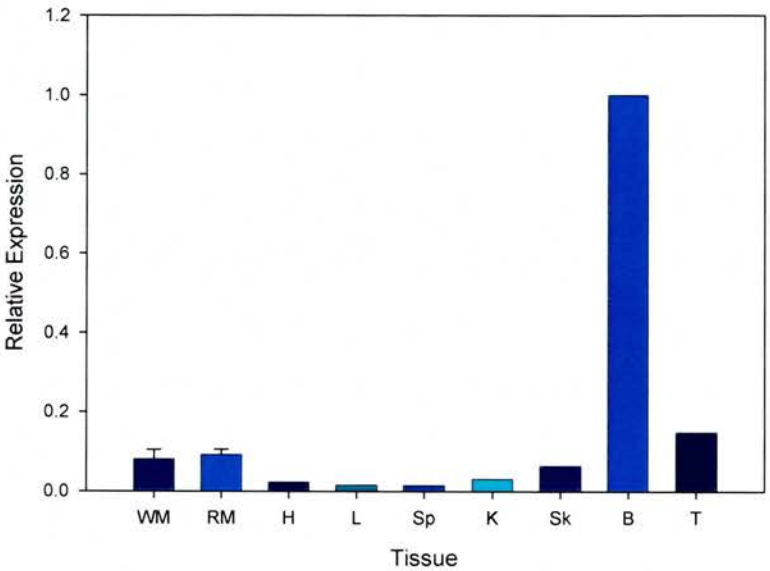


Figure 5. 4b *Sox8* expression in *T. rubripes* fast muscle tissue section

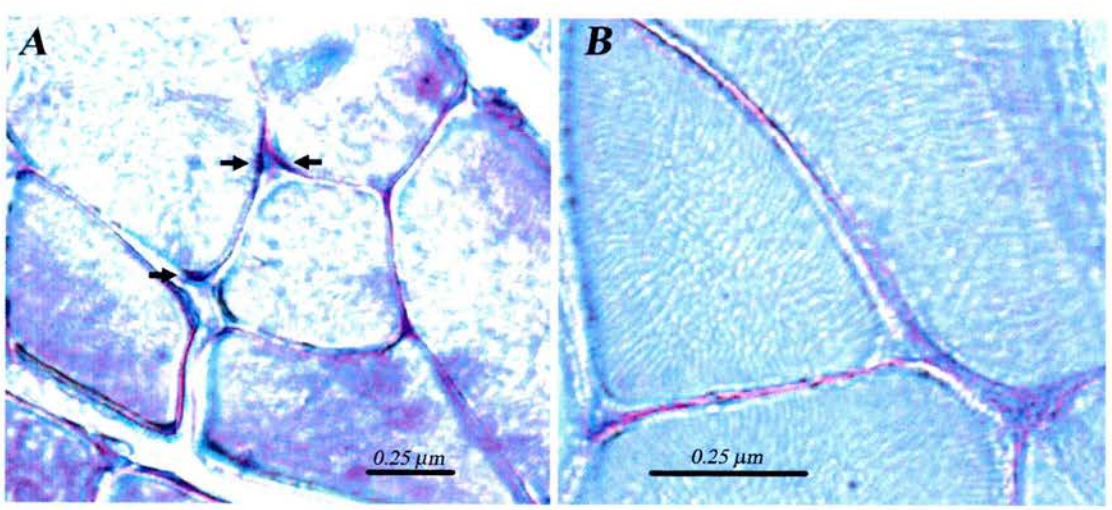


Figure 5. 4. Sox8 expression in *T. rubripes* tissues as determined by qPCR.

(a): Expression in different tissues determined by qPCR; WM (fast skeletal muscle), RM (slow skeletal muscle), H (heart), L (liver), Sp (spleen), K (Kidney), Sk (skin), B (brain) and T (testes). 18s rRNA was used as an endogenous control and relative expression was normalised against the highest expression value (Brain). N=7 for fast and slow skeletal muscle, n=2 for other tissues where means are plotted, values did not vary by > 25 %.

(b): Localisation of *Sox8* transcripts in *T. rubripes* fast muscle tissue section using antisense (A) and sense (B) riboprobes specific for *Sox8*. Some areas of potential MPCs are indicated with black arrows. The scale bars represent 0.25 μm.

Figure 5. 5 *Sox8* expression throughout *T. rubripes* development using *in situ* hybridisation

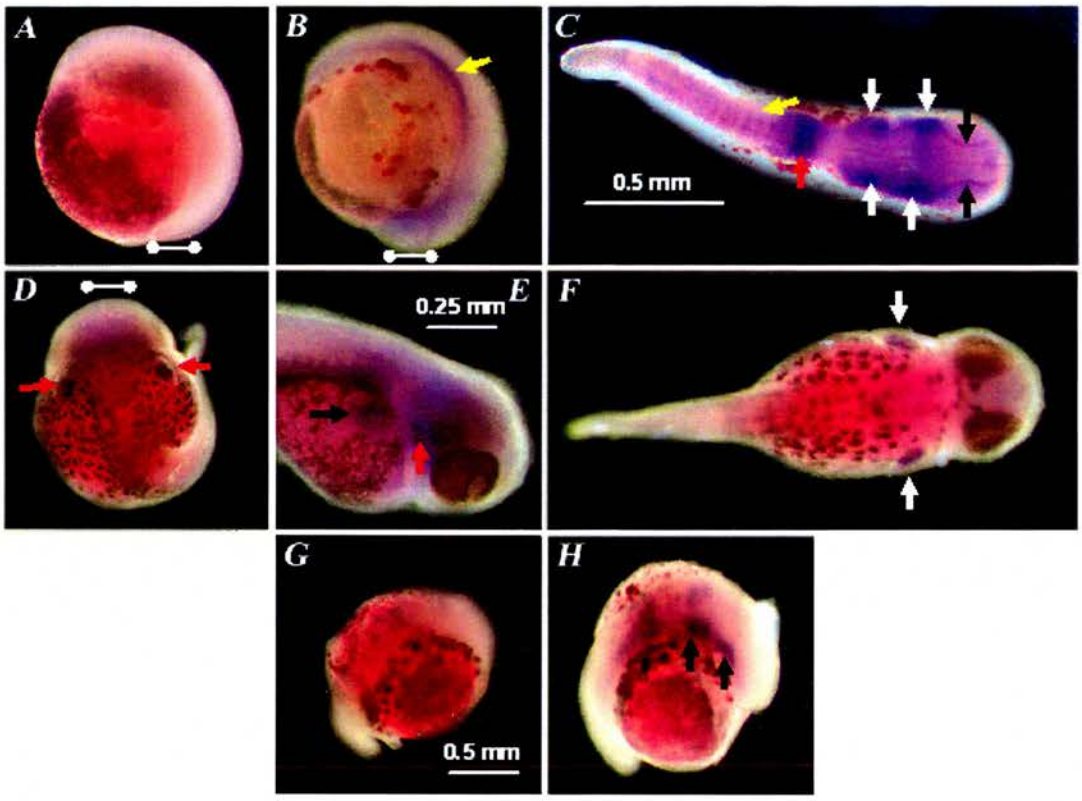


Figure 5. 5. Localisation of *Sox8* transcripts in *T. rubripes* embryos throughout development at 18°C by whole mount *in situ* hybridisation with a cRNA probe: (A) 42 hpf, bud stage of gastrulation, (B) 60 hpf, early somitogenesis, yellow arrow indicates expression in the notochord, (C) 100 hpf, 25 somite stage, somite expression is indicated with a yellow arrow, rhombomere expression with a red arrow, expression in the lateral hind and mid brain is indicated with white arrows and black arrows highlight expression in the periphery of the developing eye, (D) 162 hpf, high pec stage of the pharyngula period, red arrows indicate staining in the developing muscle of the fin buds, (E + F) hatch stage, black and red arrows in 'E' indicate expression in the fin primordia muscle and otoliths respectively. Ubiquitous expression in the brain is also noticed at this stage. White arrows in 'F' indicate expression in the muscle of the fin buds, (G) 79 hpf, 15 somite stage, sense control. (H) 79 hpf, 15 somite stage, expression in mid and hind brain is indicated with black arrows. White dumbbells indicate the anterior of the embryo in A, B and D, in all others anterior is right. Scale bar in C shows 0.5 mm for A, B, C, D and F. Scale bar in E is for 0.25 mm and the scale bar in G for 0.5 mm is for G and H.

Figure 5. 6 Embryonic expression of *Sox8* with respect to developmental temperature.

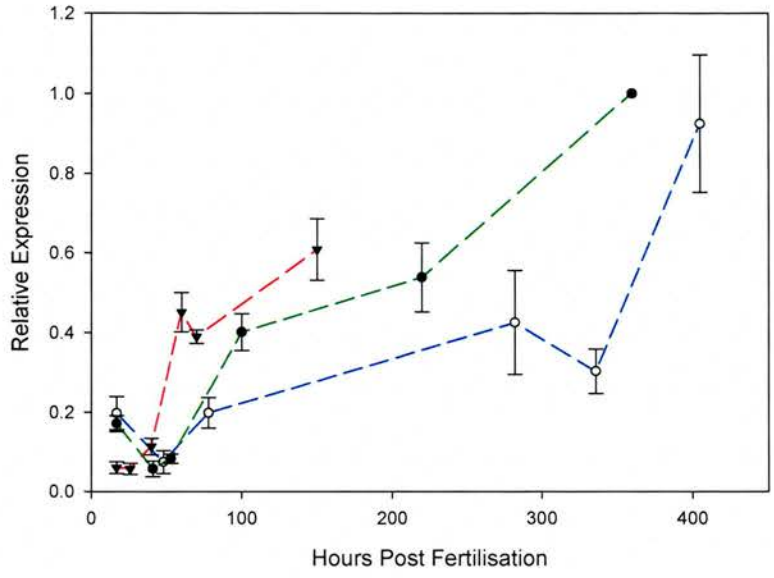


Figure 5. 6 Relative *Sox8* expression illustrating changes in transcript number during *T. rubripes* development at 3 temperatures; 15°C (blue), 18°C (green), 21°C (red) with respect to developmental time. Subjecting data to a general linear model confirmed that relative expression varied significantly throughout development at the individual temperatures (15°C p=<0.001, 18°C p=<0.001, 21°C p=<0.001).

Figure 5. 7 Relative expression of Sox8 with developmental time normalised to somite interval

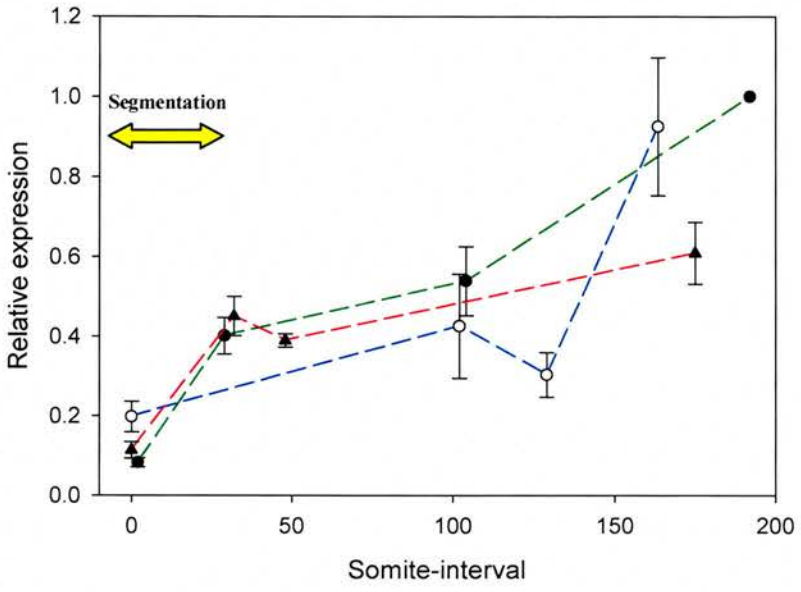


Figure 5. 7 Relative *Sox8* expression illustrating changes in transcript number during *T. rubripes* development at 3 temperatures; 15°C (blue), 18°C (green), 21°C (red) with respect to developmental time normalized somite interval. Post hoc analysis revealed no significant change in expression with respect to temperature. RNA polymerase II was used as the endogenous internal control and all points (n=5) were normalised to the highest expression value (360 hpf, 18°C).

Figure 5. 8 Relative expression of *Sox8* and *myogenin*.

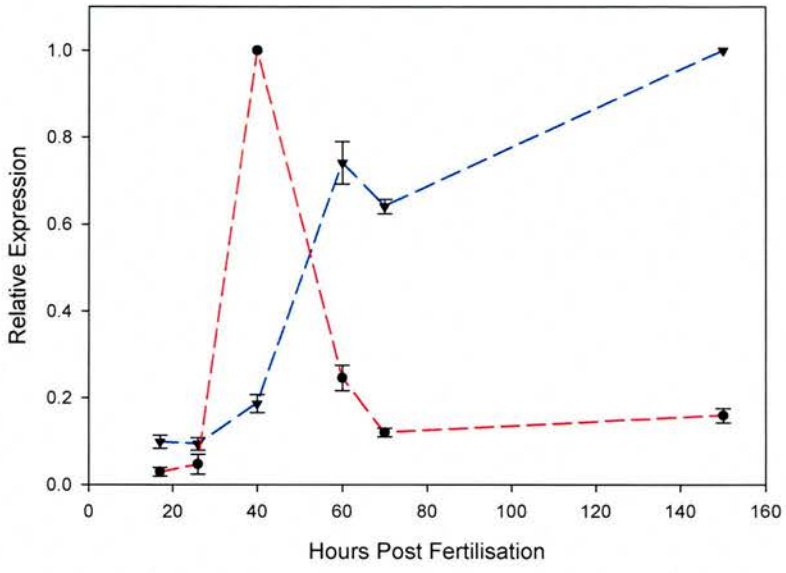


Figure 5. 8 Relative expression of *Sox8* (blue) and *myogenin* (red) during development in embryos incubated at 21°C. *Myogenin* expression decreases sharply as *Sox8* expression increases.

Chapter 6: Follistatin and two paralogues of Myostatin in *T. rubripes.*

6. 1 Abstract

Follistatin is known to antagonise the function of various members of the TGF-β family of secreted signalling factors including the master inhibitor of muscle growth, myostatin (Mstn). In the present study follistatin and two Mstn paralogues (FMstn-1 and FMstn-2) have been cloned and characterised in the puffer fish species *T. rubripes*. *FMstn-1* codes for a 376-residue putative protein and is an orthologue of zebrafish Mstn-1 (ZfMstn-1) and human myostatin sharing 76 and 60% identity respectively. *FMstn-2* codes for a 359-residue putative protein that is a paralogue of *FMstn-1* and similar to zebrafish *Mstn-2* (*ZfMstn-2*). *T. rubripes* follistatin shares a 75% identity with the human orthologue and is a 320-residue putative protein. *Follistatin* expression was wide spread in adult tissues, most specifically in skeletal muscle and brain. *FMstn-1* expression was detected in a number of adult tissues including a 2.5-fold greater expression in fast muscle compared to slow muscle. *FMstn-2* transcripts were highly expressed in adult brain tissue but not detectable in embryonic stages. During development *FMstn-1* transcripts were detected strongest during late segmentation and hatch. *Follistatin* transcripts were detected earlier than *FMstn-1* at the late gastrula stage of development. *In situ* hybridisation at this stage revealed expression in the pre-somitic mesoderm. Beyond this stage *follistatin* transcripts were localised to the developing brain and the lateral portion of the somites following a rostal to caudal progression. *FMstn-1* and *FMstn-2* expression could not be detected by *in situ* hybridisation. In addition we found developmental temperature to have a profound effect on the relative expression of *FMstn-1* and *follistatin* at the start of segmentation. *Follistatin* expression was ~2-fold and ~2.5-fold greater at 15°C compared to 21°C and 18°C respectively during early somitogenesis. At the

same developmental stage *FMstn-1* showed a similar response to temperature with a 1.5-fold and 2-fold greater expression at 15°C compared to 21°C and 18°C respectively. These changes in relative expression of *FMstn-1* and *follistatin* suggest they may have a role in creating the different phenotypes previously observed in teleost larvae developed at different incubation temperatures.

6. 2 Introduction

Myostatin is a member of the transforming growth factor β (TGF- β) super family and the most powerful inhibitor of muscle growth found to date in mammals (McPherron and Lee 1997). Amthor et al. (2004) showed *follistatin* and *myostatin* to be very closely expressed in overlapping or in very closely located domains of developing muscle and to interact directly with high affinity and proposed follistatin to be an antagonist of myostatins ability to inhibit myogenesis. Studies of myostatin in various fish species have identified two orthologues (Ostbye et al., 2001; Rescan et al., 2001; Kerr et al., 2005).

Myostatin-null mice (*myostatin* -/-) show a massive increase in both hyperplastic and hypertrophic muscle growth (McPherron et al., 1997). Similarly over expression of *follistatin* induces a phenotypic response akin to *myostatin* knock-out mice whereas a *follistatin* -/- mutant displays muscle deficiency (Matzuk et al., 1995; Lee and McPherron 2001). Myostatin is thought to function by up-regulating the cdk inhibitor p21 and decreasing the activity of cdk2, arresting cells in the G1 phase of the cell cycle (Thomas et al., 2000a). Follistatin antagonises myostatin function through direct protein interaction (Amthor et al., 2004). Further evidence for the role of follistatin in mammalian muscle development is its interactions with the bone morphogenetic proteins (BMPs) particularly BMP-7 (Amthor et al., 2002). Follistatin enhances BMP-7 action to induce muscle growth but prevents its ability to induce the apoptosis of MPCs (Amthor et al., 2002). Follistatin has also been proposed to promote *Pax-3* expression delaying muscle differentiation and holding cells in a proliferative state (Amthor et al., 1998). The mammalian *follistatin* transcript exists as two splice variants, one of which translates to a carboxy truncated follistatin (Shimasaki et al., 1988), and myostatin has 2 paralogues in teleosts (Mstn-1 and Mstn-2) as described in zebrafish (Kerr et

al., 2005) and salmonids (Ostbye et al., 2001; Rescan et al., 2001). Expression of mammalian *myostatin* is primarily restricted to skeletal muscle (McPherron et al., 1997) however the teleost *Mstn-1* orthologue shows ubiquitous expression (Ostbye et al., 2001; Roberts and Goetz 2001; Vianello et al., 2003). Salmonid Mstn-2 appears to be more specific to muscle growth and development (Rescan et al., 2001) and this has also been suggested for ZfMstn-2 with expression limited to early somitogenesis (Kerr et al., 2005).

Embryonic temperature has been shown to influence fibre number and size in teleosts (Johnston and Hall 2004). The double muscling (DM) phenotype observed in some breeds of cattle with a myostatin -/- mutation is associated with a massive increase fibre number (Wegner 1999). Further evidence relating myostatin to DM has recently been shown in murine systems with the use of a mutant myostatin allele ($Mstn^{Cmpt\text{-}dll Abc}$) introgressed into a high growth mouse cell line resulting in a 24% increase in muscle fibre number with no significant change to muscle mass. Interestingly a decrease in slow muscle fibre number was also observed (Rehfeldt et al., 2005). Teleosts also show an increase in fibre number in transgenic zebrafish expressing the myostatin pre-domain (Wargelius et al., 1999; Acosta et al., 2005). We hypothesised that the change in fibre number and size observed in teleosts developed at different incubation temperatures may be due to changes in levels of myostatin similar to those observed in myostatin mutants. This would suggest *myostatin* is a potential plasticity gene whose transcription is directly influenced by temperature. The aims of the present study were to clone and characterise both myostatin paralogues and the myostatin inhibitor follistatin in the puffer fish and investigate their potential as plasticity genes.

6.3 Results

Using cDNA from *T. rubripes* larvae as a template the full length cDNAs of two paralogues of the Transforming Growth Factor-β (TGF- β) *myostatin* have been cloned and characterised in *T. rubripes* and submitted to the genbank database under the accession numbers AY445322 and AY445321 respectively. The cDNA coding for follistatin a known antagonist of myostatin (Amthor et al., 2004) has also been cloned and characterised (GenBank DQ288127).

6.3 *i T. rubripes* Myostatin 1 and 2

The two *T. rubripes myostatin* paralogues (*FMstn-1* and *FMstn-2*) code for peptides of 376 (Q6SYV4) and 359 (Q6SYV3) residues sharing 66% and 62% identity at the nucleotide and peptide level respectively. The highest identity (83%) is found at the C-terminal of the protein following the proteolytic breakdown site (residues 264-267 of FMsnt-1, RCCR) corresponding to the functional TGF-β domain. The N-terminal latency-associated peptide (LAP) domain upstream of the proteolytic breakdown site was where most of the sequence differences between orthologues were found (Fig 6.1a). A 5-residue sequence (DYMYL) found in the TGF-β domain of ZfMstn-1 (zebrafish Myostatin 1, O42222) and not in ZfMstn-2 (zebrafish myostatin 2, Q5MNY3) is conserved in FMstn-1 and all other sequences studied, however poorly conserved in FMstn-2. A 6-residue motif (GDDSKD) also found in the LAP domain of ZfMstn-1 and not ZfMstn-2 is only found in FMstn-1 (residues 100-105) but not present in FMstn-2 (Fig 6.1b). Two myostatin paralogues have been found in *Oncorhynchus mykiss* (rainbow trout) (Q90ZD2, Q90ZD1) (Rescan et al., 2001) however they are different to the *T. rubripes* paralogues presented here. The myostatin phylogeny carried out by Kerr et

al (2005) is shown in figure 6.6. The functional domain of FMstn-2 shares an 86% identity with the functional domain of both trout myostatins and the same domain in FMstn-1 shares a 96% identity with both the trout myostatins. The human myostatin protein (O14793) shares a 60% and 58% identity with FMstn-1 and FMstn-2 respectively (Fig 6.1b). ZfMstn-1, like FMstn-1, has a slightly higher identity to human myostatin than its paralogue ZfMstn-2 and shares a 76% identity to FMstn-1 compared to 64% with FMstn-2. ZfMstn-2 and FMstn-2 share a 59% identity.

6.3 *ii* ***T. rubripes*** **Follistatin**

T. rubripes follistatin cDNA codes for a putative protein consisting of 320-residues (ABC00774), with a 75% identity to human follistatin (P19883) and 89%, 85%, 81%, 80%, 75% identity with zebrafish (Q9YHV4), goldfish (Q6R256), *Xenopus* (P31515), chick (Q90844) and mouse (P47931) respectively (Fig 6.1c). The protein contains 3 conserved Kasal domains (amino acids 119-166, 194-241, 272-318) and the two discontinuous activin binding sites previously identified in human (Wang et al., 2000) and chick (Amthor et al., 2004) at the N-terminus (amino acids 3-30, 47-61). The first activin-binding domain previously identified in mammals is poorly conserved in teleosts compared to the second domain.

6.3 *iii* **Genomic organisation**

Both *FMstn-1* and *FMstn-2* consist of 3 exons and orthologues of the *T. rubripes* genes are found on chromosomes 3 and 2 in *T. nigroviridis* respectively. *T. rubripes follistatin* consists of 5 exons (Fig 6.2) and its *T. nigroviridis* orthologue is located on chromosome 12. Exon

size and intron proportions are more similar between *FMstn-1* and human *myostatin* than with *FMstn-2* and human *myostatin*. Exon 3 of *FMstn-2* is ~40 bp bigger than both its paralogue in *T. rubripes* and orthologue in human and its second intron is more than 10-fold larger than intron 1. This is in comparison to intron 2 of *FMstn-1* and human myostatin which is only marginally bigger than the corresponding intron 1 (Fig 6.2).

Structurally the human *follistatin* gene and the *T. rubripes* orthologue are similar (Fig 6.2) with the transcripts sharing 72% identity. The exon size has remained fairly conserved throughout evolution as have exon splice sites. The intron size varies quite considerably with intron 1 of human *follistatin* being 2-fold larger. *Follistatin* splice sites are completely conserved between *T. rubripes* and human, however a splice variant in human *follistatin* contains a sixth exon not present in the *T. rubripes* transcript amplified in this study. The splice sites for exons 3/4 and 4/5 are immediately downstream of the first and second conserved Kasal domains respectively in both human and *T. rubripes*. The stop codon in *T. rubripes follistatin* is just downstream of the third conserved domain, thus the Kasal domains reside in each of the third, fourth and fifth exons.

6.3 *iv* Developmental and tissue expression of *follistatin* in *T. rubripes*.

The spatial and relative expression of *follistatin* was investigated throughout *T. rubripes* development using whole mount *in situ* hybridisation in embryos incubated at 18°C and quantitative real time PCR using embryos reared at 3 temperatures (15, 18 and 21°C). *Follistatin* transcripts were initially detected during mid/late gastrulation (42 hpf) at 18°C (Fig 6.3A) in the presomitic mesoderm lateral to the notochord. During somitogenesis (from ~50 to ~100 hpf) expression was detected in the lateral portion of developing somites, the

mid and hind brain and the periphery of the forming eye (Fig 6.3B, C and D). Expression in the somites progressed in a rostral caudal gradient and by 90 hpf had faded in the rostral somites and was confined to the developing caudal somites (Fig 6.3D). During mid-pharyngula expression was confined to the later portion of the mid and hind brain (Fig 6.3E).

Relative *follistatin* expression from qPCR analysis was plotted against somite interval (as described in chapter 4) at 3 different temperatures (15°C, 18°C and 21°C) (6.4a). A General Linear Model (GLM) with a Bonferroni post hoc test revealed expression to differ significantly throughout development at all temperatures ($p=<0.001$). The general pattern of expression remained the same for *follistatin* at the 3 temperatures with the peak of expression appearing at the start of somitogenesis at all temperatures. Comparing relative expression at an equivalent stage during early somitogenesis using a Mann Whitney U non-parametric test revealed a ~2-fold ($p=0.029$) and ~2.5-fold ($p=0.029$) greater *follistatin* expression at 15°C compared to 21°C and 18°C respectively (Fig 6.4a).

Follistatin expression was also investigated in a selection of *T. rubripes* adult tissues using qPCR (Fig 6.4b). *Follistatin* was expressed most in brain and this was chosen as the tissue to normalise against all tissues for relative expression. Expression was approximately 2-fold less in skeletal muscle (fast and slow) than brain and equivalent between the two skeletal muscle types. *Follistatin* transcripts were also detected in spleen and heart with minimal expression in liver, kidney, skin and testes.

