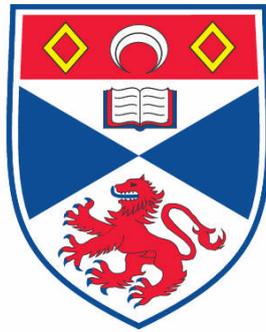


**EMBRYONIC TEMPERATURE AND THE GENES REGULATING  
MYOGENESIS IN TELEOSTS**

**Daniel John Macqueen**

**A Thesis Submitted for the Degree of PhD  
at the  
University of St. Andrews**



**2008**

**Full metadata for this item is available in the St Andrews  
Digital Research Repository  
at:**

**<https://research-repository.st-andrews.ac.uk/>**

**Please use this identifier to cite or link to this item:**

**<http://hdl.handle.net/10023/518>**

**This item is protected by original copyright**

**This item is licensed under a  
[Creative Commons License](#)**

# **Embryonic temperature and the genes regulating myogenesis in teleosts**

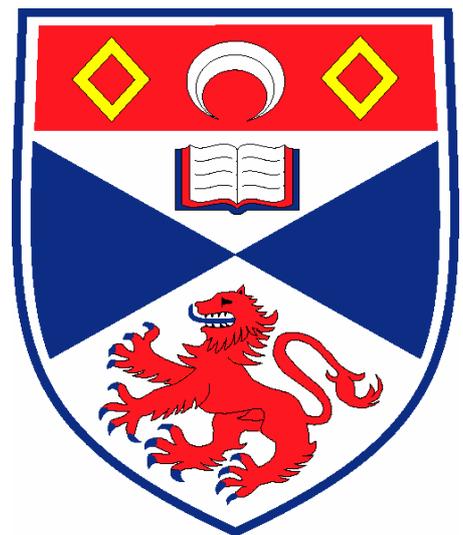
Submitted to the University of St Andrews for the degree of Doctor of Philosophy

by

**Daniel John Macqueen**

## **Project Supervisor**

Professor Ian Alistair Johnston,  
School of Biology,  
Gatty Marine Laboratory,  
University of St Andrews,  
St Andrews,  
Fife,  
KY16 8LB



**January 2008**

## *Declarations*

- i. I, Daniel John Macqueen, hereby certify that this thesis, which is approximately 65,000 words in length, was written by me, is a record of work performed by me and has not been submitted in any previous application for a higher degree.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

- ii. I was admitted as a research student in October 2004 and as a candidate for the degree of Doctor of Philosophy in October 2005; the higher study, for which this thesis is a record, was performed in the University of St Andrews from 2004-2007.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

- iii. I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulation appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date \_\_\_\_\_ 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 in accordance with the regulation of the university library for the time being in force, subject to any copyright vested in the work not being affected thereby. I understand that the title and abstract will be published, and that a copy of the work may be supplied to any *bona fide* library or research worker.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

## Acknowledgments

I would firstly like to thank my supervisor, Professor Ian Johnston for his ongoing support during my project. Ian has kept me on track, while letting me be creative and has helped me exponentially develop my writing skills. Additionally, I acknowledge the financial support of the Natural Environment Research Council who funded my studentship and tuition fees (ref NERC/S/A/2004/12435). As a CASE student, I received additional funding from EWOS Innovation.

Other past and present work colleagues from the Fish Muscle Research Group that I wish to thank are as follows: Dr Jorge Fernandes (now associate Professor at the university college of Bodø, Norway), not only for his incredible support throughout my PhD, but also for always upholding his role as the lab's central post-doc (and ethanol consumer); Dr Neil Bower (present), for sharing ideas about genes and science-stuff, for work-relieving chats about football and for being a fine drinking buddy; Hung-Tai Lee (present), for providing zebrafish samples and for interesting discussions on the teleost dermomyotome and the origin of stratified hyperplasia!; Tom Ashton (present) for kindly measuring the texture of some mislabelled and heavily oxidised salmon muscles and for taking forward the burden of the salmon calpastatin gene; Dr Matt Mackenzie (past) for supplying me with pufferfish samples, for co-inventing the game 'sweat-boy' and for providing a true comedy moment by jumping fully clothed into the sea off of Barcelona harbour with mobile phone and wallet in tow; Shelby Steele and Ørjan Hagen (both past) for being good pals and for a couple of enjoyable late nights; Hannah Chubb (past) for being a really nice person; Dr Sonia Consuegra (past) for putting up with my first PCR attempts and for shouting at me; Dr Vera Johnston (Vieira) (present) for always being friendly, helpful and thoughtful and Marguerite Abercromby (past) for her help with learning to cut cryosections and for always being in a good mood.

I would like to acknowledge EWOS Innovation for supporting the whole embryonic temperature experiment and for providing all Atlantic salmon material used in this thesis. I would specifically like to whole-heartedly thank the following EWOS Innovation associates: Dr Dave Robb for setting up and designing the project, for his continuous help with sampling and for drinking beer and wine with me on many a cold Lonnigdahl night. Also, thanks Dave for not getting mad when I forgot my Passport before a flight to Norway!; Tom Olsen for being a great guy and for running

the whole trial day to day for over three years; Linda Melstveit for her help with sampling and for keeping me and Dave on our toes; Liv Oma for help with sampling; Frode Ormhaug for husbandry of fish during freshwater stages; Aage Melstveit (and Linda again) for accommodation and for letting me watch Liverpool win the Champions league when no other T.V. was in sight for miles around as well as Viv Crampton and Dr Louise Buttle for helping to establish and design the project.

I couldn't have got through this PhD (or life generally) in one piece, without my wife Ruth Macqueen, who I would like to thank for being my best friend and for all the things only she knows she does for me. I would also like to thank my Mum and Dad for everything they have done for me in my life's journey. My Mother, Marilyn has been a rock in my life for as long as I can remember and I can't thank her enough for her endless support through all things I have done, good or bad. I thank my Dad, Jonathan, for inadvertently installing a mighty work ethic in me, despite his distance. I would also like to thank my parent in-laws Catherine and Stephen for their support in recent times. Other members of my family who I can't thank enough for always being there are my Sister, Samantha (and her other half Pricey [Princey]), my Aunties/surrogate mothers Lorraine and Karen, my cousins and pseudo-siblings Steph and Chris (the boy), my Nan, Doreen and Grandad, Jack, the beautiful collies Holly and Tess, and to those others whom I haven't mentioned or have moved on to greener pastures.

## Peer reviewed publications and author contributions

1. Macqueen, D.J. and Johnston, I.A. 2006. A novel salmonid myoD gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization. *FEBS Letters* **580**: 4996-5002.

I performed all experiments and conceived and wrote the paper together with Ian Johnston.

2. Macqueen, D.J., Robb, D., Johnston, I.A. 2007. Temperature influences the co-ordinated expression of myogenic regulatory factors during embryonic myogenesis in Atlantic salmon (*Salmo salar* L.). *Journal of Experimental Biology* **210**: 2781-2794.

Ian Johnston, Dave Robb and myself conceived the experiment. Dave Robb managed the sampling of Atlantic salmon embryos. I performed all experimental procedures and wrote the paper with Ian Johnston.

3. Macqueen, D.J. and Johnston, I.A. Evolution of follistatin in teleosts revealed through phylogenetic, genomic and expression analyses. *Development Genes and Evolution*: **218**: 1-14.

I performed all experiments and conceived and wrote the paper together with Ian Johnston.

4. Fernandes, J.M.O.<sup>1</sup> Macqueen, D.J.<sup>1</sup>, Lee, H-T and Johnston, I.A. 2007. Genomic, evolutionary and expression analyses of *cee*, an ancient gene involved in normal growth and development *Genomics: E pub before print*. DOI: 10.1016/j.ygeno.2007.10.017

<sup>1</sup> Denotes equal contribution.

\*Jorge Fernandes, Ian Johnston and myself conceived the paper. Jorge Fernandes and myself were the principal experimental contributors. Jorge Fernandes, Ian Johnston and myself wrote

the paper. Individual contributions are acknowledged in the materials and methods section of chapter 6.

5. Macqueen, D.J. and Johnston, I.A. An update on MyoD evolution in teleosts and a proposed consensus nomenclature to accommodate the tetraploidization of different vertebrate genomes. *PLoS ONE* **3**: e1567 doi:10.1371/journal.pone.0001567.

I performed all experiments and conceived and wrote the paper together with Ian Johnston

6. Macqueen, D.J., Robb, D.H.F., Olsen, T., Melstveit, L., Paxton, C. and Johnston, I.A. Temperature until the “eyed stage” of embryogenesis programs the growth trajectory and muscle phenotype of adult Atlantic salmon. Accepted in *Biology Letters*. In Press.

Ian Johnston and Dave Robb conceived the embryonic temperature experiment. Tom Olsen managed the fish throughout the whole experiment. Tom Olsen, Linda Melstveit, Dave Robb and myself sampled the fish. I performed all experimental and statistical procedures relating to establishing adult muscle growth traits, except for modelling the growth data, which was done by Charles Paxton and the probability density function analysis, which was performed by Ian Johnston and myself. I wrote the paper jointly with Ian Johnston.

## Table of Contents

	Page
<b>Abstract</b>	1-2
<b>Chapter 1. General introduction</b>	3-53
<i>1.1 Teleost fish: biodiversity and use as model organisms</i>	3
<i>1.2 Teleost muscle as a source of food</i>	4
<i>1.3 A focus on Atlantic salmon</i>	5
<i>1.4 Why study teleost muscles?</i>	6
<i>1.5 Gene and genome duplication and teleosts</i>	7
<i>1.5.1 Genome duplication: a historical perspective</i>	7
<i>1.5.2 Genome duplication who and when?</i>	8
<i>1.5.3 Genome duplication and the fate of paralogues</i>	10
<i>1.6 Myogenesis</i>	11
<i>1.6.1 Common features of myogenesis across vertebrates</i>	11
<i>1.6.2 Teleost myotomal muscle: patterns and innovations</i>	12
<i>1.6.3 Myogenic regulatory factors: the master switches for muscle specific gene transcription</i>	18
<i>1.6.4 Lessons from knockout studies of the MRFs</i>	19
<i>1.6.5 The MEF-2 transcription factors co-regulate differentiation with MRFs</i>	24
<i>1.6.6 Paired box transcription factors and satellite cells</i>	24
<i>1.6.7 Myostatin: the most potent known negative regulator of myogenesis</i>	26
<i>1.6.8 Follistatin a regulator of Mstn and of myogenesis</i>	28
<i>1.6.9 Other known regulators of Mstn</i>	29

<i>1.6.10 The Insulin like growth factors, Calcineurin and the Akt/mTOR pathway</i>	30
<i>1.6.11 Forkhead box proteins and satellite cells</i>	33
<i>1.6.12 Sox8, another marker of adult satellite cells</i>	34
<i>1.6.13 The Calpain- Calpastatin system</i>	34
<i>1.6.14 MicroRNAs and myogenesis</i>	36
<b>1.7 Phases of teleost myogenesis</b>	37
<i>1.7.1 Embryonic myogenesis: early patterns and onset of slow muscle formation</i>	37
<i>1.7.2 Slow muscle specification and hedgehog morphogens</i>	38
<i>1.7.3 Formation of embryonic fast muscle</i>	39
<i>1.7.4 The anterior somite of teleosts and continuing muscle growth</i>	40
<i>1.7.5 Stratified hyperplasia</i>	42
<i>1.7.6 Mosaic hyperplasia and maximum fibre diameter</i>	42
<i>1.7.7 What is the source of muscle growth during mosaic hyperplasia?</i>	44
<i>1.7.8 Signals regulating myotube production</i>	44
<b>1.8 Myogenesis and the environment</b>	47
<i>1.8.1 Introduction</i>	47
<i>1.8.2 Embryonic temperature and myogenesis</i>	48
<i>1.8.3 Lasting effects of early rearing temperature on myogenesis</i>	50
<b>1.9 Commercial relevance of plasticity of teleost myogenesis</b>	51
<b>1.10 Aims and goals</b>	52
<b>Chapter 2. Materials and methods</b>	54-69
<b>2.1 Introduction</b>	54
<b>2.2 Embryonic temperature experiment</b>	54

2.2.1 <i>Experimental set-up</i>	54
2.2.2 <i>Embryonic sampling</i>	55
2.2.3 <i>Post embryonic sampling</i>	56
<b>2.3 <i>Basic computational biology methods</i></b>	<b>57</b>
2.3.1 <i>Gene databases</i>	57
2.3.2 <i>DNA manipulation in silico</i>	58
2.3.3 <i>Primer design</i>	58
2.3.4 <i>Predication of gene structure</i>	59
2.3.5 <i>Sequence submission</i>	59
<b>2.4 <i>Basic experimental molecular biology</i></b>	<b>59</b>
2.4.1 <i>General working conditions</i>	59
2.4.2 <i>Extraction of DNA and RNA</i>	60
2.4.3 <i>Gel Electrophoresis</i>	60
2.4.4 <i>Quantification and quality assessment of RNA</i>	61
2.4.5 <i>Gel purification of DNA</i>	62
2.4.6 <i>Complementary DNA (cDNA) synthesis</i>	62
2.4.7 <i>Polymerase chain replication (PCR) and Reverse Transcription-PCR (RT-PCR)</i>	62
2.4.8 <i>Rapid Amplification of Complementary Ends (RACE) PCR</i>	63
2.4.9 <i>DNA cloning</i>	63
2.4.10 <i>Plasmid purification, digestion and screening</i>	64
2.4.11 <i>Sequencing of DNA</i>	65
<b>2.5 <i>Quantitative PCR (qPCR)</i></b>	<b>65</b>
2.5.1 <i>Experimental procedure</i>	65
2.5.2 <i>Dissociation analysis and primer specificity</i>	66
<b>2.6 <i>In situ hybridisation</i></b>	<b>66</b>

2.6.1 Introduction	66
2.6.2 Amplification of probe templates	66
2.6.3 Probe transcription	67
2.6.4 Embryo preparation and fixation	67
2.6.5 Probe hybridisation and detection	68
2.6.6 Embryo study, cryosectioning and photography	68
<b>Chapter 3. The evolutionary relationships of teleost myod genes revealed through comparative phylogenetic and genomic analyses</b>	<b>70-101</b>
<b>3.1 Abstract</b>	<b>70</b>
<b>3.2 Introduction</b>	<b>71</b>
<b>3.3 Materials and methods</b>	<b>72</b>
3.3.1 <i>In silico</i> mining and cloning of myod2	72
3.3.2 Testing the selective constraints across MyoD1/2 proteins	73
3.3.3 Sequence retrieval for phylogenetic reconstruction	73
3.3.4 Phylogenetic reconstruction of myod genes	74
3.3.5 Synteny analysis of teleost myod genes	76
3.3.6 Phylogenetic reconstruction of myod-neighbouring genes	76
<b>3.4 Results</b>	<b>78</b>
3.4.1 <i>In silico</i> and experimental characterisation of teleost myod2 genes	79
3.4.2 Evolutionary constraints on Acanthopterygian MyoD1 and MyoD2 proteins	79
3.4.3 Phylogenetic reconstruction of teleost myod genes within a MRF framework	79
3.4.4 Further phylogenetic reconstruction of MyoD	81
3.4.5 Genomic neighbourhood surrounding myod genes	82

3.4.6 <i>Phylogenetic reconstruction of myod-neighbouring genes</i>	83
<b>3.5 Discussion</b>	95
3.5.1 <i>MyoD2 is specifically conserved in the Acanthopterygii and is evolving rapidly relative to its paralogue</i>	95
3.5.2 <i>MyoD duplications and vertebrate polyploidy</i>	96
3.5.3 <i>A comparative genomic study of myod neighbouring genes reveals the true extent of the MyoD1/MyoD2 duplication</i>	97
3.5.4 <i>Chapter conclusion: a consensus nomenclature for vertebrate MyoD sequences</i>	99
<b>Chapter 4. Characterisation of a novel differentially expressed Atlantic salmon MyoD paralogue</b>	102-129
<b>4.1 Abstract</b>	102
<b>4.2 Introduction</b>	103
<b>4.3 Materials and methods</b>	104
4.3.1 <i>Fish sampling</i>	104
4.3.2 <i>Computational approaches and sequence retrieval</i>	104
4.3.3 <i>Cloning Atlantic salmon myoD1c</i>	105
4.3.4 <i>Genomic characterisation of myoD1c</i>	105
4.3.5 <i>In situ hybridisation</i>	106
4.3.6 <i>Quantitative RT-PCR (qPCR)</i>	106
<b>4.4 Results</b>	109
4.4.1 <i>Sequence characterisation of myoD1c</i>	109
4.4.2 <i>Characterisation of non-coding regions of myoD1c</i>	110
4.4.3 <i>Comparison of salmon MyoD1 paralogues with MyoD1 orthologues</i>	111

4.4.4	<i>Expression of myod1 paralogues during salmon embryogenesis</i>	111
4.4.5	<i>qPCR: primer specificity and validity of housekeeping genes</i>	113
4.4.6	<i>Quantification of myod1 paralogue expression in salmon fast and slow muscle</i>	114
<b>4.5</b>	<b>Discussion</b>	121
4.5.1	<i>Three paralogous myod1 genes are present in salmonid fish</i>	121
4.5.2	<i>Comparison of salmon MyoD1 paralogues</i>	122
4.5.3	<i>Atlantic salmon myod1 genes and the differentiation of embryonic fast and slow muscle</i>	124
4.5.4	<i>Differential expression of myod1 paralogues in adult muscle</i>	125
4.5.5	<i>The stabilization of myod1 paralogues in the salmonid genome</i>	126
4.5.6	<i>Chapter conclusion</i>	127
<b>Chapter 5</b>	<b>Evolution of Follistatin in teleosts revealed through phylogenetic, genomic and expression analyses: a role for fst1 in teleost myogenesis</b>	130-159
<b>5.1</b>	<b>Abstract</b>	130
<b>5.2</b>	<b>Introduction</b>	131
<b>5.3</b>	<b>Materials and methods</b>	132
5.3.1	<i>Sequence retrieval and genomic analyses</i>	132
5.3.2	<i>Construction of synteny diagram</i>	133
5.3.3	<i>Phylogenetic and evolutionary analyses</i>	134
5.3.4	<i>Testing the selective constraints across Fst proteins</i>	135
5.3.5	<i>Tissue specific mRNA expression of teleost fst genes</i>	135
5.3.6	<i>Cloning of salmon fst1</i>	136
5.3.7	<i>Embryos and whole-mount in situ hybridization</i>	137
<b>5.4</b>	<b>Results</b>	137

5.4.1	<i>Two fst paralogues are present in the Ostariophysi</i>	137
5.4.2	<i>Teleost fst genes are present on duplicated chromosomes with double conserved synteny relative to tetrapods</i>	138
5.4.3	<i>Asymmetric evolution of teleost Fst paralogues</i>	140
5.4.4	<i>Distinct mRNA tissue distribution of fst1/2 paralogues</i>	141
5.4.5	<i>Comparative promoter analysis of vertebrate fst genes</i>	141
5.4.6	<i>Expression of fst1 during embryonic myogenesis in salmon</i>	142
5.4.7	<i>fst1 expression in myogenic progenitors of the pectoral fin buds</i>	143
<b>5.5</b>	<b><i>Discussion</i></b>	<b>155</b>
5.5.1	<i>Two Fst genes were duplicated in a common teleost ancestor</i>	155
5.5.2	<i>Positive selection of Fst2 relative to other Fst proteins?</i>	155
5.5.3	<i>Differential regulation of fst1/fst2 paralogues suggests distinct evolution of regulatory regions</i>	156
5.5.4	<i>Has subfunctionalization of fst2 contributed to its retention in the Ostariophysi?</i>	157
5.5.5	<i>A conserved role for vertebrate fst genes in regulating myogenesis?</i>	158
5.5.6	<i>Chapter conclusion</i>	159
<b>Chapter 6.</b>	<b>Genomic, evolutionary and expression analyses of cee, an ancient and novel gene involved in normal growth and development</b>	<b>160-186</b>
<b>6.1</b>	<b><i>Abstract</i></b>	<b>160</b>
<b>6.2</b>	<b><i>Introduction</i></b>	<b>160</b>
<b>6.3</b>	<b><i>Materials and methods</i></b>	<b>159</b>
6.3.1	<i>Animals and sample collection</i>	162
6.3.2	<i>Computational identification of cee orthologues</i>	162
6.3.3	<i>Cloning and sequencing of cee cDNAs</i>	163

6.3.4	<i>Sequence alignments and intron-exon structure of cee</i>	163
6.3.5	<i>Phylogenetic inference and tests of selection</i>	164
6.3.6	<i>RNA probe preparation and whole-mount in situ hybridization</i>	165
6.3.7	<i>Tissue distribution of cee mRNA in adult salmon</i>	165
<b>6.4</b>	<b>Results</b>	166
6.4.1	<i>In silico identification of cee in eukaryote genomes</i>	166
6.4.2	<i>Characterisation of cee in metazoans</i>	168
6.4.3	<i>Gene structure of cee</i>	169
6.4.4	<i>Phylogenetic reconstruction of Cee</i>	170
6.4.5	<i>Selective constraints on Cee during animal evolution</i>	170
6.4.6	<i>Developmental expression of cee</i>	171
<b>6.5</b>	<b>Discussion</b>	182
6.5.1	<i>Cee is a novel protein of unknown function that is widely conserved in the eukaryotes</i>	182
6.5.2	<i>cee is a lonely gene without a family</i>	183
6.5.3	<i>Gene structure of cee</i>	184
6.5.4	<i>Developmental expression of cee</i>	185
<b>Chapter 7</b>	<b>The co-ordinated spatiotemporal expression of myogenic regulatory factors is affected by temperature during embryonic myogenesis in Atlantic salmon (<i>Salmo salar</i> L.)</b>	187-227
<b>7.1</b>	<b>Abstract</b>	187
<b>7.2</b>	<b>Introduction</b>	188
<b>7.3</b>	<b>Materials and methods</b>	189
7.3.1	<i>Embryos</i>	189
7.3.2	<i>Cloning new Atlantic salmon myogenic regulatory factors and <i>smlc1</i></i>	189

7.3.3 <i>Computational processing of new MRF sequences</i>	190
7.3.4 <i>Probe transcription and in situ hybridisation</i>	190
7.3.5 <i>Processing embryos and figure construction</i>	191
<b>7.4 Results</b>	192
7.4.1 <i>Characterisation of Atlantic salmon myogenic regulatory factors</i>	192
7.4.2 <i>Genomic organisation of salmonid MRFs</i>	194
7.4.3 <i>Characterisation of Atlantic salmon <i>smc1</i></i>	194
7.4.4 <i>MRF expression: introduction</i>	194
7.4.5 <i>MRF expression in a single maturing somite (Fig. 7.6)</i>	195
7.4.6 <i>The dynamics of rostral-caudal expression of MRFs during embryogenesis (Figs. 7.8-7.9)</i>	197
7.4.7 <i>Embryonic temperature and somitogenesis</i>	201
7.4.8 <i>Embryonic temperature and the co-ordinated expression of MRFs</i>	201
<b>7.5 Discussion</b>	217
7.5.1 <i>MRFs in Atlantic salmon and other teleosts</i>	217
7.5.2 <i><i>Mrf4</i> and myogenic specification, a trait lacking in teleosts?</i>	220
7.5.3 <i>Salmon MRFs and the external cell layer</i>	221
7.5.4 <i>Heterochronies in MRF expression at different temperatures</i>	221
7.5.5 <i>Concluding thoughts</i>	223
<b>Chapter 8. Temperature until the ‘eyed stage’ of embryogenesis programs the growth trajectory and muscle phenotype of adult Atlantic salmon</b>	227-246
<b>8.1 Abstract</b>	227
<b>8.2 Introduction</b>	228
<b>8.3 Materials and methods</b>	229

8.3.1 Embryonic temperature experiment	229
8.3.2 Fish sampling	229
8.3.3 Modelling growth data	230
8.3.4 Muscle morphometry statistics	231
8.3.5 Calculating fibre probability density functions	231
<b>8.4 Results</b>	232
8.4.1 Embryonic growth trajectory	232
8.4.2 Post embryonic growth trajectory	232
8.4.3 Embryonic temperature and adult muscle fibre morphometrics	233
8.4.4 Embryonic temperature and muscle fibre nuclear density	234
8.4.5 Temperature-induced differences in muscle fibres size distribution	235
<b>8.5 Discussion</b>	242
8.5.1 Changing temperature solely to the eyed stage programs adult salmon growth trajectory and muscle fibre phenotype	242
8.5.2 Implications of temperature induced differences in muscle fibre number	243
8.5.3 Embryonic temperature induced alterations in myonuclear density	244
8.5.4 How does embryonic temperature program adult myogenic phenotype?	245
<b>Chapter 9. General discussion</b>	247-256
9.1 A molecular tool-box for studying Atlantic salmon myogenesis	247
9.2 Characterising teleost and muscle salmon genes: patterns and paralogues	247
9.3 What role did genome duplication play in the teleost success story?	249
9.4 Taking the embryonic temperature findings onwards	251
9.5 Effect of embryonic temperature on other systems	253

<b>9.6 Thoughts for the future</b>	254
<b>Appendix I. Two colour whole mount <i>in situ</i> hybridisation using sequential alkaline phosphatase staining with chromogenic substrates of <i>Salmo salar</i> embryos</b>	259-265
<b>Appendix II. List of manufacturers addresses</b>	266
<b>References</b>	267-312

## List of Figures

	Page
Fig. 1.1. Taxonomic relationships of major model species within the phylum Chordata	9
Fig. 1.2. A simple model of post-embryonic vertebrate myogenesis	13
Fig. 1.3. The archetypal teleost muscle phenotype	17
Fig. 1.4. A simple model demonstrating the individual role of the MRFs	21
Fig. 1.5. Schematic representation of the cellular specification, differentiation and migration of embryonic MPC populations in a generalised teleost	45
Fig. 3.1 Full AA sequence alignment of MyoD1 and MyoD2 AA sequences in three species of the Acanthopterygii	86
Fig. 3.2. Average ratio of synonymous and non-synonymous substitutions in MyoD1 and MyoD2 of Acanthopterygian teleosts	87
Fig. 3.3. ML cladogram of vertebrate MRF sequences	88
Fig. 3.4 Unrooted phylograms of vertebrate MyoD AA sequences constructed using Bayesian, maximum likelihood and neighbour joining approaches	90
Fig. 3.5. Conserved synteny between the <i>myod</i> -containing chromosomes of human, chicken, zebrafish, pufferfish, stickleback and medaka	91
Fig. 3.6. Unrooted phylogenetic cladograms for AA translations of genes in proximity to <i>myod</i> that are conserved as two copies in teleosts	93
Fig. 4.1. MyoD1c at the mRNA and protein level	115
Fig. 4.2 Multiple species alignment of vertebrate MyoD1 sequences	117
Fig. 4.3. Differential expression of <i>myod1</i> genes during Atlantic salmon somitogenesis	118
Fig. 4.4. <i>myod1</i> expression at the Atlantic salmon eyed stage	119
Fig. 4.5. Quantitative expression of <i>myod1</i> paralogues in adult Atlantic salmon fast and slow muscle fibres	120
Fig. 4.6. Two scenarios could account for three salmonid <i>myod</i> paralogues conserved from the <i>myod1</i> lineage	128
Fig. 4.7. A simple scenario describing the evolution of salmonid <i>myod1</i> paralogues under the DDC/subfunctionalization model	129

Fig. 5.1. ML phylogenetic reconstruction of 33 vertebrate AA Fst sequences	144
Fig. 5.2. Several duplicated genes in the neighbourhood of zebrafish <i>fst1/2</i> were also present on two chromosomal regions in Acanthopterygian teleosts	145
Fig. 5.3. Phylogenetic reconstruction of genes in chromosomal proximity to <i>fst</i> in zebrafish and Acanthopterygian teleosts	147
Fig. 5.4. Plots of the cumulative number of synonymous and non-synonymous substitutions across Fst proteins from mammals, euteleosts and the Ostariophysi	149
Fig. 5.5. Duplex RT-PCR of <i>fst</i> and <i>ef1-<math>\alpha</math></i> genes in zebrafish and Atlantic salmon	150
Fig. 5.6. Conserved motifs in the promoter regions of teleost and tetrapod <i>fst</i>	151
Fig. 5.7. Expression of <i>fst1</i> during the segmentation period of Atlantic salmon embryogenesis	152
Fig. 5.8. Expression of <i>fst1</i> at the end of the segmentation period	153
Fig. 5.9. Expression of <i>fst1</i> in muscle progenitors of the pectoral fin buds	154
Fig. 6.1. Sequence alignment of Cee amino acid sequences from twelve vertebrates	176
Fig. 6.2 Gene organisation of <i>cee</i> across the eukaryotes	177
Fig. 6.3. Unrooted ML phylogram of Cee proteins from a broad range of metazoan taxa	178
Fig. 6.4. Synonymous and non-synonymous substitution rates in Cee.	179
Fig. 6.5. Developmental expression pattern of <i>cee</i> in Atlantic salmon	180
Fig. 7.1. Atlantic salmon Myf5 at the mRNA and AA level	204
Fig. 7.2. Atlantic salmon Myog at the mRNA and AA level	205
Fig. 7.3. Atlantic salmon Mrf4 at the mRNA and AA level	206
Fig. 7.4. Sequence alignment of all known salmonid MyoD family members with a MyoD orthologue in the cephalochordate-amphioxus	207
Fig. 7.5. Intron-exon structures of Atlantic salmon myogenic regulatory factors	208
Fig. 7.6. Schematic diagram illustrating, for several stages of Atlantic salmon embryogenesis, the mRNA expression patterns of six MRF genes, as well as <i>smlc1</i> and <i>pax7</i> in the most anterior (oldest) somite	209
Fig. 7.7. Expression of Atlantic salmon MRFs prior to somitogenesis	210

Fig. 7.8. mRNA expression patterns of MRFs and <i>smlc1</i> expression during the 30-45 somite stage of Atlantic salmon embryogenesis	211
Fig. 7.9. mRNA expression patterns of MRFs as well as <i>smlc1</i> and <i>pax7</i> at the 65ss and during the eyed stage of Atlantic salmon embryogenesis	212
Fig. 7.10. Rate of somitogenesis in Atlantic salmon reared at 2, 5 and 8°C	213
Fig. 7.11. Temperature associated heterochronies in <i>myf5</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	214
Fig. 7.12. Temperature associated heterochronies in <i>mrf4</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	215
Fig. 7.13. Temperature associated heterochronies in <i>smlc1</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	216
Fig. 7.14. Two <i>mrf4</i> genes map to two distinct genomic locations in Atlantic salmon	225
Fig. 8.1. Changing embryonic temperature solely to the ‘eyed stage’ of embryogenesis produced marked effects on the post-smoltification growth trajectory of Atlantic salmon	226
Fig. 8.2. The effect of embryonic temperature on the fast-steak cross sectional area of adult Atlantic salmon	238
Fig. 8.3. The effect of embryonic temperature on the final muscle fibre phenotype of adult Atlantic salmon	239
Fig. 8.4. Scatterplot of myonuclear density versus the fast-steak cross sectional area in adult Atlantic salmon reared at 5 and 10°C	240
Fig. 8.5. The effect of embryonic temperature on the size distribution of muscle fibres in adult Atlantic salmon	241

## List of Tables

	Page
Table 3.1. Details of teleost MyoD sequences, including their current designation, Genbank accession number/Ensembl gene ID as well as correct designations (proposed consensus nomenclature) according to the comparative-genomic and phylogenetic results of chapter 3.	101
Table 4.1. Primer details for chapter 4	106
Table 4.2. Comparison of percentage sequence identity of coding and non-coding features of <i>myod1</i> co-orthologues ( <i>myod1a/1b/1c</i> ) and <i>myod1c</i> orthologues	116
Table 6.1. List of primer pairs used to amplify <i>cee</i> in chapter 6	173
Table 6.2. Accession numbers for metazoan <i>cee</i> genes and putative proteins studied chapter 6	174
Table 7.1. Primer details for chapter 7	203
Table 8.1. Summary of mixed-model ANOVA parameters used to distinguish variation in post-smolt body mass	235
Table 8.2. Summary of general linear model ANOVA parameters used to distinguish variation in muscle fibre characteristics	237
Table 9.1. New genes characterised during the current project	257

## List of standard and non-standard abbreviations

AA: Amino acid

Abcc-8: ATP-binding cassette, sub-family C (CFTR/MRP), member 8,

ATP: Adenosine Triphosphate

BCIP: 5-Bromo-4-chloro-3-indolyl phosphate

BHLH: basic helix-loop-helix

BLAST: Basic Local Alignment and Search Tool

Blimp1: transcriptional repressor B-lymphocyte-induced maturation protein 1

BMP: bone morphogenic protein

BSA: Bovine Serum Albumen

Bya: Billion years ago

cDNA: complementary deoxyribose nucleic acid

CDS: coding sequence

Cee: conserved edge expressed

CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

cm: centimetres

cRNA: complementary nucleic acid

Cxcr4: CXC chemokine Receptor 4

DacA: *Drosophila* dachshund A homologue

DDC: duplication, degeneration, complementation

DEPC: diethylpyrocarbonate

DIG: digoxigenin

DNA: deoxyribose nucleic acid

dNTP: deoxyribonucleotide triphosphate

EDTA: ethylenediaminetetraacetic acid

EST: expressed sequence tag

EtBr: ethidium bromide

Fgf8: fibroblast growth factor 8

FLRG: follistatin related gene

Flu: fluorescein

Foxk1: forkhead box K1

Fst: follistatin

GASP1: WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2

GATA-2: GATA binding protein 2

GDF8: growth differentiation factor 8

GDF11: growth differentiation factor 11

HCl: hydrogen chloride

Hh: hedgehog

HLH: helix-loop-helix

Hox: homeobox gene

IGF: Insulin-like growth factor

JTT: Jones-Taylor-Thorton

kb: kilobase

KCl: potassium chloride

Kcnc1: potassium voltage gated channel-1

LB: Luria-Bertani

Lbx1: ladybird homeobox homolog 1

LiCl: lithium chloride

MADS-box: MCM1, AGAMOUS, DEFICIENS and SRF (serum response factor)-box

MEF2: myocyte enhancing factor 2

MgCl<sub>2</sub>: magnesium chloride

min: minutes

miRNA: microRNA

ML: maximum likelihood

mm: millimetres

mM: millimolar

MMLV: Moloney Murine Leukemia Virus

Mox1: mesenchyme homeobox 1

MPC: myogenic progenitor cell

MRF: myogenic regulatory factor

Mrf4: myogenic regulatory factor 4

mRNA: messenger RNA

MS222: tricaine methanesulphonate

Mstn: myostatin

mTOR: mammalian target of rapamycin

m/v: molecular volume

Mya: million years ago

MyoD: myogenic determination factor

Myog: myogenin

Myf5: myogenic factor 5

Myf6: myogenic factor 6

NaCl: sodium chloride

NBT: Nitro-Blue Tetrazolium Chloride

NCBI: National Center for Biotechnology Information

NFAT: nuclear factor of activated T-cells

NH<sub>4</sub>: ammonium

NJ: neighbour joining

nM: nanomolar

Nucb2: nucleobindin 2

ORF: open reading frame

Otog: Otogelin

p70S6 kinase: ribosomal S6 protein kinase p70 S6 kinase

PCNA: proliferating cell nuclear antigen

PI(3)K: phosphatidylinositol-3-OH kinase

Pik3c2a: phosphoinositide-3-kinase, class 2, alpha polypeptide

pH: power of hydrogen (a measure of the acidity or alkalinity of a substance)

Pax3: paired box transcription factor 3

Pax7: paired box transcription factor 7

PBS: phosphate buffered saline

PBT: phosphate buffered saline; 0.1% Tween 20

Pbx: pre-B cell leukaemia homeobox gene

PCR: polymerase chain replication

PHAS-1: phosphorylated heat and acid stable protein regulated by insulin

PIT Tag: passive integrator transponder tag

Plekha7: pleckstrin homology domain containing, family A member 7

ppm: parts per million

qPCR: quantitative real-time RT-PCR

RNA: ribonucleic acid.

RNase: ribonuclease

Rpm: revolutions per minute

Rps13: ribosomal protein S13

RT-PCR: reverse transcription-PCR

s: seconds

Sdf1a: Stromal Cell Derived Factor 1 alpha

SERGEF: secretion regulating guanine nucleotide exchange factor

Smlc1: Slow myosin light chain 1

ss: somite stage

SSC: sodium chloride sodium citrate

TAE: Tris-acetate-EDTA

TE: Tris-EDTA

TFBS: transcription factor binding site

TGF- $\beta$ : Transforming growth factor-beta

Tm: melting temperature

Tph1: tryptophan hydroxylase gene-1

Tris: Trizma acetate

tRNA: transfer RNA

TropI: Troponin I

TropT: Troponin T

UTR: untranslated region

v/v volume/volume

WAG: Whelan And Goldman

WGD: whole genome duplication

$\mu$ M: micromolar

## Abstract

In this study, full coding sequences of Atlantic salmon (*Salmo salar* L.) muscle genes were cloned, including myogenic regulatory factors (MRFs) (*myod1c*, *myog*, *mrf4*, *myf5*), inhibitors of Myostatin (*fst*, decorin), markers of myogenic progenitor cell (MPC) proliferation (*sox8*) and fusion (calpastatin), a marker of slow muscle fibre differentiation (*smlc1*) and a novel eukaryotic gene involved in regulating growth (*cee*). Several of these genes were then characterised using a range of experimental and computational analyses with the aim to better understand their role in myogenesis and their evolution in teleosts.

A series of experiments supported previous findings that teleosts have extra copies of many genes relative to tetrapods as a result of a whole genome duplication (WGD) event that occurred some 320-350 Mya. For example, it was shown that genes for *myod* and *fst* have duplicated in a common teleost ancestor, but were then specifically lost or retained in different lineages. Furthermore, several characterised Atlantic salmon genes were conserved as paralogues, likely from a later WGD event specific to the salmonid lineage. Phylogenetic reconstruction and comparative genomic approaches were used to characterise the evolution of teleost paralogues within a framework of vertebrate evolution. As a consequence of one experiment, a revised nomenclature for *myod* genes was proposed that is relevant to all diploid and polyploid vertebrates.

The expression patterns of multiple myogenic genes were also established in Atlantic salmon embryos using specific complementary RNA probes and *in situ* hybridization. For example, coordinated embryonic expression patterns were revealed for six salmon MRFs (*myod1a*, *myod1b*, *myod1c*, *myog*, *mrf4*, *myf5*), as well as markers of distinct MPC populations (*pax7*, *smlc1*), providing insight into the regulatory networks governing myogenesis in a tetraploid teleost.

Furthermore, it was shown that Atlantic salmon *fst1* was expressed concurrently to *pax7* in a recently characterised MPC population originating from the anterior domain of the epithelial somite, which is functionally analogous to the amniote dermomyotome. In another experiment, the individual expression domains of three Atlantic salmon *myod1* paralogues were shown to together recapitulate the expression of the single *myod1* gene in zebrafish, consistent with the partitioning of ancestral cis-acting regulatory elements among salmonid *myod1* duplicates. Additionally, the *in situ* expression of *cee* a novel and highly conserved eukaryotic gene was revealed for the first time in a vertebrate and was consistent with an important role in development including myogenesis.

Additionally, Atlantic salmon were reared at 2, 5, 8 or 10 °C solely to a defined embryonic stage, which was just subsequent to the complete pigmentation of the eye. After this time, animals were provided an equal growth opportunity. Remarkably, changing temperature during this short developmental window programmed the growth trajectory throughout larval and adult stages. While 10 and 8 °C fish were larger than those reared at 2 and 5 °C at the point of smoltification, strong compensatory growth was subsequently observed. Consequently, after 18 months of on growing, size differences among 5, 8 and 10 °C fish were not significant, although each group was heavier than 2 °C fish. Furthermore, significant embryonic-temperature induced differences were observed in the final muscle fibre phenotype, including the number, size distribution and myonuclear density of muscle fibres. A clear optimum for the final muscle fibre number was observed in 5 °C fish, which was up to 17% greater than other treatments. In a sub-sample of embryos, temperature induced heterochonies were recorded in the expression of some MRFs (*myf5*, *mrf4*) but not others (*myod1a*, *myog*). These results allowed the proposition of a potential mechanism explaining how temperature can program the muscle phenotype of adult teleosts through modification of the somitic external cell layer, a source of MPCs throughout teleost ontogeny.

# Chapter 1. General Introduction

## *1.1 Teleost fish: biodiversity and use as model organisms*

Teleost fish, the study animals in this thesis, form the dominant class of the Actinopterygii (the ray-finned fishes) and are the most diverse vertebrate group with around 25,000 species currently recognised, comprising 38 orders, 426 families and more than 4000 genera (Nelson, 2006). This group display a remarkable variety of body forms and lifestyles, inhabiting every imaginable aquatic environment from  $-2^{\circ}\text{C}$  water of the Antarctic, where small antifreeze proteins have evolved to stop the body from freezing (Fletcher et al., 2001), to hot alkaline springs where temperatures can exceed  $40^{\circ}\text{C}$  (Coe, 1966). Teleosts range in size from the miniscule *Paeodocypris progenetica*, which reaches a maximum of 10 mm body length (Kottelat et al., 2006) to the Ocean sunfish (*Mola mola*), which can reach 3 m and weigh 2500 kg, or in freshwater, the Wels catfish (*Siluris glanis*) which has been observed to reach body lengths of 5 m. Such a variety of life-style adaptations make teleosts an extremely attractive study group to approach a range of biological questions and several species have become important models. In developmental biology, zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*) are like mice, used as models for organogenesis and human disease, due to their prodigious breeding and hardiness in the laboratory, their short generation times and embryonic transparency, the availability of numerous developmental mutant lines and the routine use of morpholino-antisense RNA to ‘knockdown’ genes of interest (Grunwald and Eisen, 2002; Wittbrodt et al., 2002; Barut and Zon, 2000). The pufferfishes *Tetraodon nigroviridis* and *Takifugu rubripes* are used almost purely as models for the study of genomics and evolution due to the compactness of their genomes, characterised by short introns and a lack of sequences that are repetitive or non-protein coding (Brenner et al., 1993). Other teleosts, such as cichlid sp. and stickleback (*Gasterosteus aculeatus*) are important models for studying how speciation and adaptive radiation occur at various biological levels, owing to the presence of convenient populations of phenotypically

similar species/morphs that have evolved rapidly in replicated situations of sympatry, allopatry or parapatry (Kocher, 2004; McKinnon and Rundle, 2002). This list is far from exhaustive and studies on teleosts extend into almost every field of biology.

Currently, the use of teleosts as biological models is complemented by a growing resource of genetic and genomic resources, available for a wide range of species (reviewed in Volff, 2005; Cossins and Crawford, 2005) and ranging from expressed sequence tags to genetic maps and whole genome sequencing projects. Currently, five teleost genomes have been sequenced to completion or near completion, including the zebrafish (the *D. rerio* sequencing project: [http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/)), the pufferfishes, *T. rubripes* (Aparicio et al., 2002) and *T. nigroviridis* (Jaillon et al., 2004), the stickleback (*G. aculeatus*) (broad institute: <http://www.broad.mit.edu/tools/data/data-vert.html>) and Japanese medaka (Kasahara et al., 2007). The availability of genomic resources on this scale allows the consideration of biological questions to a depth previously unthinkable.

### ***1.2 Teleost muscle as a source of food***

Teleost skeletal muscle, the focus of this project, accounts for up to 70% of the body mass (Weatherly and Gill, 1985) and forms the majority of the edible tissue. For this reason teleosts are an important human food resource exploited from worldwide fisheries on scales ranging from the leisure fisherman to global industries. However, capture fisheries are in decline in the modern world and aquaculture, the growth of aquatic species for human consumption or use, is an expanding industry (FAO, 2003, 2006). Worldwide in 2004, 106 million tonnes of animals were either produced or captured from aquatic environments for human consumption, and ~43% came directly from aquaculture (FAO, 2006). Several teleost species form important sectors of global aquaculture industries (Naylor et al., 2000; FAO, 2006), including carp (*Cyprinus carpio*), tilapia sp., salmonid sp., Ictalurid catfish, Atlantic cod (*Gadus morhua*) and Atlantic halibut

(*Hippoglossus hippoglossus*). The production of low value fish species such as cyprinids predominates global aquaculture which is mainly accounted for by the Asian industry that produces ~90% of all cultured fish. In contrast, aquaculture in the Western world principally involves the growth of high value species like Atlantic salmon (Naylor et al., 2000). While aquaculture is of increasing importance in terms of meeting the protein demands of a growing global population (FAO, 2003), paradoxically the culture of certain animals can have a negative effect impact on fisheries, for example due to the high level of wild fish required to supplement the protein requirements of carnivorous species (Naylor et al., 2000) as well as the escape of cultured animals into the environment and subsequent ecological impacts (Naylor et al., 2001, 2005).

### ***1.3 A focus on Atlantic salmon***

The Atlantic salmon (*S. salar* L.) is historically reputed for its fight with the fly fisherman but most importantly is savoured for its delicate taste and texture. For this reason, the sale of Atlantic salmon as a food item forms a valuable global industry that has its basis in aquaculture. In fact the culture of Atlantic salmon began in the 19<sup>th</sup> century when UK rivers were stocked with parr to supplement anglers catches (FAO culture species information programme). Previous to this time, there is evidence that Atlantic salmon have been long revered by humans, having been found carved into a stone-age cave on the Vezere River in France dated to be over 20,000 years old and depicted on ancient Celtic coins (Atlantic salmon Federation, <http://asf.ca>). More recently, the FAO reported a massive drop in Atlantic salmon fishery hauls with about 4000 tonnes being caught in 2003, compared to typical annual catches exceeding 12,000 tonnes before 1990 (FAO culture species information programme). A concomitant expansion of the Atlantic salmon cultivation industry has occurred and from its infancy in 1980, when 99% of consumed salmon was still derived from fisheries, annual Atlantic salmon outputs from aquaculture have increased every year to exceed a million tonnes from 2001 (FAO culture species information

programme). Atlantic salmon are cultured in their native habitats, such as northern Europe and the east of the USA, as well as further a field in regions like Chile, the western USA and Tasmania (Naylor et al., 2005).

In nature, Adult salmon spawn in their natal river during autumn and winter months. Eggs are deposited in nests buried within the gravel and hatchling termed alevins, remain in this environment feeding from a yolk sac (Armstrong and Nisrow, 2006). Upon yolk sac absorption, juveniles are termed parr and generally spend 1-5 years in the river before they smolt (gain physiological tolerance to saltwater) and enter seawater as post-smolts during later spring and summer months (Thorpe, 1988). Adults migrate to oceanic feeding grounds for up to 5 years, before they return to their natal system upon sexual maturation (Hansen et al., 1993). Unlike the wild life cycle, the aquaculture production cycle of Atlantic salmon is controlled in a way designed to make fish grow faster, altering the timing of smoltification and restricting maturation, with a view to producing a higher quality end product in a shorter time. For example, photoperiod is controlled to influence the timing of smolting (Duston and Saunders, 1995) and to manipulate the onset of sexual maturation (Hansen et al., 1992). Additionally, in meeting consumer demands for pink salmonid flesh (Clydesdale, 1993), the carotenoid pigments astaxanthin and canthaxanthin are used as dietary supplements (Torrissen et al., 1995; Nickell and Bromage, 1998).

#### ***1.4 Why study teleost muscles?***

The end phenotype of an animal's tissue is the product of an ongoing and complex interaction between the genome and the rearing environment. Muscle is a plastic tissue and in teleosts is known to vary in phenotype with a changing environment (Johnston and Temple, 2002; Johnston, 2006), which in turn can have implications for growth characteristics (Johnston et al., 2003a) as well as flesh quality (Johnston, 2001a). Thus, understanding the molecular regulation

of muscle development and how this interacts with the environment to produce high quality flesh must be a key goal for the aquaculture industry. Additionally, teleost myogenesis is studied from a comparative perspective for example as a model for human disease (Dooley and Zon, 2000) and for establishing the evolution of myogenesis in vertebrates (e.g. Devoto et al., 2006). This introductory chapter reviews the cellular and molecular process of myogenesis in teleost fishes from embryo to adult, with particular reference to the role of the environment in shaping the muscle phenotype. Firstly, it is important to review the current literature on genome duplication, which is an inescapable phenomenon when studying teleosts at the molecular level, a fact which will become evident in following experimental chapters.

### ***1.5. Gene and genome duplication and teleosts***

#### *1.5.1 Genome duplication: a historical perspective*

Genome duplication is a massively important evolutionary mechanism with relevance to almost all biological disciplines (reviewed in Zhang, 2003; Taylor and Raes, 2004). It was observed some 70 years ago that a *Drosophila melanogaster* mutant with reduced eye size had a double chromosomal band compared to the wildtype (Bridges, 1936). Later, Stephens hypothesised that since mutations in genetic material likely brought about some impairment in function, a mechanism for the evolution of new gene functions with such a cost would be wasteful and evolutionary progress likely required the production of new genetic material by the expansion of the existing genome (Stephens, 1951). Later, Ohno suggested that leaps in evolutionary complexity from single-celled organisms to modern vertebrates could not have occurred without entirely new genes and suggested that WGDs have performed this function (Ohno, 1970). It was argued that genome duplication would allow a new copy of each gene to freely develop a new function, while the other maintained its original role (Ohno, 1970).

### 1.5.2 Genome duplication who and when?

It is known that protostomes and deuterostome ancestors of vertebrates typically have single copies of genes relative to vertebrates, which have up to four (Holland et al., 1994; Spring, 1997; Makalowski, 2001). For example, the *hox* genes, which are tandemly orientated in metazoan genomes are found as four clusters in the Sarcopterygii compared to a single cluster in protostomes and the basal deuterostome amphioxus (Garcia-Fernández and Holland, 1996). The most common explanation for such observations is referred to as the 1-to-2-to-4 rule, or 2R (2 rounds) hypothesis (reviewed by Meyer and Schartl, 1999; Meyer and Van de Peer, 2005), which proposes that two rounds of WGD have occurred in the ancestry of vertebrates, respectively prior to and after the divergence of the lamprey lineage (e.g. Holland et al., 1994; Sidow, 1996) (As illustrated in Fig. 1.1). However, the phylogenetic reconstruction of gene families has sometimes supported (e.g. Vandepoele et al., 2004) or rejected (e.g. Hughes, 1999, Martin, 2001) the expected tree topology of the 2R hypothesis, which remains a contentious issue. It was further proposed that the teleost genome duplicated again following the split of the Actinopterygii and Sarcopterygii lineages, based on the presence of seven *Hox* clusters in zebrafish relative to four in tetrapods (e.g. Amores et al., 1998). This occurrence has been named the 3R duplication or the 1-to-2-to-4-to-8 rule (Meyer and Schartl, 1999; Meyer and Van de Peer, 2005) and was supported by rigorous phylogenetic reconstructions and synteny analyses of teleost genes found as paralogues compared to Sarcopterygian relatives (e.g. Taylor et al., 2003; Vandepoele et al., 2004). These notions have since been proven true based on the genome-wide study of Jaillon and Co-Workers who provided unequivocal evidence for a teleost specific whole duplication. This work showed firstly that for ~1000 pairs of duplicated genes in *T. nigroviridis* and *T. rubripes*, there was a strong tendency for one copy to be on a single distinct chromosome compared to its paralogue, rather than distributed at random (Jaillon et al., 2004). Secondly a striking pattern of double conserved synteny existed where two *Tetraodon* chromosomal segments were observed

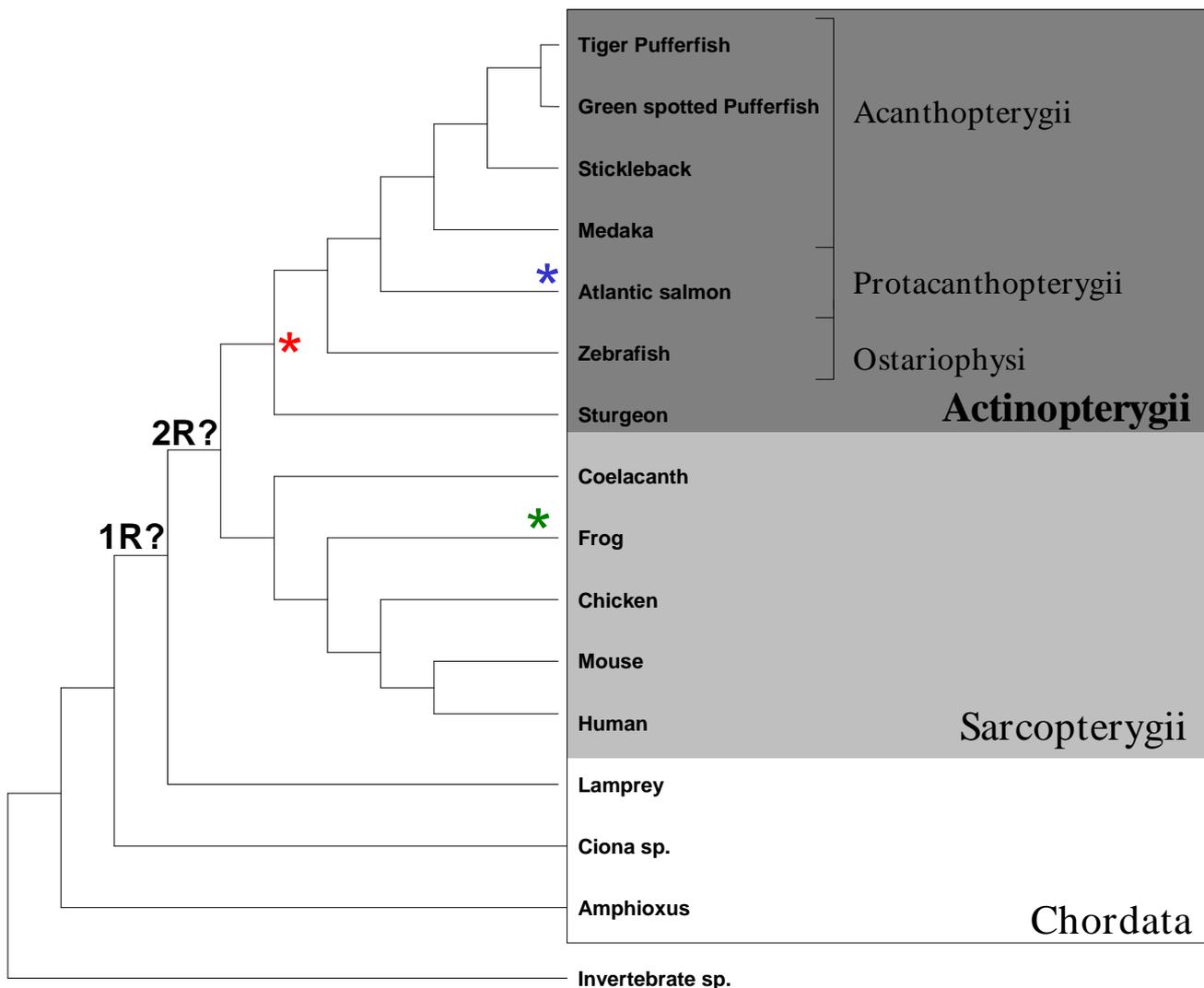


Fig. 1.1. Taxonomic relationships of major model species within the phylum Chordata and the inferred timing of duplication events during chordate evolution. Stars show the position of unequivocal polyploidization events. The red star shows the position of the teleost WGD event during basal teleost evolution (Jaillon et al., 2004). The blue star shows the position of the tetraploidization of the Salmonidae teleost lineage (Allendorf and Thorgaard, 1984) and the green star represents the allotetraploidization of the genome of the frog, *X. laevis* (Bisbee et al., 1977). 1R? and 2R? respectively shows the suspected timing of the 1R and 2R duplication events (based on Wolfe, 2001). For a full list of chordate taxa that have a history of polyploidization, refer to Otto and Whitton (2000). The relationships of vertebrate taxa is based on Benton and Donoghue (2007), except for the position of the Protacanthopterygii, which is based on (Nelson, 2006) and the position of lower chordates/invertebrate sp., which is based on Delsuc et al. (2006).

relative to mammalian relatives (Jaillon et al., 2004). The general consensus for the timing of the teleost WGD is between 320-350 Mya (Vandepoole et al., 2004; Christoffels et al., 2004).

More recently, further WGD events have occurred within several teleost lineages (reviewed in Otto and Whitton, 2000; Le Comber and Smith, 2004). Of particular relevance to the current project is the tetraploidization of the Salmonid genome (Allendorf and Thorgaard, 1984). The affected taxa include the subfamilies Coregoninae (e.g. whitefishes), Thymallinae (e.g. graylings) and Salmoninae (salmon, trout and charrs) (Phillips and Rab, 2001). It has been estimated that 50-75% of paralogues that arose during the salmonid WGD have been conserved (Bailey et al., 1978). Thus the salmonid WGD can be thought of as a 4R duplication, and the potential exists for any invertebrate/basal deuterostome gene to be retained as sixteen paralogues in salmonid species, when eight/four copies have been respectively conserved in other teleosts/diploid tetrapods. For example, it was shown that the last common ancestor of Atlantic salmon and rainbow trout had fourteen *hox* clusters, compared to seven in most non-salmonid teleosts (Moghadam et al., 2005).

### *1.5.3 Genome duplication and the fate of paralogues*

Upon genome duplication, two redundant copies of every part of the genome are present including regions that are protein coding, regulatory or otherwise. In terms of protein-coding genes, beneficial mutations are rare and deleterious mutations are common, so it should be expected that as long as one gene fulfils the role of the ancestral gene, the other gene should accumulate mutations until it becomes a pseudogene (Wagner, 1998), which is otherwise known as nonfunctionalization (Force et al., 1999). However, in rare cases when beneficial mutations occur, it is possible that entirely new functions can be conserved in gene paralogues (Ohno, 1970, Walsh, 1995), otherwise known as neofunctionalization (Force et al., 1999). Additionally, paralogues could be stabilised if a selective advantage is gained by having multiple copies of a

gene either through additional dosage or by the buffering effect of redundancy (Zhang, 2003; Nowak et al., 1997; Wagner, 1999; Gu et al., 2003). A seminal paper by Force and co-workers presented another mechanism by which duplicated genes can be retained, which was named subfunctionalization (Force et al., 1999). Subfunctionalization is hypothesised to occur under the duplication/degeneration/complementation (DDC) model (Force et al., 1999). Subfunctions can be defined as DNA features that have been duplicated in a polyploidization event and subsequently have the potential to mutate independently from one another (Force et al., 1999). The DDC model predicts that after duplication, the two descendents of a gene will each initially have a full complement of subfunctions that underlie the common function or expression pattern of their mother sequence. When degenerative mutations halt different subfunctions in the respective paralogues, the transcription of both is then necessary to fulfil the ancestral role. The theory of subfunctionalization has been experimentally supported for several teleost gene duplicates relative to a single orthologue in the ancestral lineage e.g. *engrailed* (Force et al., 1999), *myod* (Delalande and Rescan, 1999), *sox9* (Klüver et al., 2005) and *proopiomelanocortin* (de Souza et al., 2005). Furthermore, it has recently proposed that subfunctionalization functions as transition step towards neofunctionalization (He and Zhang, 2005; Rastogi and Liberles, 2005).

## **1.6 Myogenesis**

### *1.6.1 Common features of myogenesis across vertebrates*

Vertebrate skeletal muscle is a post-mitotic terminally differentiated tissue and requires a source of new nuclei for growth and nuclear turnover. The muscle fibre is the unit that makes up skeletal muscle and is formed when myoblasts fuse together to form myotubes, which can be extended by the further absorption of myoblasts, in a process termed hyperplasia (Johnston, 2006). The expansion of muscle fibres occurs by hypertrophy, where myoblasts fuse to existing

fibres and are absorbed, which acts to maintain a constant ratio of nuclear material to cytoplasm as well as to turnover myonuclei (Johnston, 2006). In adult amniotes, post-embryonic muscle growth is accounted for by satellite cells, which are undifferentiated MPCs found beneath the basal lamina of existing muscle fibres (Mauro, 1961), and originate during embryonic stages (Gros et al., 2005). Satellite cells are believed to be common to all vertebrates including teleosts (e.g. Koumans et al., 1991, Hollway et al., 2007), although undifferentiated myogenic precursors can also be found scattered throughout the myotome of adult Atlantic salmon outside of the characteristic basal lamina position (Johnston et al., 2003a). The basic cellular processes leading to muscle growth are well conserved across vertebrates and include the specification of a stem cell lineage to become myoblasts, myoblast self-replication (proliferation), myoblast-differentiation, myoblast migration and finally myoblast fusion (summarised in Fig. 1.2) (Johnston, 2006). While the phenotypic context in which these cellular processes are exhibited during development can be divergent in different vertebrates, the genes regulating these steps are generally strongly conserved (e.g. Rescan, 2001; Johnston et al., 2003b). The aim of the following section is two fold: 1. to review the current literature on the most important genes and proteins regulating teleost myogenesis, taking lessons from other vertebrate models, and 2. to describe the cellular processes underlying muscle formation throughout teleost ontogeny, with an emphasis on the accompanying molecular regulation.

### *1.6.2 Teleost myotomal muscle: patterns and innovations*

The myotomal muscle of adult teleosts is comprised of numerous myotomes that are separated by connective tissue called myosepta and arranged axially as overlapping segments in a complex 3-D pattern (reviewed in Videler, 1993). Swimming undulations are generated by the activation of myotome segments from the head downwards leading to the bending of the body profile and thrust is generated from the resistance between the body and caudal/paired fin/s and the external

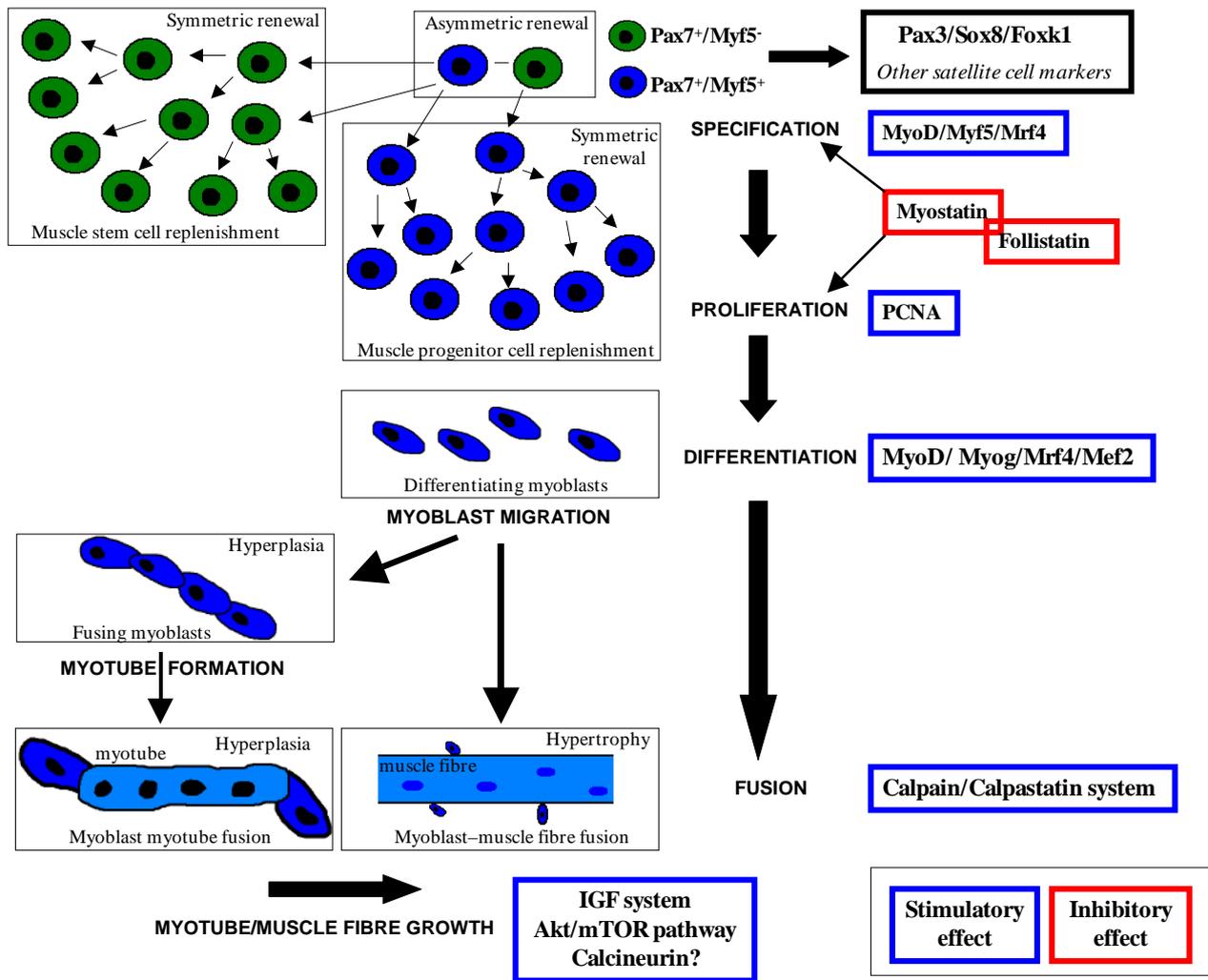


Fig. 1.2. A simple model of post-embryonic vertebrate myogenesis indicating the principal cellular events and the main regulatory genes. The model of satellite cell physiology is based on the recent findings of Kuang et al. (2007) and the model of myoblast-fusion events is based on Johnston (2006). Resident muscle satellite cells can be found between the muscle fibre basal lamina and sarcolemma and are a heterogeneous population that either express Pax7 (and have never expressed MRFs) or Pax7 and Myf5. The Pax7<sup>+</sup>/Myf5<sup>-</sup> population are stem cells, while the Pax7<sup>+</sup>/Myf5<sup>+</sup> cells differentiate more readily into muscle. Both populations are able to replenish the pool of satellite stem cells or committed muscle progenitors (see main text section 1.6.6). While the most common pattern of satellite cell proliferation is through symmetric division, with each type of satellite cell dividing to produce two clonal daughter cells, each cell type can also divide asymmetrically to produce a satellite stem cell and MPC. Other satellite cell markers include the transcription factors Sox8 and Foxk1 (sections 1.6.11-1.6.12). In addition to Myf5, its paralogues MyoD and Mrf4 are also essential myoblast specification factors (section 1.6.3-1.6.4). During myoblast proliferation,

(Fig. 1.2 continued) PCNA (proliferating cell nuclear antigen) is upregulated (Johnston, 2006). Mstn is a negative regulator of muscle growth acting to inhibit both the specification and proliferation of myoblasts by targeting MyoD family and paired box transcription factors (section 1.6.7). In turn Fst is a powerful positive regulator of skeletal myogenesis acting to directly inhibit Mstn and also functioning through other less characterised pathways (section 1.6.8). For viable differentiation of myoblasts, the MRFs Myog, MyoD and Mrf4 and their interactions with MEF2 family members are essential (section 1.6.3-1.6.5). Migrating myoblasts can then contribute to muscle growth by hyperplasia (new fibre production) or by hypertrophy (absorption of myoblasts into existing fibres for nuclear accretion). Hyperplasia, is absent (barring certain stimuli) in post-fetal mammals, but present until around 45% of the final body length in teleosts, and involves the fusion of myoblasts to each other, to form myotubes which can be extended by further fusing myoblasts. Formed myotubes begin to take on the muscle fibre phenotype upon the initiation of myofibrillargenesis. For muscle fibre growth new nuclei are required to maintain the ratio of nuclear material to cytoplasm. This happens through myoblast-muscle fibre fusion events and occurs during all growth stages where differentiated fibres are present. Myoblast fusion events are thought to involve Calpains and their inhibitory protein Calpastatin (section 1.6.13). Fibre growth occurs until muscle fibres reach a maximal size based on diffusional constraints, and this process is thought to be regulated by the IGF-I and IGF-II ligands and their interactions with the Akt/mTOR signalling pathway (section 1.6.10). The calcium activated phosphatase Calcineurin has also been suggested as a possible regulator of fibre hypertrophy (section 1.6.10).

water (reviewed by Videler, 1993; Wardle et al., 1995; Altringham and Ellerby, 1999). Fish muscle has many unique physical and biochemical properties relative to the muscles of terrestrial vertebrates that are almost certainly adaptations to producing propulsion in an aquatic medium. An increasing body velocity in water does not require a linear increase in energy expenditure as is the case when moving in a terrestrial environment (Bennett, 1978). In fact the power required for swimming increases as velocity and body size rise (Johnston, 1981) and to 'burst' through the water to catch prey or avoid being eaten requires a proportionally greater power investment than to swim at a sustained slower speed. Up to 70% of teleost body mass is myotomal muscle that is primarily recruited for burst propulsion, but also composes an energy store utilised during starvation, migration or maturation and is thus an important part of the animal's energy budget. This is a large increase compared to terrestrial animals; for example, skeletal muscle typically composes from 21-33% of total body mass in young fit humans (Kim et al., 2002). However, the overwhelming proportion of myotomal muscle in the teleost body is sustained without much negative cost due to the neutral buoyancy of fishes in water (Johnston, 1981).

The separate energy requirements for sustained and burst swimming almost certainly accounts for the anatomical separation of fish muscle fibres into distinct layers, the bulk for burst swimming and a smaller proportion for sustained swimming (reviewed in Johnston, 1981; Bone, 1978; Bone, 1989). The different muscle types can be easily identified in any muscle steak cross-section as they are arranged in distinct layers (see Fig. 1.3), unlike the muscle fibre arrangement in amniotes where different fibre-types form a mosaic pattern (Currie and Ingham, 2001). In teleosts, the main bulk of the myotome is formed of fast-twitch fibres recruited for burst swimming, metabolising ATP at a rate four times greater than slow muscle and producing higher power outputs (Johnston et al., 1977; Altringham and Johnston, 1990). Fast-twitch fibres are mainly fuelled by anaerobic metabolism and have relatively low vascularisation and mitochondrial content (Bone, 1978; Johnston, 1981). At the periphery of the fast myotome and

contributing ~0.5-30% to the total myotomal muscle (Greer-Walker and Pull, 1975), is a layer of slow-twitch fibres, which are recruited during activities requiring sustainable swimming and fuelled by aerobic metabolism (Johnston et al., 1977; Rome et al., 1984). Slow fibres are smaller in diameter than fast fibres, heavily vascularised and contain high concentrations of myoglobin and high volume densities of mitochondria (Bone, 1978). Unsurprisingly, the proportion of the myotome made up of slow fibres provides a strong indication of teleost life-style. Active pelagic families like the Scombridae have a higher proportion of slow fibres, compared to the myotomes of benthic predators and deep sea fishes which are almost entirely composed of fast muscles (Johnston, 1981). Additionally, another muscle fibre type exists (the intermediate fibres) in some teleosts groups, which are situated between, and have intermediate physiological properties to, the slow and fast fibres (Johnston et al., 1977; Johnston, 1981).

In elasmobranchs, there is a complete division of labour between fast and slow muscles for respective burst and sustained swimming speeds (Bone, 1966). The slow fibres of all representatives of the Chondrichthyes and Osteichthyes are multiply innervated and are activated through junction potentials (Bone, 1964; Johnston, 1981). The fast fibres of the Chondrichthyes as well as many Actinopterygians including the holosteans, dipnoans and certain basal teleost taxa, are focally innervated at the end of the myosepta and are activated by action potentials (Johnston, 1981; Bone, 1989). However, an innovation of the majority of teleost taxa is that fast fibres are multiply innervated and have been observed to be activated by both junction and action potentials (Hudson, 1969; Johnston, 1981; Bone, 1978, 1989). Further, several teleosts with this innervation pattern have been shown to recruit fast muscle fibres for sustained swimming movements (Johnston et al., 1977; Bone et al. 1978; Hudson, 1973). For example, the polyinnervated fast fibres of carp were recruited at moderate swimming speeds and action potentials were only recorded at higher swim speeds, whereas the focally innervated fast fibres

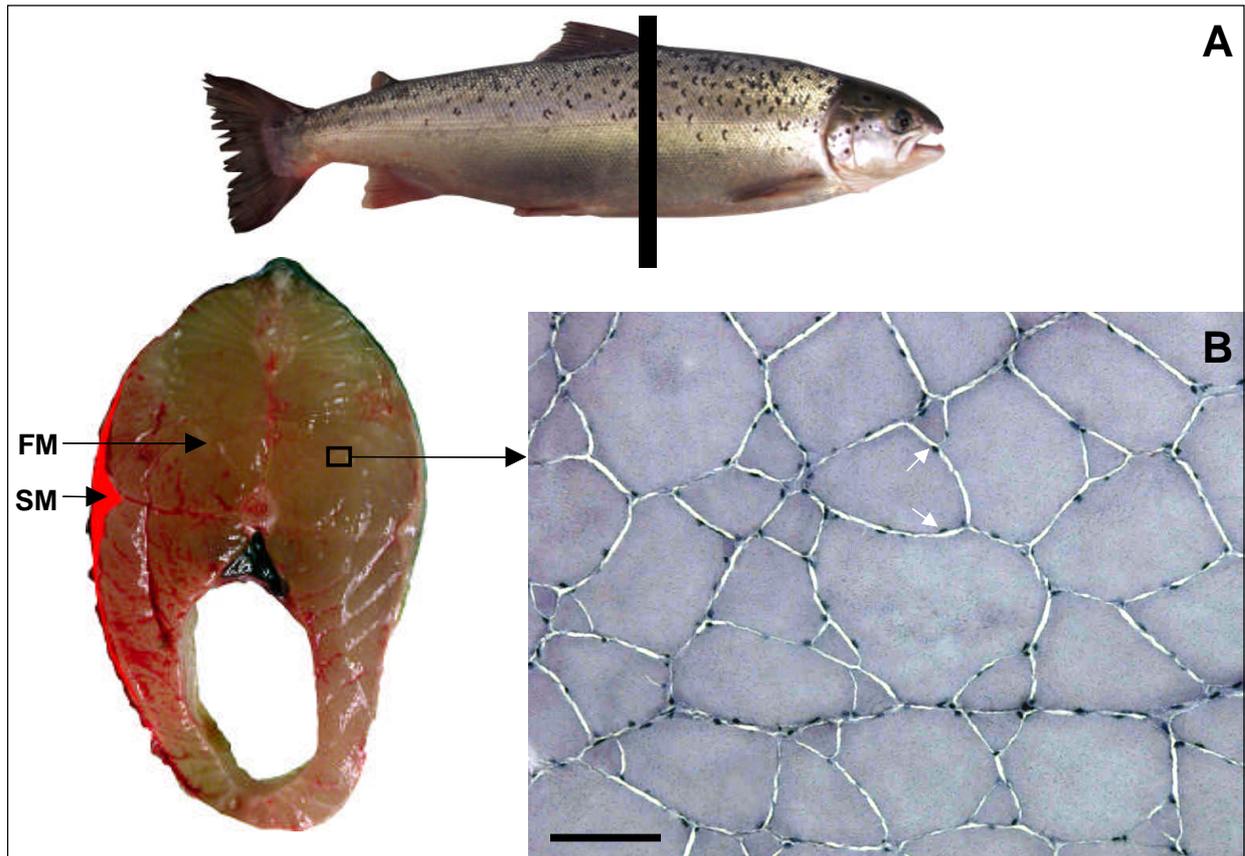


Fig. 1.3. **A.** A whole-body cross section at the level of the first dorsal ray fin of an adult Atlantic salmon shows the archetypal teleost muscle phenotype where the bulk of the muscle is formed of fast-twitch muscle fibres (FM) and is flanked peripherally by a thinner layer of slow-twitch muscle fibres (SM) (filled red). Atlantic salmon are devoid of intermediate fibres. **B.** Shows, a 7  $\mu\text{m}$  cross section taken through the steak that was stained with haematoxylin, revealing the typical gradient of muscle fibre diameters observed in adult teleosts that portray mosaic hyperplasia. Muscle fibre nuclei can be observed throughout the cross section (e.g. white arrows). Scale bar is 100  $\mu\text{m}$ .

of herring were solely recruited for burst swimming (Bone et al., 1978). However, the recruitment of polyinnervated fast fibres in sustained swimming is probably not simply a case of activation through junction potentials, since polyinnervated fibres were observed to produce solely an action potential when stimulated in the sculpin *Myoxocephalus scorpius* (Altringham and Johnston, 1988). It has also been shown that the level of vascularisation and mitochondrial content of fast fibres varies between different teleost taxa, with polyinnervated species possessing greater aerobic capacities than focally innervated species (Johnston and Moon, 1981; Johnston et al, 1983). Additionally, for certain species, such as carp and brook trout (*Salvelinus fontinalis*) the aerobic capacity of fast fibres is probably sufficient to sustain slow swimming speeds (Johnston and Moon, 1981).

### *1.6.3 Myogenic regulatory factors: the master switches for muscle specific gene transcription*

The discovery of the MyoD family of basic helix-loop helix transcription factors, otherwise known as the myogenic regulatory factors (MRFs), was a landmark for understanding the molecular basis of myogenesis (for reviews see Edmonson and Olson, 1993, Rudnicki and Jaenisch, 1995, Berkes and Tapscott, 2005). Remarkably, the ectopic expression of MyoD (Davis et al., 1987), Myog (Edmonson and Olson, 1989), Myf5 (Braun et al., 1989) and Mrf4 (also known as Myf6) (Rhodes and Konieczny, 1989) can convert cell lines including fibroblasts, hepatocytes and adipocytes into myogenic cells, albeit with different efficacies (Weintraub et al., 1991a; Edmonson and Olson, 1993; Berkes and Tapscott, 2005). This common trait is explained by their potent ability to initiate the transcription of muscle genes, via two conserved motifs, the basic region and Helix-Loop-Helix (together the bHLH), which are conserved between MRF proteins at a level of around 80% (Edmondson and Olson, 1993). The HLH domain allows MRFs to dimerise to themselves, to other MRFs, or to their cousin proteins, the bHLH containing E-Proteins, which show ubiquitous tissue expression patterns (Tapscott, 2005; Berkes and Tapscott, 2005). Examples of E-proteins to which MRFs dimerise include E2A (Murre et al., 1989), HEB

(Hu et al., 1992), ITF (Henthorn et al., 1990) and E12 (Spinner et al., 2002). These dimers generally bind (via the basic regions) to a nucleotide motif termed the E-box (CANNTG), which is conserved in the regulatory regions of almost all muscle genes (Tapscott, 2005). Interestingly, while the binding of MyoD-E-protein heterodimers to E-boxes of muscle genes requires a full complement of the arginine residue clusters conserved in the basic region (Shklover et al., 2007), MRF homodimers are also able to bind to the regulatory regions of a subset of muscle genes that form tetraplex structures (Yafe et al., 2005), as long as all arginine-clusters are not compromised (Shklover et al., 2007). The different stringency in DNA binding of these MRF homodimers versus MRF-E-protein complexes could provide MRFs with distinct muscle-gene targets (Shklover et al., 2007).

While the binding of MRFs to regulatory DNA sequences is the underlying mechanism behind the initiation and maintenance of myogenesis, this model is oversimplified, since E-boxes are prevalent throughout the genome and can be targeted by several bHLH factors apart from MRFs (Tapscott, 2005) and MyoD has also been shown to repress the transcription of muscle differentiation genes (Mal and Harter, 2003). However, the known reasons limiting promiscuous binding of MRFs to non-myogenic targets and further allowing the correct temporal binding of MRFs to muscle genes are many and mainly beyond the scope of this review (but for comprehensive reviews, see Tapscott, 2005; Berkes and Tapscott, 2005).

#### *1.6.4 Lessons from knockout studies of the MRFs*

Knockout mice are mutants carrying a non-functional form of a gene of interest, which can then be studied to better understand the function of the ablated protein coded for within the wild-type locus. A historical perspective on knockout mice can be found in Koller and Smithies (1992) and Capecchi (2005) and the following description is derived from these papers and from a commercial protocol obtained from the University of Texas Transgenic Mouse Facility

(<http://www.utmb.edu/>). Mouse embryonic stem (ES) cells can be cultured *in vitro*, where they retain their pleuropotency and can be injected into mouse blastocysts and then implanted into the uterus of a pregnant foster-mouse, where they proceed to develop normally. Subsequently, chimeric pups are born which carry some of the ES cell genome, which can be then passed on to their offspring and this process is manipulated to produce knockout mice. A mutated (non-functional) form of a gene of interest (generally held within a selectable vector, which may also have a reporter gene attached) is injected into the mouse ES cell line by electroporation, which is then inserted into the genome at the appropriate locus through homologous recombination. To ensure a successful recombination event, two homologous nucleotide sequences are required either side of the mutant-carrying vector. The rare ES cells that have uptaken the mutant gene are selected by their chemical resistance and further screened by PCR or Southern blotting to check that the mutant gene is positioned at the desired locus. The mutant ES cells are then cultured, injected into mouse blastocysts and reimplanted into a pseudopregnant mouse. The corresponding pups are chimeras and will have developed some tissues that have grown from the mutant stem cells. The chimeras are then bred to produce mice heterozygous (+/-) for the mutant gene, which are bred with one another to produce animals that are homozygous (-/-) for the mutant gene i.e. knockout mice.

Knockout studies in mice have provided important insight into the protein functions of the different MRFs *in vivo* (summarised in Fig. 1.4) Myf5 and MyoD are essential for normal myogenic determination since mice lacking myoblasts and subsequently skeletal muscle fibres were produced in *myf5/myod* double mutants (Rudnicki et al., 1993; Kablar et al., 2003). When *myf5* and *myod* are individually nulled in mice, a viable skeletal muscle phenotype is observed in each case, which is not overtly different from the wildtype (Braun et al., 1992; Rudnicki et al., 1992). Closer inspection revealed that some disparity exists within their function as *myod* -/-

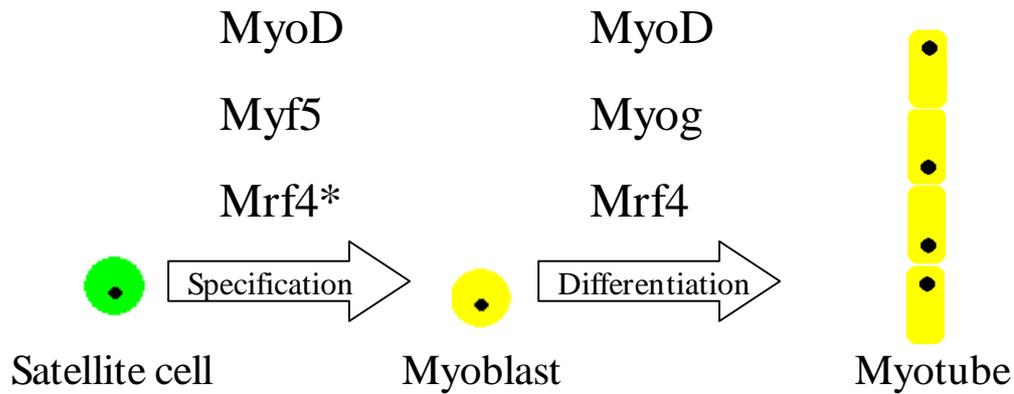


Fig. 1.4. A simple model demonstrating the individual and overlapping roles of the different MRFs in myogenesis. MyoD and Myf5 are specification factors with unique and overlapping target genes. In mice mutants lacking viable genes for either MyoD or Myf5, myogenesis proceeds normally if Mrf4 expression is not disrupted, indicating that Mrf4 can also act as a myogenic specification factor in mammals. MyoD, Myog and Mrf4 function as muscle differentiation factors. In the mammalian embryo, Myog is indispensable for myogenesis. \*In teleosts, the dual role of Mrf4 in myogenic specification and differentiation is seemingly not conserved, since zebrafish morphants lacking viable MyoD/Myf5 translation also lack a normal myogenic program and Mrf4 expression.

mice had delayed development of limb and brachial arch musculature, whereas *myf5*<sup>-/-</sup> mice have delayed development of paraspinal and rib musculature (Kablar et al., 1997). This redundancy is conserved in zebrafish, since the individual knockdown of MyoD or Myf5 protein by morpholino-antisense RNA has little effect on embryonic muscle growth, but the ablation of both proteins abolishes the myogenic program (Hammond et al., 2007). In mice mutants lacking *myog*, committed myoblasts are present, but skeletal muscle does not differentiate properly and animals die shortly after birth (Hasty et al., 1993; Nabeshima 1993). Thus, the role of Myog as a differentiation factor in early mammalian development cannot be compensated by other MRFs. Interestingly, the post-embryonic deletion of *myog* in mice is non-lethal, does not disrupt the differentiation program and only lead to a moderate (30%) reduction in body size (Knapp et al., 2006). *Mrf4* seems to have a dual role in myogenesis since cultured mice myoblasts lacking Myog expression can enter a normal program of differentiation upon the overexpression of *Mrf4* but not MyoD (Sumariwalla and Klien 2001; Myer et al., 2001). However when *Mrf4* expression is not compromised in *myf5/myod* double null mice, skeletal muscle formed normally indicating another role in myogenic determination (Kassar-Duchossoy et al., 2004). This feature of *Mrf4* is seemingly not conserved in teleosts, since *mrf4* transcripts are absent in MyoD/Myf5 double knockdown zebrafish (Hinits, et al., 2007). A hypothesis for the underlying reason for this finding is described in chapter 7 (section 7.5.2). To summarise, while the functions of the MRF are partially redundant, each protein has a specific role, with Myog being indispensable for normal muscle differentiation in early stages of growth.

In addition to the bHLH, other domains exist in MRFs that are essential to their function. At the N-terminal of MyoD, a highly acidic domain is present that functions to activate transcription (Weintraub et al., 1991b). Two regions are known to underlie the ability of MRFs to initiate endogenous gene expression and to activate a subset of target genes in inactive chromatin (including *myog*). These are 1. a histidine-cysteine-rich (H/C) domain just N-terminal to the

basic region in MyoD which is conserved in Myf5 and to a lesser extent in Myog and 2. a motif towards the C-terminal of MRFs, which forms a putative amphipathic-helix (the helix-3) (Gerber et al., 1997; Bergstrom and Tapscott, 2001; Berkes et al., 2004). It was shown specifically, that the helix-3 confers the ability of different MRFs to initiate the expression of endogenous muscle genes (Bergstrom and Tapscott, 2001). This motif is most conserved in MyoD/Myf5 and invertebrate MyoD-orthologues, partially with Mrf4 and poorly with Myog. Chimeras replacing the helix-3 of MyoD with equivalent sequences of Mrf4 and invertebrate orthologues were able to efficiently activate endogenous muscle genes, whereas the equivalent Myog chimera was not (Bergstrom and Tapscott, 2001). Thus, the helix-3 provides a mechanism whereby the different MRFs function as specification or differentiation factors (Bergstrom and Tapscott, 2001). Additionally, the binding of MyoD to the Myog promoter occurs through an interaction of the H/C and helix-3 domains with the homeodomain protein Pbx, which is bound to *myog* prior to its expression (Berkes et al., 2004).

All four MRFs are conserved in teleost fishes and have been cloned and characterised in numerous species, including representatives from a range of evolutionary diverse taxa. The characterisation of teleost MRFs has generally involved determining embryonic and adult mRNA expression patterns and has focused on either zebrafish or commercially important species. Work with the zebrafish model has established how MRFs are regulated during embryogenesis by the use of various mutant lines with aberrant myogenic phenotypes and recently by morpholino-antisense RNA 'knockdown' of MyoD and Myf5. MRF characterisation has formed one of the focal points of this project and has included 1. the characterisation of MyoD evolution in teleost fishes (chapter 3). 2. establishing the differential expression of *myod* paralogues in salmon embryos and adults (chapter 4) and 3. the potential role of temperature induced heterochonies in *myod*, *myog*, *mrf4* and *myf5* expression as a potential mechanism regulating plasticity of the muscle phenotype (chapter 7).

### *1.6.5 The MEF-2 transcription factors co-regulate differentiation with MRFs*

The myocyte enhancer factor-2 (Mef2) gene family are essential myogenic co-regulators that combine with MRFs to control the differentiation of skeletal muscle (Reviewed in Olson et al., 1995; Black and Olson, 1998; Yun and Wold, 1996). Mef2 proteins are part of a family of transcription factors termed MADS after the first 4 members to be identified (Black and Olson, 1998). Gossett et al. (1989) first discovered that Mef2 bound to an A/T rich element in the enhancer region of muscle creatine kinase and it has since been realised that virtually all muscle genes have Mef2 binding sites in their regulatory regions, including the MRFs (Black and Olson, 1998). The MADS-box contains 57 amino acids (AAs) situated toward the extreme N terminus of the MEF2 protein. This region binds DNA but requires the adjacent Mef2 domain, a 29 AA motif, for high-affinity DNA binding and dimerization (Black and Olson, 1998). In higher vertebrates, there are four distinct MEF2 genes (Mef2a, b, c, d) which each contain the MADS box and Mef2 domains. Mef2 genes are not expressed in skeletal muscle until after MRF transcripts have accumulated (Edmondson et al., 1994) and unlike MRFs, the Mef2 factors cannot alone activate myogenesis, but instead combine with MRFs through protein-protein interactions at the MADS box and bHLH to vastly improve the efficiency of the myogenic program (Kaushal et al., 1994; Molkenin et al., 1995).

### *1.6.6 Paired box transcription factors and satellite cells*

Paired box (Pax) transcription factors commonly contain a DNA binding domain, the paired box, and function as key regulators of the growth and development of a number of animal tissues including muscle (reviewed in Buckingham and Relaix, 2007). Mutant mice lacking *pax7* died two weeks post-birth, at a weight a third of wild-type littermates, and with smaller, but normally organised muscle fibres (Seale et al., 2000). Additionally, cells cultured from the mutant muscle were completely devoid of myoblasts (Seale et al., 2000). Thus, *pax7* is essential for satellite cell

function. *pax7* and *pax3* are expressed in embryonic stages of amniote development in the dermomyotome compartment of the somite. *pax7* is expressed in the central domain of the dermomyotome whereas *pax3* is initially expressed before segmentation, prior to its restriction to hypaxial/epaxial regions of the dermomyotome (Buckingham and Relaix, 2007). Recently, a Pax3/Pax7 dependent population of myogenic progenitors was identified in amniotes that originated from the central dermomyotome and are utilised in embryonic, fetal and adult muscle growth (Relaix et al., 2005; Gros et al., 2005; Kassam-Duchossoy et al., 2005). These Pax7<sup>+</sup>/Pax3<sup>+</sup> cells can be found quiescent in fetal and adult stages under the basal lamina of existing muscle fibres in the classic position of satellite cells (Buckingham and Relaix, 2007). Mammalian satellite cells are a heterogeneous population and 90% were observed to readily differentiate into muscle and could be identified by their dual expression of Pax7 and Myf5 (Kuang et al., 2007). The remaining 10% of satellite cells expressed Pax7 and had never expressed Myf5 (Kuang et al., 2007). The majority of satellite cell divisions (90%) were symmetric with the mitotic spindle running parallel to the fibre and resulted in either two Pax7<sup>+</sup>/Myf5<sup>-</sup> or two Pax7<sup>+</sup>/Myf5<sup>+</sup> cells that remained in contact with the basal lamina and plasmalemma (Kuang et al., 2007). However, 10% of divisions were observed to be asymmetric and occurred perpendicular to the fibre. These divisions produced one Pax7<sup>+</sup>/Myf5<sup>+</sup> satellite cell that lost contact with the basal lamina, but retained contact with the plasmalemma and one Pax7<sup>+</sup>/Myf5<sup>-</sup> satellite stem cell that remained in contact with the basal lamina alone (Kuang et al., 2007). These findings raise the possibility that the asymmetrically derived Pax7<sup>+</sup>/Myf5<sup>+</sup> cells directly fuse with muscle fibre (Cossu and Tajbakhsh, 2007). Additionally, the perpendicular satellite cell divisions were stimulated upon muscle injury (Kuang et al., 2007), suggesting that asymmetric/symmetric satellite cell divisions may be differentially induced in different contexts of muscle growth or injury. Thus satellite cells can, through symmetric or asymmetric divisions, quickly replenish the stem cell population or add to a population of committed muscle progenitors that can quickly differentiate when new muscle growth is required (Fig. 1.2). It has

also been suggested that during amniote embryogenesis, asymmetric divisions of Pax7<sup>+</sup>/Pax3<sup>+</sup>/Myf5<sup>-</sup> stem cells in the dermomyotome could contribute one Pax7<sup>+</sup>/Pax3<sup>+</sup>/Myf5<sup>-</sup> stem cell as well as a more committed Pax7<sup>+</sup>/Pax3<sup>+</sup>/Myf5<sup>+</sup> myoblast that enters the myotome to establish new muscle fibres (Cossu and Tajbakhsh, 2007).

*pax7* and/or *pax3* have been cloned and characterised in several teleost fishes, including zebrafish (Seo et al., 1998; Feng et al., 2006) Atlantic salmon (Gotensparre et al., 2006) and pearlfish (*Rutilus frisii*) (Steinbacher et al., 2006). The expression of these genes is indicative of a conserved role in both embryonic and post-embryonic myogenesis (see below section: 1.7). For example, *pax7* was shown by *in situ* hybridization to be specifically expressed in MPCs of adult Atlantic salmon fast muscle (Gotensparre et al., 2006). In this study, *pax7* was used as an embryonic marker of a recently discovered population of MPCs originating in the anterior somite (e.g. Hollway et al., 2007 see section 1.7.3) during Atlantic salmon embryogenesis (chapters 5, 6, 7).

#### 1.6.7 Myostatin: the most potent known negative regulator of myogenesis

A member of the transforming growth family (TGF- $\beta$ ), Myostatin (Mstn), or GDF8, has a pivotal role in regulating mammalian skeletal muscle growth (McPherron et al., 1997; reviewed in Lee, 2004). In mammals, unlike many other TGF- $\beta$  proteins, *mstn* is expressed principally in the skeletal muscle of adults and myotome of embryos (Lee, 2004). Mice lacking *mstn* have up to three times more muscle than wild-type counterparts as a result of both hyperplasia and hypertrophy, but are otherwise healthy (McPherron et al., 1997). In mammals, several reports of naturally occurring *mstn* mutations are correlated with either increased muscularity, or athletic prowess (reviewed in Lee, 2007a). For example mutations in the *mstn* gene account for double muscled breeds of cattle (Grobet et al., 1997; Kambadur et al., 1997; McPherron and Lee, 1997). Recently a human child with a mutation causing mis-spliced *mstn* mRNA and limited translation

of Mstn was reported to be exceptionally muscular and strong but otherwise healthy (Schuelke et al., 2004). The negative effect of Mstn on muscle growth occurs during myoblast proliferation/differentiation (Lee, 2004). Mstn has been shown to directly downregulate the expression of several MRFs (Langley et al., 2002; Ríos et al., 2002) as well as *pax3* (Amthor et al., 2002, 2004).

Mstn is thus an exceptionally interesting protein from the perspective of the potential treatment of human muscle disease (e.g. Bogdanovich et al., 2002) as well as possible applications to enhance meat quality outcomes in the agriculture industry. This interest was evident in the aquaculture industry and *mstn* was cloned in numerous teleost species. However, it is currently debatable whether the function of Mstn as a ‘super-inhibitor’ of myogenesis is present in the teleosts. *Mstn* is conserved as two genes in most teleosts (*mstn1* and *mstn2*), as a result of the WGD (discovered by Maccatrozzo et al., 2001). The nomenclature for teleost *mstn* genes used throughout this thesis is based on the phylogenetic study of Kerr and co-workers (Kerr et al., 2005). Salmonid teleosts have two paralogues of each of the teleost genes (named *mstn1a/1b* and *2a/2b* [*2b* is a pseudogene]) as a result of the genome tetraploidization (Kerr et al., 2005; Garikipati et al., 2006, 2007). However, unlike the muscle-specific expression of *mstn* in mammalian adults, teleost *mstn* mRNAs are expressed in a large range of tissues (e.g. Østbye et al., 2001; Garikipati et al., 2006). Additionally, whereas *mstn* mRNA is abundant in mouse and chicken embryos at limits detectable by *in situ* hybridisation (Lee, 2004; Amthor et al., 1996), *mstn* is not detectable by this approach in teleosts (e.g. Kerr et al., 2005; Xu et al., 2003). Additionally, overexpression of the Mstn propeptide (a negative regulator of Mstn, see below section 1.6.9) in zebrafish only produced slight changes in the muscle fibre phenotype (~12% increase in hyperplasia, no change in muscle fibre size and no change in MRF expression), compared to the striking phenotype produced in mice overexpressing the Mstn propeptide (Lee and McPherron, 2001). Further, vertebrate *mstn* and the closely related gene *GDF11* probably

arose in an ancient duplication event (Xu et al., 2003), supported by a single archetypal *mstn*/*GDF11* gene in the amphioxus (Xing et al., 2007). In amphioxus, this gene is expressed in multiple tissues (Xing et al., 2007) and it is possible that *mstn* genes of teleosts have retained this feature, whereas the strong-inhibitory role in mammalian myogenesis arose following the split of the Sarcopterygians/Actinopterygians. In support of this hypothesis, certain expression domains of teleost *mstn* are conserved with mammalian *mstn* and *GDF11* (Østbye et al., 2001) suggesting that this occurrence could potentially be as a result of the different portioning of regulatory subfunctions in different vertebrate lineages.

#### *1.6.8 Follistatin a regulator of Mstn and of myogenesis*

Follistatin (Fst) is a secreted glycoprotein that was first identified as a potent inhibitor of follicle stimulating hormone secretion (Esch et al., 1987; Phillips and Krestor, 1998), explained by its strong binding affinity for the TGF- $\beta$  protein Activin (Nakamura et al., 1990). Fst is expressed in multiple tissues and plays many important roles during vertebrate development, including the regulation of myogenesis (For reviews see Patel, 1998; Phillips and Krestor, 1998). As well as Activin, Fst is known to bind other TGF- $\beta$  proteins, including Mstn (Amthor et al., 2004) and Bone Morphogenic Proteins (BMPs) (Amthor et al., 2002). Several lines of evidence suggest that Fst has an important role regulating amniote myogenesis in embryos and adults. Fst is expressed in chick embryos in a spatiotemporal pattern consistent with an interaction with Mstn (Amthor et al., 1996, 2004) and was shown in this *in vivo* context to inhibit the negative effect of Mstn on the expression of *myod* and *pax3*, though direct binding (Amthor et al., 2004). Further, Fst was shown to directly promote *pax3* expression in embryonic chick wing-buds, by reversibly binding to Bmp7 and presenting this growth factor at low enough levels to stimulate myogenesis rather than at higher levels, which induced apoptosis (Amthor et al., 2002). Additionally, knockout mice lacking a functional *fst* gene showed retarded growth and had a reduced mass of diaphragm and intercostal muscles (Matzuk et al. 1995), whereas transgenic mice overexpressing

Fst showed a dramatic increase in muscle mass with a 66% increase in fibre number and 28% increase in mean fibre diameter (Lee and McPherron, 2001). Furthermore, when overexpressed in *mstn*-null mice, Fst induced an even greater increase in muscle mass, principally through fibre hypertrophy, indicating a capacity to affect other ligands involved in muscle fibre growth (Lee, 2007b). Additionally, the positive regulation of myogenesis associated with deacetylase inhibitor treatment in mice model systems occurs via upregulation of Fst (Minetti et al. 2006, Iezzi et al. 2004). Finally, the stimulatory effect of nitric oxide on the fusion of embryonic/adult myoblasts occurs through the upregulation of *fst* expression (Pisconti et al., 2006). Thus Fst is a protein of great interest from a perspective of potential therapeutic applications for human muscle diseases (e.g. Nakatani et al., 2007).

To date *fst* has been cloned in a few teleost species and characterised solely from the perspective of anterior-posterior patterning (Bauer et al., 1998; Dal-Pra, 2006). Thus, the role of Fst in teleost myogenesis has yet to be examined. In chapter 5 of this study, the *fst* gene of Atlantic salmon was cloned and characterised from the perspective of a newly characterised feature of teleost embryonic myogenesis: the anterior somite compartment.

#### *1.6.9 Other known regulators of Mstn*

In addition to Fst, several other proteins are known to inhibit Mstn. Mstn, in a common fashion for TGF- $\beta$  proteins, is synthesised as a precursor molecule that is proteolysed into a biologically active species (Lee, 2004; McPherron et al., 1997). A 24 AA signal peptide is removed and Mstn is separated into a propeptide at the N-terminal and a biologically active peptide at the C-terminal (McPherron et al., 1997). The C-terminal remains bound to the propeptide, sequestered from biological activity until the proteolysis of the propeptide, by members of the BMP1 family of proteases (Lee, 2004). The binding of the propeptide thus acts to regulate the titre of biologically active Mstn, and transgenic mice overexpressing the propeptide had a similar

phenotype to *mstn* knockout mice (Lee and McPherron, 2001). Additionally, two proteins, FLRG and GASP1 were recorded to bind to Mstn at high affinities *in vitro* and were bound to the biologically active form of Mstn in the blood of mammals (Hill et al., 2002, 2003). Transgenic mice overexpressing FLRG showed nearly double the muscle mass of wild-type littermates (Lee, 2007b). However, the *in vivo* function of GASP1 in myogenesis has yet to be established.

Decorin, a small proteoglycan that circulates in the extracellular matrix was also shown *in vitro* to bind to Mstn and to rescue its inhibitory effect on the proliferation of cultured myoblasts (Miura et al., 2006). Further it was found that decorin was downregulated in myotubes cultured from dystrophic mice, when Mstn was upregulated (Zanotti et al., 2007) and that myotubes overexpressing Decorin differentiated at a higher rate than wild-type myotubes with concurrent upregulation of MRF genes/*fst* and downregulation of *mstn* (Li et al., 2007). Additionally, it was shown that Decorin neutralized the stimulatory effect of Mstn on muscle fibrosis, again with concurrent upregulation of *fst* (Zhu et al., 2007). In teleosts, the role of decorin during myogenesis remains to be investigated. As part of this project a full-coding sequence for decorin was cloned in Atlantic salmon and submitted to GenBank for future use.

#### *1.6.10 The Insulin like growth factors, Calcineurin and the Akt/mTOR pathway*

The insulin-like growth factors (IGFs) have a focal role in regulating growth patterns in vertebrates (reviewed in Florini et al., 1996; Nakae et al., 2001; Wood et al., 2005). The IGF system is formed of two ligands, IGF-I, IGF-II, their receptors and six IGF-binding proteins. This system is known to regulate the global pattern of organism growth through the regulation of cellular growth, proliferation, migration, survival and differentiation (Wood et al., 2005). The context in which these cellular processes are induced in a tissue specific manner are dependent on the specific interactions of the different components of the IGF system as well as their interactions with other proteins from several well characterised signalling cascades (Glass, 2003,

Heszele and Price, 2004). Knockout studies have indicated the importance of the IGF system in early vertebrate development (reviewed in Wood, 1995). Mutant mice lacking genes for either IGF-I or IGF-II show severe growth retardation and are only 60% the size of wildtype littermates (DeChiara et al., 1990; Liu et al., 1993), whereas mutants lacking the gene coding for the IGF-I receptor or functional genes coding for both IGF-I and IGF-II die shortly after birth with an even more severe dwarf phenotype (Liu et al., 1993). The importance of the IGFs to embryonic myogenesis is emphasised by the identical muscle phenotype shared between IGF and Myog mutant mice (Florini et al., 1996). In addition to their essential role in prenatal development, IGFs have important roles in regulating adult myogenesis and both IGF-I and IGF-II have been implicated in regulating muscle fibre hypertrophy and satellite cell differentiation (Florini et al., 1996; Armand et al., 2004), while IGF-I has also been shown to induce satellite cell proliferation (Rommel et al., 2001; Shavlakadze et al., 2005). The IGF system is well characterised in fishes, and is highly conserved in terms of its function in growth (e.g. Méndez et al., 2001; Castillo et al., 2002, 2004). However a full discussion of the IGF system in fish is beyond the scope of this review (for a comprehensive review see Wood, 2005).

The maintenance of skeletal muscle mass occurs by hypertrophy and involves a balance of two contrasting processes, protein synthesis for muscle growth and protein breakdown during muscle atrophy. IGF-I stimulates several well-characterised signalling pathways involved in protein production, while simultaneously inhibiting atrophy inducing pathways (system reviewed in Glass, 2003, 2005; Heszele and Price, 2004). Two hypertrophy-signalling cascades downstream of IGF have been principally investigated, the Calcineurin/NFAT pathway and the Akt/mTOR signalling cascade. The calcium/Calmodulin-activated phosphatase Calcineurin has been suggested to have a role in IGF-I mediated hypertrophy as well as in establishing the phenotype of muscle fibres (reviewed in Michel et al., 2004). Two papers in the same issue of *Nature* concurrently reported that IGF-I induced hypertrophy observed in rodent muscle cell culture was

mediated through Calcineurin (Semsarian et al., 1999; Musarò et al., 1999). Semsarian and co-workers showed that IGF-I treatment activated Calcineurin through mobilisation of intracellular calcium, and was accompanied by the dephosphorylation and nuclear translocation of the NFAT transcription factor (Semsarian et al., 1999). Musarò et al. showed that IGF-I or the expression of activated Calcineurin induced the expression of the GATA-2 transcription factor in a subset of differentiated myonuclei (where Calcineurin had accumulated) and interacted with another transcription factor (NF-AT-c) to initiate muscle gene expression (Musarò et al., 1999). Additionally, it was shown *in vivo* that Calcineurin inhibition blocked the normal hypertrophy response in overloaded mice muscles (Dunn et al., 1999).

However, the exact role of Calcineurin as a direct regulator of hypertrophy has been contentious (see Glass, 2003; Dunn et al., 2002 vs. Yancopoulos and Glass, 2002). Importantly, it has been observed that Calcineurin activity decreased during fibre hypertrophy, rather than increased, which would be expected if it played a direct role in this process (Bodine et al., 2001). Additionally, whereas some groups have reported that blocking Calcineurin with Cyclosporin A (CsA) inhibits hypertrophy (Dunn et al., 1999) others have found no effect (Musarò et al., 2001; Serrano et al., 2001; Rommel et al., 2001). Further, transgenic mice have been produced that overexpress activated Calcineurin in their skeletal muscle (Naya et al., 2000; Dunn et al., 2000). The muscles of these mice had augmented numbers of slow muscle fibres but showed no evidence of Calcineurin induced hypertrophy (Naya et al. 2000; Dunn et al., 2000). This is consistent with several reports that Calcineurin plays an important role during muscle fibre-phenotype transformation (Chin et al., 1998; Naya et al., 2000; Parsons et al., 2003, 2004).

However, the Akt /mTOR (mammalian target of Rapamycin) signalling pathway has been unequivocally demonstrated to be essential for muscle fibre hypertrophy, independently of Calcineurin (Rommel et al., 2001; Bodine et al., 2001). This pathway, which is highly complex

and only discussed here briefly, is initiated by IGF-I and culminates in protein synthesis through p70S6 kinase and Elongation Factor 2B, (for exhaustive reviews see Glass, 2003; 2005). The induction of mammalian muscle hypertrophy through stimulation of IGF-I initiates a cascade of phosphorylation events activating members of the PI(3)K/Akt/mTOR signalling pathway (Rommel et al. 2001; Bodine et al., 2001). IGF-I indirectly activates PI(3)K, which has a known role regulating skeletal muscle hypertrophy (Glass, 2003). Activation of PI(3)K leads to the translocation and phosphorylation of Akt, which in turn phosphorylates many proteins with multiple biological functions (Glass, 2003, 2005). Muscle fibre hypertrophy was stimulated in transgenic mice overexpressing activated Akt (Lai et al., 2004), whereas knockout mice lacking an Akt1 gene were viable, but showed around a 20% reduction in body mass (Chen et al., 2001). Conversely, Akt is dephosphorylated in atrophic muscles (Glass, 2005). Thus Akt signalling is alone sufficient to induce muscle fibre hypertrophy. Akt is also known to directly phosphorylate and activate mTOR during muscle fibre hypertrophy and the inhibition of mTOR reduces fibre hypertrophy (Glass, 2003, 2005). Activated mTOR is a necessary mediator for protein translation during muscle hypertrophy as it is known to phosphorylate translation initiation factor 4E and the ribosomal protein 70S6 kinase while concurrently deactivating PHAS-1, a binding protein (and negative regulator) of translation initiation factor 4E (Glass, 2005; Heszele and Price, 2004).

#### *1.6.11 Forkhead box proteins and satellite cells*

The forkhead box protein Foxk1 otherwise known as Myocyte Nuclear Factor (MNF) is a transcription factor with DNA binding capacity due to the presence of a 110 AA motif termed forkhead (Bassel-Duby et al., 1994). In the mouse embryo, Foxk1 is primarily expressed in the somites, brain and myocardium, but is localized to satellite cells in adults (Garry et al., 1997). The murine *foxk1* gene is alternatively spliced to produce two mRNAs coding for the protein isoforms Foxk1 $\alpha$  and Foxk1 $\beta$  (Yang et al., 1997). Foxk1 $\alpha$  is expressed in satellite cells committed to the myogenic lineage where as Foxk1 $\beta$  is expressed in quiescent satellite cells

(Yang et al., 1997; Hawke and Garry, 2001). Mice lacking a viable *foxk1* gene have an impaired skeletal muscle where satellite cells portray G<sub>0</sub>/G<sub>1</sub> arrest in the cell cycle following upregulation of the cyclin-dependent kinase inhibitor, p21(CIP) (Garry et al., 2000; Hawke et al., 2002). The *foxk1* gene has been characterised in the pufferfish *T. rubripes*, where it is expressed as three splice-variants, one of which is expressed in MPCs of adult myotomal fast muscle (Fernandes et al., 2007). Additionally, antibodies to Foxk1 protein have also been used to successfully quantify MPCs in Arctic charr (Johnston et al., 2004). In the current project a partial cDNA for Foxk1 was cloned in Atlantic salmon and submitted to GenBank for future use.

#### *1.6.12 Sox8, another marker of vertebrate satellite cells*

Sox proteins are a family of transcription factors that have a high mobility DNA-binding domain (Wegner, 1999). Sox8 is a member of the Sox protein subgroup E and is closely related to Sox9 and Sox10 (Bowles et al., 2000). Sox8 expression is confined to satellite cells in the skeletal muscle of adult mice (Schmidt et al., 2003) where it acts to inhibit the expression of MyoD and Myog and disrupts myoblast differentiation (Schmidt et al., 2003). Thus Sox8 can be used as a marker of undifferentiated satellite cells. The first Sox8 gene to be cloned in a fish species was originally termed SoxP1 after its isolation from a rainbow trout pituitary gland cDNA library (Ito et al., 1995) but was recently reclassified (Schepers et. al., 2002). Full coding cDNAs for *sox8* has also been cloned in the pufferfishes *T. nigroviridis* and *T. rubripes* where it was shown by *in situ* hybridization of adult fast myotomal muscle to be expressed specifically in putative MPCs (Mackenzie, 2006). In the current project a full coding cDNA of *sox8* was cloned in Atlantic salmon and submitted to GenBank for future characterisation purposes.

#### *1.6.13 The Calpain-Calpastatin system*

Calpains are a group of calcium requiring proteins that are generally expressed ubiquitously and specifically degrade multiple protein substrates (Croall and DeMartino, 1991). Calpastatin is a

protein that inhibits the activity of calpain and is likewise universally expressed in most tissues (Goll et al., 2003). For an exhaustive review on the role of the Calpain system refer to Goll et al (2003) or Croall and De Martino (1991). In terms of the cellular events of myogenesis the Calpain/Calpastatin system has an important role at the point of myoblast fusion (Temm-Grove et al., 1999; Barnoy et al., 1996; 2005). Calpain is expressed throughout myogenesis and causes proteins within membranes to degrade, creating fusion-competent domains in fusing myoblasts (Barnoy et al., 1998). Conversely, Calpastatin is transiently downregulated prior to myoblast fusion, thus relaxing constraints on Calpain membrane proteolysis (Barnoy et al., 1996, 2005). The microinjection of Calpain into cultured mammalian myoblasts massively increased the rate of myoblast fusion, whereas the microinjection of Calpastatin ablated myoblast fusion (Temm-Grove et al., 1999). Similarly Calpastatin overexpression also prevented myoblast fusion and the breakdown of proteins associated with fusion (Barnoy et al., 2005). Further, Myog expression was effectively ablated by Calpastatin overexpression, indicating that myogenic differentiation was arrested (Barnoy et al., 2005). *In vivo*, transgenic mice overexpressing Calpastatin had significantly smaller muscle fibres than wild-type mice without showing any overall change in body mass (Tidball and Spencer, 2002).

The degrading action of Calpain on muscle proteins is also present post-mortem. Thus the levels of Calpain and Calpastatin in muscle can strongly contribute to post-mortem flesh quality attributes, meaning this system is of great interest to agriculture and aquaculture industries (e.g. Koohmaraie, 1996). For example, various natural mutations in *calpain/calpastatin* genes have been associated with flesh quality and tenderness in beef (e.g. Casas et al., 2006) and pigs (e.g. Ciobanu et al., 2004). However, the use of Calpain/Calpastatin as markers of fish flesh quality is in its infancy and is limited to a few papers (e.g. Salem et al., 2005). As part of the current project, seven distinct full coding cDNAs of calpastatin were cloned in Atlantic salmon, representing at least three genes and several splice variants. These sequences will contribute to

the ongoing study of the calpain/calpastatin system in our laboratory with regards to Atlantic salmon growth and flesh quality.

#### *1.6.14 MicroRNAs and myogenesis*

While not a focus of the current project, microRNAs (miRNAs) are briefly considered here due to the recent realisation of their enormous importance in regulating a wide range of physiological processes including myogenesis. MiRNAs are short (~22 nucleotide) RNA sequences found in animals and plants, and involved in regulating gene expression through cleavage of mRNAs or by inhibiting their translation (reviewed in Bartel et al., 2004; Ambros, 2004). Hundreds of miRNAs have been characterised throughout the animal kingdom, with many conserved across taxa (Ambros, 2004). Several miRNAs are expressed specifically in muscle and are targeted by MRFs during myogenic specification and differentiation (Rao et al., 2006; Rosenberg et al., 2006). For example, miR206, induces myogenic differentiation of C2C12 myoblasts by inhibiting the expression of genes that repress MRFs (Kim et al., 2006). Additionally, a single transcribed sequence codes for the miRNAs miR1 and miR133 (Chen et al., 2006), which is directly activated by muscle-specific transcription factors such as MyoD and Mef2 (Liu et al., 2007). miR1 promotes myogenesis by repressing histone deacetylase 4, with concurrent upregulation of muscle regulatory genes like MyoD, Myog and Mef2, whereas miR133 repressed serum response factor, a known inhibitor of myoblast proliferation (Chen et al., 2006). Remarkably, a single nucleotide transition was identified in the 3' UTR of *mstn* that was shown to allow the binding of miR1 and miR206 and contribute to the double muscling phenotype observed in Texel sheep (Clop et al., 2006). A further miRNA, miR214, functions during embryonic myogenesis to regulate titres of Hedgehog (Hh) morphogens, with implications for the specification of different resident myoblast populations (Flynt et al., 2007).

## ***1.7 Phases of teleost myogenesis***

### *1.7.1 Embryonic myogenesis: early patterns and onset of slow muscle formation*

The onset of muscle growth occurs prior to somitogenesis in teleosts, whereas in amniotes myogenesis is not initiated until somitogenesis is well established. This is likely a reflection of the external fertilization of teleost embryos, and the accompanying early requirement to generate movement upon hatching to avoid predation (Currie and Ingham, 2001). The unique teleost myotome phenotype, where fast and slow muscles are found in discrete compartments is established during embryogenesis (Devoto et al., 1996; Johnston and McLay, 1997). It has been suggested that this feature allows teleost larvae to simultaneously generate burst and constant swimming speeds, for respective predator evasion and food foraging (Koumans and Akster, 1995). The zebrafish model has been studied most extensively in terms of the mechanisms underlying embryonic myogenesis, due to its transparency allowing the counting and tracing of cells, the presence of several mutant lines with defective muscle development (Currie and Ingham, 2001) and the established use of morpholino antisense RNA to ‘knockdown’ genes of interest. However, the *in situ* expression patterns of myogenic genes are also well documented in several other teleost species and have generally indicated a highly conserved molecular origin for embryonic muscle.

A summary of the events leading to the formation of embryonic muscle in teleosts is summarized in Fig. 1.5. At the onset of the segmentation period, the first somites arise when cells in the rostral presomitic mesoderm condense and form epithelia around freely organised mesenchymal cells (Stickney et al., 2000; Currie and Ingham, 2001). Somites arise in this manner in a rostral to caudal direction until a final number is reached which is species dependent. For example, zebrafish have around 30 somites (Kimmel et al., 1995), whereas salmonids have in the region of 65 somites. Unlike the situation in amniotes, teleost somites are not overtly compartmentalized.

The ventral somite gives rise to the sclerotome, the future skeleton, which expressed markers such as *pax9* and *twist* (Nornes et al. 1996; Morin-Kensicki and Eisen 1997) and is greatly reduced relative to amniotes (Currie and Ingham, 2001). The main bulk of the somite will become the myotome, which is composed of various MPC populations (Fig. 1.5, A). The first MPCs become specified to myogenesis before the onset of segmentation at the end of gastrulation. In zebrafish, these so called adaxial cells can be identified prior to somite formation by their expression of *myod* (Weinberg et al., 1996), a pattern widely conserved in other teleost species (Delalande and Rescan 1999; Temple et al., 2001; Cole et al., 2004; Galloway et al., 2006) (Fig. 1.5, A, B). The adaxial cells are first observed in zebrafish as a four by five layer of notochord-adjacent, cube shaped cells that then transform into a single stack of ~twenty elongated cells (Devoto et al. 1996), under the influence of the actin regulatory protein Cyclase associated protein-1 (Daggett et al., 2007). The elongated adaxial cells migrate radially away from the notochord to form a superficial layer of slow muscle fibres (Devoto et al., 1996, Cortés et al., 2003) (Fig. 1.5, C). This cellular pattern in zebrafish is seemingly conserved in other teleosts, evidenced by an apparent migration of slow muscle specific gene transcripts across the rainbow trout myotome (Chauvigné et al., 2005). A subset of adaxial cells called the muscle pioneers, does not migrate radially, and stay close to the notochord at the level of the horizontal myoseptum (Devoto et al., 1996; Felsenfeld et al., 1991). Pioneer cells express engrailed proteins (Hatta et al., 1991; Ekker et al., 1992) and are the first fibres to show contractile activities (van Raamsdonk et al., 1978).

### *1.7.2 Slow muscle specification and Hh morphogens*

The specification of slow myoblasts in the teleost embryo is critically dependent on midline secretions of Hh morphogens in a dose dependent manner (reviewed in Ingham and Kim, 2005). Zebrafish mutants lacking a differentiated notochord show defects in slow muscle development (Currie and Ingham, 2001). The zebrafish notochord expresses, and presumably secretes

orthologues of the drosophila Hh glycoprotein called Sonic Hh (SHh) and Echidna Hh (EHh) (Currie and Ingham, 1996). When Shh is overexpressed in zebrafish embryos, slow muscle is formed at the expense of fast muscle (Currie and Ingham, 1996; Blagden et al., 1997; Du et al., 1997). Further, disruption to Hh signalling disrupts slow muscle development. cAMP-dependent protein kinase is a known blocker of Hh and its overexpression, or activation through forskolin treatment, caused a loss of slow muscle fibres (Barresi et al., 2000). Additionally, the overexpression of Patched 1, a negative regulator of Hh signalling, ablates slow muscle formation (Lewis et al., 1999). Further, zebrafish mutations affecting genes involved in the Hh signalling pathway such as *gli2* and *smoothened* lead to a near complete loss of embryonic slow muscle fibres (Lewis et al., 1999; Barresi et al., 2000). Downstream of Hh signalling, Blimp1, which is coded for within the zebrafish *u-boot* mutation, is an important regulator of adaxial cells. In *u-boot* mutants, adaxial myoblasts abort slow myogenesis and adopt a fast muscle fate (Roy et al., 2001). *blimp1* is expressed in adaxial cells after the expression of MRFs, but before muscle specific genes transcripts like slow myosin heavy chain are present (Baxendale et al., 2004). *blimp1* expression is then downregulated before adaxial cells start their radial migration suggesting a role for this gene in priming but not maintaining slow muscle development (Baxendale et al., 2004). Importantly, Blimp1 overexpression can rescue mutants devoid of normal slow-muscle development in the absence of Hh signalling (Baxendale et al., 2004).

### 1.7.3. Formation of embryonic fast muscle

The bulk of the cells of the presomitic mesoderm differentiate into the fast-twitch muscle fibres in the wake of migrating adaxial cells (Devoto et al., 1996; Cortes et al., 2003; Henry and Amacher, 2004) (Fig. 1.5, C). Fast twitch MPCs can be identified in the posterior epithelial zebrafish somite by their expression of MRFs like *myod*, *myog* and transiently, *myf5* (Weinberg et al., 1996; Coutelle et al., 2001; Devoto et al., 1996) (Fig. 1.5, A). The expression domains of *myod* and *myog* then extend anteriorly, to encompass the whole posterior-anterior domain of

the somite (Weinberg et al., 1996; Coutelle et al., 2001; Groves et al., 2005). This expression pattern marks the elongation/fusion of myoblasts to span the entire posterior-anterior length of the somite (Stellabotte et al., 2007) and is conserved in other teleost species (e.g. Delalande and Rescan, 1999; Tan and Du, 2002). Further, this event is coincident with the 90° rotation of the somite that occurs during mid-segmentation stages of zebrafish embryogenesis (Hollway et al., 2007; section 1.7.4, Fig. 1.5). An additional population of fast fibres differentiates close to the zebrafish notochord and express *engrailed* at the same time as slow-pioneer cells, although at lower levels (Wolff et al., 2003). In common with muscle pioneers and adaxial cells, the medial fast fibres are regulated by Hh morphogens, in a dose dependent manner (Wolff et al, 2003; Ingham and Kim, 2005). The differentiation of posterior-lateral fast fibres but not the medial fast fibres is dependent on the fibroblast growth factor (Fgf8) (Groves et al., 2005). Fgf8 is not required for the initiation of *myf5* expression in the posterior somite, but specifically for the progression of *myod* to *myog* expression during the differentiation of the lateral fast fibres (Groves et al., 2005). Additionally, retinoic acid was shown to induce *myod* expression in the zebrafish somite and initiate fast muscle differentiation, through the activation of *fgf8* (Hamade et al., 2006). Homeodomain Pbx proteins are specifically required for the specification and differentiation of fast, but not slow muscle cells in zebrafish embryos, likely by targeting the promoters of a subset of fast-muscle genes that also contain MyoD binding sites including, but not limited to *Myog* (Maves et al., 2007).

#### *1.7.4 The anterior somite of teleosts and continuing muscle growth*

Whereas the MPCs of the posterior somite establish the embryonic fast muscle fibres, recent studies have identified a further source of fast-MPCs originating in the anterior part of the epithelial somite, that contribute to late-embryonic and post-embryonic muscle growth (Hollway et al., 2007; Stellabotte et al., 2007). It had previously been shown that the zebrafish anterior somite expressed *pax7/pax3* before these expression domains shifted anteriorally, to encompass

the external cell layer (Groves et al., 2005; Devoto et al., 2006; Feng et al., 2006; Hammond et al., 2007), a feature previously described in zebrafish (Waterman, 1969). An inverse relationship was observed between the titres of Hh or Fgf8 and the number of Pax3/Pax7 expressing cells in the zebrafish somite (Feng et al., 2006; Hammond et al., 2007). Additionally, the ablation of Myod/Myf5 caused an increase in Pax3/Pax7 expressing cells (Hammond et al., 2007). Stellabotte et al. (2007) used single cell lineage tracing to show that Pax7<sup>+</sup> anterior cells migrate laterally to found a self-renewing MPC population in the external cell-layer that during late-embryonic and larval growth, provided cells that re-entered the myotome through the slow fibre layer and differentiated into fast muscle (Fig. 1.5). A similar cell-lineage tracing approach was used by Hollway et al., 2007, who again showed that Pax7<sup>+</sup> cells of the anterior somite founded the external cell layer, which was a source of fast myotome growth in the late embryo. Additionally, Pax7<sup>+</sup> MPCs were shown to migrate to reside under the basal lamina of existing muscle fibres in larval stages (Fig. 1.5, D) (Hollway et al., 2007). Further, it was shown that in addition to myotome growth, progenitors of the hypaxial and fin muscles as well as the skin were derived from the anterior somite (Hollway et al., 2007). The work of Hollway and co-authors also expanded on the mechanism by which the external cell layer was formed, showing that the whole cell-profile of the somite rotates 90° from its starting position (Hollway et al., 2007) (Fig. 1.5, B-C). This somite rotation, which occurred during mid-somitogenesis and was finished by the end of segmentation, was shown by a morpholino-knockdown approach to be dependent on the expression of the cytokine Sdf1a, and its receptors Cxcr4a and Cxcr4b (Hollway et al., 2007). Thus, the study of Hollway et al. (2007) is of particular importance, since it both proved and effectively extended the hypothesis of several authors (Devoto et al., 2006; Feng et al., 2006; Groves et al., 2005), whom suggested that based on the conserved embryonic expression of *pax* genes, a teleost equivalent of the amniote dermomyotome exists.

### *1.7.5 Stratified hyperplasia*

In teleosts, the growth of the embryonic myotome continues when new muscle fibres are added in localized germinal zones by a process termed stratified hyperplasia (Rowlerson and Veggetti, 2001; Rescan, 2005). This process occurs at the end of segmentation and regions of new fibre production are marked by the expression of MRF genes (Barresi et al., 2001; Steinbacher et al., 2006, 2007). New fast fibres are added at the border of the slow layer and periphery of the existing fast myotome as well as the dorsal/ventral extremities of the myotome, whereas slow fibres are added at the dorsal/ventral edge of the existing slow layer (Stellabotte and Devoto, 2007). This process leads to a gradient of increasing fast fibre sizes moving from the peripheral to medial myotome (Rowlerson and Veggetti, 2001; Stellabotte et al., 2007). It is known that the fast fibre population for stratified hyperplasia are derived from the external layer (described above) (Stellabotte and Devoto, 2007; Stellabotte et al., 2007; Hollway et al., 2007). It was also suggested that slow fibres added during stratified hyperplasia were sourced from the external cell layer (Veggetti et al., 1990). Since stratified hyperplasia of slow muscle fibres occurs in zebrafish mutants lacking Hh signalling (Barresi et al., 2001), a distinct mechanism is required for this process relative to the signals regulating embryonic adaxial cells. Considering the anatomical proximity of the slow-layer and external cell-layer, and the fact that external cells have to migrate through the slow layer (Stellabotte et al., 2007) the simplest mechanism would be that a sub-population of the external cells also contribute to the slow fibres during stratified hyperplasia.

### *1.7.6 Mosaic hyperplasia and maximum fibre diameter*

Mosaic hyperplasia is the final stage of new fibre production in fast muscle and is so called as myogenic cells scattered throughout the myotome differentiate to form myotubes at the surface of existing fibres, leading to a mosaic of different fibre sizes in a myotome cross section (Johnston, 2006; Rowlerson and Veggetti, 2001) (see Fig. 1.3). The bulk of muscle growth in

post-embryonic life-stages thus occurs by the overlapping processes of new fibre production and the concurrent expansion of existing fibres by hypertrophy (Johnston, 2006). Mosaic hyperplasia generally occurs until around 45% of the final body length (Weatherly et al., 1988) and in most species accounts for the majority of the final fibre number (e.g. Johnston et al., 2003a). For example in Atlantic salmon and zebrafish, 95 and 70% of the final number of fast fibres were respectively derived from mosaic hyperplasia (Johnston, 2006). However, in certain small teleost species, or for some species (or morphs of a particular species) under strong selective pressure, mosaic hyperplasia is absent and the final fibre number is accounted for solely by embryonic and stratified myotube production e.g. in the guppy *Poecilia reticulata* (Veggetti et al., 1992), stickleback (Johnston, I.A. unpublished results), in dwarf salmon (Johnston et al., 2005) and in species of certain Notothenioid families (Johnston et al., 2003c).

Muscle fibres absorb myonuclei until they reach a diameter where diffusional constraints become limiting (Johnston et al. 2003c). It has been proposed that the maximum diameter of a muscle fibre is a trade-off between the need to limit diffusional constraints, which increases with increasing fibre diameter, and the cost of activating ion pumps across the fibre membrane, which increases with a higher fibre surface area per unit of muscle (i.e. more smaller diameter fibres) (Johnston, 2006). Since myotomal muscle forms such a large part of the teleost body, which in turn requires a large proportion of the energy budget (over 20%) to maintain ionic homeostasis (Jobling, 1994), this trade-off has significant implications for establishing the mean fibre diameter of teleosts, which is a reflection of the fibre number per unit of muscle cross-sectional area. This hypothesis has been tested from an ecological perspective, where it has been shown that 1. muscle fibre number was a function of body size in a large number of closely related Notothenioid teleosts living in similar conditions (Johnston, 2003c) 2. that these Notothenioids, which live in extreme cold conditions and have evolved a sluggish lifestyle, were observed to have a reduced fibre number relative to non-Antarctic relatives, accounted for by a large increase

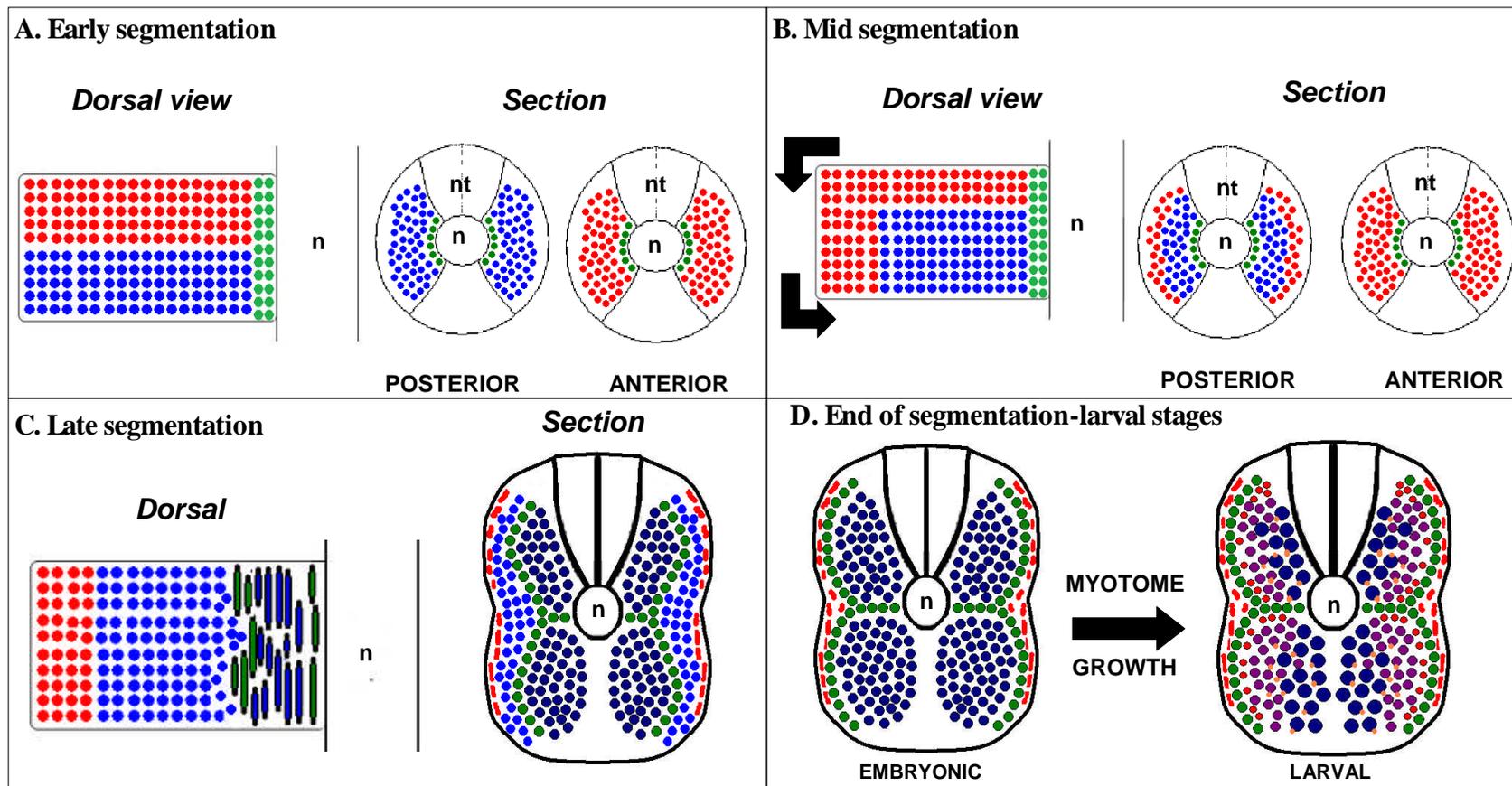
in fibre diameter (Johnston, 2003c) and 3. that salmonid populations that have evolved a dwarf phenotype, also exhibited a reduction in fibre number of 50-75% (Johnston, et al., 2004, 2005).

#### *1.7.7 What is the source of muscle growth during mosaic hyperplasia?*

Several teleost species retain a layer of undifferentiated cells external to the myotome during juvenile stages, including zebrafish (Waterman, 1969), sea bass (*Dicentrarchus labrax*) (Veggetti et al., 1990), herring (*C. Harengus*) (Johnston, 1993), trout (Steinbacher et al. 2007) and pearlfish (Stoiber et al., 1998). In zebrafish, trout and pearlfish these cells express *pax7* suggesting they are likely the daughter cells of earlier anterior somite derived *pax7*<sup>+</sup> cells and are a known source of stratified hyperplasia (Stellabotte and Devoto, 2007). It is also known that external cells migrate into the myotome to reside quiescent under the basal lamina of existing fibres in larval stages (Hollway et al., 2007). While these cells could provide some progenitors for mosaic hyperplasia, this is currently unproven and requires single-cell lineage tracing work extended beyond larval stages (Stellabotte and Devoto, 2007). Additionally, within the myotome, there are only a small number of *pax7* expressing cells, at a time when mosaic hyperplasia is prominent (Steinbacher et al., 2006; Stellabotte et al., 2007). Thus it is currently unknown whether the external cells contribute exclusively, or are a side population to the majority of muscle growth during post-embryonic growth. However, the fact that an increased fibre number under environmental manipulation lead to a concurrent increase in myonuclear content in existing fibres (Johnston, 2003a; also see chapter 8) is supportive of a single MPC population for hyperplasia and hypertrophy (Johnston, 2006).

#### *1.7.8 Signals regulating myotube production*

In teleost fishes new myotubes are assembled during both embryonic and post-embryonic stages. However, in mammals, myotube production ceases prior to birth (Rowe and Goldspink, 1969), barring stimuli such as exercise and injury, when new fibre production has been demonstrated



**KEY**

- Anterior somite cell
- Embryonic fast myoblast
- Adaxial myoblast
- Slow muscle fibres
- Fast fibre derived from embryonic fast myoblast
- ▬ External cell derived from anterior somite
- Newly differentiated fast fibre derived from the external cell layer
- Fast fibre derived from the external cell layer
- Satellite cell derived from external cell layer

Fig.1.5 Legend is on the next page

Fig. 1.5. A simplified schematic representation of the cellular specification, differentiation and migration of embryonic MPC populations in a generalised teleost. Cell sizes are not scaled and the diagram is largely based on the recent work of Hollway et al. (2007) and Stellabotte et al. (2007), but also considers the cellular fate of the adaxial cells, which was first described by Devoto et al. (1996).

**A.** During early segmentation stages, teleost somites are epithelial in nature and organised into three principal MPC populations identifiable by their characteristic positions and gene expression profiles. Next to the notochord of somites, and extending into the presomitic mesoderm (not shown) are MyoD<sup>+</sup> adaxial myoblasts. In the posterior somite a population of embryonic fast myoblasts reside that express MRFs like MyoD, Myog and transiently Myf5. In the anterior somite compartment, there resides a recently described population of Pax7 expressing MPCs.

**B.** As segmentation proceeds, the epithelial somite begins a dynamic cellular rearrangement where the whole profile of cells rotate 90° from their starting positions. It is by this mechanism that the anterior cells come to lay externally to the myotome and the fast myoblasts are spread throughout the anterior-posterior extent of the somite. During this time the adaxial cells start to terminally differentiate in their medial position within the myotome.

**C.** By late segmentation, somite rotation is complete and all anterior cells are positioned externally to the myotome in the so-called external cell layer (ECL). At a similar time the differentiated adaxial cells begin to migrate laterally across the myotome to eventually lie external to the fast myotome, but internally to the ECL. A subset of adaxial cells don't migrate (the muscle pioneer cells) and reside at the level of the horizontal myoseptum. In the wake of the adaxial cells the fast myoblasts fuse and elongate to form the embryonic fast muscle fibres.

**D.** At the end of segmentation, the adaxial cells have formed a thin layer of slow muscle at the periphery of the fast muscle fibres. From this time onwards, the Pax7<sup>+</sup> cells of the ECL either migrate through the slow muscle layer and differentiate into new fast fibres at the periphery of the fast myotome (i.e. stratified hyperplasia) or replenish the existing ECL MPC population. Stratified hyperplasia gives rise to a characteristic-increasing gradient of fibre diameter moving from the lateral to medial fast myotome at this stage. During larval stages, the ECL cells also migrate to deeper myotome positions where they reside under the basal lamina of established fibres in the characteristic satellite cell position.

(Blaveri et al., 1999; Darr and Schultz 1987). Additionally, zebrafish were shown to produce new myotubes upon muscle injury (Rowlerson et al., 1997). Accordingly, it has been suggested that in healthy teleost muscle where mosaic hyperplasia has ceased, specific inhibitory signals are present that limit the assembly of new myotubes (Johnston, 2006). Corroborating evidence for this hypothesis comes from a limited number of experiments from our laboratory. The first study used suppression subtractive hybridization in the model pufferfish *T. rubripes* to identify four genes that were strongly upregulated in life stages when fibre recruitment had ceased (Fernandes et al., 2005). Furthermore, two of these genes were also upregulated in zebrafish at the precise body length where hyperplasia had ceased (Lee, 2007). Additionally, several miRNAs are respectively up or downregulated in an inverse relationship to the up/down regulation of their target genes in growth stages where hyperplasia was active or had stopped (Parthanemon et al., unpublished). In the future, it will be important to functionally characterise these gene and miRNA candidates in both teleosts and other vertebrates, to understand the cellular levels at which they contribute to myogenesis, their targets and from a comparative perspective, to disentangle the mechanisms by which teleosts are able exhibit eternal hyperplasia. One of the genes identified in the original screen (FRC386; Fernandes et al., 2005) was further characterised in chapter 6 of this study.

## ***1.8 Myogenesis and the environment***

### *1.8.1 Introduction*

From a comparative perspective, the embryonic rearing environment of different teleosts is dependent on reproductive strategy, which is highly variable in different groups (Jalabert, 2005). This probably reflects a trade-off in energy investment between egg number and size relative to the nature of the rearing environment (Roff, 1992). For example, salmonids lay ~2000-3000 eggs

of 4-5 mm diameter in freshwater benthic nests, whereas some marine species like cod and turbot have 500,000-1,000,000 eggs of 1 mm diameter which float freely in the nekton (Jalabert, 2005). While this variation in reproductive tactic leads to a variety of physiological/behavioural experiences of embryos/juveniles from different species, a common feature of all teleost embryos is immobility prior to hatching, excluding the chance to select an environment with favourable conditions and imposing full exposure to the ambient temperature, salinity, oxygen content and pH as well as biotic factors such as predation. In some species, movement in juvenile stages is also limited; for example Atlantic salmon hatchlings stay in their gravel nest until the yolk sac is reabsorbed and even upon emergence to the river stay within a few metres of this area (Armstrong and Nislow, 2006). Thus, the ambient environment largely dictates the physical-chemical conditions of embryonic and juvenile stages of teleost ontogeny with implications for all developmental processes including myogenesis (Johnston, 2006). Recent reviews have summarised the effect of environmental conditions on myogenesis in teleosts (Johnston, 2001b; Johnston and Hall, 2004; Johnston, 2006).

### *1.8.2 Embryonic temperature and myogenesis*

Temperature is thought to be the most important extrinsic factor affecting the cellular environment of organisms (Hochachka and Somero, 2002). Teleosts are ectothermic so their embryonic body temperature is a close reflection of the temperature of the ambient water. Each species has a range of temperatures under which normal development proceeds and the rate of developmental processes tends to increase until an upper lethal limit (Johnston, 2001a). At the extreme limits of range for normal development, organism abnormalities become increasingly common as mortality rates increase (Stockard, 1921). Temperature has been observed to induce phenotypic plasticity in teleost muscle development in two ways: either in the short term, generally through

heterochronic processes or with a persistent effect on the phenotype (Johnston and Hall, 2004; Johnston, 2006).

Temperature induced plasticity in early teleost development has been extensively reviewed (Johnston and Hall, 2004; Johnston, 2006). From a myogenic perspective, heterochronic incidences of temperature-induced plasticity include the altered timing of the occurrence of the embryonic slow muscle layer (Johnston et al., 1995), an altered temporal pattern of muscle innervation (Johnston et al., 1997) as well as the altered timing of expression of developmental-specific isoforms of muscle proteins (Vieira and Johnston, 1992; Johnston et al., 1997; Johnston et al., 2001). The timing of MRF expression was also shown to vary at the mRNA and protein levels with varying temperature in rainbow trout embryos of equivalent growth stages (Xie et al., 2001). Additionally, in herring embryos it was shown that escape performance varied as a function of developmental temperature due to heterochronies in fin and fin muscle development (Johnston et al., 2001). In addition to these heterochronic events, it was also observed that the mitochondrial content of embryonic muscle fibres was dependent on embryonic temperature (Vieira and Johnston, 1992). Additionally, embryonic temperature has been shown to affect the number, size and physiology of muscle fibres in larval stages of teleost development. For example in four separate investigations, Atlantic salmon embryos reared at heated temperatures (8-11°C) versus lower temperatures (between 1.5-5 °C) had less fibres (between ~12-50%) which were of an average smaller diameter (Stickland et al., 1988; Usher et al., 1994; Nathanailides et al., 1995; Johnston and McLay, 1997) and had increased numbers of myonuclei (Nathanailides et al., 1995; Johnston and McLay, 1997). However, while this pattern is conserved across studies, it has also been observed that the temperature-muscle cellularity interaction varied within the same population at the family level (Johnston and McLay, 1997) as well as between reproductively isolated populations of salmon caught from different tributaries of the same river, which

naturally experienced different temperature regimes (Johnston et al., 2000a). Thus, temperatures effect on early Atlantic salmon myogenesis seems to be buffered by genetic and adaptive differences imposed at the individual and population level. The temperature effect on larval muscle fibre number is not limited to Atlantic salmon. For example it has been shown at low relative to high embryonic rearing temperatures that larval muscle fibre number increased in herring (*Clupea harengus*) (Vieira and Johnston, 1992) and cod (*G. morhua*) (Hall and Johnston, 2003) and conversely, decreased in sea bass (*D. labrax*) (Ayala et al., 2000), turbot (*Scophthalmus maximus*) (Calvo and Johnston, 1992) and plaice (*Pleuronectes platessa*) (Brooks and Johnston, 1993).

### *1.8.3 Lasting effects of early rearing temperature on myogenesis*

Early rearing temperature has also been shown to have a persistent impact on the teleost muscle phenotype. Johnston et al. (2003a) reared Atlantic salmon through freshwater stages at either an ambient or heated (by 1-3°C) temperature, before tagging individual fish and transferring them to identical seawater conditions for the rest of the experiment. At the point of seawater transfer the ambient fish were on average around half the size of treated fish. However, by the end of the experiment, the ambient and heated groups were of an equivalent size, owing to a faster growth rate of ambient fish (Johnston et al., 2003a). Ambient fish also showed an increased intensity of myotube production and at the end of the experiment had, per cross section, around 22% more muscle fibres with significantly more MPCs (Johnston et al., 2003a). Further, the myonuclear content of the largest fibres was increased by around 20% in the ambient treatment (Johnston et al., 2003a). Thus, some effect of the temperature treatment was programmed into the resident MPC population between embryonic and freshwater stages that persisted into seawater growth (Johnston, 2006). It is currently unknown at what stage this temperature effect was imposed: e.g. are several populations of MPCs affected during both embryonic and postembryonic growth?

Alternatively, could a short window of temperature in the embryo have an equivalent effect on final muscle fibre recruitment patterns by producing a life-long effect on a single supplying MPC population? To this end, results are presented in chapter 8 from a study investigating how embryonic temperature during a short window of embryogenesis affects the life-long growth trajectory and final muscle phenotype of Atlantic salmon. Additionally, studies on the effect of early temperature regime have generally incorporated two temperature regimes and thus rarely established the norm of reaction response of muscle phenotype to temperature (Johnston, 2006).

### ***1.9. Commercial relevance of plasticity of teleost myogenesis***

During Atlantic salmon aquaculture, downgrading of flesh occurs due to poor fillet colour, flesh softness and blood spots as well as fillet gaping (reviewed in Michie, 2001). Variations in muscle cellularity have important outcomes for fish flesh quality in terms of sensual (taste, texture, smell, colour) as well as processing attributes (reviewed in Johnston, 1999; 2001b). The consumer demand for the texture of fish flesh is firmness i.e. the reverse situation to meat agriculture where tenderness is a desirable trait. Across teleost species a significant inverse-correlation exists between perceived firmness of cooked teleost flesh and the size of muscle fibres i.e. species with smaller fibres were recorded by a taste panel to have firmer flesh (Hurling et al., 1996). This is also true for smoked farmed Atlantic salmon, where a significant positive correlation was observed between an increasing density of muscle fibres and flesh texture as measured by an expert taste panel (Johnston et al., 2000b). Additionally, smoked salmon flesh colour is influenced by the level of muscle fibre recruitment, with more fibres producing a significant increase in colour visualization in the fillet (Johnston et al., 2000b). Further, incidences of Atlantic salmon gaping were observed to cease beyond a threshold muscle fibre density of 95 muscle fibres per mm<sup>2</sup> (Johnston et al., 2002). Finally, it has been shown that 35% of the individual variation observed in growth rate of farmed salmon that had ceased mosaic

hyperplasia, could be accounted for by the final fibre number, which in turn was dependent on the freshwater temperature regime (Johnston et al., 2003a). Thus, phenotypic plasticity of the muscle fibre phenotype with changing environmental variables has implications for the growth rate of farmed salmon as well as its processing attributes and desirability to the consumer. Therefore, a goal of the salmon farming industry must be to understand how varying husbandry practices can manipulate the fibre phenotype in a manner consistent with desirable economic outcomes. In this regard, the effect of embryonic temperature on the Atlantic salmon muscle phenotype at commercially relevant body sizes and concomitant flesh quality attributes are currently uninvestigated.

### ***1.10 Aims and goals***

The aims of this project were as follows:

1. To clone genes involved in Atlantic salmon myogenesis and to develop a molecular toolbox of various cDNAs and primers for future gene expression assays.
2. To characterise a subset of these genes in Atlantic salmon and more generally teleosts. This included establishing their mRNA expression patterns during embryonic and adult myogenesis in Atlantic salmon.
3. Since teleost genes are commonly retained as multiple paralogues, another aim was to establish their evolutionary relationships as inferred by phylogenetic and comparative genomic approaches.
4. To execute a rigorous and statistically sound experiment, where Atlantic salmon embryos are reared at 2, 5, 8 or 10 °C solely from fertilization until the ‘eyed stage’ of embryogenesis and after which time fish are provided with an equal growth opportunity.

5. To use *in situ* hybridization to investigate potential heterochronies in the expression patterns of a subset of myogenic genes characterised in aims 1 and 2, in salmon embryos reared as part of aim 3.
6. To use growth modelling and digital morphometric methods to establish the norm of reaction response of growth trajectory and final muscle fibre phenotype of adult Atlantic salmon sampled from aim 3.

## **Chapter 2. General materials and methods**

### ***2.1 Introduction***

In this chapter, the main methodologies used during this thesis are described. This includes basic computational and experimental molecular biology methods, as well as the set-up of the embryonic temperature experiment. Accordingly, reference to sections of this chapter are made throughout the materials and methods sections of subsequent experimental chapters. A list of manufacturers addresses can be found in Appendix 2.

### ***2.2 Embryonic temperature experiment***

#### ***2.2.1 Experimental set-up***

Broodstock for the study were provided by Salmobreed (Norway, A/S strain) and were obtained from 60 female and 10 male fish. Fertilized eggs were incubated at AKVAFORSK (Sunndalsora, Norway) under four temperature treatments (2, 5, 8 and 10°C) (with 4 trays per treatment and ~1000 eggs per tray) from fertilization (Nov 03) until the embryonic eye became pigmented. Subsequent to the 'eyed stage' (defined below, section 2.2.2), embryos were maintained at a constant 8 °C until first feeding. At this time, fish were transferred to 50 cm diameter circular tanks and the temperature was slowly increased to 12°C over 3 days. Fry were then transferred to EWOS Innovation (Lønningdal, Norway) in May and June 04 and transferred to four replicate 200 cm diameter circular tanks. Fish were separated by treatment during freshwater growth to minimise problems associated with different size ranges of animals from different groups e.g. treatment-dependent feeding hierarchies. The temperature started at 12 °C and then ranged between 2.7 °C and 15 °C depending on ambient conditions over the year. Fish were provided

with continuous light until Jan 2005, and then twelve hours light: twelve hours dark (winter signal) until seawater transfer. Fish were fed a commercial (EWOS, Micro range) diet during freshwater stages.

Fish were transferred to seawater tanks in May 2005 following sedation in 2ppm AQUI-S (AQUI-S New Zealand Ltd) and exposure to the anaesthetic MS222 (Sigma) at 50ppm. Passive Integrator Transponder (PIT) tags (Fish Eagle) were inserted into the ventral gut cavity and one hundred animals from each treatment were randomly transferred to three replicate seawater tanks. Fish were reared under continuous light and were fed a commercial diet (EWOS Pyramid range).

### *2.2.2 Embryonic sampling*

Temperature-induced changes in the rate of development of *S. salar* embryos are proportional over a wide range of non-lethal temperatures (Gorodilov, 1996). The staging system of Gorodilov describes over one hundred morphological states of Atlantic salmon embryonic development and provides a relative age of embryos at any stage, defined by the time it takes to form a somite pair (Ts) during the somitogenesis period. Gorodilov has calculated Ts in Atlantic salmon embryos incubated from 0-13 °C at increments of 0.1 °C (Gorodilov, 1996). Thus, these values can be used to calculate the time necessary to reach a particular morphological state, or a stage when a particular number of somites have formed. Using this approach, embryos from the 2, 5, 8 and 10 °C treatments were sampled at times equivalent to the following seven stages:

1. The end of gastrulation (Ts-50)
2. 1-3 somite stage (ss) (Ts-55)
3. 10-15 ss (Ts-65)

4. 30-40 ss (Ts-90)
5. 45-50 ss (Ts-102)
6. Toward the end of segmentation (Ts-125)
7. Post-segmentation (The 'eyed stage') (Ts 165)

A small puncture was made in the chorion of embryos, which were subsequently fixed overnight in 4 % (m/v) paraformaldehyde (Sigma) in phosphate-buffered saline (PBS) (Sigma). Subsequently embryos were dehydrated in increasingly concentrated methanol (Fisher Scientific) (10 mins in 25%, 50% and 75% (m/v) methanol in PBS, followed by 2 X 10 mins in 100% methanol). Subsequently, embryos were stored for long-term preservation in 100% methanol at –70°C.

### *2.2.3 Post embryonic sampling*

For freshwater stages, 24 fish per treatment (6 per tank) were randomly sampled in Jan 2005. In seawater stages, fish were individually weighed and identified by their PIT tag 5 times from the point of smoltification until the end of the experiment (May and Nov 2005, Mar, Jun and Nov 2006). At each seawater sample point, 18 fish (6 per tank) were randomly selected per temperature treatment and sampled for muscle fibre analyses. The protocol was identical for all samples. All sampling instruments and surfaces were sterilized using both alcohol and RNase Zap (Ambion). Fish were sacrificed with a blow to the head and muscle blocks were prepared immediately. A muscle steak of 1 cm was prepared at the level of the first dorsal fin ray using a sharp knife. The steak was placed on a square of glass over a piece of graph paper and a digital image was recorded using a Nikon CP4500 camera (Nikon). A series of evenly spaced 1cm<sup>2</sup>

muscle blocks were then prepared from the fast myotomal muscle of one side of the steak ranging from 2-6 blocks (depending on fish size). Muscle blocks were mounted on corkboard, coated with cryomatrix (Thermo Electron Corp.) and frozen in isopentane (Fisher Scientific) cooled to near freezing (-159 °C) over liquid nitrogen. Blocks were wrapped in labelled foil and stored in liquid nitrogen. Additionally, fast and slow muscle samples were dissected from the dorsal epaxial myotome of each fish and flash frozen in liquid nitrogen for later nucleic acid extraction.

### ***2.3 Basic computational biology methods***

#### *2.3.1 Gene databases*

Several gene databases and various bioinformatic tools were utilised during this project. These included the National Centre for Biotechnology Information (NCBI, <http://www.ncbi.nih.gov>), the Ensembl genome browser (<http://www.ensembl.org/index.html>) as well as expressed sequence tag (EST) databases including the Salmon Genome Project (<http://www.salmongenome.no/>), the Gene Indices Project (<http://compbio.dfci.harvard.edu/tgi/>) and cGRASP (<http://web.uvic.ca/cbr/grasp/>). The basic local alignment search tool (BLAST) (Altschul et al, 1990) was also used to compare sequences of interest against these databases. The most commonly used BLAST programs were BLASTn, which compares nucleotide sequences to a nucleotide database, BLASTp, which compares protein sequences to a protein database and tBLASTn which compares a translated nucleotide sequence to a database of nucleotides translated in all six open reading frames. Details of the individual use of gene databases can be found in following chapters.

### 2.3.2 DNA manipulation *in silico*

DNAMAN (Lynnon BioSoft) was used to identify and translate nucleotide open reading frames, to position DNA start/stop codons and to identify the reverse complement of a sequence of interest. DNA and AA sequences were aligned using Bioedit (Hall, 1999), clustalX1.81 (Thompson et al., 1997), or T-coffee (Notredame et al., 2000). Further details on the use of alignments and of the optimisation of alignment parameters can be found in individual chapters.

### 2.3.3 Primer design

Primers for polymerase chain replication (PCR) experiments were designed according to a stringent set of rules. Generally, primers ranged from 18-30 nucleotides in length, had a guanine/cytosine content greater than 50% and a melting temperature ( $T_m$ ) from 50-70 °C. To increase priming efficiency, the 3' of each primer was designed where possible to end in a string of one to three G/C's. The programs NetPrimer (<http://www.premierbiosoft.com/netprimer/>) and primer 3 (<http://frodo.wi.mit.edu/>) were used to predict the likelihood of self/cross dimerisation of primers at a range of annealing temperatures. Primers were predicted to have a low dimerisation potential particularly at the 3' of sequences where primer to target DNA binding occurs.

For quantitative real-time PCR (qPCR, see section 2.5) primers, additional stringency in design was enforced. Primers were designed to have a  $T_m$  greater than 60 °C and were predicted to produce no homo/heterodimers in NetPrimer. In all cases at least one primer pair spanned an exon/intron junction to reduce the amplification of contaminating genomic DNA. When genomic information was not published for genes of interest, primers were designed to experimentally amplify introns (section 2.3.4). In cases when closely related genes were amplified (e.g. chapter

4), sequences were aligned and primers were designed to be as divergent as possible to minimize cross amplification, particularly at the primers 3' end.

#### *2.3.4 Predication of gene structure*

To characterise intron-exon boundaries, two approaches were used. If a corresponding cDNA and genomic DNA sequence were available, then the intronic splice sites were characterised by loading these sequences into Spidey (Wheelan et al., 2001) at <http://www.ncbi.nlm.nih.gov/spidey/> with default settings. If only a cDNA sequence was known, then that sequence would be aligned with orthologous cDNAs from related species, where the gene structure was known, to compare sequence conservation at the established intron-exon boundaries. If intron-exon boundaries were conserved across multiple species, and the corresponding cDNA of interest aligned well at mRNA regions flanking splice sites, then it was considered that the intron-exon boundary was likely conserved, and would accordingly design primers to experimentally validate the position of putative introns.

#### *2.3.5 Sequence submission*

Annotated cDNA and genomic DNA sequences were deposited in the NCBI database using the BANKIT program (<http://www.ncbi.nlm.nih.gov/BankIt/>).

### ***2.4 Basic experimental molecular biology***

#### *2.4.1 General working conditions*

Surfaces and pipettes were cleaned with 70% (v/v) ethanol and with RNase Zap<sup>®</sup> (Ambion) to reduce contaminating sources of bacteria and nucleases. Milli-Q water (Millipore) was used in all molecular biology procedures and was autoclaved with 0.1% diethylpyrocarbonate (DEPC)

(Sigma) for RNA work and without DEPC treatment for other applications. Further, all plastic wear was either autoclaved, or certified sterile on purchase and glasswear used for RNA work was autoclaved.

#### *2.4.2 Extraction of DNA and RNA*

Genomic DNA was extracted from 100 mg of adult tissue using the Dneasy kit and manufacturers protocol (Qiagen). For total RNA extraction, 100 mg tissue was added to FastRNA Pro Green Beads (MP Biomedicals) with 1 ml of Tri Reagent (Sigma). A FastPrep machine (MP Biomedicals) was used to homogenise tissues for 40s at a speed setting of 6.0. Samples were left at room temperature for 5 min to allow the melting of nucleic acid-protein complexes. 0.2 ml of chloroform was then added to samples that were vortexed vigorously for 15s, left to stand at room temperature for 5 mins and then centrifuged at 13,000 g for 15 min at 4 °C. The aqueous phase was then transferred to a new microcentrifuge tube and RNA was precipitated in 0.5 ml 2-propanol (Sigma). The resulting precipitate was pelleted by centrifugation at 13,000g at 4 °C and washed in 75% (v/v) ethanol/DEPC water. The pellet was air-dried to avoid ethanol being carried forward to later reactions and then dissolved in 75 µl DEPC water at 55 °C for 5 min and subsequently stored at -70 °C. The TURBO DNA-free™ (Ambion) kit was then used to eliminate genomic DNA contamination from total RNA preparations following the manufacturers instructions.

#### *2.4.3 Gel Electrophoresis*

Nucleic acids were size fractioned using agarose gel electrophoresis. Agarose gels were made by dissolving agarose powder (Bioline) to a desired molecular volume in Tris-Acetate-EDTA buffer (TAE) (40mM Tris base (Sigma), 20mM acetic acid (Sigma), 1mM EDTA (Sigma)). Agarose was melted into the buffer using a microwave and then ethidium bromide (EtBr) (Sigma) was

added to a final concentration of 0.5µg/ml when the molten agarose had cooled to ~50 °C. Gels were placed into an electrophoretic apparatus, allowed to solidify and then submersed in TAE. Experimental nucleic acids and a quantitative DNA/RNA marker (New England Biolabs) were then mixed with 1X loading buffer (Promega) and loaded into gel wells. An electrical current was then passed through the gel at 50-100v until nucleic acids were fully separated. Nucleic acids bound to EtBr were then visualised under ultraviolet light using a VersaDoc<sup>TM</sup> 3000 imaging system (Bio-Rad).

#### *2.4.4 Quantification and quality assessment of RNA*

During this study, total RNA concentration was quantified by one of two approaches. The first was using the RiboGreen RNA quantitation reagent kit (Invitrogen) following a modified version of the manufacturers 'high range' protocol and using a FluoStar fluorimeter (BMG Lab Technologies). Each reaction was performed in triplicate and contained 100 µl of total RNA (of unknown concentration, diluted to 0.01% v/v in TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and 100 µl of RiboGreen working solution. A standard curve was produced using 5 dilutions of a quantified RNA standard (final concentrations of 1µg/ml, 500, 100, 20 and 0 ng/ml RNA standard diluted in TE). Additionally, a NanoDrop (ND-1000) spectrophotometer (Nanodrop Technologies) became available in the laboratory in the later stages of this project, and was used to directly quantify 1 µl of RNA samples of interest.

The quality of RNA was qualitatively assessed by agarose gel electrophoresis (section 2.4.3). 1-2 µl of RNA was ran on a 1% RNAase free agarose gel with an RNA size marker (New England Biolabs) and the clear integrity of 18 and 28S ribosomal RNA bands was used to indicate that the total RNA was not degraded and thus that the mRNA portion was intact.

#### *2.4.5 Gel purification of DNA*

PCR products were excised from agarose gels under UV light using a sterile disposable scalpel taking care to minimise the carry-over of agarose external to the band of interest. PCR products were then purified from agarose gels using the QIAquick gel extraction kit (Qiagen) following the manufacturers protocol.

#### *2.4.6 Complementary DNA (cDNA) synthesis*

First-strand cDNA was created using the RETROscript™ kit (Ambion) following the manufacturers protocol. For first strand synthesis, a combination of oligo-dT primers and random decamers were used in a respective 2:1 ratio and reactions were performed in a thermocycler (Bio-Rad). Initially the RNA was denatured at 85 °C and then double stranded cDNA was synthesized at 55 °C for 60 min in a mix containing 100U of reverse transcriptase (MMLV-RT), 2 mM dNTPs, 10U of RNase inhibitor and 2 µl 10X reverse transcription buffer (100mM Tris-HCl (pH 8.3), 750mM KCl, 30mM MgCl<sub>2</sub>, 50mM dithiothreitol). The reaction was halted by denaturing the MMLV-RT at 92 °C for 10 min. cDNA was stored at -20 °C and diluted 5-10 times in sterile milli-Q water.

#### *2.4.7 Polymerase chain replication (PCR) and Reverse Transcription-PCR (RT-PCR)*

Nucleic acid sequences of interest were amplified by PCR or RT-PCR in 25µl reactions containing 1 µl cDNA (RT-PCR) or 1µl DNA (PCR), 1.25U of BioTaq™ DNA polymerase (Bioline), with 2.5µl of the supplied 10X NH<sub>4</sub> buffer (160 mM NH<sub>4</sub>2SO<sub>4</sub>, 670 mM Tris-HCL (pH 8.8) 0.1% Tween-20), 1.25µl 50 mM MgCl<sub>2</sub>, 0.1mM dNTPs, and generally, 4 µM of forward and reverse primer. Reactions were carried out in microcentrifuge tubes in a thermocycler (Bio-Rad) with a range of cycling conditions, specific to each primer set. Commonly, cycling parameters included a 10 minute initial denaturation of 95 °C followed by

(20-38) cycles of 95 °C for 30s, 48-65 °C for 30s (depending on application and primer properties), and 72 °C for 15s-3 min (depending on product size; around 1000 bp/min are added by DNA polymerases). Additionally, a 10-minute final extension at 72°C was used to ensure all products had the 3' adenosine residue overhangs necessary for the subsequent cloning procedure (section 2.4.9).

#### *2.4.8 Rapid Amplification of Complementary Ends (RACE) PCR*

To amplify the 3' UTR of genes of interest, 3' RACE PCR was used with the BD SMART™ RACE cDNA Amplification Kit (BD Biosciences). First strand cDNA was synthesised using the supplied 3'CDS primer (5'–3': AAGCAGTGGTATCAACGCAGAGTAC(T)<sub>30</sub> 30N-1N, where N = A,C,G or T and N-1 = A, C or G) in a reaction containing 1 µg of total RNA, 2µl 5X First Strand Buffer (250 mM Tris-HCl (pH 8.3), 375mM KCl, 30mM MgCl<sub>2</sub>), 1 µl dithiothreitol (20mM), 1.6mM dNTPs and 1µl reverse transcriptase. Reverse transcription was allowed to proceed at 42 °C for 1.5 hours and cDNA was stored at –20 °C. 3' RACE reactions contained the following: 2.5 µl 3' RACE cDNA, 5µl 10X universal primer mix, and 8µM of gene-specific primer. Touchdown PCR was used with 5 cycles of 94 °C for 30 s, 72 °C for 30s, followed by 5 cycles of 94 °C for 30 s, 70 °C for 30s, 72°C for 3 min and then 30 cycles of 94 °C for 30 s, 68°C for 30 s and 72 °C for 3 min.

#### *2.4.9 DNA cloning*

PCR products were cloned using the TOPO T/A Cloning® kit (Invitrogen), which takes advantage of the tendency of Taq DNA polymerase to add an adenine residue at the 3' of PCR products following extension (Clark, 1988). The pCR®4- TOPO vector has a 3' thymine overhang, which thus complements the overhang of the PCR product. The TOPO reaction contained 4 µl of PCR product, 1 µl of salt solution (1.2M NaCl; 0.06M MgCl<sub>2</sub>) and 1 µl of

vector and was incubated for 10 min at room temperature. The vector was then mixed with chemically competent One Shot<sup>®</sup>TOP10 *Escherichia coli* cells (Invitrogen), which were transformed by heat-shock at 42 °C for 30s. Subsequently 0.25 ml of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM sodium chloride, 2.5 mM KCl, 10 mM MgCl, 10 mM magnesium sulfate, 20 mM glucose) (Invitrogen) was added to the cells. The mixture was agitated on a horizontal shaking platform at 200 rpm for 60 min before 50-200 µl was spread on agar (Sigma) containing 100 mg/ml ampicillin (Sigma) and left overnight at 37 °C. Clearly separated colonies were handpicked in a flame-sterilised environment with a sterile toothpick/pipette tip and placed in 1-2ml of LB broth (Sigma) containing 100 mg/ml ampicillin and left overnight at 37 °C on a horizontal rocking platform at 200 rpm.