6.3*v* Developmental and tissue expression of *FMstn-1* and *FMstn-2* in *T. rubripes*.

FMstn-2 transcripts could not be detected during development using qPCR and we were unable to localise the expression of either *FMstn* orthologue using *in situ* hybridisation. This problem was shared with other groups who could not achieve localisation of *ZfMstn-1* or *ZfMstn-2* mRNA using conventional *in situ* hybridisation (Xu et al., 2003; Kerr et al., 2005).

Relative expression of *FMstn-1* was investigated during *T. rubripes* embryogenesis with respect to incubation temperature as described for *follistatin* (Fig 6.4a). A significant change in expression was revealed throughout development ($p<0.001$ for all temperatures). Relative expression at an equivalent stage between temperatures during early somitogenesis revealed a 2-fold ($p=0.029$) and 2.5-fold ($p=0.029$) greater expression at 15°C compared to 21°C and 18°C respectively.

FMstn-2 transcripts were amplified from brain but not detected in any other tissue (results not shown). *FMstn-1* was widespread throughout most tissues (Fig 6.4c). *FMstn-1* was expressed highly in brain, with high transcript levels also detected in skeletal muscle, spleen, heart and skin. Expression in kidney was comparable to that of brain however there was a large variation in relative expression in the kidney samples used. Interestingly a Mann Whitney U non-parametric test revealed expression of *FMstn-1* was significantly higher in fast muscle compared to that of slow muscle by a factor of 2.6 ($p = 0.0195$).

6.4 Discussion

In the present study, two full length, putative paralogues of myostatin (*Mstn*) (*FMstn-1* and *FMstn-2*) and one follistatin (*Flstn*) transcript have been cloned and characterised in the tiger puffer fish *T. rubripes*. The genomic organisation of the two *T. rubripes Mstns* is similar, both made up of 3 exons of relatively similar sizes, but with introns of different sizes (Fig 6. 2). FMstn-1 shared a higher identity with ZfMstn-1 and human Mstn than FMstn-2. The two Mstn paralogues previously described in trout (Rescan et al., 2001) share a 94% identity with each other, in contrast the paraloguous genes of *T. rubripes* and Zebrafish Mstns only share a 62% and 63% identity respectively (Fig 6.1b). Despite the apparent differences in the two trout Mstns when compared to the *T. rubripes* and Zebrafish paralogues (Kerr et al., 2005) their expression patterns are similar. *Mstn-1* transcripts in zebrafish and trout show ubiquitous expression in many adult tissues and throughout development whereas *Mstn-2* expression is restricted predominantly to brain and slow muscle fibres of free swimming larvae decreasing dramatically in maturing muscle (Rescan et al., 2001; Xu et al., 2003; Kerr et al., 2005). Kerr et al (2005) amplified ZfMstn-2 transcripts in slow muscle only after 70 cycles of RT-PCR showing its minimal expression in this tissue. *FMstn-2* was only identified in adult brain tissue and *FMstn-1* expression was wide spread in most adult tissues studied (Fig 6.5b). This is stark contrast to the Mstn expression in mammals where it is restricted primarily to skeletal muscle (McPherron et al., 1997; Ji et al., 1998). Interestingly we identified a significantly higher relative expression of *FMstn-1* in fast skeletal muscle compared to slow skeletal muscle in adult *T. rubripes*. This difference in expression may be due to the different growth patterns of the two fibre types. The two fibre types have different growth patterns in teleosts with slow muscle continuing to recruit fibres with increased body length (Johnston et al., 2004) and fast muscle only recruiting new fibres up to a clearly

identifiable body length (Weatherley et al., 1980; Johnston et al., 2003a; Johnston et al., 2003b; Johnston et al., 2003c). It may be of interest to further this observation by investigating *FMstn-1* expression in both muscle types at different growth stages, as FMstn-1 may play a different role in the growth of the two fibre types. Expression of *ZfMstn-1* using RT-PCR (Kerr et al., 2005) at specific developmental stages revealed a high expression during somitogensis followed by a drop in expression post-somitogenesis and a second up-regulation at hatch. Our qPCR results of *FMstn-1* during *T. rubripes* development at 18°C revealed a similar pattern of expression. We could not localise either *FMstn* transcripts using *in situ* hybridisation, a second observation shared with *ZfMstn* transcripts (Kerr et al., 2005). Xu et al. (2003) encountered similar problems previously.

Comparing the putative translations and expression patterns of the two *FMstns* with orthologues in other teleosts and mammals, a common ancestral *Mstn* gene (*Mstn-1*), human *Mstn*, *FMstn-1* and *ZfMstn-1*, arises and also a second less related paralogue found in teleosts. Phylogenetic analysis carried out by Kerr et al. (2005) suggests a divergence of *Mstn* genes due to a complete genome duplication resulting in two *Mstn* sister clades (Mstn-1 and Mstn-2) prior to the divergence of teleosts (Fig 6.6). The salmonid *Mstn* genes (Rescan et al., 2001) were additionally duplicated within salmonids after this initial divergence suggesting that the majority of fish *Mstn* genes currently described are in fact paralogues of *Mstn-1*. The ubiquitous tissue expression of *Mstn-1* orthologues in teleosts suggests a more universal role in growth and development than previously described in mammals where its primary role is the reduction of muscle growth by the inhibition of myoblast proliferation (Thomas et al., 2000a; Rios et al., 2001; Taylor et al., 2001; Langley et al., 2002).

Despite the proposed importance of Flstn in muscle development as shown in chick (Amthor et al., 2004), very little is known about its function related to muscle growth and development in teleosts. The 960 bp transcript codes for a 320-residue putative protein and is located on chromosome 12 in the closely related species *T. nigroviridis*. Despite the 75% identity with the human Flstn peptide *T. rubripes's* 5 exon transcript contains one fewer exon suggesting the possibility of splice variants. This is supported by a second larger transcript is predicted at Ensembl. Human *Flstn* has two splice variants coding for 344 and 317-residues (Shimasaki et al., 1988). The shorter of the two peptides in human arises from partial splicing resulting in a shorter carboxy truncated form of Flstn lacking 27 residues (Shimasaki et al., 1988). The zebrafish orthologue of 322 residues is also translated from a 5 exon transcript (Bauer et al., 1998). The lack of exon 6 does not interfere with the Kasal motifs of zebrafish and *T. rubripes* follistatin (Fig 6.1c).

Mammalian Flstn mRNA has been shown to be broadly distributed in adult tissues and not confined solely to tissues associated with reproduction (Patel 1998). Previous tissue analysis of *Flstn* transcripts in juvenile catfish (Gregory et al., 2004) using RT-PCR revealed expression was only shown in heart and testes. Our analysis using qPCR conflict this observation with the expression in *T. rubripes* highest in skeletal muscle and brain with significant expression also in spleen and heart. Transcripts were also detected in liver, kidney, skin and testes suggesting a wide spread biological role in fish as well as mammals. The differences between *T. rubripes* and catfish may be due to a number of factors including different growth stages used and the higher sensitivity of qPCR compared to RT-PCR.

Expression of *Flstn* during *T. rubripes* development revealed the highest expression during somitogenesis and post-hatch. *In situ* hybridisation paralleled this showing the majority of

Flstn expression coinciding with early brain and muscle differentiation during late gastrulation and somitogenesis. Expression beyond somitogenesis was confined to mid and hind brain. Amthor et al. (1996, 2002, 2004) have shown detailed expression patterns of *Flstn* particularly in developing chick somites. In the chick, *Flstn* expression is initially visible in the dorsolateral somite moving to the dorsomedial quarter. Within the dermomyotome, the muscle precursor cells of limb and tongue muscle, are *Flstn* positive (Amthor et al., 1996). Flstn induces muscle growth by enhancing BMP-7 action and prevents its ability to induce the apoptosis of MPCs (Amthor et al., 2002). *Flstn* expression in *T. rubripes* appears to follow a similar pattern to that shown in chick (Amthor et al., 1996) starting in the lateral quarter of the somite and moving medially in a rostral caudal gradient. Expression was also prominent in the mid and hind brain and the perimeter of the developing eye. This, in conjunction with the tissue expression, suggests the chief role of Flstn in *T. rubripes* is in the growth and development of skeletal muscle and neuronal tissue.

We have also revealed a substantial effect on transcript levels of *FMstn-1* and *Flstn* during development in response to changes in incubation temperature. During early somitogenesis relative expression of both transcripts was significantly higher at the lower temperature of 15°C compared to 18°C and 21°C. The expression at 21°C was marginally higher than at 18°C for both genes. Despite this there appears to be no change in the relative timing of *FMstn-1* and *Flstn* expression suggesting no heterochronic shifts. Flstn is known to be a inhibitor of Mstn protein in mammals (Amthor et al., 2004) so an increase in Flstn transcription may increase Flstn protein levels resulting in a reduction of Mstn function but not necessarily its transcription. Previous work in chick has described a close expression pattern of the two genes revealing abutting and overlapping expression. It has also been suggested in mammals recently that Mstn regulates its own expression through a negative

feedback loop (Forbes et al., 2006), therefore a decrease in function may in fact increase expression. The close proximity and timing of expression of the two genes in chick (Amthor et al., 2004) suggests a potential common transcriptional regulator that may explain the similar effect temperature has on their transcription. Analysis of their promoter regions may shed light on common binding motifs and potential transcriptional regulators. A further possibility is the increase in *Flstn* expression may result in an increase in endogenous Flstn protein and a subsequent down-regulation of Mstn function. The proposed negative feedback loop of Mstn (Forbes et al., 2006) may then be hindered resulting in an increase in transcription of *Mstn*. Several possible scenarios would therefore exist: (1) Transcription of both genes is influenced by embryonic temperature; (2) An upstream common regulator of the transcription of both genes is altered by embryonic temperature; (3) *Flstn* transcription is influenced by temperature and results in a change in *Mstn* expression through the inhibition of Mstn protein function. All these scenarios propose the transcription of *Flstn* and *Mstn* to be under the influence of temperature in a direct or indirect fashion. To address these propositions it is essential to measure protein levels under the same temperature conditions to correlate any changes in protein level with the amount of transcription. It would also be of interest to investigate the interaction kinetics of the two proteins. The changes in fibre number and size observed in teleosts when manipulating embryonic temperature (Johnston and Hall 2004) may therefore be related to the same phenomenon observed in both mammalian and teleost systems using Mstn mutants (Wargelius et al., 1999; Wegner 2000; Acosta et al., 2005; Rehfeldt et al., 2005) and over expression of *Flstn* (Matzuk et al., 1995; Lee and McPherron 2001).

6. 5 Figures

Figure 6. 1a FMstn-1 and FMstn-2 protein alignment.

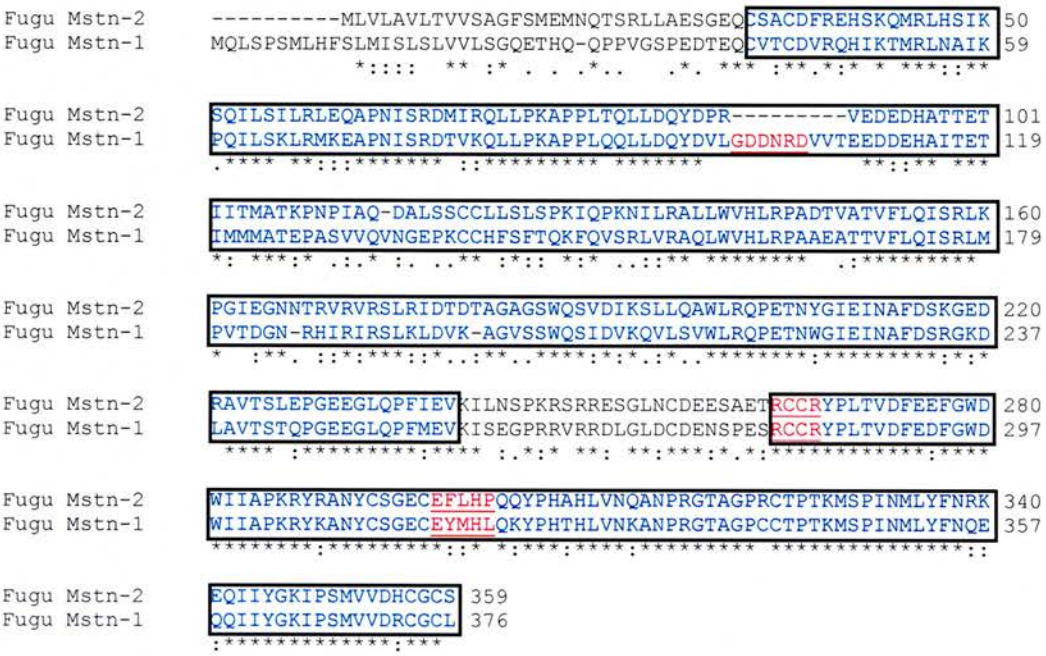


Figure 6. 1a. Sequence alignment of FMstn-1 (Q6SYV) and FMstn-2 (Q6SYV4) protein sequences. Conserved substitutions are marked with a colon, semi-conserved substitutions with a dot and global conservation is represented with an asterisk. The latency associated peptide (LAP) and TGF-β domains are boxed and highlighted blue respectively. Other sequences of interest are highlighted red and underlined.

Figure 6. 1b Myostatin protein alignment.

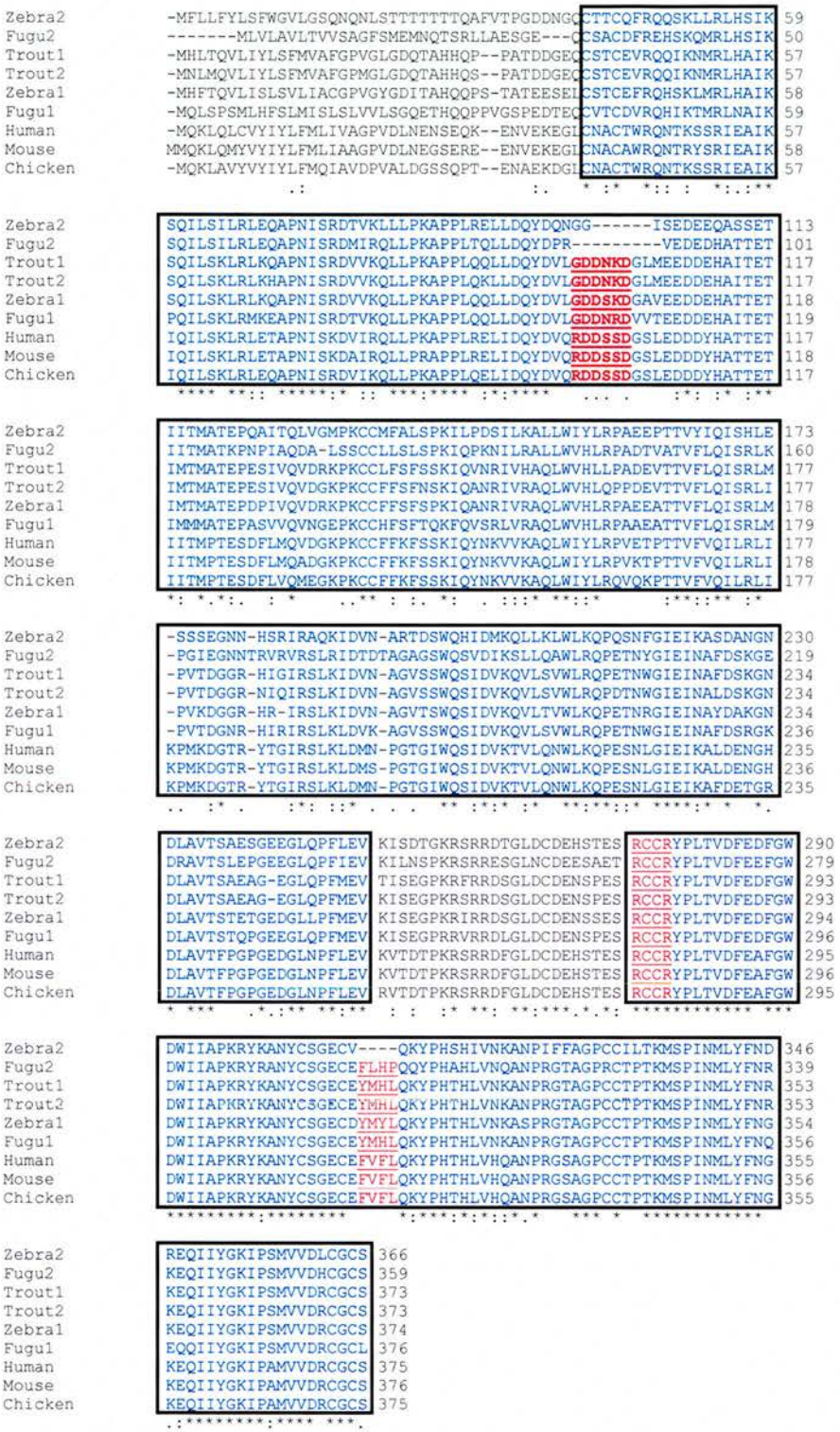


Figure 6.1b. Multiple sequence alignment of Myostatin protein sequences. The inferred FMstn-1 (Q6SYV3) and FMstn-2 (Q6SYV4) sequence are compared with those of ZfMstn-1 (Q42222), ZfMstn-2 (Q5MNY3) trout Mstn-1 (Q90ZD2), trout Mstn-2 (Q90ZD1), chick (O42220), mouse (O08689) and human (Q14793). Conserved substitutions are marked with a colon, semi-conserved substitutions with a dot and global conservation is represented with an asterisk. The latency associated peptide (LAP) and TGF-β domains (blue) are boxed respectively. Other sequences of interest are highlighted red and underlined.

Figure 6. 1c Follistatin protein alignment.

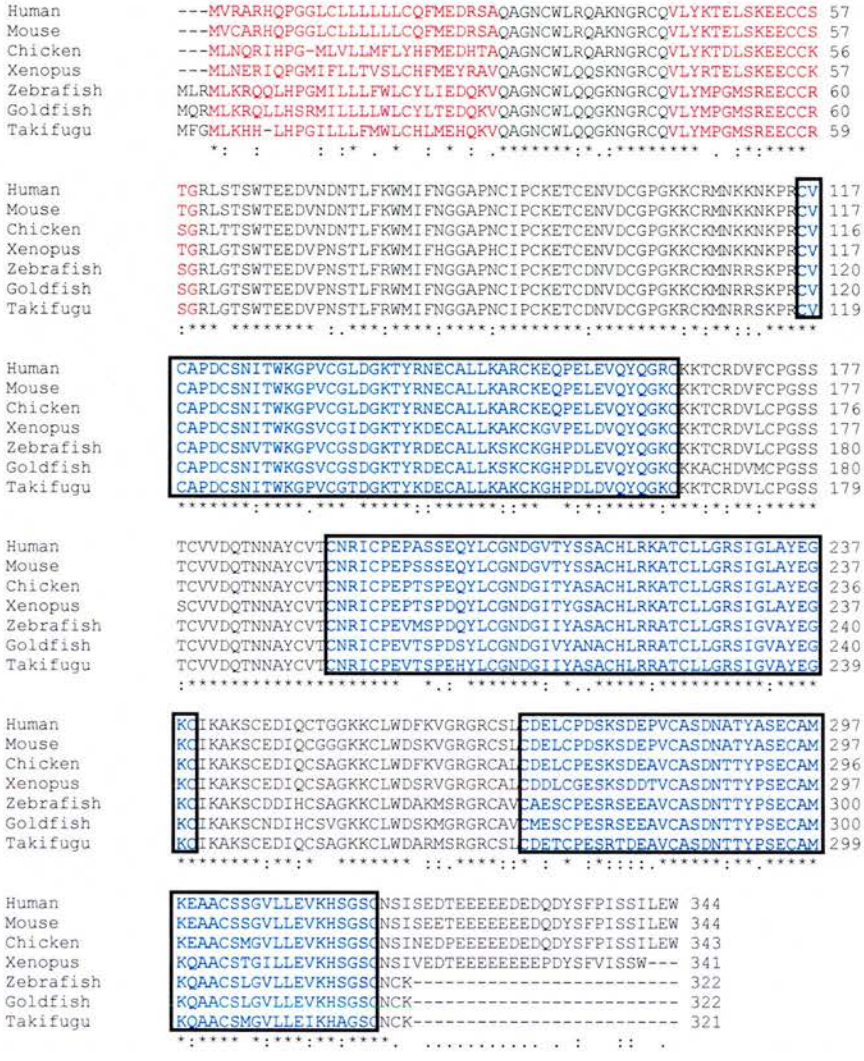


Figure 6. 1c. Multiple sequence alignment of Follistatin protein sequences. The inferred *T. rubripes* sequence is compared with those of Human (P19883), zebrafish (Q9YHV4), goldfish (Q6R256), xenopus (P31515), chick (Q90844) and mouse (P47931). Conserved substitutions are marked with a colon and global conservation is represented with an asterisk. The Kasal conserved motifs are highlighted blue and boxed. The regions corresponding to the two potential activin binding sites found in human and chick are highlighted red in all species.

Figure 6. 2 Genomic Organisation

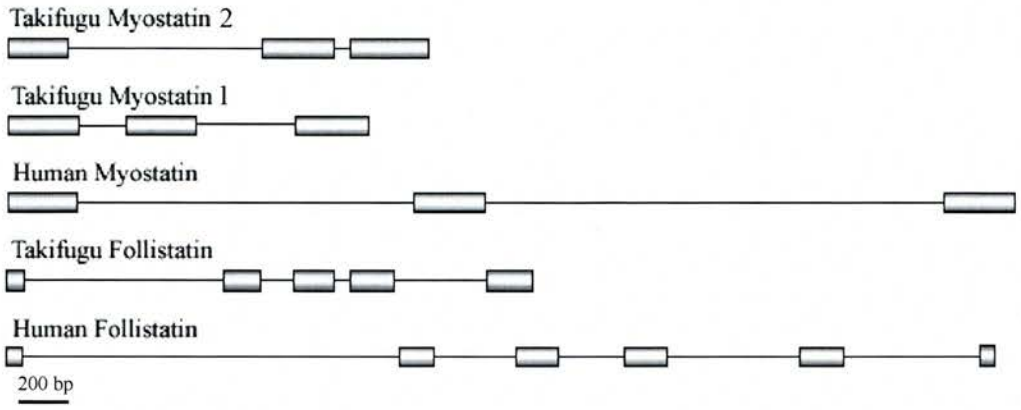


Figure 6. 2 Transcript structure of *FMstn-1*, *FMsnt-2* and *follistatin* from *T. rubripes* compaired to their human orthologues. Bars represent exons and lines represent introns. The scale bar indicates 200 bp.

Figure 6. 3 *Follistatin* expression by *in situ* hybridisation

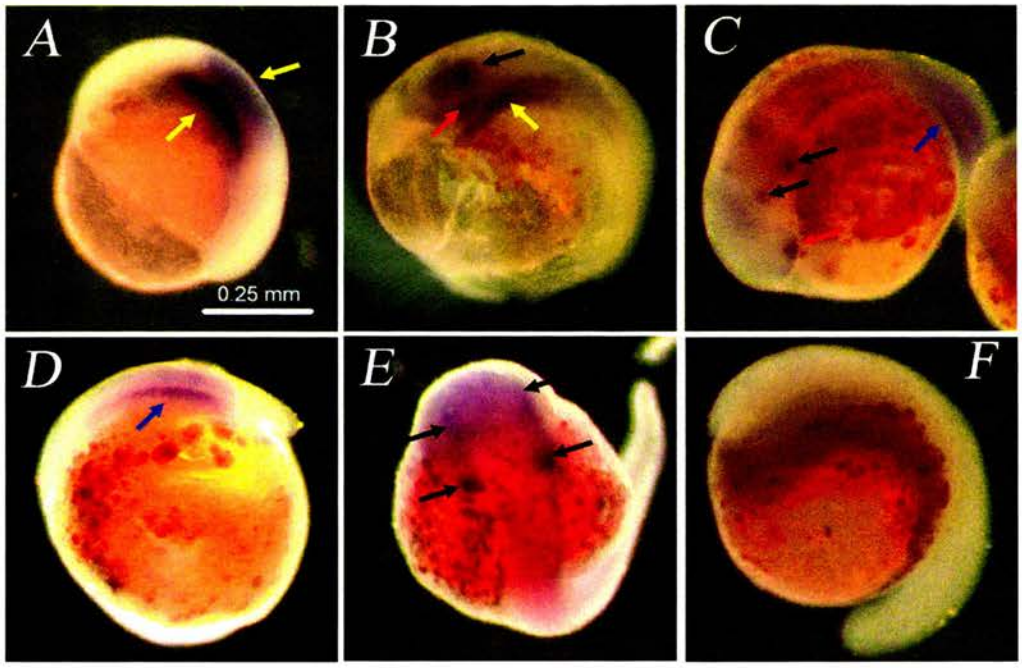


Figure 6. 3. Localisation of *Flstn* transcripts in *T. rubripes* embryos throughout development at 18°C by whole mount *in situ* hybridisation with a cRNA probe: (A) 42 hpf, bud stage of gastrulation. Strong expression in the pre-somitic mesoderm is indicated with yellow arrows. (B) 63 hpf, early-somitogenesis. Expression in the mid and hind brain is indicated with a black arrow, the yellow and red arrows indicate pre-somitic mesoderm and the otic vesicle respectively. (C) 80 hpf, mid-somitogenesis. Black arrows indicate staining in the mid and hind brain, the red arrow shows expression in the eye and staining in the somites is indicated with a blue arrow. (D) 90 hpf, late-somitogensis. Staining in the somites is shown with a blue a arrow. (E) 135 hpf, mid pharyngula. Expression restricted to the mid and hind brain. (F) 63 hpf, no staining was visible in the sense control embryos. Scale bar in A shows 0.25 mm. Anterior is left in except D and E where anterior is bottom and top respectively.

Figure 6. 4 *Follistatin* qPCR analysis

Figure 6. 4a

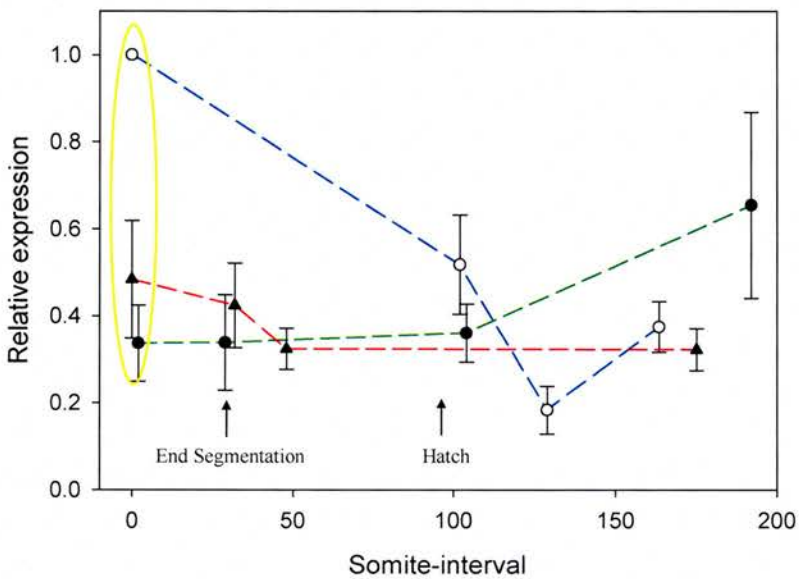


Figure 6. 4b

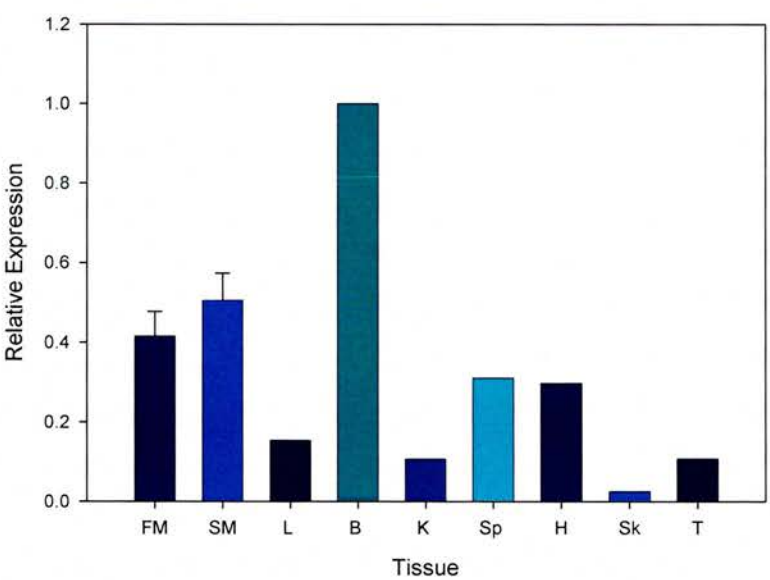


Figure 6. 4. Follistatin expression as determined by qPCR.

(a): Relative *follistatin* expression during *T. rubripes* development at 15°C (blue), 18°C (green) and 21°C (red) normalised to somite interval. A general linear model was carried out revealing significant changes in expression throughout development at all 3 temperatures ($p=<0.001$). A Mann Whitney U non parametric test revealed significant differences in *follistatin* expression with respect to temperature during early somitogenesis (15/18°C $p=0.029$, 15/21°C $p=0.029$). RNA polymerase II was used as the endogenous internal control and all points ($n=4$) were normalised to the highest expression value (78 hpf 15°C).

(b): *Follistatin* expression in *T. rubripes* adult tissues as determined by qPCR. FM, (fast skeletal muscle), SM (slow skeletal muscle), L (liver), B (brain), K (Kidney), Sp (spleen), H (heart), Sk (skin), and T (testes). 18s rRNA was used as an endogenous control and relative expression was normalised against brain. N = 7 for fast and slow skeletal muscle. Where error bars are not shown data represent an average of 2 values that did not differ by more than 25%.

Figure 6. 5 *FMstn-1* qPCR analysis

Figure 6. 5a

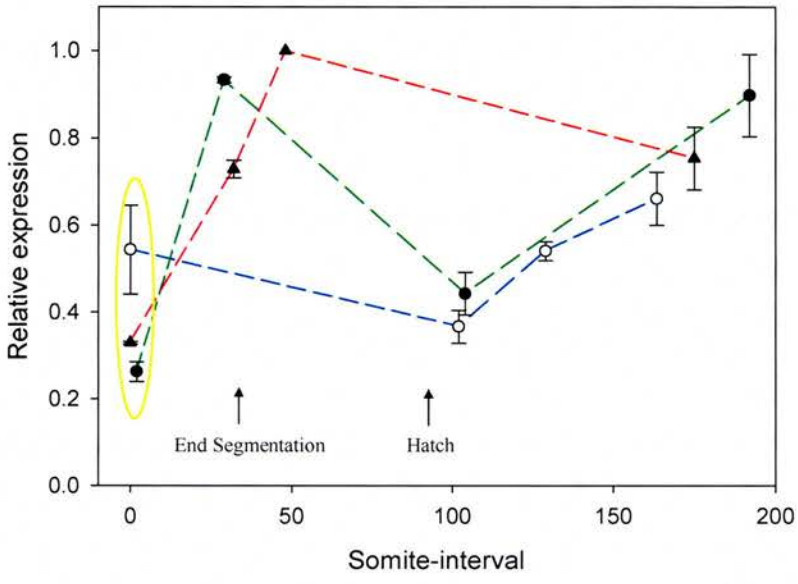


Figure 6. 5b

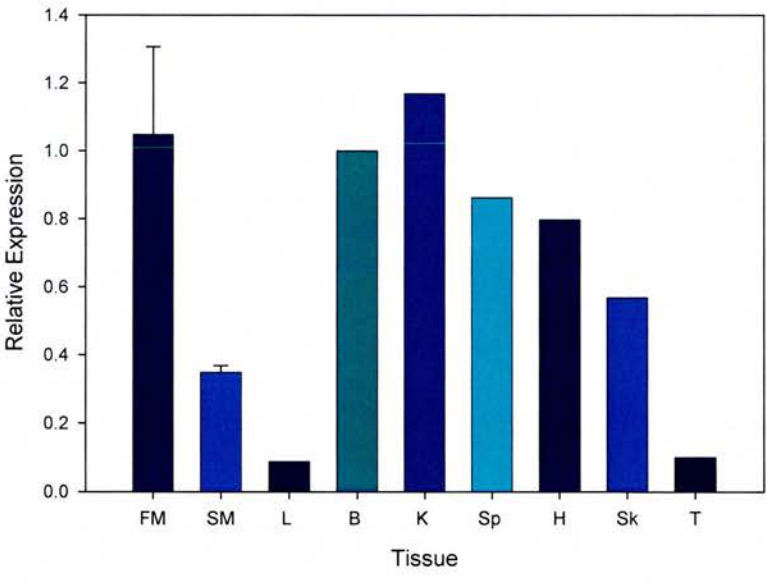


Figure 6. 5. FMstn-1 expression as determined by qPCR.

(a): Relative *FMstn-1* expression during *T. rubripes* development at 15°C (blue), 18°C (green) and 21°C (red) normalised to somite interval. A general linear model was carried out revealing significant changes in expression throughout development at all 3 temperatures ($p=<0.001$). A Mann Whitney U non parametric test revealed a 2-fold and 1.5-fold in *FMstn-1* expression with respect to temperature during early somitogenesis (15/18°C $p=0.029$, 15 /21°C $p=0.029$). RNA polymerase II was used as the endogenous internal control and all points ($n=4$) were normalised to the highest expression value (70 hpf 21°C).

(b): *FMstn-1* expression in *T. rubripes* adult tissues as determined by qPCR. FM, (fast skeletal muscle), SM (slow skeletal muscle), L (liver), B (brain), K (Kidney), Sp (spleen), H (heart), Sk (skin), and T (testes). 18s rRNA was used as an endogenous control and relative expression was normalised against brain. *FMstn-1* expression was 3.6 fold greater in FM compared to SM ($n=7$, $p=0.02$). Where error bars are not shown data represent an average of 2 values that did not differ by more than 30%.

Figure 6. 6 Myostatin phylogeny

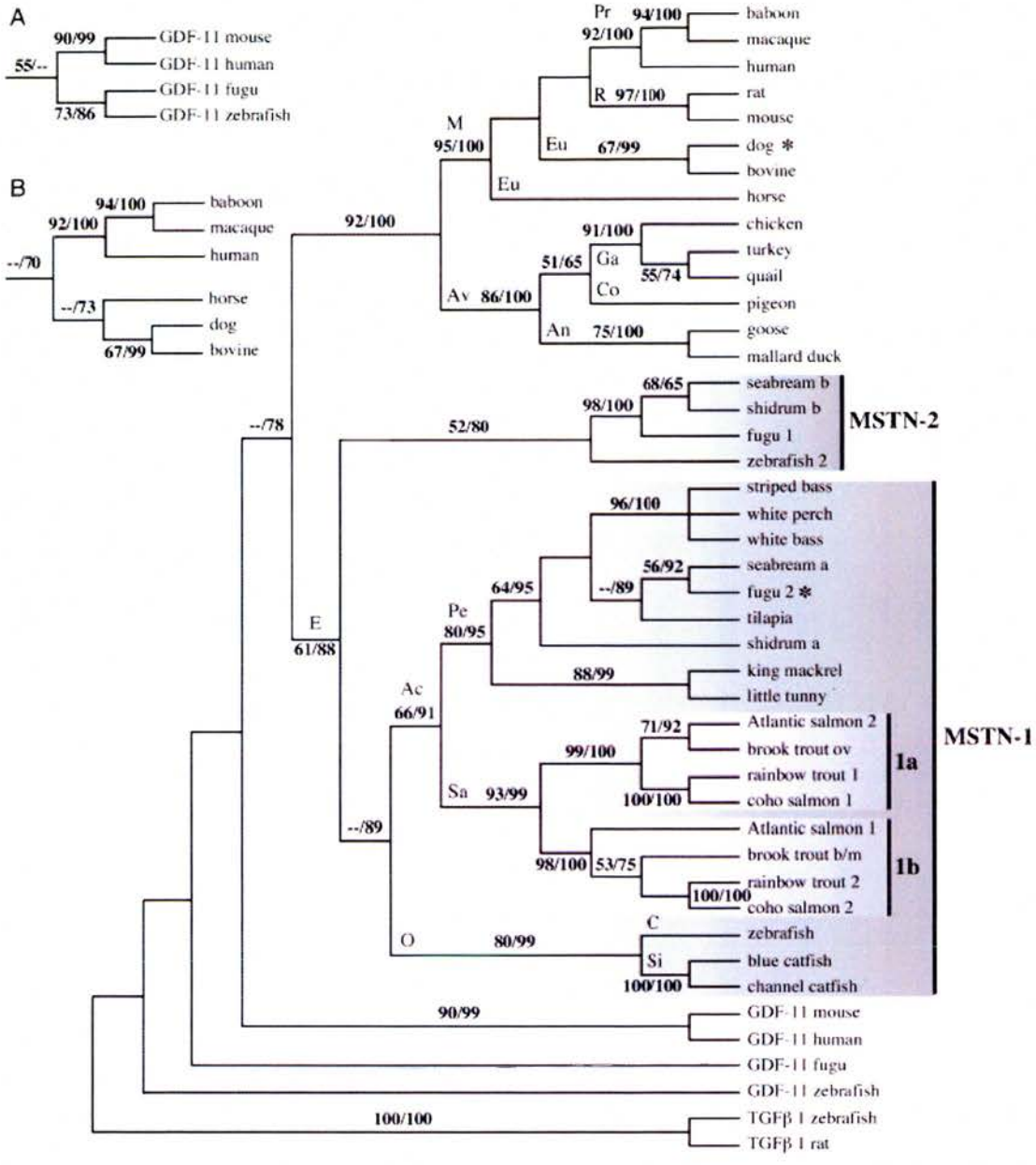


Figure 6.6. Mstn gene sub-family ML tree (Kerr et al, 2005). GDF-11 was used as an out group due to its structural similarity to Mstn. Dark gray-shaded clades refer to the two Mstn paralogues (Mstn-1/-2). Light gray shaded areas refer to the salmonid paralogues, described by Kerr et al (2005) as Mstn-1a and Mstn-1b. ML bootstrap values are indicated above branches and Bayesian posterior probability values if above 50% are below branches. Major classification units are mapped onto branches using the following abbreviations: Ac, Acanthopterygii; An, Anseriformes; Av, Aves; C, Cyprinoformes; Co, Columbiformes; E, Euteleostei; Eu, Eutherians; Ga, Galliformes; M, Mammalia; O, Ostariophysi; Pe, Perciformes; Pr, Primates; R, Rodentia; Sa, Salmoniformes; Si, Siluriformes. Species not within the in classification are marked with an asterisk. Abreviations: Mstn, Myostatin; ML, maximum likelihood; GDF, growth/differentiating factor. FMstn-1 and FMstn-2 are shown as fugu2 and fugu1 respectively. This figure was adapted from Kerr et al, 2005.

Chapter 7: A genomic approach to reveal novel genes associated with myotube formation in the model teleost, *Takifugu rubripes*.

7.1 Abstract

We adopted a genomic approach to identify genes that may have a role in myotube formation in the fast muscle of *Takifugu rubripes (T. rubripes)*. Forward and reverse subtracted cDNA libraries were prepared from the fast muscle of a 180 g (myotube +) and a 3.4 kg (myotube -) fish resulting in 1,452 differentially expressed expressed sequence tags (EST). After grouping the ESTs into non redundant clusters and removal of house keeping and structural genes, 57 sequences were chosen to investigate their differential and tissue expression with the use of qPCR. Fast muscle from independent myotube (+) and myotube (-) fish acclimated to the same environmental conditions and diet was used to clarify the significance of their differential expression before further tissue expression studies. Eleven novel genes showed consistent differential expression and four of these were found to have appropriate tissue expression. These four genes were upregulated 5-25 times in the fast muscle of myotube (-) fish relative to myotube (+) fish while their expression remained unchanged in the other tissues studied. The novel genes identified are also present in other vertebrate genomes and may play a role in the inhibition of myotube production in vertebrate muscle.

7.2 Introduction

Embryonic myogenic cells in mammals differentiate and fuse together to form myotubes and primary myofibres, the structural scaffold to which secondary muscle fibres are formed (Handel and Stickland 1987b; Ontell et al., 1988; Wilson et al., 1992; Oksbjerg et al., 2004). Muscle fibre number in mammals is defined shortly after birth and during normal muscle growth myotube production is inhibited. Subsequent muscle growth occurs through the increase in length, diameter and nuclear content of fibres already present (Rowe and Goldspink 1969; Schultz 1996). A pool of myogenic progenitors (myosatellite cells) residing between the sarcolemmal and basil lamina are responsible for providing nuclei for growth and tissue maintenance of post-embryonic muscle (Schultz and McCormick 1994; Hawke and Garry 2001). Despite this certain stimuli such as exercise (Darr and Schultz 1987) and injury (Blaveri et al., 1999) can induce myotube production in skeletal muscle.

Teleost skeletal muscle offers a good model for studying vertebrate myogenesis since its fast and slow skeletal muscle are arranged in discrete layers and myotube production is not restricted to embryonic muscle growth (Johnston 2001). Furthermore, the two fibre types have distinct post embryonic growth patterns with slow muscle continuing hyperplastic growth with fish length (Johnston et al., 2004) and fast muscle fibre number increasing until a pre-determined body length after which further muscle growth is by hypertrophy (Weatherley et al., 1988; Rowlerson et al., 1995; Johnston et al., 2003b; Johnston et al., 2004). The nuclei required for post-embryonic fast muscle hyperplasia and hypertrophy in teleosts are derived from a population of myogenic progenitor cells (MPCs), equivalent to mammalian myosatellite cells (Johnston and Hall 2004). These MPCs express the myogenic regulatory

factors (MRFs, see chapters 3a and 3b) defining the myogenic lineage and differentiation (Edmondson and Olson 1993; Kassar-Duchossoy et al., 2004).

Although progress has been made in defining these myogenic progenitor cells and their responsibility for post embryonic muscle growth and fibre recruitment, little is know about the transcriptional networks regulating their function. A splice variant of the neural cell adhesion molecule was found to be upregulated during myogenesis and induced myoblast fusion (Dickson et al., 1990). The calpain family, FOXO1a and the cytokine IL-4 have also been identified in the fusion of myoblasts (Bois and Grosveld 2003; Horsley et al., 2003; Dedieu et al., 2004). The present study takes advantage of the suppressive subtractive hybridisation assay to identify novel genes whose expression patterns correlate with the termination of hyperplastic muscle growth in fast skeletal muscle. Previous work in our lab identified the approximate threshold body size for the end of fibre recruitment (Fig 7.1) was 1.2 kg with a standard length of 35 cm and the maximum fibre diameter measured in a 4kg *T. rubripes* was ~300μm (Fernandes et al., 2005). This data was used to select appropriate fish for the subtracted libraries.

7.3 Results

7.3 *i* Construction and characterisation of two subtracted cDNA Libraries.

Two subtracted cDNA libraries were made using the fast skeletal muscle from *T. rubripes*, care was taken to avoid the intervening slow muscle in the fast muscle mass. The reverse library (myotube +) was made using cDNA from a small *T. rubripes* still recruiting fibres (180 g) as the tester and the cDNA from a large *T. rubripes* (3.4 kg) which had ceased recruiting fibres as the driver. This library contained genes with a higher expression in fish that were still recruiting fibres. A reciprocal forward library (myotube -) was made containing genes with a higher expression in fish that had stopped recruiting fibres. 1,152 clones from each library were single pass sequenced from the 5' end resulting in 804 and 648 good quality ESTs from the myotube (+) and myotube (–) library respectively. The average EST length after removal of poly A and vector sequences was ~510 bp. All ESTs were submitted to the EST database dbEST. Using partigene the ESTs were grouped into nonredundant clusters to account for multiple representatives of the same gene resulting in 293 putative genes from the myotube (-) library and 392 from the myotube (+) library, of which 175 and 262 respectively were singletons. Only 33 of these clusters were common to both libraries reflecting the efficiency of the subtractions, this is further corroborated by the overall redundancy of the libraries, defined as the ratio between the total number of sequences and the number of clusters, being 2.74 and 1.65 for the myotube (-) and myotube (+) library respectively. The common clusters contained several ESTs and corresponded to structural and motor proteins as well as some enzymes involved in glycolytic metabolism. Overrepresented genes found in the libraries included fast skeletal muscle myosin heavy chain, tropomyosins, aldolase A, G3PDH (glyceraldehydes-3-phosphate dehydrogenase) and

the muscle isoform 1 of creatine kinase which was represented by 43 sequences for a single cluster. A very small percentage of the clusters were found to represent mitochondrial or ribosomal RNA sequences (Fig 7. 2). Only 62% of the clusters from the myotube (-) library and 59% from the myotube (+) library had significant (expected e value of less than 10^{-8}) hits against the NCBI non redundant protein database (Fig 7. 2). The rest of the clusters were not accounted for in the *T. rubripes* ensemble predicted gene set. Approximately 24% of the clusters from both libraries had open reading frames as predicted by ESTscan (Iseli et al., 1999) or DECODER (Fukunishi and Hayashizaki 2001) these sequences were therefore novel sequences, sequences with hits against the nonredundant protein database were not translated. The remaining 13% and 17% of clusters from the myotube (-) and myotube (+) respectively could not be translated and were therefore most likely pseudogenes and untranslated regions (UTRs).

Putative translations of the clusters from both libraries were categorised according to their gene ontology (GO), a large proportion of the putative proteins however had no associated GO and were referred to as unclassified (Fig 7.3a, 3b). The majority of these proteins were hypothetical and had no know function or cellular location. Approximately 70% of the proteins with a GO classification were proteins with binding activity or enzymes. Comparisons with the *T. nigroviridis* genome assembly at ensembl revealed approximately 70% of the clusters from both libraries had matches with known or predicted *T. nigroviridis* genes. Only 4 clusters from each library had no hits in the current *T. nigroviridis* genome assembly.

7.3 *ii* Candidate gene selection

Both subtracted libraries were screened for putative candidates with possible involvement in muscle fibre recruitment. Housekeeping genes, common structural proteins, pseudogenes, UTRs and clusters common to both libraries were discarded and from the remaining putative genes 24 and 33 were selected from the myotube (-) and myotube (+) libraries respectively (table 7.1a, 1b). Only the results from the forward (myotube (-) library) subtracted library will be presented here.

qPCR was then used to confirm the differential expression of 24 candidates from the myotube (-) library using cDNA from the fast muscle of the fish used to make the subtracted libraries. Sequence specific primers were designed for all candidates (see tables 7.2a) and 18s rRNA. Primers used for the myotube (+) library are shown in table 7.2b. 18s rRNA was chosen as the internal standard to nomalise the results in light of its transcript levels not changing significantly between samples. Single dissociation peaks were observed for all primer sets confirming only one amplified product per reaction. Expression levels of approximately 75% of the candidates from the forward library showed significant up-regulation in the appropriate sized fish. The remaining candidates showed an unaltered expression level or insignificant change confirming further the efficiency of the library with only 2 from the myotube (-) library significantly down-regulated.

7.3 *iii* Real time PCR of candidates in fast muscle from six independent fish

The relative expression of the selected candidates, minus the 25% showing no upregulation, were analysed using a post hoc gene expression analysis by qPCR using the cDNA from the

fast muscle of 3 myotube (-) fish and 3 myotube (+) fish acclimatised for 3 weeks at the same environmental conditions and diet. This reduced the likelihood of the differentially expressed candidates being unique to the individual fish used for the library. 9 candidates from the myotube (-) library showed consistent and significant differential expression between the appropriate sample groups (the 9 genes including the 2 from the myotube (+) library are shown in table 7.3).

7.3 *iv* Tissue specificity of nine candidates

The relative differential expression of the 9 myotube (-) candidates was then investigated in various tissues using fast skeletal muscle (FM), slow skeletal muscle (SM), heart (H), liver (L), skin (Sk) and brain (B) of myotube (-) and myotube (+) fish respectively. Using this analysis the candidates were categorised into strong candidates, intermediate candidates and unlikely prospects according to certain criteria. Strong candidates had a high myotube (-)/ myotube (+) ratio in fast muscle but not in other tissues and expression was limited to fast muscle in the myotube (-) fish. Intermediate candidates have a high myotube (-)/ myotube (+) ratio in fast muscle and limited expression a limited number other tissues in myotube (-) fish. Unlikely prospects showed a high myotube (-)/ myotube (+) ratio in multiple tissues and wide spread expression in myotube (-) fish.

Four genes from the myotube (-) library, *FRC258*, *FRC363*, *FRC386*, *FRC405*, followed the criteria for strong candidates showing a higher expression in the fast muscle of myotube (-) fish by an average factor of 25, 5, 16 and 21 respectively. Upregulation in other tissues between myotube (-) and myotube (+) fish was minimal to none and the relative expression in the tissues of myotube (-) fish was at least 4-fold greater in fast muscle (Fig 7.4a). These four

genes were hypothesised to have an involvement in the inhibition of muscle fibre recruitment with their differential expression appearing to be specific to fast muscle rather than just body size *FRC167*, *FRC214* and *FRC272* were considered to be intermediate candidates with a higher expression in the fast muscle of myotube (-) by an average factor of 5, 15 and 13, however expression varied with size in other tissues (Fig 7.4b). *FRC40* and *FRC263* from the myotube (-) library consistently showed differential expression in all tissues analysed and were unlikely candidates (results not shown).

7.3 *v* Bioinformatic analysis of likely and intermediate candidates

FRC258, *FRC363*, *FRC386* and *FRC405* are all found in the current *T. rubripes* genome assembly on scaffolds 278, 237, 62 and 572 respectively. Orthologues of the genes can be found on chromosomes; 5, 14, 7, 6 of human and 12, un-random, 2 and 17 of *T. nigroviridis*. Other ensembl genomes including zebrafish and mouse have predicted orthologues of the genes. *FRC258* is made up of 9 exons coding for a putative protein of 403 residues and is an orthologue of the human cardiomyopathy- associated 5 protein (NP_705838) with many functional domains including B302, Fibronectin and SPRY. It shares a 22% and 65% identity with the corresponding protein in human and *T. nigroviridis* respectively. *FRC363* and *FRC405* contain 5 and 2 exons and code for 483 and 635 residue proteins containing BTB_POZ and Kelch repeat domains. *FRC363* has a 34% and 76% identity with its orthologues in human and *T. nigroviridis* respectively. *FRC405* however is more conserved sharing a 75% and 84% identity with its human and *T. nigroviridis* orthologues respectively. *FRC386* is a large gene with the ensembl prediction 9 exons in length and coding for a putative protein of 403 residues described as hypothetical from a family of conserved

eukaryotic proteins with undetermined function. The conserved region is known as DUF410 and is found in many species including human and yeast.

The 3 intermediate candidates (*FRC167*, *FRC214* and *FRC272*) also have orthologues in human and *T. nigroviridis*. *FRC167* matches a novel ensembl prediction composed of 3 exons and codes for a small 85 residue peptide with 93% similarity to its *T. nigroviridis* orthologue. Despite no GO classification the protein contains highly conserved tripeptide motif close to the N-terminus found in the LYR family of proteins. It is note worthy that splice variants of the gene have been predicted in most species. The predicted transcript structure of *FRC214* contains 4 exons that translate to a 430 residue protein containing the POZ/BTB domain and kelch repeats found in *FRC363* and *FRC405*. It shares a 35% and 65% identity with the corresponding human and *T. nigroviridis* proteins respectively. The FRC272 transcript is composed of 6 exons corresponding to a 144-residue peptide with orthologues in both human (58% identity) and *T. nigroviridis* (71% identity). This essentially uncharacterised protein contains a bipartite nuclear localisation signal and an A1pp processing enzyme domain.