#### *2.4.10 Plasmid purification, digestion and screening*

Plasmid DNA was purified from *E. coli* cells using the QiaPrep spin Miniprep kit (Qiagen) following the manufacturer protocol. Plasmid DNA was enzyme digested in a 20 µl mix containing 6U of ECOR1 (Promega), 0.1 µg Bovine Serum Albumen (BSA) (Promega) and 2.0 µl of 10X buffer H (Promega) made up to volume in sterile milli-Q water. The digestion was allowed to proceed for 3 hours at 37 °C. Products were run on a 1.2% agarose gel with a quantitative 1kb DNA marker (New England Biolabs) and visualised under UV light, to quantify plasmid DNA mass and to check for the presence of the expected insert. Alternatively plasmids were screened for inserts using PCR, with the T3/T7 or M13F/R primers provided with the TOPO T/A Cloning<sup>®</sup> kit. Clones containing the desired insert were catalogued, sent for sequencing (section 2.3.11), and for long time preservation were stored at -20°C.

#### *2.4.11 Sequencing of DNA*

The Sequencing Service at the University of Dundee (UK) was used to sequence plasmid DNA. Two clones of each plasmid were sequenced in sense/antisense directions using T3 and T7 primers with an ABI PRISM 377 DNA sequencer (Applied Biosystems). Sequencing results were retrieved as a sequence file that was processed in DNAMAN and as an electropherogram that was assessed using the program ChromasLite (freeware: [http://www.technelysium.com.au/chromas\\_lite.html](http://www.technelysium.com.au/chromas_lite.html)).

### **2.5 Quantitative PCR (qPCR)**

#### *2.5.1 Experimental procedure*

All qPCR plates were setup in a sterile laminar flow-hood (Microflow Class 2 Advanced Biosafety cabinet, BIOQUELL) to minimise nucleic acid contamination from aerial sources. Pipettes were routinely cleaned internally using 1% (v/v) bleach in sterile water to remove contaminating nucleic acids and filter tips (Eppendorf) were used for all reactions. RNA extraction, quantification and first-strand cDNA synthesis was performed as described above. Care was taken to ensure that the initial mass of tissue was normalised across samples and a normalised input of total RNA (by mass) was used prior to reverse transcription. Reactions were performed using an ABI Prism 7000 (Applied Biosystems) real-time thermocycler using SYBR Green (QuantiTect SYBR Green PCR kit, Qiagen). Each reaction contained 1X QuantiTect SYBR Green PCR master mix, 1µl of cDNA and 0.2 µM of each primer, made up to 20µl with sterile water. In each round of cDNA synthesis negative controls lacking reverse transcriptase (-RT) were produced from pooled samples of all RNAs being examined. In each qPCR run, negative controls were ran to assess the presence of contamination within primers (no template

control, 1 µl of water instead of cDNA) and genomic DNA carried over from the cDNA synthesis (1µl of -RT sample).

### *2.5.2 Dissociation analysis and primer specificity*

To ensure the specificity of primers a dissociation analysis from 60-90 °C was included at the end of each qPCR run to obtain an amplicon melting curve (Ririe et al., 1997). A single sharp band indicated that a single product was amplified without primer dimer formation. If undesired products, or secondary structures (i.e. primer dimers) were amplified, the qPCR assay was optimised by altering the annealing temperature until a single peak was obtained.

## ***2.6 In situ hybridisation***

### *2.6.1 Introduction*

The *in situ* hybridisation protocol used throughout this project was written by the author and is included for reference to future workers (Appendix 1). It is specific to salmon embryos and can be used for single or dual stain labelling. This protocol was based on previous methodologies that are acknowledged therein. Below is a detailed description of the methodologies.

### *2.6.2 Amplification of probe templates*

DNA used for probe transcription was amplified by PCR using products held within the pCR<sup>®</sup>4-TOPO plasmid which contains the specific binding sites required by T3/T7 RNA polymerases. Each PCR reaction contained 200-300 ng of plasmid template, 2.5µl of 10X Taq polymerase buffer (Bioline), 0.4µM of M13 forward and reverse primers (Invitrogen) and 1U of BioTaq<sup>™</sup> polymerase (Bioline) made up to volume using DEPC treated milli-Q water. A thermocycler was used (Bio-Rad) with the following conditions: 1 cycle of 95 °C for 5 min; followed by 35 cycles

of 30s at 95 °C, 30s at 56 °C and 1 min at 72 °C. Products were separated on 1.1 % agarose gels and PCR products were isolated as described in section 2.4.5.

### *2.6.3 Probe transcription*

100-200ng of PCR product was used to synthesise anti-sense and sense probes in a 20.0 µl mix containing 2.0 µl digoxigenin-UTP (DIG) 10X labelling mix (Roche), or Fluorescein 10X labelling mix (Roche), 2.0µl 10X transcription buffer (400nM Tris-HCl, pH 8.0, 60nM MgCl<sub>2</sub>, 10mM dithiothreitol, 20mM spermidine) (Roche), 0.5µl Rnasin<sup>®</sup> (Promega) and 2.0 µl of T3 or T7 RNA polymerase (Roche) made up to volume with DEPC treated milli-Q water. The mix was incubated at 37°C for 2 hours. 2 µl of TURBO Dnase (Ambion) was then added to the mix which was reincubated at 37 °C for 15 min. Transcription was halted with 1µl of 0.5M EDTA (pH 8.0). RNA probes were purified with 1.25 µl of 8M LiCl (Sigma) and then 75µl of 100% (v/v) ethanol was used to precipitate the RNA overnight at -80°C. Precipitated RNA was pelleted by centrifugation at 13,000 rpm and then resuspended in 100µl DEPC water. The integrity and size of RNA transcripts was analysed using gel electrophoresis on a 1.1% (m/v) RNase free agarose gel with a 1kb RNA marker (New England BioLabs).

### *2.6.4 Embryo preparation and fixation*

Embryos were rehydrated by successive washes in decreasingly concentrated volumes of methanol in PBS/0.1% Tween 20 (Sigma) (PBT) (5 min with 75%, 50% and 25% methanol (v/v) /PBT, 2 x 5 min in PBT). Embryos were carefully dechorionated using watchmakers forceps under a Leica MZ7.5 binocular microscope (Leica Microsystems) and were then permeabilised by digestion in 1:1000 proteinase K (Roche) (v/v)/PBT for either 5 min at room temperature (pre-somite embryos), 15 min at room temperature (1-60 ss embryos) or 20 min at 37°C

(embryos from the end of segmentation). Embryos were then washed in PBT for 2 x 5 min, refixed in 4% (m/v) paraformaldehyde/PBT fixative/ 0.2% glutaraldehyde (Sigma) for 20 min and washed for 3 x 5 min in PBT to remove remaining fixative.

#### *2.6.5 Probe hybridisation and detection*

Embryos were initially incubated in hybridisation mix (50% formamide (Fisher Scientific), 5X Sodium chloride, sodium citrate (SSC), 2% blocking reagent (Roche), 0.1% Triton X-100 (Sigma), 50µg/ml tRNA (Roche), 5mM EDTA and 50µg/ml heparin (Roche)) for 3 hours at 70 °C. Embryos were then incubated for 24-72 hours at 70 °C in hybridisation mix containing 1µg/ml DIG and/or Flu labelled probe. Following hybridisation, the embryos were washed with decreasing stringency to remove unbound probe. At 70 °C the following washes were performed: 2 x 10 min in 2X SSC, 3 x 15 min in 2X SSC, 0.1% (m/v) CHAPS (Sigma) and 3 x 20 min in 0.2X 1M SSC, 0.1% (m/v) CHAPS. Embryos were then blocked in 4 % (v/v) BSA (50mg/ml) (Sigma), 5% sheep serum (Sigma), 1 % (v/v) dimethylsulphoxide (Sigma), 90% (v/v) PBT firstly for 2 hours at room temperature and then again overnight at 4°C with respective 1:5000 or 1:2000 (v/v) dilutions of phosphate-conjugated anti-DIG or anti-Flu monoclonal antibodies (Roche). In single labeling experiments, probes were incubated with either Fast Red for Flu (Invitrogen) made up to the manufacturers instructions or 1 mg/ml<sup>-1</sup> NBT/BCIP (Roche) for DIG. Incubations were allowed to proceed until the colour reaction (red for Flu, blue/purple for DIG) had fully developed. When dual-stain experiments were used, probes were detected consecutively as described in the Appendix 1 protocol.

#### *2.6.6 Embryo study, cryosectioning and photography*

Whole-mount embryos were studied in dark/brightfield using a Leica MZ7.5 binocular microscope (Leica Microsystems). Additionally, embryos were flatmounted in PBS on a glass

microscope slide using a cover slip that was separated from the slide by droplets of silicon jelly. Flatmounted embryos were studied with differential interference contrast optics using a Leica DMRB microscope (Leica Microsystems). In preparation for cryosectioning, embryos were appropriately orientated in cryomatrix and then flash frozen in isopentane (Fisher Scientific) cooled to near its freezing point (-159 °C) over liquid nitrogen. Embryo cryosections were then cut at 18 µm on a cryostat (Leica Microsystems, CM1850). All photographs were recorded with a Nikon Coolpix 4500 digital camera (Nikon).

## **Chapter 3. The evolutionary relationships of teleost *myod* genes revealed through comparative phylogenetic and genomic analyses**

### **3.1 Abstract**

In several teleost species including salmonids and Acanthopterygians, but not zebrafish, at least two MyoD paralogues are conserved that are thought to have arisen from distinct, possibly lineage-specific duplication events. Additionally, two MyoD paralogues have been characterised in the allotetraploid frog, *X. laevis*. This has led to a confusing nomenclature since MyoD paralogues have been named outside of an appropriate phylogenetic framework. The initial aim of this chapter was to use phylogenetic reconstruction to establish the orthology and paralogy of teleost MyoD sequences within a framework of vertebrate MRFs. A maximum likelihood (ML) analysis showed that one gene, *myod1* is conserved in all teleosts but that species of the Acanthopterygii superorder have a second gene (*myod2*). Further, three salmonid *myod* genes are each recent orthologues of teleost *myod1* and arose distinctly from the duplication event from which *myod2* is conserved. However, the position of MyoD2 on the ML tree was external to all vertebrate MyoD sequences, which does not support either an Acanthopterygian or teleost specific event. Directly depicting the phylogenetic relationships of teleost MyoD sequences is hindered by the asymmetric evolutionary rate of Acanthopterygian MyoD paralogues. Thus, the next aim of this chapter was to confidently position the event from which Acanthopterygii MyoD2 arose using a comparative analysis of the chromosomal regions containing *myod* across the vertebrates. To this end it is shown that genes on the single *myod*-containing chromosome of human and chicken genomes are retained in both zebrafish and Acanthopterygian teleosts in a striking pattern of interleaved double conserved synteny. Further, phylogenetic reconstruction of these neighbouring genes using Bayesian and ML methods supported a common origin for

teleost paralogues following the split of the Actinopterygii and Sarcopterygii. These results strongly suggest that *myod* was duplicated during the basal teleost WGD event, but was subsequently lost in the Ostariophysi (zebrafish) and Protacanthopterygii lineages. Finally, a sensible consensus nomenclature is suggested for vertebrate *myod* genes that is justified from a phylogenetic perspective and should be implemented in future studies of teleost MyoD.

### 3.2 Introduction

In most diploid tetrapods, including birds, mammals, the frog *Xenopus tropicalis* as well as teleosts of the Ostariophysi superorder, a single gene represents MyoD. The allotetraploid frog, *X. laevis* has two differentially expressed MyoD paralogues that were originally named Xlmf1 and Xlmf25 (Scales et al., 1990). Teleost species of the Acanthopterygii also have two differentially expressed paralogues originally denoted MyoD1 and MyoD2 (Tan and Du, 2002). Additionally, two salmonid MyoD duplicates were characterised in rainbow trout (*Oncorhynchus mykiss*) and named MyoD and MyoD2 (Rescan and Gauvry, 1996). In chapter 4, a third salmonid MyoD sequence was characterised (also see Macqueen and Johnston, 2006). While the MyoD paralogues in salmonids and frogs can be explained by lineage specific WGD events, the scale of the event from which Acanthopterygian MyoD2 arose is currently unknown. Additionally the nomenclature for MyoD genes in the vertebrates is confusing as it is based outside of a phylogenetic framework.

In this chapter, the evolution of MyoD is investigated in teleost fish using direct phylogenetic reconstruction as well as through comparative analyses of the phylogenetic relationships and conserved synteny of genes in neighbourhood to *myod* across the vertebrates. To this end, the extent of *myod* duplications arising in different vertebrate lineages was confidently established,

including the timing of the event from which Acanthopterygian *myod2* arose. Additionally, I propose the use of a sensible nomenclature consensus for vertebrate *myod* genes, which accommodates the frequency of polyploidy observed in teleosts, and other non-diploid vertebrates.

### 3.3 Materials and methods

#### 3.3.1 *In silico* mining and cloning of *myod2*

*S. aurata* MyoD1/2 sequences were BLAST screened against zebrafish (*D. rerio*), tiger pufferfish (*T. rubripes*), green spotted pufferfish (*T. nigroviridis*), stickleback (*G. aculeatus*) and medaka (*O. latipes*) genomes at the Ensembl database ([www.ensembl.org](http://www.ensembl.org)). This approach was also used to screen ESTs in the salmon genome project ([www.salmongenome.no/](http://www.salmongenome.no/)), the Gene Indices project, (<http://compbio.dfci.harvard.edu/tgi/>) and the cGRASP database (<http://web.uvic.ca/cbr/grasp/>). The same approach was also used to screen catfish (*Ictalurus punctatus*) and fathead minnow (*P. promelas*) ESTs at the Gene Index Project (<http://compbio.dfci.harvard.edu/tgi/>).

Primers to amplify pufferfish *myod2* were designed from fragmented sequences in the *T. rubripes* and *T. nigroviridis* genomes (Ensembl gene ID's respectively: SINFRUG00000163904 and GSTENG00034775001) and their sequences were as follows: F-5'-3': ATGGATCTGTCCGAGCTGGTCTTC, R-5'-3': TCAGAGCGGCTCGTAGATCCCTG). These were used in a standard RT-PCR reaction using *T. rubripes* cDNA synthesised from total RNA that was extracted from fast-twitch myotomal muscle (kindly provided by Dr Matthew MacKenzie). The subsequent PCR products produced a single band by agarose gel electrophoresis, which was extracted, cloned and sequenced as previously described (see chapter 2, section 2.3).

### 3.3.2 Testing the selective constraints across MyoD1/2 proteins

Four full coding sequence Acanthopterygian sequences were retrieved for both MyoD1 and MyoD2. For MyoD1 this included sequences from *T. rubripes*, *G. aculeatus*, *Sparus aurata* and *Paralichthys olivaceus*. For MyoD2 this included sequences from *T. rubripes*, *G. aculeatus*, *S. aurata* and *H. hippoglossus*. MyoD1 and MyoD2 sequences were separately aligned at the AA level using ClustalX with default settings and were then loaded into PAL2NAL (<http://coot.embl.de/pal2nal/>) (Suyama et al., 2006), along with the corresponding nucleotide sequences. PAL2NAL then converted this data into a multiple codon alignment which was loaded into SNAP (Korber, 2000) at (<http://www.hiv.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html>) which estimated the average number of non-synonymous (dN) and synonymous (dS) substitutions within each codon alignment. SNAP was also used to produce a plot of the cumulative average non-synonymous and synonymous substitutions across MyoD1 and MyoD2.

### 3.3.3 Sequence retrieval for phylogenetic reconstruction

The following AA translations of MRFs were retrieved from Genbank: **MyoD**: mouse (*Mus musculus*) (NM\_010866), rat (*Rattus norvegicus*) (M84176), human (*Homo sapiens*) (X75798), chimpanzee (*Pan troglodytes*) (XP-508311), pig (*Sus scrofa*) (X56677), dog (*Canis familiaris*) (XP-854756), cattle (*Bos taurus*) (AB110599), sheep (*Ovis aries*) (XG2102), chicken (*Gallus gallus*) (L34006), Japanese quail (*Coturnix japonica*) (L16886), turkey (*Meleagris gallopavo*) (AY641567), Western clawed frog (*X. tropicalis*) (AJ579310), African clawed frog (*X. laevis*) mf25 (BC073672), mf11 (m311117), mf1 (M31116), zebrafish (Af318503), knifefish (*Sternopygus macrurus*) (AY396566), common carp (*C. carpio*) (AB012882), rainbow trout (*O. mykiss*) MyoD1a (X75798), MyoD1b (Z46924), Atlantic salmon (*S. salar*), MyoD1a (AJ618978), MyoD1b (AJ557150), MyoD1c (DQ317527), brown trout (*S. trutta*) MyoD1c

(DQ366710), flounder (*P. olivaceus*) (DQ184914), Atlantic halibut (*H. hippoglossus*) MyoD2 (AJ630127), channel catfish (*I. punctatus*) (AY534328), blue catfish (*I. furcatus*) (AY562555), white catfish (*Ameiurus catus*) (AY562556), gilthead seabream, MyoD1 (AF478568), MyoD2 (AF478569), blue tilapia (*Oreochromis aureus*) (AF270790), tiger pufferfish MyoD1 (AB235116) MyoD2 (NM\_001040062), *Ciona intestinalis* MyoD family protein J (AAB61360), amphi-MyoD1 (AY313170), amphi-MyoD2 (AB092416) **Myf5**: chicken (NM\_001030363), mouse (NP\_008656), human (NP\_005593), tiger pufferfish (NM\_001032770), zebrafish (AF253470), knifefish (DQ016032), rainbow trout (AY751283), Western clawed frog (AJ579311, knifefish (DQ0160320). **Myog**: mouse (M95800), human (NM\_002479), chicken (D90157), African clawed frog (NM\_001016725), knifefish (AY396565), Atlantic salmon (DQ294029), zebrafish (NM\_131006). **Mrf4**: mouse (NM\_008657), human (NM\_002469), chicken (D10599), western clawed frog (S84990), tiger pufferfish (AY445320), knifefish (DQ059552), zebrafish (NM\_001003982), Atlantic salmon (DQ479952).

### 3.3.4 Phylogenetic reconstruction of myod genes

The evolution of teleost MyoD was initially assessed within a phylogenetic analysis of the relationships of fully coding AA translations of vertebrate *myod*, *myf5*, *myog* and *Mrf4* genes (accession numbers listed above). These sequences were aligned in ClustalX (Thompson et al., 1997) and multiple alignment gap penalties were optimised using TuneClustalX (freeware <http://homepage.mac.com/barryghall/TuneClustalX.html>). ML was implemented in the web-based interface of PhyML (<http://atgc.lirmm.fr/phyml/>) (Guindon and Gascuel, 2003; Guindon et al., 2005) using the WAG model of AA substitution with concurrent estimation of the gamma distribution of among-site substitution rates. For comparison with a previous phylogenetic analysis of *myod* (Atchely et al., 1994), a neighbour joining (NJ) analysis was concurrently

performed on the same alignment using Mega 3.1 (Kumar et al., 2004), with the p-distance model and a uniform distribution of among site rates. Trees were reconstructed using mega 3.1.

The methods described above were used to produce the cladogram shown in Fig 3.3, which was published in Macqueen and Johnston (2006). However, an apparent anomaly was present on the tree (the position of the MyoD2 clade external to all vertebrates) that was not initially appreciated by myself or picked up during the publication review process. This was thought to be an artefact of long branch attraction (LBA) or mutational saturation (see result section: 3.4.3). Accordingly, a new alignment of MyoD was produced from 17 vertebrate species (accession numbers/genbank IDs can be found in Table 1; rationale for choice of sequences explained in results section 3.4.4). These sequences were aligned with T-coffee (Notredame et al., 2000) using a combination of Lalign and ClustalW alignments. Phylogenetic reconstruction was then performed using Bayesian, ML and NJ approaches. Bayesian analysis was performed in Mr Bayes 3.12 (Ronquist and Huelsenbeck, 2003) with a mixed AA model, sampling every 100 generations and assuming a gamma distribution of substitution rates. 500,000 generations were implemented with a burnin value corresponding to the first 150,000 generations. The runs were considered to have converged when the standard deviation of split frequencies was constantly less than 0.01 (this occurred after 150,000 generations) and trees from the burnin phase were discarded. A majority rule consensus tree was then built based on the final 3500 trees. A similar approach was also used without including a gamma distribution as a parameter. PhyML (Guindon and Gascuel, 2003) was used to perform ML with concurrent estimation of the gamma distribution of among-site substitution rates, the WAG model (which gave the best posterior probability values in MrBayes), and with 500 pseudobootstrap replicates for branch confidence. NJ was implemented in Mega 3.1 using a gamma distribution of among site substitution rates (0.66, as estimated by PhyML), the Poisson correction model and 5000 bootstrap replicates. The

same approach was also used to produce a NJ tree considering among-site substitution rates to be uniform. A NJ tree was then constructed considering solely the unsaturated fraction of substitution sites using ASATURA (Van de Peer et al., 2002). The WAG model was used and a cut off value of 2.584 was considered to remove saturated sites. Branch support was then obtained from 5000 bootstrap replicates. All trees were reconstructed in Mega 3.1.

### 3.3.5 Synteny analysis of teleost *myod* genes

Genes in neighbourhood to human *myod* were manually obtained from the Ensembl database ([www.ensembl.org](http://www.ensembl.org)) using the MultiContig View, Gene view and by using the orthologue/parologue feature, while recording strand orientation and chromosomal position relative to *myod*. Orthologues of these genes were then obtained by the same approach for chicken, zebrafish, pufferfish (*T. rubripes*), stickleback and medaka and a synteny diagram was constructed.

### 3.3.6 Phylogenetic reconstruction of *myod*-neighbouring genes

Phylogenetic analysis was used to reconstruct the relationships of genes in upstream/downstream proximity to *myod* in human relative to other species used in the synteny analysis, and also using sequences obtained from Ensembl genome databases of mouse and the diploid frog *X. tropicalis*. The criteria for gene selection was that two teleost copies were retained on two paralogous chromosomal regions, each retaining synteny to the single *myod*-containing chromosome of human/chicken genomes. Within the synteny analysis, this included genes coding for TropI, TropT, Kcnc1, Tph1, Nucb2 and Plekha7. High quality AA translations of these genes were manually obtained using the MultiContig/Geneview features at the Ensembl database. Sequences were aligned with T-coffee using a combination of Lalign and ClustalW alignments. Phylogenetic reconstruction was performed using Bayesian and ML approaches. Bayesian

analysis was performed in Mr Bayes 3.12 and the number of generations and ‘burnin’ values for different sequences analysed were: TropI: 300,000 generations, burnin of 100,000 generations, TropT: 300,000 generations, burnin of 100,000 generations, Kcnc1: 100,000 generations, burnin of 25,000 generations, Tph1: 200,000 generations, burnin of 60,000 generations, Nucb2: 100,000 generations, burnin of 25,000 generations, Plekah7: 100,000 generations, burnin of 25,000 generations. Runs were considered to have converged when the average standard deviation of split frequencies between chains remained less than 0.01. Trees from the burnin phase were discarded and majority rule consensus trees with posterior probability values were calculated from trees obtained after runs had converged. ML was performed using PhyML with the AA substitution model that gave the best posterior probability values in MrBayes (TropI: WAG, TropT: JTT, Kcnc1: JTT, Tph1: JTT, Nucb2: JTT, Plekha7: JTT), and assuming a gamma distribution of among-site substitution rates. 500 pseudobootstrap replicates were used to assess branch confidence. For TropT and Tph1, the tree topology returned by the Bayes/ML approach was inconsistent with the synteny/neighbouring genes analysis and trees retained for other *myod*-neighbouring genes. For these sequences I tested the hypothesis that mutational saturation may have affected the alignment. This was achieved in ASATURA, which was used to construct NJ trees with and without prior removal of frequently mutating residues from the alignment. The AA substitution with the highest MrBayes posterior probability values was used and branch confidence was assessed with 1000 bootstrap replicates. For the Tph1 alignment, the JTT matrix was employed and cut off values of 850 and 2348 were respectively used prior to tree reconstruction to consider all residues in the alignment and only the unsaturated fraction of sites. For the TropT alignment, the JTT matrix was used and cut off values of 610 and 2258 were respectively used prior to tree reconstruction to consider all residues in the alignment and only the unsaturated fraction of sites.

## 3.4 Results

### 3.4.1 *In silico* and experimental characterisation of teleost *myod2* genes

Using BLAST and manual searches of all available teleost Ensembl genomes, four species from the superorder Acanthopterygii were found to have two *myod* genes on two chromosomes (*T. rubripes*, *T. nigroviridis*, *G. aculeatus* and *O. latipes*). Conversely, the *D. rerio* genome contained only the single *myod* gene previously characterised (Weinberg et al., 1996). Additionally, in EST libraries for catfishes and the fathead minnow, only sequences orthologous to zebrafish *myod* were retrieved. Further, in salmonid EST libraries only *myod* genes that had previously been characterised were retrieved. To gain an additional experimentally validated sequence for phylogenetic analysis, primers were designed from fragmented *T. rubripes* and *T. nigroviridis* sequence predictions and used in an RT-PCR reaction with *T. nigroviridis* fast muscle cDNA. A single band was separated by gel electrophoresis and cloned. Subsequent sequencing revealed a 792 bp sequence encoding an open reading frame of 263 AAs. A tBLASTn search of this putative AA sequence against the NCBI public database revealed that it shared highest sequence identity with *S. aurata* MyoD2 (73% identity, E-value 5e-87), which was elevated relative to pufferfish MyoD1 (54% identity, E-value 3e-66). Thus the sequence was submitted to GenBank as MyoD2. Figure 3.1 shows an alignment of Acanthopterygian MyoD1 and MyoD2 sequences. It can be seen that the bHLH is strongly conserved across all MyoD1 and 2 proteins whereas the N-terminal and all regions C-terminal to the HLH are less well conserved. Interestingly, a serine rich region of around 20 AAs is conserved at the N-terminal of MyoD1 in pufferfish and gilthead seabream, but not MyoD1 of stickleback, and not MyoD2 of any species (Fig. 3.1). Additionally, the helix-3 domain is strongly conserved between all MyoD proteins (Fig. 3.1).

#### 3.4.2 Evolutionary constraints on Acanthopterygian MyoD1 and MyoD2 proteins

Next, the selective constraints on MyoD1/MyoD2 paralogues in the Acanthopterygians were investigated. Fig 3.2 shows the average cumulative number of synonymous and non-synonymous substitutions for MyoD1 and MyoD2 calculated from codon alignments obtained for four Acanthopterygian species (see methods, section 3.3.2). For MyoD1 proteins, the rate of change of non-synonymous substitutions (dN) at the N and C termini is comparable to synonymous changes (dS), except for the last 8 residues, which are devoid of AA changing substitutions (see Fig. 3.1). From residues ~60-240, incorporating the bHLH and amphipathic helix-3, dS generally exceeds dN in MyoD1 (steeper red line than green line on Fig. 3.2). For MyoD2, dN and dS is comparable at the N-terminal (first ~100 residues). From residues ~100-160, which incorporates the bHLH (Fig. 3.1, region marked bHLH on Fig. 3.2), the rate of increase in synonymous substitutions in MyoD2 strongly outweighs dN. However, from this point until the helix-3 (residues ~180-220 on Fig. 3.2) a steep rise in non-synonymous substitutions is observed. Similarly to MyoD1, the MyoD2 helix-3 is all but devoid of non-synonymous substitutions (flat green line marked H3 on Fig. 3.2: MyoD2) but this is followed by a steep rise in non-synonymous substitutions at the extreme C-terminus (Fig. 3.2, marked C). Additionally the highly conserved extreme 3' residues of MyoD1 proteins are absent in MyoD2.

#### 3.4.3 Phylogenetic reconstruction of teleost myod genes within a MRF framework

Phylogenetic reconstruction of teleost *myod* genes was originally performed within a framework of multiple vertebrate MRF AA sequences (focusing on teleost MyoD) using ML and for comparison with a previous analysis (Atchley et al., 1994), by NJ. The ML/supporting NJ tree is shown in 3.3 and is identical in topology to that published in Macqueen and Johnston (2006). Amphioxus and tunicate MyoD orthologues were used as outgroups with the former to root the tree. MyoD and Myf5 sequences can be seen to separately branch from an Mrf4/Myog clade. In

teleosts, MyoD branched into two separate clades; one (MyoD1) represented by all teleosts (including species of the Acanthopterygii, Ostariophysii and Protacanthopterygii) and a second (MyoD2) is represented solely by species of the Acanthopterygii (halibut, pufferfish and gilthead seabream). All salmonid MyoD paralogues (MyoD1a, b and c) can be seen to branch together from the MyoD1 lineage. Within the salmonid MyoD1 clade, it can be seen that MyoD1b/1c are sister sequences, branching from MyoD1a. The specific evolution of MyoD1 paralogues in salmonids is further discussed in the next chapter. Additionally, MyoD paralogues in the tetraploid frog *X. laevis* formed a clade internally to the single MyoD sequence found in its diploid relative, *X. tropicalis*. The single cluster of tetrapod MyoD sequences branched externally from the teleost MyoD1 clade, but internally relative to Acanthopterygian MyoD2. This topology was supported by 100% bootstrap confidence in both ML and NJ approaches. Taken literally, this topology suggests that MyoD2 sequences originated prior to the separation of the Sarcopterygii and Actinopterygii, and that MyoD2 was lost in all vertebrate lineages except the Acanthopterygii. This could be an artefact originating from the rapid evolution of teleost MyoD2 relative to MyoD1 (note the long branch lengths leading to MyoD2 in Fig 3.4). Asymmetric evolution of paralogues is known to affect tree topology through LBA (Fares et al., 2006). LBA occurs when branches on a phylogenetic tree are attracted and cluster together, often with strong support, irrespective of their true phylogenetic relationships. This happens due to the limited nature of the molecular code as a comparative tool. Phylogenetic analysis cannot distinguish whether an AA is present in two lineages as a result of shared ancestry, or by independent chance (Bergsten, 2005). Since there are only 20 AAs, two long branches may by chance have more similar AA positions to be considered as synapomorphies (residues conserved from a common ancestor), compared to a shorter branch which may be a true sister group to one of the longer branches (Bergsten, 2005). In this case the rapid divergence of Acanthopterygian MyoD2 relative to MyoD1 sequences, may explain its position as an outgroup to all vertebrate

MyoD sequences, since it may be ‘attracted’ to other long branches in the alignment, i.e. Myf5, Myog, Mrf4 and the invertebrate orthologues.

#### 3.4.4 Further phylogenetic reconstruction of MyoD

A more expected position for Acanthopterygian MyoD2 within a vertebrate MyoD tree topology would be to either branch from all teleost MyoD sequences, if it arose in a common teleost ancestor (e.g. during the teleost WGD), or from Acanthopterygian MyoD1 if a specific *myod* duplication occurred within this lineage. In an attempt to recover a more expected tree topology, a new MyoD alignment was produced (not shown here, see Fig. S1 in publication 5, page VII), with 17 MyoD sequences including paralogues found within different vertebrate taxa (salmonids, Acanthopterygians and frogs) but with reduced sequence representation of potential long-branches, including sequences for Myf5, Mrf4 and Myog as well as basal-deuterostome MyoD orthologues. This alignment was then used to depict the phylogenetic relationships of MyoD orthologues/paralogues by Bayesian, ML and NJ approaches. By all methods of reconstruction and as observed in Fig 3.3, *X. laevis* MyoD paralogues branched as a sister clade from *X. tropicalis* MyoD (Fig. 3.4 a-d). Additionally, all salmonid MyoD paralogues again branched as co-orthologues of teleost MyoD1, with the MyoD1b/1c paralogues being sister sequences branching from MyoD1a (Fig. 3.4 a-d). Interestingly, Bayesian, ML and NJ analyses placed the point of the MyoD1/MyoD2 duplication as a specific event within the Acanthopterygii when a gamma distribution of among site rate variation was used which is known to be resistant to LBA (Bergsten, 2005; Fares et al., 2006) (Fig. 3.4 a-c). Bayesian inference and ML also placed the duplication as a specific event to Acanthopterygians when among-site substitution rates were considered low or uniform (not shown). Conversely, when a NJ tree analysis was performed assuming a uniform distribution of among site rate variation, the tree topology supported a common teleost origin of MyoD1/MyoD2 paralogues (Fig. 3.4 d). Finally, by removing

frequently mutating residues from the alignment before NJ tree reconstruction, a topology was retrieved supporting an Acanthopterygian origin of paralogues (not shown). Thus, with the new alignment, most methods of phylogenetic reconstruction supported an Acanthopterygian specific event and all methods corrected the external position of MyoD2 to all vertebrate MyoD proteins.

#### 3.4.5 Genomic neighbourhood surrounding *myod* genes

The next aim was to establish the chromosomal locations of genes in proximity to *myod* in human, relative to their positions in chicken, zebrafish and three Acanthopterygian species. This information was used to construct a diagram of conserved synteny across the vertebrates (Fig. 3.5). Additionally, since *tropT* and *tropI* genes are in direct 3' proximity to all teleost *myod* genes, I also assessed their location in human and chicken genomes. A remarkable degree of synteny is retained between the *myod* containing regions of human chromosome 11 and chicken chromosome 5 (Fig. 3.5). Comparing these regions with teleosts, while some inter and intra chromosomal rearrangements have occurred, a striking pattern of double conserved synteny is observed where teleost genes are found as either single copies interspersed between two paralogous chromosomal tracts (*otog*, *abcc-8*, *kcnj11*, *pik3c2a*, *rps13*, *sergef*) or as at least two copies on both chromosomes (*tropT*, *tropI*, *tph1*, *kcnc1* [zebrafish specific], *nucb2*, *plekha7*) (Fig. 3.5). This pattern was maintained for genes found in both upstream and downstream proximity to *myod* in human/chicken and importantly, was observed in zebrafish (Ostariophysi) and the three Acanthopterygian species studied (Fig. 3.5). However, on zebrafish chromosome 5, the duplicated *myod2* gene is absent relative to its inferred position from Acanthopterygian genomes (Fig. 3.5, black arrow).

#### 3.4.6 Phylogenetic reconstruction of *myod*-neighbouring genes

Next, phylogenetic relationships were established for six genes found in proximity to *myod* in human/chicken genomes that were found as two copies on two paralogous chromosomes in teleosts. For *Kcnc1*, two copies were retained on the two paralogous chromosomes in zebrafish, but not Acanthopterygian species, which have retained this gene on a single chromosome orthologous to zebrafish chr 25 (Fig. 3.5). The Bayesian/ML analyses clustered one of the zebrafish paralogues (*Kcnc1-1*) with the Acanthopterygian sequences, and its paralogue *Kcnc1-2* (on chr 7), as an outgroup to these sequences, but internally to tetrapod orthologues (Fig. 3.6 a). *Nucb2* and *Plekha7* paralogues, which are common to all teleosts examined (Fig. 3.5), formed two teleost sister clades, branching from tetrapod orthologues (Fig. 3.6 b-c).

Fast skeletal muscle specific *tropI* genes are closely associated with *myod* genes in all teleost genomes, and appear more distally downstream of *myod* in tetrapod genomes (Fig. 3.5). In teleosts, *tropI* can be found as distinct tandem paralogues (ranging from 2-5 in number) just downstream of *myod1*, but also in proximity to Acanthopterygian *myod2* genes and the position where the *myod2* gene of zebrafish was putatively lost (black arrow on chr 7, Fig. 3.5). Conversely, fast muscle specific *tropI* appears as a single gene on chromosomes 11 and 5 in human and chicken genomes. Thus, it seems that *tropI* has been through a series of in-chromosomal (tandem) duplications and a chromosomal duplication event specifically during teleost evolution. For ease, the tandem paralogues on each teleost chromosome were designated as *a*, *b*, *c* etc, based solely on their left to right position on Fig. 3.5. To investigate their evolutionary relationships, Bayesian and ML phylogenetic trees were constructed for all teleost *TropI* sequences within the scope of the synteny analysis, which produced identical topologies (Fig. 3.6d). Interestingly, teleost sequences orthologous to zebrafish *TropI-1c* (stickleback-*TropI-1c*, medaka-*TropI-1b*, pufferfish-*TropI-1B*) clustered as an outgroup to a clade containing

all other teleost TropI sequences, with 100% branch support from both methods (Fig. 3.6d). This suggests that these TropI orthologues are the least derived relative to tetrapod TropI and are likely ancestral to all other teleost TropI paralogues, tandem or otherwise. The fact that the next teleost TropI sequences to branch internally to this clade (zebrafish TropI-1d, stickleback TropI-1d, medaka TropI-1a) are found on the same chromosome as the ‘ancestral’ TropI sequence, likely reflects an ancient teleost tandem duplication event in a common teleost ancestor (Fig. 3.6, d, marked by a  $*(T)$ ). Internal to these branches, are TropI sequences from the paralogous chromosome (i.e. zebrafish chr 7, stickleback group 2 and tiger pufferfish scaf 1) (Fig. 3.6 d). This branching likely reflects the chromosomal duplication event (black star on Fig. 3.6 d) suggested by trees constructed for other neighbouring genes (Fig. 3.6, a-c, f). Branches found internally to these sequences correspond to TropI sequences found in tandem with the ancestor TropI proteins (i.e. in zebrafish TropI-c and d). These results suggest that *tropI* duplicated in tandem prior to the teleost WGD event and other paralogues, either tandem or chromosomal are derived from these ancestral sequences

However, the Bayesian/ML trees retrieved for Tph1 and TropT paralogues, were not consistent with other trees and either branched one of the zebrafish genes as a sister group to its paralogue (TropT, not shown) or externally to all teleost genes (Tph1, Fig. 3.6, e). These are possible tree artefacts arising from the different rates of paralogue evolution between zebrafish and Acanthopterygian species. However, employing a gamma distribution of among-site rate variation in the Bayesian analysis did not change the topology of either tree, but did reduce posterior probability values at the suspected aberrant positions (not shown). To test for mutational saturation in these alignments, NJ trees were constructed considering all substitution sites and then solely the unsaturated fraction of sites. NJ considering all sites retrieved trees very similar to the Bayesian/ML analyses for both Tph1 and TropT (not shown). However, when the

unsaturated alignments were analysed, both trees changed in topology, suggesting these alignments were affected by mutational saturation. The ‘unsaturated’ Tph1 NJ tree topology was changed in a manner consistent with other trees and branched teleost duplicates into two paralogous sister clades (Fig. 3.6, f). However, the expected topology was not retrieved for the TropT alignment by this approach and the two zebrafish sequences formed a sister clade with low branch confidence (not shown).

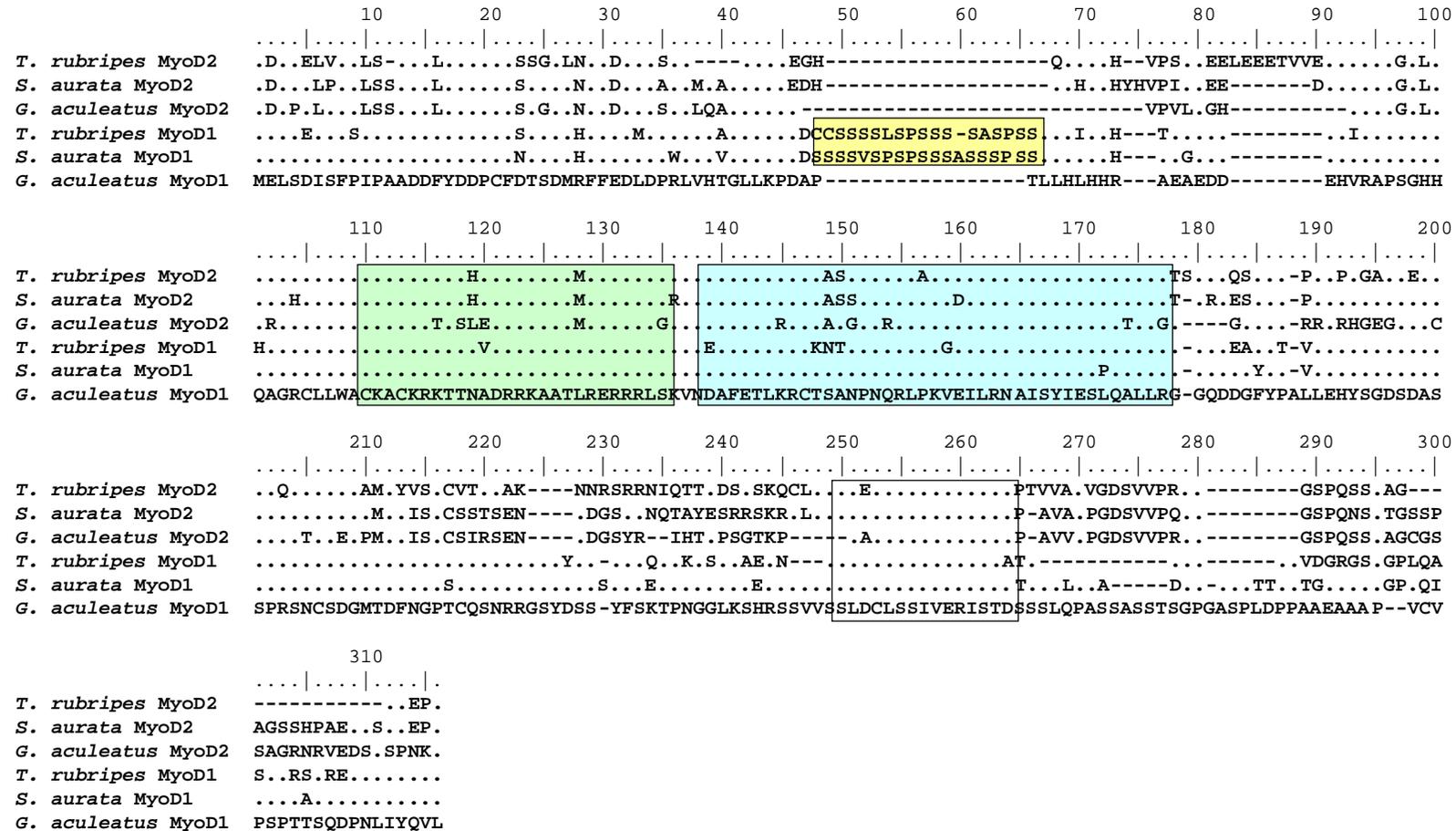


Fig. 3.1 Full AA sequence alignment of MyoD1 and MyoD2 AA sequences in three species of the Acanthopterygii (*T. rubripes*, *S. aurata* and *G. aculeatus*). AA sequences identical to stickleback MyoD1 are indicated by a dot and gaps are shown as a dash. The basic (green box) and HLH domains (blue box) are shown. Additionally, the helix-3 domain is boxed towards the C-terminal. Interestingly, the ser-rich region towards the N-terminal (shown boxed in yellow) is absent from stickleback MyoD1 as well as from all MyoD2 proteins

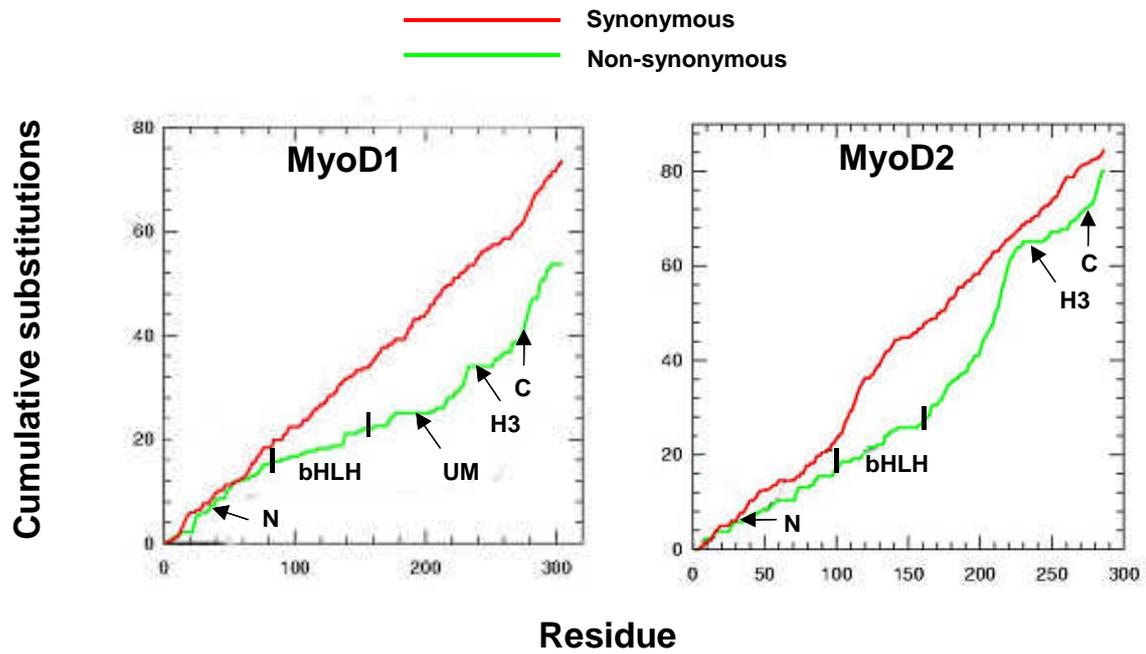


Fig. 3.2. Average ratio of synonymous and non-synonymous substitutions in MyoD1 and MyoD2 of Acanthopterygian teleosts. Regions/motifs within MyoD proteins are marked by arrows as N: N-terminus, bHLH: basic-helix-loop-helix, UM: highly conserved motif of unknown function, H3: amphipathic helix-3 and C: C-terminus.

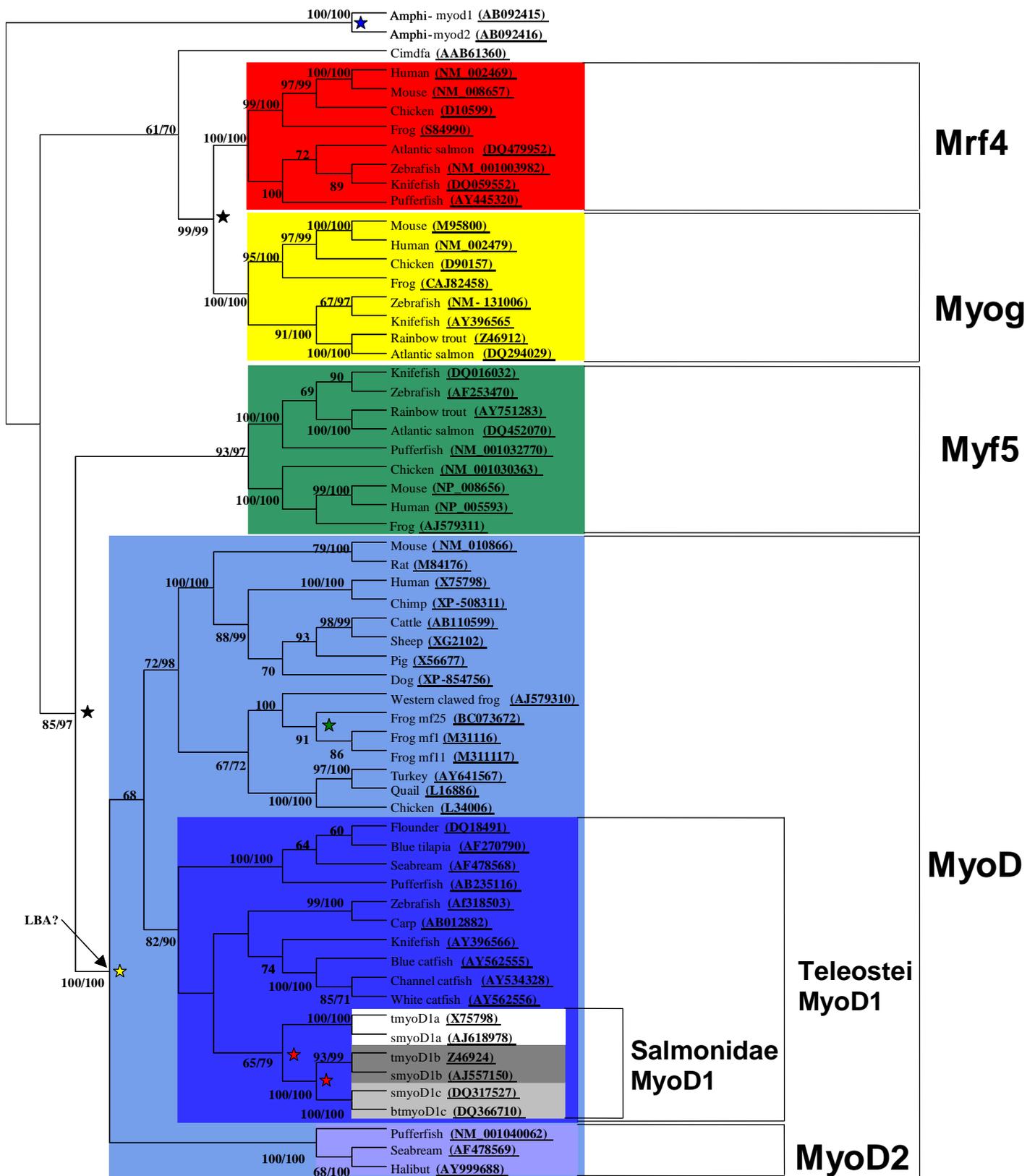
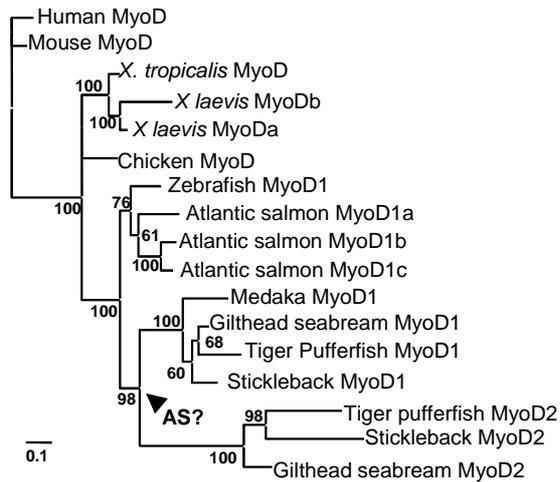


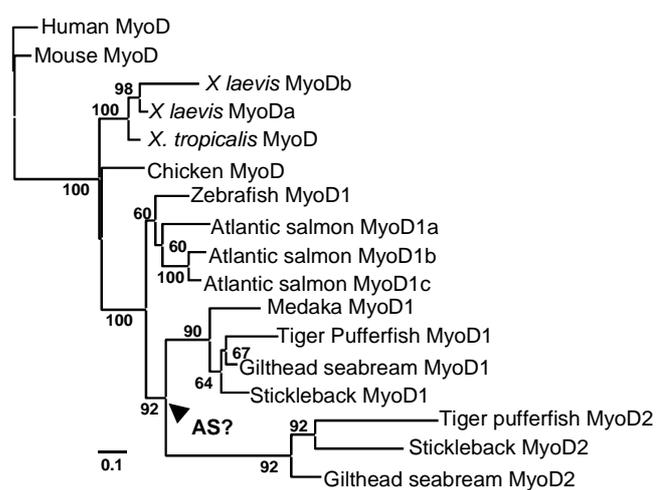
Fig. 3.3. Legend on next page

Fig. 3.3. ML cladogram showing the phylogenetic reconstruction of vertebrate MRF sequences. This tree was published in Macqueen and Johnston (2006). Stars of different colours respectively show an independent MyoD duplication in the Amphioxus lineage (blue star), the duplication events leading to the current MRFs, Myog/Mrf4 and MyoD/Myf5 (black stars), the allotetraploidization of the *X. laevis* genome (green star), duplications of MyoD1 specific to the salmonid lineage (red stars; discussed in chapter 4) and the duplication leading to the conservation of MyoD1/MyoD2 in the Acanthopterygii (yellow star). The positions of Acanthopterygian MyoD2 sequences are almost certainly incorrect and may be an artefact of long-branch attraction (marked LBA?). Branch confidence is respectively from 500 (ML) and 1000 (supporting NJ) bootstrap replicates (values >60% shown).

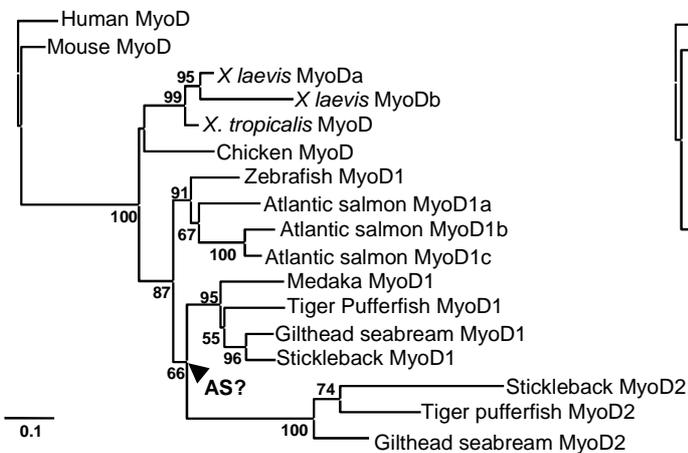
(a) Bayesian; gamma distributed



(b) Maximum likelihood; gamma distributed



(c) Neighbour Joining; gamma distributed



(d) Neighbour Joining; uniform rate

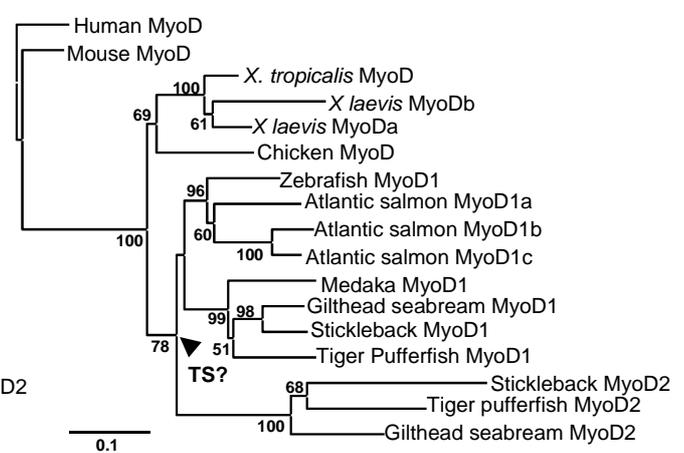
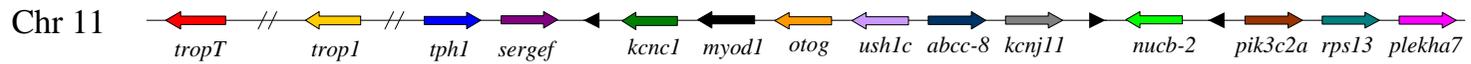
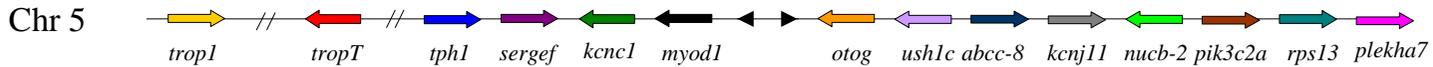


Fig. 3.4 Unrooted phylograms of vertebrate MyoD AA sequences constructed using several approaches. (a) Bayesian inference with a mixed model of AA substitutions and assuming a gamma distribution of among-site substitution rates. (b) ML with the WAG model of AA substitution and assuming a gamma distribution of among-site substitution rates (gamma distribution parameter estimated by PhyML as 0.66) with 500 bootstrap replicates. (c) NJ using the Poisson correction model and assuming a gamma distribution of among-site rates (gamma distribution parameter of 0.66) and 1000 bootstrap replicates. (d) NJ using the Poisson correction model assuming a uniform distribution of among-site substitutions rates with 1000 bootstrap replicates. The same vertebrate MyoD duplication events observed in Fig 3.3 are recorded in all trees, except that the position of Acanthopterygian MyoD2 external to all vertebrate MyoD sequence is corrected. Arrows marked AS refer to the Acanthopterygian specific (AS) positioning of the teleost MyoD1/2 duplication inferred in trees a-c. The arrow marked TS shows the teleost specific (TS) positioning of the teleost MyoD1/2 duplication event inferred by tree d. Scale bars show the number of substitutions per site. Branch confidence values >50% are shown.

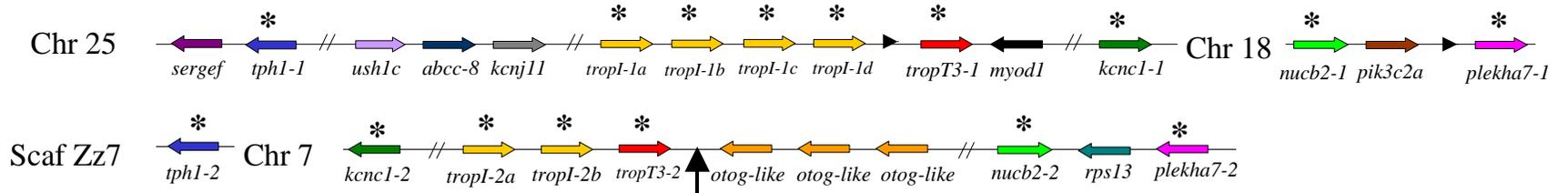
# Human



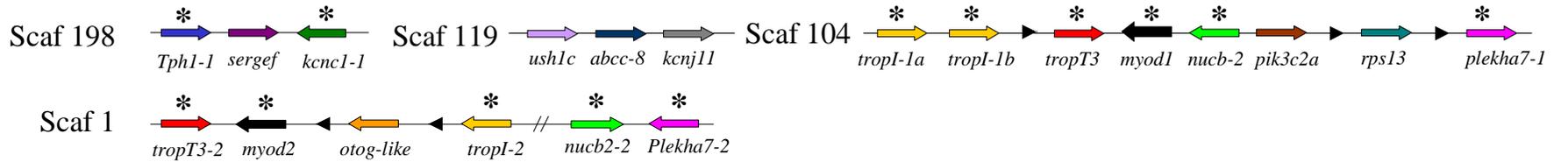
# Chicken



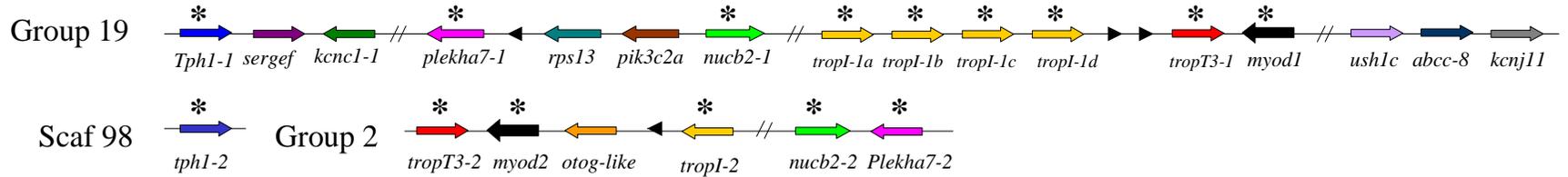
# Zebrafish



# Pufferfish



# Stickleback



# Medaka

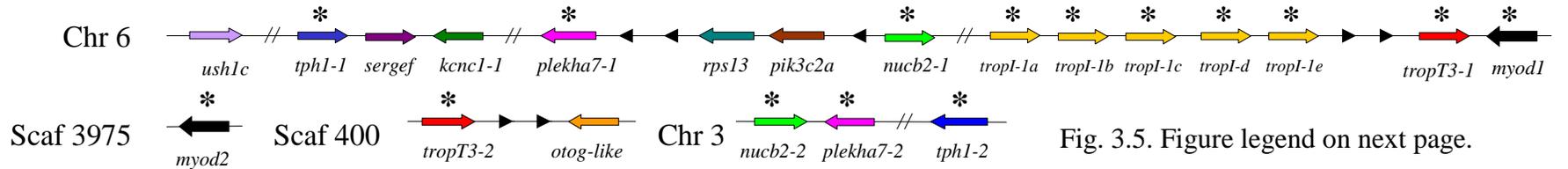


Fig. 3.5. Figure legend on next page.

Fig. 3.5. Depicts the synteny conserved between the *myod*-containing chromosomes of human, with that of chicken, zebrafish, pufferfish, stickleback and medaka. A striking pattern of interleaved double conserved synteny can be seen where teleost genes are distributed between two regions as either single copies or paralogues. This, in contrast to the direct depiction of MyoD phylogenetic relationships (Fig. 3.4), suggests that a *myod*-containing chromosome duplicated in a common teleost ancestor. Genes are not scaled by size and are represented by arrows (identifying the direction of transcription) coloured by their orthology to human genes. Black arrowheads represent genes not conserved between humans and other species on the chromosomal region investigated. Double diagonal lines represent a gap of more than three genes. Teleost genes found on the two paralogous chromosomal regions are marked with a black star. The black arrow on zebrafish chromosome 7 marks the putative position where *myod2* was non-functionalized. Teleost genes orthologous to those on zebrafish chromosome 25 and 7 are respectively designated as Gene-1 and Gene-2, to identify their common paralogy. Multiple tandem *tropI* genes present on duplicated teleost chromosomes are labelled as *a*, *b*, *c*, etc based on their left to right position and not by their inferred paralogy/orthology from phylogenetic reconstruction (Fig. 3.5d).

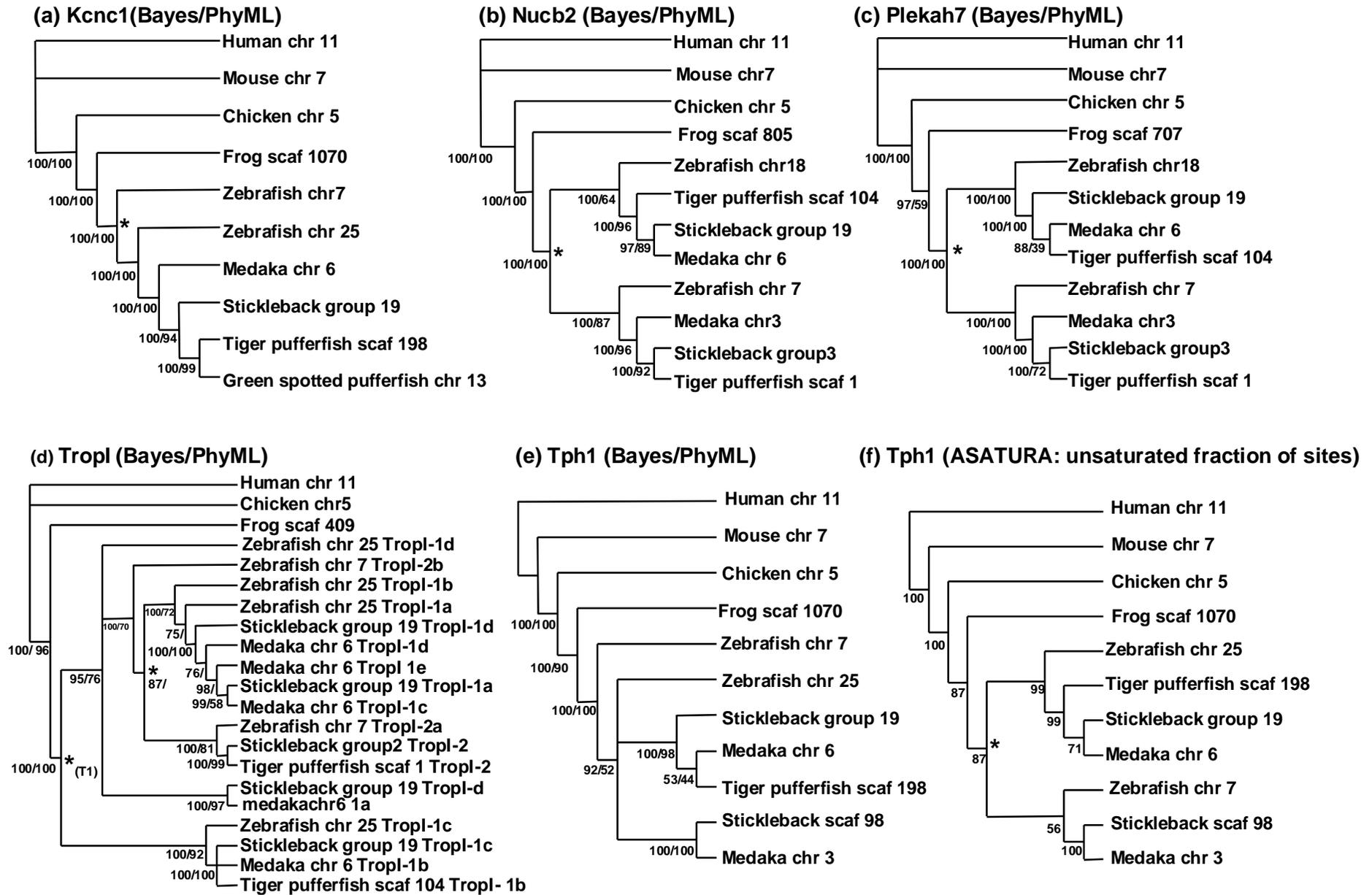


Fig. 3.6. Legend is on the next page

Fig. 3.6. Unrooted phylogenetic cladograms for AA translations of genes in proximity to *myod* that are conserved as two copies on two paralogous chromosomal regions in teleosts. Branch confidence values from different phylogenetic reconstruction methods are shown in the order they are bracketed. (a) *Kcnc1* (Bayesian/ML topology). (b) *Nucb2* (Bayesian/ML topology). (c) *Plekh7* (Bayesian/ML topology). (d) *TropI* (Bayesian/ML topology). \* represents a chromosomal duplication event arising in a common teleost ancestor. \*(T1) represents the presumed first tandem duplication of *TropI* (e) *Tph1* (Bayesian/ML topology). (f) *Tph1* (topology corrected for mutational saturation). Branch confidence values >50% from the different reconstruction methods are shown.

## 3.5 Discussion

### 3.5.1 *MyoD2* is specifically conserved in the Acanthopterygii and is evolving rapidly relative to its paralogue

It has previously been observed that teleosts of the superorder Acanthopterygii, including the gilthead seabream and Atlantic halibut (experimentally supported: Tan and Du, 2002, Galloway et al., 2006) as well as pufferfishes, sticklebacks and medaka (genome predictions) have two MyoD paralogues (Fernandes et al., 2007). However, in zebrafish genomes as well as EST libraries for other representatives of the Ostariophysi (catfish, fathead minnow) a single *myod* gene was retrieved. This suggests that *myod2* either arose by duplication in the Acanthopterygii lineage, or has been retained specifically in this group following a common teleost event. Here the *myod2* gene of the tiger pufferfish *T. rubripes* was also characterised. MyoD2 has clearly evolved at a faster rate than MyoD1 as the phylogenetic branch lengths leading to MyoD2 following the MyoD1/MyoD2 duplication are attenuated in all analyses (Fig. 3.4a-d). A more detailed explanation of the asymmetric evolution of MyoD paralogues in different vertebrate groups can be found in the associated publication to this chapter (publication 5, page VII). Interestingly, the serine-rich motif found in *T. rubripes* and *S. aurata* MyoD1 proteins (Fig. 3.1) is absent from MyoD1 of stickleback as well as all MyoD2 sequences. It has been suggested that this motif has a regulatory role through post-translational phosphorylation (Fernandes et al., 2007). This motif is also conserved in halibut (*H. hippoglossus*), flounder (*P. olivaceus*) and tilapia (*O. aureus*), suggesting that its lack in stickleback MyoD1 is an anomaly among the Acanthopterygii. The presence of the ser-rich motif, as well as the conservation of MyoD2 specifically within the Acanthopterygii suggests that myogenesis in this group may have some unique transcriptional and post-translational levels of regulation compared to other teleosts.

To investigate the selective constraints on MyoD proteins during the evolution of the Acanthopterygii, plots were produced of the cumulative number of substitutions that were

synonymous or AA changing across MyoD1/MyoD2 in four species. Both proteins have evolved under selective pressure to avoid AA substitutions in the bHLH and helix-3, which is unsurprising considering their essential nature for normal MyoD function (Tapscott, 2005, Bergstrom and Tapscott, 2001). However, for both MyoD1/2 the rate of protein changing substitutions is attenuated at the N/C termini (excluding the last 8 residues of MyoD1), suggesting these regions have evolved under relaxed constraints. A striking difference between MyoD1/2 was observed in the ~80 residues between the bHLH and helix-3 of MyoD2, which showed a rapid rise in AA changing substitutions suggesting that this region has evolved under positive selection. Conversely, the equivalent region in MyoD1 is strongly conserved (Fig. 3.1) and under more stringent pressure to avoid AA changing nucleotide substitutions (Fig. 3.2). It is possible that this region contributes to differences in protein function between MyoD1/2 paralogues in the Acanthopterygii.

### 3.5.2 *MyoD* duplications and vertebrate polyploidy

In this study a ML phylogenetic reconstruction was initially used to establish the relationships of vertebrate MyoD paralogues and orthologues within a scaffold of other MRFs (Fig. 3.3). This tree supported a previous analysis that suggested that the four MRFs arose from a common ancestral gene that duplicated twice, to first produce the ancestor genes to Myf5/MyoD and Mrf4/Myog and subsequently to produce the current MRFs (Atchley et al., 1994, Fig. 3.2, black stars). Additionally this tree indicated that MyoD has duplicated several times in different animal taxa producing paralogues in the amphioxus, the allotetraploid frog, *X. laevis*, salmonid teleosts and Acanthopterygian teleosts. It is clear from the sister grouping of paralogues in amphioxus, frogs and salmonids that these duplicates arose through lineage specific events, i.e. the allotetraploidization of the *X. tropicalis* genome (Bisbee et al., 1977), the tetraploidization of the salmonid genome (Allendorf and Thorgaard, 1984) and likely through an independent duplication in the amphioxus. However, the Acanthopterygian MyoD paralogues clustered

externally to all vertebrate MyoD sequences (Fig. 3.3), a topology supporting a common vertebrate MyoD duplication followed by loss in every vertebrate lineage except the Acanthopterygii. However, by testing an alignment containing only MyoD sequences, and thus excluding possible ‘long branches’, Bayesian/ML/NJ trees were constructed that consistently corrected this aberrant topology (Fig. 3.3, a-d). Thus, LBA may have affected the original ML tree reconstruction (Fig. 3.3). Furthermore, most trees consistently supported an Acanthopterygian specific origin of MyoD2 (Fig. 3.4), including all Bayesian and ML reconstructions, which are known to be more accurate than NJ methods at reconstructing distantly related or quickly evolving sequences (Holder and Lewis, 2003) such as MyoD.

### *3.5.3 A comparative genomic study of myod neighbouring genes reveals the true extent of the MyoD1/MyoD2 duplication*

Thus, when I employed a comparative genomic approach to study the relationships of genes in neighbourhood to *myod* in several teleosts and two diploid tetrapods, I fully expected that some signal of *myod* duplication would be retrieved specifically in the Acanthopterygii. However, the common pattern of interleaved-double conserved synteny observed in teleosts relative to tetrapods (Fig. 3.5), is most consistent with the duplication of a *myod*-containing chromosome in a common ancestor to zebrafish (Ostariophysi) and the Acanthopterygii, but not tetrapods, the most parsimonious explanation being during the WGD of basal teleost evolution (Jaillon et al., 2004). This finding was supported by reconstructing the phylogenetic relationships of genes in neighbourhood to *myod* in diploid tetrapods, which were found as duplicates in teleosts. These trees generally (in 6/7 cases) formed two paralogous teleost sister clades internal to a single tetrapod clade (Fig. 3.6). Thus, the position of the teleost MyoD1/MyoD2 duplication supported by direct phylogenetic analysis (Fig. 3.4) is almost certainly incorrect. These results highlight the importance of avoiding the use of single gene phylogenies when inferring the origin of gene paralogues and advocate the importance of studying the conserved synteny between, and phylogenetic relationships of, neighbouring genes in duplicated and non-duplicated lineages.

Another interesting finding was the multiple tandem and chromosomal paralogues of fast muscle *tropI* that were found in proximity to *myod* genes in all teleosts, but not mammals or birds (Fig 3.5). It is known that MyoD binds to and regulates the expression of fast muscle *tropI* genes through interactions with E-proteins (Lin et al., 1991). The presence of multiple tandem fast-muscle *tropI* paralogues in close association with *myod* in teleosts but not tetrapods suggests that a selective advantage has arisen in teleost evolution for the tight regulation of multiple copies. Embryonic *in situ* expression data is available for one zebrafish fast skeletal muscle *tropI* gene. The zebrafish probe used by Thisse et al. (2001) (denoted *tnni2*) shares 100% identity to the putative Ensembl transcript of the *tropI-1d* gene (Fig. 3.5) and from mid-somitogenesis accumulated in muscles of the somite, fin buds and head (Thisse et al., 2001) which overlaps spatially and temporally with *myod1* transcripts (Weinberg et al., 1996). Additionally, in Atlantic cod (*G. morhua*) a cRNA probe orthologous to zebrafish *tropI-1d* was similarly expressed throughout the developing myotome during embryogenesis (Hall et al., 2003). These findings suggest that this *tropI* gene is likely regulated by *myod1* during embryonic myogenesis. *In situ* expression data is not available for other fast-skeletal *tropI* genes. To gain insight into their regulation I performed tBLASTn searches of the EST database at GenBank using full AA translations of each zebrafish *tropI* gene within Fig. 3.5. A cut-off of 98-100% sequence identity was considered a positive hit from the returned sequences. Positive hits were returned for each *tropI* gene, confirming that each paralogue is transcribed into an mRNA product. Consistent with the *in situ* data, several hundred positive hits for zebrafish *tropI-1d* were retrieved solely from EST libraries representing embryonic zebrafish tissues. Interestingly, other *tropI* genes were not limited to embryonic tissues and were abundant in cDNA libraries obtained from adult zebrafish brain (*tropI-1c, 1a, 2a, 2b*), skin (*tropI-1c*), eye (*tropI-1b, 2a*), gill (*tropI-1c*), intestine (*tropI-1c*), gut (*tropI-1a*) and cultured myoblasts (*tropI-2b*). Similarly, BLAST searches of several Atlantic salmon EST databases using the various zebrafish TropI AA sequences retrieved

multiple salmon *tropI* ESTs from tissue-specific cDNA libraries including fast muscle, slow muscle, gill, heart, skin, brain and eye. These findings suggest that the multiple ‘fast-muscle’ specific *tropI* paralogues found in teleosts are not solely involved in the assembly of fast skeletal muscle. Further their expression in multiple tissues is clearly not limited to regulation by muscle-specific transcription factors like *myod*. A more detailed examination of the expression patterns of teleost fast skeletal *tropI* duplicates would be a fruitful future experiment to gain insight into the evolution of cis-acting regulation of paralogues following gene duplication.

The phylogenetic and genomic findings presented here require that *myod2* was lost in zebrafish and since this gene is not represented in salmonid, minnow or catfish EST libraries, this notion can be tentatively extended to the Ostariophysii and Protacanthopterygii lineages. The differential retention/loss of paralogues in different teleost lineages following the WGD is surprisingly common. For example, it was shown that ~50% of zebrafish paralogues were retained as single copies in pufferfish genomes (Woods et al., 2005; Taylor et al., 2003). Additionally, a reverse situation is presented in chapter 5, where two WGD paralogues coding for Fst, are shown to be retained specifically in the Ostariophysii, but not other lineages.

#### 3.5.4 Chapter conclusion: a consensus nomenclature for vertebrate *MyoD* sequences

Here I have provided strong evidence that a chromosomal region containing *myod* duplicated in a common teleost ancestor, but that *myod2* was lost in non-Acanthopterygian lineages. The current vertebrate nomenclature is generally author specific and based on the timing of MyoD discovery and does not account for evolutionary relationships of paralogues that have arisen in different vertebrate lineages. These results suggest that a consensus nomenclature should be implemented that is relevant to all vertebrate *myod* genes. I suggest that for all teleost species that have arisen subsequent to the WGD, *myod* paralogues should be first identified by their orthology to either *myod1* or *myod2* and then more recently derived copies discovered within specific lineages

should be named within this framework as *myod1(a/b/etc)* or *myod2(a/b/ etc)*. For other vertebrates that did not go through the teleost WGD, including tetrapods, and basal Actinopterygian groups such as the Acipenseridae and Lepisosteidae, *myod* orthologues retained as a single copy should be simply denoted *myod*, whereas lineage specific paralogues should be called *myoda/b/ etc* (e.g. *X. laevis* MyoDa/b). This evolutionary relevant nomenclature, which is highlighted in Table 3.1 provides the simplest way of distinguishing between *myod* paralogues arising from the teleost WGD and those arising from lineage-specific duplication events. Furthermore, considering the frequency of polyploidy in fishes, amphibians and reptiles (Otto and Whitton, 2000; Le Comber et al., 2004) and the importance of the ongoing study of MyoD, it is likely that many more paralogues will be characterised in the future.

Table 3.1. Details of teleost MyoD sequences, including their current designation, Genbank accession number/Ensembl gene ID as well as correct designations (proposed consensus nomenclature) according to the comparative-genomic and phylogenetic results of this study.

Species	Current designation	Sequence	GenBank accession	Ensembl gene ID	Suggested designation
<b>Tetrapoda</b>					
<i>Homo sapiens</i>	MyoD	complete cds	CAA40000	ENSG00000129152	MyoD
<i>Mus musculus</i>	MyoD	complete cds	X56677	ENSMUSG00000009471	MyoD
<i>Gallus gallus</i>	MyoD	complete cds	X16189	ENSGALG00000006216	MyoD
<i>Xenopus tropicalis</i>	MyoD	complete cds	AJ579310	ENSXETG00000001320	MyoD
<i>Xenopus laevis</i>	Mf1	complete cds	M31116	n/a	MyoDa
	Mf25	complete cds	M31118	n/a	MyoDb
<b>Ostariophysi</b>					
<i>Danio rerio</i>	MyoD	complete cds	NM_131262	ENSDARG00000030110	MyoD1
<i>Cyprinus carpio</i>	MyoD	complete cds	AB012882	n/a	MyoD1
<i>Sternopygus macrurus</i>	MyoD	complete cds	AY396566	n/a	MyoD1
<b>Protacanthopterygii</b>					
<i>Salmo salar</i>	MyoD1	complete cds	AJ557148	n/a	MyoD1a
	MyoD2	complete cds	AJ557149	n/a	MyoD1b
	MyoD1c	complete cds	DQ317527	n/a	MyoD1c
<i>Salmo trutta</i>	MyoD1c	complete cds	DQ366710	n/a	MyoD1c
<i>Oncorhynchus mykiss</i>	TMyoD	complete cds	X75798	n/a	MyoD1a
	TMyoD2	complete cds	Z46924	n/a	MyoD1b
	unnamed	EST: partial	CX137438	n/a	MyoD1c
<b>Acanthopterygii</b>					
<i>Takifugu rubripes</i>	MyoD1	complete cds	NM_001032769	SINFRUG00000154785	MyoD1
	MyoD2	complete cds	NM_001040062	SINFRUG00000163904	MyoD2
<i>Tetraodon nigroviridis</i>	MyoD1	partial cds	AY616520	GSTENG00003954001	MyoD1
	MyoD2	genomic fragmented	n/a	GSTENG00034775001	MyoD2
<i>Oryzias latipes</i>	MyoD1	genomic complete	n/a	ENSORLG0000000694	MyoD1
	MyoD2	genomic fragmented	n/a	UTOLAPRE05100109983	MyoD2
<i>Gasterosteus aculeatus</i>	MyoD1	genomic complete	n/a	ENSGACG00000008444	MyoD1
	MyoD2	genomic complete	n/a	ENSGACG00000017350	MyoD2
<i>Hippoglossus hippoglossus</i>	MyoD1	partial cds	AY999688	n/a	MyoD1
	MyoD2	complete cds	AJ630127	n/a	MyoD2
<i>Sparus aurata</i>	MyoD1	complete cds	AF478568	n/a	MyoD1
	MyoD2	complete cds	AF478569	n/a	MyoD2

## Chapter 4. Characterisation of a novel differentially expressed Atlantic salmon MyoD paralogue

### 4.1 Abstract

In this chapter a novel *myod* gene (*myod1c*) was characterised and shown to be conserved in Atlantic salmon (*S. salar*), brown trout (*S. trutta*) and rainbow trout (*Oncorhynchus mykiss*). In chapter 3, phylogenetic reconstruction showed that Atlantic salmon *myod1c* is a paralogue of *myod1a/myod1b* and that all three genes are co-orthologues of the teleost *myod1* gene, sharing no direct evolutionary heritage with the *myod2* gene of Acanthopterygians. *myod1c* is more related to *myod1b* than to *myod1a*, evidenced by a thorough genomic comparison of coding, intronic and untranslated sequences and supported by phylogenetic reconstruction. The most likely explanation for these findings is that two-*myod* duplications have occurred during salmonid evolution, one where *myod1a* and an ancestor gene to *myod1b/1c* arose, and a second duplication of the *myod1b/1c* ancestor gene to produce the current *myod1b* and *myod1c* genes. The tetraploidization of the whole genome accounts for one of these events, but a second event of unknown scale is required for the presence of a third *myod1* paralogue.

An experimental approach was then used to study the roles played by the different *myod* paralogues during myogenesis at different growth stages of Atlantic salmon development. In salmon embryos, the spatio-temporal expression patterns of *myod1* paralogues were distinct but overlapping, as revealed by dual-stain *in situ* hybridisation. *myod1a* was expressed from just prior to the onset of somitogenesis, in notochord adjacent adaxial myoblasts of the presomitic mesoderm, whereas *myod1b* mRNA accumulated in fast muscle progenitors of the posterior-lateral epithelial somite. Interestingly, the *myod1c* expression domain had characteristics of both

its paralogues and was detected contemporaneously to *myod1a* in somitic adaxial cells and to *myod1b* in the posterior-lateral somite. These overlapping domains are likely a reflection of the partitioning of cis-acting regulatory regions between independently evolving promoters of each paralogue, since the combined expression profiles of *myod1a/1b/1c* recapitulated the expression domain of the single zebrafish *myod1* gene. In adult salmon, it was shown by qPCR that *myod1a* was the predominant gene expressed in fast-twitch muscle, whereas *myod1c* was upregulated in slow muscle fibres relative to *myod1a/1b*. Finally, a simple model to explain the conservation and differential expression of three *myod1* paralogues in the salmonid genome is presented within the framework of the classical subfunctionalization model.