7.4 Discussion

This study has resulted in the identification of four novel genes (*FRC258*, *FRC363*, *FRC386* and *FRC405*) that are consistently and specifically up-regulated in the fast skeletal muscle of *T. rubripes* no longer recruiting muscle fibres. The temporal and tissue expression of these candidate gene transcripts suggests strongly that they have an involvement in the inhibition of myotube formation in teleosts and indeed other vertebrates. *FRC258* has a 25-fold upregulation in the fast muscle of myotube minus fish and is an orthologue of human cardiomyopathy- associated 5 protein, not previously identified in teleosts. The conserved domains of this protein (B302, SPRY, fibronectin) suggest many possible functions including transcriptional regulation, cell migration, differentiation and adhesion (Dean et al., 1987; Ponting et al., 1997). FRC363 and FRC405 show a 5 and 21-fold upregulation in myotube (-) fish and both translate into previously unidentified proteins containing BTB_POZ domains and Kelch repeats. These domains propose a few possible functions of the proteins and are found in a large family of proteins with wide spread function know as the Kelch family. Kelch is a 50-residue motif first identified in drosophila and is found in a number of different proteins including the mouse MIPP protein, Drosophila egg chamber regulatory protein, a number of Pox viruses (Xue and Cooley 1993). A likely function of Kelch containing proteins is a cytoskeletal function as shown in the developing Drosophila oocyte where it is localised to the actin-rich ring canals that connect the 15 nurse cells to the developing ooctye (Way et al., 1995). BTB_POZ domains are found near the N terminus of a number of Zinc finger proteins and in families containing the Kelch motif and some pox virus proteins. The domain mediates homomeric dimerisation and in some instances heteromeric dimerisation (Bardwell and Treisman 1994; Zollman et al., 1994). POZ domains from Zinc finger proteins have been shown to mediate transcriptional repression (Deweindt et al., 1995) and interact

with co-repressors such as SMRT and N-CoR (Huynh and Bardwell 1998). It can be proposed that *FRC363* and *FRC405* may have an involvement in protein/ protein interactions possibly repressing transcription cooperatively with a co transcriptional regulator. These protein/ protein interactions leading to transcriptional repression are common in many developmental processes including the cell cycle where accurate regulation is essential. For example retinoblastoma tumour suppressor protein (pRb) represses transcription of genes essential for cell cycle progression by interacting and ultimately inhibiting a family of transcription factors known as the E2Fs and their associated heterodimer DP-1 protein. (Harbour and Dean 2000; Wu et al., 2001). This activity promotes cell cycle arrest and subsequent exit into cell differentiation (Walsh and Perlman 1997).

FRC386 has a 16-fold upregulation in the fast muscle of myotube minus fish. The corresponding Ensembl prediction is a hypothetical open reading frame (ORF), 9 exons in length and codes for a 403 residue protein. The protein has no conserved domains with recognised function but has an 85% identity with a novel human protein know as CGI-20 (Q9Y309) identified using Comparative Genomics Identification (CGI) with the *C. elegans* genome protein sequences (proteome) as a scaffold for identifying novel human genes (Lai et al., 2000). This comparative genomics analysis proves an evolutionary conservation of the gene. The conserved region, although functionally unknown, has the identification DUF410.

The possible functions of these candidate genes suggest a number of ways in which myotube formation ceases in teleosts. The Kelch domains suggest a more transcriptional regulation where transcription of crucial genes may be activated or indeed repressed by a protein complex. The cardiomyopathy- associated 5 protein protein appears more likely to play a role in cell migration and adhesion perhaps having an involvement in muscle hypertrophy

whereby it becomes active once fibre recruitment has ceased and subsequent muscle growth relies on the elongation and expansion of fibres already present. Further functional work would be required to confirm their actual role, however it can be assumed the process is a complex series of events ultimately leading to the inhibition of muscle fibre recruitment.

The three intermediate genes (*FRC167*, *FRC214* and *FRC272*) were classed as such due to their up-regulation in myotube (-) fish not specific to fast muscle. For example *FRC214* showed a 15-fold increase in expression in fast muscle and little relative expression when compared to fast muscle in myotube (-) fish however showed a significant up-regulation in liver in myotube (-) fish. Tissue specificity was important in categorising the candidates to eliminate the likelihood of differential expression being solely due to body size. Slow skeletal muscle was included in this comparison due to its different growth pattern in teleosts when compared to fast muscle. Slow skeletal muscle in teleosts continues recruiting fibres with increasing fish length (Johnston et al., 2004).

FRC167 showed a 5-fold increase in expression in the fast muscle of myotube (-) fish. Its full length transcript has 3 exons and codes for a short 85 residue protein. The complex 1 domain suggests it is a member of the LYR family so named after a highly conserved tripeptide motif close to the N-terminus of these proteins. GO of this motif suggests localisation to the inner mitochondrial membrane and an involvement in electron transport with oxidoreductase activity, acting on NADH or NADPH (Walker et al., 1992). Its translated product has a 93% identity with its orthologue in *T. nigroviridis* and a number of splice variants have been predicted in many ensembl genomes. Interestingly it has a 64% identity with two possible paralogues found on chromosomes 12 and 22 in human.

The transcript of FRC214 increased 15-fold in the fast muscle of the myotube (-) fish. The 403 residue translated product of its 4 exons contain a BTB_POZ and Kelch domain suggesting a similar function to FRC363 and FRC405 and shares a low 35% and 65% identity with its closest human and *T. nigroviridis* orthologue respectively. Finally the FRC272 transcript was upregulated 13 times in myotube (-) fish and composed of 6 short introns coding for a 144 residue protein. The protein is uncharacterised but has predicted orthologues in both Human and *T. nigroviridis* (58% and 71% identity respectively) and contains an A1pp domain and a bipartite nuclear localisation signal. This suggests the protein may be a cyclic phosphodiesterase (Martzen et al., 1999).

Despite the differential expression of the remaining 4 candidates from the myotube (-) library (FRC40, FRC263) and myotube (+) library (FRC177, FRC300) in fast muscle this differential expression was relatively ubiquitous among all tissues. Body size can have a major effect on various cellular processes during growth. In teleosts increasing skeletal muscle size has a negative effect on aerobic enzyme activity, while glycolytic enzyme activity increases with body mass (Somero and Childress 1980). These four genes were concluded to be involved in cellular processes and any expressional changes were due to body size, playing no specific part in myotube production and were discarded as possible candidates for this study.

The 7 candidate genes potentially involved in the inhibition of myotube production were identified amongst 1451 ESTs generated from 2 subtracted libraries. The Fwd myotube (-) library provided all the final candidates. Programmes that predict the presence of genes in sequenced genomes such as GENSCAN (Burge and Karlin 1997) have become increasingly useful in predicting possible genes structures amongst the ever escalating sequencing of

complete genomes. Despite these programmes the methods are not always able to predict correct exon/ intron boundaries in genomic DNA, particularly with alternatively spliced transcripts (Rogic et al., 2001) posing a problem for genome annotation. Experimental data can be used to assist such programmes particularly EST data where actual expressed sequences from cDNA can be used to experimentally gain information on exon/ intron structure and the presence of alternatively spliced transcripts. Suppressive Subtractive Hybridisation (SSH) not only reduces the reduncany of the libraries but improves the chances of identifying rare transcripts. A total of 178 EST clusters (41%), without any significant hit against the nr protein database were identified from our libraries with a high percentage of these not possessing an associated GO term, reflecting the fact their biological processes and molecular function have not yet been identified. The importance of ESTs has become apparent in our study with the majority of the uncharacterised clusters having no corresponding prediction at ensembl and in some cases identifying errors in exon structure of predicted transcripts. These ESTs may be UTRs, non conserved exons, pseudogenes or novel genes. Of these 178 clusters 168 could be translated using ESTscan or DECODER and therefore may represent a resource of previously unidentified genes with potential interest.

Interestingly we have also found a number of clusters coding for putative proteins with uncharacterised function that have orthologues in other species from *C. elegans* (Lai et al., 2000) to mouse (Okazaki et al., 2002). These uncharacterised genes have been conserved throughout evolution and are likely to have important yet unknown biological roles. Comparisons of our partial *T. rubripes* transcriptome with the *T. nigroviridis* genome assembly (Jaillon et al., 2004) revealed a total of 198 clusters that could not be found in the current *T. nigroviridis* assembly. Of these 109 had matches to *T. rubripes* predicted genes and only 8 could not be found in the *T. nigroviridis* genomic sequence. It is likely that they

account for errors in *T. nigroviridis* gene prediction but may also represent genuine differences in gene repertoire. Despite the relative short evolutionary distance, 18-30 million years, between the two species (Hedges 2002) there are some drastic physiological differences. The most obvious being size and preferred environmental habitat with the marine species *T. rubripes* growing up to ~70cm in length and the brackish water species *T. nigroviridis* attaining a maximum size of ~15cm.

The physical aspects of myotube production during teleost growth (Weatherley et al., 1988; Rowlerson et al., 1995; Johnston et al., 2003b; Johnston et al., 2004) and in response to certain stimuli such as exercise (Darr and Schultz 1987) and injury (Blaveri et al., 1999) is well documented however the current knowledge of the genetic networks regulating the process is limited. A few studies have been carried out on mammalian myotube production during embryogenesis, Dickson et el. (Dickson et al., 1990) found that alternative splicing of the neural cell adhesion molecule was regulated during myogenesis and constitutive expression of the shorter splice variant induced the fusion of myoblasts. It has also been shown recently that proteases from the calpain family are involved in myoblast migration and adhesion as differentiated myoblasts over expressing the protease inhibitor calpistatin showed no migratory competence (Dedieu et al., 2004). Two other molecules that have been shown to have an involvement in murine myoblast fusion are the cytokine IL-4 which acts on the transcription factor NFATc2 (Horsley et al., 2003) and the phosphorylated isoform of the forkhead transcription factor FOX01a (Bois and Grosveld 2003). Environmental factors such as embryonic temperature (Johnston et al., 2003b) and photo period (Johnston et al., 2003c) have a profound effect on myotube formation and hypertrophy altering the number of myotubes formed while producing proportional changes in the myonuclei content of the muscle fibres. Two potential scenarios exist: the most likely with these results is the presence

of a single population of Myogenic Precursor Cells (MPCs) for both hypertrophic muscle growth and myotube formation however the possibility of 2 separate populations of MPCs cannot be ruled out.

We have shown the power and ability of the SSH technique in creating small transcriptomes of organisms and identifying novel genes relating the analysis to specific developmental processes. The four strong candidates described in the present study were all expressed significantly higher in the fast muscle of fish that had stopped recruiting fibres and may form part of a genetic network regulating the inhibition of myotube production. Future studies in the functional aspects of the proteins these genes code for will need to be carried out to confirm this hypothesis. With the use of model teleost species (*Tetraodon nigroviridis* and *Danio rerio*) pinpointing the precise stage of growth that these genes are up-regulated and correlating this with the cessation of myotube production will be crucial to shed further light on this proposal. Functional studies using an in vitro myotube formation assay would also be of great interest to the future prospects of this study. For example, over expressing any one of the four strong candidates in a myogenic cell line may interfere with the ability of the cells to form myotubes in vitro. This would potentially confirm their role in the inhibition of myotube production. The proposed orthologues present in other vertebrate species suggest a possible role for these genes in myotube formation inhibition throughout vertebrates.

7. 5 Figures

Figure 7. 1 Fibre Number against body mass

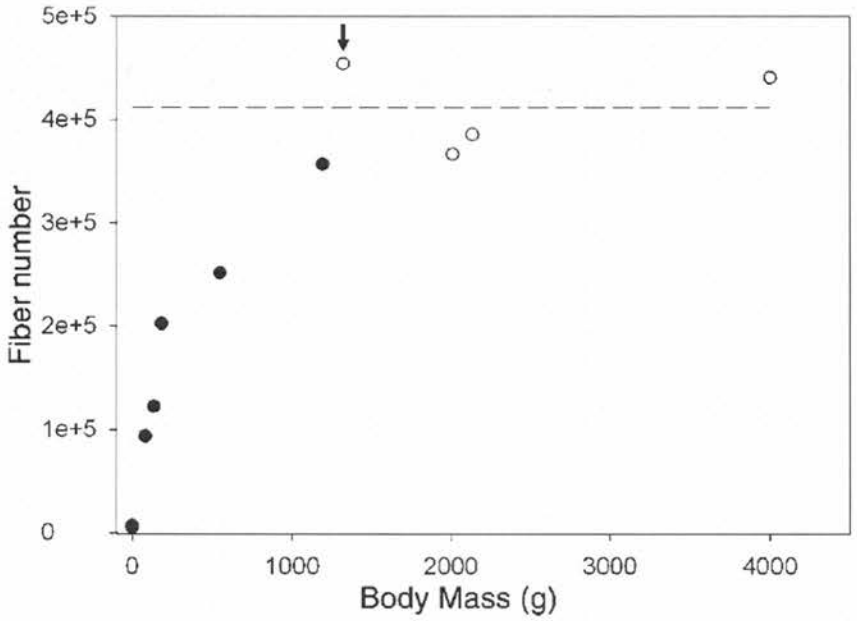


Figure 7. 1 Relationship between the body mass and the number of fast muscle fibres per trunk cross section at 0.7 standard length (SL) in *T. rubripes*. Figure adapted from Fernandes et al (2005).

Figure 7. 2 Global Classification of EST clusters

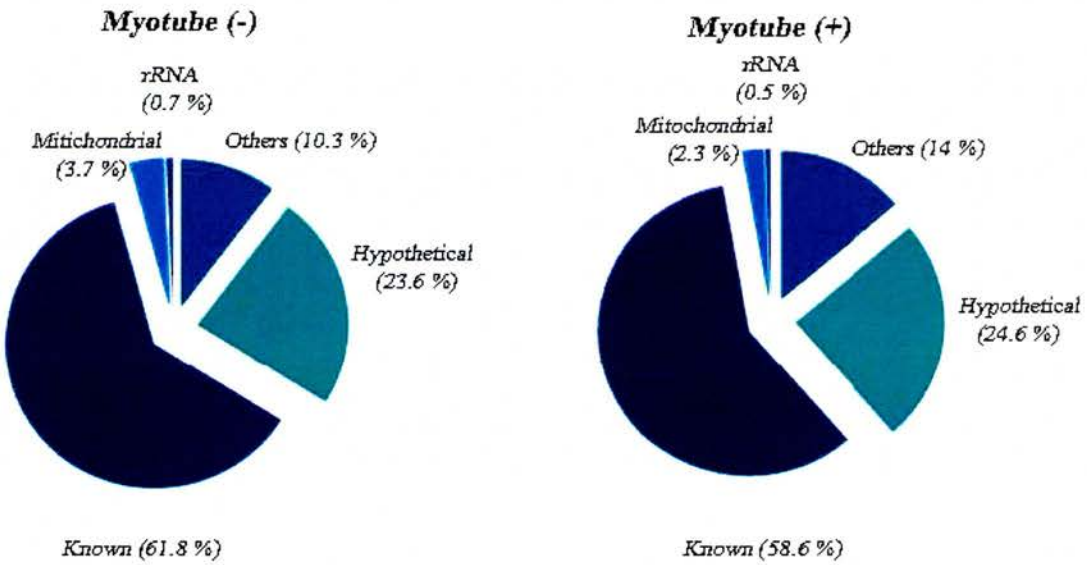


Figure 7. 2. Global classification of non redundant EST clusters from the myotube (-) and myotube (+) libraries. 'Known' clusters refer to groups of ESTs that have significant matches to entries in the nonredundant database of proteins. The clusters described as 'hypothetical' are novel, putative peptide sequences predicted by ESTscan or DECODER. ESTs predicted not to code for functional proteins including untranslated regions (UTR) and pseudogenes were classified as 'others'. Proportions of mitochondrial and rRNA sequences are also shown.

Figure 7. 3a Gene ontology from myotube (-) library

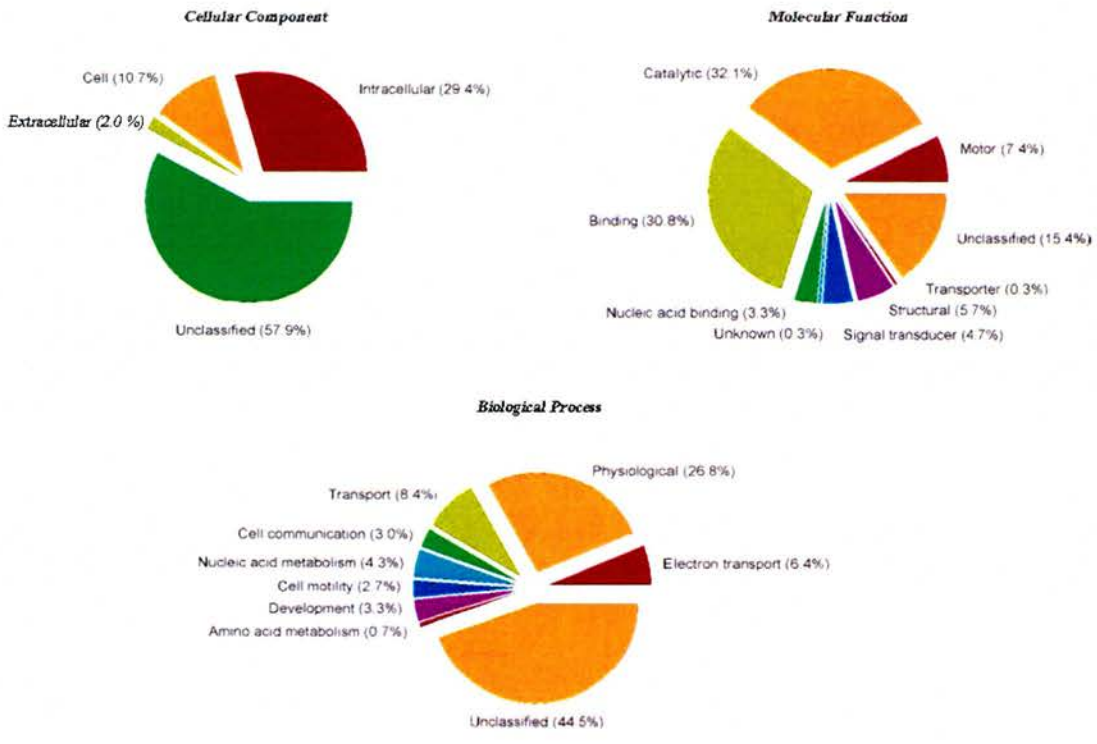


Figure 7.3a Gene ontology (GO) classification of non redundant clusters from the myotube (-) library. Translated sequences corresponding to the clusters were annotated according to GO classification terms. Percentages of the clusters representative of particular category are shown. Clusters referred to as 'unclassified' are those without an associated GO term.

Figure 7. 3b Gene ontology from myotube (+) library

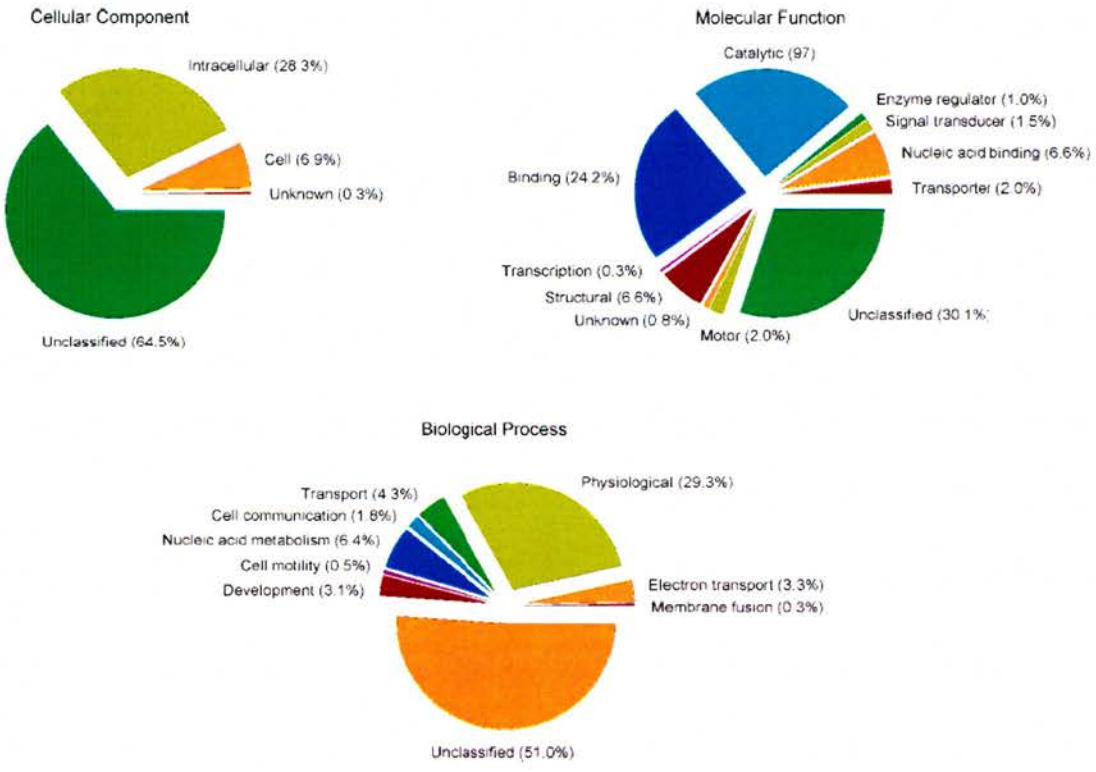


Figure 7. 3b Gene ontology (GO) classification of non redundant clusters from the myotube (+) library. Translated sequences corresponding to the clusters were annotated according to GO classification terms. Percentages of the clusters representative of particular category are shown. Clusters referred to as 'unclassified' are those without an associated GO term.

Figure 7. 4a Tissue expression of strong candidates

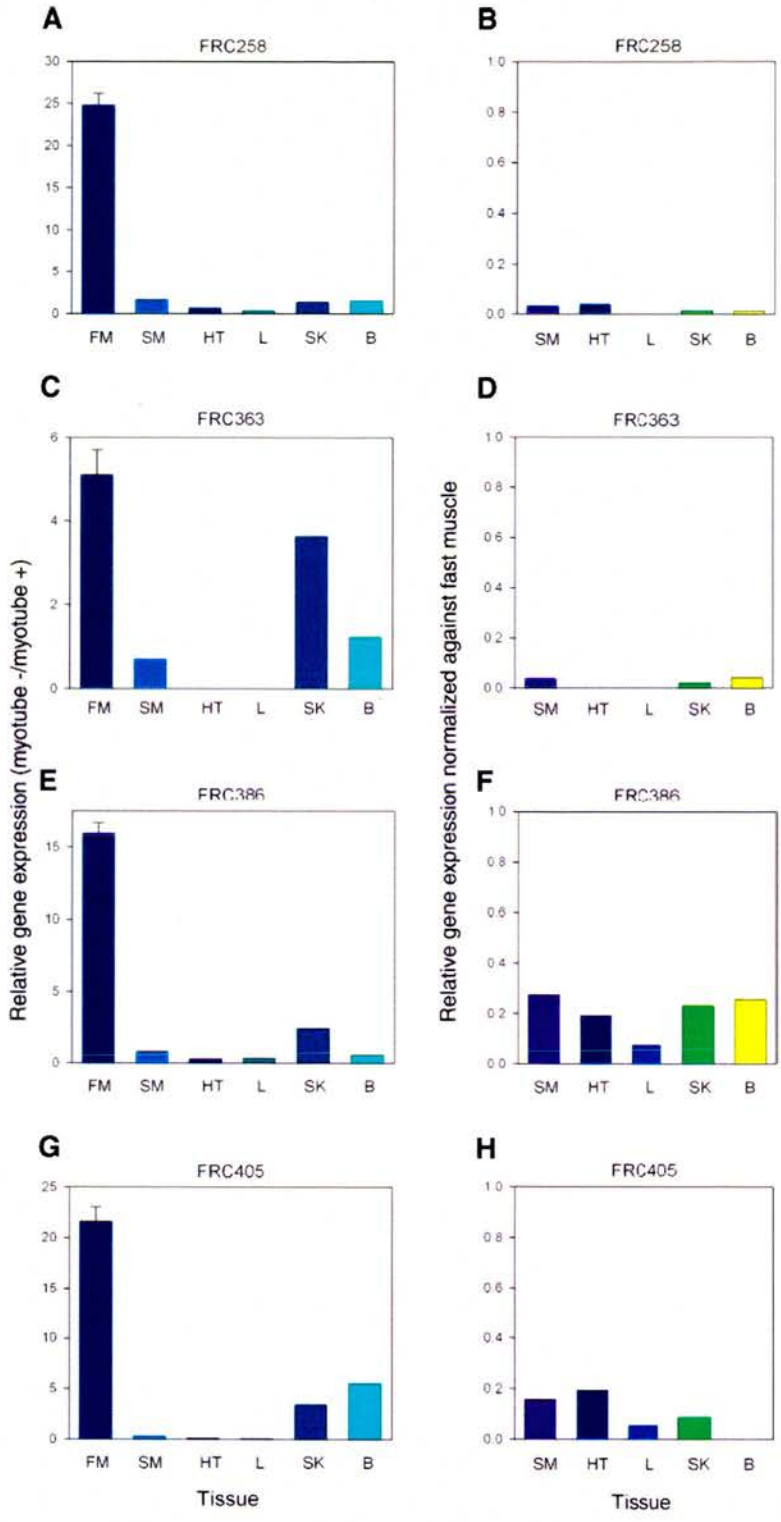


Figure 7. 4a. Relative tissue expression of the 4 strong candidate genes from the myotube (-) library. A, C, E, G: relative gene expression of FRC258, FRC363, FRC386, FRC405 in fast muscle (FM), slow muscle (SM), heart (HT), Liver (L), skin (SK) and brain (B) of *T. rubripes*. Data is represented as a ration of myotube (-)/ myotube (+) average expression normalised against 18s rRNA transcript levels. B, D, F, H: relative special expression of the strong candidates in myotube (-) fish. Mean transcript levels in each tissue was normalised against expression in fast muscle with 18s rRNA as the internal standard. Error bars indicate the standard error (n=4) of expression in fast muscle, all other data is the average of 2 values which did not differ by > 30%.

Figure 7. 4b Tissue expression of intermediate candidates

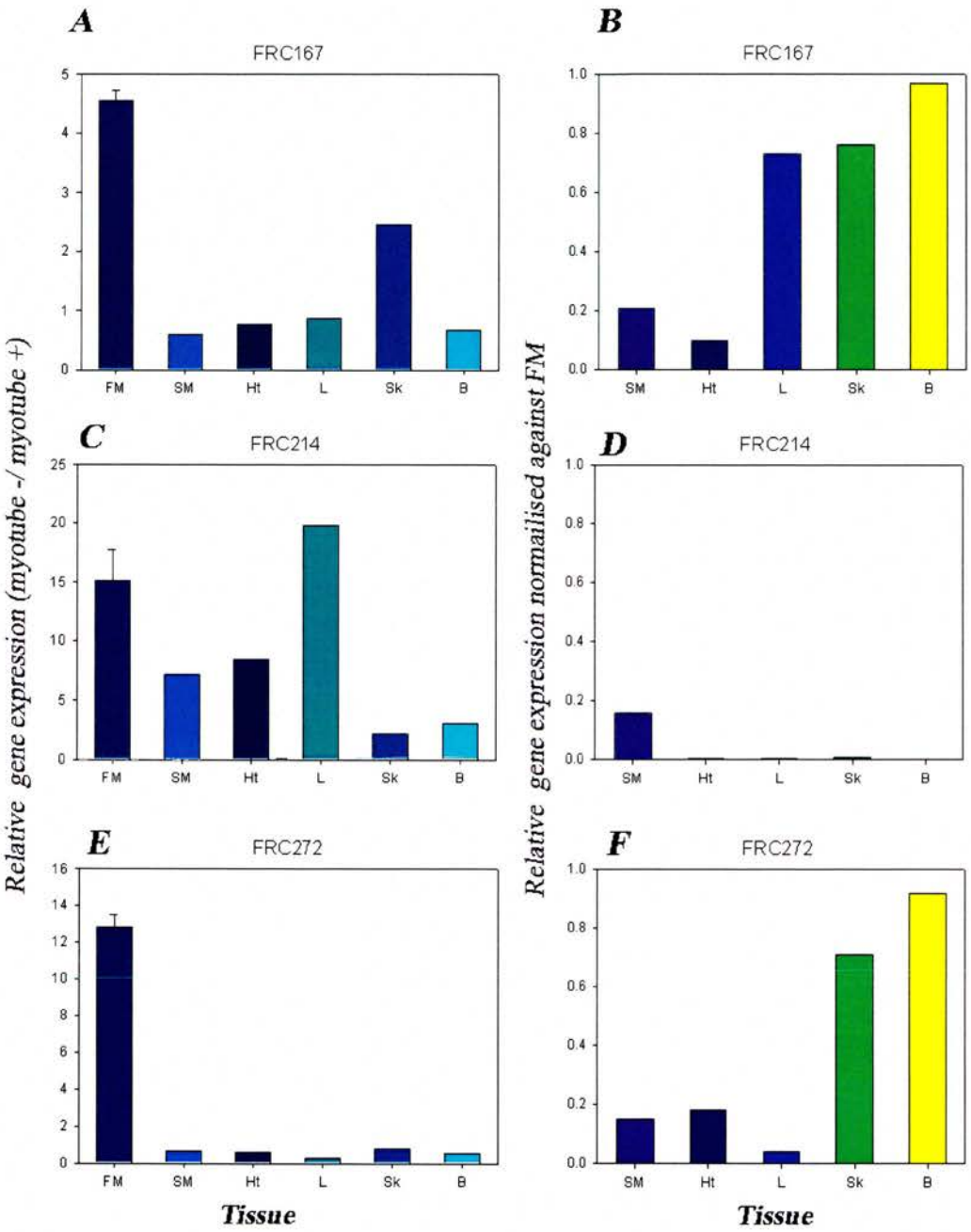


Figure 7. 4b. Relative tissue expression of the 3 intermediate candidate genes from the myotube (-) library. A, C, E: relative gene expression of FRC167, FRC214, FRC272 in fast muscle (FM), slow muscle (SM), heart (HT), Liver (L), skin (SK) and brain (B) of *T. rubripes*. Data is represented as a ration of myotube (-)/ myotube (+) average expression normalised against 18s rRNA transcript levels. B, D, F: relative special expression of the strong candidates in myotube (-) fish. Mean transcript levels in each tissue was normalised against expression in fast muscle with 18s rRNA as the internal standard. Error bars indicate the standard error (n=4) of expression in fast muscle, all other data is the average of 2 values which did not differ by > 30%.

7. 6 Tables

Table 7. 1a Candidate gene information from the myotube (-) library.

Cluster ID	BLASTX result (Accession)	E value	Description
FRC00040	No significant hit		Uncharacterised protein
FRC00102	Nebulin (P20929)	6e-77	Muscle-specific protein
FRC00125	RIKEN cDNA 9830160G03 (Q8BZ52)	1e-37	Predicted Zn-finger protein
FRC00148	Hypothetical protein (XP_215378.1)	3e-28	Uncharacterised haematopoietic stem cells protein
FRC00163	RIKEN cDNA 1700127B04 (O42545)	2e-35	Transcription regulator/nuclease
FRC00167	No significant hit found		Complex 1 protein (LYR family). Unknown function
FRC00181	RIKEN cDNA 4930429H24 (Q9D5K3)	1e-22	Contains BTB/POZ and Kelch domains
FRC00214	Kelch-like 4 isoform 2 (Q9C0H6)	4e-15	Kelch repeat Broad-Complex, possible transcription factor
FRC00220	CD146 (P43121)	2e-25	Adhesion molecule
FRC00258	Hypothetical protein (XP_424765)	5e-12	Potential zinc finger transcription factor
FRC00263	No significant hit		
FRC00272	Expressed sequence AW558560 (Q925I6)	3e-50	Contains a domain found in macrohistone 2
FRC00311	Putative homeodomain transcription factor 2 (Q7TPX6)	2e-43	Putative transcription factor
FRC00321	B-cell receptor-associated protein BAP29 (Q9UHQ4)	3e-17	May be involved in CASP8-mediated apoptosis
FRC00333	Homeobox protein A10b (Q8AWY2)	1e-19	Homeobox protein
FRC00345	Unnamed protein product	1e-05	BTB/POZ and Kelch domains
FRC00358	Reverse transcriptase (BAC82595)	7e-04	LINE-1 reverse transcriptase homologue
FRC00363	No significant hit		
FRC00378	Hypothetical protein zgc77785 (Q6P0T8)	5e-21	Calcineurin-binding protein
FRC00386	CGI-20 protein (Q9Y309)	3e-42	Uncharacterised protein
FRC00391	α-NAC gp220 (P70670)	1e-57	Transcription factor
FRC00405	Kelch-like protein Klhl (Q6Q7X9)	2e-92	Zinc finger protein
FRC00415	RIKEN cDNA 2310038H17 (Q8R2Y7)	3e-34	RNA processing and modification
FRC00429	No significant hit		

Table 7. 1b Candidate gene information from the myotube (+) library

Cluster ID	BLASTX result (Accession)	E value	Description
FRC00001	Ghitm-prov protein (Q8AVX3)	2e-25	Uncharacterised protein
FRC00005	DEAD/H box polypeptide (Q6Y244)	2e-99	ATP- dependent helicase
FRC00034	UNR protein (O75534)	1e-87	RNA-binding protein
FRC00043	Defender against cell death 1 (P46966)	3e-36	Has anti-apoptotic activity
FRC00095	Nsfl1c-prov protein (Q7ZYT4)	2e-44	Contains UBA-like and UBX domains
FRC00098	Sequestosome 1 protein (Q9BUV7)	4e-35	Potentially involved in signal transduction
FRC00149	Hypothetical protein (NP_060749)	2e-13	Uncharacterised protein
FRC00152	Db83 protein (Q803C2)	2e-44	Uncharacterised protein
FRC00173	F-box only protein 32 (NP_957211)	2e-35	Involved in skeletal muscle atrophy
FRC00176	Receptor for activated protein kinase C (O42249)	e-130	Involved in signal transduction
FRC00177	Protein kinase C and casein kinase substrate in neurons protein 3 (Q9UKS6)	2e-26	Regulates of endocytosis
FRC00179	Hypothetical protein HSPC039 (Q9Y5U9)	7e-19	Uncharacterised protein
FRC00193	Leydig cell tumor 10 kDa protein (Q05310)	1e-19	Uncharacterised protein
FRC00194	No significant hit		
FRC00207	Hypothetical PTD015 protein (Q96AQ4)	2e-43	Uncharacterised protein
FRC00212	No significant hit		
FRC00214	No significant hit		Has some similarity to Pax transcription activation domain
FRC00231	S-phase kinase-associated protein 1A (P34991)	4e-62	Involved in the ubiquitin conjugation pathway
FRC00239	No significant hit		
FRC00251	Hypothetical protein MGC51029 (NP_776155)	5e-37	Uncharacterised protein
FRC00254	No significant hit		
FRC00282	Hypothetical protein MGC73223 (NP_957126)	8e-17	Belongs to Spot14 family
FRC00300	1700018O18Rik protein (Q80ZW8)	3e-31	Uncharacterised protein
FRC00327	Casein kinase I alpha (P48729)	2e-66	Participates in Wnt signalling
FRC00344	CGI-49 protein (Q8TAR0)	4e-39	Uncharacterised protein
FRC00347	Hypothetical protein MGC10540 (NP_115729)	4e-21	Uncharacterised protein
FRC00349	Hypothetical protein MGC77169 (Q6P128)	2e-38	Uncharacterised protein
FRC00354	C21orf19-like protein (NP_956088)	6e-13	Uncharacterised protein
FRC00359	SAM domain and HD domain-containing protein 1 (Q60710)	2e-72	Has GTPase activity
FRC00383	Protein C14orf129 homologue (Q7ZWI4)	4e-38	Expressed in $CD34^-$ cells; unknown function
FRC00386	Hypothetical protein MGC73111 (NP_957010)	1e-46	Uncharacterised protein
FRC00400	Ras homologue gene family, member F (NP_061907)	3e-19	Plays a role in controlling cell morphology and cell adhesion
FRC00403	BAG-1 (Q99933)	7e-41	Has anti-apoptotic activity

Table 7. 2a Candidate gene primers sequences from forward subtracted library, myotube (-)

Cluster ID	Forward Primer (5'→3')	Reverse Primer (5'→3')	Size (bp)
FRC00040	CGGGCATCCAGAGGTCAG	ACAAACGAATGATTCCAGGTC	193
FRC00102	CACGTTCACTCTGGATTTTG	CACTCAAGTGGTGGATTCAC	395
FRC00125	GGTACGAGTTTCTCTCCGTG	CCATAGTAGGTGTGCCCAC	390
FRC00148	CAACAAACAGACGCGGCTTC	GTAAAGACAGCCCCAAAGTGG	625
FRC00163	CACGGAAGTGGACTCGTTC	CTAAACAATGTCGCCACCTG	585
FRC00167	CTCACAAGTCATGGCACCG	CTGCCGTAAAAGAGGACACAG	533
FRC00181	GGAGAAGGACTTTCACAGAG	TAGCGCTCCCAGAACCTC	360
FRC00214	AGACGAGCCTCATGTGCC	TGATGAAGGAGATGGAGTGG	238
FRC00220	ACCCCTCCACTGCTGTCAAG	AGCGTCCCGTTGGTCTGC	508
FRC00258	CAATGTCAGGCTGCAGATACC	GCATCTTCAGGACCAGGCT	240
FRC00263	TACCTGTTTGCGTTCTGGAC	TACCTGTTTGCGTTCTGGAC	204
FRC00272	ACGTTTGTTGATACATCCGG	TGCAGTGTGTCTTCATGTCC	535
FRC00311	TACAGCAACACCTCCATCCTG	TGATCTTCCACAGCTTCAGG	268
FRC00321	CATCATGCGTCGGATATTCAG	TGTTTCTTCATGGCCTCCAC	276
FRC00333	ACTTTACCCTGCAACAATCTG	GCTCAGGCAGTCTCAGGAG	506
FRC00345	CTTCGCCATGATGCCCAG	TCCGAGTTCCTCAGAGTCCAG	199
FRC00358	GGTAACCATGAGGAGGCAGC	GTCCTTCCATGATAGCCCCTC	433
FRC00363	ACAGATCTGGATGGTATGGC	TCTGGTCTTGTGGTCTTTCC	499
FRC00378	GAGTCAGACCCTGAAGCCTC	GGCATTTCTGGGTGGATC	398
FRC00386	TGCTGTCTTCCTGGGCCTC	TGGTGCTTCCGTGGTGTTC	338
FRC00391	GAATGATAAGGGGTCCGG	GTACTGTTGGCGTCTGTGTG	428
FRC00405	AGGATGTTCTGTTGGTATGGC	TGGCACCTTAGTCAAGATGTG	456
FRC00415	GCAACCTTAAAGCAAACGTG	CATGTGACCCATTAGAAGGC	398
FRC00429	TGCACATTGTGGAGAACGG	TGGCATCAGAGGAGAGGAG	257

Table 7. 2b Candidate gene primers sequences from reverse subtracted library, myotube (+)

Cluster ID	Forward Primer (5'→3')	Reverse Primer (5'→3')	Size (bp)
FRC00001	ATGTTTGTGGCGAGGCTGAC	CACGCACATGTAGGTGGAGTG	399
FRC00005	GGCGTATTCGTGCTCTTATCC	ACCTCTCCTCCTGTGCCATG	459
FRC00034	AGGTTACCAAAGTGGCTCCAG	CTGTGCGCTGGTTGAGGAC	450
FRC00043	ATGTCGAACTCCGTAATATCGG	GGCCTGTAGATTGTTGTGAGC	453
FRC00095	CAAGATGTCCACGTGGTGCTG	GTACCGTCTGCCAATCGAATC	386
FRC00098	TTCAGAATGAGAGGGGTGAGG	TACTGGATGGCATCAAGGGC	392
FRC00149	ATGTCTGCGGTGAAATCCG	CCAGAGTAGGCACTGATCGTG	166
FRC00152	CAGCCATTATTTCCAAAGATGG	CTTCTTGAAGAGCTTGTGCCAG	292
FRC00173	GGAAGAGGGATTCTGCTTGAC	GTACAATAGCCGTGGCGC	381
FRC00176	TGCTCTGTCTGGAGCTTGGG	AGGTGCTTGCCCTCATTCAG	450
FRC00177	ATGTCCATCAACGAGCAAGATG	GTAGAAGCCCTCCTTCCCTCC	312
FRC00179	CCTTTGATTGCAGTGAACTCG	GTGGCCCATGACATCAAATTAG	179
FRC00193	CTCAGCGCCCAGGAATATC	GCCACTTGTGATCATCAGCTC	324
FRC00194	GGGCAGGTACAGCACTTTCATC	TCCCTGGTGGATGATGAGAAC	490
FRC00207	GGACTGTAAAGTCTGGCCTGG	TCAGCAGGTGGAATGGAAGAC	298
FRC00212	AGATGCAATGGTTCTGGCTATG	GGCTGAGCTTCTGCGAAATG	331
FRC00214	GAACAACCCCAGCAGCAGC	TGTCTGGGACCTTGACCATTG	405
FRC00231	GATGCCCACGATAAAGCTGC	GGTCTTCCTGATCTCCTCGGG	415
FRC00239	CGGGCTGTCCCTGTTGTTC	CCTCTGGAGGTTCTTTCACTGC	596
FRC00251	GTGTTCCGAAATGGCGAGTC	CATATGCCCTCATCTTCTGCC	511
FRC00254	GTACAGGCCCAAGTACTGCG	GAATGCTGACGTATCCACAGC	410
FRC00282	GATGATGCAGCTCCTAGAAACC	TCTCCTTGTCCTCCTCCAGC	146
FRC00300	CATCCAGGGGCAGATCGTCG	GGATCCGACACGGCAGAACTC	247
FRC00327	GGAGCACCAACGAGTCCACG	GGTCTGCCAGCATCAGCACTG	487
FRC00344	AAGCTCAAACGCAGGGGTGC	GAGCAGGTTACGTCCAAAACGG	246
FRC00347	CCAGAAATGGGATGGTCAACG	ACCACCAGTTGCCTGACCAGC	257
FRC00349	TGTGATGGCTTGTGGAGCTAC	CCTGGTAGAGCTAAGCAACTGG	312
FRC00354	GGAGCTGAGGAAGTCGGGC	GGGCTGGATCAGCTCATGG	175
FRC00359	CTCTTCTTCCCCGGAACTGG	GCTGGACCATTCGAGCTTCAG	498
FRC00383	CGTCTCCCAAATGCACGTG	GCACAAACCACCAATCCTGAG	501
FRC00386	CAAACCACCCAGCACAGATTC	ACTGTGTCCAGCCACTGTCTTC	384
FRC00400	GGTTGCAAGACTGACCTCAGG	GCAGATTTCGACGCTAGTGG	562
FRC00403	AGCGGAACAGTCCAGAAGAAG	ATAGGACATGGCGAGGAAGAC	407

Table 7. 3 Candidates consistently and significantly differentially expressed in fast muscle of *T. rubripes* at two distinct growth phases.

Cluster ID	*Library*	*GenBank Accession No.*	*BLASTX Hit (GenBank Accession No.)*	*Conserved Domains (CDD Accession No.)*
FRC00040	Myotube (–)	**CK829265**	No significant hit found	No putative domains detected
FRC00167	Myotube (–)	**CK829482**	*Mus musculus* 4930469P12Rik protein (AAH21522)	LYR family (pfam05347)
FRC00214	Myotube (–)	**CK829703**	*Pan troglodytes* hypothetical gene (AK044523)	Kelch (smart00612)
FRC00258	Myotube (–)	**CK829660**	*Gallus gallus* hypothetical protein (XP_424765)	SPla and RYanodine receptor (smart00449)
FRC00263	Myotube (–)	**CK829990**	*Tetraodon nigroviridis* unnamed protein product (CAF96211)	No putative domains detected
FRC00272	Myotube (–)	**CK829695**	*Mus musculus* expressed sequence AI314976 (NP_997102)	Appr-1-p processing enzyme (smart00506)
FRC00363	Myotube (–)	**CK829877**	No significant hit found	BTB/Kelch (IP000210)
FRC00386	Myotube (–)	**CK829928**	*Gallus gallus* CGI-20 protein (XP_414758)	No putative domains detected
FRC00405	Myotube (–)	**CK829961**	*Danio rerio* kelch-like protein Klhl (AAS84610)	BTB/Kelch (IP000210)
FRC00177	Myotube (+)	**CK830424**	*Homo sapiens* protein kinase C and casein kinase substrate in neurons 3 (NP_057307)	Src homology 3 (CD00174)
FRC00300	Myotube (+)	**CK830205**	*Mus musculus* 1700018O18Rik protein (AAH46793)	Na^+/melibiose symporter (COG2211)

Chapter 8: General discussion

8.1 Main Project findings

This thesis describes the identification and characterisation of the muscle regulatory factors (Myf5, MyoD, Myogenin and Myf6), Sox8, Mstn-1, Mstn-2, follistatin and candidate genes involved in the cessation of myotube production in puffer fish. The main findings were *myogenin*, *Mstn-1* and *follistatin* transcript levels during development varying with temperature. Sox8 was identified as a possible marker of myogenic progenitor cells (MPCs) in teleosts. In addition four genes, not previously characterised in fish, were identified as potential candidates involved in the cessation of myotube production in teleosts.

8.2 Functional intronic and intergenic sequences

Brenner et al (1993) revealed the *T. rubripes* genome to be approximately an 8^{th} the size of the human genome and with less than 10% repetitive DNA. They also showed both *T. rubripes* and human to have approximately the same amount of coding sequence. An explanation for the small genome may therefore lie in the intergenic and intronic regions. The puffer fish genes described in this study all show a large reduction in intron size compared to mammalian sequences. Untranslated regions (UTRs) of DNA in the past have been referred to as "junk DNA" without a function, however they have been noted to contain important regulatory functions such as enhancers and silencers for example as observed in the expression of *regA* in somatic cell differentiation (*Volvox carteri*) (Stark et al., 2001). Cis-regulatory sequences are non-coding elements of DNA consisting of multiple transcription

factor binding sites that act in combination influencing gene expression (Arnonone and Davidson 1997). McEwen et al, (2006) have recently carried out a large study identifying 124 families of duplicated elements corresponding to cis-regulatory sequences when comparing the *T. rubripes* and human genomes. 3' UTRs have also been studied and conserved regions similar to cis-motifs found in promoter regions have been identified (Hobert 2004). These binding sites are targets of short interfering RNA (siRNA) suggesting post-transcriptional control of gene expression through mRNA degradation or inhibition of transcription via 3'UTR (Brennecke et al., 2003; Heinrich and Pagtakhan 2004; Roca et al., 2006). Interestingly binding sites for reporter genes in the 3' UTR of the mammalian TSP-1 (Thrombospondin-1) mRNA have been identified with a function to increase the stability of the transcript in response to heat shock (Kang et al., 2006). A subset of conserved motifs termed "pyknons" have recently been found in the UTRs and coding regions of almost all known human genes with properties suggesting a link between coding and non coding DNA (Rigoutsos et al., 2006). In addition to UTR several intronic functions such as intron dependent gene expression (Callis et al., 1987), enhancement of protein stability (Noe et al., 2003), influencing accurate polyadenylation and mRNA formation (Collis et al., 1990; Nesic et al., 1993) and increasing the stability of transcripts in the nucleus (Ryu and Mertz 1989) have been identified. Some genes undergo the phenomenon of intron loss and this has been identified in puffer fish as well as some plant species (genus *Gossypium* and *Leavenworthia*) and mould (*Trichoderma viride*). Such genes include alcohol dehygrogenase (*Adh*) (Charlesworth et al., 1998; Small and Wendel, 2000), silent antifungal protein (*AFP*)-like gene (Hao et al., 2000), *Rhodopsin* and *SART1* (Fitzgibbon et al., 1995; Venkatesh et al., 1999). *Rhodopsin* in particular completely lacks introns in certain fish but several intronless genes have been reported in mammals. Therefore although some introns are removed during evolution giving rise to the "junk DNA" hypothesis, introns and UTR also have important

functions suggesting they contain both useless DNA as well as functional sequences. The reduction in intron size and intergenic regions of *T. rubripes* genes compared to mammalian sequences suggest puffer fish may have lost the majority of "junk DNA" and retained the functional aspects of the intronic sequences. The *T. rubripes* genome will greatly assist research in the area of intronic functions and fish-mammal genomic comparisons have already proved to be a powerful tool in identifying the functional and conserved aspects of non coding elements without the noise of the un-functional DNA (McEwan et al., 2006).

8.3 Changes in mRNA levels in relation to temperature during development.

Environmental conditions including incubation temperature are known to impact all aspects of development with a potential influence on somatic growth, muscle morphology, physiology and even mortality in developing teleosts (Johnston and Hall 2004). Conflicting evidence between species leads to some confusion as to how levels of transcription may influence the phenotypic changes observed as a result of temperature plasticity. Xie et al (2001) indicated a more advanced state of muscle differentiation in trout embryos reared at higher temperatures with an elevation in *myogenin* and myosin heavy chain transcript at the eye stage. Temple at al (2001) identified changes in *myogenin* expression level but no significant effect on the timing of expression in Atlantic herring. In agreement with Temple et al (2001) we have identified a significant change in expression level of *myogenin*, *FMstn-1* and *follistatin* with respect to temperature during early somitogenesis. Due to the lack of developmental stages sampled a relative change in timing of expression was difficult to identify. *Myogenin* had a 3-fold increase in expression at 21°C compared to 15°C at the start of somitogenesis. This may be explained as a mechanism for the cells to cope with an

increase in mRNA breakdown as mRNAs are less stable at higher temperatures. Perhaps unexpectedly then, the opposite was true with *Mstn-1* and *Flstn* expression during early somitogenesis significantly higher at the lower temperature of 15°C. *Sox8* revealed no change in relative expression throughout development with respect to temperature. The apparent redundancy between different genes suggests that temperature may only affect specific developmental pathways. The fact *myogenin*, *Mstn-2* and *follistatin* have major roles in promoting and inhibiting myogenic differentiation suggests these changes in transcription levels have the potential to alter the timing of muscle differentiation and therefore the time window for myoblast proliferation. Johnston et al (unpublished) identified a 20% higher fibre number in *T. rubripes* larvae reared at 15°C and 18°C compared to the higher temperature of 21°C. This maybe explained by a reduced cell cycle window resulting in less proliferation and a higher proportion of cells exiting the cell cycle into the differentiation programme. We identified an acceleration of development at 21°C consistent with this. The higher *myogenin* expression at elevated temperatures implies an increase in differentiated cells with a higher proportion of cells exited from the cell cycle. Contrary to this the higher *Mstn-1* expression at the lower temperature suggests an increase in the inhibition of myogenic proliferation not consistant with the higher fibre number at 15°C. A possible explanation for this lies in the mRNA levels of *follistatin* a known inhibitor of Mstn (Amthor et al, 2003). We identified an increase in transcription of *follistatin* at 15°C more so than the rise in *Mstn-1* expression. This proposes a mechanism to counteract the higher levels of *Mstn* and normalise the system. Ideally functional expression studies where the genes are overexpressed in myogenic cell lines would be required to establish the mechanisms.

The redundancy of the myogenic transcriptional networks makes these results difficult to interpret and, despite the significance, mRNA levels alone are insufficient to explain the

nature of the biological systems that ultimately rely on protein levels (Lee et al, 2003). It would be of interest to investigate the abundance of protein in the systems and correlate this with mRNA levels and indeed the ultimate phenotype. Several mechanisms have been employed to estimate specific protein levels including isotope-coded affinity tag system (ICAT) in conjunction with mass spectrometry (Gygi et al., 1999) and 2 dimensional gel electrophoresis coupled with mass spectrometry (Lahm and Langen 2000). Another important aspect to bridge the gap between transcribed mRNA and translated protein levels is to investigate the ribosomal involvement (Zong et al., 1999). Actively translated mRNAs are associated with multiple ribosomes forming structures known as polysomes whereas translationally inactive mRNAs are coupled with individual ribosomes (monosomes) or sequestered in messenger ribonucleoprotein (mRNP) particles. Levels of these structures may reveal an operational distinction between translated and untranslated mRNAs (Ruan et al., 1997; Zong et al., 1999) and would be of use to isolate the functional relevance, if any, related to the changes in mRNA levels we identified as a result of developmental temperature.