## 4.2 Introduction

MyoD is possibly the most important transcription factor governing muscle growth in vertebrates and forms the crux of a transcriptional cascade leading to normal myogenesis (reviewed in Tapscott, 2005). Several vertebrate taxa have retained the *myod* gene as duplicated copies (chapter 3) adding another level of complexity to the underlying regulation of myogenesis in these groups. For example, in salmonids, two *myod* genes have been characterised and were named *myod* and *myod2* in rainbow trout (Rescan and Gauvry, 1996) and *myod1* and *myod2* in Atlantic salmon (Gotensparre, 2004). In this chapter a third novel *myod* gene was discovered, that is conserved across salmonids. This gene was shown in chapter 3 to be a recent paralogue of Atlantic salmon *myod1* and *myod2*. Further a consensus nomenclature was proposed that salmon *myod1/myod2* genes should be renamed *myod1a/myod1b*, while the novel gene should be named *myod1c*. The main aim of this chapter was to experimentally characterise *myod1c*.

During embryogenesis in zebrafish (*D. rerio*), *myod1* is initially expressed in notochord adjacent adaxial cells of the presomitic mesoderm before it is activated across the entire posterior domain

of the somite (Weinberg et al., 1996). These two waves of *myod* expression respectively mark separate slow and fast muscle-precursors (Devoto et al., 1996; Stellabotte et al., 2007). In rainbow trout embryos, it was shown that the joint expression of *myod1a/1b* mRNAs recapitulated these zebrafish expression domains (Delalande and Rescan, 1999). An initial aim of this chapter was to compare the embryonic expression of *myod1c* with its paralogues to establish its potential role in embryonic myogenesis and to aid in establishing an evolutionary scenario whereby three *myod* genes have been conserved in salmonids. A further goal was to use quantitative real-time PCR to establish the expression of the different *myod1* paralogues in adult Atlantic salmon muscle fibres of the fast/slow-twitch phenotype at growth stages where myotube production is active (myotube<sup>+</sup> stage) or inactive (myotube<sup>-</sup> stage).

### 4.3 Material and Methods

#### 4.3.1 Fish sampling

The *S. salar* embryos used in this study were derived from the 10°C treatment described in chapter 2 (section 2.2.2). Adult *S. salar* and *S. trutta* were respectively obtained from EWOS Innovation (Lønningdal, Norway) and the Fisheries Research Service (Pitlochry, UK). Twelve adult salmon were selected, six representing myotube<sup>-</sup> growth stages and six representing myotube<sup>+</sup> growth stages (respective mean mass ± S.D. = 304.8 ± 34.9 g and 4297 ± 600.7 g). Fish were humanely sacrificed and pure samples of fast or slow twitch muscle were dissected from the dorsal epaxial myotome and then frozen and stored in liquid nitrogen.

#### 4.3.2 Computational approaches and sequence retrieval

Salmonid ESTs in the salmon genome project (<http://www.salmongenome.no/cgi-bin/sgp.cgi>), TGI (<http://compbio.dfci.harvard.edu/tgi/>) and cGRASP databases (<http://web.uvic.ca/cbr/grasp/>) were screened for MyoD sequences using AA translations of the *myod1a/1b* salmon sequences

and tBLASTn searches. This revealed a *myod* EST in *O. mykiss* (accession number: CX137438) that translated into a MyoD sequence divergent from other salmonid MyoD sequences published to date.

#### 4.3.3 Cloning Atlantic salmon *myod1c*

Total RNA was extracted and processed from 100 mg of Atlantic salmon fast muscle, as described in chapter 2 (sections 2.4.2-2.4.4). First-strand Atlantic salmon cDNA was then synthesised from 1 µg of total RNA as described previously (chapter 2 section 2.4.6). Primers were designed to flank the whole CDS of *myod1b*/putative *myod1c* (Table 4.1, *myod1c-1* and 2) and were used in a standard RT-PCR reaction using fast-twitch muscle cDNA for *S. salar* and *S. trutta*. PCR products were separated by gel electrophoresis, extracted, purified, cloned, *ECOR1*-digested and sequenced as described in chapter 2 (section 2.4). To amplify the 3'UTR of salmon *myod1c*, a sense primer was designed in a region of high divergence between *myod1b/1c* (Table 4.1, *myod1c-5*), and used in a 3' RACE PCR as described in chapter 2 (section 2.4.8). The RACE PCR product was cloned and sequenced as described above.

#### 4.3.4 Genomic characterisation of *myod1c*

To characterise the genomic organisation of *myod1c*, specific primers were designed in untranslated regions flanking the whole CDS (Table 4.1, *myod1c-3* and 4) and used in a standard PCR reaction using Atlantic salmon genomic DNA, which was extracted as described in chapter 2 (section 2.4.2). The intron-exon organisation of salmon *myod1* genes was assessed by comparing the *myod1* cDNAs with their corresponding genomic sequences using the program Spidey (Wheelan et al., 2001).

#### 4.3.5 *In situ* hybridisation

Primers to amplify cDNA templates for *myod1a/1b* RNA probes can be found in Table 4.1 and amplified less conserved regions of the three paralogues (including the 3' of the CDS and part of the 3' untranslated region). The cDNA template for the *myod1c* probe was the 888 bp sequence created by the 3' RACE PCR reaction. Plasmid DNA containing the probe inserts were then amplified by PCR with T3/T7 primers as described in chapter 2 (section 2.6.2). The cDNA products were used as templates to transcribe sense/antisense digoxigenin and fluorescein labelled cRNA probes with T3/T7 RNA polymerases and dual stain *in situ* hybridisation was performed as described in chapter 2 (section 2.6.2 and 2.6.4) using the protocol in Appendix 1.

#### 4.3.6 Quantitative real time RT-PCR

To compare the relative expression of *myod1* paralogues between fast and slow-twitch myotomal muscle and between myotube<sup>+</sup> and myotube<sup>-</sup> growth stages, qPCR was used. RNA was extracted from 100 mg of fast and slow muscle derived from the six myotube<sup>+</sup> and six myotube<sup>-</sup> fish described above (section 4.3.1) with concurrent elimination of contaminating genomic DNA as described in chapter 2 (section 2.4.2). First-strand cDNA was synthesised using 0.3µg of total RNA as a template. A negative control was included that lacked reverse transcriptase, using a pool of all RNAs. Reactions were performed as described in chapter 2 (section 2.5.1) using an ABI Prism 7000 thermocycler (Applied Biosystems) and one primer pair was designed to span an exon/exon boundary (primers for *myod1a/1b/1c* and the housekeeping genes *β-actin* and elongation-factor1- $\alpha$  (*efl- $\alpha$* ) can be found in Table 4.1). Primers for *myod1* paralogues were designed in divergent regions (particularly at the 3' of primers) to minimise paralogue cross-amplification. Expression of the housekeeping genes *β-actin* and *efl- $\alpha$*  were used to normalise gene expression and were ran on each qPCR plate. The stability of housekeeping gene expression was confirmed by comparing their relative expression between different treatments i.e. between fast/slow fibre types and in myotube<sup>+</sup>/myotube<sup>-</sup> fish. To normalise SYBR Green fluorescence

between plates the fluorescent dye ROX was used. To assess the specificity of primers and to check for primer dimers, a dissociation analysis was used to produce a DNA melting curve from 60-90°C (chapter 2, section 2.5.2). PCR efficiencies were estimated by regressing the logarithm of SYBR green fluorescence against cycle number for every PCR reaction using the program LinRegPCR (Ramakers et al., 2003). The normalised ratio of expression of each *myodl* gene was calculated against the expression of *myod1a* for fast and slow muscle separately and for each *myodl* paralogue in myotube<sup>+</sup> fish relative to myotube<sup>-</sup> fish in both fast and slow muscles. Cycle threshold (ct) values were converted to normalized expression ratios using REST-384<sup>®</sup> (Pfaffl et al., 2002). This program calculates expression ratios based on a mathematical model (see below) (Pfaffl, 2001) where E *gene* and E *standard* are the respective efficiencies of experimental (*myod1a/1b/1c*) or standard (*efl-α* and *β-actin*) genes and Δ ct is the difference between threshold cycles of two treatments. Statistical support for observed differences in gene expression was provided using a pairwise fixed reallocation test in REST-384<sup>®</sup>.

$$\text{ratio} = \frac{(E_{\text{gene}})^{\Delta \text{ct}_{\text{gene}} (\text{Mean control} - \text{mean sample})}}{(E_{\text{standard}})^{\Delta \text{ct}_{\text{standard}} (\text{Mean control} - \text{mean sample})}}$$

Table 4.1. Primer details for chapter 4. Note the prefix's *s* and *bt* respectively indicate a product in Atlantic salmon and brown trout.

Primer name	Product	Primer Sequence (5'-3')
<i>myod1c</i> 1	<i>s/btmyod1c</i> whole cds	f: ATGGAGTTGTCCGGATATTTTCG
<i>myod1c</i> 2	"	r: TCATAGCACTTGGTAGATGGGGTC
<i>myod1c</i> 3	<i>smyod1c</i> gene	f: GACAGTGAGATAGAGATGGAGTTG
<i>myod1c</i> 4	"	r: ATGGCAAGGGAAAGAAAGTGGTC
<i>myod1c</i> 5	<i>smyod1c</i> 3' UTR (RACE PCR)	f: CTACTIONCCTTCGCTGGAGCACTACAACG
<i>myod1c</i> 6	163bp <i>smyod1c</i> (qPCR)	f: CCCTTCGCTGGAGCACTACAACG
<i>myod1c</i> 7	"	r: GCTTCTGGCATCAGCATTGGAG
<i>myod1b</i> 1	172bp <i>smyod1b</i> (qPCR)	f: CGGCGAGAACTACTACCCTATGT
<i>myod1b</i> 2	"	r: GGCACCAGCATTTGGAGTTTC
<i>myod1a</i> 1	112bp <i>smyod1a</i> (qPCR)	f: CCAAATAGTTCCAGACGCAAG
<i>myod1a</i> 2	"	r: ACAGCGGGACAGGCAGAGG
$\beta$ -actin 1	146 bp $\beta$ -actin (qPCR)	f: TGACCCAGATCATGTTTGAGACC
$\beta$ -actin 2	"	r: CTCGTAGATGGTACTGTGTGGG
<i>ef1</i> - $\alpha$ 1	141bp <i>ef1</i> - $\alpha$ (qPCR)	f: GAATCGGCTATGCCTGGTGAC
<i>ef1</i> - $\alpha$ 2	"	r: GGATGATGACCTGAGCGGTG
<i>myod1b</i> 3	<i>smyod1b</i> probe template	f: GACGCATCCAGTCCACAGTCCAAC
<i>myod1b</i> 4	"	r: GATGACGATGACAACACACACAC
<i>myod1a</i> 3	<i>smyod1a</i> probe template	f: CGGACAGGAGGGCAACTAT
<i>myod1a</i> 4	"	r: GACCTTCGCAAGTCTTTGGT

## 4.4 Results

### 4.4.1 Sequence characterisation of *myod1c*

Following initial BLAST searches a rainbow trout EST (CX137438) was recovered that was distinct from rainbow trout/Atlantic salmon *myod1a/b*. Next, RT-PCR was used with salmon fast muscle cDNA and primers designed specifically to amplify the whole CDS of *myod1b*/putative novel *myod*. This produced a single band that was cloned and twelve colonies were cultured in LB broth. Plasmid DNA was extracted from these cultures and six plasmids containing the expected insert size were sequenced, revealing the presence of two similar sized amplicons of 831 bp and 819 bp. BLASTn/tBLASTn searches revealed that the 831 bp amplicon was the *S. salar* orthologue of rainbow trout *myod1b* whereas the 819 bp sequence encoded a single open reading frame of 272 AAs (Fig. 4.1). Phylogenetic reconstruction (chapter 3), direct sequence comparison, and expression data (see following sections) strongly suggested that the 819 bp novel sequence was a new *myod* gene. Accordingly, it was initially submitted to Genbank as *myod3*, but was renamed *myod1c* according to a revised nomenclature (see chapter 3; section 3.5.4) (accession number DQ317527). In Atlantic salmon, *myod1c* shares a higher identity in coding regions to *myod1b* than *myod1a* (Table 4.2). Fast muscle cDNA from *S. trutta* was then used in a PCR reaction with the same primer set and another 819 bp amplicon was sequenced translating into an AA sequence which varied by a only few residues compared to the salmon MyoD1c sequence. This sequence was submitted to Genbank as MyoD1c and assigned the accession number DQ366710. Thus, *myod1c* is a novel gene conserved in at least three salmonid species.

#### 4.4.2 Characterisation of non-coding regions of *myod1c*

The *myod1b* and *myod1c* nucleotide sequences were aligned and a sense primer was designed in a highly divergent region and used in a 3' RACE PCR reaction with Atlantic salmon fast muscle cDNA. This produced an 888 bp product containing the complete 3' untranslated region of *myod1c* including a polyadenylation tail, and polyadenylation signal (AATAAA) (Fig. 4.1). The 3' UTRs of *myod* genes shared less sequence identity than in coding regions (Table 4.2). Next a primer pair was designed from the more disparate UTRs of *myod1c*, which were used to amplify the *myod1c* gene by PCR using *S. salar* genomic DNA as a template. This produced a 1581 bp *myod1c* amplicon that contained three exons and two introns (Fig. 4.1). This sequence was joined *in silico* with the 3' RACE product and a 1815 bp genomic sequence was submitted to GenBank (accession number DQ366709).

Comparative percentage identities of intronic regions of *S. salar myod1* genes can be seen in Table 4.2. *myod1a/1b/1c* have retained a similar genomic structure with three exons of similar size and two introns that are more variable (diagrammatically shown in chapter 7, Fig. 7.5). *myod1b* and *myod1c* shared conserved splice sites, whereas only the intronic donor (GT) of *myod1a* was conserved with *myod1b/1c* genes (not shown). Intron 1 and 2 of *myod1a* are respectively shorter and longer compared to *myod1b* and *1c* (shown in chapter 7, Fig. 7.5). In turn, intron 1 of *myod1b* is larger than *myod1c* owing mainly to several insertions/deletions of 11-34 bp (not shown). The size of intron 2 is virtually conserved between *myod1b/1c* (chapter 7, Fig. 7.5) and this region shares a higher sequence identity compared to intron 1 (Table 4.2). Thus at the genomic and mRNA level *myod1c* and *myod1b* are more similar to each other than to *myod1a*.

#### 4.4.3 Comparison of salmon *MyoD1* paralogues with *MyoD1* orthologues

Fig. 4.2 shows an alignment of the three Atlantic salmon *MyoD1* sequences with other vertebrate *MyoD* orthologues. All vertebrate *MyoD* proteins are strongly conserved at the N-terminal activation domain, the bHLH region, the cis/his rich region just N-terminal to the basic region, the helix-3 motif and in a conserved motif of unknown function just downstream of the HLH. The least conserved regions globally are the extreme N terminus and the C terminus, excluding the last five residues, which are entirely conserved across vertebrates. Additionally, the salmon *MyoD1b/1c* proteins have several AA substitutions in the HLH that are unique among the vertebrates as well as a single site in the basic region that is represented by a different residue in each paralogue, but entirely conserved in other vertebrates (Fig. 4.2). Further, while the helix-3 of salmon *MyoD1c* is entirely conserved with all other vertebrate *MyoD* proteins, *MyoD1a* and *MyoD1c* respectively have one and two unique substitutions.

#### 4.4.4 Expression of *myod1* paralogues during salmon embryogenesis

The expression patterns of *myod1a/1b/1c* were examined in salmon embryos using dual and single stain *in situ* hybridisation (Figs 4.3/4.4). Additionally, in chapter 7 (Fig. 7.6), a diagrammatical summary of the mRNA expression patterns of all Atlantic salmon MRFs is presented covering the whole somitogenesis period. This figure is accompanied by a more detailed description of MRF expression domains including *myod1a/1b/1c* and also includes the expression domains of markers of known differentiation events (chapter 7, Fig. 7.6). Thus the description below is not meant to be exhaustive, but mainly elicits the differences and similarities in expression patterns between the *myod1* paralogues.

At no point did embryos incubated with sense cRNA produce any colour staining other than faint background (Fig. 4.3, A, E, Fig. 4.4, Bii). The first salmon *myod1* paralogue to be detected was

*myod1a*, at the end of the gastrulation period, just prior to segmentation, in two stripes of cells either side of the nascent notochord (chapter 7, Fig. 7.7). From the onset of somitogenesis, *myod1a* transcripts were then detected directly adjacent to the notochord of the newly formed somites and the presomitic mesoderm (Fig. 4.3 B, J). Conversely *myod1b/1c* mRNA was not expressed until around 20 somites had been formed (not shown). *myod1b* transcripts accumulated in all posterior-lateral cells of the newest caudal somites (Fig. 4.3 B, C, F). Interestingly, the expression domain of *myod1c* had characteristics of both *myod1a/1b*. From its first detection until the end of segmentation, *myod1c* co-localised with *myod1a* in notochord-adjacent cells of caudal somites (Fig. 4.3, D, H), but was also present in more lateral regions of the posterior somite until mid-somitogenesis (Fig. 4.3, D). However, *myod1c* expression was never detected as far laterally across the posterior aspect of caudal somites as *myod1b* (e.g. Fig. 4.3, C vs. D). As somites matured, *myod1b* transcripts extended anteriorly to encompass the entire posterior-anterior domain of the somite (Fig. 4.3, C). From around the 30 ss *myod1a/1c* transcripts initially began to spread dorso-ventrally in the maturing medial somite (Fig. 4.3, F) before this domain migrated laterally away from the notochord to span the bulk of the remaining the somite (Fig. 4.3 F, G, I). In chapter 7 (Fig. 7.6), it is shown that this *myod1a/1c* expression domain migrates concurrently to a similar domain of transcripts for the adaxial cell differentiation marker slow myosin light chain-1 (*smlc1*). By the end of segmentation *myod1b* labelling had retracted in the inner myotome and became restricted to the lateral edge of the myotome (not shown, see chapter Fig. 7.6) and as embryos matured further and developed pigmented eyes, *myod1b* transcripts were restricted to the dorsal and ventral extremes of the somite as well as directly adjacent to the horizontal myoseptum at the myotomes periphery (Fig. 4.4, Aii). Conversely at the end of segmentation and early eyed stage, *myod1c/1a* were initially maintained in the bulk of the myotome (Fig. 4.5, Ai) but as embryos developed further (evidenced by growth of pectoral fin buds), transcripts were increasingly reduced in the inner myotome, but maintained in the

lateral myotome edge, particularly in dorsal/ventral regions and adjacent to the horizontal septum (not shown here see chapter 7, Fig. 7.6).

During the eyed stage, each salmon *myod1* gene co-localised in several extraocular muscles (Fig. 4.4. \*, Bi, C), the developing muscles of the hyoid and mandibular arches (Fig. 4.4 \*, Bi, D), the medial region of branchial arches 1-5 (Fig. 4.4 \*, Bi, E) and in two stripes in the developing fin buds (Fig. 4.4, \*, F). There were no differences in the expression of *myod1a/1b/1c* genes in non-myotomal regions, either by comparing embryos stained with one probe, which produced either purple/blue (DIG) or red (Flu) products, or by dual labelling, where a distinct combined colour (deep red), was produced when both probes co-localised (e.g. Fig. 4.4, Bi, F).

#### 4.4.5 qPCR: primer specificity and validity of housekeeping genes

qPCR was used to investigate the relative expression patterns of the three *myod1* paralogues in fast and slow muscles during myotube<sup>+</sup> and myotube<sup>-</sup> stages of *S. salar* muscle development. To ensure that non-specific amplification of the different paralogues did not occur, several steps were taken. Firstly, primers were designed in the most distinct regions possible, particularly at the primers 3' end where binding occurs. *myod1* products were ran on 2% agarose gels, which resulted in a single product of the desired length (*myod1a*, 104 bp; *myod1b*, 172 bp; *myod1c*, 163 bp) (not shown). However, if non-specific amplification of the different paralogues had occurred, the product sizes would be highly similar and difficult to separate by electrophoresis. Therefore a dissociation protocol was run from 60-90°C. A dissociation curve shows the SYBR-green fluorescence of an amplified sample as a function of a temperature range providing a melting curve of the amplified PCR product. The T<sub>m</sub> of the product is a function not only of length, but also GC vs. AT content as well as the GC distribution and very small differences in T<sub>m</sub> between similar sized products produce distinct peaks during DNA dissociation (Ririe et al., 1997). The

T<sub>m</sub> and GC content of each *myod1* paralogue product (*1a/1b/1c*) was calculated and varied by 1-3°C (not shown). Since each *myod1* paralogue produced a unique single sharp dissociation peak it is safe to assume that each primer set amplified a single specific product.

#### 4.4.6 Quantification of *myod1* paralogue expression in salmon fast and slow muscle

Fig. 4.5 shows the mean quantitative expression of *myod1* genes in the fast and slow muscles of six juvenile (~300g) and six adult Atlantic salmon (~4.3 kg) respectively reflecting periods of active myotube production (myotube<sup>+</sup> stage) and a growth stage when new myotube production had ceased (myotube<sup>-</sup> stage). The result was near identical when data was normalised using individual (not shown) or combined housekeeping genes using the REST-384© software. There was no statistical difference (p>0.05) in the expression of any *myod1* paralogue between myotube<sup>+</sup> or myotube<sup>-</sup> stages in either fast or slow muscle. However, there was strong evidence for the differential regulation of *myod1* paralogues in muscle fibres of the fast or slow twitch phenotype. For example, in fast muscles, *myod1a* was strongly and significantly upregulated (p<0.001) by ~12/4 times relative to *myod1b/1c* in myotube<sup>-</sup> fish and by ~ 4/4 times in myotube<sup>+</sup> fish. Conversely in slow muscle fibres of myotube<sup>-</sup> fish, *myod1c* was significantly upregulated (p<0.01) relative to *myod1a/1b* by ~2.2/3.25 times and significantly upregulated (p<0.05) by 2.3/2.2 times relative to *myod1a/1b* in myotube<sup>+</sup> fish. Additionally, *myod1a* was respectively upregulated by 1.5/2.2 times relative to *myod1b* in slow muscle fibres at myotube<sup>-</sup> (significant result p<0.01) and myotube<sup>+</sup> stages (not significant). Thus, *myod1a* was strongly predominantly expressed over *myod1b/1c* in fast muscles of myotube<sup>+</sup> and myotube<sup>-</sup> fish whereas *myod1c* was the most abundant mRNA in slow muscles at both growth stages.

▼ exon 1

1 ***ATGGAGTTGT***CGGATATTTTCGTTCCCTATCACCTCCGCTGATGACTTTTATGACGACCT 20  
M E L S D I S F P I T S A D D F Y D D P

61 TGCTTCAACACCAGCGACATGCATTTTTTCGAAGACATGGACCCCGGTTAGTCCATGTT 40  
C F N T S D M H F F E D M D P R L V H V

121 GGTCTCCTCAAGCCAGACGACCATCATCATAACGAAGACGAGCACATCAGGGCCCCAAGC 60  
G L L K P D D H H H N E D E H I R A P S

181 GGGCACCACCAAGCCGGTAGGTGCCTCCTCTGGGCATGCAAAGCCTGCAAGAGGAAAAAC 80  
G H H Q A G R C L L W A **C K A C K R K T**

241 ACCAACACCGACCGGAGGAAGGCTGCTACCATGCGGGAGAGGAGGAGGCTGGGGAAGGTC 100  
**T N T D R R K A A T M R E R R R L G K V**

301 AACGACGCCTTCGAGAACCTGAAGAGATGCACGTCGAACAACCCCAATCAGAGGCTTCCA 120  
**N D A F E N L K R C T S N N P N Q R L P**

361 AAGGTGGAGATCCTGAGAAATGCCATCAGCTACATCGAGTCTCTGCAGTCTCTGCTCAGG 140  
**K V E I L R N A I S Y I E S L** Q S L L R

421 GGCCAGGACGGCGAAAACTACTACCCTTCGCTGGAGCACTACAACGGGGACTCTGACGCA 160  
G Q D G E N Y Y P S L E H Y N G D S D A

▼ exon 2

481 TCCAGCCCACGGTCCAACCTGCTCTGATGGAATGATGGAATATAATGCCCCGACGTGCACG 180  
S S P R S N C S D G M M E Y N A P T C T

▼ exon 3

541 TCCGCAAGACGAAGCAGCTATGAAAGCTCTTATTTTCGCGGAGACTCCAAATGCTGATGCC 200  
S A R R S S Y E S S Y F A E T P N A D A

601 AGAAGCAAAAAGAACGCAGTCATCTCCAGTTTGGATTGTCTATCCAGCATCGTGGAGAGA 220  
R S K K N A V I S S L D C L S S I V E R

661 ATCTCAACAGACACGTCCGCGTGCACCTATGTTATCAGTTCAGGAGGGTAGCCCCTGCTCT 240  
I S T D T S A C T M L S V Q E G S P C S

721 CCCCAGAGGGATCTATCCTGAGCGAGACAGGGGCAACCGTGCCGTACCGACCAAGTGC 260  
P Q E G S I L S E T G A T V P S P T K C

781 CCACAGCCCTCCCATGACCCCATCTACCAAGTGCTATGAA**AAAGACAGT**GATAGTATATGG 280  
P Q P S H D P I Y Q V L

841 **CTACATTCATGAAAATGAATCGATTTCATGTAAATGAAGAAAACAACGAACGGACAGGGTT**  
901 **GCAGAATGTTGCTGTACTTTAAGAATAGGATGGCAAAGTTGCCTTTCCTTCTTTAATAG**  
961 **AACGTTTTAAATCTTTGAAATTCATTCATTGGCATCCCATGGGTCACGTTTTATAAGAT**  
1021 **ATAATTTACCGATTTATTTATGTAAATGTCATCGTATATGAAATATTTGACCACTTTCTT**  
1081 **TCCCTTGCCATTTGAATTGTTGATGACAATAGAGGACATTTTGTATAGGTCTATGTG**  
1141 **AAATAAGAGGTGAAACGAAAAATGTTTTAATGTATTTAATGTTGCCTGTTTATATTCA**  
1201 **GGGTGCGTGGTTGTTGTTGTTGAGAAATGTTAACTTTATATTTATACATCTTAAAT**  
1261 **TTGTGAAAGTCGTTTCGATGTTATTTATAAATAAATGACTATGACTAGAAAAA**  
1321 **AAAAA**

Fig. 4.1. MyoD1c at the mRNA and protein level. The sense nucleotide strand is shown (number to left of figure) with codons shown above translated AAs (numbers to the right of figure). Exon-exon boundaries are identified (▼) and the translation start and stop codon are underlined in bold and italic font respectively. The 3' untranslated region is shown in red font and a polyadenylation signal (AATAAA) and poly-A tail are respectively shown in blue and green font. Also shown at the AA level are the basic region (light grey shading), HLH (dark grey shading) and helix-3 domains (underlined).

Table 4.2. Comparison of percentage sequence identity of coding and non-coding features of *myod1* co-orthologues (*myod1a/1b/1c*) and *myod1c* orthologues. The prefix's *s* and *bt* respectively indicate an Atlantic salmon or brown trout sequence. Comparisons are at the nucleotide level, except the coding sequence (marked with a star), which indicates percentage sequence identity at the AA level.

<b>Region</b>	<i>smyod1a</i> vs. <i>smyod1b</i>	<i>smyod1a</i> vs. <i>smyod1c</i>	<i>smyod1b</i> vs. <i>smyod1c</i>	<i>smyod1c</i> vs. <i>btmyod1c</i>
Exon 1	82.6	81.3	92.6	n/a
Intron 1	23.9	33.3	57.5	n/a
Exon 2	65.1	66.3	88.6	n/a
Intron 2	40.7	35.3	78.1	n/a
Exon 3	66.3	67.6	87.9	n/a
3' UTR	41.4	45.5	55.7	n/a
Coding sequence *	78.6	78.2	90.6	99.3



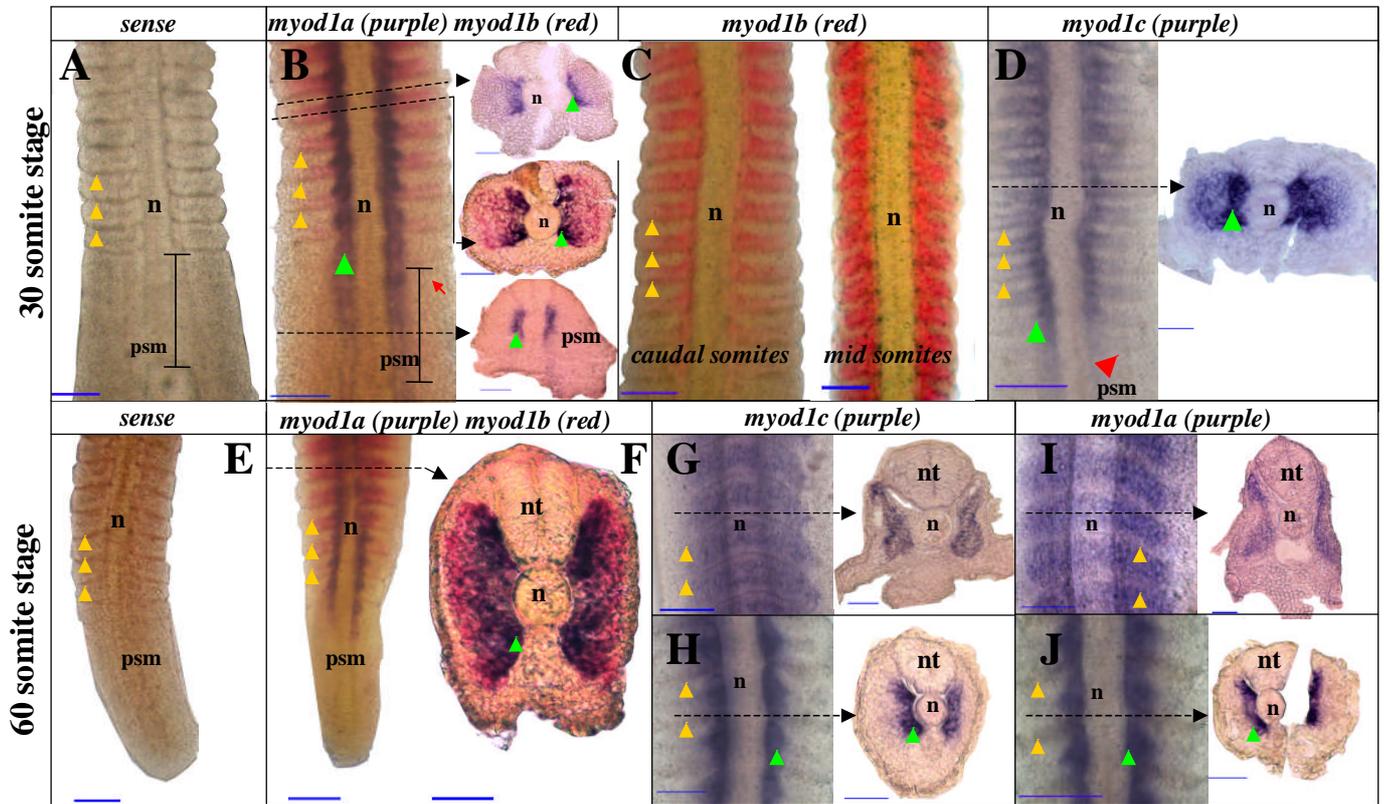


Fig. 4.3. Differential expression of *myod1* genes during Atlantic salmon somitogenesis. **A-D**. Dorsal flatmounts of ~30-somite embryos showing the most caudal somites unless otherwise indicated. **A**. Example of an unstained sense control. **B-D**. *myod1* paralogue staining. **E-J**. Flatmounts of embryos at the end of segmentation. **E**. Sense control (dorsal view of caudal somites). **F, H, J**. *myod1* staining in caudal somites (dorsal view). **G, I**. *myod1* staining in rostral somites (lateral view). Dotted lines indicate the position of somite cross sections. Yellow arrows show somite boundaries. Green arrows show adaxial cells. Red arrows show the forming furrow of the newest somite. Abbreviations: n, notochord; nt, neural tube; psm, presomitic mesoderm. Scale bars are 50µm on sections, 100µm on flatmounts.

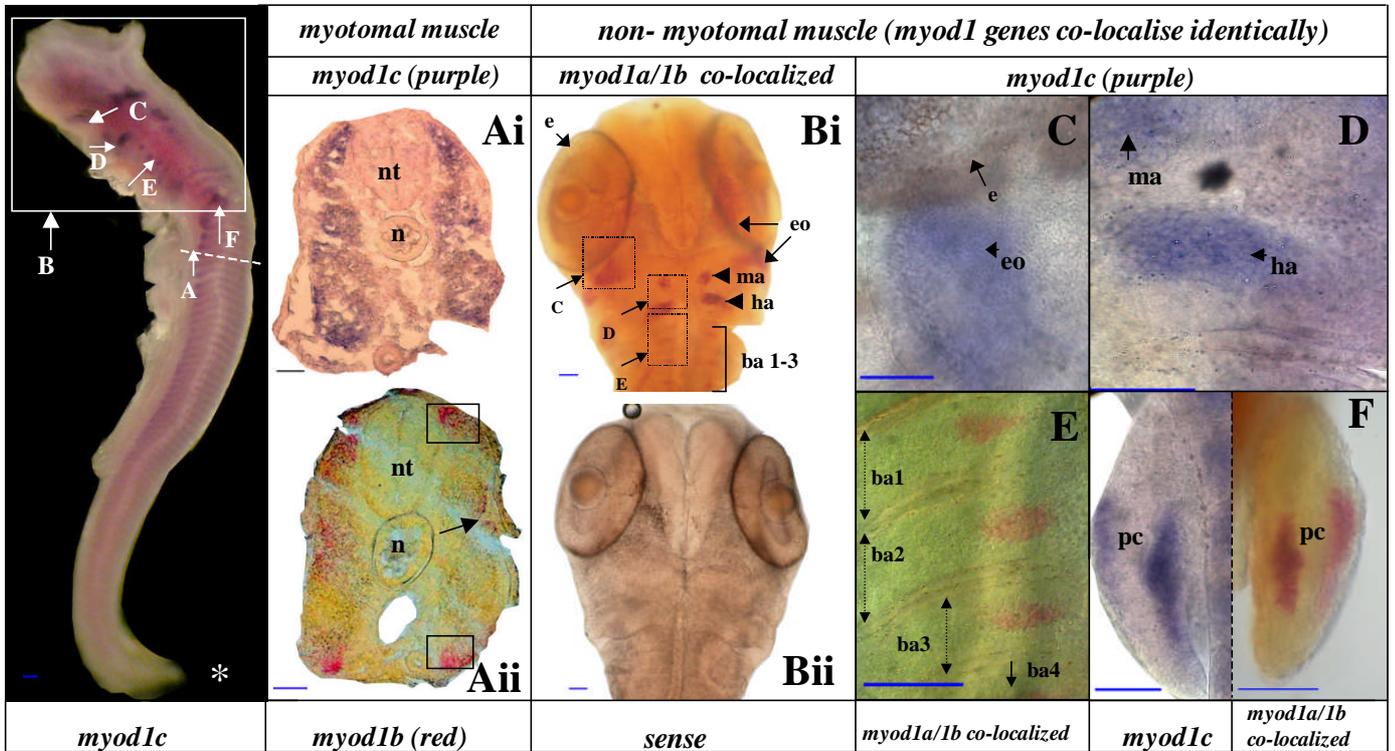


Fig. 4.4. *myod1* expression at the Atlantic salmon eyed stage. \* Darkfield image showing a whole embryo stained with *myod1c* and directing the reader to higher magnification images (white arrows). **Ai**. Myotome cross section stained with *myod1c* (*myod1a* staining was identical) **Aii**. Myotome cross section stained with *myod1b*. Boxes show *myod1b* staining in regions equivalent to germinal zones of stratified hyperplasia. **B-F**. Non-myotomal muscles where *myod1* mRNAs co-localised. **Bi**. Ventral view of embryos head region. Boxes direct reader to higher magnification images of staining in **C**, the superior rectus extraocular muscle, **D**, the hyoid and mandibular arches, **E**, the branchial arches and **F**, the fin buds. Abbreviations: n, notochord; nt, neural tube; e, eye; eo, extraocular muscle; ph, pharyngeal arches; pc, pre-cartilage core; ma, mandibular arch; ha, hyoid arch; ba, branchial arch. Scale bars are 50µm on sections, 100µm on flatmounts.

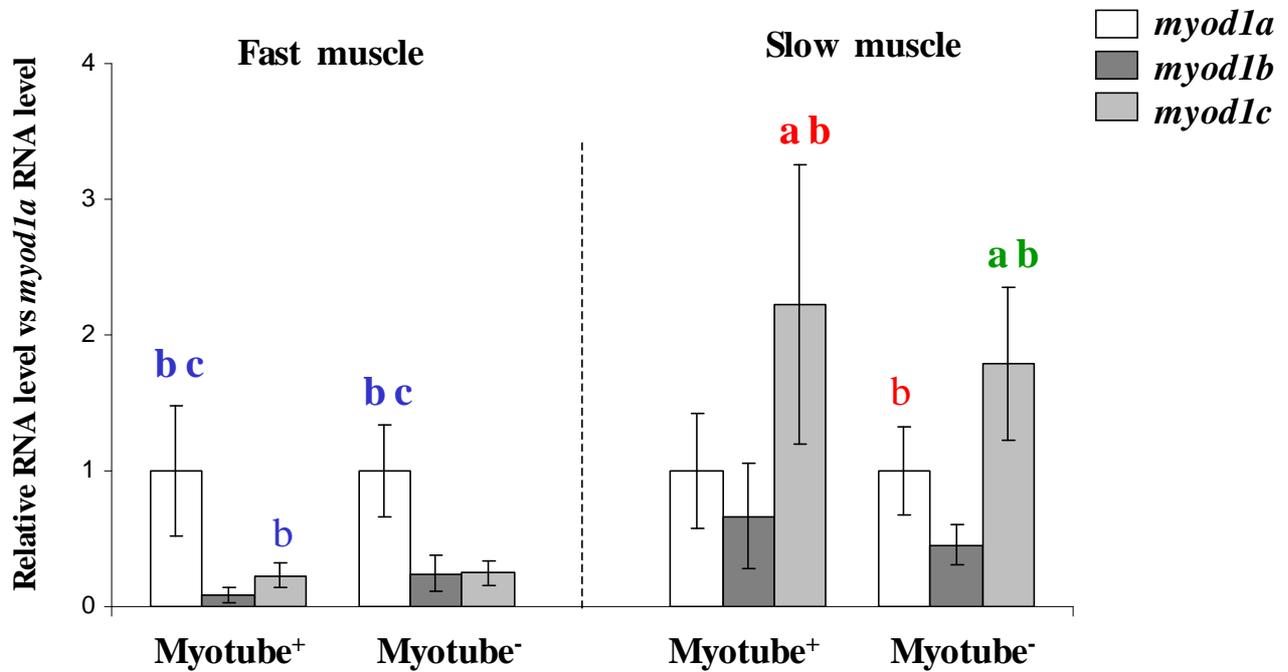


Fig. 4.5. Quantitative expression of *myod1* paralogues in adult Atlantic salmon fast and slow muscle fibres. Data are the mean expression ratios of each gene vs. mean *myod1a* expression, normalised by mean  $\beta$ -actin and *ef1- $\alpha$*  expression. Error bars show the standard error of the mean (n = 6). Letters (a, b, c) indicate that the marked *myod1* paralogue is significantly greater than: a, *myod1a*, b, *myod1b* c, *myod1c*. Blue, green and red letters respectively indicate a significant difference at the (p<0.001), (p<0.01) and (p<0.05) levels.

## 4.5 Discussion

### 4.5.1 Three paralogous *myod1* genes are present in salmonid fish

Several lines of evidence suggest that *myod1c* is not an allelic variant of *myod1b/1a*, but is in fact a paralogue. For example, there is a large disparity in sequence identity between *myod1c* compared to *myod1a/myod1b* in coding regions (~78/90% respectively) that is attenuated in intronic and untranslated regions (~33-46%/55-78% respectively). If *myod1c* was a conserved allele of *myod1b/1a*, then less sequence divergence would be expected over the whole gene, and since coding and regulatory (non-coding) regions would be under similar selective constraints, differences in sequence identity would not be exaggerated in non-coding regions. Additionally, *myod1c* is highly conserved in three salmonid species, which would not be expected if *myod1c* was an allele of *myod1a/1b*. The most likely explanation for the presence of three salmonid *myod1* genes is through gene/genome duplication. The teleost WGD occurred around 320-350 Mya (Van de Peer, 2004) and later the salmonid genome was duplicated (25-100 Mya) (Allendorf and Thorgaard, 1984). This means when two teleost paralogues are conserved as a consequence of the WGD, four copies can potentially be conserved in salmonids. For example, the *mstn* gene is present as two paralogues in most non-salmonid teleosts (*mstn1* and *mstn2*) (Kerr et al., 2005) and salmonids have two copies of each gene i.e. *mstn1a/1b* and *mstn2a/2b* (Garikipati et al., 2007).

During the WGD, a *myod* gene duplicated in teleosts to produce *myod1* (universally conserved in teleosts) and *myod2* (retained specifically in the Acanthopterygii) (chapter 3). Further, phylogenetic reconstruction showed that each of the three salmon *myod* paralogues were co-orthologues of the teleost *myod1* gene and thus originated as a consequence of two salmonid specific events in the *myod1* lineage post WGD. The salmonid tetraploidization almost certainly accounts for the presence of two of the paralogues, but a further duplication is necessary to

explain the presence of the third paralogue. Comparing salmon *myod1* paralogues at the genomic or coding level, it is clear that *myod1b/1c* are more closely related to each other than to *myod1a*. For example in the coding sequence they share about 91% identity to each other and ~78-79% with *myod1a*. This pattern was confirmed by phylogenetic analysis that placed MyoD1b/1c as sister sequences branching from MyoD1a (chapter 3). It was originally suggested that *myod1a/1b* arose during the salmonid tetraploidization (Rescan and Gauvry, 1996), but this was based on the knowledge of two, rather than three genes. Currently, it cannot be confidently distinguished whether *myod1a/1b* or *myod1b/1c* arose as a consequence of the genome tetraploidization (see Fig. 4.6). However, if *myod1* duplicated in a salmonid ancestor before the tetraploidization, then four paralogues would be initially present subsequent to the tetraploidization (Fig. 4.6, scenario 1). This would require that one paralogue has been lost or is currently undiscovered (Fig. 4.6 scenario 1). The absence of an uncharacterised paralogue in EST libraries, where *myod* sequences are well represented, suggests that the former (non-functionalization) is more likely. A more parsimonious explanation, without the requirement of gene loss, is that *myod1* duplicated during the genome tetraploidization to produce *myod1a* and the ancestor gene to *myod1b/myod1c* which then duplicated again independently to produce the current *myod1b* and *myod1c* genes (Fig. 4.6, scenario 2). Two *myod* genes were shown to map to two distinct duplicated linkage groups in brook trout (*Salvelinus fontinalis*) (Gharbi et al. 2006), but were identified only as *MYODi* and *MYODii* so it was unclear which of the three paralogues were identified. It would be interesting to map the genomic location of all *myod* paralogues in salmon to establish the extent of conserved synteny between regions containing the different paralogues and to assess the scale of the two *myod* duplication events.

#### 4.5.2 Comparison of salmon MyoD1 paralogues

The Atlantic salmon MyoD1 paralogues are well conserved throughout the protein but particularly in the basic, HLH, cis/his rich, Helix-3 and activation domains (Fig. 4.2). These

domains are essential for the normal function of MyoD as a muscle specific transcription factor (Tapscott, 2005), suggesting that MyoD1a/1b/1c proteins probably perform a similar general function. It is notable, that several unique AA substitutions are present at sites in the HLH of salmon that are conserved in other vertebrates, including zebrafish (Fig. 4.2). However it is unknown whether these mutations will affect the dimerisation capacity of salmon MyoD1 proteins. Additionally, the helix-3 of salmon MyoD1a has two substitutions that are not present in the equivalent region of other MyoD1 paralogues or vertebrate MyoD proteins. The helix-3 region accounts for the difference in capacity of the different MRFs for myogenic specification versus differentiation (Bergstrom and Tapscott, 2001). A MyoD-chimera containing the helix-3 of Mrf4, which varies by six residues compared to the helix-3 of MyoD, could initiate the *in vitro* expression of endogenous muscle genes as efficiently as wild-type MyoD. In light of these results, it is unlikely that two substitutions will strongly contribute to differences in specification vs. differentiation function of salmon *myod1* paralogues. It is also notable, that between the HLH and helix-3 exists a highly conserved motif of around 20 residues that is ser-rich (6/20 amino acids) that has currently been assigned no function.

However, at all other regions C-terminal to the bHLH (except the conserved motif of unknown function), salmonid MyoD1 proteins are less conserved, particularly comparing MyoD1a to MyoD1b/1c (Fig. 4.2). Interestingly, Ishibashi et al. (2005) showed that whilst MyoD and Myf5 share many target genes, the former is greatly more efficient at inducing those involved in myogenic differentiation. This was accounted for by an interaction between the region of MyoD C-terminal to the bHLH and the N-terminal activation domain, which is lacking in the Myf5 protein (Ishibashi et al., 2005). Thus it is possible that the differences in these regions in salmonid MyoD1 proteins could affect their efficiency at initiating muscle genes involved in specification or differentiation.

#### 4.5.3 Atlantic salmon *myod1* genes and the differentiation of embryonic fast and slow muscle

The spatiotemporal expression patterns of salmon *myod1* paralogues was coincident with events leading to the differentiation of embryonic muscle fibres with different phenotypes. *myod1b/1c* staining in the posterior-lateral domain of the epithelial somite was reminiscent of an equivalent portion of the zebrafish *myod1* expression domain (Weinberg et al., 1996; Groves et al., 2005). These zebrafish posterior cells differentiate into the embryonic fast-twitch myotome (Stellabotte et al., 2007; Hollway et al., 2007) a process regulated by retinoic acid signalling and mediated through Fgf8 (Hamade et al., 2006; Groves et al., 2005). *myod1a/1c* transcripts were strongly expressed in the medially located adaxial cells, before this domain spread and migrated laterally in maturing somites from the 30 ss This is coincident with the medial-lateral migration of *smlc1*, which marks differentiating adaxial cells in rainbow trout and Atlantic salmon (Chauvigné et al., 2005, chapter 7). From the end of segmentation onwards, several genes involved in slow and fast muscle fibre differentiation became repressed in the inner and superficial trout myotome respectively (Chauvigné et al., 2005). At this time, *myod1b* was repressed in the inner myotome of rostral somites, but was detected at the lateral edge of the myotome, possibly in the single layer of slow fibres as previously suggested for trout (Delalande and Rescan, 1999), or alternately, marking the newly differentiating fast muscle fibres from the external cell layer as suggested for *myf5* in chapter 7 (Fig. 7.6). Conversely, in rostral somites at the end of segmentation, *myod1a/1c* transcripts were maintained in the inner fast myotome, before being gradually downregulated to zones of stratified myotube production, as embryos developed further (see chapter 7, Fig. 7.6). Thus the data presented here is consistent with a distinct embryonic role for *myod1* paralogues in the specification/differentiation of different salmon muscle fibre types. Additionally, the different *myod1* paralogues may differentially mark the level of differentiation of myoblasts originating from the external cell layer. These differences in expression pattern are almost certainly a reflection of the independent accumulation of mutations

in distinct regulatory regions of each paralogue (section 4.5.5). However, each salmon *myod1* paralogue was also expressed concurrently in non-somite regions of muscle development i.e. in embryonic extraocular, pharyngeal and fin muscles (Fig. 4.4). This suggests that the regulatory elements governing these expression fields have probably been conserved among the different *myod1* promoters. It is possible that a dose-dependent selective advantage is present for the transcription of all *myod1* genes in these regions.

#### 4.5.4 Differential expression of *myod1* paralogues in adult muscle

In 300-500g rainbow trout, *myod1a* mRNA was shown by northern blotting to be abundant in fast and slow muscle, but *myod1b* was only detected in slow muscle (Delalande and Rescan, 1999). Here, qPCR was used in adult Atlantic salmon to show that each *myod1* gene was expressed to a greater or lesser extent in both fibre types. Consistent with previous findings *myod1a* was upregulated many times in fast muscle relative to *myod1b/1c* in both juvenile (myotube<sup>+</sup>) and large adult (myotube<sup>-</sup>) fish. However, *myod1c* was the predominant paralogue expressed in slow muscle at both growth stages. This difference may reflect the higher sensitivity of the qPCR approach used here, but it is also possible that *myod1c* and *myod1b* mRNA (which are >90% similar in the coding sequence) both hybridised to the *myod1b* probe used previously (Delalande and Rescan, 1999). As for embryonic stages, these results are consistent with the differential expression of *myod1* paralogues playing a role in the specification of fast vs. slow muscle fibre types in adult fish. It has been observed at the mRNA level, that *myod* and *myog* accumulate in comparatively higher amounts in mammalian fast and slow muscle respectively (Hughes et al., 1993). Thus, the differential regulation of distinct muscle fibre phenotypes, by the expression of different MRF genes may be a common feature of vertebrate myogenesis. Conversely, the alternate splicing of the pufferfish *myod1* gene was not implicated in the regulation of teleost myogenesis by fibre phenotype (Fernandes et al., 2007). Additionally, while there was no evidence for the differential contribution of *myod1* paralogues to the myotube<sup>-</sup> and

myotube<sup>+</sup> phenotype, there was evidence for different ratios of paralogues in fast and slow muscles at different growth stages. For example, while *myod1a* was present at 12 times the abundance of its paralogues in juvenile fast muscles, in adult muscles it was only ~4 times as abundant. Interestingly, Fernandes et al. (2007) showed that in *T. rubripes* two splice variants of *myod1*, denoted *myod1a* and *myod1γ* were respectively significantly upregulated and downregulated in myotube<sup>+</sup> fish compared to myotube<sup>-</sup> fish. Thus the *myod* gene of teleosts seems to regulate myogenesis at multiple levels.

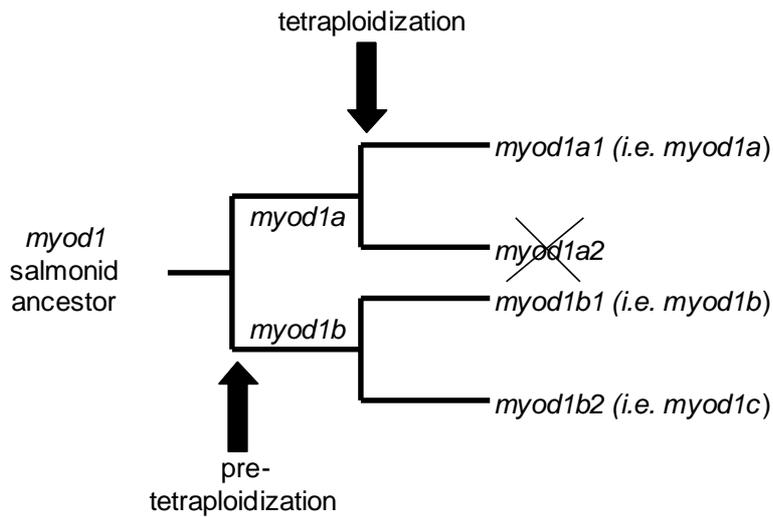
#### 4.5.5 The stabilization of *myod1* paralogues in the salmonid genome

The dual expression pattern of *myod1a/1b* in Atlantic salmon and rainbow trout embryos (Delalande and Rescan, 1999) is clearly reminiscent of the single *myod1* gene in zebrafish (Weinberg et al., 1996). It has been suggested that the stabilization of *myod1a/1b* in the salmonid genome is consistent with the DDC/subfunctionalization model (Rescan, 2001). Interestingly, while the expression pattern of salmon *myod1c* is distinct from *myod1a/1b*, it does not spatially encompass novel regions in the embryo, and in fact has characteristics of both its paralogues. Fig. 4.7 shows a theoretical scenario to account for the preservation of *myod1c* within the framework of the DDC model (Force et al., 1999) (following Scenario 2 of Fig. 4.6). It should be noted that the actual evolution of salmonid *myod1* genes was probably far more complex, and that the subfunctions described here, while feasible, are unlikely to represent the whole regulatory intricacy accounting for the expression of the ancestral *myod1* gene. The ‘function’ of the gene ancestral to salmonid *myod1* paralogues can be thought of as the sum of the regulatory elements governing its somitic expression pattern. This can be split into four categories based on distinct spatiotemporal expression domains of *myod1a/1b/1c* in the salmon embryo (Fig. 4.7, A). After the first round of *myod1* duplication, *myod1a* and the ancestor gene to *myod1b/1c*, would initially share identical coding and regulatory features (Fig. 4.7, B). Subsequently, *myod1a* accumulated degenerative mutations in the regulatory elements governing expression fields in

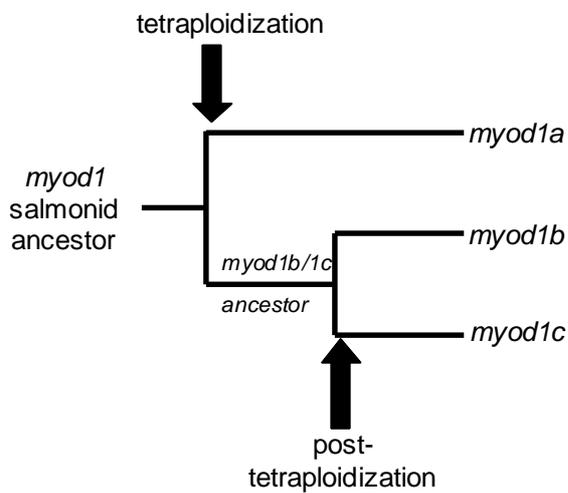
the posterior somite and at the myotomes periphery at the end of segmentation (respective subfunctions 3 and 4) (Fig. 4.7, C). During this time, the *myod1b/1c* ancestor accumulated degenerative mutations within the sequence features governing adaxial cell expression in adaxial myoblasts of the presomitic mesoderm (subfunction 1). Thus, at this stage, both genes would be required to preserve the function of the ancestral *myod1* gene (Fig. 4.7 D). When *myod1b/1c* later arose they too would have initially shared an identical complement of regulatory elements (Fig. 4.7, E). Following the duplication of the *myod1b/1c* ancestor gene (Fig. 4.7, E), the current observed expression profile of *myod1b* and *1c* would then be consistent with the respective loss of subfunctions 2 and 4 respectively (Fig. 4.7, F).

#### 4.5.6 Chapter conclusion

In this chapter a novel *myod* gene was characterised in Atlantic salmon that is also conserved in brown and rainbow trout. This gene was named *myod1c* following the consensus nomenclature proposal of chapter 3. *myod1c* had a unique embryonic expression compared to its paralogues, *myod1b/myod1a*, although it did not encompass novel expression domains and together the paralogues recapitulated the expression of the single *myod1* gene of zebrafish. Further, in adult fish, *myod1c* was upregulated in slow-twitch myotomal muscle relative to its paralogues, and *myod1a* was strongly predominant in fast muscle. Taken together, these results are consistent with the differential regulation of the different *myod1* paralogues playing a role in the specification/differentiation of muscle fibres of distinct phenotypes. Further, it is likely that these effects are under the control of independently mutable regulatory regions that have subfunctionalized the role of the ancestral *myod1* gene.



## Scenario 1



## Scenario 2

Fig. 4.6. Two scenarios could account for three salmonid *myod* paralogues conserved from the *myod1* lineage. Scenario 1 requires that prior to the genome tetraploidization, *myod1* duplicated in a salmonid ancestor to produce two *myod1* genes (*1a/1b*) that went on to duplicate again during the genome tetraploidization. To account for the current finding of three *myod1* paralogues, with two being clearly more related (*myod1b/1c*) it is then necessary for a paralogue to be non-functionalized in the *myod1a* lineage. Scenario 2 is more parsimonious, in that it does not require gene loss. In this case a *myod1* gene duplicated in a salmonid ancestor to produce the *myod1a* gene and an ancestor gene to *myod1b/1c*. Subsequently this gene duplicated to produce *myod1b* and *myod1c* genes. In both scenarios, the scale of the *myod1* duplication external to the tetraploidization event must not be genome-wide to account for the fact that salmonid genes are generally found as duplicates relative to normal teleosts and not as four copies.

# **Embryonic temperature and the genes regulating myogenesis in teleosts**

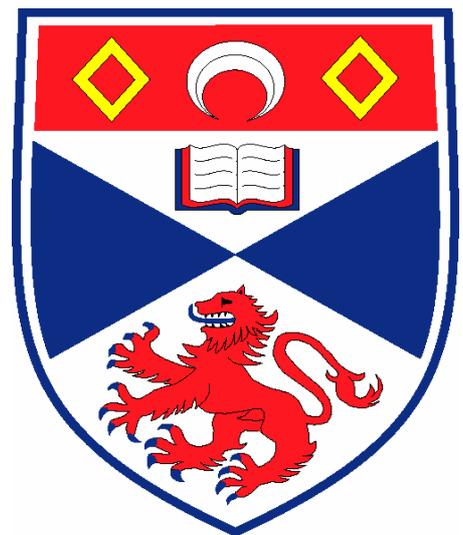
Submitted to the University of St Andrews for the degree of Doctor of Philosophy

by

**Daniel John Macqueen**

## **Project Supervisor**

Professor Ian Alistair Johnston,  
School of Biology,  
Gatty Marine Laboratory,  
University of St Andrews,  
St Andrews,  
Fife,  
KY16 8LB



**January 2008**

## *Declarations*

- i. I, Daniel John Macqueen, hereby certify that this thesis, which is approximately 65,000 words in length, was written by me, is a record of work performed by me and has not been submitted in any previous application for a higher degree.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

- ii. I was admitted as a research student in October 2004 and as a candidate for the degree of Doctor of Philosophy in October 2005; the higher study, for which this thesis is a record, was performed in the University of St Andrews from 2004-2007.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

- iii. I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulation appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date \_\_\_\_\_ 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 in accordance with the regulation of the university library for the time being in force, subject to any copyright vested in the work not being affected thereby. I understand that the title and abstract will be published, and that a copy of the work may be supplied to any *bona fide* library or research worker.

Date \_\_\_\_\_ Signature of candidate \_\_\_\_\_

## Acknowledgments

I would firstly like to thank my supervisor, Professor Ian Johnston for his ongoing support during my project. Ian has kept me on track, while letting me be creative and has helped me exponentially develop my writing skills. Additionally, I acknowledge the financial support of the Natural Environment Research Council who funded my studentship and tuition fees (ref NERC/S/A/2004/12435). As a CASE student, I received additional funding from EWOS Innovation.

Other past and present work colleagues from the Fish Muscle Research Group that I wish to thank are as follows: Dr Jorge Fernandes (now associate Professor at the university college of Bodø, Norway), not only for his incredible support throughout my PhD, but also for always upholding his role as the lab's central post-doc (and ethanol consumer); Dr Neil Bower (present), for sharing ideas about genes and science-stuff, for work-relieving chats about football and for being a fine drinking buddy; Hung-Tai Lee (present), for providing zebrafish samples and for interesting discussions on the teleost dermomyotome and the origin of stratified hyperplasia!; Tom Ashton (present) for kindly measuring the texture of some mislabelled and heavily oxidised salmon muscles and for taking forward the burden of the salmon calpastatin gene; Dr Matt Mackenzie (past) for supplying me with pufferfish samples, for co-inventing the game 'sweat-boy' and for providing a true comedy moment by jumping fully clothed into the sea off of Barcelona harbour with mobile phone and wallet in tow; Shelby Steele and Ørjan Hagen (both past) for being good pals and for a couple of enjoyable late nights; Hannah Chubb (past) for being a really nice person; Dr Sonia Consuegra (past) for putting up with my first PCR attempts and for shouting at me; Dr Vera Johnston (Vieira) (present) for always being friendly, helpful and thoughtful and Marguerite Abercromby (past) for her help with learning to cut cryosections and for always being in a good mood.

I would like to acknowledge EWOS Innovation for supporting the whole embryonic temperature experiment and for providing all Atlantic salmon material used in this thesis. I would specifically like to whole-heartedly thank the following EWOS Innovation associates: Dr Dave Robb for setting up and designing the project, for his continuous help with sampling and for drinking beer and wine with me on many a cold Lonnigdahl night. Also, thanks Dave for not getting mad when I forgot my Passport before a flight to Norway!; Tom Olsen for being a great guy and for running

the whole trial day to day for over three years; Linda Melstveit for her help with sampling and for keeping me and Dave on our toes; Liv Oma for help with sampling; Frode Ormhaug for husbandry of fish during freshwater stages; Aage Melstveit (and Linda again) for accommodation and for letting me watch Liverpool win the Champions league when no other T.V. was in sight for miles around as well as Viv Crampton and Dr Louise Buttle for helping to establish and design the project.

I couldn't have got through this PhD (or life generally) in one piece, without my wife Ruth Macqueen, who I would like to thank for being my best friend and for all the things only she knows she does for me. I would also like to thank my Mum and Dad for everything they have done for me in my life's journey. My Mother, Marilyn has been a rock in my life for as long as I can remember and I can't thank her enough for her endless support through all things I have done, good or bad. I thank my Dad, Jonathan, for inadvertently installing a mighty work ethic in me, despite his distance. I would also like to thank my parent in-laws Catherine and Stephen for their support in recent times. Other members of my family who I can't thank enough for always being there are my Sister, Samantha (and her other half Pricey [Princey]), my Aunties/surrogate mothers Lorraine and Karen, my cousins and pseudo-siblings Steph and Chris (the boy), my Nan, Doreen and Grandad, Jack, the beautiful collies Holly and Tess, and to those others whom I haven't mentioned or have moved on to greener pastures.

## Peer reviewed publications and author contributions

1. Macqueen, D.J. and Johnston, I.A. 2006. A novel salmonid myoD gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization. *FEBS Letters* **580**: 4996-5002.

I performed all experiments and conceived and wrote the paper together with Ian Johnston.

2. Macqueen, D.J., Robb, D., Johnston, I.A. 2007. Temperature influences the co-ordinated expression of myogenic regulatory factors during embryonic myogenesis in Atlantic salmon (*Salmo salar* L.). *Journal of Experimental Biology* **210**: 2781-2794.

Ian Johnston, Dave Robb and myself conceived the experiment. Dave Robb managed the sampling of Atlantic salmon embryos. I performed all experimental procedures and wrote the paper with Ian Johnston.

3. Macqueen, D.J. and Johnston, I.A. Evolution of follistatin in teleosts revealed through phylogenetic, genomic and expression analyses. *Development Genes and Evolution*: **218**: 1-14.

I performed all experiments and conceived and wrote the paper together with Ian Johnston.

4. Fernandes, J.M.O.<sup>1</sup> Macqueen, D.J.<sup>1</sup>, Lee, H-T and Johnston, I.A. 2007. Genomic, evolutionary and expression analyses of *cee*, an ancient gene involved in normal growth and development *Genomics: E pub before print*. DOI: 10.1016/j.ygeno.2007.10.017

<sup>1</sup> Denotes equal contribution.

\*Jorge Fernandes, Ian Johnston and myself conceived the paper. Jorge Fernandes and myself were the principal experimental contributors. Jorge Fernandes, Ian Johnston and myself wrote

the paper. Individual contributions are acknowledged in the materials and methods section of chapter 6.

5. Macqueen, D.J. and Johnston, I.A. An update on MyoD evolution in teleosts and a proposed consensus nomenclature to accommodate the tetraploidization of different vertebrate genomes. *PLoS ONE* **3**: e1567 doi:10.1371/journal.pone.0001567.

I performed all experiments and conceived and wrote the paper together with Ian Johnston

6. Macqueen, D.J., Robb, D.H.F., Olsen, T., Melstveit, L., Paxton, C. and Johnston, I.A. Temperature until the “eyed stage” of embryogenesis programs the growth trajectory and muscle phenotype of adult Atlantic salmon. Accepted in *Biology Letters*. In Press.

Ian Johnston and Dave Robb conceived the embryonic temperature experiment. Tom Olsen managed the fish throughout the whole experiment. Tom Olsen, Linda Melstveit, Dave Robb and myself sampled the fish. I performed all experimental and statistical procedures relating to establishing adult muscle growth traits, except for modelling the growth data, which was done by Charles Paxton and the probability density function analysis, which was performed by Ian Johnston and myself. I wrote the paper jointly with Ian Johnston.

## Table of Contents

	Page
<b>Abstract</b>	1-2
<b>Chapter 1. General introduction</b>	3-53
<i>1.1 Teleost fish: biodiversity and use as model organisms</i>	3
<i>1.2 Teleost muscle as a source of food</i>	4
<i>1.3 A focus on Atlantic salmon</i>	5
<i>1.4 Why study teleost muscles?</i>	6
<i>1.5 Gene and genome duplication and teleosts</i>	7
<i>1.5.1 Genome duplication: a historical perspective</i>	7
<i>1.5.2 Genome duplication who and when?</i>	8
<i>1.5.3 Genome duplication and the fate of paralogues</i>	10
<b>1.6 Myogenesis</b>	11
<i>1.6.1 Common features of myogenesis across vertebrates</i>	11
<i>1.6.2 Teleost myotomal muscle: patterns and innovations</i>	12
<i>1.6.3 Myogenic regulatory factors: the master switches for muscle specific gene transcription</i>	18
<i>1.6.4 Lessons from knockout studies of the MRFs</i>	19
<i>1.6.5 The MEF-2 transcription factors co-regulate differentiation with MRFs</i>	24
<i>1.6.6 Paired box transcription factors and satellite cells</i>	24
<i>1.6.7 Myostatin: the most potent known negative regulator of myogenesis</i>	26
<i>1.6.8 Follistatin a regulator of Mstn and of myogenesis</i>	28
<i>1.6.9 Other known regulators of Mstn</i>	29

<i>1.6.10 The Insulin like growth factors, Calcineurin and the Akt/mTOR pathway</i>	30
<i>1.6.11 Forkhead box proteins and satellite cells</i>	33
<i>1.6.12 Sox8, another marker of adult satellite cells</i>	34
<i>1.6.13 The Calpain- Calpastatin system</i>	34
<i>1.6.14 MicroRNAs and myogenesis</i>	36
<b>1.7 Phases of teleost myogenesis</b>	37
<i>1.7.1 Embryonic myogenesis: early patterns and onset of slow muscle formation</i>	37
<i>1.7.2 Slow muscle specification and hedgehog morphogens</i>	38
<i>1.7.3 Formation of embryonic fast muscle</i>	39
<i>1.7.4 The anterior somite of teleosts and continuing muscle growth</i>	40
<i>1.7.5 Stratified hyperplasia</i>	42
<i>1.7.6 Mosaic hyperplasia and maximum fibre diameter</i>	42
<i>1.7.7 What is the source of muscle growth during mosaic hyperplasia?</i>	44
<i>1.7.8 Signals regulating myotube production</i>	44
<b>1.8 Myogenesis and the environment</b>	47
<i>1.8.1 Introduction</i>	47
<i>1.8.2 Embryonic temperature and myogenesis</i>	48
<i>1.8.3 Lasting effects of early rearing temperature on myogenesis</i>	50
<b>1.9 Commercial relevance of plasticity of teleost myogenesis</b>	51
<b>1.10 Aims and goals</b>	52
<b>Chapter 2. Materials and methods</b>	54-69
<b>2.1 Introduction</b>	54
<b>2.2 Embryonic temperature experiment</b>	54

2.2.1 <i>Experimental set-up</i>	54
2.2.2 <i>Embryonic sampling</i>	55
2.2.3 <i>Post embryonic sampling</i>	56
<b>2.3 <i>Basic computational biology methods</i></b>	<b>57</b>
2.3.1 <i>Gene databases</i>	57
2.3.2 <i>DNA manipulation in silico</i>	58
2.3.3 <i>Primer design</i>	58
2.3.4 <i>Predication of gene structure</i>	59
2.3.5 <i>Sequence submission</i>	59
<b>2.4 <i>Basic experimental molecular biology</i></b>	<b>59</b>
2.4.1 <i>General working conditions</i>	59
2.4.2 <i>Extraction of DNA and RNA</i>	60
2.4.3 <i>Gel Electrophoresis</i>	60
2.4.4 <i>Quantification and quality assessment of RNA</i>	61
2.4.5 <i>Gel purification of DNA</i>	62
2.4.6 <i>Complementary DNA (cDNA) synthesis</i>	62
2.4.7 <i>Polymerase chain replication (PCR) and Reverse Transcription-PCR (RT-PCR)</i>	62
2.4.8 <i>Rapid Amplification of Complementary Ends (RACE) PCR</i>	63
2.4.9 <i>DNA cloning</i>	63
2.4.10 <i>Plasmid purification, digestion and screening</i>	64
2.4.11 <i>Sequencing of DNA</i>	65
<b>2.5 <i>Quantitative PCR (qPCR)</i></b>	<b>65</b>
2.5.1 <i>Experimental procedure</i>	65
2.5.2 <i>Dissociation analysis and primer specificity</i>	66
<b>2.6 <i>In situ hybridisation</i></b>	<b>66</b>

2.6.1	<i>Introduction</i>	66
2.6.2	<i>Amplification of probe templates</i>	66
2.6.3	<i>Probe transcription</i>	67
2.6.4	<i>Embryo preparation and fixation</i>	67
2.6.5	<i>Probe hybridisation and detection</i>	68
2.6.6	<i>Embryo study, cryosectioning and photography</i>	68
<b>Chapter 3. The evolutionary relationships of teleost myod genes revealed through comparative phylogenetic and genomic analyses</b>		70-101
3.1	<b><i>Abstract</i></b>	70
3.2	<b><i>Introduction</i></b>	71
3.3	<b><i>Materials and methods</i></b>	72
3.3.1	<i>In silico mining and cloning of myod2</i>	72
3.3.2	<i>Testing the selective constraints across MyoD1/2 proteins</i>	73
3.3.3	<i>Sequence retrieval for phylogenetic reconstruction</i>	73
3.3.4	<i>Phylogenetic reconstruction of myod genes</i>	74
3.3.5	<i>Synteny analysis of teleost myod genes</i>	76
3.3.6	<i>Phylogenetic reconstruction of myod-neighbouring genes</i>	76
3.4	<b><i>Results</i></b>	78
3.4.1	<i>In silico and experimental characterisation of teleost myod2 genes</i>	79
3.4.2	<i>Evolutionary constraints on Acanthopterygian MyoD1 and MyoD2 proteins</i>	79
3.4.3	<i>Phylogenetic reconstruction of teleost myod genes within a MRF framework</i>	79
3.4.4	<i>Further phylogenetic reconstruction of MyoD</i>	81
3.4.5	<i>Genomic neighbourhood surrounding myod genes</i>	82

3.4.6 <i>Phylogenetic reconstruction of myod-neighbouring genes</i>	83
<b>3.5 Discussion</b>	95
3.5.1 <i>MyoD2 is specifically conserved in the Acanthopterygii and is evolving rapidly relative to its paralogue</i>	95
3.5.2 <i>MyoD duplications and vertebrate polyploidy</i>	96
3.5.3 <i>A comparative genomic study of myod neighbouring genes reveals the true extent of the MyoD1/MyoD2 duplication</i>	97
3.5.4 <i>Chapter conclusion: a consensus nomenclature for vertebrate MyoD sequences</i>	99
<b>Chapter 4. Characterisation of a novel differentially expressed Atlantic salmon MyoD paralogue</b>	102-129
<b>4.1 Abstract</b>	102
<b>4.2 Introduction</b>	103
<b>4.3 Materials and methods</b>	104
4.3.1 <i>Fish sampling</i>	104
4.3.2 <i>Computational approaches and sequence retrieval</i>	104
4.3.3 <i>Cloning Atlantic salmon myoD1c</i>	105
4.3.4 <i>Genomic characterisation of myoD1c</i>	105
4.3.5 <i>In situ hybridisation</i>	106
4.3.6 <i>Quantitative RT-PCR (qPCR)</i>	106
<b>4.4 Results</b>	109
4.4.1 <i>Sequence characterisation of myoD1c</i>	109
4.4.2 <i>Characterisation of non-coding regions of myoD1c</i>	110
4.4.3 <i>Comparison of salmon MyoD1 paralogues with MyoD1 orthologues</i>	111

4.4.4 <i>Expression of myod1 paralogues during salmon embryogenesis</i>	111
4.4.5 <i>qPCR: primer specificity and validity of housekeeping genes</i>	113
4.4.6 <i>Quantification of myod1 paralogue expression in salmon fast and slow muscle</i>	114
<b>4.5 Discussion</b>	121
4.5.1 <i>Three paralogous myod1 genes are present in salmonid fish</i>	121
4.5.2 <i>Comparison of salmon MyoD1 paralogues</i>	122
4.5.3 <i>Atlantic salmon myod1 genes and the differentiation of embryonic fast and slow muscle</i>	124
4.5.4 <i>Differential expression of myod1 paralogues in adult muscle</i>	125
4.5.5 <i>The stabilization of myod1 paralogues in the salmonid genome</i>	126
4.5.6 <i>Chapter conclusion</i>	127
<b>Chapter 5. Evolution of Follistatin in teleosts revealed through phylogenetic, genomic and expression analyses: a role for fst1 in teleost myogenesis</b>	130-159
<b>5.1 Abstract</b>	130
<b>5.2 Introduction</b>	131
<b>5.3 Materials and methods</b>	132
5.3.1 <i>Sequence retrieval and genomic analyses</i>	132
5.3.2 <i>Construction of synteny diagram</i>	133
5.3.3 <i>Phylogenetic and evolutionary analyses</i>	134
5.3.4 <i>Testing the selective constraints across Fst proteins</i>	135
5.3.5 <i>Tissue specific mRNA expression of teleost fst genes</i>	135
5.3.6 <i>Cloning of salmon fst1</i>	136
5.3.7 <i>Embryos and whole-mount in situ hybridization</i>	137
<b>5.4 Results</b>	137

5.4.1	<i>Two fst paralogues are present in the Ostariophysi</i>	137
5.4.2	<i>Teleost fst genes are present on duplicated chromosomes with double conserved synteny relative to tetrapods</i>	138
5.4.3	<i>Asymmetric evolution of teleost Fst paralogues</i>	140
5.4.4	<i>Distinct mRNA tissue distribution of fst1/2 paralogues</i>	141
5.4.5	<i>Comparative promoter analysis of vertebrate fst genes</i>	141
5.4.6	<i>Expression of fst1 during embryonic myogenesis in salmon</i>	142
5.4.7	<i>fst1 expression in myogenic progenitors of the pectoral fin buds</i>	143
<b>5.5</b>	<b><i>Discussion</i></b>	<b>155</b>
5.5.1	<i>Two Fst genes were duplicated in a common teleost ancestor</i>	155
5.5.2	<i>Positive selection of Fst2 relative to other Fst proteins?</i>	155
5.5.3	<i>Differential regulation of fst1/fst2 paralogues suggests distinct evolution of regulatory regions</i>	156
5.5.4	<i>Has subfunctionalization of fst2 contributed to its retention in the Ostariophysi?</i>	157
5.5.5	<i>A conserved role for vertebrate fst genes in regulating myogenesis?</i>	158
5.5.6	<i>Chapter conclusion</i>	159
<b>Chapter 6.</b>	<b>Genomic, evolutionary and expression analyses of cee, an ancient and novel gene involved in normal growth and development</b>	<b>160-186</b>
<b>6.1</b>	<b><i>Abstract</i></b>	<b>160</b>
<b>6.2</b>	<b><i>Introduction</i></b>	<b>160</b>
<b>6.3</b>	<b><i>Materials and methods</i></b>	<b>159</b>
6.3.1	<i>Animals and sample collection</i>	162
6.3.2	<i>Computational identification of cee orthologues</i>	162
6.3.3	<i>Cloning and sequencing of cee cDNAs</i>	163

6.3.4	<i>Sequence alignments and intron-exon structure of cee</i>	163
6.3.5	<i>Phylogenetic inference and tests of selection</i>	164
6.3.6	<i>RNA probe preparation and whole-mount in situ hybridization</i>	165
6.3.7	<i>Tissue distribution of cee mRNA in adult salmon</i>	165
<b>6.4</b>	<b>Results</b>	166
6.4.1	<i>In silico identification of cee in eukaryote genomes</i>	166
6.4.2	<i>Characterisation of cee in metazoans</i>	168
6.4.3	<i>Gene structure of cee</i>	169
6.4.4	<i>Phylogenetic reconstruction of Cee</i>	170
6.4.5	<i>Selective constraints on Cee during animal evolution</i>	170
6.4.6	<i>Developmental expression of cee</i>	171
<b>6.5</b>	<b>Discussion</b>	182
6.5.1	<i>Cee is a novel protein of unknown function that is widely conserved in the eukaryotes</i>	182
6.5.2	<i>cee is a lonely gene without a family</i>	183
6.5.3	<i>Gene structure of cee</i>	184
6.5.4	<i>Developmental expression of cee</i>	185
<b>Chapter 7</b>	<b>The co-ordinated spatiotemporal expression of myogenic regulatory factors is affected by temperature during embryonic myogenesis in Atlantic salmon (<i>Salmo salar</i> L.)</b>	187-227
<b>7.1</b>	<b>Abstract</b>	187
<b>7.2</b>	<b>Introduction</b>	188
<b>7.3</b>	<b>Materials and methods</b>	189
7.3.1	<i>Embryos</i>	189
7.3.2	<i>Cloning new Atlantic salmon myogenic regulatory factors and <i>smlc1</i></i>	189

7.3.3 <i>Computational processing of new MRF sequences</i>	190
7.3.4 <i>Probe transcription and in situ hybridisation</i>	190
7.3.5 <i>Processing embryos and figure construction</i>	191
<b>7.4 Results</b>	192
7.4.1 <i>Characterisation of Atlantic salmon myogenic regulatory factors</i>	192
7.4.2 <i>Genomic organisation of salmonid MRFs</i>	194
7.4.3 <i>Characterisation of Atlantic salmon <i>smc1</i></i>	194
7.4.4 <i>MRF expression: introduction</i>	194
7.4.5 <i>MRF expression in a single maturing somite (Fig. 7.6)</i>	195
7.4.6 <i>The dynamics of rostral-caudal expression of MRFs during embryogenesis (Figs. 7.8-7.9)</i>	197
7.4.7 <i>Embryonic temperature and somitogenesis</i>	201
7.4.8 <i>Embryonic temperature and the co-ordinated expression of MRFs</i>	201
<b>7.5 Discussion</b>	217
7.5.1 <i>MRFs in Atlantic salmon and other teleosts</i>	217
7.5.2 <i><i>Mrf4</i> and myogenic specification, a trait lacking in teleosts?</i>	220
7.5.3 <i>Salmon MRFs and the external cell layer</i>	221
7.5.4 <i>Heterochronies in MRF expression at different temperatures</i>	221
7.5.5 <i>Concluding thoughts</i>	223
<b>Chapter 8. Temperature until the ‘eyed stage’ of embryogenesis programs the growth trajectory and muscle phenotype of adult Atlantic salmon</b>	227-246
<b>8.1 Abstract</b>	227
<b>8.2 Introduction</b>	228
<b>8.3 Materials and methods</b>	229

8.3.1 Embryonic temperature experiment	229
8.3.2 Fish sampling	229
8.3.3 Modelling growth data	230
8.3.4 Muscle morphometry statistics	231
8.3.5 Calculating fibre probability density functions	231
<b>8.4 Results</b>	232
8.4.1 Embryonic growth trajectory	232
8.4.2 Post embryonic growth trajectory	232
8.4.3 Embryonic temperature and adult muscle fibre morphometrics	233
8.4.4 Embryonic temperature and muscle fibre nuclear density	234
8.4.5 Temperature-induced differences in muscle fibres size distribution	235
<b>8.5 Discussion</b>	242
8.5.1 Changing temperature solely to the eyed stage programs adult salmon growth trajectory and muscle fibre phenotype	242
8.5.2 Implications of temperature induced differences in muscle fibre number	243
8.5.3 Embryonic temperature induced alterations in myonuclear density	244
8.5.4 How does embryonic temperature program adult myogenic phenotype?	245
<b>Chapter 9. General discussion</b>	247-256
9.1 A molecular tool-box for studying Atlantic salmon myogenesis	247
9.2 Characterising teleost and muscle salmon genes: patterns and paralogues	247
9.3 What role did genome duplication play in the teleost success story?	249
9.4 Taking the embryonic temperature findings onwards	251
9.5 Effect of embryonic temperature on other systems	253

<b>9.6 Thoughts for the future</b>	254
<b>Appendix I. Two colour whole mount <i>in situ</i> hybridisation using sequential alkaline phosphatase staining with chromogenic substrates of <i>Salmo salar</i> embryos</b>	259-265
<b>Appendix II. List of manufacturers addresses</b>	266
<b>References</b>	267-312

## List of Figures

	Page
Fig. 1.1. Taxonomic relationships of major model species within the phylum Chordata	9
Fig. 1.2. A simple model of post-embryonic vertebrate myogenesis	13
Fig. 1.3. The archetypal teleost muscle phenotype	17
Fig. 1.4. A simple model demonstrating the individual role of the MRFs	21
Fig. 1.5. Schematic representation of the cellular specification, differentiation and migration of embryonic MPC populations in a generalised teleost	45
Fig. 3.1 Full AA sequence alignment of MyoD1 and MyoD2 AA sequences in three species of the Acanthopterygii	86
Fig. 3.2. Average ratio of synonymous and non-synonymous substitutions in MyoD1 and MyoD2 of Acanthopterygian teleosts	87
Fig. 3.3. ML cladogram of vertebrate MRF sequences	88
Fig. 3.4 Unrooted phylograms of vertebrate MyoD AA sequences constructed using Bayesian, maximum likelihood and neighbour joining approaches	90
Fig. 3.5. Conserved synteny between the <i>myod</i> -containing chromosomes of human, chicken, zebrafish, pufferfish, stickleback and medaka	91
Fig. 3.6. Unrooted phylogenetic cladograms for AA translations of genes in proximity to <i>myod</i> that are conserved as two copies in teleosts	93
Fig. 4.1. MyoD1c at the mRNA and protein level	115
Fig. 4.2 Multiple species alignment of vertebrate MyoD1 sequences	117
Fig. 4.3. Differential expression of <i>myod1</i> genes during Atlantic salmon somitogenesis	118
Fig. 4.4. <i>myod1</i> expression at the Atlantic salmon eyed stage	119
Fig. 4.5. Quantitative expression of <i>myod1</i> paralogues in adult Atlantic salmon fast and slow muscle fibres	120
Fig. 4.6. Two scenarios could account for three salmonid <i>myod</i> paralogues conserved from the <i>myod1</i> lineage	128
Fig. 4.7. A simple scenario describing the evolution of salmonid <i>myod1</i> paralogues under the DDC/subfunctionalization model	129

Fig. 5.1. ML phylogenetic reconstruction of 33 vertebrate AA Fst sequences	144
Fig. 5.2. Several duplicated genes in the neighbourhood of zebrafish <i>fst1/2</i> were also present on two chromosomal regions in Acanthopterygian teleosts	145
Fig. 5.3. Phylogenetic reconstruction of genes in chromosomal proximity to <i>fst</i> in zebrafish and Acanthopterygian teleosts	147
Fig. 5.4. Plots of the cumulative number of synonymous and non-synonymous substitutions across Fst proteins from mammals, euteleosts and the Ostariophysi	149
Fig. 5.5. Duplex RT-PCR of <i>fst</i> and <i>ef1-<math>\alpha</math></i> genes in zebrafish and Atlantic salmon	150
Fig. 5.6. Conserved motifs in the promoter regions of teleost and tetrapod <i>fst</i>	151
Fig. 5.7. Expression of <i>fst1</i> during the segmentation period of Atlantic salmon embryogenesis	152
Fig. 5.8. Expression of <i>fst1</i> at the end of the segmentation period	153
Fig. 5.9. Expression of <i>fst1</i> in muscle progenitors of the pectoral fin buds	154
Fig. 6.1. Sequence alignment of Cee amino acid sequences from twelve vertebrates	176
Fig. 6.2 Gene organisation of <i>cee</i> across the eukaryotes	177
Fig. 6.3. Unrooted ML phylogram of Cee proteins from a broad range of metazoan taxa	178
Fig. 6.4. Synonymous and non-synonymous substitution rates in Cee.	179
Fig. 6.5. Developmental expression pattern of <i>cee</i> in Atlantic salmon	180
Fig. 7.1. Atlantic salmon Myf5 at the mRNA and AA level	204
Fig. 7.2. Atlantic salmon Myog at the mRNA and AA level	205
Fig. 7.3. Atlantic salmon Mrf4 at the mRNA and AA level	206
Fig. 7.4. Sequence alignment of all known salmonid MyoD family members with a MyoD orthologue in the cephalochordate-amphioxus	207
Fig. 7.5. Intron-exon structures of Atlantic salmon myogenic regulatory factors	208
Fig. 7.6. Schematic diagram illustrating, for several stages of Atlantic salmon embryogenesis, the mRNA expression patterns of six MRF genes, as well as <i>smlc1</i> and <i>pax7</i> in the most anterior (oldest) somite	209
Fig. 7.7. Expression of Atlantic salmon MRFs prior to somitogenesis	210

Fig. 7.8. mRNA expression patterns of MRFs and <i>smlc1</i> expression during the 30-45 somite stage of Atlantic salmon embryogenesis	211
Fig. 7.9. mRNA expression patterns of MRFs as well as <i>smlc1</i> and <i>pax7</i> at the 65ss and during the eyed stage of Atlantic salmon embryogenesis	212
Fig. 7.10. Rate of somitogenesis in Atlantic salmon reared at 2, 5 and 8°C	213
Fig. 7.11. Temperature associated heterochronies in <i>myf5</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	214
Fig. 7.12. Temperature associated heterochronies in <i>mrf4</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	215
Fig. 7.13. Temperature associated heterochronies in <i>smlc1</i> expression in Atlantic salmon embryos incubated at 2 or 8 °C	216
Fig. 7.14. Two <i>mrf4</i> genes map to two distinct genomic locations in Atlantic salmon	225
Fig. 8.1. Changing embryonic temperature solely to the ‘eyed stage’ of embryogenesis produced marked effects on the post-smoltification growth trajectory of Atlantic salmon	226
Fig. 8.2. The effect of embryonic temperature on the fast-steak cross sectional area of adult Atlantic salmon	238
Fig. 8.3. The effect of embryonic temperature on the final muscle fibre phenotype of adult Atlantic salmon	239
Fig. 8.4. Scatterplot of myonuclear density versus the fast-steak cross sectional area in adult Atlantic salmon reared at 5 and 10°C	240
Fig. 8.5. The effect of embryonic temperature on the size distribution of muscle fibres in adult Atlantic salmon	241

## List of Tables

	Page
Table 3.1. Details of teleost MyoD sequences, including their current designation, Genbank accession number/Ensembl gene ID as well as correct designations (proposed consensus nomenclature) according to the comparative-genomic and phylogenetic results of chapter 3.	101
Table 4.1. Primer details for chapter 4	106
Table 4.2. Comparison of percentage sequence identity of coding and non-coding features of <i>myod1</i> co-orthologues ( <i>myod1a/1b/1c</i> ) and <i>myod1c</i> orthologues	116
Table 6.1. List of primer pairs used to amplify <i>cee</i> in chapter 6	173
Table 6.2. Accession numbers for metazoan <i>cee</i> genes and putative proteins studied chapter 6	174
Table 7.1. Primer details for chapter 7	203
Table 8.1. Summary of mixed-model ANOVA parameters used to distinguish variation in post-smolt body mass	235
Table 8.2. Summary of general linear model ANOVA parameters used to distinguish variation in muscle fibre characteristics	237
Table 9.1. New genes characterised during the current project	257

## List of standard and non-standard abbreviations

AA: Amino acid

Abcc-8: ATP-binding cassette, sub-family C (CFTR/MRP), member 8,

ATP: Adenosine Triphosphate

BCIP: 5-Bromo-4-chloro-3-indolyl phosphate

BHLH: basic helix-loop-helix

BLAST: Basic Local Alignment and Search Tool

Blimp1: transcriptional repressor B-lymphocyte-induced maturation protein 1

BMP: bone morphogenic protein

BSA: Bovine Serum Albumen

Bya: Billion years ago

cDNA: complementary deoxyribose nucleic acid

CDS: coding sequence

Cee: conserved edge expressed

CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

cm: centimetres

cRNA: complementary nucleic acid

Cxcr4: CXC chemokine Receptor 4

DacA: *Drosophila* dachshund A homologue

DDC: duplication, degeneration, complementation

DEPC: diethylpyrocarbonate

DIG: digoxigenin

DNA: deoxyribose nucleic acid

dNTP: deoxyribonucleotide triphosphate

EDTA: ethylenediaminetetraacetic acid

EST: expressed sequence tag

EtBr: ethidium bromide

Fgf8: fibroblast growth factor 8

FLRG: follistatin related gene

Flu: fluorescein

Foxk1: forkhead box K1

Fst: follistatin

GASP1: WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2

GATA-2: GATA binding protein 2

GDF8: growth differentiation factor 8

GDF11: growth differentiation factor 11

HCl: hydrogen chloride

Hh: hedgehog

HLH: helix-loop-helix

Hox: homeobox gene

IGF: Insulin-like growth factor

JTT: Jones-Taylor-Thorton

kb: kilobase

KCl: potassium chloride

Kcnc1: potassium voltage gated channel-1

LB: Luria-Bertani

Lbx1: ladybird homeobox homolog 1

LiCl: lithium chloride

MADS-box: MCM1, AGAMOUS, DEFICIENS and SRF (serum response factor)-box

MEF2: myocyte enhancing factor 2

MgCl<sub>2</sub>: magnesium chloride

min: minutes

miRNA: microRNA

ML: maximum likelihood

mm: millimetres

mM: millimolar

MMLV: Moloney Murine Leukemia Virus

Mox1: mesenchyme homeobox 1

MPC: myogenic progenitor cell

MRF: myogenic regulatory factor

Mrf4: myogenic regulatory factor 4

mRNA: messenger RNA

MS222: tricaine methanesulphonate

Mstn: myostatin

mTOR: mammalian target of rapamycin

m/v: molecular volume

Mya: million years ago

MyoD: myogenic determination factor

Myog: myogenin

Myf5: myogenic factor 5

Myf6: myogenic factor 6

NaCl: sodium chloride

NBT: Nitro-Blue Tetrazolium Chloride

NCBI: National Center for Biotechnology Information

NFAT: nuclear factor of activated T-cells

NH<sub>4</sub>: ammonium

NJ: neighbour joining

nM: nanomolar

Nucb2: nucleobindin 2

ORF: open reading frame

Otog: Otogelin

p70S6 kinase: ribosomal S6 protein kinase p70 S6 kinase

PCNA: proliferating cell nuclear antigen

PI(3)K: phosphatidylinositol-3-OH kinase

Pik3c2a: phosphoinositide-3-kinase, class 2, alpha polypeptide

pH: power of hydrogen (a measure of the acidity or alkalinity of a substance)

Pax3: paired box transcription factor 3

Pax7: paired box transcription factor 7

PBS: phosphate buffered saline

PBT: phosphate buffered saline; 0.1% Tween 20

Pbx: pre-B cell leukaemia homeobox gene

PCR: polymerase chain replication

PHAS-1: phosphorylated heat and acid stable protein regulated by insulin

PIT Tag: passive integrator transponder tag

Plekha7: pleckstrin homology domain containing, family A member 7

ppm: parts per million

qPCR: quantitative real-time RT-PCR

RNA: ribonucleic acid.