8.4 Control of muscle fibre number

Perhaps the most significant result presented in this study is the identification of four genes we propose to have a role in the inhibition of muscle fibre recruitment in teleosts. The genetic mechanism by which fish cease recruiting fibres at ~40% body length (Weatherley et al., 1988) is unknown and these initial results provide a platform from which future studies can be based. The zebrafish is an ideal model to investigate biological systems with successful gene knock-out mutants (Amali et al., 2005), gene silencing by RNAi (Acosta et al., 2005) and the culture of embryonic stem cells (Fan et al., 2004) all attainable. These techniques are

essential if the function of the candidates is to be accurately identified. The transcripts do show expression in other tissues suggesting a role in other biological systems and the potential to result in non-viable homozygous knock outs. Perhaps the most suitable functional study for these genes would be in cell culture. The over expression of these genes in zebrafish myogenic cell lines should present a model for investigating their role in myoblast fusion and myotube production. It will also be of interest to investigate the interactions of these candidates with other proteins as protein-protein interactions play a key role in many biological systems. Yeast two-hybrid screens are a common method of identifying protein-protein interactions (Uetz and Hughes 2000) and still used in mammalian systems as recently described by Liu et al (Liu et al., 2006). There are also a number of computational techniques for identifying protein-protein interactions based on predictions (Qi et al., 2006). As the four candidates all have orthologues in the human genome it is plausible to suggest they have a similar role in the cessation of muscle fibre production in mammals. Muscle growth in mammals is very different to teleosts with the final muscle fibre number characterised at birth (Rowe and Goldspink 1969) however various stimuli including exercise (Darr and Schultz 1987), stretch (Rosenblatt et al., 1994) and injury (Blaveri et al., 1999) can trigger myotube formation. The expression of the genes identified from the subtracted libraries in response to these stimuli will be of interest. They may for example show down-regulation in response to injury suggesting one system potentially controlling myotube formation in growth and regeneration across the vertebrate population.

8.5 Final Words

The sequencing of the *T. rubripes* and *T. nigroviridis* genomes has been an invaluable resource for this work assisting in the identification of gene orthologues and the putative characterisation of the candidate ESTs identified from the subtracted library. The genes characterised have provided information on transcriptional regulation during development and growth. However, to address the functional aspects of the biological system the recognition of key translational features such as ribosome activity, mRNA and protein degradation are essential.

The results presented draw us to two important conclusions: (i) The transcription of some myogenic factors in teleosts is susceptible to changes in environmental temperature during development and may be responsible for altering the rate of transition between proliferation and differentiation during myogenesis; (ii) The process of myotube formation is regulated by an inhibitory pathway and the putative proteins of the candidate genes identified are likely to be involved in protein-protein interactions, adhesion, cell migration and transcriptional regulation.

Chapter 9: References

Acosta J., Carpio Y., Borroto I., Gonzalez O. and P. E. M. (2005). *Myostatin* gene silenced by RNAi show a zebrafish giant phenotype. *Journal of Biotechnology* **119**(4): 324-331.

Agarkova I., Ehler E., Lange S., Schoenauer R. and Perriard J.-C. (2003). M-band: a safeguard for sarcomere stability? *Journal of Muscle Research and Cell Motility* **24**(2 - 3): 191-203.

Agarkova I. and Perriard J.-C. (2005). The M-band: an elastic web that crosslinks thick filaments in the center of the sarcomere. *Trends in Cell Biology* **15**(9): 477-485.

Alberts B., Bray D., Johnson A., Lewis J., Raff M., Roberts K. and Walter P. (1998). Essential Cell Biology, An introduction to the molecular biology of the cell. New York, Garland Publishing, Inc.

Alexander R. M. (1969). The orientation of muscle fibers in the myomeres of fishes. *Journal of the Marine Biological Association of the United Kingdom.* **49**: 263–290.

Altschul S. F., Madden, TL, Schaffer A., Zhang J., Zhang Z., Miller W. and Lipman D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17): 3389-3402.

Amali A. A., Lin C. J., Chen Y. H., Wang W. L., Gong H. Y., Lee C. Y., Ko Y. L., Lu J. K., Her G. M., Chen T. T. and Wu J. L. (2005). Up-regulation of muscle-specific transcription factors during embryonic somitogenesis of zebrafish (Danio rerio) by knock-down of myostatin-1. *Developmental Dynamics* **229**(4): 847-856.

Amores A., Force A., Yan Y.-L., Joly L., Amemiya C., Fritz A., Ho R. K., Langeland J., Prince V., Wang Y.-L., Westerfield M., Ekker M. and Postlethwait J. H. (1998). Zebrafish hox Clusters and Vertebrate Genome Evolution. *Science* **282**(5394): 1711-1714.

Amthor H., Christ B., Rashid-Doubell F., Kemp C. F., Lang E. and Patel K. (2002). Follistatin Regulates Bone Morphogenetic Protein-7 (BMP-7) Activity to Stimulate Embryonic Muscle Growth. *Developmental Biology* **243**(1): 115-127.

Amthor H., Christ B., Weil M. and Patel K. (1998). The importance of timing differentiation during limb muscle development. *Current Biology* **8**(11): 642-652.

Amthor H., Connolly D., Patel K., Brand-Saberi B., Wilkinson D. G., Cooke J. and Christ B. (1996). The Expression and Regulation of Follistatinand a Follistatin-like Gene during Avian Somite Compartmentalization and Myogenesis. *Developmental Biology* **178**(2): 343-362.

Amthor H., Nicholas G., McKinnell I., Kemp C. F., Sharma M., Kambadur R. and Patel K. (2004). Follistatin complexes Myostatin and antagonises Myostatin-mediated inhibition of myogenesis. *Developmental Biology* **270**(1): 19-30.

Aparicio S., Chapman J., Stupka E., Putnam N., Chia J. M., Dehal P., Christoffels A. and Rash S., Hoon, S., Smit, A., Gelpke, M. D., Roach, J., Oh, T., Ho, I. Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S. F., Clark, M. S., Edwards, Y. J., Doggett, N., Zharkikh, A., Tavtigian, S. V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y. H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science.* **297**(5585): 1301-10.

Arber S., Halder G. and Caroni P. (1994). Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. *Cell* **79**(2): 221-231.

Arnonone M. I. and Davidson E. H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**(10): 1851-64.

Atchley W., Fitch W. and Bronner-Fraser M. (1994). Molecular Evolution of the MyoD Family of Transcription Factors. *PNAS* **91**(24): 11522-11526.

Bai C., Richman R. and Elledge S. J. (1994). Human cyclin F. *EMBO J.* **13**(24): 6087-98.

Bardwell V. J. and Treisman R. (1994). The POZ domain: a conserved protein-protein interaction motif. *Genes and Development* **8**(14): 1664-77.

Barker D. J. P. (1992). Fetal and infant origins of adult disease. London, British Medical Journal Publishing.

Barker D. J. P. (1994). Mothers, babies and disease in later life. London, British Medical Journal Publishing.

Barker D. J. P. and Clark P. M. (1997). Fetal undernutrition and disease in later life. *Reviews of Reproduction* **2**: 105-112.

Barresi M. J. F., D'Angelo J. A., Hernandez L. P. and Devoto S. H. (2001). Distinct mechanisms regulate slow-muscle development. *Current Biology* **11**(18): 1432-1438.

Bauer H., Meier A., Hild M., Stachel S., Economides A., Hazelett D., Harland R. M. and Hammerschmidt M. (1998). Follistatin and Noggin Are Excluded from the Zebrafish Organizer. *Developmental Biology* **204**(2): 488-507.

Bell K. M., Western P. S. and Sinclair A. H. (2000). SOX8 expression during chick embryogenesis. *Mechanisms of Development* **94**(1-2): 257-260.

Benezra R., Davis R. L., Lockshon D., Turner D. L. and Weintraub H. (1990). The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61**(1): 49-59.

Beranger F., Mejean C., Moniot B., Berta P. and Vandromme M. (2000). Muscle Differentiation Is Antagonized by SOX15, a New Member of the SOX Protein Family. *Journal of Biological Chemistry* **275**(21): 16103-16109.

Bergstrom D. A. and Tapscott S. J. (2001). Molecular Distinction between Specification and Differentiation in the Myogenic Basic Helix-Loop-Helix Transcription Factor Family. *Molecular and Cell Biology* **21**(7): 2404-2412.

Black B. L. and Olson E. N. (1998). Transcriptional control of muscle development by Myocyte enhancer factor-2 (MEF2) proteins. *Annual Review of Cell and Developmental Biology* **14**(1): 167-196.

Blagden C. S., Currie P. D., Ingham P. W. and Hughes S. M. (1997). Notochord induction of zebrafish slow muscle mediated by Sonic hedgehog. *Genes and Development* **11**: 2163-2175.

Blaveri K., Heslop L., Yu D. S., Rosenblatt J. D., Gross J. G., Partridge T. A. and Morgan J. E. (1999). Patterns of repair of dystrophic mouse muscle: Studies on isolated fibers. *Developmental Dynamics* **216**(3): 244-256.

Boddeke R., Alijper, E. J., Van der Stelt, A. (1959). Histological characteristics of the body musculature of fishes in connection with their way of life. *Proceedings of the Koninklijke Nederlandskie Akademie van Wetenschappen Series C: Biological and medical sciences* **62**: 576-588.

Bois P. R. and Grosveld G. C. (2003). FKHR (FOXO1a) is required for myotube fusion of primary mouse myoblasts. *EMBO J.* **22**(5): 1147-57.

Bokdawala F. D. and George J. C. (1967). A histochemical study of the red and white muscles of the carp, *Cirrhina mrigala*. *Journal of Animal Morphology and Physiology* **14**: 60-68.

Bone Q. (1964). Patterns of innervation in the lower chordates. *International Review of Neurobiology* **6**: 99-147.

Bone Q. (1975). Muscular and energenic aspects of fish swimming. *Swimming Flying Nature* **2**: 494-258.

Bone Q. (1978). Locomotor muscle. *Fish Physiology* **7**: 361-424.

Bowles J., Schepers G. and Koopman P. (2000). Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. *Developmental Biology* **227**(2): 239-255.

Brainerd E. L., Slutz S. S. and Hall E. K. (2001). Patterns of genome size evolution in Tetraodontiform fishes. *Evolution* **55**(11): 2363-2368.

Brand-Saberi B. (2005). Genetic and epigenetic control of skeletal muscle development. *Annals of Anatomy* **187**(3): 199-207.

Breitbart R. E., Liang C. S., Smoot L. B., Laheru D. A., Mahdavi V. and Nadal-Ginard B. (1993). A fourth human MEF2 transcription factor, hMEF2D, is an early marker of the myogenic lineage. *Development* **118**(4): 1095-1106.

Brennan T., Chakraborty T. and Olson E. (1991). Mutagenesis of the Myogenin Basic Region Identifies an Ancient Protein Motif Critical for Activation of Myogenesis. *PNAS* **88**(13): 5675-5679.

Brennecke J., Hipfner D. R., Stark A., Russel R. B. and Cohen S. M. (2003). Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**(1): 25-36.

Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B. and Aparicio S. (1993). Characterisation of the pufferfish (*Fugu*) genome as a compact vertebrate genome. *Nature* **366**(6452): 265-8.

Brosius J. (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* **107**(1 - 3): 209-238.

Brown J. H. and Cohen C. (2005). Regulation of Muscle Contraction by Tropomyosin and Troponin: How Structure Illuminates Function Advances in Protein Chemistry, Academic Press: 121-159.

Buchberger A., Ragge K. and Arnold H. H. (1994). The *myogenin* gene is activated during myocyte differentiation by pre- existing, not newly synthesized transcription factor MEF-2. *Journal of Biological Chemistry* **269**(25): 17289-17296.

Buckingham M. (2001). Skeletal muscle formation in vertebrates. *Current Opinion in Genetics & Development* **11**(4): 440-448.

Burge C. and Karlin S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**(1): 78-94.

Callis J., Fromm M. and Walbot V. (1987). Introns increase gene expression in cultured maize cells. *Genes and Development* **1**(10): 1183-200.

Carr C. S. and A. S. P. (1990). A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. *Molecular and Cell Biology* **10**(8): 4384-8.

Champagne F. A., Francis D. D., Mar A. and Meaney M. J. (2003). Variations in maternal care in the rat as a mediating influence for the effects of environment on development. *Physiology & Behavior* **79**(3): 359-371.

Charge S. B. P. and Rudnicki M. A. (2004). Cellular and Molecular Regulation of Muscle Regeneration. *Physiol. Rev.* **84**(1): 209-238.

Chauvigne F., Ralliere C., Cauty C. and Rescan P. Y. (2006). In situ hybridisation of a large repertoire of muscle-specific transcripts in fish larvae: the new superficial slow-twitch fibres exhibit characteristics of fast-twitch differentiation. *Journal of Experimental Biology* **209**(2): 372-379.

Chen Y. H., Lee W. C., Cheng C. and Tsai H. J. (2000). Muscle regulatory factor gene: zebrafish (*Danio rerio*) *myogenin* cDNA. *Comparative Biochemistry and Physiology. Part B, Biochemistry and Molecular Biology* **127**(1): 97-103.

Chen Y. H., Lee W. C., Liu C. F. and Tsai H. J. (2001). Molecular structure, dynamic expression, and promoter analysis of zebrafish (*Danio rerio*) *myf-5* gene. *Genesis* **29**(1): 22-35.

Cheng J. and Blickhan R. (1994). NOTE ON THE CALCULATION OF PROPELLER EFFICIENCY USING ELONGATED BODY THEORY. *Journal of Experimental Biology* **192**(1): 169-177.

Choi J., Costa M. L., Mermelstein C. S., Chagas C., Holtzer S. and Holtzer H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proceedings of the National Academy of Sciences of the United States of America* **87**(20): 7988-7992.

Christensen M., Oksbjerg N., Henckel P. and Jorgensen P. F. (2000). Immunohistochemical examination of myogenesis and expression pattern of myogenic regulatory proteins (myogenin and myf-3) in pigs. *Livestock Production Science* **66**(2): 189-195.

Christoffels A., Koh E. G. L., Chia J. M., Brenner S., Aparicio S. and Venkatesh B. (2004). Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. *Molecular Biology and Evolution* **21**(6): 1146-1151.

Christy B. A., Sanders L. K., Lau L. F., Copeland N. G., Jenkins N. A. and Nathans D. (1991). An Id-Related Helix-Loop-Helix Protein Encoded by a Growth Factor-Inducible Gene. *PNAS* **88**(5): 1815-1819.

Cole N. J., Hall T. E., Martin C. I., Chapman M. A., Kobiyama A., Nihei Y., Watabe S. and Johnston I. A. (2004). Temperature and the expression of myogenic regulatory factors (MRFs) and myosin heavy chain isoforms during embryogenesis in the common carp *Cyprinus carpio L. Journal of Experimental Biology* **207**(24): 4239-4248.

Collis P., Antoniou M. and Grosveld F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* **9**(1): 233-40.

Conlon R., Reaume A. and Rossant J. (1995). Notch1 is required for the coordinate segmentation of somites. *Development* **121**(5): 1533-1545.

Cossins A. R. and Bowler K. (1987). Temperature Biology of Animals. *Cambridge: Cambridge University Press*: 339.

Coutelle O., Blagden C. S., Hampson R., Halai C., Rigby P. W. J. and Hughes S. M. (2001). Hedgehog Signalling Is Required for Maintenance of *myf5* and *myoD* Expression and Timely Terminal Differentiation in Zebrafish Adaxial Myogenesis. *Developmental Biology* **236**(1): 136-150.

Currie P. D. and Ingham P. W. (1998). The generation and interpretation of positional information within the vertebrate myotome. *Mechanisms of Development* **73**(1): 3-21.

Darr K. C. and Schultz E. (1987). Exercise-induced satellite cell activation in growing and mature skeletal muscle. *Journal of Applied Physiology* **63**(5): 1816-1821.

Davis R. L., Cheng P. F., Lassar A. B. and Weintraub H. (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60**(5): 733-46.

Dean D. C., Bowlus C. L. and Bourgeois S. (1987). Cloning and Analysis of the Promoter Region of the Human *Fibronectin* Gene. *PNAS* **84**(7): 1876-1880.

Dedieu S., Poussard S., Mazeres G., Grise F., Dargelos E., Cottin P. and Brustis J. J. (2004). Myoblast migration is regulated by calpain through its involvement in cell attachment and cytoskeletal organization. *Experimental Cell Research* **292**(1): 187-200.

Dekens M. P. S., Santoriello C., Vallone D., Grassi G., Whitmore D. and Foulkes N. S. (2003). Light Regulates the Cell Cycle in Zebrafish. *Current Biology* **13**(23): 2051-2057.

Delalande J. M. and Rescan P. Y. (1998). Expression of a cysteine-rich protein (CRP) encoding gene during early development of the trout. *Mechanisms of Development* **76**(1-2): 179-83.

Delalande J. M. and Rescan P. Y. (1999). Differential expression of two nonallelic *MyoD* genes in developing and adult myotomal musculature of the trout (*Oncorhynchus mykiss*). *Development Genes And Evolution* **209**(7): 432-437.

Devoto S., Melancon E., Eisen J. and Westerfield M. (1996). Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development* **122**(11): 3371-3380.

Devoto S. H., Stoiber W., Hammond C. L., Steinbacher P., Haslett J. R., Barresi M. J. F., Patterson S. E., Adiarte E. G. and Hughes S. M. (2006). Generality of vertebrate developmental patterns: evidence for a dermomyotome in fish. *Evolution & Development* **8**(1): 101-110.

Deweindt C., Albagli O., Bernardin F., Dhordain P., Quief S., Lantoine D., Kerckaert J. P. and Leprince D. (1995). The LAZ3/BCL6 oncogene encodes a sequence-specific transcriptional inhibitor: a novel function for the BTB/POZ domain as an autonomous repressing domain. *Cell Growth and Differentiation* **6**(12): 1495-1503.

Dickson G., Peck D., Moore S. E., Barton C. H. and Walsh F. S. (1990). Enhanced myogenesis in NCAM-transfected mouse myoblasts. *Nature* **344**(6264): 348-51.

Dietrich F. S., Voegeli S., Brachat S., Lerch A., Gates K., Steiner S., Mohr C., Pohlmann R., Luedi P., Choi S., Wing R. A., Flavier A., Gaffney T. D. and Philippsen P. (2004). The Ashbya gossypii Genome as a Tool for Mapping the Ancient *Saccharomyces cerevisiae* Genome. *Science* **304**(5668): 304-307.

Dietz T. J. and Somero G. N. (1992). The threshold induction of the 90-kDa heat shock protein is subject to acclimatization in eurythermal goby fishes (genus Gillichthys). *Proceedings of the National Academy of Sciences of the United States of America* **89**(8): 3389-93.

Dirik L., Bohm T. and Nasmyth K. (1995). Roles of regulation of Cln-cdc28 kinases at the start of the cell cycle of *saccharomyces cerevisiae*. *EMBO* **14**: 4803-4813.

Dockter J. and Ordahl C. (2000). Dorsoventral axis determination in the somite: a re-examination. *Development* **127**(10): 2201-2206.

Du S. J., Devoto S. H., Westerfield M. and Moon R. T. (1997). Positive and Negative Regulation of Muscle Cell Identity by Members of the hedgehog and TGF-beta Gene Families. *Journal of Cell Biology* **139**(1): 145-156.

Edmondson D. G. and Olson E. N. (1993). Helix-Loop-Helix Proteins as Regulators of Muscle-specific Transcription. *Journal of Biochemistry* **268**(2): 755-8.

Ekholm S. V. and Reed S. I. (2000). Regulation of G1 cyclin dependent kinases in the mammalian cell cycle. *Current opinions in Cell Biology* **12**: 676-684.

Elgar G., Clark M. S., Meek S., Smith S., Warner S., Edwards Y. J. K., Bouchireb N., Cottage A., Yeo G. S. H., Umrania Y., Williams G. and Brenner S. (1999). Generation and Analysis of 25 Mb of Genomic DNA from the Pufferfish *Fugu rubripes* by Sequence Scanning. *Genome Research* **9**(10): 960-971.

Elgar G., Sandford R., Aparicio S., Macrae A., Venkatesh B. and Brenner S. (1996). Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *Trends in Genetics* **12**(4): 145-150.

Ennion S., Gauvry L., Butterworth P. and Goldspink G. (1995). Small-diameter white myotomal muscle fibres associated with growth hyperplasia in the carp (*Cyprinus carpio*) express a distinct myosin heavy chain gene. *Journal of Experimental Biology* **198**(7): 1603-1611.

Ewing B., Hillier L., Wendl M. C. and Green P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research* **8**(3): 175-185.

Fainsod A., Deissler K., Yelin R., Marom K., Epstein M., Pillemer G., Steinbeisser H. and Blum M. (1997). The dorsalizing and neural inducing gene follistatin is an antagonist of BMP-4. *Mechanisms of Development* **63**(1): 39-50.

Fan L., Crodian J. and Collodi P. (2004). Culture of embryonic stem cell lines from zebrafish. *Methods in Cell Biology* **76**: 151-60.

Fernandes J. M. O., MacKenzie M. G., Elgar G., Suzuki Y., Watabe S., Kinghorn J. R. and Johnston I. A. (2005). A genomic approach to reveal novel genes associated with myotube formation in the model teleost, *Takifugu rubripes*. *Physiological Genomics* **22**: 327-338.

Fischer R. S. and Fowler V. M. (2003). Tropomodulins: life at the slow end. *Trends in Cell Biology* **13**(11): 593-601.

Fitzgibbon J., Hope A., Slobodyanyuk S. J., Bellingham J., Bowmaker J. K. and Hunt D. M. (1995). The rhodopsin-encoding gene of bony fish lacks introns. *Gene* **164**(2): 273-7.

Forbes D., Jackman M., Bishop A., Thomas M., Kambadur R. and Sharma M. (2006). Myostatin auto-regulates its expression by feedback loop through Smad7 dependent mechanism. *Journal of Cellular Physiology* **206**(1): 264-272.

Foster J. W., Dominguez-Steglich M. A., Guioli S., Kwok C., Weller P. A., Stevanovic M., Weissenbach J., Mansour S., Young I. D., Goodfellow P. N., Brook J. D. and Schafer A. J. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**(6506): 525-530.

Franzini-Armstrong C. (1973). The structure of a simple Z line. *The Journal Of Cell Biology* **58**(3): 630-642.

Fukunishi Y. and Hayashizaki Y. (2001). Amino acid translation program for full-length cDNA sequences with frameshift errors. *Physiolical Genomics* **5**(2): 81-87.

Garry D. J., Meeson A., Elterman J., Zhao Y., Yang P., Bassel-Duby R. and Williams R. S. (2000). Myogenic stem cell function is impaired in mice lacking the forkhead/winged helix protein MNF. *PNAS* **97**(10): 5416-5421.

Gerber A. N., Klesert T. R., Bergstrom D. A. and Tapscott S. J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes and Development* **11**(4): 436-50.

Goldstein M., Michael L., Schroeter J. and Sass R. (1988). Structural states in the Z band of skeletal muscle correlate with states of active and passive tension. *Journal of General Physiology* **92**(1): 113-119.

Goldstein M., Schroeter J. and Sass R. (1977). Optical diffraction of the Z lattice in canine cardiac muscle. *Journal of Cell Biology* **75**(3): 818-836.

Gregory D. J., Waldbieser G. C. and Bosworth B. G. (2004). Cloning and characterization of myogenic regulatory genes in three Ictalurid species. *Animal Genetics* **35**(6): 425-430.

Grimaldi A., Tettamanti G., Martin B. L., Gaffield W., Pownall M. E. and Hughes S. M. (2004). Hedgehog regulation of superficial slow muscle fibres in *Xenopus* and the evolution of tetrapod trunk myogenesis. *Development* **131**(14): 3249-3262.

Gros J., Manceau M., Thome V. and Marcelle C. (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* **435**(7044): 954-958.

Grounds M. D. and Yablonka-Reuveni Z. (1993). Molecular and cell biology of skeletal muscle regeneration. *Molecular and Cell Biology of Human Disease Series* **3**: 210-56.

Gu X., Wang Y. and Gu J. (2002). Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nature Genetics* **31**(2): 205-209.

Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Munsterberg A., Vivian N., Goodfellow P. and Lovell-Badge R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**(6281): 245-250.

Guengerich F. P., Wu Z.-L. and Bartleson C. J. (2005). Function of human cytochrome P450s: Characterization of the orphans. *Biochemical and Biophysical Research Communications* **338**(1): 465-469.

Guo K., Wang J., Andres V., Smith R. C. and Walsh K. (1995). MyoD-induced expression of p21 inhibits cyclin-dependent kinase activity upon myocyte terminal differentiation. *Mol. Cell. Biol.* **15**(7): 3823-3829.

Gurdon J. B. (1988). A community effect in animal development. *Nature* **336**(6201): 772-774.

Gurdon J. B., Lemaire P. and Kato K. (1993a). Community effects and related phenomena in development. *Cell* **75**(5): 831-834.

Gurdon J. B., Tiller E., Roberts J. and Kato K. (1993b). A community effect in muscle development. *Current Biology* **3**(1): 1-11.

Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H. and Aebersold R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**(10): 994-9.

Halevy O., Novitch B. G., Spicer D. B., Skapek S. X., Rhee J., Hannon G. J., Beach D. and Lassar A. B. (1995). Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD. *Science* **267**(5200): 1018-21.

Hall T. E., Cole N. J. and Johnston I. A. (2003). Temperature and the expression of seven muscle-specific protein genes during embryogenesis in the Atlantic cod *Gadus morhua L. Journal of Experimental Biology* **206**(Pt 18): 3187-200.

Hammonds-Odie L. P., Jackson T. R., Profit A. A., Blader I. J., Turck C. W., Prestwich G. D. and Theibert A. B. (1996). Identification and cloning of centaurin-[alpha]. A novel phosphatidylinositol 3,4,5-trisphosphate-binding protein from rat brain. *Journal of Biological Chemistry* **271**(31): 18859-18868.

Handel S. E. and Stickland N. C. (1987a). The growth and differentiation of porcine skeletal muscle fibre types and the influence of birthweight. *Journal of Anatomy* **152**: 107-119.

Handel S. E. and Stickland N. C. (1987b). The growth and differentiation of porcine skeletal muscle fibre types and the influence of birthweight. *Journal of Anatomy* **152**: 107-19.

Hao J. J., Ye J. Q., Yang Q., Gong Z. Z., Liu W. Y. and Wang E. D. (2000). A silent antifungal protein (AFP)-like gene lacking two introns in the mould Trichoderma viride. *Biochemica et Biophysica Acta.* **1475**(2): 119-24.