RNase: ribonuclease

Rpm: revolutions per minute

Rps13: ribosomal protein S13

RT-PCR: reverse transcription-PCR

s: seconds

Sdf1a: Stromal Cell Derived Factor 1 alpha

SERGEF: secretion regulating guanine nucleotide exchange factor

Smlc1: Slow myosin light chain 1

ss: somite stage

SSC: sodium chloride sodium citrate

TAE: Tris-acetate-EDTA

TE: Tris-EDTA

TFBS: transcription factor binding site

TGF- $\beta$ : Transforming growth factor-beta

Tm: melting temperature

Tph1: tryptophan hydroxylase gene-1

Tris: Trizma acetate

tRNA: transfer RNA

TropI: Troponin I

TropT: Troponin T

UTR: untranslated region

v/v volume/volume

WAG: Whelan And Goldman

WGD: whole genome duplication

$\mu$ M: micromolar

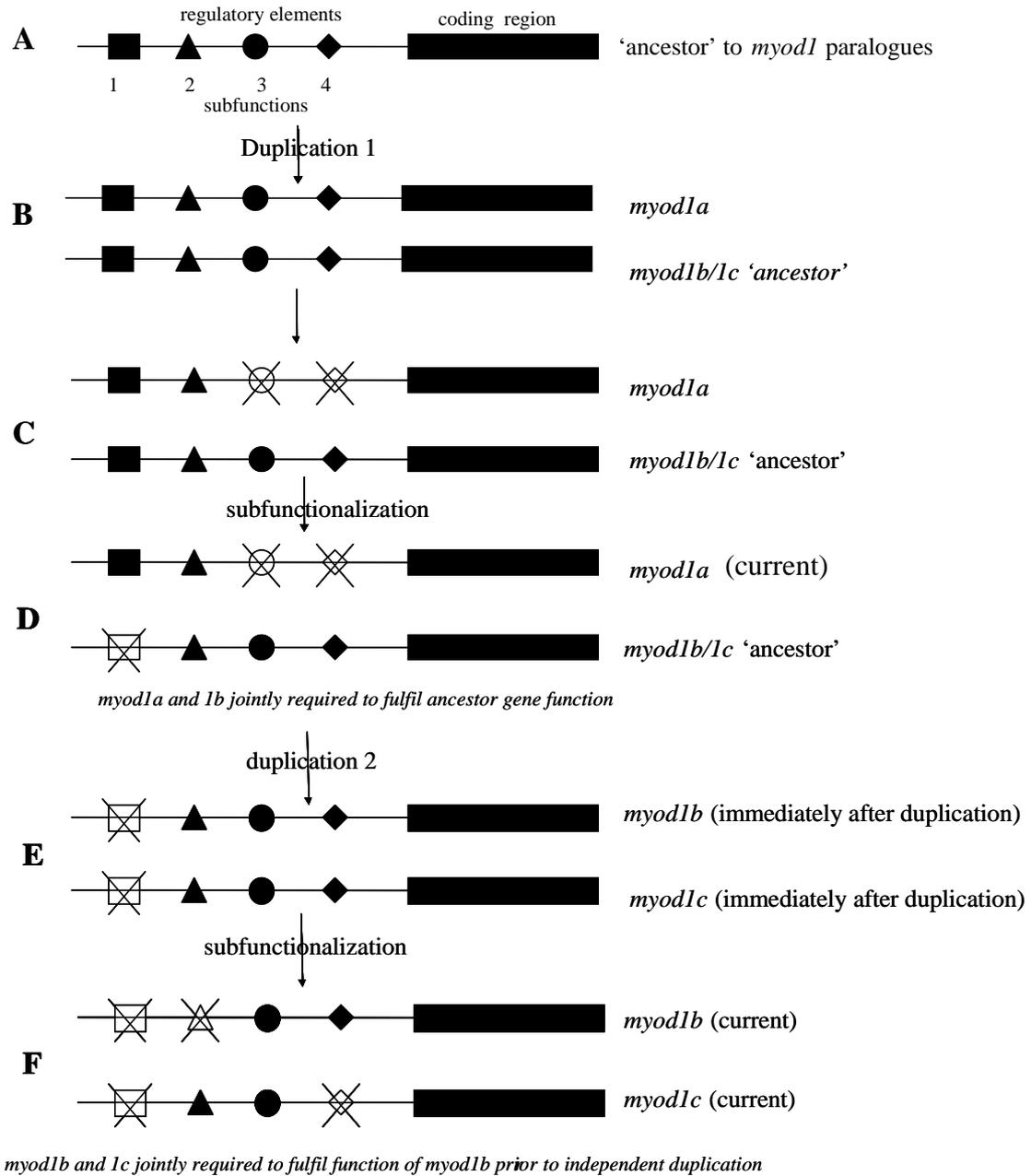


Fig. 4.7. A simple scenario describing the evolution of salmonid *myod1* paralogues under the DDC/subfunctionalization model. Note that scenario 2 from Fig. 4.6 is considered here. Subfunctions can be thought of as the regulatory elements governing the total expression domain of the ancestral *myod1* gene and are split as: 1. PSM/early adaxial cell expression. 2. Somitic adaxial cell expression 3. Expression in posterior domain of the epithelial somite. 4. Expression at the peripheral myotome at the end of segmentation.

## **Chapter 5. Evolution of Follistatin in teleosts revealed through phylogenetic, genomic and expression analyses: a role for *fst1* in teleost myogenesis**

### **5.1 Abstract**

Follistatin (Fst) is an important regulator of amniote myogenesis. In this chapter, a comparative phylogenetic, genomic and experimental approach was used to study its evolution in teleosts and its potential role in myogenesis during Atlantic salmon embryogenesis. A ML analysis showed that one *fst* gene (*fst1*) is common to euteleosts, but a second gene (*fst2*) is conserved specifically within the Ostariophysi. However, the ML tree topology suggested that these *fst* duplicates arose in a teleost specific event. Zebrafish *fst1/2* appear on chromosomes 5/10 in two genomic regions, each with conserved synteny to a single region in tetrapods. Interestingly, other teleosts have two corresponding chromosomal regions with a similar repertoire of paralogues. Phylogenetic reconstruction generally clustered these gene duplicates into two sister clades branching from tetrapod sequences. These findings indicate that an ancestral *fst*-containing chromosome was duplicated during the teleost WGD, but that *fst2* was lost in lineages external to the Ostariophysi. Interestingly, the Fst1 protein of teleosts/mammals has evolved under strong purifying selection, but the N-terminal of Fst2 has likely evolved under positive selection. Furthermore, the tissue distribution of zebrafish *fst2* was restricted to fewer tissues than zebrafish *fst1* and the *fst1* gene of Atlantic salmon, suggesting that two independently evolving regulatory regions govern the expression of these paralogues. Zebrafish *fst1/2* may have subfunctionalized relative to non-duplicated vertebrate lineages since several regions in the *fst* promoter of tetrapods were conserved with one paralogue, but not both. Finally, the embryonic expression of *fst1* was examined in Atlantic salmon to establish its potential role in myogenesis. During early-mid segmentation, *fst1* and *pax7* were concurrently expressed in the anterior compartment of the epithelial somite, but were excluded from muscle progenitors that strongly expressed MRFs. At

the end of segmentation, *fst1* and *pax7* mRNA became co-restricted to the external cell layer and further, *fst1* was expressed in myogenic progenitors of the pectoral fin buds in rostral somites. These expression results imply an unappreciated role for teleost Fst1 in myogenic cells originating from the anterior somite, which is functionally equivalent to the dermomyotome of amniotes.

## 5.2 Introduction

The *fst* gene of amniotes has a vital role in regulating myogenesis (reviewed in chapter 1, section 1.6.8). For example, in the chick embryo, Fst is expressed in myogenic progenitors of the dermomyotome (Amthor et al., 1996, 2004), where it has been shown to positively regulate muscle growth by antagonising Myostatin's negative effect on the expression of muscle transcription factors like MyoD and Pax3 (Amthor et al., 2004), and modulating the titres of bone morphogenic proteins (Amthor et al., 2002). Research into the *fst* gene in teleosts is limited. The zebrafish (*D. rerio*) has two *fst* genes, *fst1* and *fst2*, that were characterised in the context of dorso-ventral patterning (Dal-Pra et al., 2006; Bauer et al., 1998). However, it is yet to be established whether *fst* plays a role in regulating the fate of the dermomyotome-equivalent anterior somite compartment of teleosts. Further, since teleost fish underwent WGD in the early evolution of the ray-finned fish (Actinopterygii) (Jaillon et al., 2004), it is possible that like zebrafish, many teleost genomes have two *fst* genes.

In this chapter it is initially shown that the cyprinid *Pimephales promelas* (fathead minnow), like zebrafish, has two *fst* genes. Additionally, phylogenetic analysis indicated that three siluriform (catfish) *fst* genes recently characterised (Gregory et al., 2004) are orthologues of one of the zebrafish/minnow genes (*fst2*), and that the other *fst* gene (*fst1*) is universal among teleosts, and less derived from the single *fst* gene of tetrapods. By establishing the genomic neighbourhood

surrounding teleost and tetrapod *fst* genes, the ratio of synonymous to non-synonymous substitutions in teleost Fst1/Fst2 and mammalian Fst1 as well as the tissue specific expression patterns of teleost *fst* genes, I propose an evolutionary scenario to explain the presence of two *fst* paralogues within the Ostariophysi. Finally, *fst1* expression was investigated in a teleost fish external to this group (Atlantic salmon) to reveal a possible role in the function of cells derived from the anterior somite compartment.

## 5.3 Material and Methods

### 5.3.1 Sequence retrieval and genomic analyses

Complete Fst AA sequences were retrieved from GenBank for the following species: Human (*H. sapiens*: AAH04107), rat (*R. norvegicus* NP\_036693), cow (*B. taurus*: AAA305), mouse (*M. musculus*: NP\_032072), pig (*S. scrofa*: NP\_001003662), horse (*Equus caballus*: NP\_001075280), chicken (*G. gallus*: NP\_990531), frog (*X. laevis* AAB30638), zebrafish (Fst1: AAD09175, Fst2: ABC48670), tiger pufferfish (*T. rubripes*: DQ288127), tilapia (*O. mossambicus*: ABC69147), largemouth bass (*Micropterus salmoides*: ABL95955), goldfish (*Carassius auratus*: AAR99335), grass carp (*Ctenopharyngodon idella*: ABC72407), white catfish (*A. catus*: AAS88751), blue catfish (*I. furcatus*: AAS88750), channel catfish (*I. punctatus*: AAS48082), Atlantic salmon (*S. salar*) (ABA29021) and marine lamprey (*Petromyzon marinus*: ABD49691).

A rainbow trout (*O. mykiss*) EST sequence for *fst* was obtained from a tBLASTn search in the Salmon Genome Project database ([www.salmongenomeproject.no](http://www.salmongenomeproject.no)) (accession number: CA36476). Two distinct *fst* sequences were obtained for the fathead minnow (*P. promelas*) from tBLASTn searches of the TGI database (<http://compbio.dfci.harvard.edu/tgi/>) and were both partial at the 3' (accession numbers: DT261339 and DT272487).

Manual screening and tBLASTn searches (using human *Fst* as a probe) of several Ensembl genome databases ([www.ensembl.org](http://www.ensembl.org)) identified the following high quality *fst* predictions and the Ensembl gene ID and genomic location are shown in brackets: Macaque (*Macaca mulatta*: ENSMMUP00000006461, complete sequence, located on chromosome 6 at nucleotides 51,141,906-51,147,384), cat (*Felis catus*, ENSFCAP00000002232, partial at 5', located on GeneScaffold\_1792 at nucleotides 63,515-80,323), dog (*C. familiaris*, ENSCAFG00000018405, partial, located on chromosome 4, at nucleotides 64,982,844-64,988,119), rabbit (*Oryctolagus cuniculus*: ENSOCUG00000004687, partial at 5', located on scaffold\_189546 at nucleotides 7,024-9,641), European hedgehog, (*Erinaceus europaeus*, ENSEEUG00000014055 partial at 5' end, located on scaffold-305006 at nucleotides 10,877-13,858), opossum (*Monodelphis domestica*, ENSMODG00000019463, partial at 3', located on chromosome 3 at nucleotides 15,752,774-15,758,423) platypus (*Ornithorhynchus anatinus*, ENSOANP00000017746, partial at 5', located on chromosome 1 at nucleotides 116,977-124,688), stickleback (*G. aculeatus*: ENSGACG00000004583, complete sequence, located on group 13 at nucleotides 1,936,787-1,941,260), medaka (*O. latipes*, ENSORLP00000002355, complete sequence, located on chromosome 9 at nucleotides 4,945,916-4,950,698), green-spotted pufferfish (*T. nigroviridis*, GSTENG00029600001, complete sequence, located on chromosome 12 at nucleotides 5,833,182-5,835,972).

### 5.3.2 Construction of synteny diagram

The genomic neighbourhoods surrounding *fst* genes in zebrafish (*fst1* and *fst2*), stickleback, pufferfishes (*T. nigroviridis*, and *T. rubripes*), chicken and human, were manually obtained using the Multicontigview in the Ensembl database. MatInspector and DialignTF (Cartharius et al. 2005) were used to analyse the proximal promoter (2KB upstream of the first *fst* exon), for shared transcription-factor binding sites (TFBS's) and other regions conserved between

zebrafish, *X. tropicalis* and *M. musculus*. Pairwise comparisons using nucleotide sequences and AA translations were performed in DNAMAN.

### 5.3.3 Phylogenetic and evolutionary analyses

The 33-Fst sequences obtained were aligned in clustalX (Thompson et al. 1997) using default penalty parameters for pairwise alignments (gap opening: 10, gap extension 0.1) and respective multiple alignment penalties of 2.0 and 0.5 for gap opening and extension which gave the highest Q-score in TuneClustalX (<http://www.homepage.mac.com/barryghall/Software.html>) versus a series of other penalties.

ML was performed on this alignment using PhyML (Guindon and Gascuel, 2003) with a gamma distribution of among site substitution rates, the WAG model for AA substitution, and 500 pseudobootstrap iterations for branch support. Mega 3.1 (Kumar et al. 2004) was used to produce a concurrent NJ tree using the JTT substitution model and 1000 bootstrap replicates for branch support with a gamma distribution of among site substitution rates. Additionally a NJ tree was produced using the JTT substitution matrix in ASATURA (Van de Peer et al. 2002), which was used to remove frequently mutated amino-acid positions from the alignment prior to tree reconstruction (with 1000 bootstrap replicates). All trees were rooted using the marine lamprey orthologue of Fst.

For genes in proximity to *fst*, phylogenetic analysis was performed on AA sequences retrieved from the Ensembl database and aligned in ClustalX using default parameters for pairwise alignments and multiple Gap penalties optimised using TuneClustalX. NJ analyses were then performed on these alignments considering all substitution sites in Mega 3.1 using a gamma distribution of among site substitution rates and after excluding frequently mutated amino-acid

positions using ASATURA (in both cases using the JTT substitution model and 1000 bootstrap replicates). Details of outgroups used to root these trees can be found in the Fig. 5.3 legend.

#### 5.3.4 Testing the selective constraints across *Fst* proteins

To investigate the selective constraints affecting *Fst* proteins during evolution, experimentally validated fully-coding *fst* mRNA sequences were first split into four groups: 1. Teleost *fst1* (*D. rerio*, *T. rubripes* and *S. salar*), 2. Teleost *fst2* (*D. rerio*, *P. promelas* and *A. catus*), and 3. Mammalian *fst1* (*M. musculus*, *H. sapiens* and *S. scrofa*). Putative translations of these groups were aligned at the AA level and along with the corresponding nucleotide sequences, loaded into PAL2NAL (Suyama et al. 2006), which converted this data into a multiple codon alignment. This was loaded into SNAP (Korber et al., 2000), which estimated the average number of non-synonymous (dN) and synonymous (dS) substitutions within each codon alignment. SNAP was used to produce a plot of the cumulative average non-synonymous and synonymous substitutions across the *Fst* proteins from the different groups.

#### 5.3.5 Tissue specific mRNA expression of teleost *fst* genes

To assess the tissue specific expression of zebrafish *fst1/2* and salmon *fst1*, reverse transcription PCR (RT-PCR) was used with cDNA derived from eight different tissues: heart, brain, liver, spleen, ovary, skin, slow-twitch myotomal muscle and fast-twitch myotomal muscle. For zebrafish, dissections of these tissues were taken from 3-10 adults (25mm standard length) and stored in RNAlater (Ambion). For Atlantic salmon, the same tissues were dissected from two adult fish (3.8 and 4kg) and flash frozen in liquid N<sub>2</sub>. Total RNA was extracted from each set of salmon tissues, and from pooled tissues for zebrafish, using the method described in chapter 2 (section 2.4.2) The quantity of total RNA was recorded using a Nanodrop ND-1000 spectrophotometer (Nanodrop technologies, UK) and its quality confirmed by assessing the integrity of 18S and 28S ribosomal RNA bands by gel electrophoresis. The QuantiTect Reverse

Transcription kit (Qiagen) was then used to transcribe cDNA from 500 ng of total RNA, with a concurrent gDNA removal step. Controls lacking reverse transcriptase were included on a pool of total RNAs. Duplex RT-PCR reactions were then performed using 1µl of each cDNA (5X diluted) as a template, with concurrent amplification of the housekeeping gene *ef1-α* (primer sequences for both zebrafish and salmon: sense, 5'-3': ATGGGCTGGTTCAAGGGATG; antisense, 5'-3': GGGTGGGTCGTTCTTGCTGT) and each target gene: zebrafish *fst1* (primer sequences: sense, 5'-3': CGCTGCTCGTCTCTACTCTTTC; antisense, 5'-3': GCAACATTCCTCCCGACTCATC), zebrafish *fst2* (primer sequences: sense, 5'-3': AGACATCAACTGCCGAGAGGG, antisense, 5'-3': CAGGAGCCCGAGTGTTTGACTTC) and salmon *fst1* (primer sequences: sense, 5'-3': CCCGATACCTCGTTCACTTGTTTC, antisense 5'-3': CCCAGCCTCCCGCTTCTAC). In each case 37 cycles were performed with 30s at 95 °C, 30s at 60 °C and 30s at 72 °C, and reverse transcriptase and no-template (1 µl sterile water) controls were included. Salmon cDNAs from both fish were tested separately and produced highly comparable amplification profiles.

### 5.3.6 Cloning of salmon *fst1*

To clone salmon *fst1*, total RNA was initially extracted from the heart ventricle of a juvenile adult Atlantic salmon (~300g) following the expression results of Gregory et al. (2004). The concentration of total RNA was then quantified, and its integrity assessed as described in chapter 2 (section 2.4.2). First strand cDNA was synthesised from 1µg of RNA using the RETROscript kit with subsequent enzymatic elimination of genomic DNA (as described in chapter 2, section 2.4.5). Standard RT-PCR reactions were ran using the following primers (f: 5-3' CCCGATACCTCGTTCACTTG; r: 5-3'-GGAAGGGAATGAAGAGGCG) with 1µl of heart cDNA as a template. A band corresponding to the expected size of *fst1* was amplified, cloned and sequenced as previously described (chapter 2). This revealed that the PCR product encoded the whole coding sequence of the Atlantic salmon *fst1* gene.

### 5.3.7 Embryos and whole-mount *in situ* hybridization

Atlantic salmon embryos used in this chapter were from the 10°C group of the temperature experiment (chapter 2, section 2.2). The salmon *fst1* PCR product (including 65 nucleotides of 5' UTR and nucleotides 1-963 of the coding sequence) contained in the TOPO-vector was used to transcribe sense/antisense digoxigenin-labelled cRNA probes with T3/T7 polymerases (Roche). *In situ* hybridization was performed with six embryos from each developmental stage as described in chapter 2 (section 2.6). Embryos were cryosectioned and photographed as described in chapter 2 (section 2.6.6).

## 5.4 Results

### 5.4.1 Two *fst* paralogues are present in the *Ostariophysi*

Computational searches revealed that in most vertebrate species, *fst* is represented as a single gene. It was recently shown that zebrafish have two *fst* genes, namely *fst1* and *fst2* (Dal-Pra et al. 2006). Here two distinct EST sequences were discovered in the fathead minnow that were each orthologous to one of the two-zebrafish genes (DT261339 and zebrafish Fst1 share 98% AA identity; DT272487 and zebrafish Fst2 share 94.5% AA identity, whereas zebrafish Fst1 and Fst2 share ~70% AA identity). To establish the phylogenetic relationships of vertebrate *fst* genes, ML and NJ analyses were used on an alignment of teleost Fst AA sequences (n = 17), with representative orthologues from several tetrapods (n = 15) and the marine lamprey Fst orthologue as an outgroup (Fig. 5.1). NJ analyses were also performed firstly considering all AA substitution sites, but also when frequently mutated AA positions were removed. This was achieved manually using a plot of substitution frequency versus evolutionary distance in ASATURA (Van de Peer et al., 2002) and occurred at a cut off value of 1766 using the JTT matrix. Removing frequently mutated AA positions from the Fst alignment had no effect upon

the subsequent NJ tree topology. Trees produced by all methods showed that tetrapod *Fst* sequences branched together as a single clade (Fig. 5.1). However, the sampled teleost *Fst* sequences branched into two clades (100% ML bootstrap support/100% NJ bootstrap support/94% 'unsaturated' NJ bootstrap support). One clade was represented by several teleost superorders including the Ostariophysi, Acanthopterygii and Protacanthopterygii and was more closely related to the single tetrapod *Fst* clade. *fst1* is an appropriate designation for this gene in vertebrates. A second gene, *fst2*, formed a second teleost *Fst* clade and was represented solely by species of the Ostariophysi superorder, including three catfish species, and two cyprinids (zebrafish and fathead minnow) (Fig. 5.1). Interestingly, in all trees the *Fst2* clade branched away from *Fst1* before the divergence of the Ostariophysi from other teleost lineages (Fig. 1, 100% ML bootstrap support), which is thought to have occurred around 150 MYA (Benton and Donoghue, 2007).

#### *5.4.2 Teleost fst genes are present on duplicated chromosomes with double conserved synteny relative to tetrapods*

To further explore the *fst* duplication in the Ostariophysi I retrieved the genomic neighbourhood of the chromosomal regions surrounding zebrafish *fst1* and *fst2*, and then searched for corresponding synteny with other teleosts and two tetrapod species (summarised in Fig. 5.2). Zebrafish *fst1* and *fst2* were respectively present on chromosomes 5 and 10, supporting a paralogous non-allelic relationship (Fig. 5.2). In pufferfish (*T. nigroviridis*), stickleback, chicken and mouse, a single chromosomal region containing *fst1* can respectively be found on chromosome 12, group 13, chromosome Z and chromosome 13. On zebrafish chromosome 5, *fst1* appears twice, in two inverted blocks also containing genes coding for NADH dehydrogenase, Hspb3, Snag1 and Arl15 (Fig. 5.2, black boxes on chromosome 5). The inverted regions have retained a similar order of genes, except that Arl15 and Snag1 have swapped positions (Fig. 5.2). On chromosome 5, the genomic region containing the two-zebrafish *fst1* variants (marked a and b on Fig. 5.2), shared 98.5% nucleotide identity, and differed by a few

SNPs that were mainly synonymous at the AA level. Additionally, on chromosome 10, two genes just downstream to zebrafish *fst2* (*mbsp-4* and *snag-1*) are also present twice in an inverted region (Fig. 5.2, black box on chromosome 10).

A clear signal of duplication was detected between zebrafish chromosomes 5 and 10, as several genes including zebrafish *fst1/2* were present on both (Fig. 5.2). To establish whether this duplicated region was specific to zebrafish, I used tBLASTn and manual searches of several teleost/tetrapod genome databases for the genes in proximity to zebrafish *fst1/2*. Many of the zebrafish paralogues are also conserved as two copies in pufferfishes and stickleback on two chromosomes corresponding to zebrafish chromosomes 5 and 10 (Fig. 5.2, marked by \*). These duplicated teleost chromosomal regions each retained conserved synteny with single chromosomal segments in tetrapods (Fig. 5.2). Next, high quality AA translations of teleost paralogues were retrieved from the duplicated chromosomes along with their single corresponding tetrapod orthologues and phylogenetic reconstruction was used to establish their evolutionary relationships. In most cases (5/7), NJ trees constructed in Mega 3.1 and ASATURA supported a teleost specific origin of paralogues since two teleost sister clades branched away from a single tetrapod clade, following the expected chromosomal orthology established by the genomic analysis (Fig. 5.3, a-e). For future nomenclature purposes, genes orthologous to zebrafish chromosome 5 and 10 were respectively denoted as (GENE)- $\alpha$  and (GENE)- $\beta$ . However, in two cases (for *Golph3* and *Paqr3*) the expected tree topology was not obtained by NJ analysis in Mega 3.1 or when frequently saturated AA positions were removed by ASATURA. For these genes, one of the teleost clades appeared as a sister group to tetrapods (not shown), although the two teleost clades were still split according to the expected chromosomal orthology established by the genomic analyses (not shown). These results suggest that a *fst*-containing chromosome duplicated in a common ancestor to the Ostariophysi and Acanthopterygii, likely during the teleost WGD but that the *fst2* paralogue was lost in

Acanthopterygians. Further, when salmonid EST libraries were screened for *fst*, a single gene was retrieved, orthologous to *fst1* (Fig. 5.1), suggesting that *fst2* was also lost in the Protacanthopterygii lineage.

#### 5.4.3 Asymmetric evolution of teleost Fst paralogues

By studying the ML tree branch lengths from the point of the putative WGD (star on Fig. 5.1) it is clear that Fst1/2 have evolved at different rates. The genetic distance to the Fst1 and Fst2 clades is over 10-fold different, respectively ~0.015 and ~0.18 substitutions per site. Thus, Fst2 is evolving at a faster rate compared to its paralogue, reflected in lower percentage identities with tetrapod Fst1 proteins e.g. at the AA level from ~67-69% for zebrafish Fst2 compared to ~70-80% for zebrafish Fst1. Additionally, the branch length separating the cypriniform and siluriform clusters within the Fst2 clade is extended (~0.23 substitutions per site) relative to the maximum combined branch lengths within the Fst1 clade, (maximum of 0.14 substitutions in the medaka from the divergence of Fst1 between the Acanthopterygii and Ostariophysii), despite the fact that the sampled species in this group are less related than catfish and cyprinids. To test whether this advanced rate of evolution was accompanied by altered selective constraints, the cumulative change in synonymous (dS, silent) and non-synonymous (dN, AA-changing) substitutions was compared across Fst proteins in mammals and teleosts, for both Fst1 and Fst2 (Fig. 5.4). For Fst1 of mammals and teleosts, dS substitutions exceed dN substitutions across the whole protein (Fig. 5.4, a b). For Fst2, the number of protein changing substitutions exceeds that of synonymous substitutions from residues 1-65 (Fig. 5.4, c). After this region, the number of synonymous substitutions outweighs those that are AA altering, but not to the extent of Fst1 (Fig. 5.4).

#### 5.4.4 Distinct mRNA tissue distribution of *fst1/2* paralogues

RT-PCR was then used to investigate the tissue specific expression patterns of zebrafish *fst1/2* and the single *fst* gene of Atlantic salmon. Primers were designed in highly divergent regions to ensure that no cross amplification would occur. All samples were ran in identical conditions at the same time and as duplex reactions with the housekeeping gene *ef1- $\alpha$*  as a reference between *fst* primer sets. After 37 PCR cycles, salmon *fst1* and zebrafish *fst1* were detected in all eight tissues tested (Fig. 5.5, a vs. c) whereas zebrafish *fst2* expression was virtually undetected in liver and spleen, greatly reduced in heart and slow skeletal muscle, but strongly expressed in skin and brain, (Fig. 5.5, b vs. a). Thus zebrafish *fst1* and salmon *fst1* share a more similar tissue specific expression pattern compared to *fst2* of zebrafish.

#### 5.4.5 Comparative promoter analysis of vertebrate *fst* genes

Next, the proximal promoter (2KB upstream of exon 1, retrieved from the Ensembl database) of zebrafish *fst1/2* was compared with the equivalent region upstream of *fst1* in two tetrapods (*X. tropicalis* and *M. musculus*) using DiAlignTF (Cartharius et al. 2005). In the whole alignment, there were three predicted TFBS's recognised as conserved between all vertebrate *fst* promoters (e.g. Fig. 5.6, a) and each was situated toward the 3' of the promoter sequences. This included a TATA- binding site, a Gli-zinc finger family binding site, and an E-box motif all between 18-150 nucleotides upstream of exon 1 (not shown). However, there were ten instances when regions in the promoter alignment were conserved between zebrafish *fst1* and/or the frog/mouse promoter but not with zebrafish *fst2* (example in Fig. 5.6, b, d). Conversely, there were six instances when a zebrafish *fst2* promoter sequence was conserved with either or both tetrapods, but not zebrafish *fst1* (example in Fig. 5.6, c).

Additionally, MatInspector was used to predict TFBS in the proximal promoter of zebrafish *fst1* and *2* separately and 82 TFBS were common to both. However, there were several instances

when a TFBS was either absent, or more abundant in one promoter. As an example from the perspective of myogenesis, the zebrafish *fst1/2* promoters respectively had 3/0, 4/0, 2/0 and 4/1 predicted binding sites for the transcription factors MyoD1, Paraxis, Blimp1, and Pax3.

#### 5.4.6 Expression of *fst1* during embryonic myogenesis in salmon

Salmon *fst1* was first detected during early segmentation (~5 ss) in the presumptive cephalic mesoderm, anterior to the first few epithelial somites (Fig. 5.7, a, a1). The most posterior reach of expression was the anterior tip of somite-1 (Fig. 5.7, a2). In zebrafish, *fst1* was similarly expressed in the cephalic mesoderm from 50% epiboly and then in somites from the onset of segmentation (Thisse et al. 2001). A certain degree of somite maturity was required for the initiation of salmon *fst1* expression, meaning that during segmentation the most posterior few somites were unstained (Fig. 5.7, b, c). At all stages, *fst1* expression was excluded from the medial somite (Fig. 5.7, b2), where adaxial cells reside until around the 30ss. Additionally no *fst1* staining was present in the posterior somite, but instead was detected throughout most of the anterior width of the somite (Fig. 5.7, b1, b2), a near identical domain to *pax7* at this time (Fig. 5.7, labelled box within b1).

By the 45ss, a clear rostrally-directed decrease in *fst1* signal was recorded in the somites (Fig. 5.7, c). While *fst1* transcripts accumulated in the caudal somites as described for the 30ss, (Fig. 5.7, magnified blue box in c), expression in maturing somites was increasingly restricted to more lateral and ventral regions of the anterior myotome (Fig. 5.7, c1, c2-marked by red arrows). In the most rostral somites (1-5) at this time, *fst1* staining was further reduced in the bulk of the somite, but strongly maintained in ventral regions (shown in Fig. 5.9, b, c).

By the end of segmentation, *fst1* transcripts were mainly restricted to the external cell layer of somites (Fig. 5.8, a), which also expressed *pax7* mRNA (Fig. 5.8, b). However, whereas *pax7*

was expressed along the length of each somite (not shown), *fstl* expression was limited to the region adjacent to posterior border of the somite (Fig. 5.8, a). Concurrently, the MRF *myf5* was also expressed at the posterior somite border, but at the lateral edge of the myotome, bordering the external layer (Fig. 5.8, c) at a time when *myod1a* mRNA was expressed throughout the myotome bulk (Fig. 5.8, d).

#### 5.4.7 *fstl* expression in myogenic progenitors of the pectoral fin buds

In the most rostral few somites of 45ss salmon embryos, *fstl* expression was strongest in ventral regions at a time when *myog* was expressed throughout the myotome (Fig. 5.9, a-c). At the end of segmentation, the pectoral fin buds were evident as simple oval structures budding from the ventral portion of somites 1-6. At this time, no myoblasts were specified in the fin buds, evidenced by a lack of MRF expression, although transcripts were detected in adjacent myotomes (*myog* shown, Fig. 5.9, d). However, *fstl* was broadly expressed throughout lateral and ventral regions of fin bud-adjacent somites, including the external cell layer, but not the fin buds themselves (Fig. 5.9, e, f). As each pectoral fin bud matured, its developing musculature was marked by the expression of *myog* (Fig. 5.9, g) as well as *myod1a/1b/1c* (Chapter 4, Fig. 4.5, F). At this stage, *fstl* mRNA was even more restricted to the ventral-lateral region of pectoral fin adjacent somites and was also expressed within the ventral fin bud muscle (Fig. 5.9, h, i).

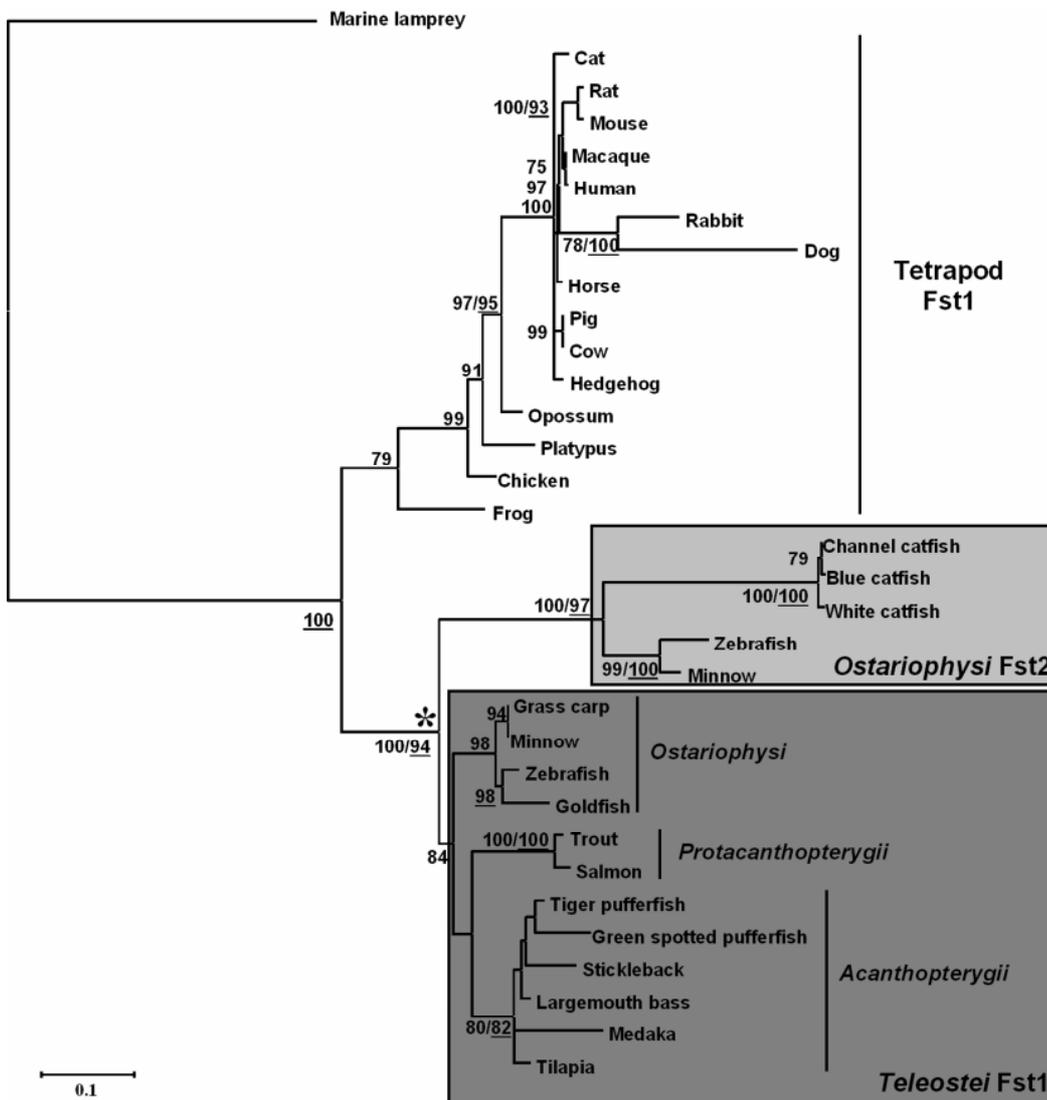


Fig. 5.1. ML phylogenetic reconstruction of 33 vertebrate AA Fst sequences produced in PhyML considering a gamma distribution of among site substitution rates. The Fst orthologue of *Petromyzon marinus*, the marine lamprey, was used to root the tree. Two concurrent NJ trees were produced (not shown), the first considering all substitution sites and the second considering only rare substitution sites (frequently mutated AA positions were removed in ASATURA). Both reconstructions were highly consistent with the ML analysis and each tree showed that in teleosts, the *fst* gene forms two sister clades, one represented by all Euteleosts (Fst1) and another solely by species of the Ostariophysi (Fst2). ML branch confidence values/supporting ASATURA-NJ bootstrap values (underlined) >75% are shown and were respectively obtained from 500 pseudobootstrap/1000 bootstrap replications. The scale bar shows the number of substitutions per site from the ML analysis.

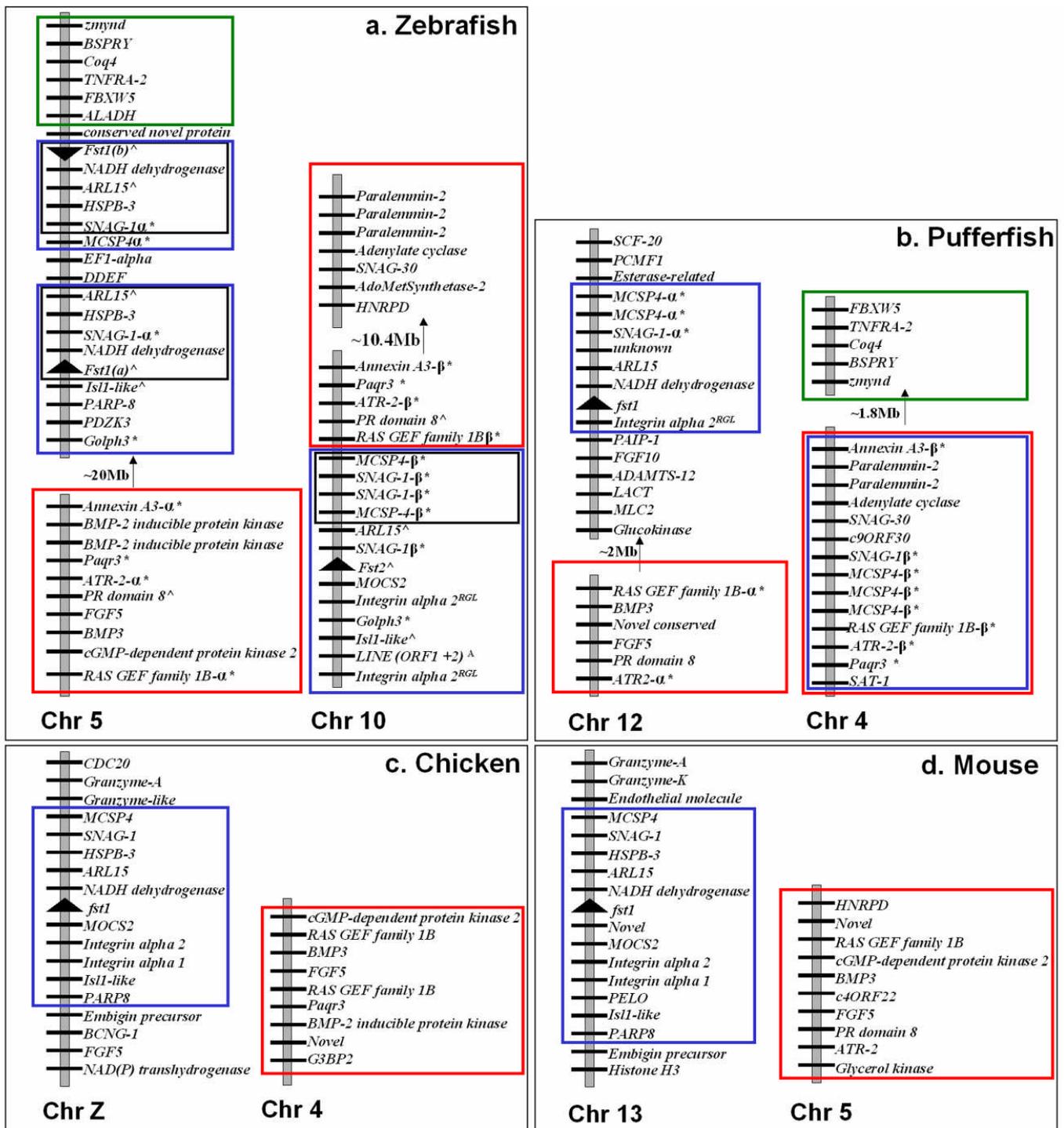
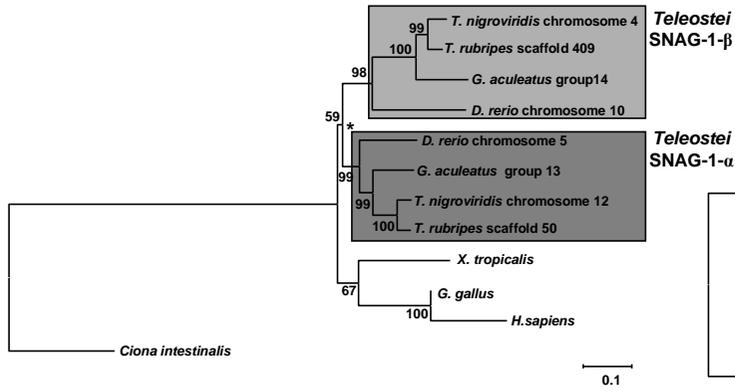


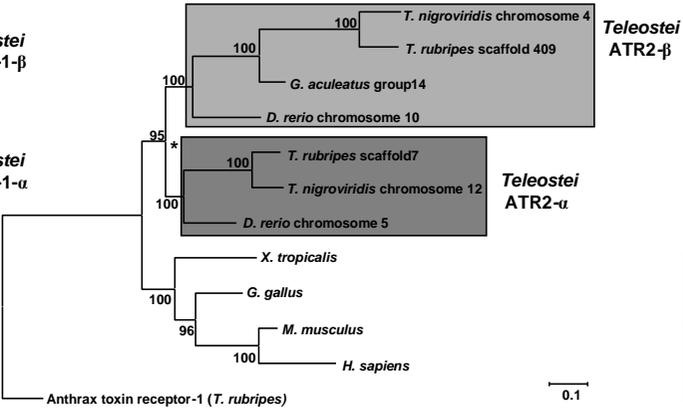
Fig. 5.2. Legend is on the next page

Fig. 5.2. Several duplicated genes in the neighbourhood of zebrafish *fst1/2* (**a**) were also present on two chromosomal regions in Acanthopterygian teleosts (*T. nigroviridis* shown) (**b**) each with double conserved synteny relative to a single region in the tetrapods *G. gallus* (**c**) and *M. musculus* (**d**). Genes retained on these two chromosomes in all teleosts are marked \*. Gene paralogues denoted as either  $\alpha$  or  $\beta$  were shown by phylogenetic reconstruction (see Fig. 5.3) to have duplicated after the split of the Actinopterygii and Sarcopterygii, likely during the teleost WGD. While paralogues always appeared on the corresponding chromosomes in different teleosts, intrachromosomal rearrangements have moved some genes outside the scale of the diagram for pufferfish e.g. for annexin A3 and *golph3*. Integrin alpha-2 was likely retained by reciprocal gene loss after the WGD (see discussion text) and is marked by <sup>RGL</sup>. Genes retained as duplicates in zebrafish, but not other teleost genomes are marked ^. The black boxes on zebrafish chromosomes 5 show inverted genomic regions. Coloured boxes identify corresponding blocks of conserved synteny between paralogous and orthologous chromosomes. Black arrowheads show the direction of *fst* transcription.

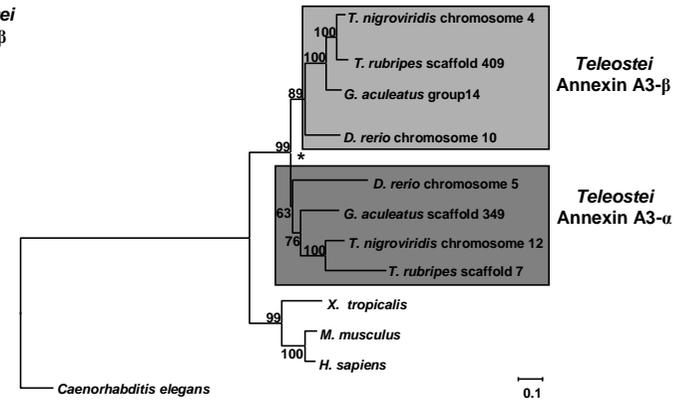
a. SNAG-1



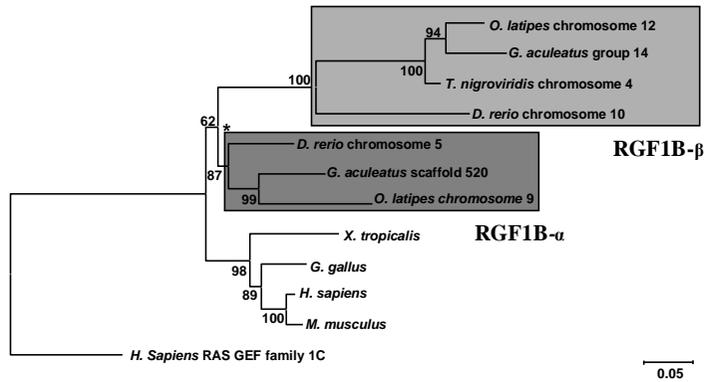
b. Anthrax toxin receptor-2 (ATR2)



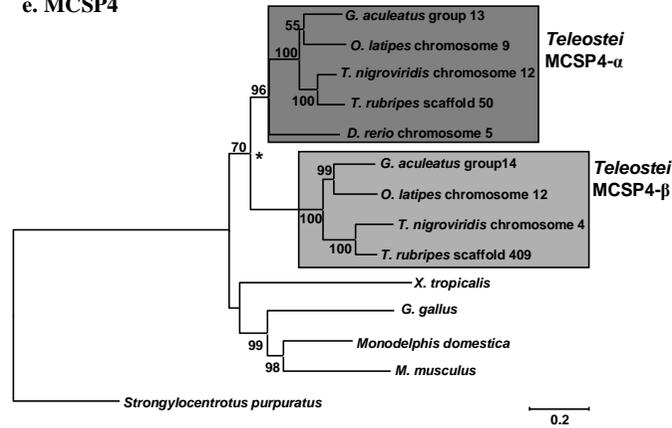
c. Annexin-A3



d. RAS GEF family 1B (RGF1B)



e. MCSP4



f. Integrin alpha-2 (reciprocal gene loss)

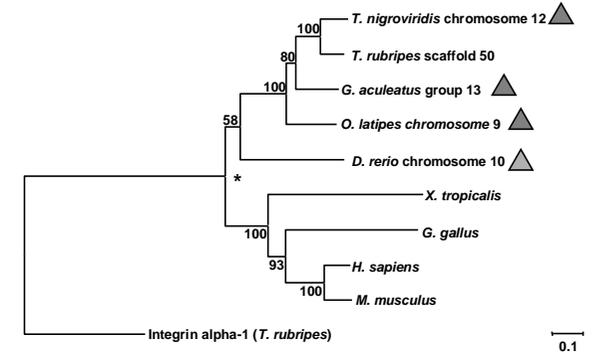


Fig. 5.3 Legend is on the next page

Fig. 5.3. **a-e**. Phylogenetic reconstruction of evolutionary relationships between duplicated genes in chromosomal proximity to *fst* in zebrafish and several Acanthopterygian teleosts. Trees were reconstructed using NJ analyses (in Mega 3.1 and ASATURA; see methods) of AA sequences retrieved from the Ensembl database for the following genes: **a**. *snag-1*, **b**. anthrax toxin receptor-2, **c**. annexin A3 **d**. RASGEF family member 1B **e**. *mcspl4*. For each tree, an outgroup was retrieved from Ensembl databases and used to root the tree. This was either an invertebrate orthologue (*C. intestinalis* for **a**, *Caenorhabditis elegans* for **c** and *Strongylocentrotus purpuratus* for **e**), or a gene from the same family (Anthrax toxin receptor-1 for **b**, RASGEF family member 1c for **e**). Teleost WGD paralogues conserved on orthologous regions to zebrafish chromosome 5 and 10 are respectively denoted as either  $\alpha$  or  $\beta$  for future nomenclature purposes (dark and light grey boxes respectively). **f**. Phylogenetic relationships of the single teleost integrin alpha-2 gene. The tree was rooted using a vertebrate AA sequence coding for integrin alpha-1 and shows that despite the abnormal chromosomal location of the zebrafish gene (on the opposite-paralogous chromosome compared to other teleosts), all teleost integrin alpha-2 genes form a single clade and are thus true orthologues. \* indicates the suggested timing of the WGD event. Bootstrap values greater than 50% are shown.

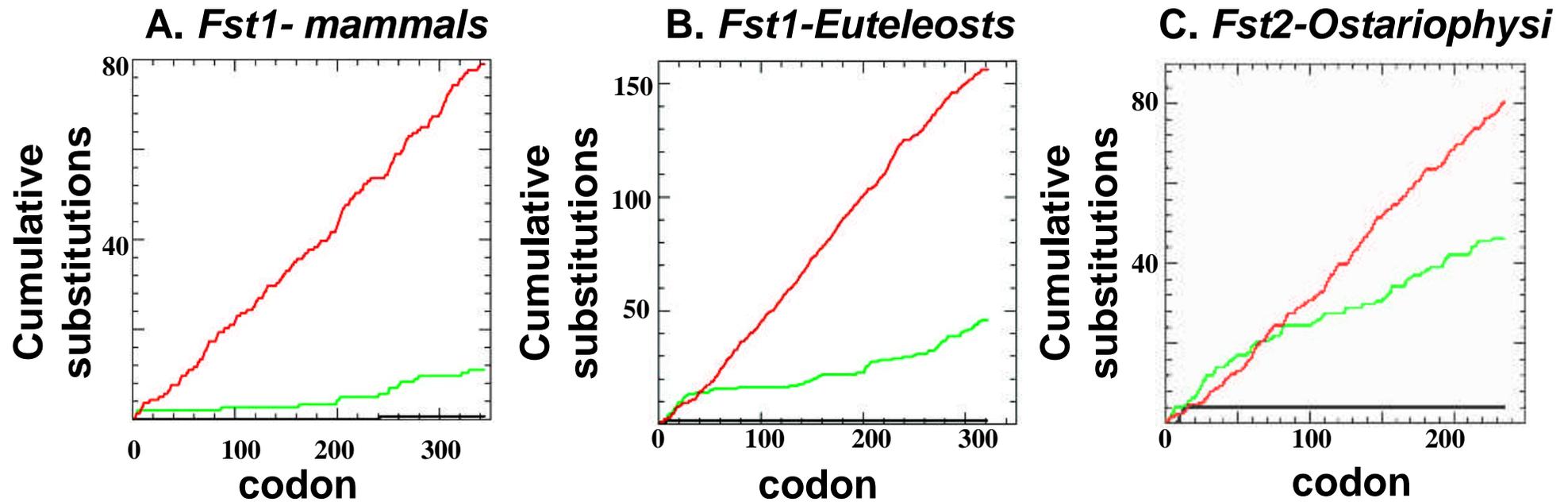


Fig. 5.4. Plots of the cumulative number of synonymous and non-synonymous substitutions (respective dotted red and solid green lines) across Fst proteins from 3 groups: **a.** Fst1 in mammals. **b.** Fst1 in euteleosts **c.** Fst2 in the Ostariophysi. The number of synonymous substitutions is generally high compared to non-synonymous substitutions across Fst1 of mammals and euteleosts. However the first 65 residues of Fst2 have accumulated more protein-changing substitutions than silent changes indicating that this region is under different selective constraints than Fst1 of teleosts and mammals and may have evolved under positive selection. Further, for the rest of the Fst2 protein, the difference in rate of accumulation of synonymous and non-synonymous changes is less exaggerated compared to Fst1, suggesting the protein has evolved under relatively relaxed constraints.

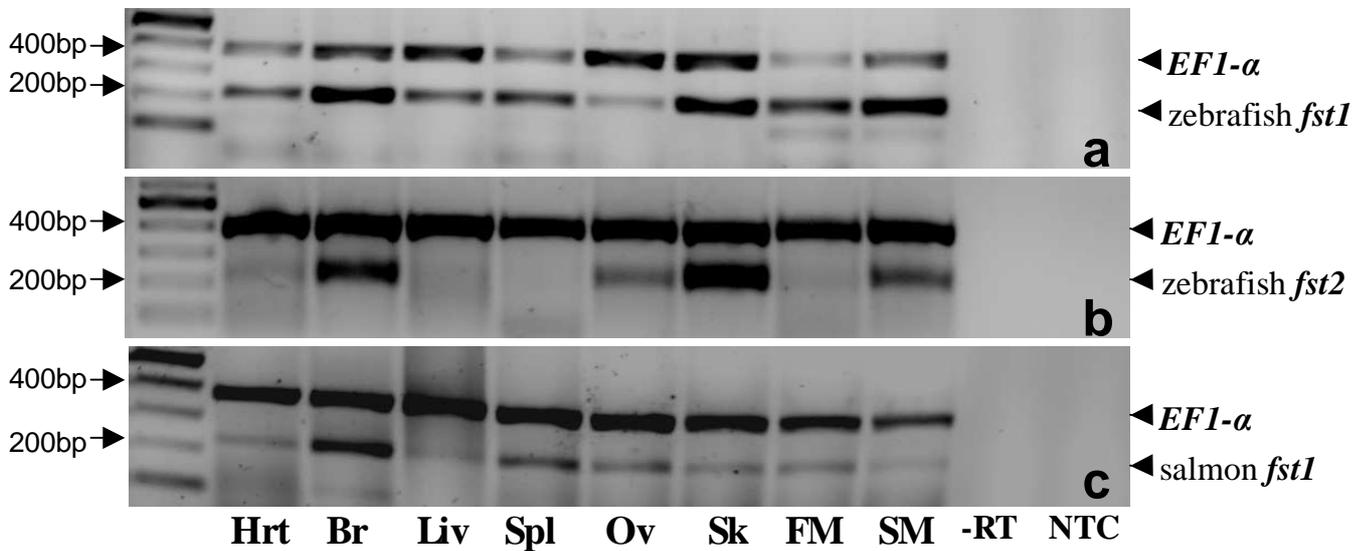


Fig. 5.5. Duplex RT-PCR analyses using the housekeeping gene *ef1-α* and primers specific to (a) zebrafish *fst1*, (b) zebrafish *fst2* and (c) salmon *fst1* in 8 different adult tissues: Heart (Hrt), brain (Br), liver (Liv), spleen (Spl), ovary (Ov) skin (Sk), fast-twitch muscle (FM) slow-twitch muscle (SM) Other lanes are no reverse transcriptase (-RT) and no template (NTC) controls. Bands corresponding to *ef1-α* and *fst* mRNA are shown to the right of the image by an arrowhead. The size of the DNA marker is shown to the left of the image.

**a**

```

fst1 1893 GAAGACCGCCCACACCACAGACAGTCGGCCCGAGACCCCTTATACACTTA
fst2 1906 GAAGACCGCCCAC--CACAGGCATCCAGCTCGAGACCCCATATAGATTTA
Frog 1917 GAAGACCGCCCACACCAGACagctaAGAC-----CCCCTTATAGATTTA
Mouse 1934 GAAGACCGCCCACACCAAACctcggAGAC-----CCCCGTCTAGATTTA
          ***** * * * * *

```

**b**

```

fst1 240 ACAAAAAGAAAGTGAATGAAT
fst2 3 -----
frog 213 CCAGAAACAAACAGAATGATT
Mouse 138 ACAGAAACAGACAAAATGAAT
          ** *** * * *****

```

**c**

```

fst1 1748 -----
fst2 1154 AAAATTAAAACATTTTTCTTATTTTTTTCATTTATTTTTTTA
frog 1280 ACACTTAACTACTGTTTCAGGTTCTGGGAGTTATATTTTTA
mouse 1213 -----TTAAACATTTTTGTTTGCTTCTGACTTGT-----
          * * * * * * * * * * * * * * * *

```

**d**

```

fst1 344 CAAAAGAGAAAAGGATTCAAGTCTGAAATT
fst2 7 -----
frog 258 CAAAAGGAACTATGAATTAAGTTAGGAATT
mouse 192 -----
          ***** * * * * * * * * * *

```

Fig. 5.6 The 2KB regions upstream of the first exon of zebrafish *fst1/fst2*, and *fst1* of frog (*X. tropicalis*) and mouse (*M. musculus*) were aligned using DiAlignTF (see methods). **a-d** are example extracts from the alignment showing conserved regions that existed between all species (**a**), between the promoter of zebrafish *fst1* and tetrapod *fst1*, but not zebrafish *fst2* (**b**), between the promoter of zebrafish *fst2* and tetrapods but not zebrafish *fst1* (**c**) and between the promoter of zebrafish *fst1/2* and either the frog or mouse promoter (**d**). The underlined red and blue font in **a** show two conserved TFBS's. The numbers to the left of the alignment show the position of the first nucleotide on that alignment relative to the 2000bp of promoter sequence (5'-3'). Stars and dashes respectively highlight conserved nucleotides and gaps in the alignment.

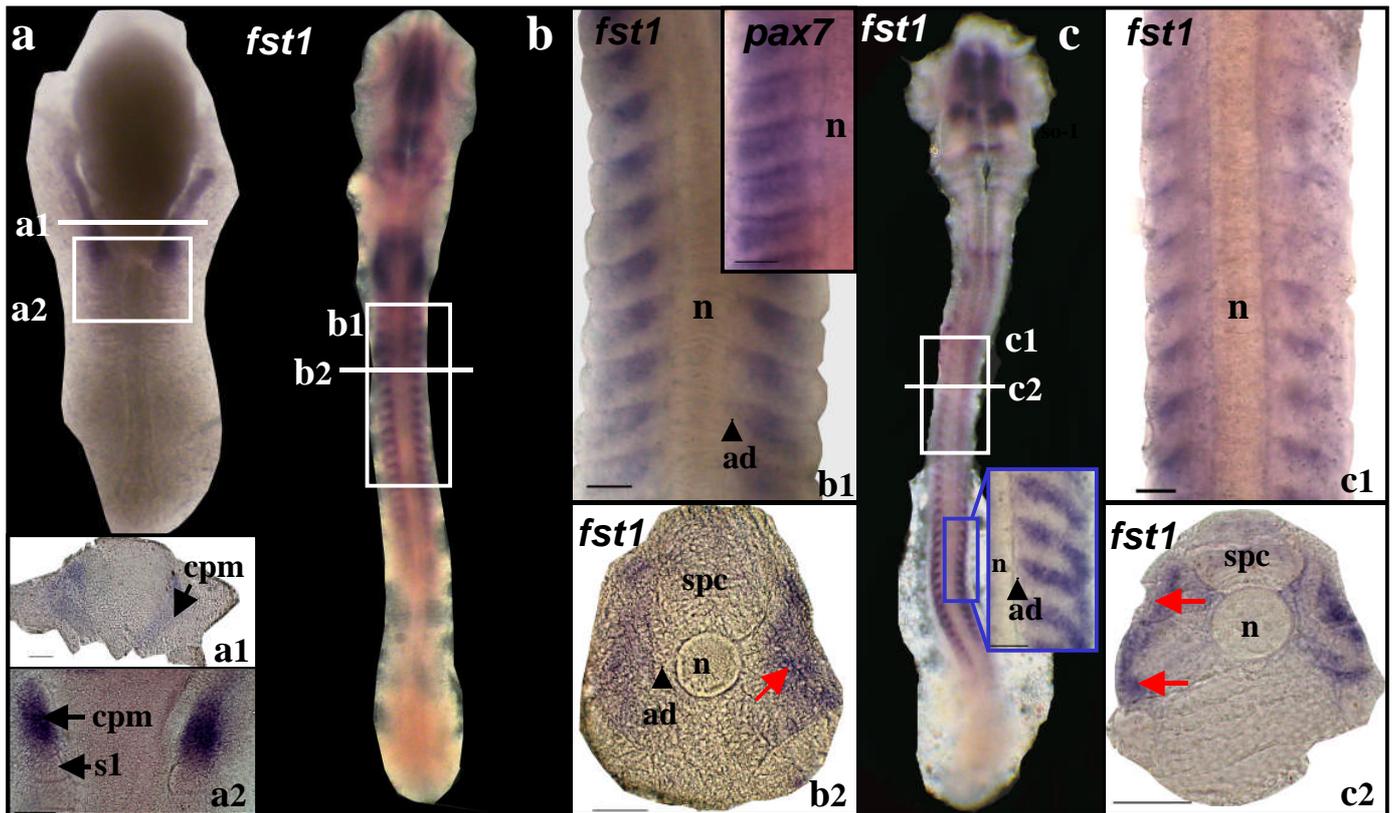


Fig. 5.7. Salmon *fst1* was expressed in the anterior compartment of the epithelial somite during the early segmentation period. **a.** *fst1* was first detected at the 5ss in the cephalic mesoderm and most rostral somite (a1-a2). **b.** Transcripts extended into the somites as segmentation proceeded (30ss shown) and expression was present throughout most of the anterior-lateral extent of most somites but was excluded from medial adaxial myoblasts (b1-b2). Also shown is the comparable expression of *pax7* in the anterior somite at this stage (labelled box in b1). **c.** At the 45ss, *fst1* expression was strongly present across the anterior region of caudal somites as observed for the 30ss (blue box shows dorsal flatmount), but in maturing somites was increasingly restricted to more lateral regions of the anterior myotome (c1-c2). White lines and boxes respectively show corresponding images of sections and dorsal flatmounts. Abbreviations: ad: adaxial myoblasts, cpm: cephalic mesoderm, n: notochord, s1: somite-1, spc: spinal cord. Scale bars are 50  $\mu\text{m}$ .

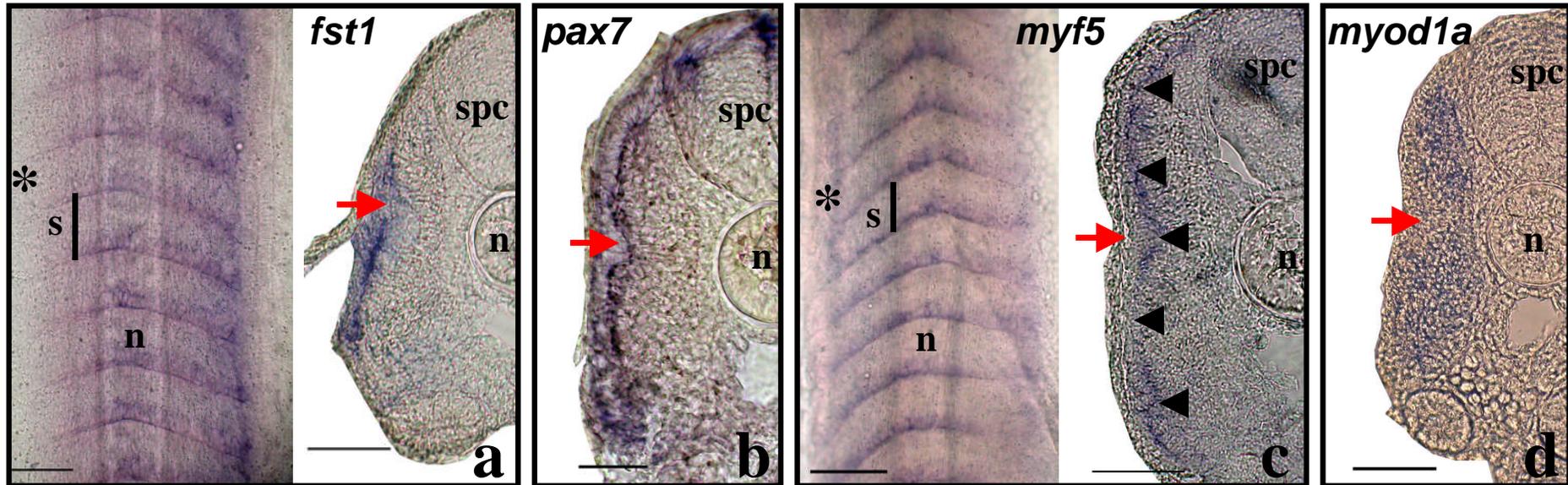


Fig. 5.8. **a.** *fst1* expression was mainly restricted to the external cell layer of the posterior somite at the end of salmon segmentation. **b.** *pax7* was expressed throughout the external cell layer along the length of the somite at this time. **c.** *myf5* expression was, like *fst1*, expressed along the posterior somite border, except that transcripts were detected at the boundary between the external cell-layer and lateral myotome (black arrowheads). **d.** At this time *myod1a* was expressed in the bulk of the myotome. Abbreviations are as in Fig. 5.7. \* shows the approximate somite at which the corresponding section was taken. Red arrows show the position of the external cell layer. Scale bars are 50  $\mu$ m.

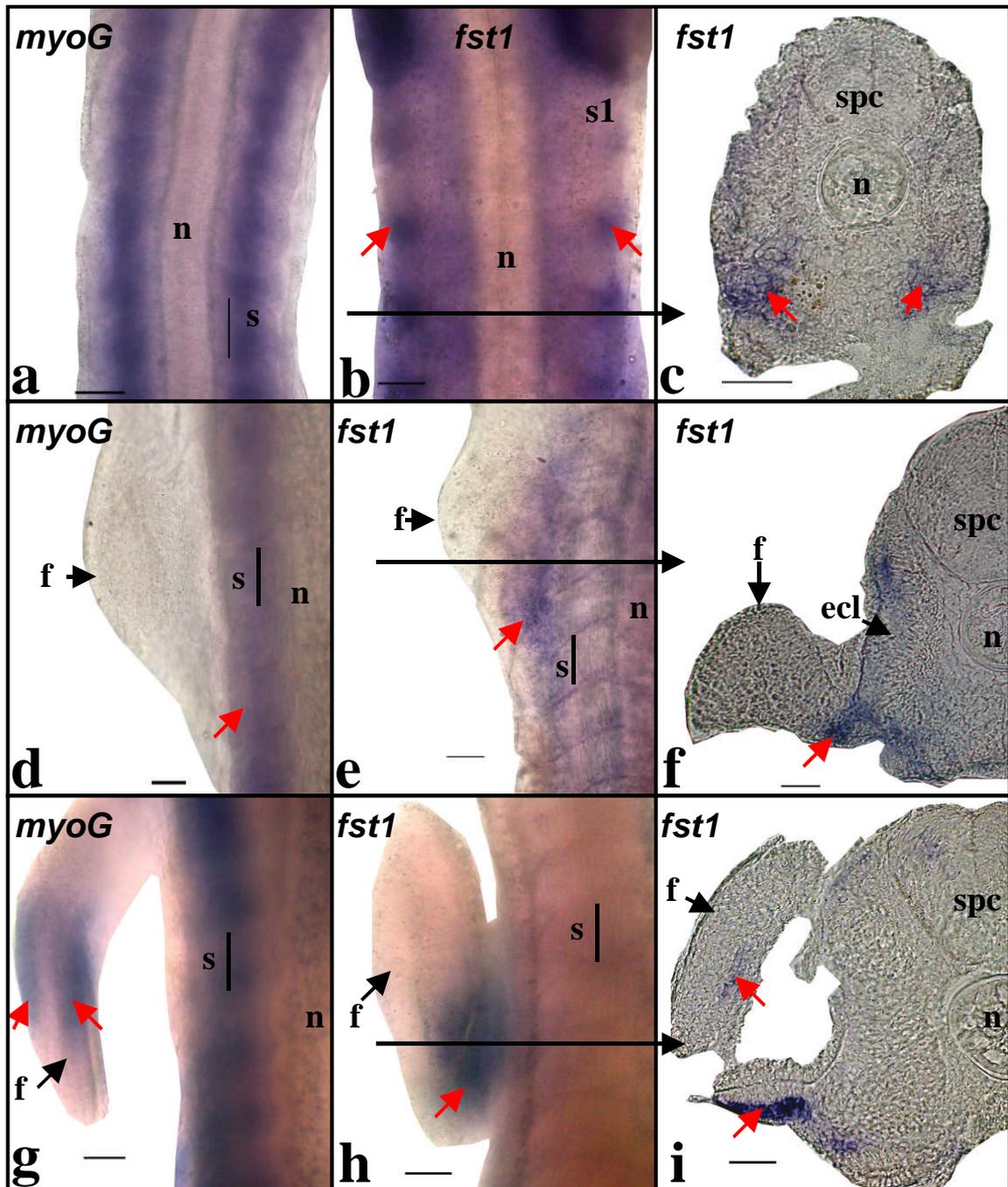


Fig. 5.9. The expression of *fst1* in rostral somites of salmon embryos during late segmentation suggests a role in pectoral fin bud myogenesis. **a**. At the 45ss, the muscle differentiation marker *myog* was expressed throughout the myotome. **b-c**. Concurrently, *fst1* was mainly restricted to the ventral-lateral somite. **d**. At the end of segmentation, when the fin buds were evident budding from somites 1-6, *myog* expression was still restricted to the myotome. **e-f**. At this stage, *fst1* expression was strongest in the ventral domain of the somite and within the external cell layer. **g**. Following segmentation, the fin buds expressed *myog* in the ventral and dorsal muscle masses. **h-i**. Concomitantly, *fst1* was restricted to the ventral somite but was also expressed within the musculature of the fin buds. Abbreviations are as in Fig. 5.7 except for f: fin bud, ecl: external cell layer. Black arrows demonstrate the level at which imaged sections were taken. Scale bars are 50  $\mu\text{m}$ .

## 5.5 Discussion

### 5.5.1 Two *Fst* genes were duplicated in a common teleost ancestor

The computational, phylogenetic and comparative-genomic results presented in this study provide strong support for a duplication of a *fst*-containing chromosome in a common teleost ancestor. The ML/NJ position of teleost *Fst2* external to teleost *Fst1* (100% bootstrap support) but internally to tetrapod sequences is consistent with a common duplication event in teleosts, rather than a lineage specific duplication. The position of *Fst2* was not altered by removing frequently mutating residues from the alignment prior to NJ tree reconstruction (94% bootstrap support), suggesting that the alignment was not affected by mutational saturation (Van de Peer et al., 2002). Furthermore, the presence of interleaved-double conserved chromosomal synteny of genes in neighbourhood to *fst* in teleost genomes relative to tetrapods is again, a clear signal of a common event in teleost evolution, the most parsimonious explanation being the teleost WGD (Jaillon et al. 2004). It was estimated that ~15-25% of paralogues have been retained from this event (Jaillon et al. 2004, Brunet et al. 2006) and the differential retention or loss of gene paralogues in different teleost lineages was estimated to be 50% when zebrafish and pufferfish genomes were compared, in two separate studies (Woods et al., 2005, Taylor et al., 2003). In this respect it should be noted that other genes in proximity to *fst* (*arl15* and *Isl1-like*) were similarly only retained as paralogues on the duplicated zebrafish chromosomes (Fig. 5.2 marked by ^). Conversely, in a previous study (Chapter 3), I found that two *myod* paralogues were retained in Acanthopterygian species, again as a consequence of the WGD, but lost in zebrafish.

### 5.5.2 Positive selection of *Fst2* relative to other *Fst* proteins?

In Fig 5.4 it was shown that mammalian and teleost *Fst1* proteins have evolved under strict purifying selection to avoid AA changing nucleotide substitutions throughout the whole propeptide. However, an increased number of non-synonymous (dN) compared to synonymous

substitutions (dS) exists in the N-terminal of Fst2 suggesting that this region has evolved under positive selection. Further, the rest of the Fst2 protein, while still subject to purifying selection as evidenced by a higher ratio of dS to dN substitutions (Fig. 5.4), has evolved under more relaxed constraints than Fst1 (compare Fig. 5.4, b with c). Residues 30-95 of mammalian Fst1 contain a 63 AA domain (the N-terminal domain) that performs an essential role for Activin binding (Thompson et al., 2005; Sidis et al., 2001). On an AA alignment of teleost and tetrapod Fst proteins, residues ~30-65 of teleost Fst2 fall within the N-terminal domain of human Fst1 (not shown). Thus, positive selection in residues 1-65 of Fst2 may have altered its intrinsic binding activity to Activin and by inference to other TGF- $\beta$  proteins such as Myostatin, compared to Fst1 of teleosts and mammals. This in turn could have contributed to its retention within the Ostariophysi.

### *5.5.3 Differential regulation of *fst1/fst2* paralogues suggests distinct evolution of regulatory regions*

Here it was shown by RT-PCR that zebrafish *fst2* has a restricted mRNA tissue distribution compared to *fst1* in zebrafish (its paralogue) and *fst1* of Atlantic salmon. Additionally, the two zebrafish genes are differentially expressed during embryogenesis (Bauer et al., 1998; Thisse et al., 2001; Dal-Pra et al., 2006). The restriction of zebrafish *fst2* expression suggests that some cis-acting regulatory regions shared by zebrafish *fst* and salmon *fst1* were lost in *fst2*. The teleost WGD event is believed to have occurred 320-350 MYA (Van de Peer, 2004), long prior to the separation of the Ostariophysi and euteleosts, some 150 MYA (Benton and Donoghue, 2007). Assuming that *fst* was duplicated during the WGD, then the regulatory regions of *fst1* and *fst2* would have had ~170-200 Mya to evolve independently in a common ancestor to zebrafish and salmon, before these lineages separated and *fst2* was lost in the latter. This means the regulatory regions of salmon and zebrafish *fst1* are separated by less evolutionary time, share a closer heritage and are likely to be under more similar selective constraints compared to zebrafish *fst2*. Accordingly, it could be expected that the tissue-specific expression of salmon *fst1* and zebrafish

*fst1* would be more similar than either gene to *fst2*. Additionally, the differential expression of zebrafish *fst1/2* genes in zebrafish embryos and adults could be a function of distinct regulatory elements. For example, binding sites for several important transcription myogenic factors, including MyoD1, Blimp1 and Paraxis were found in the promoter of *fst1* but were absent in the *fst2* promoter. These genes were each shown to be first expressed during zebrafish gastrulation (respectively: Weinberg et al., 1996; Shanmugalingam and Wilson, 1998; Baxendale et al., 2004). Similarly, zebrafish *fst1* was expressed from the onset of gastrulation (Thisse et al., 2001), whereas zebrafish *fst2* was not detected until mid somitogenesis (Dal-Pra et al., 2006) suggesting these binding sites may contribute to *fst1* transcription during early embryogenesis.

#### 5.5.4 Has subfunctionalization of *fst2* contributed to its retention in the Ostariophysi?

One mechanism by which duplicated genes are retained is through subfunctionalization, via the duplication-degeneration-complementation (DDC) model (Force et al., 1999). Under the DDC model it could be expected that important motifs in a single ancestral cis-acting regulatory region (here, any non-WGD basal Actinopterygii or Sarcopterygii lineage) should be conserved between two complementary regions in the WGD-daughter lineage (here, the Ostariophysi). A comparative analysis of putative zebrafish promoters with those of tetrapod genomes revealed three common transcription factor binding sites which are likely crucial to the normal regulation of *fst* across the vertebrates. Interestingly, while there were several other motifs, that were conserved between tetrapods promoters and both zebrafish *fst* paralogues, there were also numerous instances when a region was conserved between one paralogue and tetrapods, but not the other (Fig. 5.6). While these regions were not detected as conserved TFBS's their conservation suggests they may have a role in regulating *fst* expression. These results suggest that the function of zebrafish *fst1* and *fst2* may have subfunctionalized relative to tetrapods. However, to confirm this experimentally, the *in situ* expression of *fst* transcripts in a diploid

tetrapod embryo such as the mouse or chick would have to be compared to *fst1/fst2* in zebrafish embryos.

#### 5.5.5 A conserved role for vertebrate *fst* genes in regulating myogenesis?

In chick embryos, *fst* is expressed at the dermomyotomal lips in *pax3* expressing myogenic progenitors (Amthor et al., 2004). Results presented here suggest that salmon *fst1* is expressed in myogenic progenitors cells originating in the anterior somite, the teleost equivalent of the dermomyotome (Hollway et al., 2007). During early segmentation, *fst1* is expressed in the anterior epithelial somite concurrently to *pax7* (Fig 5.7), which is activated in zebrafish MPCs originating from this region (Stellabotte et al., 2007; Hollway et al., 2007). Conversely, the posterior domain of the epithelial somite expressed MRFs like *myod1b* and *myog* (chapter 4, chapter 7). These MRF expression domains extended anteriorly to encompass the whole rostral-caudal axis of the somite presumably marking the differentiation of embryonic fast muscle fibres (Stellabotte et al., 2007). As this occurred, *pax7* mRNA migrated from the anterior somite to encompass the external cell layer, a domain mirrored by *fst1* (Fig. 5.8). Concurrently, *myf5* was expressed in peripheral cells of the myotome (Fig. 5.8) and may mark the earliest differentiation of myogenic cells originating from the external cell layer. Additionally, *fst1* was expressed in the ventral domain of rostral somites during late-somitogenesis onwards (Fig. 5.9). In zebrafish it was shown that mesenchymal cells migrated from the ventral portion of somites 2-4 (particularly 4), to form muscles of the pectoral fin bud (Neyt et al., 2000). The anterior somite compartment of zebrafish was also shown to be a source of MPCs that contributed to pectoral fin musculature from somite 4 (Hollway et al., 2007). Zebrafish orthologues of amniote dermomyotome-expressed transcription factors, including *lhx1*, *mox1*, *pax3* and *dacA*, were progressively restricted to the ventral-lateral region of pectoral fin adjacent somites before accumulating in the fin bud musculature (Neyt et al., 2000, Hollway et al., 2007), as recorded here for *fst1* (Fig. 5.9).

### 5.5.6 Chapter conclusion

The results presented in this chapter are consistent with a common duplication of *fst* in teleosts, on a chromosomal scale indicative of the WGD. In the teleost orders sampled, only the Ostariophysi lineage has retained both *fst* paralogues, which may have diverged in protein function at the Activin-binding N-terminal and are regulated by two promoters that have evolved independently as evidenced by differential tissue mRNA distributions. Additionally, the expression pattern recorded for the single *fst1* gene of Atlantic salmon indicates an unappreciated role for teleost *fst1* in the regulation of the anterior somite cells destined to become muscle progenitors of the external cell layer and pectoral fin buds. These findings re-enforce the molecular and functional link conserved between the amniote dermomyotome and anterior somite of teleosts.

## **Chapter 6. Genomic, evolutionary and expression analyses of *cee*, an ancient and novel gene involved in normal growth and development**

### **6.1 Abstract**

In this chapter results are presented from experiments that involved characterising *cee* (conserved edge expressed protein), a gene that was recently discovered in a transcriptome-wide screen of mRNAs upregulated concurrent to the end of myotube production (Fernandes et al., 2005). Comparative genomic analyses indicate that *cee* arose 1.6-1.8 billion years ago (Bya) and remarkably shares no homology to any characterised gene family, and further, has no AA motifs with any assigned function. Accordingly, *cee* is found as a single copy in most eukaryotic genomes examined including all teleosts and mammals. The Cee protein is particularly conserved in the vertebrates, sharing more than 80% AA sequence identity between any two compared species. Further, the gene structure of *cee* is conserved as nine exons and eight introns in all vertebrates examined and more simply in non-vertebrate eukaryotes. The ratio of synonymous substitutions in the *cee* coding region far outweigh those that are AA changing indicating that the Cee protein has evolved under strong purifying selection. In Atlantic salmon embryos, *cee* is mainly expressed in the superficial layers of developing tissues and from the end of segmentation is also expressed concurrently to *pax7* in the external cell layer and to MRFs in zones of stratified hyperplasia. These data, together with functional screens in yeast and *Caenorhabditis elegans*, indicate that *cee* plays an important and uncharacterised role in the normal growth and development of eukaryotes, including myogenesis.

### **6.1 Introduction**

During teleost mosaic hyperplasia, MPCs fuse to form myotubes on the surface of existing fast muscle fibres giving rise to a mixture of fibre diameters in subsequent growth stages (Rowlerson

and Veggetti, 2001). This process continues until a genetically defined body length is attained, whereas the production of slow muscle fibres continues to occur in discrete zones until near to the final body size (Johnston et al., 2004). In the model pufferfish species *T. rubripes* a recent experiment in our laboratory used subtracted cDNA libraries to identify a number of candidate myotube inhibitory genes that were specifically upregulated in fast muscle concomitant with the cessation of myotube production in fast muscle (Fernandes et al., 2005). One of these genes, originally denoted *FRC386*, was of particular interest, since it translated into an uncharacterised protein that was found by a comparative search of eukaryote genomes to be strongly conserved in a wide range of taxa. *FRC386* mRNA transcripts in pufferfish were upregulated 15-fold specifically in fast muscle subsequent to the cessation of fibre recruitment and were present at concentrations more than five times greater than several other tissues (Fernandes et al., 2005). Further, large-scale RNAi screens in *Caenorhabditis elegans* revealed that disrupting the function of the *FRC386* orthologue resulted in a retardation of growth and development and sterility (Kamath et al., 2003; Simmer et al., 2003).

The first aim of this chapter was to clone the Atlantic salmon (*S. salar* L.) orthologue of *FRC386* and to investigate its embryonic mRNA expression pattern. Since *FRC386* mRNA was generally detected localized on the surfaces of specific tissues and organs during development, it was renamed *cee*, for conserved edge expressed protein. The next goal was to clone complete coding sequences of *cee* in other teleost species and to use this data, in conjunction with an additional 29 metazoan sequences retrieved by *in silico* mining of genome databases, to analyse the phylogeny, structure and evolution of Cee in multicellular animals.

## 6.3 Materials and Methods

### 6.3.1 Animals and sample collection

Tiger pufferfish (*T. rubripes*) and medaka (*O. latipes*) samples for this experiment were provided by Dr Jorge Fernandes (JMOF) and zebrafish (*D. rerio*) were obtained by Mr Hung-Tai Lee (HT-L). A wild-caught tiger pufferfish of 1.4 kg was purchased from a local fish market in Maisaka (Shizuoka, Japan) and medaka/zebrafish adults were respectively bred in captivity at the Ocean Research Institute (The University of Tokyo, Japan) and at the Gatty Marine Laboratory (University of St Andrews, UK). Additionally, two adult Atlantic salmon (*S. salar*) (body mass of 3.8 and 4.0 kg) were obtained from EWOS Innovation (Lønningdal, Norway). Pure dissections of fast muscle were either stored in RNAlater or flash frozen in liquid nitrogen. For RT-PCR tissue analysis, samples of heart ventricle, brain, liver, spleen, intestine, ovaries, skin, fast muscle and slow muscle were also dissected from the two Atlantic salmon and stored in liquid nitrogen. Atlantic salmon embryos were part of the 10°C group described in chapter 2 (section 2.2) and the stages (described in section 2.2.2) were sampled and prepared as described therein.

### 6.3.2 Computational identification of *cee* orthologues

The *in silico* mining of *cee* orthologues described here was performed by JMOF. The putative translation product of *FRC386* from *T. rubripes* (GenBank accession CK829928) was used as a probe in tBLASTn searches to identify homologous metazoan sequences in the following databases: non-redundant sequence and EST databases at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>), Uniprot (<http://www.expasy.uniprot.org/>), WormBase (<http://www.wormbase.org/>) and Ensembl (<http://www.ensembl.org/>). These predictions were further analysed with the gene structure prediction software Genebuilder (<http://125.itba.mi.cnr.it/~webgene/genebuilder.html>) and manually refined. For comparison

purposes, *cee* orthologues were also retrieved in the yeast (*Saccharomyces cerevisiae*), the social amoeba (*Dictyostelium discoideum*) and several other protists (*Leishmania major*, *Plasmodium* sp. and *Trypanosoma* sp.) from SGD (<http://www.yeastgenome.org/>; accession reference YOR164C), dictyBase (<http://dictybase.org/>; accession reference DDB0218329) and the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/>), respectively. The *Giardia lamblia* (<http://gmod.mbl.edu/>), *Euglena gracilis* (<http://tbestdb.bcm.umontreal.ca/>) and microbial databases at NCBI were also screened for *cee* orthologues.

### 6.3.3 Cloning and sequencing of *cee* cDNAs

A whole coding sequence of *cee* was obtained for four teleost species: Atlantic salmon (by DJM), medaka, tiger pufferfish (both by JMOF) and zebrafish (by JMOF and HT-L). Total RNA was extracted from fast muscle, quantified and used for cDNA synthesis as described in chapter 2 (section 2.4.5). Controls lacking reverse transcriptase were also included. To amplify full coding sequences of *cee*, 1 µl of cDNA from Atlantic salmon, medaka, tiger pufferfish or zebrafish was used in standard PCR reactions containing the following respective primer pairs *cee-Ss* 1, *cee-Ol* 1, *cee-Tr* 1 and *cee-Dr* 1 (Table 6.1). PCR products were cloned and plasmids were extracted, purified and sequenced as described in chapter 2 (sections 2.4.9-2.4.11).

### 6.3.4 Sequence alignments and intron-exon structure of *cee*

The sequence analyses described here were performed by JMOF. Nucleotide sequences were translated with DNAMAN (Lynnon Biosoft) and the putative proteins aligned by CLUSTALW at the Kyoto Bioinformatics server (<http://align.genome.jp/>) using a BLOSSUM matrix with default parameters and then manually optimised. Pairwise protein sequence comparisons were performed with BioEdit (Hall, 1999). The ScanProsite software (<http://www.expasy.ch/prosite/>) was used to identify structural and functional motifs in the Cee protein sequences.

The intron/exon structures of *cee* in various species were determined with Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>) (Wheelan et al., 2001) by JMOF and DJM. The construction of the intron-exon figure was performed by DJM.

#### *6.3.5 Phylogenetic inference and tests of selection*

Phylogenetic reconstruction, and assembly of the corresponding Figures was performed jointly by JMOF and DJM. Bayesian inference of phylogeny was performed with MrBayes (Ronquist and Huelsenbeck, 2003) with a mixed AA model of protein evolution and considering a uniform rate of among site substitution rates. 1,000,000 generations were used with sampling every 100 generations and a burnin value equivalent to 100,000 generations (1000 trees). The runs had converged by 100,000 generations and a majority consensus tree with Bayesian posterior probabilities was constructed from the final 9,000 trees. A concurrent ML analysis was performed with PhyML (Guindon and Gascuel, 2003) using a WAG model of AA evolution and assuming a gamma distribution of among site substitution rates. The reliability of this tree was tested using a bootstrap test with 500 pseudoreplicates. Neighbour-joining (AA model with Poisson correction) and maximum parsimony trees were reconstructed in MEGA 3.1 (Kumar et al., 2004) using the JTT model and considering among-site substitutions rates to be uniform using 10,000 bootstrap replicates to assess branch confidence.

Tests to establish the selective constraints on Cee proteins during metazoan evolution were performed by JMOF. Coding sequences corresponding to *cee* in various species were retrieved (databases listed in Table 6.2) and grouped by taxon, as follows: nematodes, platyhelminthes, insects, teleosts, amphibians, tunicates and mammals. PAL2NAL (Suyama et al., 2006) was used to align the coding sequences within each group according to the respective protein sequence alignment. The average number of synonymous (dS) and non-synonymous (dN) substitutions,

insertions and deletions in the codon alignments were determined with SNAP (Korber et al., 2000).

#### 6.3.6 RNA probe preparation and whole-mount *in situ* hybridization

All procedures relating to the *in situ* hybridisation of *cee* in Atlantic salmon embryos were performed by DJM. A 1157 bp amplicon containing 189 bp of the 5' untranslated region and the full coding sequence of Atlantic salmon *cee* was amplified by PCR using the primer pair *cee-Ss 2* (Table 6.1) and cloned as described in chapter 2 (section 2.6.2). This PCR product was used to synthesize sense and anti-sense DIG-labelled *cee* RNA probes as described in chapter 2 (section 2.6.3). Whole mount *in situ* hybridization was performed as described in chapter 2 (sections 2.6.4-2.6.5) using six embryos per developmental stage studied. Embryos were either flatmounted on a glass microscope slide or cryosectioned and then photographed as described in chapter 2 (section 2.6.6).

#### 6.3.7 Tissue distribution of *cee* mRNA in adult salmon.

DJM performed all experiments to establish the mRNA tissue distribution of *cee* in adult Atlantic salmon tissues. Total RNA was extracted from the nine Atlantic salmon tissues described above (section 6.2.1) and was used to synthesise cDNA using the method detailed in chapter 2 (section 2.4.6). To assess the presence of *cee* mRNA relative to the housekeeping gene *ef1- $\alpha$*  a duplex RT-PCR assay was used with the concurrent use of the primers *cee-Ss-3* and *ef1- $\alpha$ -Ss* (Table 6.1). 1 $\mu$ l of cDNA from each tissue was used as a template and amplifications were performed using the following thermocycling parameters: initial denaturation for 5 min at 95 °C, then 35 cycles for 30s at 95 °C, 30s at 60 °C, 30s at 72 °C, and a final extension for 5 min at 72 °C. The PCR was ran for both fish identically, and in each case no template controls were included.

## 6.4 Results

### 6.4.1 *In silico* identification of *cee* in eukaryote genomes

*cee* was first discovered in a previous study as a gene consistently upregulated in the fast muscle of tiger pufferfish of a body size that had stopped producing new myotubes, compared to smaller fish in a growth phase of active muscle fibre recruitment (Fernandes et al., 2005). The original clone containing *cee* was denoted *FRC386* (CK829928) and preliminary analyses revealed that it was both uncharacterised and widely conserved in eukaryotes. Following *in situ* hybridization experiments in Atlantic salmon embryos, we decided to call this gene *cee* (conserved edge expressed protein), based on its developmental expression pattern (see below section 6.4.6).

Exhaustive BLAST searches were performed to identify *cee* in multiple eukaryote cDNA and genome databases. With two exceptions, a single *cee* gene was present in all metazoan taxa examined (Table 6.2), including insects (yellow fever and malaria mosquitoes [respectively: *Aedes aegypti* and *Anopheles gambiae*] European honey bee [*Apis mellifera*], fruit flies [*Drosophila* sp.] and red flour beetle [*Tribolium castaneum*]), nematodes (*Caenorhabditis* sp.), platyhelminthes (*Schistosoma* sp.), echinoderms (purple sea urchin: *S. purpuratus*), teleosts (tiger pufferfish, green-spotted pufferfish, medaka, Atlantic salmon, stickleback and zebrafish), the amphibian Western clawed frog (*X. tropicalis*), birds (chicken [*G. gallus*]), tunicates (*C. intestinalis* and *C. savignyi*) and mammals (human [*H. sapiens*], chimp [*P. troglodytes*], macaque [*M. mulatta*], mouse [*M. musculus*], rat [*R. norvegicus*], pig [*S. scrofa*], guinea pig [*Cavia porcellus*], shrew [*Tupaia belangeri*], cow [*B. taurus*], dog [*C. familiaris*], cat [*F. catus*], elephant [*Loxodonta Africana*], opossum [*M. domestica*], platypus [*O. anatinus*], bushbaby [*Otolemur garnettii*], armadillo [*Dasypus novemcinctus*], European hedgehog [*E. europaeus*] Lesser hedgehog tenrec [*Echinops telfairi*] and microbat [*Myotis lucifugus*]. Two *cee* sequences that shared 95% identity at the nucleotide level were found in the African clawed frog *X. laevis*

(Table 6.2). Additionally, the mosquito *A. aegypti* had two paralogues that were present on distinct chromosomal regions (AAEL002521 in supercont1.59 and AAEL012936 in supercont1.765) but their coding sequences were 99.5% identical and coded for an identical AA sequence.

No apparent orthologue of *cee* was identified in archaea and eubacterial genomes. To gain insight into the probable evolutionary origin of *cee*, all available protist genomes were also screened. *Cee* was not conserved in the amitochondriate eukaryote *Giardia lamblia*, which occupies a basal position in the phylogeny of protists and is thought to have diverged from other eukaryotes around 2.2 billion years ago (Bya) (Hedges et al., 2001). Additionally, *cee* orthologues were not identified in the genomes of euglenida (*Euglena gracilis*) and kinetoplastida euglenozoans (*Leishmania major*, *Trypanosoma brucei*, *T. vivax* and *T. congolense*). According to recent studies regarding the phylogeny of protists (see Hedges, 2002) the most primitive eukaryotes where *cee* was present were the alveolata. This taxon (phylum Apicomplexa) includes the malarial parasites *Plasmodium berghei* (GenBank XM\_675171), *P. chabaudi* (XM\_730628), *P. falciparum* (XM\_001348503), *P. yoelii* (XM\_721059) and the tropical theileriosis parasite *Theileria annulata* (XM\_950161). *cee* was also found in amoebzoa (*D. discoideum*, XM\_635525), fungi (*S. cerevisiae*, YOR164C) and plants (*Arabidopsis thaliana*, AK176227). Taken together, these results suggest that *cee* arose sometime after the most recent symbiotic event from which mitochondria arose in eukaryotes (which is thought to have occurred around 1.8 Bya: Hedges et al., 2001) and prior to the divergence of animals/fungi and plants, which is thought to have occurred some 1.6 Bya (Blair et al., 2005).

Additionally, complete coding sequences for *cee* were obtained experimentally in four teleost fishes from the orders salmoniformes (Atlantic salmon), beloniformes (medaka), cypriniformes (zebrafish) and tetraodontiformes (tiger pufferfish). These nucleotide sequences were submitted

to GenBank as *cee*, in conformity to the guidelines proposed by the zebrafish nomenclature committee (accession numbers shown in Table 6.2). For further characterization and evolutionary analysis of *cee* in metazoans, only complete coding sequences derived from high quality predictions or with experimental support were used (Table 6.2).

#### 6.4.2 Characterisation of *cee* in metazoans

The putative protein coded by *cee* ranged in size from 307 residues in chicken to 362 AAs in *C. elegans*. There was a notable degree of conservation between Cee orthologues from different vertebrate taxa (Fig. 6.1), which shared an overall identity of at least 80% when any two species were compared (not shown). Despite this high level of similarity among Cee orthologues, no recognised motifs or conserved domains could be identified beyond the domain assigned the name DUF410, which basically corresponds to the majority of the Cee protein, barring the 40-50 most N and C terminal residues. Using the Conserved Domain Architecture Retrieval Tool (Geer et al., 2002) at NCBI (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>) the DUF410 domain was found to be present in 85 eukaryotic proteins, each orthologous to *cee*. In vertebrates, differences within Cee are generally distributed throughout the entire protein and many of the AA substitutions were isofunctional replacements (Fig. 6.1). The regions of highest variability when all taxa are considered corresponded to the extreme N and C termini of Cee and to residues 91-102 and 179-186 in the zebrafish sequence (Fig. 6.1). Interestingly, the predicted chicken Cee protein (which is derived from an experimental sequence) has a deletion in a region (residues 151-166) that is entirely conserved in other vertebrates (Fig. 6.1). The primary structure of Cee from invertebrates was rather more diverse and shared approximately 30 to 40% sequence identity at the protein level with their mammalian orthologues (alignment not shown here: see Fernandes, Macqueen et al., 2007: supplementary Fig. S1, supplementary Table S2). Additionally, the sequence identity conserved between vertebrate Cee proteins and their orthologues in *P. chabaudi*, *D. discoideum*, *S. cerevisiae* and *C. elegans* were 19, 29, 26 and

23%, respectively. Despite the relatively low degree of similarity among invertebrate Cee orthologues, a conserved region corresponding to residues 39-52 could be identified within the invertebrate sequences (see Fernandes, Macqueen et al., 2007: supplementary Fig. S1). This domain was also highly conserved in all vertebrate species except the platypus, which has six substitutions within this region (Fig. 6.1). It is noteworthy that the motif YYEAHQ was also present in *Plasmodium* Cee (not shown), suggesting that this might be an ancient functional domain.

#### 6.4.3 Gene structure of *cee*

The genomic structure of *cee* was identical amongst all the vertebrate orthologues examined and comprised nine exons and eight introns that ranged from 3.6 Kb in tiger pufferfish to 43.6 Kb in zebrafish (Fig. 6.2). The lengths of all exons and the location of their splice junctions were strongly conserved across the vertebrates. Despite some diversity of intron sizes within vertebrates, intron I-II was generally the largest, with the exception of zebrafish, where intron VII-VIII spanned approximately 30Kb (Fig. 6.2). The intron/exon structure of the *cee* gene was not conserved amongst invertebrates and its complexity varied from two exons in the yellow fever mosquito to up to eight exons in the purple sea urchin *S. purpuratus* (Fig. 6.2). In this echinoderm, exons one to six and exon eight are of a similar size to vertebrate counterparts and further, splice sites are conserved between exons 1/2, 2/3, 4/5, 5/6 and 6/7, as are the exon/ intron boundaries between the last two exons. Further, the size of exon seven in the purple sea urchin (149 bp) is roughly equivalent to the combined sizes of exons seven and eight in vertebrates. Conversely the structure of the *cee* gene in the tunicate *C. intestinalis* shares no similarity with vertebrate *cee* and is represented as an intronless gene, comparable to bakers yeast (Fig. 6.2).

#### 6.4.4 Phylogenetic reconstruction of *Cee*

The phylogenetic reconstruction of *Cee* orthologues was performed using ML, Bayesian inference, maximum parsimony and NJ. Fig 6.3 shows the ML reconstruction with supporting Bayesian posterior probability values. All methods of reconstruction produced similar topologies and most branches had strong bootstrap support. The branch lengths in all trees were relatively short, particularly within vertebrate clades (Fig. 6.3), reflecting the strength of *Cee* conservation amongst these taxa. *Cee* from coleopteran (beetles), hymenopterans (bees) and dipterans (flies and mosquitoes) formed a monophyletic group as did vertebrate *Cee* orthologues (Fig. 6.3). However, the vertebrate clade branched internally from sea urchin, which in turn branched internally to tunicates (*C. sp.*), which does not support accepted taxonomic relationships among the deuterostomes (e.g. Delsuc et al., 2006). The topology of the teleost branching followed the currently accepted phylogenetic relationship between cypriniformes, salmoniformes, beloniformes and tetraodontiformes (Nelson, 2006). It can also be seen that the two *Cee* sequences of *X. laevis* branched from the *X. tropicalis* orthologue as sister sequences (Fig. 6.3). Additionally the two *X. laevis cee* genes have evolved asymmetrically and the paralogue we have designated B (GenBank BC074468) is less derived relative to the ancestral gene (Fig. 6.3, note its longer branch length).

#### 6.4.5 Selective constraints on *Cee* during animal evolution

Next we established the cumulative number of non-synonymous (dN) and synonymous substitutions (dS) across *cee* coding regions and calculated their dN/dS ratios in the following metazoan taxa: insects (*A. aegypti*, *A. gambiae*, *A. mellifera*, *D. melanogaster*, *D. pseudoobscura* and *T. castaneum*), platyhelminthes (*S. japonicum* and *S. mansoni*), nematodes (*C. briggsae*, *C. elegans* and *C. remanei*), teleosts (*D. rerio*, *G. aculeatus*, *O. latipes*, *S. salar* and *T. rubripes*), amphibians (*X. laevis* and *X. tropicalis*) and mammals (*B. taurus*, *C. familiaris*, *H. sapiens*, *M. mulata*, *M. domestica*, *M. musculus*, *O. anatinus* and *R. norvegicus*). In teleosts (Fig. 6.4, A),

mammals (Fig. 6.4, B) and amphibians (not shown) the number of synonymous substitutions per site was approximately constant in all exons and much higher than the number of non-synonymous mutations in teleosts and mammals. The coding sequences of *cee* in nematodes, platyhelminthes and insects had an overall higher number of synonymous substitutions compared to vertebrate taxa, which corresponded to a lesser degree of *cee* conservation within these taxonomic groups (not shown). The average dN/dS ratios of all pairwise comparisons in insects, platyhelminthes, nematodes, teleosts, amphibians and mammals were respectively 0.09, 0.06, 0.07, 0.04, 0.02 and 0.03.

#### 6.4.6 Developmental expression of *cee*

Next *in situ* hybridization was used to establish the mRNA expression of *cee* throughout the embryonic development of Atlantic salmon. Sense controls were used for all stages and never produced specific staining. A signal for *cee* mRNA was absent in salmon at the end of gastrulation and from the 0-10 ss (not shown). However, *cee* was consistently detected throughout segmentation from the 25 ss onwards (Fig. 6.5, A-L). Although the pattern of staining often appeared to be more or less ubiquitous (whole-mount embryos were often a homogenous purple colour), flatmounting and sectioning revealed more specific sites of expression. At the 30ss, *cee* was strongly expressed along the entire cranial-caudal axis of each embryo in three stripes marking the lateral edges and midline of the entire brain and neural tube/spinal cord (Fig. 6.5, A, sense control included for comparison). Cross sections revealed that this staining extended through the entire dorsal-ventral axis of the brain and developing spinal cord (Fig. 6.5, B, F). Concurrently, *pax7*, a marker of the development of neural tissues as well as muscle (Lang et al., 2007) was also expressed at the midline and lateral edges of the dorsal spinal cord but was less restricted than *cee*, being present to a greater or lesser extent throughout the width of this region (Fig. 6.5, F<sub>1</sub>). By the eyed stage, *cee* transcripts were still present in the spinal cord but were no longer restricted to the edge and midline, and co-localized with *pax7* across the spinal

cords dorsal width (not shown). At this stage, *cee* was also expressed as a broad band surrounding the superficial edge of the entire cranial region (Fig. 6.5, H) and at the borders of several structures within the developing brain (not shown). During somitogenesis *cee* was expressed at the boundaries of the somites as they developed from simple oval shaped structures (Fig. 6.5, D, marked by blue arrowheads, sense control included for comparison) to chevron shaped structures with elongated muscle fibres at the eyed stage (Fig. 6.5, K). *cee* transcripts were also detected diffusely throughout the myotome and ventral regions of somites during most of the segmentation period (Fig. 6.5, F). A longitudinal section through the epithelial caudal somites at the 45 ss revealed that staining was mainly clustered between cells at the superficial-lateral border of the somite (Fig. 6.5, E, black arrows). *cee* was down-regulated in the medial myotome during late and post-segmentation stages when elongated muscle fibres were clearly present, but was expressed at the outer edge of the myotome and somite, particularly in dorsal and lateral regions as well as the external cell-layer (Fig. 6.5, K, L). During this time *pax7* mRNA was expressed throughout the external cell layer (Fig. 6.5, L<sub>1</sub>), whereas *myog* (Fig. 6.5, L<sub>2</sub>) and other MRFs (chapter 7, Fig. 7.6) were expressed at the lateral edge of the myotome particularly in dorsal and ventral regions and at the level of the horizontal myoseptum in zones of new muscle fibre production. Thus, *cee* is expressed concomitantly with *pax7* in the external cell layer and MRFs in the myotomal compartment (Fig. 6.5, L).

As the eye developed during segmentation and post-segmentation stages, *cee* was expressed on the innermost and outermost surfaces of the retina, bordering the lens and retinal epithelium, respectively (Fig. 6.5, A, C, D). *cee* transcripts could be found at the boundaries of several other structures throughout their embryonic development, such as the otolith nuclei (Fig. 6.5, G), branchial arches and fin buds (not shown). At the eyed stage, unrestricted *cee* staining was also present throughout several structures of the developing gut (Fig. 6.5, J). Consistent with its expression in multiple tissues during Atlantic salmon embryonic development, *cee* transcripts

were present in all eight tissues examined by RT-PCR (Fig. 6.5, M). This pattern was identical in both fishes tested and the strongest *cee* cDNA band was observed in liver, and the weakest in brain (Fig. 6.5, M).

Table 6.1. List of primer pairs used to amplify *cee* and corresponding amplicon sizes. This table was modified from Fernandes, Macqueen et al. (2007).

Primer Pair	Forward Primer (5'→3')	Reverse Primer (5'→3')	Size (bp)
<i>cee-Dr</i> 1	GTTCGGTTGGTCGGAGCAG	TTAATTATGCTCAATCACACCTC	1715
<i>cee-Ol</i> 1	GCGGAGAAGGATCGACCATGTC	CGGCTGTTGGGTCAGTCCAG	1000
<i>cee-Ss</i> 1	ATGTCGGAGCAGGAGGCTCTG	TCAGTCCAGCTCAATGGGGC	969
<i>cee-Tr</i> 1	GCAACGATGTCGGAACAAGAATC	TTTATCTTTGTCCTGAGGTGGG	1001
<i>cee-Ss</i> 2	GAACGGATGCTCAGCTTTATAGC	TCAGTCCAGCTCAATGGGGC	1158
<i>cee-Ss</i> 3	ACGCAGAAACACCCCTCAATAG	CTTCCTCTCCCTCCTCATCCTC	292
<i>ef1a-Ss</i>	GAATCGGCTATGCCTGGTGAC	GGATGATGACCTGAGCGGTG	141

Table 6.2. Accession numbers for metazoan *cee* genes and putative proteins studied in this chapter. This table was adapted from Fernandes, Macqueen et al. (2007).

<b>Species</b>	<b>Nucleotide</b>	<b>Protein</b>
<i>Aedes aegypti</i>	DV237030 <sup>a</sup>	EAT34855 <sup>f</sup>
<i>Anopheles gambiae</i>	ENSANGG00000017446 <sup>b</sup>	ENSANGP00000019935 <sup>b</sup>
<i>Apis mellifera</i>	XM_395262 <sup>a</sup>	XP_395262 <sup>f</sup>
<i>Bos taurus</i>	NM_001076525 <sup>a</sup>	NP_001069993 <sup>f</sup>
<i>Caenorhabditis briggsae</i>	CBG18187 <sup>c</sup>	BP:CBP19167 <sup>c</sup>
<i>Caenorhabditis elegans</i>	U12964 <sup>a</sup>	Q19824 <sup>g</sup>
<i>Caenorhabditis remanei</i>	cr01.sctg26.wum.38.1 <sup>c</sup>	cr01.sctg26.wum.38.1 <sup>c</sup>
<i>Canis familiaris</i>	ENSCAFG00000011309 <sup>b</sup>	ENSCAFP00000016657 <sup>b</sup>
<i>Ciona intestinalis</i>	ENSCING00000007232 <sup>b</sup>	ENSCINP00000014826 <sup>b</sup>
<i>Ciona savignyi</i>	ENSCSAVG00000006834 <sup>b</sup>	ENSCSAVP00000011661 <sup>b</sup>
<i>Danio rerio</i>	XM_001334114 <sup>a</sup>	XP_001334150 <sup>f</sup>
<i>Drosophila melanogaster</i>	NM_141207 <sup>a</sup>	NP_649464 <sup>f</sup>
<i>Drosophila pseudoobscura</i>	Dpse\GA22074 <sup>d</sup>	EAL28754 <sup>f</sup>
<i>Gallus gallus</i>	NM_001006159 <sup>a</sup>	NP_001006159 <sup>f</sup>
<i>Gasterosteus aculeatus</i>	ENSGACG00000004961 <sup>b</sup>	ENSGACP00000006557 <sup>b</sup>
<i>Homo sapiens</i>	NM_015949 <sup>a</sup>	NP_057033 <sup>f</sup>
<i>Loxodonta africana</i>	ENSLAFG00000016008 <sup>b</sup>	ENSLAFP00000013431 <sup>b</sup>
<i>Macaca mulatta</i>	ENSMMUG00000017063 <sup>b</sup>	ENSMMUP00000022458 <sup>b</sup>
<i>Monodelphis domestica</i>	ENSMODG00000008643 <sup>b</sup>	ENSMODP00000010752 <sup>b</sup>
<i>Mus musculus</i>	NM_026269 <sup>a</sup>	NP_080545 <sup>f</sup>
<i>Ornithorhynchus anatinus</i>	ENSOANG00000006995 <sup>b</sup>	ENSOANP00000011144 <sup>b</sup>
<i>Oryzias latipes</i>	EF177382 <sup>a</sup>	ABM53480 <sup>f</sup>
<i>Rattus norvegicus</i>	ENSRNOG00000001293 <sup>b</sup>	ENSRNOP00000001745 <sup>b</sup>
<i>Schistosoma japonicum</i>	AY813663 <sup>a</sup>	Q5DFL1 <sup>g</sup>
<i>Schistosoma mansoni</i>	Smp_050130 <sup>e</sup>	Smp_050130 <sup>e</sup>

<i>Strongylocentrotus purpuratus</i>	XM_781267 <sup>a</sup>	XP_786360 <sup>f</sup>
<i>Salmo salar</i>	EF036472 <sup>a</sup>	ABK35126 <sup>f</sup>
<i>Tribolium castaneum</i>	XM_962279 <sup>a</sup>	XP_967372 <sup>f</sup>
<i>Takifugu rubripes</i>	EF445943 <sup>a</sup>	ABO32372 <sup>f</sup>
<i>Tetraodon nigroviridis</i>	GSTENG00008219001 <sup>b</sup>	GSTENP00008219001 <sup>b</sup>
<i>Xenopus laevis</i>	BC070733 <sup>a</sup>	AAH70733 <sup>f</sup>
<i>Xenopus laevis</i>	BC074468 <sup>a</sup>	AAH74468 <sup>f</sup>
<i>Xenopus tropicalis</i>	NM_001016708 <sup>a</sup>	NP_001016708 <sup>f</sup>

---

Superscript letters indicate the databases from which sequences can be accessed: <sup>a</sup> GenBank (NCBI); <sup>b</sup> Ensembl genome assembly, <sup>c</sup> Wormbase, <sup>d</sup> Flybase, <sup>e</sup> TreeFam, <sup>f</sup> GenPept (NCBI) and <sup>g</sup> Uniprot.



### A. Yeast and invertebrates: variable gene structure

*Saccharomyces cerevisiae* (chromosome XV)



*Caenorhabditis elegans* (chromosome III)



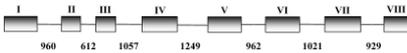
*Aedes aegypti* (supercontig 1.59)



*Drosophila melanogaster* (chromosome 3R)



*Strongylocentrotus purpuratus* (chromosome 3R)



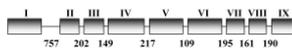
*Ciona intestinalis* (chromosome 13q)



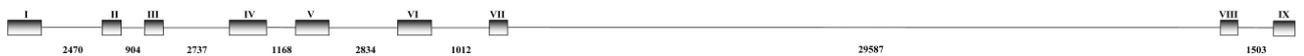
100 bp

### B. Vertebrate taxa: conserved intron/exon structure

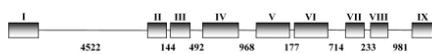
*Takifugu rubripes* (scaffold 3)



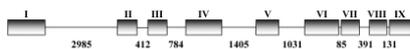
*Danio rerio* (chromosome 12)



*Xenopus tropicalis* (scaffold 850)



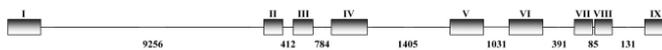
*Gallus gallus* (chromosome 14)



*Ornithorhynchus anatinus* (ultracontig 333)



*Homo sapiens* (chromosome 7)



100 bp

Fig. 6.2 Gene organisation of *cee* across the eukaryotes. **A.** The structure of the *cee* gene is highly conserved in vertebrates as nine exons and eight introns. **B.** Non-vertebrate taxa, including lower chordates (tunicates) have conserved *cee* in various less complex forms. Exons are represented by bars labelled with roman numerals. Introns are downscaled by ten times but their real size (in nucleotide basepairs) is indicated. Only the gene regions corresponding to the coding region are shown in this diagram.

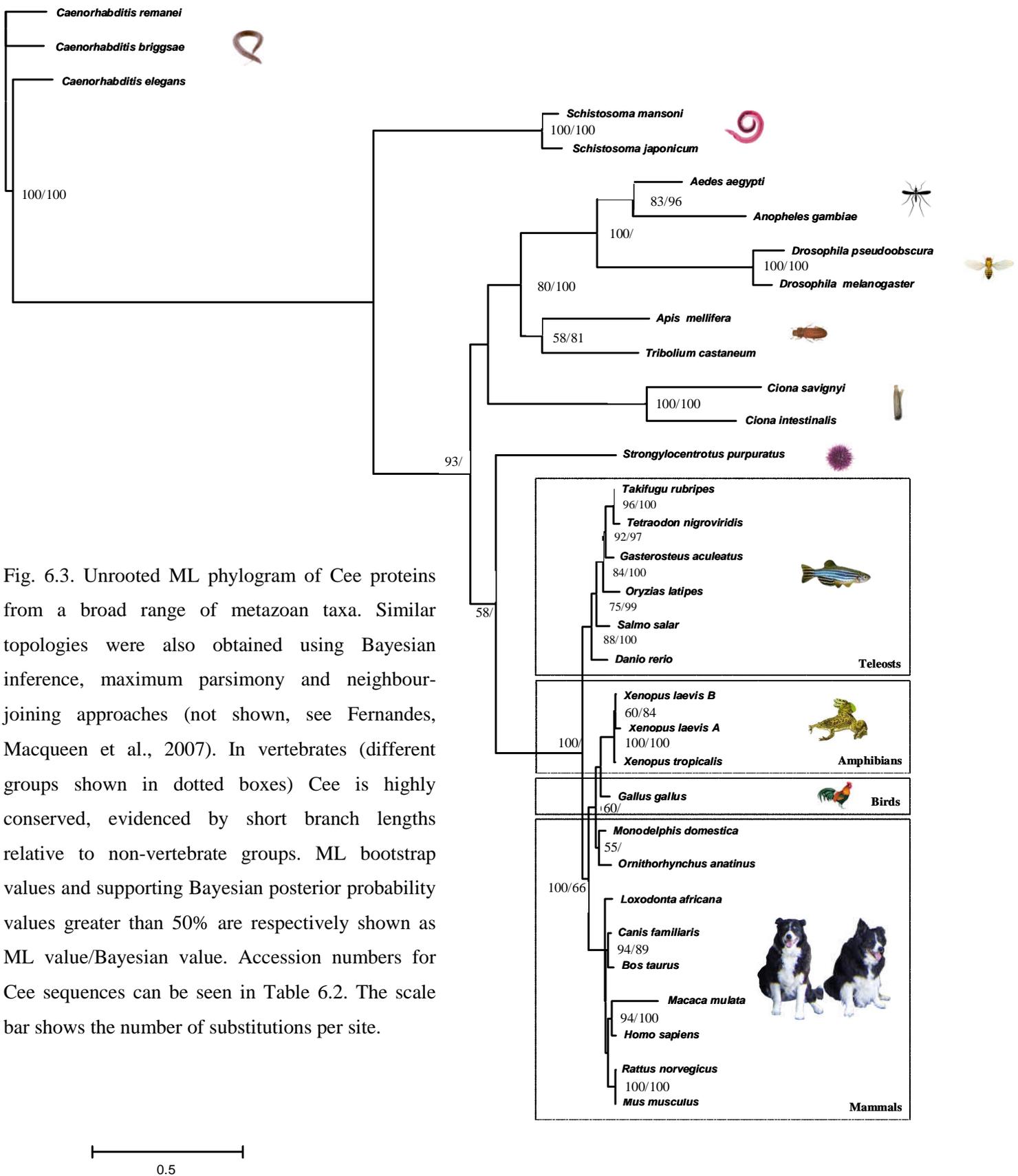


Fig. 6.3. Unrooted ML phylogram of Cee proteins from a broad range of metazoan taxa. Similar topologies were also obtained using Bayesian inference, maximum parsimony and neighbour-joining approaches (not shown, see Fernandes, Macqueen et al., 2007). In vertebrates (different groups shown in dotted boxes) Cee is highly conserved, evidenced by short branch lengths relative to non-vertebrate groups. ML bootstrap values and supporting Bayesian posterior probability values greater than 50% are respectively shown as ML value/Bayesian value. Accession numbers for Cee sequences can be seen in Table 6.2. The scale bar shows the number of substitutions per site.

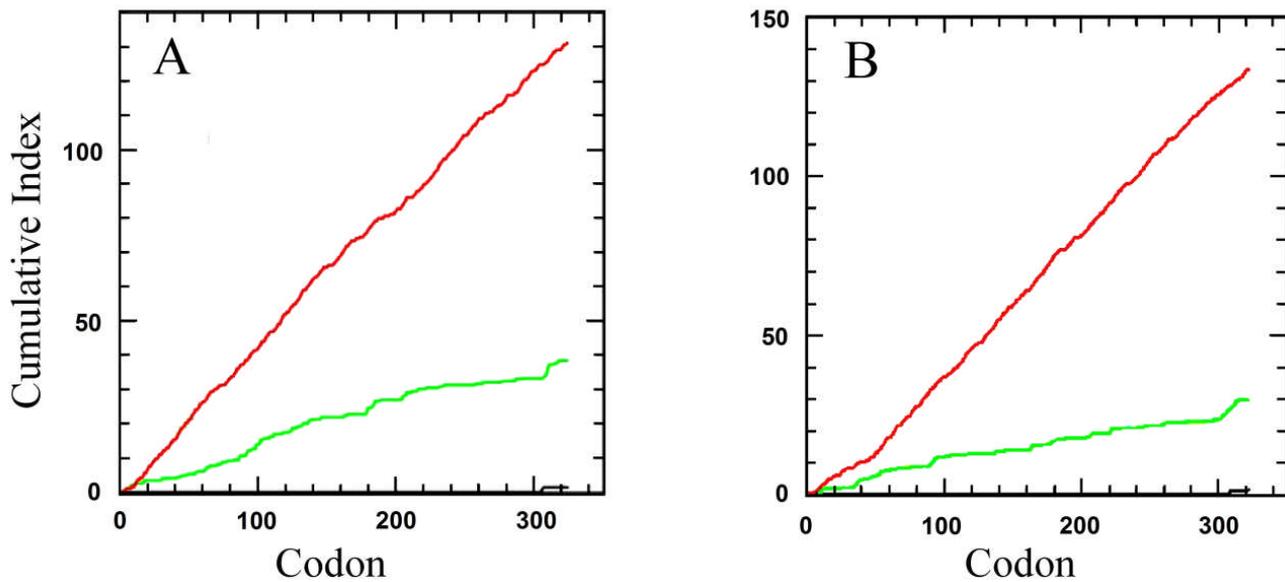


Fig. 6.4. Synonymous and non-synonymous substitution rates in *Cee*. Cumulative indices of average synonymous (red line) and non-synonymous (green line) substitutions, as well as insertions and deletions (black line) in the *cee* coding sequences are plotted against the aligned *Cee* protein sequences from teleost (**A**) and mammalian groups (**B**) (see methods section 6.2.4 for details on these groups). The number of synonymous substitutions per site is approximately constant throughout the *cee* coding sequence and markedly higher than the number of non-synonymous mutations, suggesting that the protein is under strong purifying selection. Figure from Fernandes, Macqueen et al. (2007).

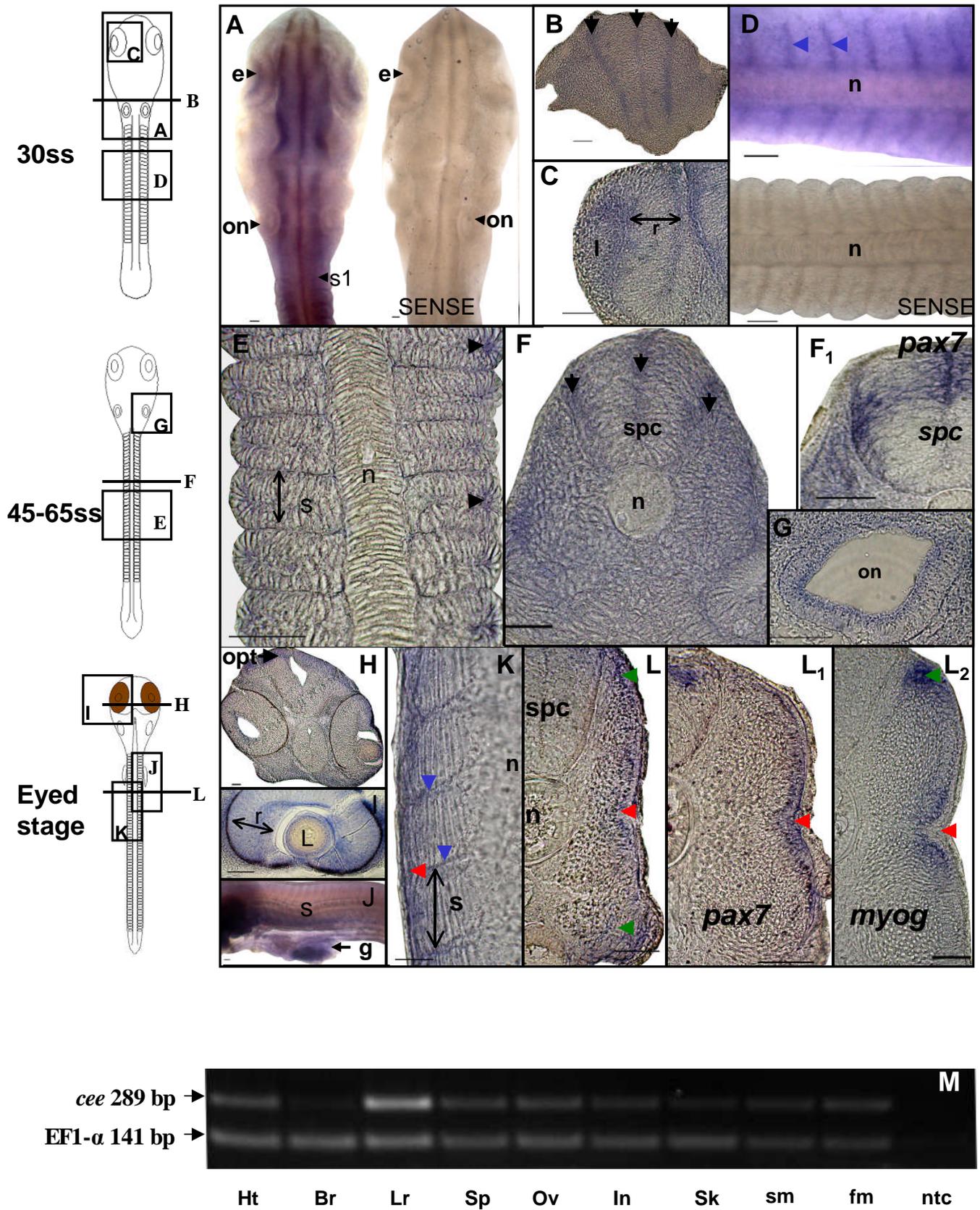


Fig. 6.5. Figure legend is on the next page

Fig. 6.5. Developmental expression pattern of *cee* in Atlantic salmon embryos (A-L) and adult tissues (M). Schematic images of embryos to the left of the figure show the position of higher magnification flat mounts and sections (letters A-L). **A.** Dorsal flat mounts showing the rostral region of a 30 ss embryo incubated with *cee* antisense or sense mRNA (marked SENSE). **B.** Cross section through the mid brain (at the level of the optic tectum) at the 30 ss. Black arrows show *cee* expression at the midline and borders of the entire dorsal-ventral brain. **C.** Cross section through the eye of a 30 ss embryo. *cee* was expressed at the lens-retina border throughout the segmentation period. **D.** Dorsal flat mount of the somite region of 30 ss embryos showing *cee* staining at the somite borders. An equivalent sense control embryo (labelled SENSE) is shown for comparison. **E.** Longitudinal section through the epithelial somites of a 45 ss embryo. *cee* mRNA clustered within cells on the lateral somite border and at the somite boundaries. **F.** Somite cross-section from an anterior somite at the 45 ss. Note the triple stripe of spinal cord expression marked by black arrows. Also shown as an inset, is *pax7* staining at an equivalent stage (**F<sub>1</sub>**). **G.** Longitudinal section through the otolith nuclei; *cee* stained the internal edge. This image is representative of *cee* staining during segmentation and eyed stages. **H.** Cross section through the mid-brain of an eyed embryo at the level of the optic tectum. **I.** Transverse section through the pigmented eye. *cee* expression persisted at the boundaries of the lens, and retina beyond segmentation. **J.** Lateral perspective flat mount of an eyed embryo showing the unrestricted expression of *cee* in the developing gut. **K.** Longitudinal section along the somites of an eyed-stage embryo. *cee* was excluded from the medial myotome but was expressed at its lateral edge, in the external cell layer (red arrows) and along the somite borders (blue arrow). **L.** Eyed stage myotome cross-section at the level just below the fin buds. *cee* was expressed in the external cell layer (red arrow) concomitant with *pax7* (**L<sub>1</sub>**) and was also present in the myotome in regions of new muscle production (green arrows), where muscle specific markers like *myog* (**L<sub>2</sub>**) were expressed. Abbreviations: e: eye, g: gut, l: lens, on: otolith nuclei, n: notochord, opt: optic tectum, r: retina, s: somites and spc: spinal cord. Scale bars represent 50  $\mu$ m. **M.** Tissue distribution of *cee* mRNA in adult Atlantic salmon tissue. A duplex RT-PCR was performed using cDNA derived from nine tissues and primers specific to *cee* and the housekeeping gene *efl- $\alpha$* . Abbreviations: Ht (heart), Br (brain), Lr (Liver), Sp (spleen), Ov (ovary), In (intestine), Sk (skin), sm (slow muscle), fm (fast muscle), ntc (no template control). The size of each amplicon is shown to the left of the figure.

## 6.5 Discussion

### 6.5.1 *Cee is a novel protein of unknown function that is widely conserved in eukaryotes*

In this chapter results have been presented that are the first to characterise *cee*, a highly conserved eukaryotic gene. The Cee protein lacks any currently recognised functional motifs so it is difficult to speculate on its developmental function. The limited information available is mainly derived from high-throughput studies using model yeast and nematodes species. In the yeast *S. cerevisiae* Cee protein (YOR164C) is located in the cytoplasm (Huh et al., 2003) and it was shown by affinity capture-mass spectrometry and two hybrid experiments that Cee interacts with the proteins Mdy2 (Fleischer et al., 2006) and Get3 (Ito et al., 2001). Mdy2 has been characterised in yeast and has a ubiquitin-like domain, which associates with ribosomes and is required for efficient mating (Hu et al., 2006). Get3 is an ATPase necessary for transporting proteins from the Golgi apparatus to the endoplasmic reticulum (Schuldiner et al., 2005). Cee mutants in yeast were viable but exhibited sensitivity at five generations to the antifungal compound nystatin (Giaever et al., 2002). Further, when RNA interference was used to inhibit the function of *cee* in *C. elegans*, its development was retarded (Kamath et al., 2003; Simmer et al., 2003). Additionally, it was found in 25 separate microarray experiments listed in the Array Express database (<http://www.ebi.ac.uk/microarray-as/aer>), that *cee* (the human *cee* orthologue is known as C7orf20, CG020 or CGI-20) was the most highly differentially expressed gene in multiple human tissues in a range of cancerous or diseased states including sufferers of Huntington's disease, breast cancer and leukaemia. These experimental data are limited but insightful and along with the fact the *cee* has been conserved from very early in eukaryote evolution suggest that the Cee protein plays some fundamental physiological role in regulating cellular growth, protein binding, intracellular traffic or translation. Since there are considerable sequence differences between the vertebrate and invertebrate orthologues, it is also possible that

Cee has additional molecular functions and is involved in other biological processes in vertebrates.

### 6.5.2 *Cee is a lonely gene without a family*

Consistent with its lack of recognised functional motifs, *cee* does not form part of any larger gene family. This is unusual given that there is a tendency for genes found as single copies in protostomes or basal deuterostomes to be present in multiple paralogous copies in the vertebrate lineage and even more in ray-finned fish (reviewed in chapter 1, section 1.5). However, *cee* was conserved as a single gene in almost all taxa examined, including the teleosts which are tetraploid in relation to most vertebrates (Jaillon et al., 2004), as well as salmonids, which have been through a further duplication (Allendorf and Thorgaard, 1984). Conversely, several genes in the vicinity of *cee* had multiple paralogues and formed parts of ancient gene families (not shown). The existence of a single *cee* gene in most species is consistent with the corresponding low dN/dS values (Fig. 6.4) and indicates that Cee is under strong selective pressure to resist changes in propeptide structure and protein function. Thus, following the multiple duplications of the *cee* gene that will have occurred throughout Eukaryote evolution, it is likely that selection has consistently and quickly eradicated one paralogue through non-functionalization. Additionally, the expression of *cee* may require tight regulation, which would be difficult to achieve with two independently evolving promoters. However there are two species where *cee* is present as two paralogues. In the yellow fever mosquito (*A. aegypti*), two *cee* genes sharing 99.5% identity within their coding sequences are found in distinct chromosomal regions. The lack of divergence between these paralogues is consistent with a very recent species-specific gene duplication event, which is not contrary to the presence of a single *cee* copy in other metazoans; it can be suggested that in time, one paralogue will probably degenerate into a pseudogene. The tetraploid frog *X. laevis* has also retained two *cee* paralogues that are 95% similar and branched as sister sequences from the *X. tropicalis* (a diploid relative) sequence in

ML/Bayesian phylogenetic reconstructions (Fig. 6.3). This indicates that the frog paralogues arose in a lineage specific event, probably during the allotetraploidization event that occurred in the *Xenopus* lineage approximately 30 Mya (Evans et al., 2004). Since the frog paralogues are evolving asymmetrically it is possible that the more derived copy is also in the process of non-functionalization.

### 6.5.3 Gene structure of *cee*

Consistent with the strong sequence conservation of the Cee protein in vertebrate taxa (Fig. 6.1), the genomic organisation of *cee* is also conserved across this group as nine exons and eight introns (Fig. 6.2). Conversely, the structure of *cee* in invertebrates and fungi was more simple and varied from a single exon in the yeast *S. cerevisiae* to a gene with a highly similar structure to the vertebrate genes in the basal deuterostome *S. purpuratus* (purple sea urchin). The genomic organisation of vertebrate *cee* in fact shares remarkable similarities with the 8-exon structure found in the purple sea urchin (results section 6.3.3). Considering this, it was surprising to find that *cee* was conserved in *C. intestinalis/sauvigni* as a single exon gene, since tunicates are more closely related to vertebrates than echinoderms (Delsuc et al., 2006). It is feasible that the intronless version of *cee* found in the tunicate has been created by reverse transcription of the processed mRNA followed by genome integration (retrotransposition), a common molecular mechanism of gene formation in eukaryotes (Babushok et al., 2007). However, we found no evidence supporting the existence of a putative parental *cee* gene in the genome of the tunicates *C. intestinalis* or *C. sauvigni*. It would also be interesting to establish the genomic structure of *cee* in the cephalochordate amphioxus, to establish whether the tunicate gene structure is an anomaly specific to this group.

#### 6.5.4 Developmental expression of *cee*

In Atlantic salmon embryos, *cee* mRNA transcripts accumulated in multiple organs and tissues including at the borders of the somites, in a triple stripe marking the edges and midline of the entire brain and neural tube, at the borders of several structures in the developing eye, and at the peripheral regions of the otolith nuclei and branchial arches (Fig. 6.5, A-L). From a comparative perspective, recent findings in our laboratory suggest that this embryonic expression pattern is conserved in zebrafish (Lee, 2008). However, it is unknown whether this expression pattern is conserved with other vertebrates, although this likely to be the case, considering the fact that *cee* has a strongly conserved function and has no related family genes/paralogues where cis-acting regulatory sequences could be distributed between independently evolving promoters. Based on the spatiotemporally complexity of embryonic *cee* expression, it is difficult to speculate further on its specific role.

From the end of segmentation to the eyed stage of Atlantic salmon development, *cee* was expressed in the external cell layer at the same time as *pax7* and in the peripheral myotome concurrent to MRFs, and particularly in dorsal and ventral zones of stratified hyperplasia (Fig.6.5, L). Additionally, the external cell-layer expression of Cee protein was shown to persist until adult zebrafish growth stages (Lee, H-T and Johnston, I.A; unpublished results). The *cee* expression domain in zones of stratified hyperplasia is to some extent unexpected, since the pufferfish *cee* orthologue was shown by qPCR to be strongly upregulated in fast muscles where muscle fibre production had ceased (Fernandes et al., 2005). Considering the lack of information on Cee protein function, it is not possible to comment on the significance of these seemingly contrasting findings at the mRNA level.

*cee* mRNA transcripts were present in multiple Atlantic salmon tissues and were not obviously upregulated in fast myotomal muscle at a body size (~4kg) where myotube production had

ceased (Fig. 6.5, M). A similar finding was observed in adult zebrafish using a similar RT-PCR approach (Lee, 2008). Additionally, *cee* was upregulated by huge magnitudes of order in multiple human tissues from individuals suffering from various cancers (see above section 6.4.1). Thus the role played by *cee* in vertebrates is clearly not limited to regulating myotube production. Further, it currently cannot be excluded that the upregulation of *cee* in pufferfish at a body size where myotube formation had ceased (Fernandes et al., 2005) was a coincidental response to some other physiological event. Additional experiments are required to establish the localisation of Cee protein in adult skeletal muscle during the transition from myoblast to differentiated fibre. It is also necessary to investigate how the overexpression of Cee protein influences the *in vitro* differentiation of myoblasts in culture. More generally, the knockout of the *cee* gene in mice, or ablating its translation in zebrafish using morpholino antisense RNA might elucidate the developmental pathways to which Cee contributes and considering our findings, it is unlikely that any genetic redundancy would exist to mask the phenotype of these prospective mutants.

# **Chapter 7. The co-ordinated spatiotemporal expression of myogenic regulatory factors is affected by temperature during embryonic myogenesis in Atlantic salmon**

## **7.1. Abstract**

In this chapter, potential molecular mechanisms regulating developmental plasticity to temperature were investigated in Atlantic salmon embryos (*S. salar* L.). Initially six salmon orthologues of the four myogenic regulatory factors (*myod1a/1b/1c*, *myf5*, *myog* and *mrf4*) were cloned and comparatively characterised at the mRNA/genomic level. *In situ* hybridisation was then performed with specific cRNA probes to establish a comprehensive record of the co-ordinated expression pattern of each gene during embryonic myogenesis. To place this MRF data in the context of known muscle fibre differentiation events, the expression of slow myosin light chain-1 (*smlc1*), a marker of adaxial cell differentiation and *pax7*, a marker of anterior somite MPCs, was concurrently investigated. Adaxial myoblasts expressed *myod1a* prior to and during somitogenesis followed by *myod1c* (20 somite-stage, ss), and *mrf4/smlc1* (25-30ss) before migrating laterally across the myotome. *myf5* was detected broadly in the segmental plate prior to somitogenesis, but not in the adaxial cells in contrast to other teleosts studied to date. The expression domains of *myf5*, *myod1b* and *myog* were not confined to the *smlc1* expression field indicating a role in fast muscle myogenesis. From the end of segmentation, each MRF was expressed to a greater or lesser extent in zones of new muscle fibre production, the precursor cells for which probably originated from the *pax7* expressing cell layer external to the single layer of *smlc1*<sup>+</sup> fibres. *myod1a* and *myog* showed similar expression patterns with respect to ss at three different embryonic temperatures (2, 5 and 8 °C) in spite of different rates of somite formation (one somite added each 5, 8 and 15 h at 8, 5 and 2 °C respectively). In contrast, the expression of *myf5*, *mrf4* and *smlc1* was retarded with respect to somite-stage at 2°C compared to

8°C, potentially resulting in heterochronies in downstream pathways influencing later muscle phenotype.

## 7.2 Introduction

Teleosts produce new myotubes throughout larval, juvenile and adult stages. The final fibre number in adult Atlantic salmon can be modified by around 15-20% according to the temperature experienced during the early life history stages (chapter 8; Johnston et al., 2003a). The majority of teleost muscle fibres are added in adult stages during mosaic hyperplasia; however, the origin of the MPCs for this growth phase is unknown. In amniotes it has recently been reported that the satellite cells, the principal MPCs for juvenile and adult muscle growth, are derived from the dermomyotome in the embryonic somite (Gros et al., 2005). In zebrafish, the anterior part of the epithelial somite has been shown to be functionally equivalent to the amniote dermomyotome in terms of the progenitor cells it supplies (Hollway et al., 2007). For example, this region provides *pax7* expressing cells that form a layer of self-renewing undifferentiated MPCs external to the myotome that are utilized for myotube production during stratified hyperplasia in larvae and possibly during mosaic hyperplasia in adult stages (Hollway et al., 2007; Stellabotte et al., 2007; Stellabotte and Devoto, 2007). Based on the conserved expression pattern of *pax7* in teleosts, it seems that the myogenic potential of the external cell layer is conserved throughout this group (Devoto et al., 2006; Steinbacher et al., 2006). If temperature affects the number of post-embryonic myogenic precursors originating from the external cell layer, this could provide a plausible explanation for later changes in the final fibre number in adult fish.

It is now well documented that early rearing temperature has a profound influence on the muscle fibre phenotype in a diverse range of teleost species (reviewed in chapter 1 section 1.8).

However, the molecular-level origin of phenotypic plasticity to temperature is poorly understood. To study the effect of embryonic temperature on teleost myogenesis, previous studies have used *in situ* hybridization of MRFs in embryos reared at different temperatures. The majority of studies have found no difference in the relative timing or intensity of *myod* or *myog* expression with respect to somite-stage in embryos reared at a range of temperatures (Atlantic cod (*G. morhua*), Hall et al., 2003; Atlantic herring (*C. harengus*), Temple et al., 2001; common carp (*C. carpio*), Cole et al., 2004 and Atlantic halibut (*H. hippoglossus*; Galloway et al., 2006). However, in rainbow trout (*O. mykiss*) it was reported that *myod1a* and *myog* expression was more intense at the mRNA and protein levels and also more advanced with respect to ss in embryos incubated at 12 versus 4 °C (Xie et al., 2001). In this chapter it was thought worthwhile to re-examine potential developmental plasticity of MRF expression with respect to temperature in Atlantic salmon embryos, extending the range of MRF markers to include all four family members.

## 7.3 Materials and methods

### 7.3.1 Embryos

*S. salar* embryos were reared and sampled as part of the embryonic temperature trial described in chapter 2 (section 2.2) and were selected from the 2, 5, and 8 °C treatments. Six embryos per stage (chapter 2, section 2.2.2) were selected for each temperature treatment. Embryos were fixed in 4% (m/v) paraformaldehyde/PBS and then dehydrated by consecutive washes in increasingly concentrated methanol (until 100% m/v) and stored at –80 °C until later use.

### 7.3.2 Cloning new Atlantic salmon myogenic regulatory factors and *smlc1*

The cloning of Atlantic salmon *myod1a/1b/1c* is described in chapter 4. Here the cloning of salmon *myf5*, *myog*, *mrf4* and *smlc1* is described. Total RNA was extracted from 100 mg of fast-

twitch myotomal muscle from juvenile Atlantic salmon provided by EWOS Innovation (n = 6; mean weight = 291 ± 36 g, mean forklength = 263 ± 27 mm) and using the protocol described in chapter 2 (section 2.4.2). The RNA sample was then DNase digested, quantified and its quality assessed as described in chapter 2 (sections 2.4.2 and 2.4.4). First strand cDNA was synthesised using 1 µg of this total RNA as described in chapter 2 (section 2.4.5). Genomic DNA was extracted from 50 mg of spleen tissue as described in chapter 2 (section 2.4.2). The primers shown in Table 7.1 were then used to amplify Atlantic salmon full coding sequences of *myog*, *myf5* and *smlc1*, and a partial *mrf4* sequence, using several standard PCR reactions with genomic DNA (MRFs) and cDNA (MRFs and *smlc1*). To obtain the 3' of the *mrf4* gene (plus full coding sequence, cds), a 3' RACE PCR reaction was performed as described in chapter 2 (section 2.4.8). PCR products were separated by gel electrophoresis, and then isolated purified, cloned and sequenced as described in chapter 2 (section 2.4).

### 7.3.3 Computational processing of MRF sequences

A consensus nucleotide and AA translation of each sequenced cDNA/genomic DNA was constructed using DNAMAN (Lynnon Biosoft). The identity of putative genes was confirmed against the complete non-redundant NCBI database using BLASTn and tBLASTn searches and each gene was submitted to the GenBank public database. The intron-exon structure of MRFs was assessed by aligning cDNA and corresponding genomic DNA sequences in Spidey (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>) (Wheelan et al., 2001).

### 7.3.4 Probe transcription and in situ hybridisation

To make DNA templates for RNA probe synthesis, PCR was used with T3/T7 primers and as a template, a pCR4- TOPO T/A plasmid containing the cDNA products of *myf5*, *myog*, *smlc1* and *mrf4* (Table 7.1) excluding the *mrf4* RACE product. The *myod1a/1b/1c* probe templates were as described in chapter 4. Additionally, to make a DNA template for the Atlantic salmon orthologue

of *pax7*, nucleotides 502-1119 of the previously characterised mRNA (Gotensparre et al., 2006) were amplified by RT-PCR, cloned and sequenced as described above (primers in Table 7.1).

Each cRNA probe was synthesised in sense and antisense directions using T3/T7 RNA polymerases with concurrent incorporation of digoxigenin or fluorescein labelling (chapter 2, section 2.6.3). Hybridization of embryos with probes was performed as described in chapter 2 (section 2.6.5) with single-probe hybridisation. Different temperature treatments were incubated in each solution for identical time periods. This ensured that differences recorded between temperature groups in the colour development step were attributable to differences in gene expression rather than unequal sample treatment.

#### *7.3.5 Processing embryos and figure construction*

All embryos from each temperature treatment and stage were studied using both a DMRB compound, and a Leica MZ7.5 binocular microscope (Leica Ltd). When DIC optics was used, embryos were flat mounted with a cover slip on a clear microscope slide and orientated to a dorsal or lateral perspective (chapter 2, section 2.6.6) Embryos were staged by counting the somite number and photographs were recorded on a Nikon P4500 camera. Subsequently, serial sections of embryos were cut following the protocol described in chapter 2 (section 2.6.6). Differences in gene expression patterns between temperature treatments were considered reliable when replicated in each embryo at each stage ( $n = 6$ ). When figures were constructed, representative images of embryos from equivalent somite stages were selected from each temperature treatment. Differences in temperature groups were not considered in developmental windows when embryos could not be accurately staged i.e. prior to somite formation and after the completion of segmentation.

## 7.4. Results

### 7.4.1 Characterisation of Atlantic salmon myogenic regulatory factors

In Atlantic salmon, *myod* is represented by three paralogues (*myod1a*, *1b* and *1c*) that were characterised in chapter 4. Here, full coding sequences of all other Atlantic salmon *myod* family genes have been obtained. Using primers designed from rainbow trout *myf5* (AY751283), a complete coding sequence (cds) of Atlantic salmon *myf5* (DQ452070) was obtained incorporating 720 bp that translated into an ORF of 239 AAs (Fig. 7.1). The percentage identity conserved between salmon *myf5* and other vertebrate orthologues at the respective nucleotide/protein level was 97.9/96.2% with rainbow trout, 75.9/73.0% with pufferfish (*T. rubripes*) (NM\_001032770), 71.7/76.3% with zebrafish (*D. rerio*) (AF253470), 59.2/54.8% with frog (*X. laevis*) (AJ579311), 60.7/56.3% with chicken (*G. gallus*) (NM\_001030363), and 63.4/54.5% with human (*H. sapiens*) (NP\_005593).

Using primers designed from rainbow trout *myog* (Z46912), a complete cds corresponding to Atlantic salmon *myog* (DQ294029) was obtained which was 789 bp long, and translated into an ORF of 254 AA (Fig. 7.2). The percentage identity conserved between salmon *myog* and other vertebrate *myog* orthologues at the respective nucleotide/protein level was 97.9/98.4% with rainbow trout, 76.2/77.6 % with pufferfish (*T. rubripes*) (AY566282), 72.7/73.5% with zebrafish (NM\_131006), 61.9/58.1% with frog (NM\_001016725), 64.1/56.4% with chicken (D90157), and 65.4/53.2% with human (NM\_002479).

*mrf4* had not previously been cloned in any salmonid fish. For this reason, primers used to amplify *mrf4* were initially based on an EST sequence (DN165140) obtained from a tBLASTn search of the salmon genome project ([www.salmongenomeproject.no](http://www.salmongenomeproject.no)) using the translated *D. rerio mrf4* mRNA (NM\_001003982) as a probe. A reverse primer was designed from this sequence and was used with a forward primer (both Table 7.1) designed in the start region of

*mrf4* based on alignments with several vertebrate sequences to amplify nucleotides 1-649 of the salmon coding sequence. Finally, a 3' RACE primer was designed at the 3' of the confirmed *mrf4* sequence (Table 7.1) and this was used in a 3' RACE reaction to obtain a whole cds for *mrf4* and a complete 3' untranslated region, with a poly-A tail and one polyadenylation signal (AATAAA) (Fig. 7.3). *mrf4* shared closest homology to its orthologue in the Knifefish, *S. macrurus* (DQ059552) with 75.0/76.9 % nucleotide/AA identity. The percentage identity conserved between *mrf4* and other vertebrate orthologues at the nucleotide/protein level was 70.9/71.6% with pufferfish (AY445320), 73.9/76.0 % with zebrafish, 63.8/59.3% with frog (S84990), 62.8/60.9% with chicken (D10599), and 62.7/62.1% with human (NM\_002469).

*Addendum: it has recently been shown that a second mrf4 sequence mapped to a distinct chromosomal location in the Atlantic salmon genome compared to the mrf4 gene characterised in this chapter (see Moghadam et al., 2007; accession number of new mrf4 genomic sequence; EF450078). This finding and how it relates to results in this chapter is discussed in depth in the discussion (section 7.5.1). Furthermore, an alignment of the two sequences can be seen in Fig. 7.14.*

Fig. 7.4 shows an AA alignment of all known salmonid MRFs with an ancient MyoD homologue in the cephalochordate *Branchiostoma belcheri*. The bHLH domain and cis-his-rich region (just N-terminal to the basic region) are strongly conserved across all sequences. Additionally, the helix-3 of MyoD1 paralogues (AA- 206-221 of MyoD1a) is most similar to cephalochordate MyoD (5/15 substitutions vs. MyoD1a) >salmonid-Myf5 proteins (6/15 substitutions vs. MyoD1a) >salmon Mrf4 (8/15 substitutions vs. MyoD1a) >salmonid Myog proteins (10/15 substitutions vs. myoD1a). Additionally a highly conserved motif of unknown function is present in salmonid MyoD1 paralogues (Fig. 7.4 shown in bold italics on MyoD1a) and other vertebrate MyoD proteins (not shown), which is partially conserved in amphi-MyoD1 but not in other

MRFs (Fig. 7.4). The extreme N-terminal and all regions C-terminal to the HLH (excluding the HLH and motif of unknown-function) are the least conserved regions of the salmonid MRF proteins.

#### 7.4.2 Genomic organisation of salmonid MRFs

The intron-exon structures of all known Atlantic salmon MRFs are presented in Fig. 7.5. Common to all vertebrate MRFs, each salmonid gene is represented as three exons and two introns. For each gene, exon 1 is the largest, incorporating the N-terminal activation domain, basic and HLH motifs, and in vertebrate *myod* genes, a highly conserved region that has no assigned function currently. Exon 2 is the smallest for each MRF, and exon 3 incorporates the helix-3 domain.

#### 7.4.3 Characterisation of Atlantic salmon *smlc1*

Primers to amplify a full cds of Atlantic salmon slow myosin light chain-1 (*smlc1*) were designed from the rainbow trout sequence previously reported (EST: (BX076946; Chauvigné et al., 2005). The cds of salmon *smlc1* (DQ916288) was 561 bp that translated into an ORF of 185 AA. The percentage identity conserved between salmon *smlc1* and other vertebrate *smlc1* orthologues at the respective nucleotide/protein level is 99.1/99.5% with rainbow trout, 78/81% with the pufferfish *T. nigroviridis* (putative: predicted within CAAE01014556), 80/83% with zebrafish (NP\_956810), 65/67% with frog (EST: AAI28964), 66/69% with chicken (P02606) and 64/67% with human (NP\_002467).

#### 7.4.4 MRF expression: introduction

The mRNA expression patterns of six MRF genes were initially recorded throughout salmon embryogenesis at 8 °C. To place the expression of each MRF in the context of known muscle fibre differentiation events, I also examined the expression of *smlc1*, which is expressed in

rainbow trout adaxial cells as they differentiate (Chauvigné et al., 2005), and *pax7*, which is expressed in the myogenic precursors of the external layer (e.g. Devoto et al., 2006; Stellabotte et al., 2007, Hollway et al., 2007). The expression data is presented here in two formats. Firstly, a schematic diagram shows the progressive expression of each gene in the most anterior somite of salmon embryos during segmentation and post-segmentation stages of embryogenesis (Fig. 7.6). This excludes the complexity generated when considering the embryos rostral-caudal axis and associated gradient in expression patterns due to changing somite maturity. Fig. 7.6 enables the reader to quickly establish the spatio-temporal correlations between the expression patterns of the six MRFs with *smlc1* and *pax7* in a single maturing somite. Next, a detailed inventory of expression images was constructed for salmon *myod1a*, *myf5*, *myog*, *mrf4* and *smlc1* from the 30-45 ss (Fig. 7.7) and these genes plus *pax7* at the end of segmentation and at the eyed stage (Fig. 7.8). Only one *myod1* paralogue (*1a*) was considered in figs 7.7 and 7.8, considering the recent description of the expression of *myod1a/1b/1c* (Chapter 4). Each of the descriptions has been written for independent use, and thus some overlap exists between them.

#### 7.4.5 MRF expression in a single maturing somite (Fig. 7.6)

When So-1 arose from the unsegmented mesoderm, *myod1a* and *myf5* were respectively expressed in the adaxial myoblasts flanking the notochord (A1) and throughout its entire lateral width, excluding the most anterior quarter (B1). *myf5* was not expressed in So-1 adaxial cells during any period of embryogenesis (B1-6) in contrast to other teleosts studied to date (e.g. Coutelle et al., 2001; Cole et al., 2004). At the point when there were around 20 newer somites caudal to So-1, *myod1a* began to extend dorso-ventrally in the medial somite and *myf5* did not change significantly (A2 and B2). However, at this time, three other MRFs were turned on in So-1. *myod1b* expression was similar to *myf5* extending through the entire posterior domain of So-1, including the undifferentiated adaxial cells (C2 vs. B2). *myod1c* had a comparable expression pattern to *myod1a* and *1b*: transcripts were detected in the adaxial myoblasts (D2 vs. A2) and

diffusely in the posterior-lateral region of So-1 (D2 vs. C2). At this time, *myog* was also widely expressed across So-1 (E2).

The first expression of *mrf4* was present in So-1 at the 25ss in the adaxial cells adjacent to the notochord (F3), just before or contemporaneously to an identical expression field for *smlc1* marking the differentiation of adaxial cells (G3). At this time *myf5* expression was downregulated in So-1 (B3), whilst *myod1b* and *myog* extended anteriorly (C3 and E3), coinciding with the differentiation of fast muscle fibres. In contrast *myod1a/1c* expression spread dorso-ventrally and laterally in the medial somite, maintaining a signal to the medial *smlc1* expressing adaxial cells (A3 and D3). This phase of expression just preceded the dorso-ventral extension and lateral migration of *smlc1* transcripts in adaxial cells that were observed spatially as a triangular wave throughout the middle of So-1 at the 45ss (G4). At this point *myod1a/1c* and *mrf4* expression was comparable to *smlc1* and no longer present in the medial myotome of So-1 (A4, D4 and F4). In contrast, *myod1b* and *myog* were detected in the entire length and width of So-1 at this time (C4 and G4). Additionally, *myf5* expression re-accumulated at the superficial edge of the posterior region of So-1, before the completion of adaxial cell migration (B4).

At the end of segmentation (60-65 ss), So-1 had fully acquired the chevron-shaped phenotype, and the adaxial cells had spanned the myotome to form a single layer of slow-fibres, evidenced by *smlc1* expression (G5). *pax7* expression was present external to this layer, presumably marking myogenic progenitors of the external cell layer (H5). At the end of segmentation, *myod1a*, *myod1c*, *myog* and *mrf4* were each expressed throughout the bulk of the myotome of So-1, presumably in differentiating fast muscle fibres (A5, D5, E5, F5). Conversely *myf5* expressed was limited to the lateral edge of the myotome, in the posterior domain of So-1 (B5). From the 45ss-end of segmentation *myod1b* was rapidly downregulated in all but the superficial myotome throughout the entire length of So-1 (C5). As So-1 matured further, *smlc1* and *pax7*

expression respectively remained in the single-slow layer and external cell layer (G6 and H6). At this time, each MRF was expressed most strongly in superficial regions of the So-1 myotome, particularly in dorsal and ventral regions and at the level of the horizontal septum (A6-E6). *myf5* staining was still restricted to the posterior region of So-1, faintly along the whole superficial edge of the myotome, and more strongly in the dorsal-ventral-zones (B6). *myod1b* expression was very similar to *myf5* in any cross-section, although the staining was present throughout the length of So-1 (C6). *myod1a/1c*, *mrf4* and *myog* expression was not entirely restricted to the superficial myotome and each was also present in the deeper fast muscle fibres (A6, D6, E6, F6), although *myog* expression was comparatively fainter in ventral regions of the myotome (E6).

#### 7.4.6 The dynamics of rostral-caudal expression of MRFs during embryogenesis (Figs. 7.8-7.9)

Several MRFs were expressed in the adaxial myoblasts before *smlc1*. *myod1a* was expressed in a bi-lateral strip flanking the nascent notochord of some pre-somitic embryos (Fig. 7.7), although more often in adaxial progenitors of the PSM/somites from the 0-10ss and then maintained here in the newest somites/PSM throughout segmentation (Fig. 7.8, Bii, Dii, Fig. 7.9, Bii). *myf5* was expressed before or contemporaneously to *myod1a*, in two triangular fields of the PSM either side of the notochord, but did not co-localize with *myod1a* in pre-somitic adaxial myoblasts at this stage (Fig. 7.7). During early-mid segmentation, *myf5* was expressed throughout the mid-posterior domain of the newest somites, and in the anterior PSM displaying a pattern of interspersed strong and reduced signal where the newest two somites arose (Fig. 7.8, Bi, Di). *myf5* staining was also present throughout the tailbud, terminating adjacent to the notochord's end (Fig. 7.8, Bi, Di), but unlike other teleosts (e.g. Coutelle et al., 2001; Cole et al., 2004), was absent in the adaxial myoblasts of the anterior PSM/caudal somites (Fig. 7.8, Bi, Di) until the end of segmentation when a residual PSM remained (Fig. 7.9, Bi: faint signal: also see Fig. 7.11, B (2°C)).

As somitogenesis progressed, other MRFs were expressed in adaxial myoblasts of somites, but never in the PSM as for *myod1a*. From the ~20ss *myod1c* co-localised with *myod1a* in somitic adaxial cells (see chapter 4, Fig. 4.3, D) but also with *myf5/myod1b* in the posterior domain of the newest somites (not shown). As somites matured, *myod1b* spread anteriorly to encompass the whole myotome (chapter 4, Fig. 4.3, C), whereas *myf5* was initially downregulated and barely detected in the rostral somites at the 30ss (Fig. 7.8, Ai). *myog* mRNA was also detected at the 20ss and was present in the adaxial myoblasts of the final few caudal somites (Fig. 7.8, Biii, Diii, Fig. 7.9, Biii), before rapidly spreading to encompass the whole myotome of more anterior somites (Fig. 7.8, Aiii, Ciii, Fig. 7.9, Biii). The final *myod* family member expressed before the adaxial cells differentiated was *mrf4* at ~25ss, in a faint transient wave of rostral-caudal expression in adaxial myoblasts (mid-caudal somites of 30ss stage shown: Fig. 7.8, Biv).