Harbour J. W. and Dean D. C. (2000). Rb function in cell cycle regulation and apoptosis. *Nature Cell Biology* **2**: 65-67.

Harcus Y., Parkinson J., Fernandez C., Daub J., Selkirk M., Blaxter M. and Maizels R. (2004). Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites. *Genome Biology* **5**(6): R39.

Harris M. A., Clark J., Ireland A. and Lomax J., Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**(90001): D258-261.

Hasty P., Bradley A., Morris J. H., Edmondson D. G., Venuti J. M., Olson E. N. and Klein W. H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**(6437): 501-506.

Hawke T. J. and Garry D. J. (2001). Myogenic satellite cells: physiology to molecular biology. *Journal of Applied Physiology* **91**(2): 534-551.

Hawke T. J., Meeson A. P., Jiang N., Graham S., Hutcheson K., DiMaio J. M. and Garry D. J. (2003). p21 is essential for normal myogenic progenitor cell function in regenerating skeletal muscle. *American Journal of Physiology. Cell Physiology* **285**(5): C1019-1027.

Hedges S. B. (2002). The origin and evolution of model organisms. *Nature Reviews Genetics* **3**(11): 838-849.

Heinrich G. and Pagtakhan C. J. (2004). Both 5' and 3' flanks regulate Zebrafish brain-derived neurotrophic factor gene expression. *BMC Neurosciences* **5**: 5-19.

Henry C. A. and Amacher S. L. (2004). Zebrafish Slow Muscle Cell Migration Induces a Wave of Fast Muscle Morphogenesis. *Developmental Cell* **7**(6): 917-923.

Hess F. and Videler J. J. (1984). Fast continuous swimming of saithe (Pollachius virens): A dynamic analysis of bending movements and muscle power. *Journal of Experimental Biology* **109**: 229-251.

Hinegardner R. (1968). Evolution of celullar DNA content in Teleost fishes. *American Naturalist Cell Biology* **102**: 517-523.

Hobert O. (2004). Common logic of transcription factor and microRNA action. *Trends in Biochemical Sciences* **29**(9): 462-8.

Holland P. W., Garcia-Fernandez J., Williams N. A. and Sidow A. (1994). Gene duplications and the origins of vertebrate development. *Development. Supplement*: 125-33.

Hollenberg S. M., Cheng P. F. and Weintraub H. (1993). Use of a conditional MyoD transcription factor in studies of MyoD trans-activation and muscle determination. *Proceedings of National Academic Sciences United States of America* **90**(17): 8028-32.

Horsley V., Jansen K. M., Mills S. T. and Pavlath G. K. (2003). IL-4 Acts as a Myoblast Recruitment Factor during Mammalian Muscle Growth. *Cell* **113**(4): 483-494.

Howard A. and Pelc S. R. (1951). Synthesis of nucleoprotein in bean root cells. *Nature* **167**(4250): 599-600.

Huxley A. F. (1957). Muscle structure and theories of contraction. *Progress in Biophysics and Biophysical Chemistry* **7**: 255-318.

Huxley A. F. and Niedergerke R. (1954). Structural changes during contraction: interference microscopy of living muscle fibres. *Nature* **173**: 971-973.

Huxley H. E. (1969). The mechanism of muscular contraction. *Science* **164**(886): 1356-65.

Huxley H. E. and Hanson J. (1954). Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* **173**: 973-976.

Huynh K. D. and Bardwell V. J. (1998). The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **17**(19): 2473-2484.

Iseli C., Jongeneel C. V. and Bucher P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*: 138-148.

Ito M., Ishikawa M., Suzuki S., Takamatsu N. and Shiba T. (1995). A rainbow trout SRY-type gene expressed in pituitary glands. *FEBS Letters* **377**(1): 37-40.

Iyer V. R., Eisen M. B., Ross D. T., Schuler G., Moore T., Lee J. C. n. F., Trent J. M., Staudt L. M., Hudson J., Jr., Boguski M. S., Lashkari D., Shalon D., Botstein D. and Brown P. O. (1999). The Transcriptional Program in the Response of Human Fibroblasts to Serum. *Science* **283**(5398): 83-87.

Jager R. J., Anvret M., Hall K. and Scherer G. (1990). A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. **348**(6300): 452-454.

Jaillon O., Aury J. M., Brunet F., Petit J. L., Stange-Thomann N., Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C. and Bernot A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy. J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigo, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., Roest Crollius, H. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431.**(7011): 946-57.

Jen Y., Manova K. and Benezra R. (1996). Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. *Developmental Dynamics* **207**(3): 235-52.

Ji S., Losinski R. L., Cornelius S. G., Frank G. R., Willis G. M., Gerrard D. E., Depreux F. F. S. and Spurlock M. E. (1998). *Myostatin* expression in porcine tissues: tissue specificity and developmental and postnatal regulation. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* **275**(4): R1265-1273.

Johnson R. T. and Rao P. N. (1971). Nucleo-cytoplasmic interactions in the acheivement of nuclear synchrony in DNA synthesis and mitosis in multinucleate cells. *Biological Reviews of the Cambridge Philosophical Society* **46**(1): 97-155.

Johnston I. A. (1977). a comparative study of glycolysis in red and white muscle of the trout (*Salmo gairdnerii*) and mirror carp (*Cyprinus carpio*). *Journal of Fish Biology* **11**: 575-588.

Johnston I. A. (1980). Specialisations of fish muscle. *Development and specialisation of muscle*. Goldspink D. F., Cambridge University Press. **7:** 123-148.

Johnston I. A. (1981). Structure and function of fish muscle. *Symposium of the zoological Society of London* **48**: 71-113.

Johnston I. A. (1987). Respiratory characteristics of muscle fibres in a fish (*Chaenocephalus aceratus*) that lacks haem pigments. *Journal of Experimental Biology* **133**: 415-28.

Johnston I. A. (1999). Muscle development and growth in fish: imlications for flesh quality. *Aquaculture* **177**: 99-115.

Johnston I. A. (2001). Genetic and environmental determinants of muscle growth patterns. *Fish Physiology. Muscle development and growth.* Johnston I. A. London, Academic. **18:** 141-184.

Johnston I. A., Abercromby M., Vieira V. L. A., Sigursteindottir R. J., Kristjansson B. K., Sibthorpe D. and Skulason S. (2004). Rapid evolution of muscle fibre number in post-glacial populations of Arctic charr *Salvelinus alpinus*. *Journal of Experimental Biology* **207**(25): 4343-4360.

Johnston I. A. and Cole N. (1998). Embryonic temperature modulates muscle growth characteristics in larval and juvenile herring. *Journal of Experimental Biology* **201**(Pt 12): 623-46.

Johnston I. A., Fernandez D. A., Calvo J., Vieira V. L. A., North A. W., Abercromby M. and Garland T., Jr (2003a). Reduction in muscle fibre number during the adaptive radiation of notothenioid fishes: a phylogenetic perspective. *Journal of Experimental Biology* **206**(15): 2595-2609.

Johnston I. A. and Hall T. E. (2004). Mechanisms of muscle development and responses to temperature change in fish larvae. *The Development of Form and Function in Fishes and the Question of Larval Adaption*. Govoni J. J. Bethesda, Maryland, American Fisheries Society Symp. **40:** 85-116.

Johnston I. A., Manthri S., Alderson R., Smart A., Campbell P., Nickell D., Robertson B., Paxton C. G. M. and Burt M. L. (2003b). Freshwater environment affects growth rate and muscle fibre recruitment in seawater stages of Atlantic salmon (*Salmo salar L.*). *Journal of Experimental Biology* **206**(8): 1337-1351.

Johnston I. A., Manthri S., Smart A., Campbell P., Nickell D. and Alderson R. (2003c). Plasticity of muscle fibre number in seawater stages of Atlantic salmon in response to photoperiod manipulation. *Journal of Experimental Biology* **206**(19): 3425-3435.

Johnston I. A. and McLay H. A. (1997). Temperature and family effects on muscle cellularity at hatch and first feeding in Atlantic salmon (*Salmo salar L.*). *Canadian Journal of Zoology* **75**: 64-74.

Johnston I. A., McLay H. A., Abercromby M. and Robins D. (2000). Early thermal experience has different effects on growth and muscle fibre recruitment in spring- and autumn-running Atlantic salmon populations. *Journal of Experimental Biology* **203**(Pt 17): 2553-64.

Johnston I. A., Patterson, S., Ward, P. S., Goldspink, G. (1974). The histchemical demonstration of myofibrillar adenosine triphosphatase activity in fish muscle. *Canadian Journal of Zoology* **52**: 871-877.

Johnston I. A. and Temple G. K. (2002). Thermal plasticity of skeletal muscle phenotype in ectothermic vertebrates and its significance for locomotory behaviour. *Journal of Experimental Biology* **205**(15): 2305-2322.

Johnston I. A., Vieira V. V. and Abercromby M. (1995). Temperature and myogenesis in embryos of the Atlantic herring *Clupea harengus*. *Journal of Experimental Biology* **198**(Pt 6): 1389-403.

Kadesch T. (1993). Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth and differentiation* **4**(1): 49-55.

Kan Z., Rouchka E. C., Gish W. R. and States D. J. (2001). Gene Structure Prediction and Alternative Splicing Analysis Using Genomically Aligned ESTs. *Genome Research* **11**(5): 889-900.

Kang J. H., Kim S. A. and Hong K. J. (2006). Induction of TSP1 gene expression by heat shock is mediated via an increase in mRNA stability. *FEBS Letters* **580**(2): 510-6.

Kassar-Duchossoy L., Gayraud-Morel B., Gomes D., Rocancourt D., Buckingham M., Shinin V. and Tajbakhsh S. (2004). Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature* **431**(7007): 466-471.

Kellis M., Birren B. W. and Lander E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**(6983): 617-624.

Kelly A. M. and Rubinstein N. A. (1980). Why are fetal muscles slow? *Nature* **288**(5788): 266-9.

Kerr T., Roalson E. H. and Rodgers B. D. (2005). Phylogenetic analysis of the myostatin gene sub-family and the differential expression of a novel member in zebrafish. *Evolution & Development* **7**(5): 390-400.

Kidwell M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**(1): 49-63.

Kikuno R., Nagase T., Ishikawa K.-i., Hirosawa M., Miyajima N., Tanaka A., Kotani H., Nomura N. and Ohara O. (1999). Prediction of the Coding Sequences of Unidentified Human Genes. XIV. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in vitro. *DNA Research* **6**(3): 197-205.

Kimmel C. B., Ballard W. W., Kimmel S. R., Ullmann B. and Schilling T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics* **203**(3): 253-310.

Kitazono A. A. and Kron S. J. (2002). An essential function of yeast cyclin-dependent kinase Cdc28 maintains chromosome stability. *Journal of Biological Chemistry* **277**(50): 48627-34.

Knappeis G. G. and Carlsen F. (1968). The ultrastructure of the M line in skeletal muscle. *Journal of Cell Biology* **38**: 202-211.

Kobiyama A., Nihei Y., Hirayama, Y., Kikuchi K., Suetake H., Johnston I. A. and Watabe S. (1998). Molecular cloning and developmental expression patterns of the MyoD and MEF2 families of muscle transcription factors in the carp. *Journal of Experimental Biology* **201**(Pt 20): 2801-13.

Kong Y., Flick M. J., Kudla A. J. and Konieczny S. F. (1997). Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD. *Molecular and Cellular Biology* **17**(8): 4750-4760.

Koopman P., Schepers G., Brenner S. and Venkatesh B. (2004). Origin and diversity of the Sox transcription factor gene family: genome-wide analysis in *Fugu rubripes*. *Gene* **328**: 177-186.

Koumans J. T. M. and Akster H. A. (1995). Myogenic cells in development and growth of fish. *Comparative biochemistry and physiology* **110A**(1): 3-20.

Kryvi H. (1977). Ultrastructure of the different fibre types in axial muscles of the sharks Etmopterus spinax and Galeus melastomus. *Cell Tissue Research* **184**(3): 287-300.

Lahm H. W. and Langen H. (2000). Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* **21**(11): 2105-14.

Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S. and Lin W.-c. (2000). Identification of Novel Human Genes Evolutionarily Conserved in *Caenorhabditis elegans* by Comparative Proteomics. *Genome Research* **10**(5): 703-713.

Lamatsch D. K., Steinlein C., Schmid M. and Schartl M. (2000). Noninvasive determination of genome size and ploidy level in fishes by flow cytometry: detection of triploid *Poecilia formosa*. *Cytometry* **39**(2): 91-95.

Lander E. S., Linton L. M. and Birren B. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Langley B., Thomas M., Bishop A., Sharma M., Gilmour S. and Kambadur R. (2002). Myostatin inhibits myoblast differentiation by down regulating *MyoD* expression. *Journal of Biological Chemistry* **277**(51): 49831-40.

Laudet V., Stehelin D. and Clevers H. (1993). Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Research* **21**(10): 2493–2501.

Le Blanc J. F., McLane K. E., Parren P. W. H. I., Burton D. R. and Ghazal P. (1998). Recognition Properties of a Sequence-Specific DNA Binding Antibody. *Biochemistry* **37**(17): 6015 - 6022.

Lee H., Goring W., Ochs M., Muhlfeld C., Steding G., Paprotta I., Engel W. and Adham I. M. (2004). Sox15 Is Required for Skeletal Muscle Regeneration. *Molecular and Cell Biology* **24**(19): 8428-8436.

Lee S. J. and McPherron A. C. (2001). Regulation of myostatin activity and muscle growth. *PNAS* **98**(16): 9306-9311.

Lefaucheur L., Edom F., Ecolan P. and Butler-Browne G. S. (1995). Pattern of muscle fiber type formation in the pig. *Developmental Dynamics* **203**(1): 27-41.

Lewis K. E., Currie P. D., Roy S., Schauerte H., Haffter P. and Ingham P. W. (1999). Control of Muscle Cell-Type Specification in the Zebrafish Embryo by Hedgehog Signalling. *Developmental Biology* **216**(2): 469-480.

Lin H., Yutzey K. E. and Konieczny S. F. (1991). Muscle-specific expression of the troponin I gene requires interactions between helix-loop-helix muscle regulatory factors and ubiquitous transcription factors. *Molecular and Cell Biology* **11**(1): 267–280.

Lin Z. Y., Dechesne C. A., Eldridge J. and Paterson B. M. (1989). An avian muscle factor related to MyoD1 activates muscle-specific promoters in nonmuscle cells of different germ-layer origin and in BrdU-treated myoblasts. *Genes & Development* **3**(7): 986-996.

Lindon C., Albagli O., Domeyne P., Montarras D. and Pinset C. (2000). Constitutive Instability of Muscle Regulatory Factor Myf5 Is Distinct from Its Mitosis-Specific Disappearance, Which Requires a D-Box-Like Motif Overlapping the Basic Domain. *Molecular and Cell Biology* **20**(23): 8923-8932.

Lindon C., Albagli O. and Pinset C. (2001). Cell density dependent induction of endogenous Myogenin (Myf4) gene expression by Myf5. *Development biology* **240**: 574-584.

Lindon C., Montarras, D., Pinset, C. (1998). Cell cycle -regulated expression of the muscle determination factor Myf5 in proliferating myoblasts. *The Journal of Cell Biology* **140**(1): 111-118.

Littlewood T. D., Evan, G. I. (1995). Transcription factors 2: helix-loop-helix. *Protein Profile* **2**(6): 621-702.

Liu D., Zhang N., Du J., Cai X., Zhu M., Jin C., Dou Z., Feng C., Yang Y., Liu L., Takeyasu K., Xie W. and Yao X. (2006). Interaction of Skp1 with CENP-E at the mid body is essential for cytokinesis. *Biochemical and Biophysical Research Communications* (in press).

Luther P. K. (1991). Three-dimensional reconstruction of a simple Z-band in fish muscle. *The Journal Of Cell Biology* **113**(5): 1043-1055.

Lymn R. W. and Taylor E. W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* **10**(25): 4617-24.

Lynch M. and Conery J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290**(5494): 1151-1155.

Martin J. F., Miano J. M., Hustad C. M., Copeland N. G., Jenkins N. A. and Olson E. N. (1994). A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Molecular and Cell Biology* **14**(3): 1647-56.

Martzen M. R., McCraith S. M., Spinelli S. L., Torres F. M., Fields S., Grayhack E. J. and Phizicky E. M. (1999). A Biochemical Genomics Approach for Identifying Genes by the Activity of Their Products. *Science* **286**(5442): 1153-1155.

Mascarello F., Stecchini M. L., Rowlerson A. and Ballocchi E. (1992). Tertiary myotubes in postnatal growing pig muscle detected by their myosin isoform composition. *Journal of animal science* **70**(6): 1806-1813.

Matzuk M. M., Lu N., Vogel H., Sellheyer K., Roop D. R. and Bradley A. (1995). Multiple defects and perinatal death in mice deficient in follistatin. *Nature* **374**(6520): 360-363.

McEwan G. K., Woolfe A., Goode D., Vavouri T., Callaway H. and Elgar G. (2006). Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. *Genome Research* **16**(4): 451-65.

McKinsey T. A., Zhang, C. L., Olson, E. N. (2002). MEF2: a calcium-dependent regulator of cell division, differentiation and death. *Trends in Biochemical Sciences* **27**(1): 40-47.

McKoy G., Leger M.-E., Bacou F. and Goldspink G. (1998). Differential expression of myosin heavy chain mRNA and protein isoforms in four functionally diverse rabbit skeletal muscles during pre- and postnatal development. *Developmental Dynamics* **211**(3): 193-203.

McLysaght A., Hokamp K. and Wolfe K. H. (2002). Extensive genomic duplication during early chordate evolution. **31**(2): 200-204.

McPherron A. C., Lawler A. M. and Lee S. J. (1997). Regulation of skeletal muscle mass in mice by a new TGF-b superfamily member. *Nature* **387**: 83-90.

McPherron A. C. and Lee S.-J. (1997). Double muscling in cattle due to mutations in the *myostatin* gene. *PNAS* **94**(23): 12457-12461.

Molkentin J. D., Black B. L., Martin J. F. and Olson E. N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83**(7): 1125-1136.

Molkentin J. D. and Olson E. N. (1996). Defining the regulatory networks for muscle development. *Current Opinions in Genetics and Development* **6**(4): 445-53.

Morin-Kensicki E. and Eisen J. (1997). Sclerotome development and peripheral nervous system segmentation in embryonic zebrafish. *Development* **124**(1): 159-167.

Mosse P. R. L. and Hudson R. C. L. (1977). The functional roles of different muscle fibre types identified in the myotomes of marine teleosts: a behavioural, anatomical and histochemical study. *Journal of Fish Biology* **11**: 417-430.

Mousseau T. A., Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology and Evolution* **13**: 403-407.

Mullins M. C. (1999). Embryonic axis formation in the zebrafish. *Methods Cell Biol* **59**: 159-78.

Murre C., McCaw P. S., Vaessin H., Caudy M., Jan L. Y., Jan Y. N., Cabrera C. V., Buskin J. N., Hauschka S. D. and Lassar A. B. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**(3): 537-44.

Nabeshima Y., Hanaoka K., Hayasaka M., Esumi E., Li S., Nonaka I. and Nabeshima Y. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**(6437): 532-535.

Naruse K., Fukamachi S., Mitani H., Kondo M., Matsuoka T., Kondo S., Hanamura N., Morita Y., Hasegawa K., Nishigaki R., Shimada A., Wada H., Kusakabe T., Suzuki N., Kinoshita M., Kanamori A., Terado T., Kimura H., Nonaka M. and Shima A. (2000). A Detailed Linkage Map of Medaka, Oryzias latipes: Comparative Genomics and Genome Evolution. *Genetics* **154**(4): 1773-1784.

Nasmyth K. (1996). View point, putting the cell cycle in order. *Science* **274**: 1643-1645.

Neafsey D. E. and Palumbi S. R. (2003). Genome Size Evolution in Pufferfish: A Comparative Analysis of Diodontid and Tetraodontid Pufferfish Genomes. *Genome Research* **13**(5): 821-830.

Nesic D., Cheng J. and Maquat L. E. (1993). Sequences within the intron function in RNA 3'-end formation in cultured cells. *Molecular and Cell Biology* **13**(6): 3359-69.

Noe V., MacKenzie S. and Ciudad. C. J. (2003). An intron is required for dihydrofolate reductase protein stability. *Journal of Biological Chemistry* **278**(40): 38292-300.

Norton J. D., Deed R. W., Craggs G. and Sablitzky F. (1998). Id helix-loop-helix proteins in cell growth and differentiation. *Trends in Cell Biology* **8**(2): 58-65.

Ohi R. and Gould K. L. (1999). Regulating the onset of mitosis. *Current Opinion in Cell Biology* **11**(2): 267-273.

Okazaki Y., Furuno M., Kasukawa T., Adachi J. and Bono H K. S., Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schonbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume DA, Quackenbush J, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake JA, Bradt D, Brusic V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani TA, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson IJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons PA, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Pertea G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple CA, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**(6915): 563-573.

Oksbjerg N., Gondret F. and Vestergaard M. (2004). Basic principles of muscle development and growth in meat-producing mammals as affected by the insulin-like growth factor (IGF) system. *Domestic Animal Endocrinology* **27**(3): 219-240.

Olson E. N. and Klein W. H. (1994). bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes and Development* **8**(1): 1-8.

Ontell M., Hughes D. and Bourke D. (1988). Morphometric analysis of the developing mouse soleus muscle. *American Journal of Anatomy* **181**(3): 279-88.

Orgel L. E. and Crick F. H. C. (1980). Selfish DNA: The Ultimate Parasite. *Nature* **284**: 604-607.

Ostbye T., Galloway T. F., Nielsen C., Gabestad I., Bardal T. and Andersen O. (2001). The two myostatin genes of Atlantic salmon (Salmo salar) are expressed in a variety of tissues. *European Journal of Biochemistry* **268**(20): 5249-5257.

Panopoulou G., Hennig S., Groth D., Krause A., Poustka A. J., Herwig R., Vingron M. and Lehrach H. (2003). New Evidence for Genome-Wide Duplications at the Origin of Vertebrates Using an Amphioxus Gene Set and Completed Animal Genomes. *Genome Research* **13**(6a): 1056-1066.

Parkinson J., Guiliano D. and Blaxter M. (2002). Making sense of EST sequences by CLOBBing them. *BMC Bioinformatics* **3**(1): 31.

Patel K. (1998). Follistatin. *The International Journal of Biochemistry & Cell Biology* **30**(10): 1087-1093.

Patterson S. and Goldspink G. (1972). The fine structure of red and white myotomal muscles of the Coalfish (*Gadus virens*). *Zeitschrift für Zellforschung mikroskopische Anatomie* **133**: 463-474.

Pedersen R. A. (1971). DNA content, ribosomal gene multiplicity, and cell size in fish. *Journal of Experimental Zoology* **177**(1): 65-78.

Petrov D. A. (2001). Evolution of genome size: new approaches to an old problem. *Trends in Genetics* **17**(1): 23-28.

Pfaffl M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**(9): 45.

Pollock R. and Treisman R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes and Development* **5**(12A): 2327-41.

Ponting C., Schultz J. and Bork P. (1997). SPRY domains in ryanodine receptors (Ca2+-release channels). *Trends in Biochemical Sciences* **22**(6): 193-194.

Punkt K., Naupert A. and Asmussen G. (2004). Differentiation of rat skeletal muscle fibres during development and ageing. *Acta Histochemica* **106**(2): 145-54.

Puri P. L., Avantaggiati M. L., Balsano C. and Sang N. (1997). p300 is required for MyoD-dependent cell cycle arrest and muscle specific gene transcription. *EMBO* **16**(2): 369-383.

Qi Y., Bar-Joseph Z. and Klein-Seetharaman J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins* **63**(3): 490-500.

Ralliere C., Chauvigne F. and Rescan P. Y. (2004). The genes for the helix-loop-helix proteins Id6a, Id6b, Id1 and Id2 are specifically expressed in the ventral and dorsal domains of the fish developing somites. *Journal of Experimental Biology* **207**(15): 2679-2684.

Ramakers C., Ruijter J. M., Deprez R. H. and Moorman A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**(1): 62-6.

Rehfeldt C., Ott G., Gerrard D., Varga L., Schlote W., Williams J., Renne U. and Bunger L. (2005). Effects of the Compact Mutant Myostatin Allele $Mstn^{Cmpt\text{-}dl1Abc}$ Introgressed into a High Growth Mouse Line on Skeletal Muscle Cellularity. *Journal of Muscle Research and Cell Motility* **26**(2 - 3): 103-112.

Rescan P. Y. (1997). Identification in a fish species of two Id (inhibitor of DNA binding/differentiation)-related helix-loop-helix factors expressed in the slow oxidative muscle fibers. *European Journal of Biochemistry* **247**(3): 870-6.

Rescan P. Y. and Gauvry L. (1996). Genome of the Rainbow Trout (*Oncorhynchus mykiss*) Encodes Two Distinct Muscle Regulatory Factors with Homology to MyoD. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **113**(4): 711-715.

Rescan P. Y., Gauvry L. and Paboeuf G. (1995). A gene with homology to *myogenin* is expressed in developing myotomal musculature of the rainbow trout and in vitro during the conversion of myosatellite cells to myotubes. *FEBS Letters* **362**(1): 89-92.

Rescan P. Y., Gauvry L., Paboeuf G. and Fauconneau B. (1994). Identification of a muscle factor related to MyoD in a fish species. *Biochimica et Biophysica Acta* **1218**(2): 202-4.

Rescan P.Y., Jutel I. and Ralliere C. (2001). Two *myostatin* genes are differentially expressed in myotomal muscles of the trout (*Oncorhynchus mykiss*). *Journal of Experimental Biology* **204**(20): 3523-3529.

Riechmann V., Van Cruchten I. and Sablitzky F. (1994). The expression pattern of *Id4*, a novel dominant negative helix-loop-helix protein, is distinct from *Id1*, *Id2* and *Id3*. *Nucleic Acids Research* **22**(5): 749-55.

Rigoutsos I., Huynh T., Miranda K., Tsirigos A., McHardy A. and Platt D. (2006). Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. *Proceedings of the National Academy of Science United States of America* **103**(17): 6605-10.

Rios R., Carneiro I., Arce V. M. and Devesa J. (2001). Myostatin regulates cell survival during C2C12 myogenesis. *Biochemical and Biophysical Research Communications* **280**(2): 561-6.