*smlc1* marks the differentiation of adaxial cells to slow muscle myocytes, which started in the rostral somites of 25-30ss embryos and progressed in a caudal direction as newer somites matured (Fig. 7.8, Av). The progression of *smlc1* expression could be correlated with that of some *myod* family members whereas others seemed independent. For example, at the 25 ss, *mrf4*, expression was present in adaxial cells of the rostral-somites, immediately before/contemporaneously to *smlc1* expression (not shown and Fig. 7.8, Av) and similarly progressed caudally at this time. However, the rostral-caudal progression of *mrf4* was initially transient, disappearing in more rostral somites as it accumulated in newer somites. The timing of *mrf4* seemed to just precede *smlc1* so at the 30ss, when adaxial cells differentiated in rostral somites (Fig. 7.8, Av) *mrf4* had been downregulated at this site (Fig. 7.8, Aiv), but was expressed in the mid-caudal somites (Fig. 7.8, Biv), prior to *smlc1* expression here (Fig. 7.8, Bv). In the rostral somites at the 30ss, *myod1a/1c* transcripts had spread laterally away from the medial somite but this domain still overlapped with *smlc1* expression in differentiating medial adaxial cells (*myod1a* shown: Fig. 7.8, Aii). As somites matured, the adaxial cells migrated laterally,

evidenced by a wave of *smlc1* transcripts in the rostral somites of 45ss embryos (Fig. 7.8, Cv). By the end of segmentation, this migration was occurring from around the 10th most caudal somite (Fig. 7.9, Bv) and was completed in the rostral somites (Fig. 7.9, Av). During adaxial cell migration, *myod1a/1c* and *mrf4* transcripts moved away from the notochord and at the 45ss, mRNA for each gene was present in a broad v-shaped domain similar to *smlc1* expression (e.g. *myod1a*: Fig. 7.8, Cii, *mrf4*: Fig. 7.8, Civ). Concurrently, each of these MRFs remained in the adaxial myoblasts of the caudal somites, co-expressed with *myog*, before *smlc1* expression (e.g. Fig. 7.8, Dii, Diii Div, Dv, Fig. 7.9, Bii, Biii, Biv).

In contrast to the 30ss, where *myf5* was downregulated in maturing somites (Fig. 7.8, Ai), by the 45ss, *myf5* had accumulated in the rear quarter of the rostral somites at the superficial myotome, before the adaxial cells had completed their migration (Fig. 7.8, Ci, Cv). This pattern was maintained, so that at the end of segmentation (60-65ss), *myf5* was expressed along the entire outer edge of the myotome at the rear border of the rostral-mid somites (Fig. 7.9, Ai). *myf5* transcripts were present at this site before the adaxial cells had completed migrating, making it unlikely that this domain was limited to the slow layer. Instead, it is possible that *myf5* expression marks the earliest production of myotubes sourced from the external cell layer. In support of this, *pax7* was concurrently expressed in the external cell layer of the rostral somites (Fig. 7.9, Avi). In more caudal somites, where *myf5* had not reached the myotome border (Fig. 7.9, Bi), *pax7* was distributed throughout the somite, and particularly strongly in cells of the anterior domain (Fig. 7.9, Bvi). Thus, the migration of *pax7* mRNA to a position external to the myotome occurred at a similar time to the restriction of *myf5* mRNA at the posterior border of the myotome.

The expression domains of *myog* and *myod1b* from the 30-65ss also suggest a role for these transcription factors that is independent of adaxial cell migration. For example, both genes were

unchangingly present across the width/length of the myotome in all but the most caudal somites at the 20-65ss, irrespective of the migration state of adaxial cells (*myog* shown: Fig. 7.8, Aiii, Ciii, Fig. 7.9, Aiii, see also, *myod1b*: chapter 4; Fig. 4, C). Additionally, whereas the extension of *myod1b/myog* transcripts occurred in an anterior direction during somite maturation, the adaxial cells migrated laterally.

From segmentation until the eyed stage, the eyes developed pigmentation, the fin buds lengthened and all somites developed the chevron shaped. During this time *smlc1* expression remained in the single superficial layer of slow-twitch fibres (e.g. Fig. 7.9, Av, Cv) but were absent from caudal somites at the eyed stage (e.g. Fig. 7.9, Dv). By the eyed stage, *pax7* mRNA was detected throughout the external cell layer and dorsal spinal cord along the embryos rostral-caudal axis (Fig. 7.9, Cvi and Dvi). At this time, *myf5* expressing cells were present in the rear portion of all somites, mainly in the dorsal and ventral superficial fast myotome, adjacent to the horizontal septum and more faintly adjacent to the single slow muscle layer (Fig. 7.9, Ci and Di). *myod1b* was also expressed in similar regions at the superficial myotome, but was maintained along each somites length (not shown). Conversely, at the end of segmentation and the early-eyed stage, *myod1a/1c* and *mrf4* transcripts were detected to a greater or lesser extent throughout the entire myotome presumably in differentiating fast muscle (as in Fig. 7.9, Aii and Aiv). As embryos matured further (evidenced by increasing fin bud length), staining for these MRFs was reduced in the medial myotome but strongly maintained or upregulated in more superficial regions of the myotome, particularly in dorsal/ventral regions (*myod1a*: Fig. 7.9, Cii and Dii, *mrf4*: Civ and Div). Similarly at the early eyed stage *myog* expression was present to a greater or lesser extent throughout the myotome, but as embryos matured, expression was reduced in the medial myotome but maintained at the dorsal (and faintly at the ventral) edge of the myotome and adjacent to the horizontal myoseptum (Fig. 7.9. Ciii and Diii).

#### 7.4.7 Embryonic temperature and somitogenesis

Fig. 7.10 shows the relationship between the rate of Atlantic salmon somitogenesis and embryonic temperature. Segmentation proceeded from around 750-1700, 425-960 and 250-600 hours post-fertilization (hpf) at 2, 5 and 8°C respectively. A first order linear regression was fitted to data of developmental time versus somite number during the linear phase of somitogenesis, which occurs from the 0ss until the last few somites are added as segmentation is completed (Gorodilov, 1996). Using the regression equation from each plot, it was calculated that somitogenesis proceeded at a respective rate of 1 somite added each 15, 8 and 5 h at 2, 5 and 8°C.

#### 7.4.8 Embryonic temperature and the co-ordinated expression of MRFs

The expression of *myod1a*, *myog*, *myf5*, *mrf4* and *smlc1* was then investigated at three embryonic temperatures (2, 5 and 8 °C). *myod1a* and *myog* expression showed no consistent variation between temperature treatments for corresponding somite-stages (results not shown). In contrast, at several equivalent somite stages, replicated differences (in 6 embryos per stage) were recorded in the mRNA expression profiles of *myf5*, *mrf4* and *smlc1* with respect to somite stage. The expression patterns of each gene at 5°C was approximately intermediate between that observed at 2 and 8°C (not shown). *In situ* hybridization cannot be used as a quantitative tool for comparative analysis and therefore, only cases in which differences in staining intensity bordered on the presence or absence of transcripts are highlighted.

At the 30ss and 45ss, *myf5* staining was intense in the newly formed caudal somites, presomitic mesoderm and tailbud at 8°C, but faint at 2°C (45ss shown, Fig. 7.11, A). In embryos approaching the end of segmentation (with ~63 somites), *myf5* staining had reached somite number 58 at 8°C, but was almost absent from somites 59-63 (Fig 7.11, B). In contrast, at 2°C, an mRNA signal for *myf5* was detected in somites 58-63 and within the residual presomitic

mesoderm (Fig 7.11, B, see arrows in corresponding cross sections). These results can be interpreted to show that *myf5* expression was retarded with respect to somite-stage at 2°C, with staining in the caudal somites and PSM peaking and subsequently retracting earlier at 8°C compared to lower temperatures.

In somites 30-45 of 45ss embryos, *mrf4* transcripts were detected in the medial somite at both temperatures (not shown), but as somites matured, staining was more advanced at 8 °C. For example, in somites 20-25, *mrf4* transcripts were starting to extend laterally away from the notochord at 8°C but not 2°C (Fig. 7.12, B). Furthermore, *mrf4* staining had advanced into somites 1-15 at 8°C, but not 2°C (Fig. 7.12, A, see arrow on cross sections). Towards the end of segmentation, while the most caudal somites (53-63) had *mrf4* transcripts in adaxial cells at both temperatures (Fig. 7.12, D), in more rostral somites (numbers 43-50) the medial compartment showed a strong *mrf4* signal at 8°C, but was virtually unstained at 2°C (Fig. 7.12, C, see arrowheads on cross sections). These results indicate that the wave of *mrf4* expression in adaxial cells was retarded with respect to somite stage at lower temperatures.

As segmentation reached completion, the most newly formed somite with *smlc1* expression in the adaxial cells at 2°C was number 52-53, compared to 56-57 at 8°C (Fig. 7.13, C, D). Thus at an equivalent somite stage, *smlc1* expression was delayed by 4-5 somites at 2°C (illustrated by blue arrowhead in Fig. 7.13, C, D: also see arrow head on cross sections through equivalent somite number of 2 and 8°C embryos). In more rostral somites, a clear wave of *smlc1* transcripts could be seen migrating laterally away from the notochord between somite 43-48 at 8 but not 2°C (Fig. 7.13, A, B). In rostral somites (numbers 1-20) an *smlc1* signal was detected in the superficial slow-layer at 8 °C, but not 2 °C (not shown). Thus, consistent with the retardation of *mrf4* expression in adaxial cells, *smlc1* expression was delayed at 2 compared to 8°C.

Table 7.1. Primer details for chapter 7.

<b>Primer name</b>	<b>Product</b>	<b>Related accession</b>	<b>Primer sequence (5'-3')</b>
<i>myf5</i> F1	whole cds	DQ452070	f: ATGGATGTCTTCTCCCAGTCC
<i>myf5</i> R1	“	“	r: TCACAATACGTGGTACACAGGTC
<i>mrf4</i> F1	Nucleotides 1-648	DQ479952	f: ATGATGGACCTTTTTGAGACC
<i>mrf4</i> R1	“	“	r: GATTGATGACAGGCCGAAGAAG
<i>mrf4</i> Race	3' cds/UTR <i>mrf4</i>	DQ479952	f:GAGTCTTCAGCGTCCACCAGCCTTCTTCG
<i>myog</i> F1	whole cds	DQ294029	f: CTAGCGTCGACCAGTATGGAG
<i>myog</i> R1	“	“	r: CTCTGGGTTTATTTGGGAATG
<i>smlc1</i> F1	whole cds	DQ916288	f: CTGTCCTCCTGTGGCTCCTG
<i>smlc1</i> R1	“	“	r: TTAAGATGCCATGACGTGTTTTAC
<i>pax7</i> F1	Nucleotides 502-1119	AJ618975	f: CTGTGAGTTCCATCAGCCGAG
<i>pax7</i> R1	“	“	r: TGGGGTACTCAGGATGCTC

▼ exon 1

1 ATGGATGTCTTCTCCAGTCCCAGATTTTCTATGACAGCGCCTGTGCCTCCTCGCCAGAG 20  
M D V F S Q S Q I F Y D S A C A S S P E

61 GACCTGGACTTCGGCCCCGGGAACTGGATGGCTCAGAGGAGGACGAGCACGTCCGGGTC 40  
D L D F G P G E L D G S E E D E H V R V

121 CCTGGGACTCCTCACCAGGCGGGTCACTGCCTTCAGTGGGCCTGCAAGGCCTGCAAGCGT 60  
P G T P H Q A G H C L Q W A C K A C K R

181 AAGTCCAGCACGGTGGACCGGCGGGCTGCCACCATGAGGGAACGACGCCGGCTGAGA 80  
K **S S T V D R R R A A T M R E R R R** L R

241 AAGGTGAACCACGGCTTCGAGGCTCTGAGGCGCTGCACCTCAGCCAACCACAGCCAGAGG 100  
**K V N H G F E A L R R** C T S **A N H S Q R**

301 CTGCCTAAGGTGGAGATCCTGCGCAACGCCATCCAGTACATCGAGAGCCTCCAGGAGCTG 120  
**L P K V E I L R N A I Q Y I E S L Q E L**

361 CTCCATGAGCATGTGGAGAACTACTACGGCCTTCTGGGGAGAGCAGCTCAGAGCCTGGG 140  
**L** H E H V E N Y Y G L P G E S S S E P G

▼ exon 2

421 AGCCCCCTCGTCCAGCCGCTCCGACAGCATGGTTGACTGTAACATTCCTGTTGTGTGGCCT 160  
S P S S S R S D S M V D C N I P V V W P

▼ exon 3

481 CAGATGAACACAAGCTATGGCAACAACACTACAGTTATACTAAGAATGTGAGCTCTGGAGAG 180  
Q M N T S Y G N N Y S Y T K N V S S G E

541 AGAGGTGCTGGTGCCTCCAGCCTGGCCCCGCTGTCTAACATAGTAGATCGCCTCTCCTCG 200  
R G A G A S S L A R L S N I V D R L S S

601 GTGGATGCCAGTGCCCCAGCAGGGCTCAGAGATATGCTTACCTTCTCGCCCTCCAGCACC 220  
V D A S A P A G L R D M L T F S P S S T

661 GACTCCCAGCCTTGCACTACAGAAAGCCCCGGGACCAGACCTGTGTACCACGTATTGTGA 240  
D S Q P C T T E S P G T R P V Y H V L \*

Fig. 7.1. Atlantic salmon Myf5 at the mRNA and AA level. The sense nucleotide stand is shown (numbers to left of figure) with codons shown above translated AAs (numbers to right of figure). Exon-exon boundaries are identified (▼) and the translation start and stop codon are underlined in bold and italics respectively. Also shown at the AA level are the basic region (light grey shading) and the HLH (dark grey shading).

▼ exon 1

1 ATGGAGCTTTTTGAGACCAACCCCTACTTCTTCCCGACCAACGCTTCTACGAAGGGGGG  
M E L F E T N P Y F F P D Q R F Y E G G 20

61 GACAACTTCTACCAGTCCCGGCTGCCGGGTGGGTATGACCAGGGGGGCTACCAGGAGCGC  
D N F Y Q S R L P G G Y D Q G G Y Q E R 40

121 GGGGGTTCCATGATGGGGCTTTGTGGGGGTCTATCCGGGAGGGTTGGGGTAGGGTTGGGT  
G G S M M G L C G G L S G R V G V G L G 60

181 GGAGGCATGGAGGACAAGGCAACCCCTCCGGTCTCTCGCCCCACCCGGAGCCCCACTGC  
G G M E D K A T P S G L S P H P E P H C 80

241 CCCGGCCAGTGCCTACCCTGGGCCTGCAAGCTGTGCAAACGCAAGACTGTGACCATGGAC  
P G Q C L P W A C K L C K R K T V T M D 100

301 CGACGGAAAGCGGCCACAATGCGGGAGAAGAGGAGGCTGAAGAAGGTGAACGAGGCATTC  
R R K A A T M R E K R R L K K V N E A F 120

361 GAGGCCCTGAAGAGGAGCACCCCTGATGAACCCCAACCAGAGGCTGCCCAAGGTGGAGATC  
E A L K R S T L M N P N Q R L P K V E I 140

421 CTGAGGAGTGCCATCCAGTACATTGAGAGGCTGCAGGCACTTGTCTCCTCCCTCAACCAG  
L R S A I Q Y I E R L Q A L V S S L N Q 160

481 CAGGAGAACGACCAGGGAACACAGGGCTTACACTACCGCACCGGACCTGCCCAACCCAGG  
Q E N D Q G T Q G L H Y R T G P A Q P R 180

▼ exon 2

541 GTCTCGTTCGTCGAGTGAGCAGGGATCAGGCAGCACCTGCTGTAGCAGCCCAGAGTGGAGC  
V S S S S E Q G S G S T C C S S P E W S 200

▼ exon 3

601 AACACCTCAGACCACTGTACCCAGAGCTACAGCAACGAGGACCTCCTGAGTGCAGACTCT  
N T S D H C T Q S Y S N E D L L S A D S 220

661 CCAGAGCAGACTAACCTGCGCTCTCTGACGTCCATCGTGGACAGCATCACAGCAGCAGAG  
P E Q T N L R S L T S I V D S I T A A E 240

721 GGGGCTCCGGTGGCCTACCCTGTACCTGTGGACATTCCCAAATAAACCCAGAGA  
G A P V A Y P V P V D I P K \* 260

Fig. 7.2. Atlantic salmon Myog at the mRNA and AA level. The sense nucleotide stand is shown (numbers to left of figure) with codons shown above translated AAs (numbers to right of figure). Exon-exon boundaries are identified (▼) and the translation start and stop codon are underlined in bold and italics respectively. The 3' untranslated region is shown in red font. Also shown at the AA level are the basic region (light grey shading) and the HLH (dark grey shading).

▼ exon 1

1 **ATG***GATGGACCTTTTTGAGACCCACACTTATTTCTTCAACGATCTGCGCTATCTCGAGGGA*  
M M D L F E T H T Y F F N D L R Y L E G 20

61 ATCATGGACCATTGCAACATTTGGACATGGCGGGGGTGTCCCCTTTGTACCACGGGAAT  
D H G P L Q H L D M A G V S P L Y H G N 40

121 GACAGCCCGTTGTCACCTGGGGGGGATCCGTCCGAGACTGGATGTGACAGCAGCGGAGAG  
D S P L S P G G D P S E T G C D S S G E 60

181 GAGCATGTCCTCGCACCCCCTGGTCTTCAGCCGCACTGCGAGGGACAGTGCCTCATCTGG  
E H V L A P P G L Q P H C E G Q C L I W 80

241 GCTTGTAAGGTTTGTAAAAGAAAGTCTGCACCGACCGACAGGCGCAAAGCGGCCACTCTC  
A C K V C K R K **S A P T D R R K A A T L** 100

301 AGAGAAAGAAGGCGGCTCAAGAGGATCAATGAAGCATTCGATGCGTTGAAGAAAAAGACC  
**R E R R R** L K **R I N E A F D A L K K K T** 120

361 GTGCCCAATCCGAACCAGCGGCTGCCCAAAGTGGAGATTTTACGCAGCGCCATAAACTAC  
**V P N P N Q R L P K V E I L R S A I N Y** 140

421 ATCGAGCAATTGCAGGACCTGTTGCATACACTGGATGAGCAAGAAAACCCCCACAAAAT  
**I E Q L Q D L L** H T L D E Q E N P P Q N 160

▼ exon 2

481 GGCTATAACGTGAAAAGAACACCATGCGTCCAATAAGGAGTACCATTGGAAGAAGAAGTGT  
G Y N V K E H H A S N K E Y H W K K N C 180

▼ exon 3

541 CAAAAGTGGCAGACCTCAGCTGATCATTCCAATGCACCAATGACGAATCAGAGAGAAGGC  
Q N W Q T S A D H S N A P M T N Q R E G 200

601 TTCACTGAGTCTTCAGCGTCCACCAGCCTTCTTCGCCTGTCATCAATCGTTGACAGCATC  
F T E S S A S T S L L R L S S I V D S I 220

661 TCAAGTGAAGAGAAACCGACTTGCAACGAAGAAGTCTCAGAAAAAT**AA*****TGCATGATTTAT***  
S S E E K P T C N E E V S E K \* 240

721 **TGGAATTTTGTAGCCTGTATAAGCGACGTCAGCATTTTCGTATTTCCATTGTCTATTTCGA**  
**AATTATTTTCACTTCTTTATTCATATGTTTTAGTTTCCATTATTATATTCATTTGTACAA**  
**ATTAACGGGCATTTTGTGGTCACTGTTTTCTTTTAAATGTATGAATGGTCAATATTTTC**  
**TTCTAAAGTATGACAACAATGGAACATAATATTTTATTTCCAAATGGACATTTGATAA**  
**TTGTAATATTTTCTAATATATTAACAACCTAATTTATTTTACATATAAAAGGCAGAAAC**  
**AAATCCTCAAGAGATTCATTGTTGTATTTGAATATTTCTATATGACCTGACCAACAATA**  
**AATTGTAAGCATTGTATCC****AAAAAAAAAAAAAAAAAAAAAAAAAAAA**

Fig. 7.3. Atlantic salmon Mrf4 at the mRNA and AA level. The sense nucleotide stand is shown (numbers to left of figure) with codons shown above translated AAs (numbers right of figure). Exon-exon boundaries are identified (▼) and the translation start and stop codon are underlined in bold and italics respectively. The 3' untranslated region is shown in red font and a polyadenylation signal (AATAAA) and poly-A tail are respectively shown in blue and green font. Also shown at the AA level are the basic region (light grey shading) and the HLH (dark grey shading).



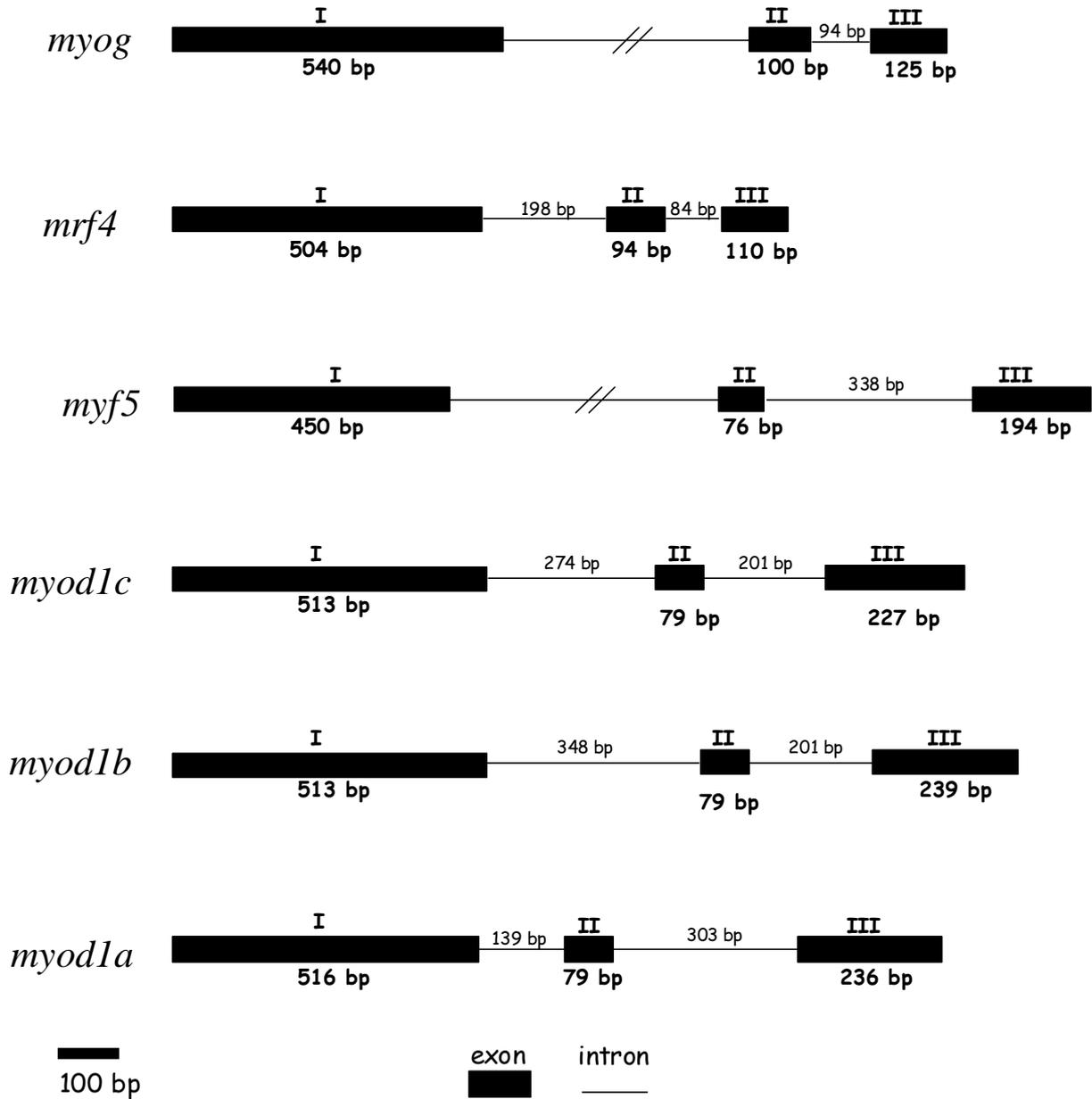


Fig. 7.5. Intron-exon structures of Atlantic salmon myogenic regulatory factors. Each gene is represented by three exons (black boxes) and two introns (lines). The known sizes of exons and introns are shown. Introns with a double line are of unknown size, but in each case were shown by sequencing to be greater than 1kb. However, all intron-exon boundaries are supported experimentally.

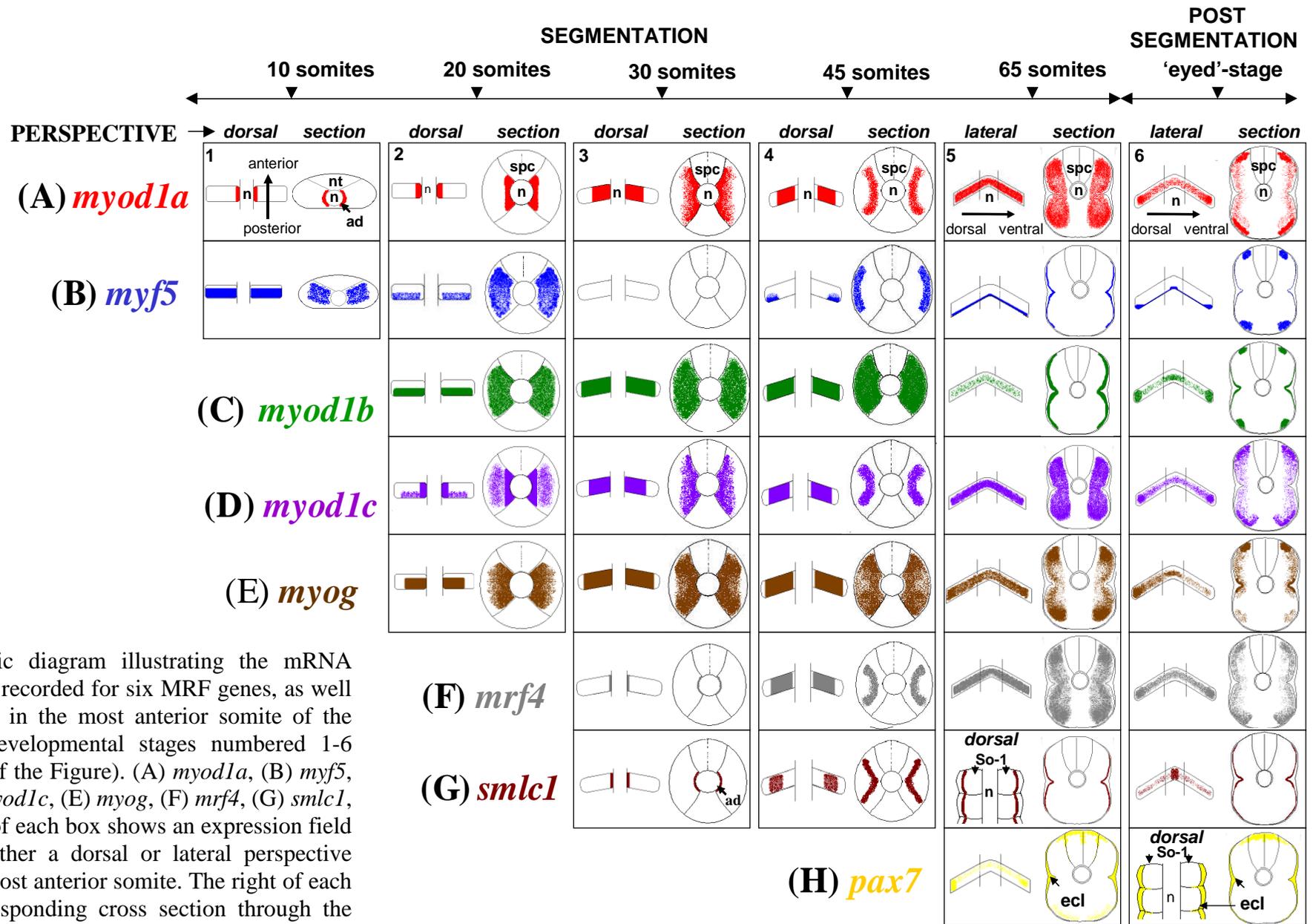


Fig. 7.6. Schematic diagram illustrating the mRNA expression patterns recorded for six MRF genes, as well as *smlc1* and *pax7* in the most anterior somite of the Atlantic salmon developmental stages numbered 1-6 (shown at the top of the Figure). (A) *myod1a*, (B) *myf5*, (C) *myod1b*, (D) *myod1c*, (E) *myog*, (F) *mrf4*, (G) *smlc1*, (H) *pax7*. The left of each box shows an expression field as viewed from either a dorsal or lateral perspective (indicated) of the most anterior somite. The right of each box shows a corresponding cross section through the region of expression in that somite. Abbreviations: ad: adaxial cells; ecl: external cell layer, n: notochord, nt, neural tube, spc: spinal cord, So-1: somite 1.

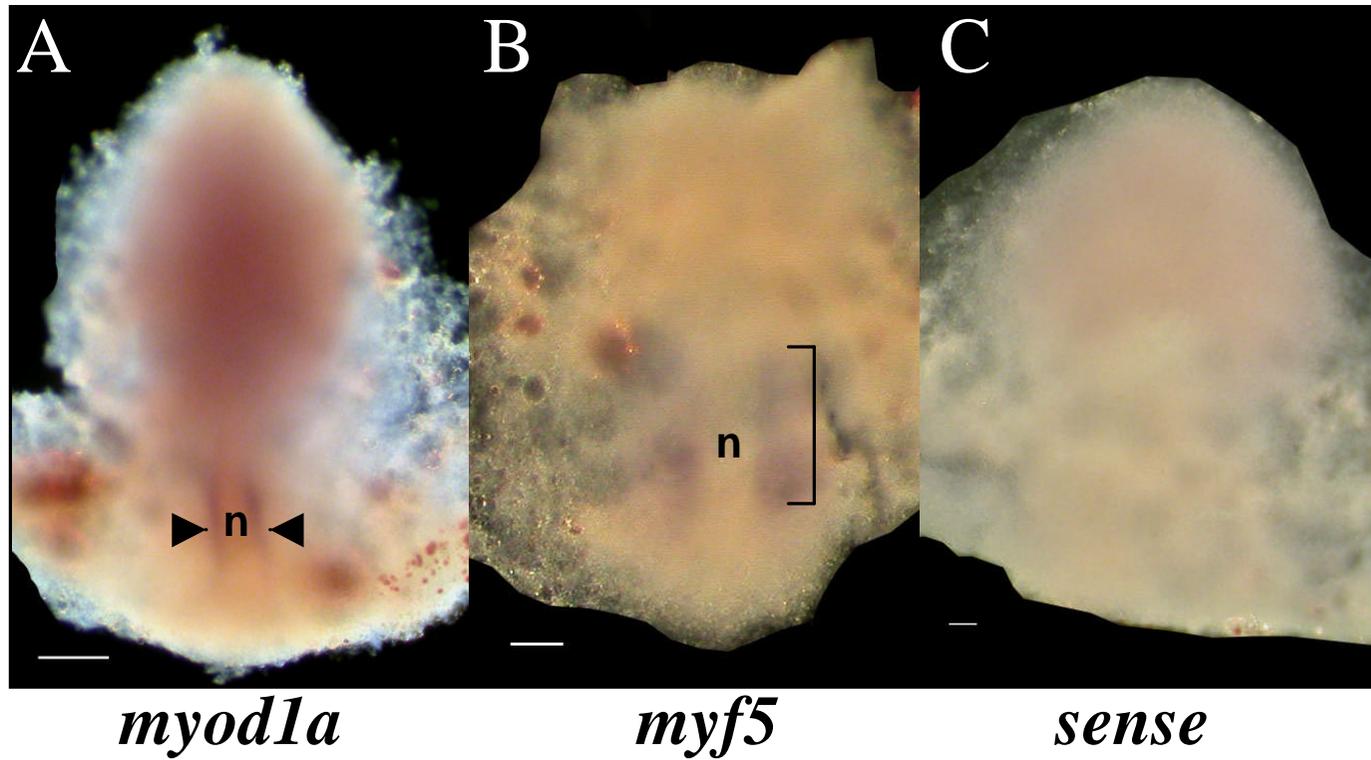


Fig. 7.7. **A.** Prior to the onset of segmentation, *myod1a* was expressed in adaxial myoblasts either side of the nascent notochord. **B.** *myf5* was expressed broadly in the presomitic mesoderm, but not in the adaxial myoblasts **C.** Sense controls produced no staining. Embryos were manually removed from the yolk-sac and photographed in darkfield. Abbreviation: n: notochord.

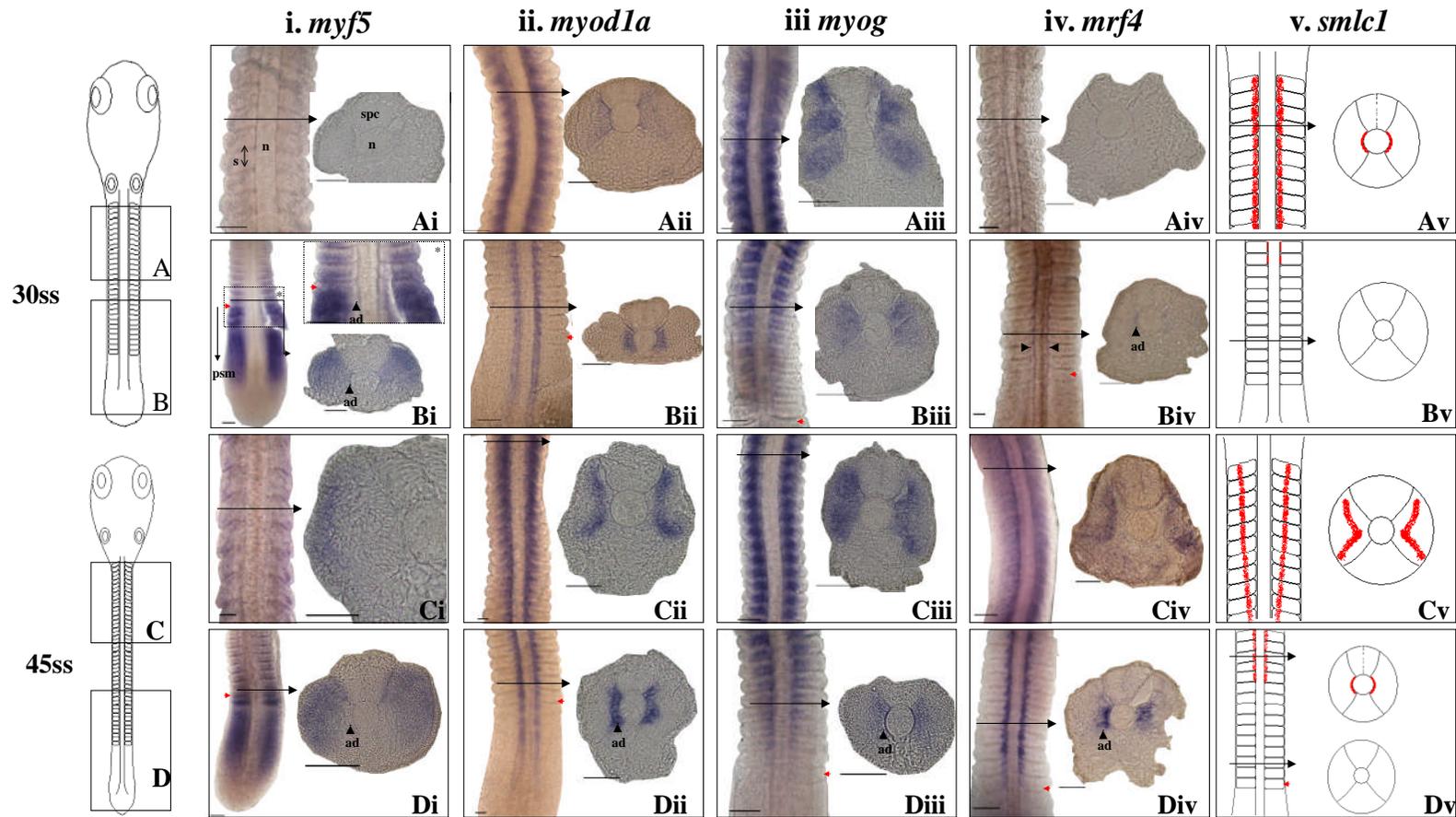


Fig. 7.8. mRNA expression patterns of MRFs and a schematic representation of *smlc1* expression during the 30-45 ss. *smlc1* was included schematically, as staining was faint at these stages. Numbers i-v represent cRNA probes for: i. *myf5*, ii. *myod1a*, iii. *myog*, iv. *mrf4* and v. *smlc1*. Letters A-D represent specific regions marked on schematic drawings of whole embryos from different stages (diagram on left side of figure). Images on the left of each box are dorsal perspective flatmounts. Images on the right of each box are 18  $\mu\text{m}$  cryosections from the region identified by a black arrow. Red arrows show the position of the last somite. \* Shows a magnified flatmount of *myf5* to show the lack of expression in adaxial myoblasts. Abbreviations are as in Fig. 7.6 with the addition of: s: somite, psm: presomitic mesoderm. Scale bars are 50 $\mu\text{m}$ .

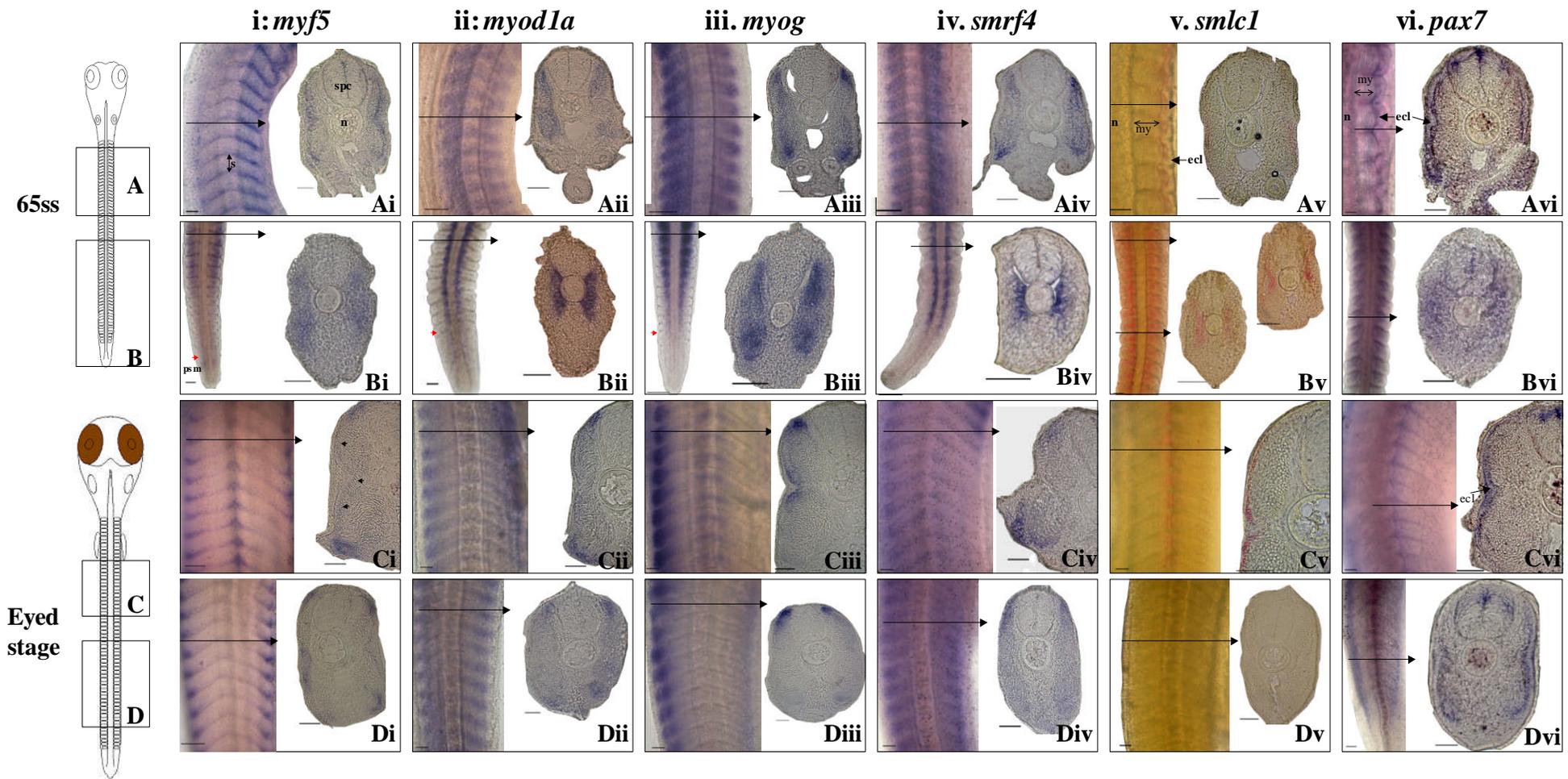


Fig. 7.9. mRNA expression patterns of MRFs as well as *smlc1* and *pax7* at the 65ss and during the eyed stage. The numbering and lettering system is equivalent to that used in Fig. 7.8, except that vi. represents the *pax7* cRNA probe. Images on the left of each box are lateral perspective flatmounts, except for Av-Avi and Bv-Bvi, which are mounted from the dorsal perspective. Images on the right of each box are 18  $\mu$ m cryosections from the region identified by a black arrow. Abbreviations and red arrows are as in Fig. 7.6/7.8. Scale bars are 50 $\mu$ m.

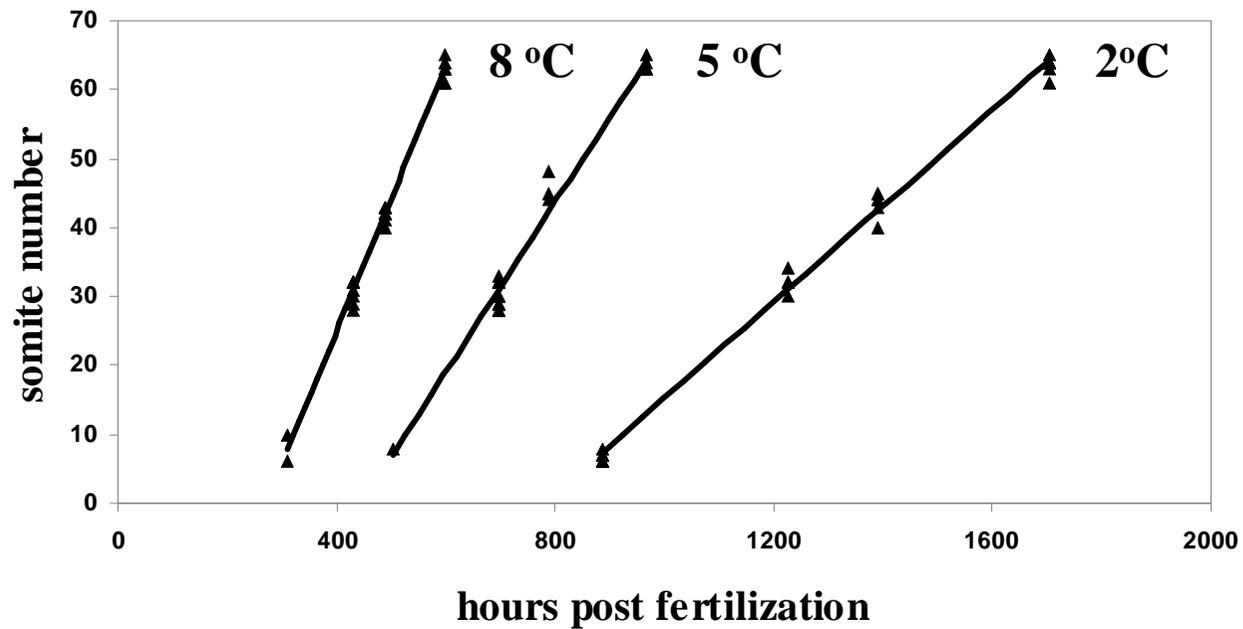


Fig. 7.10. Rate of somitogenesis in Atlantic salmon reared at 2, 5 and 8°C. First order linear regressions were fitted to each group and the following equations were obtained: 2°C: somite number =  $-54.32 + 0.0696 * \text{hpf}$ ,  $R^2 = 99.5\%$  (n = 23). 5°C: somite number =  $-54.7 + 0.123 * \text{hpf}$ ,  $R^2 = 98.2\%$ , (n = 21). 8°C: somite number =  $-51.8 + 0.192 * \text{hpf}$ ,  $R^2 = 99.3\%$ , (n = 28).

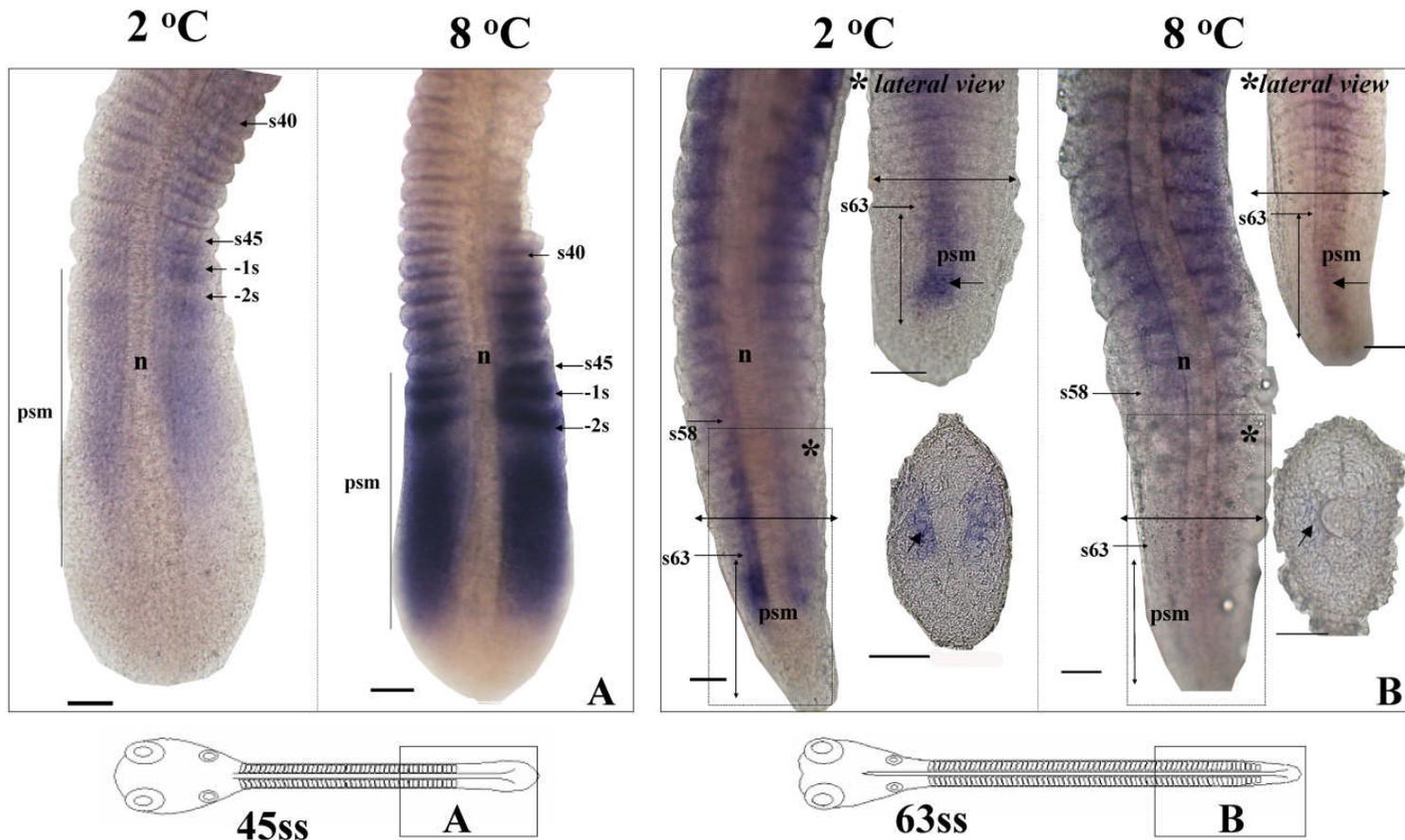


Fig. 7.11. Representative images showing the temperature associated heterochronies observed in *myf5* expression in Atlantic salmon embryos incubated at 2 or 8 °C. Letters A-B within the boxed regions on the schematic embryo show the position of corresponding images. Flatmount images are viewed from the dorsal perspective except where indicated by a star in boxes B<sup>2</sup>/B<sup>8</sup>. Somite number is shown as s(n) where s = somite, n = number and the most caudal somite is the highest numerically. Abbreviations are as in Fig. 7.8. Scale bars are 50μm.

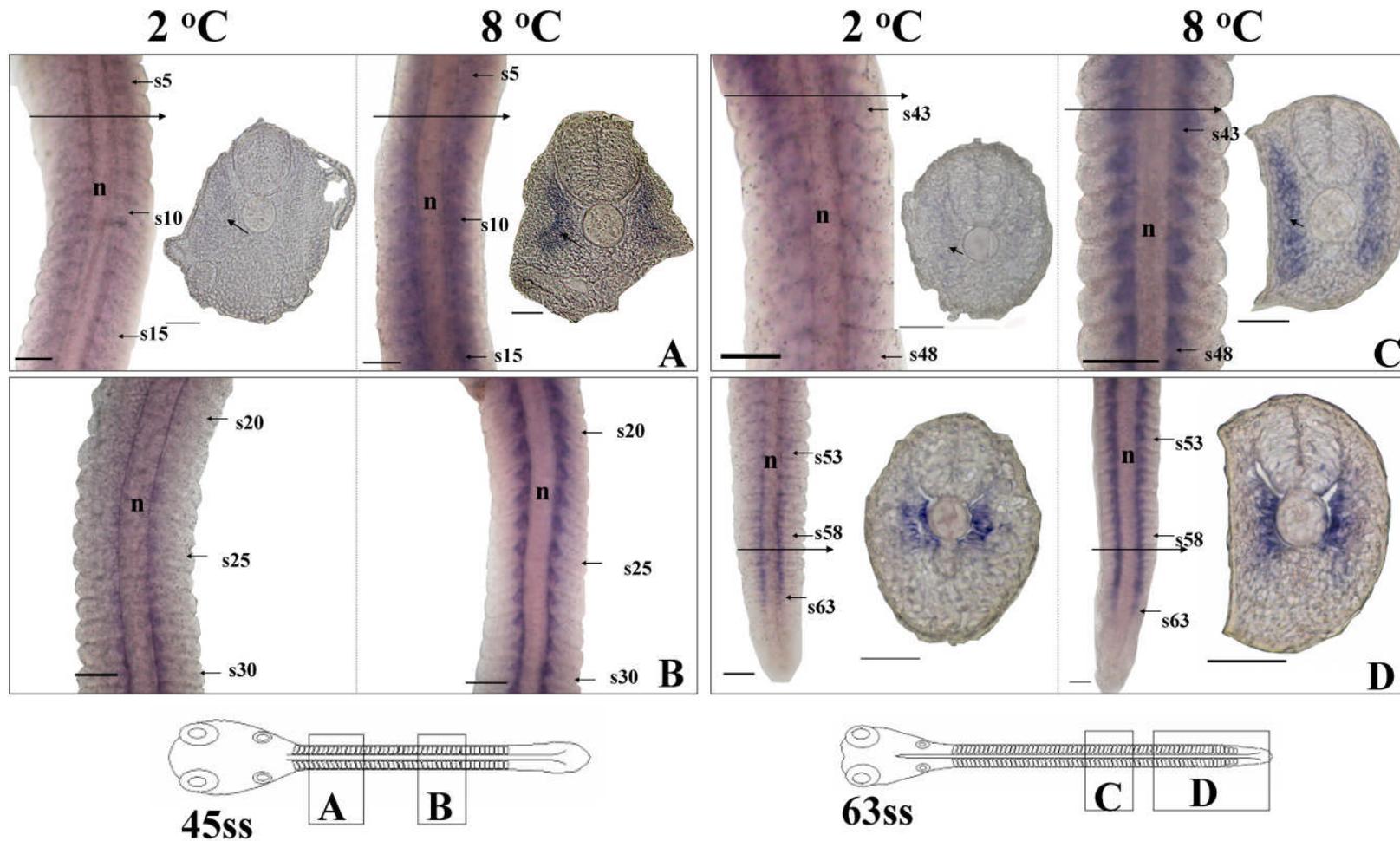


Fig. 7.12. Representative images showing the temperature associated heterochronies observed in *mrf4* expression in Atlantic salmon embryos incubated at 2 or 8 °C. Lettering and numbering system is the same as in Fig. 7.11. Flatmount images are viewed from the dorsal perspective. Scale bars are 50µm.

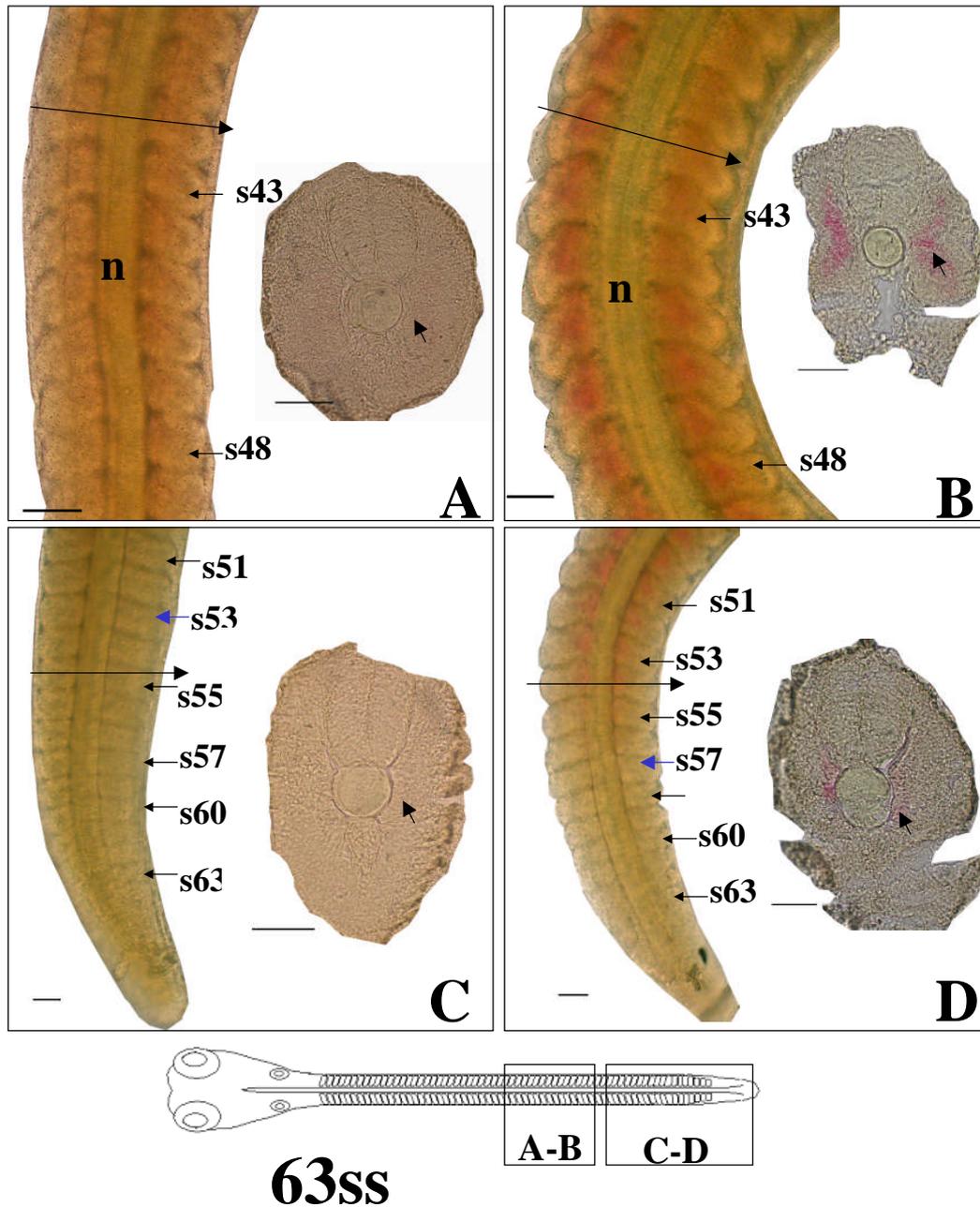


Fig. 7.13. Representative images showing the temperature associated heterochronies observed in *smlc1* expression in Atlantic salmon embryos incubated at 2 or 8 °C. Lettering and numbering system is the same as in Fig. 7.11. Flatmount images are viewed from the dorsal perspective. The blue arrow shows the last somite considered to have *smlc1* expression. Abbreviations are as in Fig. 4. Scale bars are 50µm.

## 7.5 Discussion

### 7.5.1 MRFs in Atlantic salmon and other teleosts

In this chapter, the coordinated expression of six Atlantic salmon MRFs was initially studied during embryonic myogenesis. In zebrafish, *myod* and *myf5* are the first MRFs expressed in adaxial myoblasts and fast MPCs followed by *myog* in differentiating muscles (Weinberg et al., 1996; Coutelle et al., 2001). Although the order that these MRF transcripts appear in Atlantic salmon (section 7.4.6) is similar to that described in zebrafish, there are some notable differences in expression patterns, which are probably related to the tetraploid nature of the salmonid genome. The lineage leading to modern salmonids has undergone two WGDs relative to the common tetrapod ancestor (Jaillon et al., 2004; Allendorf and Thorgaard, 1984). The salmonid-specific genome duplication is thought to have occurred 25-100 Mya (Allendorf and Thorgaard, 1984) and ~25-50% of the duplicated genes were lost from the genome and are represented by a single paralogue (Bailey et al., 1978). A single *myog* gene has been described in zebrafish (Weinberg et al., 1996), common carp (*Cyprinus carpio*) (Cole et al., 2004), flounder (*P. olivaceus*) (Xu et al., 2007), striped seabass (*Morone saxatilis*) (Tan et al., 2002) as well as rainbow trout (Delalande and Rescan, 1999) and Atlantic salmon (present study). The highly conserved expression pattern of *myog* during embryonic myogenesis in all these species suggests that *myog* is retained as a single gene in salmonids. In other cases, duplicated genes have been retained. For example, salmonid fish have three *myod* paralogues, which based on phylogenetic and experimental evidence (chapter 3, 4), are thought to have arisen from a single gene orthologous to zebrafish *myod1* via a whole genome and subsequent local duplication. The *myod1* paralogues are differentially expressed and individually recapture separate expression domains of the single *myod1* gene of zebrafish (chapter 4). Additionally, while Atlantic salmon

*myf5* was expressed in the posterior domain of recently formed somites, the anterior PSM and tailbud (Fig. 7.8, Bi, Di), as described in zebrafish (Coutelle et al., 2001), common carp (Cole et al., 2004) and flounder (*P. olivaceus*) (Tan et al., 2006), in contrast to these teleosts, it was not expressed in adaxial cells during early-late segmentation (Fig. 7.8, Bi, Di). Considering the importance of the adaxial cell *myf5* expression domain (Coutelle et al., 2001), this finding is most consistent with the presence of two subfunctionalized *myf5* paralogues. To examine the possibility of a second Atlantic salmon *myf5* paralogue, primers were designed in conserved regions of *myf5* to amplify intron 2. A single band was obtained by PCR using a genomic DNA template and despite multiple sequencing attempts, a single sequence orthologous to the *myf5* gene characterised here was consistently recovered. Furthermore, a single *myf5* orthologue was retrieved when BLAST searches were performed at the salmon genome project (<http://www.salmongenome.no/cgi-bin/blast.cgi>), TGI (Atlantic salmon/rainbow trout databases at: <http://compbio.dfci.harvard.edu/tgi/>) and cGRASP (<http://web.uvic.ca/cbr/grasp/>) databases. An interesting alternative possibility is that the two *myf5* genes produced during the tetraploidization of the salmonid genome became subfunctionalized before one paralogue (expressed in adaxial myoblasts) was lost, perhaps because of the abundance of transcribed *myod1* paralogues in adaxial cells (chapter 4) and known redundancy of Myf5/MyoD in myogenic specification (Rudnicki et al., 1993; Hammond et al., 2007). If a second Atlantic salmon *myf5* paralogue does not fulfil the known role of teleost *myf5* in adaxial cell specification (Coutelle et al., 2001), then embryonic slow muscle development in salmonids is likely to vary significantly to other teleosts. A morpholino-based knockdown of individual salmonid MRFs would be informative in this respect.

No previous *mrf4* expression pattern had been described in teleosts during the write up of the concurrent paper to this chapter (Macqueen et al., 2007). However, when the paper was *in press* a contribution from Simon Hughes laboratory was published that rigorously described the

zebrafish *mrf4* expression domain during embryonic myogenesis, taking advantage of several muscle-mutant lines and multiple mRNA markers of fast/slow differentiation (Hinits et al., 2007). The salmon and zebrafish *mrf4* expression pattern share several similarities, but some obvious differences. Clear similarities include that both *mrf4* expression domains are initially restricted specifically to slow myoblasts of the somite, and never overlap with *myod1* in progenitors to these cells in the presomitic mesoderm. Additionally, like salmon, zebrafish *mrf4* is expressed in differentiating migrating adaxial cells and subsequently is upregulated in differentiating fast muscle cells (Hinits et al., 2007). However, whereas salmon *mrf4* is first initiated when ~25-30/65 somites have formed i.e. at ~40% of the final somite number, zebrafish *mrf4* is first expressed when 5/30 somites are formed i.e. at ~15% of the final somite number. This initial phase of zebrafish expression is Hedgehog dependent and is mainly limited to muscle pioneer (MP) cells, evidenced by wildtype expression and the absence of expression in MP mutants (Hinits et al., 2007). This early expression pattern is completely lacking in salmon, similar to the *myf5* adaxial expression domain. Again, this finding suggests the presence of another uncharacterised *mrf4* gene in salmonids that is a paralogue of the gene studied in this chapter. It could then be predicted that the putative *mrf4* paralogue would be expressed in MP cells prior to adaxial cell differentiation and together with the paralogue characterised here, would recapitulate the expression pattern of zebrafish *mrf4*. Interestingly, during the early part of this project, the presence of two Atlantic salmon *mrf4* paralogues was suspected, when in a real-time PCR assay for salmon *mrf4* a 'shoulder' was observed in the dissociation analysis, indicating the presence of a second highly similar PCR product. Accordingly, more clones for the *mrf4* genomic DNA product amplified by primers *mrf4 F1* and *R1* (Table 7.1) were sequenced, revealing a second distinct *mrf4* product that was conserved to such a high extent that it was considered unlikely to be a paralogue. Thus, the further characterisation of this second *mrf4* sequence was deemed unnecessary and accordingly, it was not submitted to GenBank. However, it has very recently been shown that an *mrf4* sequence homologous to this second

genomic sequence, mapped to a distinct chromosomal location in the Atlantic salmon genome compared to the *mrf4* gene characterised in this chapter (Moghadam et al., 2007; accession number; EF450078). Fig. 7.14 shows an alignment of the two Atlantic salmon MRF4 paralogues at the genomic level and I have tentatively denoted them *mrf4a* (current gene) and *mrf4b* (novel gene). Considering the high sequence identity conserved between the *mrf4* paralogues, it is possible that the cRNA probe used for *mrf4* in this study cross reacted with mRNA transcripts of both genes. In which case, the differences in expression between salmon and zebrafish remain unexplained and it cannot be excluded that a further *mrf4* paralogue is conserved in salmonids.

### 7.5.2 *Mrf4* and myogenic specification, a trait lacking in teleosts?

When *Mrf4* expression was not compromised in *myod/myf5* double mutant mice, normal myogenesis occurred, indicating that *Mrf4* can substitute for *Myf5/myoD* in initiating muscle growth (Kassar-Duchossoy et al., 2004). This dual role for *mrf4* is likely not conserved in the teleosts, since zebrafish lacking translation of *MyoD1/Myf5* proteins show no *mrf4* expression and lack muscle (Hinits et al., 2007; Hammond et al., 2007) and further, *mrf4* is expressed in differentiating muscles only (current chapter; Hinits et al., 2007). It is interesting to note that the Helix-3 of salmon *Mrf4* is more distinct from *MyoD1a* (8/15 substitutions: Fig. 7.4), than in a comparable alignment of mouse *Mrf4* vs. mouse *MyoD* (6/15 substitutions: see Bergstrom and Tapscott, 2001). Substituting the helix-3 of mouse *MyoD* with the equivalent *Mrf4* region resulted in a chimera that activated endogenous muscle specific genes as efficiently as wildtype *MyoD* (Bergstrom and Tapscott, 2001). The equivalent region of mouse *Myog* (with 8/15 substitutions i.e. the same as salmon *Mrf4*) could not replace the original *MyoD* motif. It is possible that the increased number of substitutions in the helix-3 of *Mrf4* compared to mammalian *Mrf4* has resulted in a reduced potency for myogenic specification whilst maintaining a role in differentiation.

### 7.5.3 Salmon MRFs and the external cell layer

Following the end of segmentation, each MRF was expressed in zones of new myotube production that occur at the lateral edge of the fast myotome (stratified hyperplasia), particularly in dorsal and ventral areas and adjacent to the horizontal myoseptum. *myf5* was expressed at the superficial edge of the myotome in rostral somites from the 45ss, initially prior to the completion of adaxial cell migration and was thus independent of the first wave of slow muscle differentiation. It is possible that *myf5* marked the onset of stratified hyperplasia, which began at a similar stage of development in the closely related salmonid, *S. trutta*, evidenced by *myod/myog* expression (Steinbacher et al., 2007). *myod1a/1c*, *mrf4* and *myog* expression in the bulk of the myotome was reduced from the end of segmentation onwards, but maintained (or upregulated) at the lateral edge of the fast myotome at either the dorsal and/or ventral extremes and/or adjacent to the horizontal myoseptum. The source of additional embryonic fast muscle fibres is likely to be the external cell layer, which is marked by *pax7* expression at this stage (Fig. 4, H5 and H6 and Fig. 7.9, Avi, Cvi, Dvi) (Hollway et al., 2007; Stellabotte et al., 2007; Steinbacher et al., 2006; Devoto et al., 2006). Unfortunately, due to the lack of staging criteria at the eyed stage, it was not possible to establish the presence of temperature induced MRF heterochronies in these zones of stratified hyperplastic growth.

### 7.5.4 Heterochronies in MRF expression at different temperatures

In this chapter it was shown that altering egg incubation temperature produces heterochronies in the expression of some MRFs, but not others. Thus, whereas *myod* and *myog* expression showed no consistent differences with temperature with respect to developmental-stage, the expression of *mrf4* and *myf5* were retarded at 2°C compared to 8°C. Consistent with the delayed expression of *mrf4* in adaxial cells at 2°C, the mRNA signal for the adaxial cell differentiation marker *smc1* was also delayed. Since *mrf4* was also expressed in differentiating fast muscle fibres subsequent

to segmentation (when heterochronies were not examined), it is also likely that the fast myotome was affected by embryonic temperature. The finding that the relative timing of *myod* and *myog* expression was independent of temperature parallels observation in Atlantic cod (Hall et al., 2003), Atlantic herring (Temple et al., 2001), common carp (Cole et al., 2004) and Atlantic halibut (Galloway et al., 2006), but differs from the result reported in rainbow trout (Xie et al., 2001). In zebrafish the expression of *myod1* and *myog* in fast MPCs of the posterior somite is activated by retinoic acid and regulated through Fgf8 (Hamade et al., 2006; Groves et al., 2005). However, the transient expression of *myf5* in these fast MPCs occurs independently of Fgf8 (Groves et al., 2005), as does the expression of *mrf4* in differentiating adaxial cells (Himits et al., 2007). Thus, the embryonic heterochronies in *myf5* and *mrf4* likely occurred through pathways independent of the Fgf8 regulation of *myod* and *myog*. Additionally, since the expression of salmon *myod1a* was not temperature dependent in pre-differentiated adaxial myoblasts, temperatures affect on slow-myogenesis may be limited to adaxial cell differentiation.

As a consequence of the heterochronies in *mrf4* and *myf5* expression, the ratio of the individual MRFs at each developmental stage was a function of environmental temperature. It is known that the different MRF proteins vary in their intrinsic abilities to initiate myogenesis or promote muscle differentiation (Bergstrom and Tapscott, 2001; Ishibashi et al., 2005). For example, whilst Myf5 and MyoD targeted a similar array of genes involved in myogenic specification, MyoD was markedly more efficient at inducing muscle differentiation genes (Ishibashi et al., 2005). Functional analysis in mouse has shown that MyoD strongly up-regulates *Capn2*, a protease required for myoblast-myotube fusion whereas Myog has a weak effect and Myf5 no effect (Dedieu et al., 2003). Using a combination of genome-wide transcriptional factor binding and expression profiling in the mouse a total of 126 genes were identified which bound MyoD (Blais et al., 2005). Many of these genes were transcription factors that propagate and amplify signals initiated by the MRFs (Blais et al., 2005). MyoD and Myog occupied 91 and 137

promoters in differentiating myotubes indicating the MRFs recognise distinct, but overlapping targets (Blais et al., 2005). Of particular interest was the finding that MRFs bind a set of genes involved in synapse specification and the function of the neuromuscular junction (Blais et al., 2005). In Atlantic herring, embryonic temperature has been shown to produce major changes to the timing of development of neuromuscular junctions in the myotomal and fin muscles (Johnston et al., 1997; 2001). For example, the development of dorsal and anal fin ray muscles and their neuromuscular junctions occurred at shorter body lengths at 12 °C resulting in improved fast-start swimming performance relative to a 5 °C group (Johnston et al., 2001).

#### 7.5.5 Concluding thoughts

In the next chapter, the final muscle fibre phenotype is established for adult Atlantic salmon reared on from the same batch of embryos used in the current experiment. Following the eyed stage, fish from different embryonic treatments were provided an equal growth opportunity and it was found that the final fibre number in fish of ~2-6 kg varied by a maximum mean of ~17% between embryonic temperature treatments. Thus, the current study provides the first direct link between temperature heterochronies in MRF expression in teleost embryos and life-persisting alterations in the subsequent muscle fibre phenotype.

Morpholino knockdown experiments of *myod* and *myf5* in the zebrafish resulted in an increase in the number of *pax3/7*-expressing external cells on the lateral surface of the somite (Hammond et al., 2007). These cells are a source of fast muscle growth during post-embryonic zebrafish growth (Hollway et al., 2007; Stellabotte et al., 2007). Thus, heterochronies in MRF expression could provide a potential mechanism to explain some of the persistent changes in muscle phenotype that occur with variations in developmental temperature, including changes in muscle fibre number and nuclear density (chapter 8). The inverse relationship between embryonic titres of MRFs and number of *pax7* expressing cells in the external layer (Hammond et al., 2007)

makes it unlikely that increasing MRF mRNAs in the teleost embryo (e.g. by temperature manipulation), would equal an equivalent increase in the later MPC population from the external cell-layer. However, a temperature-induced advance in the mRNA signal of MRFs is equivalent to an increased state of muscle differentiation in salmonid embryos (Xie et al., 2001), which likely imposes a greater ability to produce swimming thrust from the trunk muscles at hatch. Presuming the external cell layer is a common and important contributing source of MPCs during adult teleost growth stages, it can be suggested that selection has imposed a temperature for each teleost species, where the state of embryonic muscle differentiation (as reflected in the stage-specific expression of MRFs) is optimised between the need to produce sufficient swimming propulsion during late embryonic stages, and the ongoing requirement of MPCs from the external-cell layer for post-embryonic muscle growth. However, this model is probably oversimplified since it presumes firstly that the external cell layer provides all (or a significant proportion) of the MPCs utilised during adult teleost growth, a feature that remains to be investigated (Stellabotte and Devoto, 2007). Secondly this model does not consider temperatures effects on other factors that may interact with *pax7/3* independently of MRFs or any other genes that regulate the external-cell-layer.

```

mrf4b ATGATGGACCTTTTTGAGACCAACACTTATTTCTTCAACGATCTGCGCTATCTCGAGGGA
mrf4a ATGATGGACCTTTTTGAGACCCACACTTATTTCTTCAACGATCTGCGCTATCTCGAGGGA
*****

mrf4b GATCATGGACCATTGCAGCACTTGGACATGGCGGGGTGTCCCCTTTGTACCACGGGAAT
mrf4a GATCATGGACCATTGCAACATTTGGACATGGCGGGGTGTCCCCTTTGTACCACGGGAAT
*****

mrf4b GACAGCCCGTTGTCACTGGGGGGGATCCGTCCGAGACTGGATGTGACAGCAGCGGAGAG
mrf4a GACAGCCCGTTGTCACTGGGGGGGATCCGTCCGAGACTGGATGTGACAGCAGCGGAGAG
*****

mrf4b GAGCATGTCCTCGTACACCCGGGTCTTCAGCCGCACTGCGAGGGCAATGCCTCATCTGG
mrf4a GAGCATGTCCTCGCACCCCTGGTCTTCAGCCGCACTGCGAGGGACAGTGCCTCATCTGG
*****

mrf4b GCTTGTAAGTTTGTAAAAGAAAGTCTGCACCGACCGACAGGCGCAAAGCGGCCACTCTC
mrf4a GCTTGTAAGTTTGTAAAAGAAAGTCTGCACCGACCGACAGGCGCAAAGCGGCCACTCTC
*****

mrf4b AGGGAAGAAGGCGGCTCAAGAAGATCAGTGAAGCATTTCGATGCGTTGAAGAAAAAGGCC
mrf4a AGAGAAAGAAGGCGGCTCAAGAGGATCAATGAAGCATTTCGATGCGTTGAAGAAAAAGACC
**

mrf4b GTGCCAATCCGAACCAGCGGCTGCCCAAAGTGGAGATTTTACGCAGCGCCATAAACTAC
mrf4a GTGCCAATCCGAACCAGCGGCTGCCCAAAGTGGAGATTTTACGCAGCGCCATAAACTAC
*****

mrf4b ATCGAGCAATTGCAGGACCTGTTGCATACACTGGATGAGCAAGAAAAACGCCCCAAAT
mrf4a ATCGAGCAATTGCAGGACCTGTTGCATACACTGGATGAGCAAGAAAAACGCCCCAAAT
*****

mrf4b GGGTCATATAACTATAACGTGAAAGAACCACATGTAAGCTTCATTTGAGAAATATTTGCA
mrf4a GG-----CTATAACGTGAAAGAACCACATGTAAGCTTCATTTGAGTAATTTTTCGA
**

mrf4b TCTGCATTTTTT--GTCACTTTATGCGTACAATAATTGGCTTTACACTTAAAGGGAAAT
mrf4a TCTGCATTTTTTTATGTCACTTTATGCGTACAATAATTGGCTTTATACTTAAAGGGAAAT
*****

mrf4b TGCCATAATGAGAAATATCTTACTGTGCATTATACTTGTAGGCTTACTACTTTGTATAAT
mrf4a TGCCATAATGGGAAATATCTTACTGTGATTATACTTGTAGGCCTACTTCTTTGTATAAT
*****

mrf4b AACACGTGTGTTAGTAGACACTTAAAGCTAATACATGTGTGTGTACGTGAAGGCCTCCAAT
mrf4a AACACGTGTGTTAGTAGACACTTAAAGCTAATACATGTGTGTGTACGTGAAGGCCTCCAAT
*****

mrf4b AAGGGGTACCATTGGAAGAAGAAGTGTCAAACTGGCAGACCTCAGCTGATCATTTCCAAT
mrf4a AAGGAGTACCATTGGAAGAAGAAGTGTCAAACTGGCAGACCTCAGCTGATCATTTCCAAT
****

mrf4b GCACCAATGACGAATCAGAGAGAAGGTTGGTGCCAATTATGAATAGTAAATTAGATCTAA
mrf4a GCACCAATGACGAATCAGAGAGAAGGTTGGTGCCAATTAAGGACAGTAAATTAGATCTAA
*****

mrf4b ATTGCTATTTTGCAACTCATCAACTGATTGTACGTCTTATATATTTGATTTCTATATTTC
mrf4a ATTGCTCTTTTGCAACTCATCAACTGATTGTAGAT-----TTTCTATTTTC
*****

mrf4b AGGCTTCACGGAGTCTTCAGCGTCCACCAGCCTTCTTCGCCTGTC-----
mrf4a AGGCTTCACGAGTCTTCAGCGTCCACCAGCCTTCTTCGCCTGTCATCAATCGTTGACAG
*****

mrf4b -----
mrf4a CATCTCAAGTGAAGAGAAACCGACTTGCACGAAGAAGTCTCAGAAAAATAA

```

Fig. 7.14. Figure legend is on the next page.

Fig. 7.14. Two *mrf4* genes map to two distinct genomic locations in Atlantic salmon. *mrf4a* (DQ479951) has been characterized in the current chapter and is highly conserved (~94% sequence identity) compared to *mrf4b* (EF450078; Moghadam et al., 2007) throughout the whole gene. At the AA level the paralogues also share ~94% sequence identity. Additionally, *mrf4a/b* paralogues share conserved splice sites and differences in intronic sequence identity are comparable to coding regions. The most obvious differences between the paralogues is an insertion of nine base pairs (GTCATATAA) in exon 1 of *mrf4b* and an insertion of thirteen base pairs (CTTATATATTTGA) in intron 2 of *mrf4b*. While the simplest explanation for these paralogues is that they arose during the salmonid genome tetraploidization, it should be noted, if this is the case then salmonid Mrf4 proteins are under considerably higher selective pressure than salmonid MyoD protein paralogues (MyoD1a and 1b share 78/<60% respective sequence identity at the AA/gene level).

## **Chapter 8. Temperature until the ‘eyed stage’ of embryogenesis programs the growth trajectory and muscle phenotype of adult Atlantic salmon**

### **8.1 Abstract**

The aim of this chapter was to investigate how the adult growth trajectory and muscle fibre phenotype of Atlantic salmon (*Salmo salar* L.) responds to temperature during a short window of embryogenesis. Fertilised eggs from several hundred families were divided between four-replicated temperature treatments (2, 5, 8 or 10 °C) until a defined stage corresponding to the complete pigmentation of the eye, after which time embryos were transferred to 8°C until hatching. Treatment groups were subsequently reared in replicated freshwater tanks under identical conditions and with an ambient temperature regime. Following smoltification, salmon were PIT-tagged and randomly assigned to one of three replicate saltwater tanks for on growing for 18 months. Fish exposed to 2 and 5°C until the ‘eyed stage’ of embryogenesis were smaller at hatching and smoltification than groups at higher temperatures, but showed substantial compensatory catch-up growth. Remarkably, altering temperature during this short window of embryogenesis dictated the muscle phenotype three years later with significant treatment effects on the final number, maximum diameter, nuclear density and size-distribution of muscle fibres. The norm of reaction response for final fibre number (FFN) was bell-shaped and FFN was highest at 5°C (8.91 x 10<sup>5</sup> fibres), and reduced by 17% and 14 % at 10 and 2°C respectively. Additionally, myonuclear density was significantly higher for fish of the same cross-sectional area at 5 than 10 °C. These findings require direct temperature effects on embryonic tissues, such as the

external cell layer, which provides MPCs for postembryonic growth, explaining persistent developmental influences on the adult phenotype.

## 8.1 Introduction

The development of teleost embryos is dictated by ambient environmental conditions. Sub-lethal temperature stress during embryogenesis can strongly modify developmental outcomes in the short term and with persistent effects (Johnston, 2006). An established example is muscle fibre phenotype in larval and adult teleosts, which is sensitive to early environmental temperature (reviewed in chapter 1, section 1.8). In teleosts, fast muscle myotubes are produced during embryonic, larval and adult stages of the life cycle. Two principal embryonic fast-twitch MPC populations are recognised. The posterior cells of the undifferentiated somite express MRFs and generate solely embryonic muscle fibres whereas the anterior cells express Pax7 and establish the external cell layer (ECL), which generates new muscle fibres during late-embryonic and larval stages, and MPCs that migrate into the myotome to reside as satellite cells (Hollway et al., 2007; chapter 1, section 1.7.4). The most significant source of fast myotube production during adult life stages is through mosaic hyperplasia, where MPCs scattered through the myotome differentiate next to existing fibres (Koumans and Akster, 1995). This process continues into adult life, after which time the fibre number is fixed (Johnston et al., 2004). Since mosaic hyperplasia accounts for ~95% of the final muscle fibre number in Atlantic salmon (Johnston, 2006), any factor that influences the number or physiology of the supplying MPCs could alter the intensity and/or duration of myotube production. MPCs also provide nuclei for fibre expansion in length and diameter during growth and for nuclear turnover at all stages in the life cycle. It should be pointed out

that it is currently unknown whether the external cell-layer is a significant source of myogenic stem cells for postembryonic growth (Stellabotte and Devoto, 2007).

To date, studies on the effect of embryonic temperature on teleost myogenesis have typically utilised two temperature treatments without knowledge of the associated norm of reaction (Johnston, 2006) and have rarely recorded growth through adult stages. Understanding the normal myogenic response of salmon to embryonic temperature could inform the aquaculture industry about optimising growth and flesh quality outcomes, which are profoundly affected by muscle fibre characteristics (Johnston et al., 2000b). In this chapter, results are presented examining the growth trajectories and final muscle fibre phenotype of Atlantic salmon reared at four embryonic temperatures chosen to stimulate the maximum sub-lethal range of phenotypic responses.

## **8.3 Materials and methods**

### *8.3.1 Embryonic temperature experiment*

Full details on the embryonic temperature experiment can be found in chapter 2 (section 2.2).

### *8.3.2 Fish sampling*

During seawater growth, fish were sampled at five time points from May 2005-Nov 2006 (dates in section 2.2). 72 fish (18 per treatment), from the final sample date (Nov 06) were sampled according to the protocol described in chapter 2 (section 2.2.3). A random subsample of animals was selected (2°C, N=13, 5 °C, N=13, 8 °C N=12, 10 °C N=12), ensuring equal tank representation. Muscle blocks from these fish were equilibrated to –20 °C and 7

$\mu\text{m}$  cryosections were cut on a Leica CM1850 cryostat (Leica Microsystems) stained with Mayer's haematoxylin (Sigma) before 10-20 digital photographic fields were recorded per section using a Zeiss imaging system (Zeiss). The outlines of 1000 muscle fibres per fish (equally represented between blocks) were digitised using SigmaScan (SPSS Inc), recording cross-sectional area and diameter. The total cross-sectional area of the fast-twitch muscle steak (FCSA) was measured by the same approach. The fibre number (FN) was estimated by the following established formula:  $\text{FN} = 1000000 \times (\text{FCSA} \times \text{number of fibres measured} / \text{cumulative area of fast fibres})$ . The maximum fibre diameter (Dmax) was estimated as the mean diameter of the 20 largest fibres. Muscle fibre nuclear density was estimated for 5 versus 10 °C fish (N= 9 and 10 respectively, randomly selected) by randomly selecting three fields per section and counting all the haematoxylin stained muscle fibre nuclei within a randomly drawn rectangular box of 0.4 mm<sup>2</sup>. For each fish, 900-1200 nuclei were counted in total within 12-18 regions representing the whole half-steak.