Roberts S. B. and Goetz F. W. (2001). Differential skeletal muscle expression of *myostatin* across teleost species, and the isolation of multiple myostatin isoforms. *FEBS Letters* **491**(3): 212-216.

Robinson-Rechavi M., Marchand O., Escriva H., Bardet P.-L., Zelus D., Hughes S. and Laudet V. (2001a). Euteleost Fish Genomes are Characterized by Expansion of Gene Families. *Genome Research* **11**(5): 781-788.

Robinson-Rechavi M., Marchand O., Escriva H. and Laudet V. (2001b). An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Current Biology* **11**(12): R458-R459.

Roca F. J., Cayuela M. L., Secombes C. J., Meseguer J. and Mulero V. (2006). Post-transcriptional regulation of cytokine genes in fish: A role for conserved AU-rich elements located in the 3'-untranslated region of their mRNAs. *Mol Immunol* (in press).

Rogic S., Mackworth A. K. and Ouellette F. B. F. (2001). Evaluation of Gene-Finding Programs on Mammalian Sequences. *Genome Research* **11**(5): 817-832.

Rosenblatt J. D., Yong D. and Parry D. J. (1994). Satellite cells activity is required for hypertrophy of overloaded adult rat muscle. *Muscle and Nerve* **17**: 608-613.

Rowe R. W. E. and Goldspink G. (1969). Muscle growth in 5 different muscles in both sexes of mice. *Journal of Anatomy* **132**: 557-579.

Rowlerson A., Mascarello F., Radaelli G. and Veggetti A. (1995). Differentiation and growth of muscle in the fish *Sparus aurata* (L): II. Hyperplastic and hypertrophic growth of lateral muscle from hatching to adult. *Journal of Muscle Research and Cell Motility (Historical Archive)* **16**(3): 223-236.

Rowlerson A. and Veggetti A. (2001). Cellular Mechanisms of Post-Embryonic Muscle Growth in Aquaculture species. *Muscle Development and Growth, Fish Physiology Series*. Johnston I. A. San Diego, Academic Press.

Ruan H. J., Brown C. Y. and Morris D. R. (1997). *Analysis of mRNA Formation and Function*. Richter J. D. New York, Academic: 305-321.

Rudnicki M. A., Schnegelsberg P. N., Stead R. H., Braun T., Arnold H. H. and Jaenisch R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**(7): 1351-9.

Ryu W. S. and Mertz J. E. (1989). Simian virus 40 late transcripts lacking excisable intervening sequences are defective in both stability in the nucleus and transport to the cytoplasm. *Journal of Virology* **63**(10): 4386-94.

Saga Y., Hata N., Koseki H. and Taketo M. M. (1997). *Mesp2*: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes and Development* **11**(14): 1827-39.

Sanger F., Air G. M., Barrell B. G., Brown N. L., Coulson A. R., Fiddes C. A., Hutchison C. A., Slocombe P. M. and Smith M. (1977). Nucleotide sequence of bacteriophage lambda DNA. *Nature* **265**(5596): 687-95.

Sanger F., Coulson A. R., Hong G. F., Hill D. F. and Petersen G. B. (1982). Nucleotide sequence of bacteriophage lambda DNA. *Journal of Molecular Biology* **162**(4): 729-73.

Sawai S. and Campos-Ortega J. A. (1997). A zebrafish Id homologue and its pattern of expression during embryogenesis. *Mechanisms of Development* **65**(1-2): 175-85.

Schepers G. E., Bullejos M., Hosking B. M. and Koopman P. (2000). Cloning and characterisation of the Sry-related transcription factor gene Sox8. *Nucleic Acids Research* **28**(6): 1473-1480.

Schepers G. E., Teasdale R. D. and Koopman P. (2002). Twenty Pairs of Sox: Extent, Homology, and Nomenclature of the Mouse and Human Sox Transcription Factor Gene Families. *Developmental Cell* **3**(2): 167-170.

Schmidt K., Glaser G., Wernig A., Wegner M. and Rosorius O. (2003). Sox8 Is a Specific Marker for Muscle Satellite Cells and Inhibits Myogenesis. *Journal of Biological Chemistry* **278**(32): 29769-29775.

Schmidt K., Schinke T., Haberland M., Priemel M., Schilling A. F., Mueldner C., Rueger J. M., Sock E., Wegner M. and Amling M. (2005). The high mobility group transcription factor Sox8 is a negative regulator of osteoblast differentiation. *Journal of Cell Biology* **168**(6): 899-910.

Schultz E. (1996). Satellite Cell Proliferative Compartments in Growing Skeletal Muscles. *Developmental Biology* **175**(1): 84-94.

Schultz E. and McCormick K. M. (1994). Skeletal muscle satellite cells. *Rev Physiol Biochemical Pharmacology* **123**: 213-57.

Schwob E., Bohm T., Mendenhall M. D. and Nasmyth K. (1994). The B-type cyclin inhibitor $p40^{sic1}$ controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**: 233-244.

Seale P., Sabourin L. A., Girgis-Gabardo A., Mansouri A., Gruss P. and Rudnicki M. A. (2000). Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**(6): 777-86.

Seeburg P. H., Shine J., Martial J. A., Ullrich A., Goodman H. M. and Baxter J. D. (1977). Nucleotide sequence of a human gene coding for a polypeptide hormone. *Transactions of the Association of American Physicians* **90**: 109-16.

Sellers J. R. (2000). Myosins: a diverse superfamily. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1496**(1): 3-22.

Shimasaki S., Koga M., Esch F., Cooksey K., Mercado M., Koba A., Ueno N., Ying S. Y., Ling N. and Guillemin R. (1988). Primary structure of the human follistatin precursor and its genomic organization. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **85**(12): 4218-4222.

Shulenin S., Nogee L. M., Annilo T., Wert S. E., Whitsett J. A. and Dean M. (2004). ABCA3 Gene Mutations in Newborns with Fatal Surfactant Deficiency. *New England Journal of Medicine* **350**(13): 1296-1303.

Shvarts A., Bazuine M., Dekker P., Ramos Y. F. M., Steegenga W. T., Merckx G., van Ham R. C. A., van der Houven van Oordt W., van der Eb A. J. and Jochemsen A. G. (1997). Isolation and Identification of the Human Homolog of a New p53-Binding Protein, Mdmx. *Genomics* **43**(1): 34-42.

Sidow A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Current Opinions in Genetics and Development* **6**(6): 715-22.

Sigma (1999). TRI Reagent: Technical Bulletin. St Louis, Missouri, USA.

Skapek S. X., Rhee J., Spicer D. B. and Lassar A. B. (1995). Inhibition of myogenic differentiation in proliferating myoblasts by cyclin D1-dependent kinase. *Science* **267**(5200): 1022-1024.

Sock E., Schmidt K., Hermanns-Borgmeyer I., Bosl M. R. and Wegner M. (2001). Idiopathic Weight Reduction in Mice Deficient in the High-Mobility-Group Transcription Factor Sox8. *Molecular and Cell Biology* **21**(20): 6951-6959.

Somero G. N. and Childress J. J. (1980). A violation of the metabolism-size scalin paradigm: activities of glycolytic enzymes in muscle increase in large fish. *Physiological Zoology* **53**: 322–337.

Southard-Smith E. M., Kos L. and Pavan W. J. (1998). SOX10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nature. Genetics* **18**(1): 60-64.

Spiller M. P., Kambadur, R., Jeanplong, F., Thomas, M., Martyn, J. K., Bass, J. J., Sharma, M. (2002). The *Myostatin* Gene Is a Downstream Target Gene of Basic Helix-Loop-Helix Transcription Factor MyoD. *Molecular and Cell Biology* **22**(20): 7066-7082.

Stark K., Kirk D. L. and Schmitt R. (2001). Two enhancers and one silencer located in the introns of *regA* control somatic cell differentiation in *Volvox carteri*. *Genes and Development* **15**(11): 1449-60.

Stearns S. (1991). The Evolution of Life Histories. *Oxford University Press*. New York.

Sterky F. and Lundeberg J. (2000). Sequence analysis of genes and genomes. *Journal of Biotechnology* **76**(1): 1-31.

Stickland N. C. (1983). Growth and development of muscle fibres in the rainbow trout (*Salmo gairdneri*). *Journal of Anatomy* **137**(Pt 2): 323-33.

Stoiber W. and Sanger A. M. (1996). An electron microscopic investigation into the possible source of new muscle fibres in teleost fish. *Anatomy and Embryology* **194**(6): 569-79.

Stolt C. C., Lommes P., Friedrich R. P. and Wegner M. (2004a). Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development* **131**(10): 2349-2358.

Stolt C. C., Lommes P., Friedrich R. P. and Wegner M. (2004b). Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy *Development* **131**(10): 2349-2358.

Stone J., de Lange T., Ramsay G., Jakobovits E., Bishop J. M., Varmus H. and Lee W. (1987). Definition of regions in human c-myc that are involved in transformation and nuclear localization. *Molecular and Cell Biology* **7**(5): 1697-709.

Sun X. H., Copeland N. G., Jenkins N. A. and Baltimore D. (1991). Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. *Molecular and Cell Biology* **11**(11): 5603-11.

Tajbakhsh S. (2003). Stem cells to tissue: molecular, cellular and anatomical heterogeneity in skeletal muscle. *Current Opinion in Genetics & Development* **13**(4): 413-422.

Tajbakhsh S. and Buckingham M. (2000). The birth of muscle progenitor cells in the mouse: spatiotemporal considerations. *Current Topics in Developmental Biology* **48**: 225-68.

Takada S. and Koopman P. (2003). Origin and possible roles of the Sox8 transcription factor gene during sexual development. *Cytogenetic and Genome Research* **101**(3-4): 212-8.

Takahashi K. and Hattori A. (1989). Alpha-actinin is a component of the Z-filament, a structural backbone of skeletal muscle Z-disks. *Journal of Biochemistry* **105**(4): 529-536.

Tan X., Hoang L. and Du S. (2002). Characterization of Muscle-Regulatory Genes, *Myf5* and *Myogenin*, from Striped Bass and Promoter Analysis of Muscle-Specific Expression. *Marine Biotechnology* **4**(6): 537-545.

Tanaka K., Imajoh-Ohmi S., Sawada T., Shirai R., Hashimoto Y., Iwasaki S., Kaibuchi K., Kanaho Y., Shirai T., Terada Y., Kimura K., Nagata S. and Fukui Y. (1997). A target of phosphatidylinositol 3,4,5-trisphosphate with a zinc finger motif similar to that of the ADP-ribosylation-factor GTPase-activating protein and two pleckstrin homology domains. *European Journal of Biochemistry* **245**(2): 512-9.

Tapon N., Moberg K. H. and Hariharan I. K. (2001). The coupling of cell growth to the cell cycle. *Current Opinions in Cell Biology* **13**(6): 731-7.

Taylor W. E., Bhasin S., Artaza J., Byhower F., Azam M., Willard D. H., Jr., Kull F. C., Jr. and Gonzalez-Cadavid N. (2001). Myostatin inhibits cell proliferation and protein synthesis in C2C12 muscle cells. *American Journal of Physiology. Endocrinology and Metabolism* **280**(2): E221-228.

Temple G. K., Cole N. J. and Johnston I. A. (2001). Embryonic temperature and the relative timing of muscle-specific genes during development in herring (*Clupea harengus L.*). *Journal of Experimental Biology* **204**(Pt 21): 3629-37.

Thisse B., Stoetzel C., Gorostiza-Thisse C. and Perrin-Schmitt F. (1988). Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early Drosophila embryos. *EMBO Journal* **7**(7): 2175-83.

Thomas M., Langley B., Berry C., Sharma M., Kirk S., Bass J. and Kambadur R. (2000). *Myostatin*, a negative regulator of muscle growth, functions by inhibiting myoblast proliferation. *Journal of Biological Chemistry* **275**(51): 40235-43.

Ticho B. S., Stainier D. Y. R., Fishman M. C. and Breitbart R. E. (1996). Three zebrafish MEF2 genes delineate somitic and cardiac muscle development in wild-type and mutant embryos. *Mechanisms of Development* **59**(2): 205-218.

Tskhovrebova L. and Trinick J. (2003). Titin: properties and family relationships. *Nature Reviews Molecular Cell Biology* **4**: 679-689.

Tu Y. and Wu C. (1999). Cloning, expression and characterization of a novel human Ras-related protein that is regulated by glucocorticoid hormone. *Biochimica et Biophysica Acta* **1489**(2-3): 452-6.

Tyler J. C. (1980). Osteology, phylogeny, and higher classification of the fishes of the order Plectognathi (Tetraodontiformes). *NOAA Technical Report NMFS* **434:** 1–422.

Uetz P. and Hughes R. E. (2000). Systematic and large-scale two hybrid screens. *Current Opinions in Microbiology* **3**(3): 303-8.

Van Leeuwen J. (1999). A mechanical analysis of myomere shape in fish. *Journal of Experimental Biology* **202**(23): 3405-3414.

Van Raamsdonk W., Mos W., Smit-Onel M. J., Van der Laarse and W. J. F., R. (1983). The development of the spinal motor column in relation to the myotomal muscle fibers in the zebrafish (Brachydanio rerio). I. Posthatching development. *Anatomy and Embryology (Berlin)* **167**(1): 125-39.

Van Raamsdonk W., Pool C. W. and Te Kronnie G. (1978). Differentiation of muscle fibre types in the teleost Brachydanio rerio. *Anatomy and Embryology (Berlin)* **153**: 137-155.

Van Raamsdonk W., Van Der Stelt A. and Diegenbach P. C. (1974). Differentiation of the musculature of the teleost *Brachydanio rerio*. I. Myotome shape and movements in the embryo. *Zeitschrift für Anatomie und Entwicklungsgeschichte* **145**(3): 321-342.

Venkatesh B. and Gilligan P. (2000). *Fugu*: a compact vertebrate reference genome. *FEBS* **476**: 3-7.

Venkatesh B., Ning Y. and Brenner S. (1999). Late changes in spliceosomal introns define clades in vertebrate evolution. *Proceedings of the National Academy of Science. The United States of America* **96**(18): 10267-71.

Venkateswarlu K., Brandom K. G. and Lawrence J. L. (2004). Centaurin-{alpha}1 Is an in Vivo Phosphatidylinositol 3,4,5-Trisphosphate-dependent GTPase-activating Protein for ARF6 That Is Involved in Actin Cytoskeleton Organization. *Journal of Biological Chemistry* **279**(8): 6205-6208.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., Gocayne J. D., Amanatides P., Ballew R. M., Huson D. H., Wortman J. R., Zhang Q., Kodira C. D., Zheng X. H., Chen L., Skupski M., Subramanian G., Thomas P. D., Zhang J., Gabor Miklos G. L., Nelson C., Broder S., Clark A. G., Nadeau J., McKusick V. A., Zinder N., Levine A. J., Roberts R. J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Francesco V. D., Dunn P., Eilbeck K., Evangelista C., Gabrielian A. E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T. J., Higgins M. E., Ji R.-R., Ke Z., Ketchum K. A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G. V., Milshina N., Moore H. M., Naik A. K., Narayan V. A., Neelam B., Nusskern D., Rusch D. B., Salzberg S., Shao W., Shue B., Sun J., Wang Z. Y., Wang A., Wang X., Wang J., Wei M.-H., Wides R., Xiao C., Yan C., Yao A., Ye J., Zhan M., Zhang W., Zhang H., Zhao Q., Zheng L., Zhong F., Zhong W., Zhu S. C., Zhao S., Gilbert D., Baumhueter S., Spier G., Carter C., Cravchik A., Woodage T., Ali F., An H., Awe A., Baldwin D., Baden H., Barnstead M., Barrow I., Beeson K., Busam D., Carver A., Center A., Cheng M. L., Curry L., Danaher S., Davenport L., Desilets R., Dietz S., Dodson K., Doup L., Ferriera S., Garg N., Gluecksmann A., Hart B., Haynes J., Haynes C., Heiner C., Hladun S., Hostin D., Houck J., Howland T., Ibegwam C., Johnson J., Kalush F., Kline L., Koduru S., Love A., Mann F., May D., McCawley S., McIntosh T., McMullen I., Moy M., Moy L., Murphy B., Nelson K., Pfannkoch C., Pratts E., Puri V., Qureshi H., Reardon M., Rodriguez R., Rogers Y.-H., Romblad D., Ruhfel B., Scott R., Sitter C., Smallwood M., Stewart E., Strong R., Suh E., Thomas R., Tint N. N., Tse S., Vech C., Wang G., Wetter J., Williams S., Williams M., Windsor S., Winn-Deen E., Wolfe K., Zaveri J., Zaveri K., Abril J. F., Guigo R., Campbell M. J., Sjolander K. V., Karlak B., Kejariwal A., Mi H., Lazareva B., Hatton T. and Narechania A. (2001). The Sequence of the Human Genome. *Science* **291**(5507): 1304 - 1351.

Vianello S., Brazzoduro L., Dalla Valle L., Belvedere P. and Colombo L. (2003). *Myostatin* expression during development and chronic stress in zebrafish (*Danio rerio*). *J Endocrinology* **176**(1): 47-59.

Volff, Hornung and Schartl (2001). Fish retroposons related to the Penelope element of *Drosophila virilis* define a new group of retrotransposable elements. *Molecular Genetics and Genomics* **265**(4): 711-720.

Volff J.-N., Bouneau L., Ozouf-Costaz C. and Fischer C. (2003). Diversity of retrotransposable elements in compact pufferfish genomes. *Trends in Genetics* **19**(12): 674-678.

Wada M. R., Inagawa-Ogashiwa M., Shimizu S., Yasumoto S. and Hashimoto N. (2002). Generation of different fates from multipotent muscle stem cells. *Development* **129**(12): 2987-95.

Wainwright P. C., Turingan R. G. and Brainerd E. L. (1995). Functional mophology of pufferfish inflation: mechanism of the buccal pump., Copeia 1995**:** 614-625.

Walker J. E., Arizmendi J. M., Dupuis A., Fearnley I. M., Finel M., Medd S. M., Pilkington S. J., Runswick M. J. and Skehel J. M. (1992). Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from bovine heart mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction. *Journal of Molecular Biology* **226**(4): 1051-72.

Walsh K. and Perlman H. (1997). Cell cycle exit upon myogenic differentiation. *Current opinions in genetics and development* **7**: 597-602.

Wang K., McClure J. and Tu A. (1979). Titin: major myofibrillar components of striated muscle. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **76**(8): 3698-3702.

Wang K. and Wright J. (1988). Architecture of the sarcomere matrix of skeletal muscle: immunoelectron microscopic evidence that suggests a set of parallel inextensible nebulin filaments anchored at the Z line
10.1083/jcb.107.6.2199. *J. Cell Biol.* **107**(6): 2199-2212.

Wang Q., Keutmann H. T., Schneyer A. L. and Sluss P. M. (2000). Analysis of Human Follistatin Structure: Identification of Two Discontinuous N-Terminal Sequences Coding for Activin A Binding and Structural Consequences of Activin Binding to Native Proteins *Endocrinology* **141**(9): 3183-3193.

Wargelius A., Ellingsen S. and Fjose A. (1999). Double-Stranded RNA Induces Specific Developmental Defects in Zebrafish Embryos. *Biochemical and Biophysical Research Communications* **263**(1): 156-161.

Way M., Sanders M., Chafel M., Tu Y., Knight A. and Matsudaira P. (1995). beta-Scruin, a homologue of the actin crosslinking protein scruin, is localized to the acrosomal vesicle of Limulus sperm. *Journal of Cell Science* **108**(10): 3155-3162.

Weatherley A. H. and Gill G. S. (1987). The Biology of Fish Growth. *Academic press.* London.

Weatherley A. H. and Gill H. S. (1985). Dynamics of increase in muscle fibres in fishes in relation to size and growth. *Experientia* **41**: 353-354.

Weatherley A. H., Gill H. S. and Lobo A. F. (1988). Recruitment and maximal diameter of axial muscle fibres in teleosts and their relationship to somatic growth and ultimate size. *Journal of Fish Biology* **33**: 851-859.

Weatherley A. H., Gill H. S. and Rogers S. (1979). Growth dynamics of muscle fibres, dry weight and condition in relation to somatic growth rate in yearling rainbow trout (*Salmo gairdner*). *Canadian Journal of Zoology* **57**: 2385-2392.

Weatherley A. H., Gill H. S. and Rogers S. (1980). The relationship between mosaic muscle fibres and size in rainbow trout (*Salmo gairdneri*). *Journal of Fish Biology* **17**: 603-610.

Wegner J., Albrecht, E., Fiedler, I., Teuscher, F., Papstein, H. J., Ender, K. (2000). Growth- and breed-related changes of muscle fiber characteristics in cattle. *Journal of Animal Science* **78**(6): 1485-1496.

Wegner M. (1999). From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Research* **27**(6): 1409-1420.

Wei Q. and Paterson B. M. (2001). Regulation of MyoD function in the dividing myoblast. *FEBS* **490**: 171-178.

Weihs D. (1973). The mechanism of rapid starting of slender fish. *Biorheology* **10**(3): 343-50.

Weinberg E. S., Allende M. L., Kelly C. S., Abdelhamid A., Murakami T., Andermann P., Doerre O. G., Grunwald D. J. and Riggleman B. (1996). Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development* **122**(1): 271-280.

Weintraub H., Tapscott S. J., Davis R. L., Thayer M. J., Adam M. A., Lassar A. B. and Miller A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of *MyoD. Proceedings Of The National Academy Of Sciences Of The United States Of America* **86**(14): 5434-5438.

Weltzien F. A., Norberg B., Helvic J. V., Andersen O., Swanson P. and Andersson E. (2003). Identification and localization of eight distinct hormone-producing cell types in the pituitary of male Atlantic halibut (*Hippoglossus hippoglossus* L.). *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* **134**(2): 315-27.

Wilkes D., Xie S. Q. and Stickland N. C. (2001). Temperature and myogenic factor transcript levels during early development determines muscle growth potential in rainbow trout (*Oncorhynchus mykiss*) and sea bass (*Dicentrarchus labrax*). *Journal of Experimental Biology* **204**: 2763-2771.

Wilson S. J., McEwan J. C., Sheard P. W. and Harris A. J. (1992). Early stages of myogenesis in a large mammal: formation of successive generations of myotubes in sheep tibialis cranialis muscle. *Journal of Muscle Research and Cell Motility* **13**(5): 534-50.

Wolf J. B., Brodie E. D., Cheverud J. M., Moore A. J. and Wade M. J. (1998). Evolutionary consequences of indirect genetic effects. *Trends in Ecology and Evolution* **13**: 64-69.

Wolfe K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2**(5): 333-341.

Wolff C., Roy S. and Ingham P. W. (2003). Multiple Muscle Cell Identities Induced by Distinct Levels and Timing of Hedgehog Activity in the Zebrafish Embryo. *Current Biology* **13**(14): 1169-1181.

Wright W. E., Sassoon D. A. and Lin V. K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**(4): 607-17.

Wu L., Timmers C. and Maiti B. (2001). The E2F1-3 transcription factors are essential for cellular proliferation. *Nature* **414**: 457-462.

Xie S. Q., Mason P. S., Wilkes D., Goldspink G., Fauconneau B. and Stickland N. C. (2001). Lower environmental temperature delays and prolongs myogenic regulatory factor expression and muscle differentiation in rainbow trout (*Onchrhynchus mykiss*) embryos. *Differentiation.* **68**(2-3): 106-14.

Xu C., Wu G., Zohar Y. and Du S.-J. (2003). Analysis of myostatin gene structure, expression and function in zebrafish. *Journal of Experimental Biology* **206**(22): 4067-4079.

Xue F. and Cooley L. (1993). Kelch encodes a component of intercellular bridges in Drosophila egg chambers. *Cell* **72**(5): 681-693.

Yu Y. T., Breitbart R. E., Smoot L. B., Lee Y., Mahdavi V. and Nadal-Ginard B. (1992). Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. *Genes and Development* **6**(9): 1783-98.

Yuan W., Condorelli G., Caruso M., Felsani A. and Giordano A. (1996). Human p300 protein is a coactivator for the transcription factor MyoD. *Journal of Biological Chemistry* **271**: 9009–9013.

Yun K. and Wold B. (1996). Skeletal muscle determination and differentiation: story of a core regulatory network and its context. *Current Opinions in Cell Biology* **8**(877-889).

Zemlickova E., Dubois T., Kerai P., Clokie S., Cronshaw A. D., Wakefield R. I. D., Johannes F.-J. and Aitken A. (2003). Centaurin-[alpha]1 associates with and is phosphorylated by isoforms of protein kinase C. *Biochemical and Biophysical Research Communications* **307**(3): 459-465.

Zhang J.-M., B. M. P. Qin Wei, et al. (1999). Coupling of the cell cycle and myogenesis through the cyclin D1-dependent intraction of MyoD with Cdk4. *EMBO* **18**(4): 926-933.

Zhang J. M., Wei Q., Zhao X. and Paterson B. M. (1999). Coupling of the cell cycle and myogenesis through the cyclin D1-dependent interaction of MyoD with cdk4. *The EMBO Journal* **18**(4): 926-933.

Zhang M., Koishi K. and McLennan I. S. (1998). Skeletal muscle fibre types: Detection methods and embryonic determinants. *Histology and Histopathology* **13**(1): 201-207.

Zhang N. and Gridley T. (1998). Defects in somite formation in lunatic fringe-deficient mice. *Nature* **394**(6691): 374-377.

Zimmers T. A., Davies M. V., Koniaris L. G., Haynes P., Esquela A. F., Tomkinson K. N., McPherron A. C., Wolfman N. M. and Lee S. (2002). Induction of Cachexia in Mice by Systemically Administered Myostatin. *Science* **296**(5572): 1486-1488.

Zollman S., Godt D., Prive G., Couderc J. and Laski F. (1994). The BTB Domain, Found Primarily in Zinc Finger Proteins, Defines an Evolutionarily Conserved Family that Includes Several Developmentally Regulated Genes in Drosophila. *PNAS* **91**(22): 10717-10721.

Zong Q., Schummer M., Hood L. and Morris D. R. (1999). Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proceeding of the National Academy of Science. The United States of America* **96**(19): 10632-6.

Zylka M. J., Shearman L. P., Weaver D. R. and Reppert S. M. (1998). Three period homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. *Neuron* **20**(6): 1103-10.