### *8.3.3 Modelling growth data*

Dr Charles Paxton modelled data relating to treatment differences in somatic growth trajectory. Body mass measurements from the seawater samples were analysed in a mixed model framework using the nlme library (Pinheiro & Bates 2000) in R (R Development Core Team, 2007) with fish mass as a dependent variable. Sample date and embryonic temperature were fixed factorial treatments whereas tank and individuals were treated as random effects with individuals nested within tanks. The data was log transformed to remove data heterogeneity. The number of individuals measured for each treatment at each sampling date can be found in the Fig. 8.1 legend. Variation in body mass between individual temperature treatments at each sampling date was assessed by calculating confidence intervals for each sample date-temperature combination.

#### 8.3.4 Muscle morphometry statistics

Muscle fibre morphometry statistics were performed in Minitab 13 (Minitab Inc). Data for each measured parameter (mass, fork length, FCSA, FN, mean fibre area, mean fibre diameter, Dmax and myonuclei per mm<sup>2</sup>) was continuous, normally distributed (as assessed by Anderson Darling's test) and homogenous (as assessed by Levine's test) so parametric statistics were appropriate. To test differences between measured parameters, a general linear model (GLM) ANOVA was used with sequential sum of squares and considering temperature, tank and tank\*temperature as fixed factors. Fishers least comparison test was used *post hoc* to establish the source of significant variation among treatments. To test the null-hypothesis that the myonuclear density of muscle fibres was not affected by treatment at an equivalent FCSA, a GLM ANOVA was employed using sequential sum of squares and considering FCSA, temperature and FCSA\*temperature as fixed factors.

#### 8.3.5 Calculating fibre probability density functions

To compare the distribution of muscle fibre sizes in fish groups matched for body length (see results, section 8.4.5) a non-parametric method was used to fit smoothed probability density functions (PDFs) to 990 measurements of fibre diameter per fish using a kernel function (Silverman, 1986) within the S-plus computing environment utilising the sm library (Bowman and Azzalini, 1997, smoothing coefficient  $h = 0.18$ ) (method described in Johnston et al., 1999). Bootstrapping was used as a visual tool to assess random variation in diameter distribution from true treatment differences. The non-parametric Kolmogorov-Smirnov two-sample test statistic was used to test the null hypothesis that PDFs of muscle fibre diameter were identical between temperature treatments.

## 8.4. Results

### 8.4.1 Embryonic growth trajectory

Embryos were reared at 2, 5, 8 or 10 °C from fertilisation until the relative age of 165 Ts (Ts defined in chapter 2, section 2.2.2) as defined by Gorodilov (Gorodilov, 1996), which is just subsequent to the period when the eye becomes completely pigmented (the 'eyed stage'). This respectively took 94, 53, 33 and 25 days at 2, 5, 8 and 10 °C. After this stage, embryos were given an equal growth opportunity and reared at 8 °C until hatching, which took 46, 32, 26 and 28 days at 2, 5, 8 and 10 °C. Therefore, the embryonic treatment encompassed different proportions of the total period of embryogenesis at 2, 5, 8 and 10 °C, respectively 67%, 62%, 56%, 47%. Further, while the rate of embryonic development rose with increasing temperature during the embryonic treatment, once transferred to 8 °C, embryos reared at 10 °C required longer to reach hatch than at 8 °C and the relative time for 5 and 2 °C treatments to hatch was reduced as a proportion of the total embryonic period compared to higher temperatures.

### 8.4.2 Post embryonic growth trajectory

In freshwater stages, juveniles from the 10 °C treatment grew faster than counterparts from 8 > 5 > 2°C treatments and the respective mass of fish from each treatment in the Jan 2005 sub-sample was 285, 218, 132 and 76 g (N=24). Since fish were not PIT tagged at this stage, it was not feasible to weigh individuals from the main population. Once fish were transferred to seawater they were PIT tagged and sampled five times over ~18 months, recording the individual weights of all fish in each treatment. The number of fish weighed at each sample date can be found in the Fig. 8.1 legend and decreased as the experiment proceeded due to the

removal of fish during each sub-sample (chapter 2, section 2.2) or due to disease, sexual maturation traits or mortality. An ANOVA, with a repeated measures design was used and indicated that variation in body mass due to the treatment, sample date and a treatment\*sample date interaction was highly significant ( $p < 0.0001$ ) (Table 8.1). Thus, embryonic temperature significantly altered the growth rate of adult salmon (plotted in Fig. 8.1). At the point of seawater transfer, 10°C fish were heavier than 8°C fish (by 10%), which in turn were heavier than 5, and 2 °C fish (respectively, by 32% and 68%). However, during on-growing, 5 and 2 °C fish grew faster than higher temperature treatments (Fig. 8.1, note the convergence of body masses between treatments towards the right of the plot) (Fig. 8.1, a, b). By the final sample, 5, 8 and 10 °C fish were of an equivalent size (varying by a maximum of 9% and between a mean 3732-4083g) and each heavier than 2 °C fish (by 28-33 %), which reached a mean final mass of 2753g (Fig. 8.1, b).

#### 8.4.3 Embryonic temperature and adult muscle fibre morphometrics

For all muscle fibre measurements, variation observed between tanks and from temperature-tank interactions was not significant (see F values, Table 8.2). The mean mass of the sub-sample was comparable to that of each fish population (see Fig. 8.1, b, versus sub-sample: 2°C = 2540 ± 591g, 5°C = 3832 ± 895g, 8°C = 4403 ± 916g, 10°C = 4016 ± 976g; Means ± SD). A highly significant effect of temperature on mass was recorded ( $p < 0.0001$ , Table 8.2). *Post hoc* testing revealed no statistical difference between the mean mass of 5, 8 and 10°C treatments, but each group was significantly heavier than 2°C animals ( $p < 0.001$ ). FCSA was also significantly affected by temperature ( $p < 0.0001$ , Table 8.2). *Post hoc* testing revealed that 5, 8 and 10 °C treatments had a significantly greater FCSA than 2 °C at the  $p < 0.001$  level and 8 °C greater than 5 °C at the  $p < 0.05$  level (Fig. 8.2, a). However, these differences in treatments can be accounted for by differences in body mass, as a highly significant

interaction ( $p < 0.0001$ ) was observed between mass and FCSA that was constant across treatments. Additionally, fast muscle fibre size attributes were affected by treatment with mean fibre diameter/area varying significantly at the  $p \leq 0.0001$  level (Table 8.2). Further significant treatment induced differences were observed in the mean final FN (Table 8.2:  $p = 0.017$ , Fig. 8.3) with the 5 °C treatment being optimal ( $8.91 \times 10^5 \pm 1.4 \times 10^5$  fibres) and 7% greater than 8 °C ( $8.29 \times 10^5 \pm 1.1 \times 10^5$  fibres), 14% greater than 2 °C ( $7.70 \times 10^5 \pm 1.5 \times 10^5$  fibres) and 17% greater than 10 °C ( $7.38 \times 10^5 \pm 1.0 \times 10^5$  fibres) (Fig. 8.3, a). *Post hoc* testing revealed that 5 °C fish had a significantly greater final FN than those reared at 2 °C and 10 °C ( $p < 0.001$ ), but that other group differences were not significant (Fig. 8.3, a). Furthermore, Dmax was significantly greater ( $p < 0.0001$ ) at 8 and 10 °C than 5 or 2 °C, although differences between 8 and 10 °C or 5 and 2 °C treatments were not significant (Fig. 8.3, b).

#### 8.4.4 Embryonic temperature and muscle fibre nuclear density

The nuclear density of muscle fibres from 5 and 10 °C fish ( $N = 9$  and  $10$  respectively, randomly selected) was then compared to reflect the two treatments that showed the largest difference in final FN. No significant difference was observed between treatments in mean nuclear density (not shown). However, a highly significant interaction between FCSA and nuclear density was observed (Table 8.2), where nuclear density decreased with FCSA. When mean nuclear density was plotted against FCSA, 5 °C fish had a significantly higher nuclear density at an equivalent steak cross-sectional area relative to the 10 °C treatment (Fig. 8.4,  $p = 0.036$ ). It should also be noted that removing the outlying point on the plot (Fig. 8.4, 10 °C group) did not alter the significant difference recorded between treatments ( $p = 0.042$ , not shown).

#### 8.4.5 Temperature-induced differences in muscle fibre size distribution

Fig. 8.5, a shows the distribution of 990 muscle fibre diameters per fish for the 5 and 10 °C treatment groups (N=11 per treatment). As fibre diameter is inherently dependent on body length, the smallest fish in each of these groups was excluded to reduce the variance in fibre diameter (two excluded from 5 °C, one from 10 °C), bringing the mean fork lengths under comparison to within 4 cm. The probability density functions (PDFs) of fibre diameters were significantly different between 5 and 10 °C fish using a Kolmogorov-Smirnov test ( $p < 0.01$ ). The difference between treatments is most noticeable at the left-hand side of the plot where 5 °C fish have a higher PDF for diameters in the smaller range (up to ~75 µm). Additionally, the PDF of fibre diameters from 100µm upwards is greater (line shifted to the right) in 10 °C fish. These differences in fibre size distribution are visualised in Fig. 8.5, b, where fibres were colour coded by diameter. More muscle fibres from the smaller diameter range (up to 75 µm) were observed throughout the myotomal cross-section in the 5 °C treatment relative to 10 °C (Fig. 8.5, b).

Table 8.1. Summary of mixed-model ANOVA parameters used to distinguish variation in post-smolt body mass.

<b>Variable</b>	<b>df</b>	<b>F</b>	<b>P</b>
(Intercept)	1	973110.8	<0.0001
Sample date	4	40256.2	<0.0001
Temperature	3	536.6	<0.0001
Sample date*Temperature	12	358.3	<0.0001

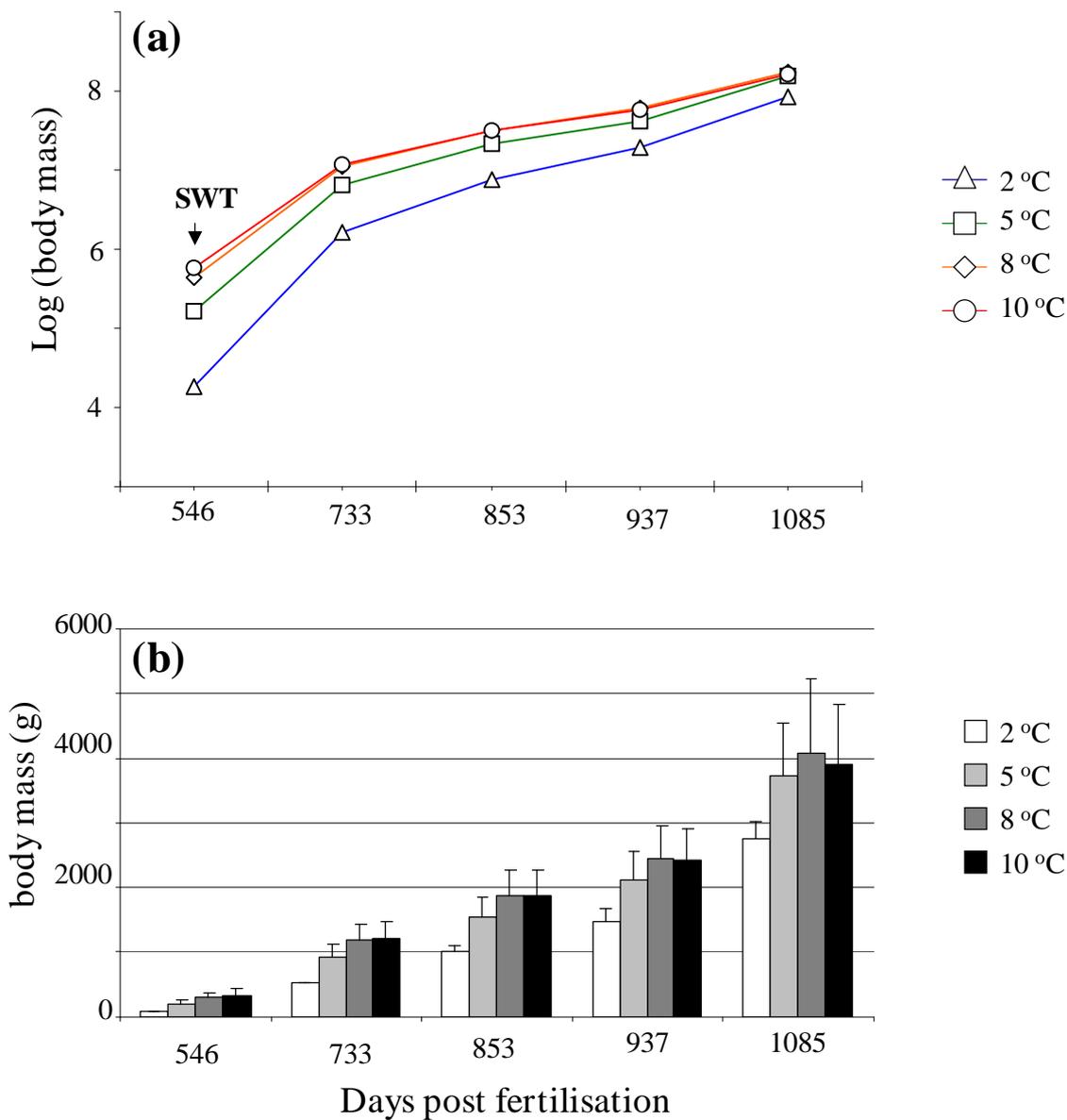


Fig. 8.1. Changing embryonic temperature solely to the ‘eyed stage’ of embryogenesis produced marked and significant (Table 8.1) effects on the post-smoltification growth trajectory of Atlantic salmon. Fish were individually weighed in May 2005, (546 days post fertilisation, N= 306, 294, 299 and 300 for 2, 5, 8 and 10 °C respectively), Nov 2005, (733 days post fertilisation, N= 280, 247, 224 and 253 for 2, 5, 8 and 10 °C respectively), Mar 2006 (853 days post fertilisation, N= 127, 125, 127 and 125 for 2, 5, 8 and 10 °C respectively), June 2006 (937 days post fertilisation, N= 74, 78, 79 and 85 for 2, 5, 8 and 10 °C respectively) and Nov 2006 (1085 days post fertilisation, N= 56, 54, 52 and 57 for 2, 5, 8 and 10 °C respectively). (a) Shows the mean of log-transformed data of individual weights, modelled in R, using a mixed model ANOVA. SWT shows the point of seawater transfer. (b). Shows the actual mean body mass of each temperature group at the 5 sampling points + SD. In both plots, substantial compensatory growth can be observed in the 5 and 2 °C treatments.

Table 8.2. Summary of general linear model ANOVA parameters used to distinguish variation in muscle fibre characteristics.

<b>Variable</b>	<b>df</b>	<b>Seq SS</b>	<b>Seq MS</b>	<b>F</b>	<b>P</b>
<b><i>Mass</i></b>					
Temperature	3	26020217	8673406	11.78	<0.0001
Tank	2	433439	216719	0.29	0.747
Temperature-Tank	6	4629072	771512	1.05	0.410
<b><i>Fork Length</i></b>					
Temperature	3	94511	31504	13.68	<0.0001
Tank	2	738	280	0.12	0.886
Temperature-Tank	6	9726	1621	0.70	0.648
<b><i>Fast muscle steak area</i></b>					
Temperature	3	72520817	24173606	10.63	<0.0001
Tank	2	325867	162934	0.07	0.931
Temperature-Tank	6	17354994	2892499	1.27	0.293
<b><i>Mean fibre area</i></b>					
Temperature	3	102209598	34069866	9.30	<0.0001
Tank	2	1929389	964694	0.26	0.770
Temperature-Tank	6	6610203	2768367	0.76	0.609
<b><i>Mean fibre diameter</i></b>					
Temperature	3	4050.6	1350.2	12.02	<0.0001
Tank	2	114.4	57.2	0.51	0.605
Temperature-Tank	6	459.4	76.6	0.68	0.665
<b><i>Mean final fibre number</i></b>					
Temperature	3	3 1.5842E+11	5.2806E+10	3.87	0.017
Tank	2	2 3.8770E+10	1.9385E+10	1.42	0.254
Temperature-Tank	6	1.1720E+11	1.9533E+10	1.43	0.229
<b><i>Max fibre diameter</i></b>					
Temperature	3	8299.9	2766.6	6.84	0.001
Tank	2	1071.4	535.7	1.32	0.278
Temperature-Tank	6	1508.6	251.4	0.62	0.712
<b><i>Nuclear density</i></b>					
Steak area	1	10578.6	10578.6	28.51	<0.0001
Temperature	1	1976.5	1976.5	5.33	0.036
Steak area-Temperature	1	83.2	83.2	0.22	0.643

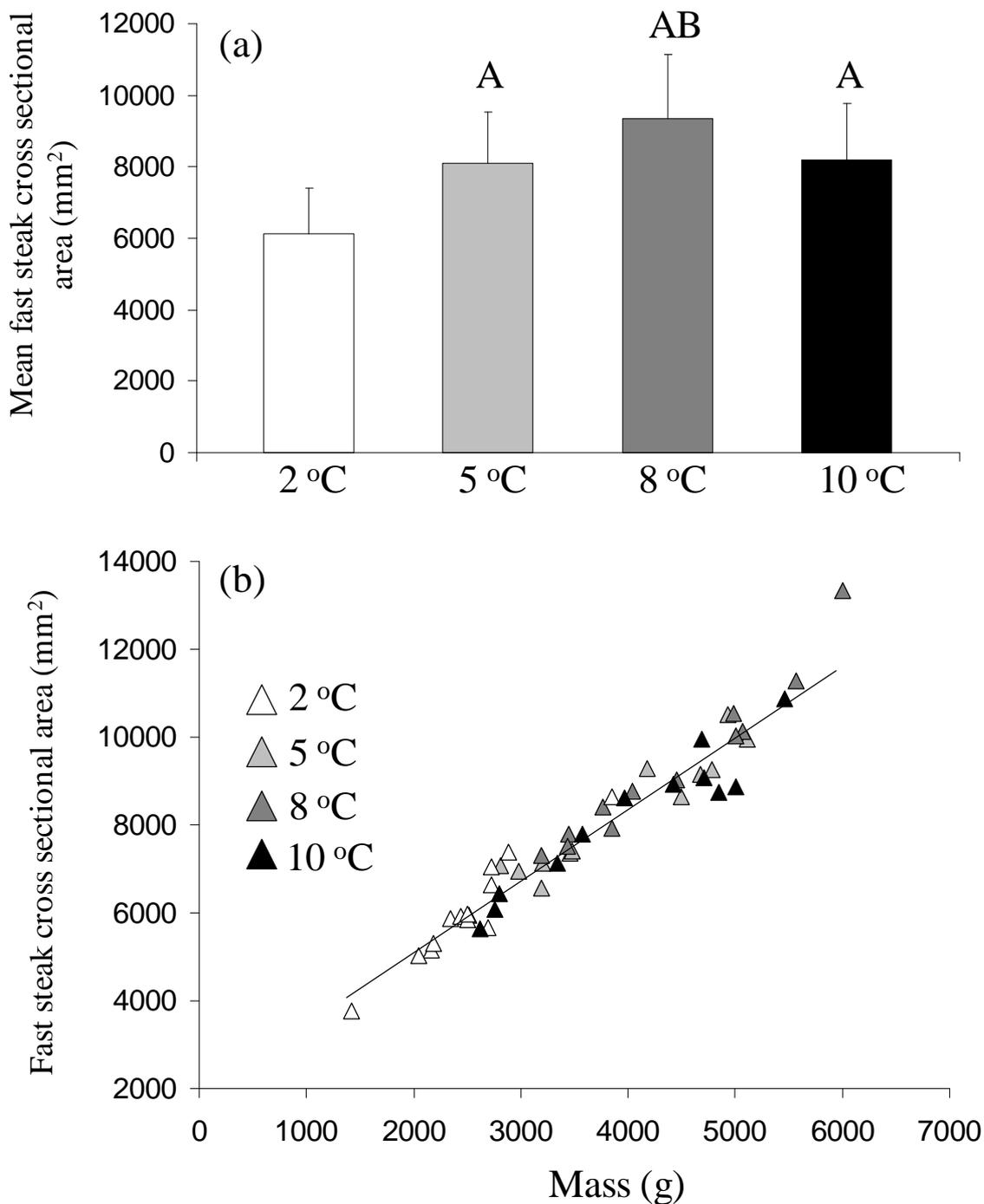


Fig. 8.2. The effect of embryonic temperature on the FCSA of adult Atlantic salmon. (a) Shows the mean FCSA + SD. A and B respectively indicate a significant difference to 2 and 5 °C (see section, 8.4.2 for details). (b) Shows a scatterplot of individual measurements of FCSA versus body mass. It is clear that body mass can account for differences observed between FCSA from different temperature treatments. A first order regression was fitted to the data and the following equation was obtained  $FCSA = 1715.1 + 1.679 * mass, R^2 = 93.7\%, p < 0.0001$ .

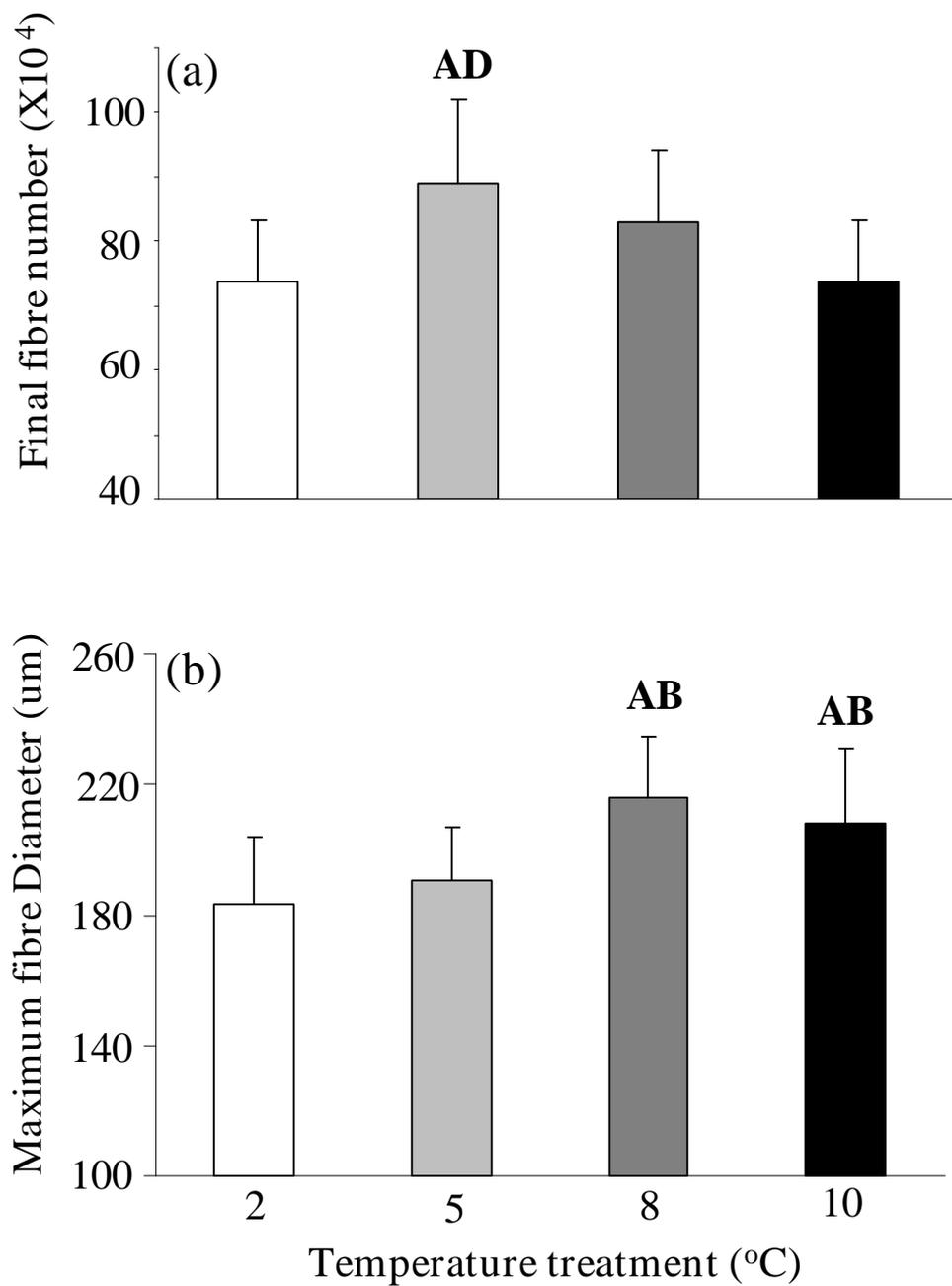


Fig. 8.3. The effect of embryonic temperature on the final muscle fibre phenotype of adult Atlantic salmon. All values are means + SD (a) shows the change in mean final muscle fibre number; A and D respectively indicate a significant difference ( $p < 0.001$ ) compared to 2°C and 10°C (b) shows the change in Dmax; A and B respectively indicate a significant difference ( $p < 0.001$ ) compared to 2°C and 5°C.

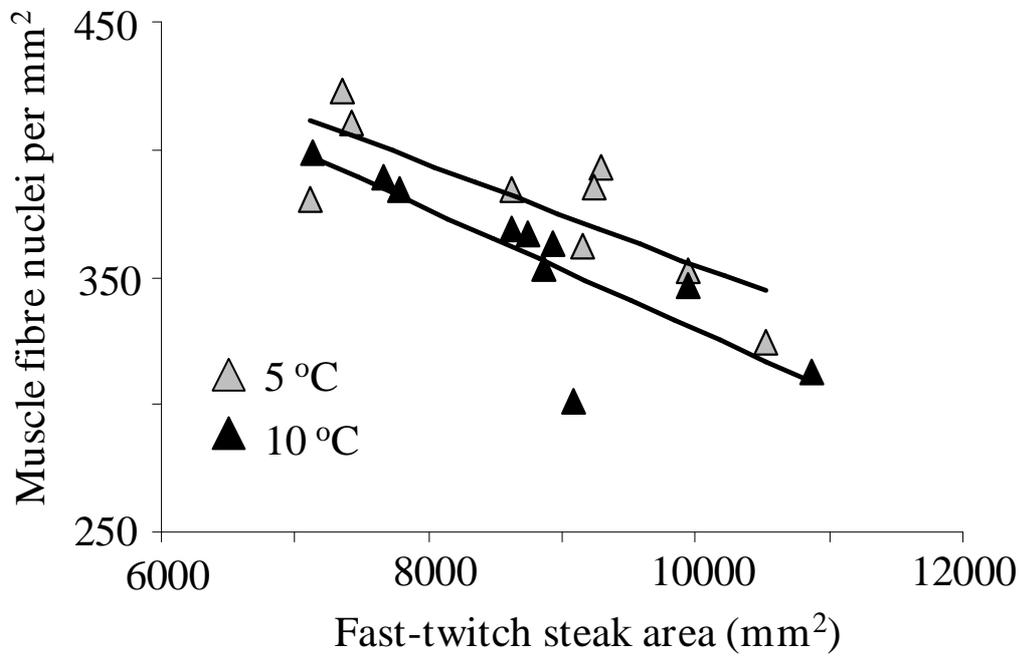


Fig. 8.4. Scatterplot of myonuclear density versus FCSA at 5 and 10°C. The number of myonuclei per mm<sup>2</sup> at an equivalent was significantly different between 5 and 10 °C treatments (Table 8.2). Importantly, removing the low outlying point at 10 °C (around 9000 mm<sup>2</sup> on the x axis) did not alter this result.

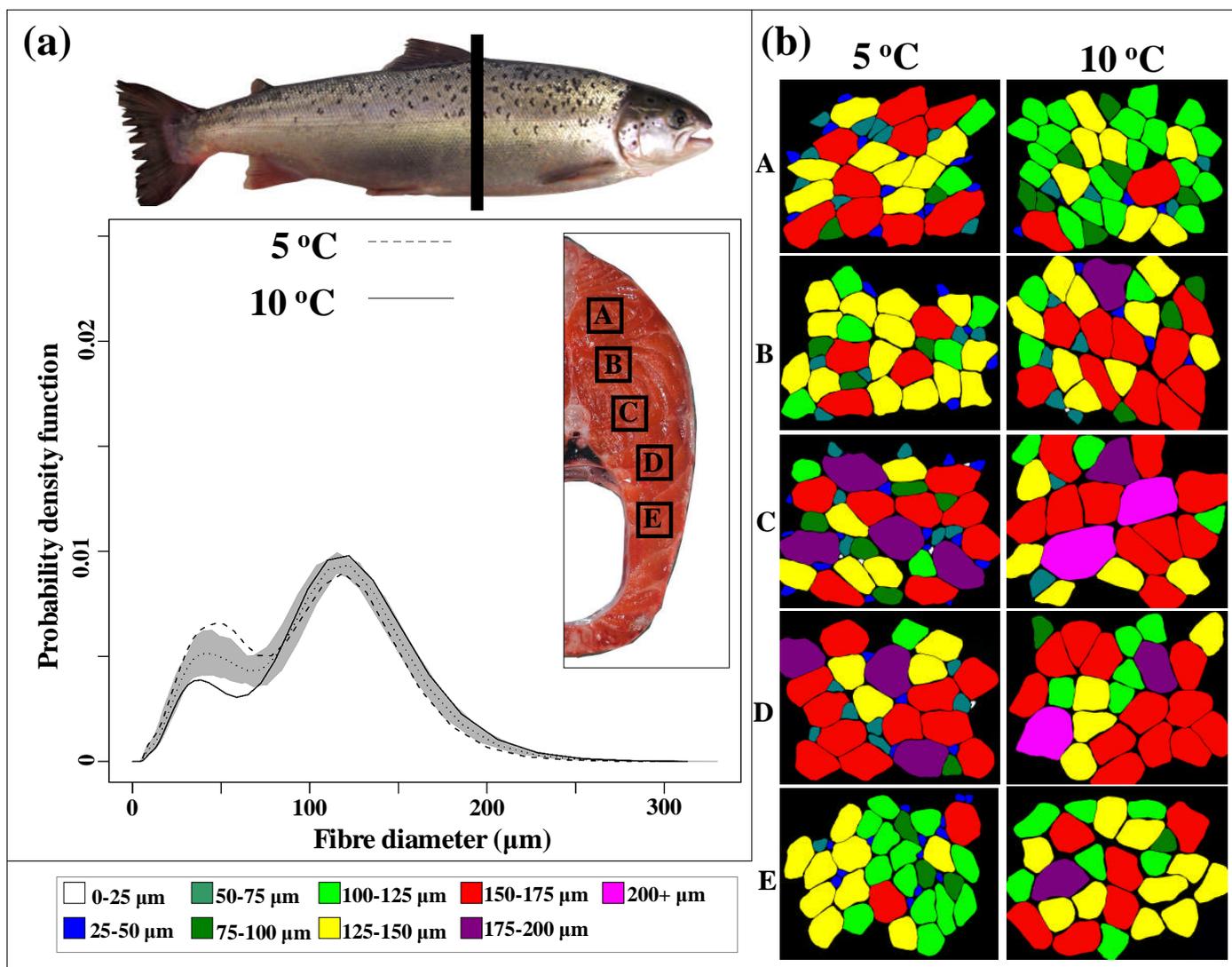


Fig. 8.5. The image of an adult salmon shows the region where the muscle steak was cut. From half of this steak (pictured), ~200 muscle fibres were digitised in 5-6 blocks, recording diameter and area. The plot shows the distribution of muscle fibre size in length-matched 5 and 10 °C fish (N=11 per treatment). The dashed and solid lines respectively show the average probability density function (PDFs) for 5 and 10 °C treatments. The dotted line central to the shaded area is the average PDF for combined 5 and 10 °C groups. The shaded area shows 100 bootstrap estimates from combined populations of fibre diameter. (b) Three randomly selected photographed fields (10X magnification) from each block were then sampled for the three largest (approximately size matched) fish in the 5 and 10 °C treatments. Each fibre was colour coded by diameter (shown to figures left). A similar pattern was obtained for each fish and the images shown are representative.

## 8.5 Discussion

### 8.5.1 Changing temperature solely to the 'eyed stage' programs adult salmon growth trajectory and muscle fibre phenotype

The main finding of the experiment was that the temperature until the 'eyed stage' of embryogenesis produced effects on somatic growth and muscle fibre phenotype that persisted into adult life. Embryonic developmental rate increased with increasing temperature as previously reported for Atlantic salmon incubated across a sub-lethal range (Gorodilov, 1996). At the point of seawater transfer the 10 °C fish were heavier than 8 °C > 5 °C > 2 °C fish (Fig. 8.1, a, b). However, during on-growing, fish from 2 and 5 °C treatments showed strong compensatory growth. Consequently, at the end of the experiment, 5 °C fish were of an equivalent size to fish from higher temperature treatments (Fig. 8.1, a, b). However, 2 °C fish never reached an equivalent body size to fish reared at higher temperature treatments. These results require that growth rate during adult life-stages was imprinted during embryonic growth. Temperature induced differences in the mRNA expression of *IGF-II* as well as *IGF-I* receptors and growth hormone receptor-1 (GHR1) were recorded by qPCR in rainbow trout embryos (Gabillard et al., 2003, 2006), as observed for salmon MRFs (chapter 7). While such findings have been suggested to regulate differences in embryonic growth rate (Gabillard et al., 2003), their relevance to adult growth trajectory remains uninvestigated.

In fish of all treatments the lack of muscle fibres in the 5-10 µm diameter range suggested that the final FN had been reached. Thus, the rate of embryonic development defined the duration of adult myotube production. The norm of reaction for this response was bell-shaped, peaking at 5 °C and decreasing towards the temperature extremes, to a maximum of 17%. Such large differences in fibre number are likely to impact flesh texture traits (chapter 1, section 1.9) and

could have accounted for the faster growth rates of fish at 5 °C, since differences in final fibre number have previously been correlated to differences in somatic growth rate (Johnston et al., 2003a). However, it should be noted that 2 °C fish showed substantial catch up growth with a similar final muscle fibre number to the 10 °C group.

Another interesting finding of this study was that fish reared at 2 and 5 °C had a significantly lower mean maximum fibre diameter (Dmax) than 8 or 10 °C treatments. For 2 °C fish, this can probably be accounted for by differences in body size, but 5 °C fish were of an equivalent size to higher temperatures and their mean Dmax was reduced by 8-9%. Dmax in adults is probably limited by diffusional constraints of oxygen across the fibre membrane limiting the build up of an anoxic core (Johnston et al., 2004). It is interesting then that this trait is sensitive to embryonic conditions, and may indicate persistent temperature-induced alterations in metabolic rate.

The current experiment also reinforces the importance of establishing a reaction norm when studying developmental plasticity of phenotype. If this study had been designed to investigate the influence of two temperatures only, for example the 2 and 10 °C groups, an incorrect conclusion could have been made, for example that temperature modified growth rate, with little affect on the final muscle phenotype.

#### *8.5.2 Implications of temperature induced differences in muscle fibre number*

Johnston and co-workers, incubated Atlantic salmon at either ambient, or heated (by 1-3 °C) temperature throughout embryonic and freshwater growth stages before fish were PIT tagged and then grown in seawater with equal growth opportunity (Johnston et al., 2003a). Similar to this study, the heated group were heavier at seawater transfer, but the ambient fish showed

substantial compensatory growth and after 15 months on-growing, differences in body mass were not significant (Johnston et al., 2003a). Additionally, the final fibre number was higher in the ambient group (Johnston et al., 2003a). While this experiment is the closest in design to the current study, differences in experimental set-up and genetic background of the fish make the comparison of the phenotypic responses observed difficult. For example differences in Atlantic salmon final FN are evident in different strains and families (Johnston et al. 2000c). It cannot currently be determined whether the effect of temperature on adult muscle phenotype is imposed solely during embryonic stages or is supplemented during freshwater stages. Furthermore, the influence of embryonic temperature on the number and size of fibres at hatching in Atlantic salmon was shown to vary between fish spawning in upland and lowland tributaries of the same river system, suggesting local adaptation (Johnston et al., 2000a).

Our experiment shows that changing temperature during a restricted window of salmon embryogenesis can alter growth and fibre characteristics throughout the whole adult life cycle independently of the final body size. From an industrial perspective, these results suggest that the careful manipulation of embryonic temperature is a promising area for optimising desirable growth and muscle fibre traits of adult fish. However, considering the genetic influence on FN and growth rate, it is likely that the reaction norms to embryonic temperature would have to be established for multiple strains. In this regard, a more fruitful approach might be to identify strains with desirable phenotypic traits (high growth rates and high fibre number) and then further modify these characteristics with environmental manipulation.

### *8.5.3 Embryonic temperature induced alterations in myonuclear density*

It has previously been reported for Atlantic salmon that myonuclear content was altered by manipulating temperature from fertilisation to hatching (Johnston et al., 2003a; Johnston and

McLay, 1997; Nathanailides et al., 1995). In this experiment the myonuclear content of muscle fibres was greater at 5 than 10 °C for fish of the same myotomal cross-sectional area (Fig. 8.4). Myonuclear density per unit area rises as fibre density increases (Johnston et al., 2003a). Thus the higher FN in 5 °C fish could have contributed to increased myonuclear density. However, treatment induced alterations of myonuclear density of >20% were recorded in single muscle fibres, which removes this confounding effect (Johnston et al., 2003a). This suggests that alterations in myonuclear content from the temperature treatment were greater than differences due to changes in fibre density. Since final FN and myonuclear density were affected by treatment in the embryo, the underlying mechanism likely involves an effect on a common pool of MPCs, even if they subsequently had different fates.

#### *8.5.4 How does embryonic temperature program adult myogenic phenotype?*

The results presented here require an embryonic mechanism to explain alterations in final muscle fibre phenotype including the final FN. The final FN was also significantly altered in adult zebrafish incubated solely during embryogenesis at a range of temperatures covering normal development (Lee, 2008). Further it has been demonstrated for numerous phylogenetically diverse teleosts, that muscle fibre recruitment and hypertrophy are sensitive to embryonic temperature (chapter 1, section 1.8). This suggests that the underlying embryonic mechanism may be conserved across teleosts. In chapter 7, temperature-induced heterochronies were observed in the expression of two important muscle-specific transcription factors (*myf5* and *mrf4*) in embryos sub-sampled as part of the current experiment. Furthermore, a feasible hypothesis was presented to explain how embryonic temperature imposes later changes in the muscle fibre phenotype (described in chapter 7, section 7.5.5), based on recent findings on the external cell layer (ECL) (Hollway et al., 2007; Stellabotte et al., 2007), a common feature of teleosts (Devoto et al., 2006). Under this theory, the results presented in this chapter can be

explained if the modifications observed in embryonic MRF expression (chapter 7), altered the subsequent titres of Pax7 changing the number or proliferation of ECL MPCs that contributed to adult muscle growth. In any future test of this hypothesis, the contribution of the ECL to adult muscle hyperplasia and hypertrophy would need to be established. If the ECL population is not temperature-dependent, or its final contribution to the adult muscle fibre phenotype is relatively small, then some other embryonic mechanism is required to account for the results observed here.

## Chapter 9. General Discussion

### 9.1 A molecular tool-box for studying Atlantic salmon myogenesis

During the first part of this project, multiple myogenic genes were amplified by RT-PCR either as complete or partial cDNAs and then cloned (Table 9.1). Many were previously uncharacterised in Atlantic salmon and included markers of myoblast specification (*myod1c*, *myf5*), differentiation (*myog*, *mrf4*) or fusion (calpastatin), of satellite cells (*sox8*, *foxk1*, *pax7*), inhibitors of Mstn (*fst*, decorin, *fstl1*, *GASP-1*) and a novel gene with a putative role in regulating growth (*cee*). For several of these genes, assays were developed to characterise their mRNA expression patterns, either by synthesising complementary RNA probes for *in situ* hybridization, or designing primers for qPCR or RT-PCR (Table 9.1). Thus, a molecular toolbox was developed for the study of Atlantic salmon myogenesis, and all plasmid-held cDNAs, cRNA probes and qPCR/RT-PCR primers have been catalogued and appropriately stored for future researchers.

### 9.2 Characterising teleost and muscle salmon genes: patterns and paralogues

In chapters 3, 4, 5, 6 and 7, several of the genes described above were characterised using a combination of expression, genomic and evolutionary analyses. In chapter 3, it was shown that *myod* duplicated during the teleost WGD event, but that the resulting paralogues were retained in the Acanthopterygii, whereas other teleosts lost one copy (*myod2*). The protein coded for by *myod2* has evolved rapidly relative to *myod1* and its role in myogenesis is almost entirely uncharacterised. An interesting future study would be to knockdown *myod2* using morpholino antisense RNA in an Acanthopterygian species to see whether it functions redundantly with *myod1*, performs a unique role, or is disposable for myogenesis and on the path to becoming a pseudogene.

In chapter 4, a novel *myod* paralogue was characterised that is specific to the Salmonidae, a group that went through a genome tetraploidization event after the teleost WGD. This gene (*myod1c*) arose from two lineage-specific duplications of *myod1*, along with its paralogues, *myod1a* and *myod1b*. The three salmon *myod1* co-orthologues were likely retained by subfunctionalization and together recapitulated the expression pattern of *myod1* in non-salmonid teleosts. Additionally, the three genes were differentially expressed in MPC (MPC) populations of distinct phenotype. In the future, the promoter regions of *myod1a*, *1b* and *1c* could be used as a study model to delineate regulatory elements governing the expression of the single teleost *myod1* gene in different MPC populations, which is an active area of research.

In a later chapter (7), the concurrent expression of six Atlantic salmon MRFs was investigated, providing valuable insight into the regulatory networks governing myogenesis in a tetraploid teleost. While the expression of these genes was generally analogous to non-salmonid teleosts, certain previously characterised expression domains for MRFs were absent in Atlantic salmon. For example, *myf5* mRNA was lacking in pre-differentiated adaxial cells and *mrf4* was expressed markedly later with respect to its zebrafish orthologue. Considering the importance of these expression domains to teleost myogenesis (Coutelle et al., 2001; Hinitz et al., 2007) it is improbable that they have been completely lost in salmon. Instead, I suggested that these findings could be explained under the subfunctionalization hypothesis (Force et al., 1999), where uncharacterised paralogues for *mrf4* (*mrf4b*, see chapter 7, Fig. 7.14, other paralogues potentially possible) and *myf5* (paralogue(s) not yet identified) fulfil the missing expression fields (see chapter 7, section 7.5.1). However, this hypothesis requires further experimental validation, and more salmonid MRF paralogues may be discovered in the future. Another point to note is that apart from *myod2* in the Acanthopterygii, no other paralogues for *myog*, *myf5* or *mrf4* could be identified in the available genomes of non-salmonid teleosts, including zebrafish, pufferfishes, stickleback and medaka.

In chapter 5, it was shown that *fst*, a gene of crucial importance to multiple biological processes including myogenesis (chapter 1, 1.6.8), was duplicated during the teleost WGD but was specifically retained as two copies in the Ostariophysi lineage. Thus, paradoxically, the zebrafish is not an ideal model to study this gene, since the duplicated copies (*fst1* and *fst2*) may have diverged in function and are differentially regulated (Chapter 5, sections 5.5.2, 5.5.3). In Atlantic salmon embryos, *fst1* was expressed in MPCs of the epithelial anterior somite, external cell-layer and fin buds. While these expression domains are analogous to the expression of *fst* in the amniote dermomyotome (Amthor et al., 1996), *mstn* mRNA is not strongly expressed in teleost embryos (Kerr et al., 2005; Xu et al., 2003) as is the case for amniotes (Amthor et al., 1996; Lee, 2004). These results suggest an early role for *Fst1* in teleosts, independent of *Mstn*. The further study of *fst1* in teleost fish could help characterise the *mstn*-independent pathways through which vertebrate *fst* functions in myogenesis (e.g. Lee, 2007).

In contrast to previous chapters and for many other teleost genes cloned during the study (Table 9.1), *cee*, a novel and highly conserved gene, was shown in chapter 6 to be retained as a single copy in all teleost genomes examined as well as in virtually every other metazoan. Future work with this gene is necessary to identify its role in development, including myogenesis, but considering its strong evolutionary conservation, lack of redundancy with other genes and the retarded phenotype of yeast and nematodes lacking *cee*, it is probably an important growth regulator.

### 9.3 What role did genome duplication play in the teleost success story?

The findings in this study support the notion that teleost genes are retained as more copies than in tetrapods (Meyer and Van de Peer, 2005). Vertebrates are phenotypically more complex than lower chordates and invertebrates, as a potential consequence of the evolutionary exploitation of

genomic material gained through WGD events (reviewed in chapter 1, section 1.5). So is the success story of the teleosts in terms of their species number (half of all vertebrates are teleosts) and remarkable habitat exploitation, a consequence of the teleost WGD? It was estimated that 75-85% of paralogues from the WGD are retained as a single copy (Jaillon et al., 2004; Brunet et al., 2006). Once nonfunctionalized, a gene can play no direct part in the evolution of phenotype. Paradoxically, nonfunctionalized gene duplicates still have a hand to play in evolution, through a process termed divergent resolution, which may be an even more important mechanism underlying speciation than gene retention (Taylor et al., 2001; Postlethwait et al., 2004). Divergent resolution occurs when paralogues are differentially lost in different lineages, as shown for *myod* and *fst* (chapter 3, 5). This process, as well as the differential partitioning of gene subfunctions (chapter 1 section 1.5.3) can occur between two allopatric populations of the same species and upon their chance re-acquaintance, creates reproductive barriers where speciation is more favourable to selection than hybridization (model described in Lynch and Force, 2000; Taylor et al., 2001; Postlethwait et al., 2004; Volff, 2005). Interestingly, the consequences of these processes are evident in the Salmonidae (Taylor et al., 2001). This tetraploid family has around seventy species, whereas its non-duplicated sister family, the Osmeridae has only ten (Taylor et al., 2001).

Thus, theoretically, the teleost WGD could have promoted speciation though gene loss as well as by lineage-specific neofunctionalization or subfunctionlization. It has been suggested that the timing of the teleost WGD correlates well with the teleost radiation (Hoegg et al., 2004). However, when extinct taxa are accounted for this pattern is not supported, as hidden depths of species diversity for non-duplicated Actinopterygians are present in the fossil record (Donoghue and Purnell, 2005). Additionally, the majority of species diversity is present in two taxa, the Acanthomorpha and Ostariophysii, with respectively ~16,000 and 6500 recognised species (Nelson, 2006). These groups are thought to have respectively arisen 195-295 and 75-105 Mya

after the WGD (calculated from an estimate of the timing of WGD [Van de Peer, 2004] and from estimated timings of the origin of the Acanthomorpha [Maissey, 1996] and Ostariophysi [Inoue et al., 2005]). Thus, it is difficult to reconcile the differences in timing of the WGD and the teleost radiation. So how was the additional genomic complexity established by WGD exploited during evolution? Crow et al. observed that several vertebrate WGD events are preceded by a multitude of extinct lineages: e.g. ~93% families of non-teleost Actinopterygians are extinct, whereas the same can be said for only 21% of teleost families (Crow et al., 2006). The authors modelled the likelihood of extinction following mass extinction events, and showed that the post-WGD teleost lineage was more than five times less likely to become extinct than non-duplicated ancestors (Crow et al., 2006). Thus, the benefits of WGD may buffer organisms from extinction, meaning the current species diversity of teleosts does not need to be explained by an immediate and explosive post-WGD radiation. Instead, the benefit of genetic redundancy and potential for neofunctionalization was realised under particular extrinsic ecological pressures, which were unique for individually evolving lineages (Crow et al. 2006). For example, it was suggested that the neofunctionalization of a vitellogenin paralogue retained in the Acanthomorpha, but lacking in other teleosts, allowed this group to spawn their eggs in seawater and subsequently radiate at a time when oceanic predation and competition was low (Finn and Kristoffersen 2007). By such means, a whole battery of gene duplicates could have been retained that have contributed to the unique evolution of different teleost lineages.

#### *9.4 Taking the embryonic temperature findings onwards*

The results presented during the temperature experiment have provided insight into the potential physiological mechanisms that underlie embryonic temperature-induced plasticity in adult muscle phenotype. However, a further experiment could be used to identify the exact developmental windows that are sensitive to temperature and their individual contributions to the final phenotype.

I suggest an experiment using a single strain of salmon, with family differences randomised. Eggs could be split into three temperature groups during embryonic stages to cover the normal range of development. Eggs could then be removed from these temperature treatments at multiple developmental stages and maintained at a common temperature for the rest of the trial. The most insightful stages to transfer embryos would be 1. during early blastulation, which is prior to somitogenesis and to maternal mRNA degradation (Pelegri, 2003), 2. at the 20 ss, which is during somitogenesis but prior to somite rotation and adaxial cell migration, 3. at the end of segmentation (65 ss), which is during somitogenesis and subsequent to somite rotation/external cell formation/adaxial cell migration and 4. at a later embryonic stage, post-somitogenesis. Remaining embryos could then be maintained by treatment throughout freshwater growth. At the point of smoltification, all fish could be PIT tagged and randomly transferred to seawater tanks. Subsequently, regular individual measurements of body mass could be obtained and fish from each treatment sampled to analyse the muscle fibre phenotype. While this experiment would be labour intensive, the benefits gained would be substantial. The data returned could statistically confirm the windows of time in embryonic and juvenile stages that are most influential for the adult muscle phenotype. It is possible that the majority of adult developmental plasticity is programmed during short developmental windows, such as somite rotation, or the mid-blastula transition, when zygotic gene expression is activated (Pelegri, 2003). If this were the case then it might be possible to manipulate temperature during critical embryonic events, to optimise final muscle fibre number, with no penalty, or even an improvement in growth rates. It should also be noted, that this experiment could be performed with relative ease in the model zebrafish with significant savings in time and costs, although its direct benefit to the aquaculture industry could only be realised by repeating the study in individual species.

### 9.5 Effect of embryonic temperature on other systems

An animal's phenotype is a consequence of the sum of its physiology functioning in, and interacting with, a changing environment. Extrinsic variables including food, temperature, pH, oxygen, salinity, photoperiod, disease or infection as well as other chemical and behavioural cues interact with multiple organs and through uncountable, overlapping physiological pathways involving the neuro-endocrine, circulatory, digestive, osmoregulatory, and immune systems (see Johnston, 2006). The interactions of these systems can modify behavioural outcomes, for example, by stimulating migration to a new environment, or alternatively can cause plasticity in developing tissues (Johnston, 2006). Considering the complexity of physiological responses to environmental change it can be suggested that our embryonic temperature treatment, in addition to modifying the adult muscle phenotype would also have affected many other developing tissues. For example, juvenile Atlantic salmon incubated at 5 °C during embryogenesis showed more active feeding behaviour than those reared at 10 °C (Albokhadaim et al., 2007). Heterochronic or persistent effects on the developing neuro-endocrine system could explain such findings. For example, temperature induced differences in the mRNA expression of important growth factors including growth hormone receptor-1, *IGF-II* and IGF1 receptors were observed in rainbow trout embryos (Gabillard et al., 2003, 2006) as recorded for MRFs (chapter 7). It is feasible that the embryonic titres of certain hormones or growth factors during critical developmental windows, could 'program' molecular pathways affecting adult traits with the potential to modulate growth trajectory including behaviour (e.g. foraging activity, appetite, competitiveness, aggression level) and physiology (e.g. thermoregulation and energy expenditure, digestive efficiency, ratio of protein production versus atrophy). However, further work is necessary to understand the lasting effect of embryonic temperature on 'whole animal' performance, to assist any hypothesis on the underlying molecular regulation.

### 9.6 Thoughts for the future

From the current study and the work of others it is clear that the effect of embryonic temperature on the mRNA expression of developmental genes is not an ‘all or nothing’ response, with the whole genome responding with a retardation or advance in signal. Instead certain genes respond to differences in thermal regime whereas others are unaffected. For example, temperature induced differences in the mRNA expression of certain salmonid MRFs (*myf5*, *mrf4*, current study) growth hormone receptors (*GHR1*, Gabillard et al., 2006), IGF receptors (IGF receptor-1 genes, Gabillard et al., 2003) and IGF genes (*IGF-II*, Gabillard et al., 2003) were recorded, whereas other closely related genes, or genes from the same system were unchanged (current study; Gabillard et al., 2003, 2006). There is a dearth of information on the effect of embryonic temperature on the expression of genes in a genome wide context, even in model animals. A goal for future researchers will be to identify not just single responding genes, but entire temperature dependent pathways. Furthermore, it is necessary to identify the underlying regulatory mechanisms by which certain genes are affected by changing embryonic temperature. The subsequent challenge will be to disentangle those responses that underlie short-term adaptation and those that produce persistent effects on later phenotype.

To answer such questions, future studies should be increasingly directed towards transcriptome wide approaches. Oligonucleotide and cDNA microarrays have been used to successfully identify differentially expressed genes in numerous experimental contexts including for adult teleosts exposed to thermal stress or temperature fluctuations (Podrabsky and Somero, 2004; Gracey et al., 2004). However, this method has certain problems associated with bias in identifying differentially expressed genes. Spurious hybridization of abundant RNAs to probes for unexpressed or lowly expressed genes can lead to false positives and make it difficult to identify truly differentially expressed genes (Draghici et al., 2006). These problems will be

augmented in a recently tetraploid species such as Atlantic salmon, where uncharacterised paralogues will be expressed. For example, for a closely related paralogue pair where only one of the duplicates is upregulated, cross-hybridization of the single copy to probes for both paralogues could feasibly occur. Additionally, microarray experiments are limited to characterised genes and in non-model species, probes will be biased towards those mRNAs most strongly represented in EST libraries meaning important low-copy mRNAs may be overlooked. Less biased approaches for establishing genome wide responses are available which are not limited to previously identified genes. These include 'digital transcriptome' methods, such as MPSS (massive parallel signature sequencing) (Brenner et al., 2000) and the newly developed PMAGE (polony multiplex analysis of gene expression) (Kim et al., 2007). Such approaches can be used to directly sequence millions of short 'tagged' cDNAs with concurrent enumeration of each read, providing a quantitative measure of every transcript expressed (Velculescu and Kinzler, 2007; Reinartz et al., 2002). However, these approaches are of limited use for *de novo* sequencing since the short read lengths make contig reassembly difficult (Hall, 2007). As sequencing technology continues to improve, read lengths from such high-throughput sequencing methods are increasing (Shendure et al., 2005; Margulies et al., 2005) and importantly, associated costs are decreasing (Service, 2006). For example, the Roche 454 GSflx genome sequencer can, in a single run, sequence 100 Mb, incorporating 400,000 reads of up to 250 bases, that are digitally recorded to measure the abundance of each cDNA transcript (University of Liverpool Advanced Genomics Facility, <http://www.liv.ac.uk/agf/applications.html>). Such methods can provide a depth of coverage whereby the entire transcriptome could be sequenced as full-length cDNAs with concurrent 'digital' recording of transcript abundance. In the near future digital sequencing technologies will allow experiments that will provide invaluable insight into how the genome and environment interact to produce phenotypic traits. In the case of tetraploid species such as the Atlantic salmon, a transcriptome of full-length cDNAs would also be an invaluable

evolutionary and comparative-genomic resource for investigating the retention and differential expression of gene-paralogues following WGD.

Table 9.1. New genes characterised during the current project. All genes listed have been catalogued and are held in a plasmid.

Gene	Fragment Cloned	Size	Species	GenBank Accession	Assays	Paralogues
<i>myod1c</i>	complete cds and 3'UTR	1325 bp <sup>a</sup> 1815 bp <sup>b</sup>	<i>S. salar</i>	DQ317527 <sup>a</sup> DQ366709 <sup>b</sup>	RT-PCR, qPCR, <i>in situ</i> hybridization	<i>myod1a</i> (AJ557148) <i>myod1b</i> (AJ557149)
<i>fst1</i>	complete cds and partial 5'UTR	1028 bp	<i>S. salar</i>	DQ186633 <sup>a</sup>	RT-PCR <i>in situ</i> hybridization	non identified <sup>c</sup>
<i>mrf4a</i>	complete cds and 3' UTR	1125bp <sup>a</sup> 990bp <sup>b</sup>	<i>S. salar</i>	DQ479952 <sup>a</sup> DQ479951 <sup>b</sup>	RT-PCR, qPCR <i>in situ</i> hybridization	<i>mrf4b</i> (EF450078)
<i>myf5</i>	complete cds	720bp <sup>a</sup>	<i>S. salar</i>	DQ452070 <sup>a</sup>	<i>in situ</i> hybridization RT-PCR	non identified <sup>c</sup> , but suspected <sup>d</sup>
<i>myog</i>	complete cds	789 bp <sup>a</sup>	<i>S. salar</i>	DQ294029 <sup>a</sup>	RT-PCR, qPCR, <i>in situ</i> hybridization	non identified <sup>c</sup>
<i>sox8</i>	complete cds	1383 bp <sup>a</sup>	<i>S. salar</i>	DQ294028 <sup>a</sup>	RT-PCR	non identified <sup>c</sup>
decorin	complete cds	1195 bp <sup>a</sup>	<i>S. salar</i>	DQ452069 <sup>a</sup>	RT-PCR	non identified <sup>c</sup>
calpastatin	complete cds	531 bp-1506 bp <sup>f</sup>	<i>S. salar</i>	not submitted <sup>g</sup>	RT-PCR	see below <sup>h</sup>
<i>smlc1</i>	complete cds and partial 5'UTR	609 bp <sup>a</sup>	<i>S. salar</i>	DQ916288 <sup>a</sup>	<i>in situ</i> hybridization RT-PCR	not investigated
<i>cee</i>	complete cds and partial 5'UTR	1158 bp <sup>a</sup>	<i>S. salar</i>	EF036472	<i>in situ</i> hybridization RT-PCR	non identified <sup>c</sup>
<i>smhc1</i>	partial cds	684 bp <sup>a</sup>	<i>S. salar</i>	DQ369355 <sup>a</sup>	RT-PCR	not investigated
<i>foxk1</i>	partial cds	457 bp <sup>a</sup>	<i>S. salar</i>	DQ317528 <sup>a</sup>	RT-PCR	not investigated

Table 9.1. Continued on the next page.

Gene	Fragment Cloned	Size	Species	GenBank Accession	Assays	Paralogues
<i>GASPI</i>	partial cds	565 bp <sup>a</sup>	<i>S. salar</i>	DQ317526	RT-PCR	not investigated
<i>fstl-1</i>	partial cds	490 bp <sup>a</sup>	<i>S. salar</i>	DQ294031	RT-PCR	not investigated
<i>zic1</i>	partial cds	557 bp <sup>a</sup>	<i>S. salar</i>	DQ186634	RT-PCR	not investigated
<i>pgc1-α</i>	partial cds	310 bp <sup>a</sup>	<i>S. salar</i>	DQ317529	RT-PCR	not investigated
<i>myod1c</i>	complete cds	819 bp <sup>a</sup>	<i>S. trutta</i>	DQ366710	RT-PCR	<i>myod1a</i> <sup>e</sup> <i>myod1b</i> <sup>e</sup>
<i>myod2</i>	complete cds	792 bp <sup>a</sup>	<i>T. rubripes</i>	DQ413000	RT-PCR	<i>myod1</i> (AB235116)

Abbreviation: cds, coding sequence. <sup>a</sup> Indicates a cDNA sequence. <sup>b</sup> indicates genomic sequence (all introns sequenced) <sup>c</sup> No significant BLAST hits against publicly available salmonid EST libraries. <sup>d</sup> based on expression data in chapter 7. <sup>e</sup> sequence not cloned in this species. <sup>f</sup> Seven distinct full cds cDNAs were cloned of 531 bp, 585 bp, 801 bp, 873 bp, 915 bp, 1350 bp and 1506 bp. <sup>g</sup> These calpastatin sequences were not further characterised and accordingly were not submitted to GenBank. However, research with these sequences is being continued in our laboratory. <sup>h</sup> Preliminary phylogenetic analysis suggested that at least three distinct calpastatin genes are present in salmonids.

# **Appendix I. Two colour whole mount *in situ* hybridisation using sequential alkaline phosphatase staining with chromogenic substrates of *S. salar* embryos**

**Daniel J. Macqueen**

Fish Muscle Research Group, Gatty Marine Laboratory, School of Biology, University of St Andrews, Fife, KY16 8LB

## **\* key references:**

A. Q. Xu And D.G. Wilkinson. 2001. Fixation and pre-treatment of embryos for *in situ* hybridization. *Oxford Practical Approach Series, Oxford University Press*

B. Q. Xu And D.G. Wilkinson. 2001. Whole mount hybridization, washing and detection of probe (method 1). *Oxford Practical Approach Series, Oxford University Press*

C. T. Jowett. 2001. Two colour *in situ* hybridization – sequential alkaline phosphatase staining with chromogenic substrates of zebrafish embryos. *Oxford Practical Approach Series, Oxford University Press*

**General Note:** cRNA probes are very sensitive to RNase degradation and working conditions should be free from RNase until after hybridisation step is complete (then RNA is double stranded and resistant to RNase degradation). Solutions should be autoclaved and RNase Zap (Ambion) should be used to clean working surfaces and equipment.

## **Reagents**

\* *general stocks*

**DEPC water.** 0.1% v/v diethylpyrocarbonate in milli-Q water. For 1L add 1ml to 999ml milliQ water. Autoclave.

**PBS** (phosphate buffered saline). Dissolve 5 tablets (Sigma) in 1L DEPC water and autoclave OR make from 10X PBS solution (50 tablets).

**PBT.** PBS, 0.1% Tween 20 (Sigma).

**20X SSC.** For 1L:

175g 3M NaCl

88g 0.3M Na<sub>3</sub>citrate.2H<sub>2</sub>O

Dissolve in 900 ml with milliQ water and then adjust pH to 7.0 with 1M HCl and fill to 1L with milliQ water.

**2X SSC and 0.2X SSC.** Make dilutions in sterile water.

**1M NaCl.** Dissolve 58.4g in 1L deionised water and autoclave.

**1M Tris.** Dissolve 121.14g Trizma-base in 1L deionised water and autoclave. Adjust to required pH with concentrated HCl.

**10% Tween 20 (Sigma) and 10% Triton-X.**

**10% CHAPS** (store at  $-20^{\circ}\text{C}$ ).

***DIG and fluorescein probe synthesis***

**PCR reagents with M13 or T3/T7 primers.** Probe sized cDNA (700-1000 bp optimal, but 400-1500 should work fine) in vector with M13 or recognition site for T3/T7 RNA polymerases (see below)

***In vitro* transcription reagents** (see mix below)

***Embryo pre-fixation***

**Paraformaldehyde fixative (PFA)** Dissolve 4% PFA powder (Sigma) (m/v) in PBS at  $65^{\circ}\text{C}$  under fumehood. Once powder is almost dissolved add a drop of 3M NaOH to clear solution. Cool to  $4^{\circ}\text{C}$ . Note. **PFA fumes are toxic so make in fumehood!**

**Methanol** (molecular grade). 100% & 75%, 50% and 25% dilutions in PBT.

***Whole mount hybridisation, washing, and dual detection of probes***

**PFAGA** (Paraformaldehyde/glutaraldehyde) fixative. Dilute glutaraldehyde stock solution (70%) to 0.2% in PFA. Add Tween 20 to 0.1%. Store frozen in 50 ml aliquots for future use. **Make in fumehood.**

**Proteinase K** (PCR grade, 20mg/ml, Roche). 1:1000 (v/v) in PBT.

**Yeast tRNA** (Roche). 50mg/ml in DEPC water. Store in 100  $\mu\text{l}$  aliquots ( $-20^{\circ}\text{C}$ ).

**Heparin** 50mg/ml in DEPC water. Store in 100  $\mu\text{l}$  aliquots ( $-20^{\circ}\text{C}$ ).

**1M levamisole** (Sigma).

**Anti-DIG-AP AND Anti-Flu-AP fab-fragments** (Roche). Store at  $4^{\circ}\text{C}$ .

**NBT (nitro blue tetrazolium)** 100mg/ml (From Roche).

**BCIP (5-bromo-4chloro-3-indolyl phosphate)** 50mg/ml (from Roche).

**Hybridization mix:** For 100ml:

Blocking reagent	2g
10% Triton X	1ml
10% CHAPS	1ml
20X SSC	25ml
Formamide	50ml
tRNA(as above)	100 $\mu\text{l}$
heparin (as above)	100 $\mu\text{l}$

0.2M EDTA	2.5ml
DEPC water	20.3 ml

**Sheep Serum.** Heat at 56°C to inactivate and then store in 50 ml aliquots at -20°C

**Colouration buffer** For 100ml:

5ml 1M MgCl<sub>2</sub>  
10ml 1M Tris pH 9.5  
10ml 1M NaCl<sub>2</sub>  
100µl Levamisole  
10ml 10% Tween 20  
To 100ml with DEPC water

**Blocking Solution.** For 100ml

4ml BSA (50mg/ml) (need 2mg/ml i.e. 1ml/25)  
5ml sheep serum  
1ml DMSO (dimethylsulophoxide, Sigma)  
90ml PBT

**Washing solution.** For 100ml

4ml BSA (50mg/ml) (need 2mg/ml i.e. 1ml/25)  
1ml DMSO  
95ml PBT

**Pre-stain buffer:** 100mM Tris-HCl, pH 8.2, 0.1% Tween 20.

**Sigma Fast™ Fast Red tablets (Invitrogen).** To develop fluorescein stain. Follow manufacturers instructions.

**AP inactivation solution:** 0.1M glycine-HCl pH 1.2, 0.1% tween 20. Adjust glycine solution to pH 2.2 with HCl and add tween 20 to 0.1%

4.5 µl/ml NBT and 3.5µl/ml BCIP made up in coloration buffer

## Protocol

### A. *Creation of DIG and fluorescein probes*

Amplify probe DNA using PCR in following mix:

10X PCR buffer	2.5 µl
10mM dNTPs	0.5µl
20µM m13 F	0.5µl
19µM m13 R	0.5µl
5U/µl Taq polymerase	0.2 µl
Plasmid DNA (1:10)	0.5µl
DEPC water	20.3µl

Cycles:

5 minute denaturation @ 95°C

Then 35 cycles of:

95°C 30s

56°C 30s  
72°C 1 minute

### **NO FINAL EXTENTION**

Run on agarose gel containing 1 µl/100ml EtBr. Cut band, ensuring agarose is minimal, and then extract DNA from agarose.

Quantify products on a 1% agarose gel using quantitative 1kb DNA marker (e.g. from New England Biolabs). A typical yield is 10-30 ng/µl.

### **THEN TRANSCRIBE PROBES *in vitro***

Prepare the following mix

100-200 ng PCR product	Xµl (depending on DNA concentration)
10X DIG labelling mix OR fluorescein mix	2µl
10X transcription buffer (vortex to resuspend)	2.0µl
40U/µl Rnasin (Promega)	0.5µl
T3 or T7 polymerase (Roche)	2.0µl

Make up to 20 µl with DEPC water.

Incubate at 37°C for 2 hr. Then add 2 µl DNase (Ambion) and reincubate at 37°C for 15 minutes.

Stop reaction: add 2ul of 0.2M EDTA (pH 8.0).

Purify labelled RNA: Add 2.5 µl 4M LiCl (Sigma) and 75 µl (-20°C) ethanol. Leave at -70°C overnight to allow precipitate to form.

Centrifuge at 13,000 g for 5 minutes at 4°C.

Rinse pellet with 50µl ice-cold 100% 70% ethanol (v/v in DEPC water).

Centrifuge at 13,000 g for 15 minutes at 4°C. Discard supernatant, dry pellet for 10 minutes in a fumehood. Resuspend pellet in 100µl fresh DEPC water by incubating at 37°C for 30 minutes.

Save 5 µl for electrophoresis and save 1-2 µl for quantification. Freeze remaining probe at -70°C in 20µl aliquots.

Quantify probe and check integrity with gel electrophoresis. Mix 1-2µl with 8-9 µl RNA buffer. Incubate at 65°C for 5-10 minutes to denature 2° RNA structures. Run gel with semi quantitative ladder RNA marker and quantify RNA mass. The cRNA should appear as a discrete band of the expected size. Typical mass = 1µg per 10µl

### **B. Embryo pre-fixation**

Fix embryos in PFA for 30 minutes. Make a small puncture in the chorion and incubate at 4°C overnight in PFA.

Rinse embryos with ice-cold PBT for 2 X 5 minutes on a rocking platform.

Dehydrate embryos with the following washes:

25% methanol/PBT	10 minutes
50% methanol/PBT	10 minutes
75% methanol/ PBT	10 minutes
Pure methanol	2X 10 minutes

\*Embryos can be stored in methanol at  $-20^{\circ}\text{C}$  or  $-70^{\circ}\text{C}$  and are stable for many months. At room temperature embryos are stable for a matter of days

### *C. Whole mount hybridisation, washing, and dual detection of probes*

1. Heat hybridisation mix to  $60-70^{\circ}\text{C}$ , thaw PFAGA, label 8ml sterile bijoux vials (if using) with a sticker and pencil

2. Rehydrate embryos with following washes. Note methanol and PFA are toxic. Complete procedures in fumehood and dispose in appropriate fashion.

75% methanol/ PBT	5 minutes
50% methanol/PBT	5 minutes
25% methanol/PBT	5 minutes
PBT	2 X 5 minutes

3. Carefully dechorionate using watchmaker's forceps and a binocular microscope under bright field illumination. Be Careful! Embryos are delicate and easy to break. You need intact embryos to get nice images!

4. Permeabilise the embryos using proteinase K diluted 1:1000 in PBT.

For **salmonid** embryos digest for the following periods according to somite number:

Pre-segmentation 5 minutes at room temperature  
1-40 somites 10 minutes at room temperature  
40+ somites 15 minutes at  $37^{\circ}\text{C}$

5. Wash off proteinase refix embryos with the following washes:

PBT	2 x 5 minutes
In Fumehood: PFAGA	20 minutes
PBT	3 x 5 minutes

6. Pre-hybridization: immerse embryos in pre-heated hybridization mix for 1 hour at  $70^{\circ}\text{C}$  **OR** overnight at  $4^{\circ}\text{C}$ .

7. Hybridization: immerse embryos in pre-heated hybridization mix containing  $1\mu\text{g/ml}$  anti-sense DIG labelled probe for GENE X and  $1\mu\text{g/ml}$  anti-sense Fluorescein labelled probe for GENE Y. Use DIG and fluorescein sense RNA together as a control.

Use autoclave tape to stop lid from popping and then incubate the mix for 2-3 days at  $60-70^{\circ}\text{C}$  (higher temperature = higher stringency of mRNA binding).

8. Heat stringency wash Solution to desired temperature (see 8) and thaw Sheep Serum.

9. Perform the following washes of decreasing stringency (higher temp means higher stringency i.e. less specifically bound RNAs will be washed away)

2 X SSC	2 X 5 minutes
2 X SSC/ 0.1% CHAPS	3 X 15 minutes
0.2X SSC/0.1% CHAPS	3 X 20 minutes

10. Wash embryos in PBT for 2 X 10 minutes at room temperature.

11. Replace PBT with blocking solution and leave for at least 1 hour at room temperature. This will reduce background staining.

12. Replace blocking solution with a 1:5000 dilution of anti-FLU-AP fab fragments in blocking solution. Incubate for 2 hours at room temperature or overnight at 4°C.

13. Wash embryos for 8 X 15 minutes in washing solution.

14. Wash embryos for 4 X 5 minutes in colouration buffer OR Pre-stain buffer.

15. Develop red colouration (anti-flourescein) with Fast Red. Reaction should be allowed to proceed in the dark at room temperature or overnight at 4°C. Stop reaction by washing embryos with PBT or DEPC water.

16. Incubate in AP inactivation solution for 30 minutes to stop the AP. Then wash several times in PBT.

**This marks the completion of the first staining and GENE X expression should be visible as red staining. Move onto to 2<sup>nd</sup> stain for GENE Y (DIG)**

17. Fix the embryos again by immersion in PFA for 20 minutes.

18. Replace the PFA with blocking solution for 2 quick washes and then for at least an hour at room temperature. Again, this will reduce background staining.

19. Incubate the embryos for 2 hours with 1:2000 anti-DIG-AP in blocking solution at room temperature on rocking platform.

20. Wash 5 X 15 minutes in PBT to remove unbound anti-body. Embryos can be left o/n during one of the later PBT washes. .

21. Wash embryos for 4 X 5 minutes in colouration buffer.

22. Visualize the DIG signal using colouration buffer containing 4.5 µl/ml NBT and 3.5µl/ml BCIP. Incubate embryos in the dark (use tin foil) at 4°C. Allow development to proceed until desired signal is strongest with minimal background.

Wash embryos in PBT. If background signal is strong, embryos can be left in PBT at 4°C overnight or longer. This will also develop the colour of the specific DIG signal in relation to background staining.

Fix embryos for 1 hour in PFA.

Rinse in PBS and then store in PBS. Embryos can be stored in PBS for several weeks, but for longer storage, Sodium Azide should be added to 0.1%. DIG label is very stable (can last for months to years without NaAz treatment) but Flu staining can fade in short time.

## **Appendix II: List of manufacturers addresses**

Ambion, Cambridgeshire, UK

Applied Biosystems, Cheshire, UK

BD Biosciences, Oxford, UK

Bioline, London, UK

Bio-Rad, Hertfordshire, UK

BIOQUELL, Hampshire, UK

BMG Lab Technologies, Buckinghamshire, UK

Eppendorf, Cambridgeshire, UK

Fish Eagle Company, Gloucester, UK

Fisher Scientific, Leicestershire UK

Invitrogen, Renfrewshire, UK

Leica Microsystems, Buckinghamshire, UK

Millipore, Hertfordshire, UK

MP Biomedicals, Cheshire, UK

NanoDrop Technologies, Delaware, USA

New England Biolabs, Hertfordshire, UK

New Zealand Ltd, Lower Hutt, New Zealand

Nikon, Surrey, UK

Promega, Hampshire, UK

Qiagen, West Sussex, UK

Roche, Sussex, UK

Sigma, Dorset, UK

Thermo Electron Corporation, Waltham, Massachusetts, USA

Zeiss, Warwickshire, UK

## References

- Albokhadaim, I., Hammond, C.L., Ashton, C., Simbi, B.H., Bayol, S., Farrington, S. and Stickland, N. 2007. Larval programming of post-hatch muscle growth and activity in Atlantic salmon (*Salmo salar*). *The Journal of Experimental Biology* **210**: 1735-1741.
- Allendorf, F.W. and Thorgaard, G.H. 1984. Tetraploidy and evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (Turner, B.J., ed) pp1-53. Plenum Publishing Corporation, New York.
- Altringham, J.D. and Johnston, I.A. 1988. Activation of multiply innervated, fast and slow myotomal muscle fibres of the teleost *Myoxocephalus scorpius*. *The Journal of Experimental Biology* **140**: 313-324.
- Altringham, J.D. and Johnston, I.A. 1990. Modelling muscle power output in a swimming fish. *The Journal of Experimental Biology* **148**: 395-402.
- Altringham, J.D. and Ellerby, D.J. 1999. Fish swimming: patterns in muscle function. *The Journal of Experimental Biology* **202**: 3397-3403.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431**: 350-355.
- 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**: 1711-1714.
- Amthor, H., Connolly, D., Patel, K., Brand-Saberi, B., Wilkinson, D.G., Cooke, J., and Christ, B. 1996. The expression and regulation of follistatin and a follistatin-like gene during avian somite compartmentalization and myogenesis. *Developmental Biology* **178**: 343-362.

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**: 115-127.

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

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., 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. and Brenner, S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**: 1301-1310.

Armand, A.S., Lécolle, S., Launay, T., Pariset, C., Fiore, F., Della Gaspera, B., Birnbaum, D., Chanoine, C. and Charbonnier, F. 2004. IGF-II is up-regulated and myofibres are hypertrophied in regenerating soleus of mice lacking FGF6. *Experimental Cell Research* **297**: 27-38.

Armstrong, J.D. and Nislow, K.H. 2006. Critical habitat during the transition from maternal provisioning in freshwater fish, with emphasis on Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*). *Journal of Zoology* **269**: 403-413.

Atchley, W.R., Fitch, W.M. and Bronner-Fraser, M. 1994. Molecular evolution of the MyoD family of transcription factors. *Proceedings of the National Academy of Sciences, USA* **91**: 11522-11526.

Ayala, M.D., López-Albors, O., Gil, F., Latorre, R., Vázquez, J.M., García-Alcázar, A., Abellán, E., Ramírez, G. and Moreno, F. Temperature effect on muscle growth of the axial musculature of the sea bass (*Dicentrarchus labrax* L.). *Anatomia Histologia Embryologia* **29**: 235-241.

Babushok, D.V., Ostertag, E.M. and Kazazian, H.H. Jr. 2007. Current topics in genome evolution: molecular mechanisms of new gene formation. *Cellular and Molecular Life Sciences* **64**: 542-554.

Bailey, G.S., Poulter, R.T. and Stockwell, P.A. 1978. Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of Sciences, USA* **75**: 5575–5579.

Barnoy, S., Glasner, T. and Kosower, N.S. 1996. The role of calpastatin (the specific calpain inhibitor) in myoblast differentiation and fusion. *Biochemical and Biophysical Research Communications* **220**: 933-938.

Barnoy, S., Glaser, T. and Kosower, N.S. 1998. The calpain–calpastatin system and protein degradation in fusing myoblasts. *Biochimica et Biophysica Acta* **1402**: 52–60.

Barnoy, S., Maki, M. and Kosower, N.S. 2005. Overexpression of calpastatin inhibits L8 myoblast fusion. *Biochemical and Biophysical Research Communications* **332**: 697-701.

Barresi, M.J., Stickney, H.L. and Devoto, S.H. 2000. The zebrafish slow-muscle-omitted gene product is required for Hedgehog signal induction and the development of slow muscle identity. *Developmental Biology* **127**: 2189-2199.

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

Barut, B.A. and Zon, L.I. 2000. Realizing the potential of zebrafish as a model for human disease. *Physiological Genomics* **2**: 49-51.

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

Bassel-Duby, R., Hernandez, M.D., Yang, Q., Rochelle, J.M., Seldin, M.F. and Williams, R.S. 1994. Myocyte nuclear factor, a novel winged-helix transcription factor under both developmental and neural regulation in striated myocytes. *Molecular and Cellular Biology* **14**: 4596-4605.

- 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**: 488-507.
- Baxendale, S., Davison, C., Muxworthy, C., Wolff, C., Ingham, P.W. and Roy, S. 2004. The B-cell maturation factor specifies slow-twitch muscle fiber identity in response to hedgehog signalling. *Nature Genetics* **36**: 88-93.
- Bennett, A.F. 1978. Activity metabolism of the lower vertebrates. *Annual Review of Physiology* **40**: 447-469.
- Benton, M.J. and Donoghue, P.C. 2007. Paleontological evidence to date the tree of life. *Molecular Biology and Evolution* **24**: 26:53.
- Bergsten, J. 2005. A review of long-branch attraction. *Cladistics* **21**: 163-193.
- 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 Cellular Biology* **21**: 2404-2412.
- Berkes, C.A. and Tapscott, S.J. 2005. MyoD and the transcriptional control of myogenesis. *Seminars in Cell and Developmental Biology* **16**: 585-595.
- Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S. and Tapscott, S.J. 2004. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Molecular Cell* **14**: 465-477.
- Bisbee, C.A., Baker, M.A., Wilson, A.C., Haji-Azimi, I., Fischberg, M. 1977. Albumin phylogeny for clawed frogs (*Xenopus*). *Science* **195**: 785-787.
- 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**: 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.
- Blair, J.E., Shah, P. and Hedges, S.B. 2005. Evolutionary sequence analysis of complete eukaryote genomes. *BMC Bioinformatics* **6**: 53.
- Blais, A.B., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y. and Dynlacht, B.D. 2005. An initial blueprint for myogenic differentiation. *Genes and Development* **19**: 553-569.
- 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**: 244-256.
- Bodine, S.C., Stitt, T.N., Gonzalez, M., Kline, W.O., Stover, G.L., Bauerlein, R., Zlotchenko, E., Scrimgeour, A., Lawrence, J.C., Glass, D.J. and Yancopoulos, G.D. 2001. Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. *Nature Cell Biology* **3**: 1014-1019.
- Bogdanovich, S., Krag, T.O., Barton, E.R., Morris, L.D., Whittemore, L.A., Ahima, R.S. and Khurana, T.S. 2002. Functional improvement of dystrophic muscle by myostatin blockade. *Nature* **420**: 418-421.
- Bone, Q. 1964. Patterns of muscular innervation in the lower chordates. *International Reviews of Neurobiology* **6**: 99-147.
- Bone, Q. 1966. On the function of the two types of myotomal muscle fibres in elasmobranch fish. *Journal of the Marine Biological Association of the UK* **46**: 321-349.
- Bone, Q. 1978. Locomotor Muscle. In *Fish Physiology Series Volume 7: Locomotion* (Hoar, W.S. and Randall, D.J., eds), pp 361-424. Academic Press, New York.
- Bone, Q. 1989. Evolutionary patterns of axial muscle systems in some invertebrates and fish. *American Zoologist* **29**: 5-18.

- Bone, Q., Kiceniuk, J. and Jones, D.R. 1978. On the role of the different fibre types in fish myotomes at intermediate swimming speeds. *Fishery Bulletin U.S.* **76**: 691-699.
- 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**: 239-255.
- Bowman, A.W. and Azzalini, A. 1997. *Applied smoothing techniques for data analysis. The Kernal approach with S-plus illustrations*. Oxford UK, Oxford University Press.
- Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., and Arnold, H.H. 1989. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *The EMBO Journal* **8**: 701-709.
- Braun, T., Rudnicki, M.A., Arnold, H.H. and Jaenisch, R. 1992. Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell* **71**: 369-382.
- Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., Aparicio, S. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**: 265-268.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* **18**: 630-634.
- Bridges, C.B. 1936. The bar 'gene' a duplication. *Science* **83**: 210-211.
- Brooks, S. and Johnston, I.A. 1993. Influence of development and rearing temperature on the distribution, ultrastructure and myosin sub-unit composition of myotomal muscle fibre types in the plaice, *Pleuronectes platessa*. *Marine Biology* **117**: 501-513.

- Brunet, F.G., Roest Crolius, H., Paris, M., Aury, J., Gibert, P., Jaillon, O., Laudet, V., and Robinson-Rechavi M. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Molecular Biology and Evolution* **23**: 1008-1016.
- Buckingham, M. and Relaix, F. 2007. The role of pax genes in the development of tissues and organs: pax3 and pax7 regulate muscle progenitor cell functions. *Annual Review of Cell and Developmental Biology* **23**: 645-73.
- Calvo, J. and Johnston, I.A. 1992. Influence of rearing temperature on the distribution of muscle fibre types in turbot *Scophthalmus maximus* at metamorphosis. *Journal of Experimental Biology and Ecology* **161**: 45-55.
- Capecchi, M.R. 2005. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews Genetics* **6**: 507-512.
- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and Werner, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**: 2933-2942.
- Casas, E., White, S.N., Wheeler, T.L., Shackelford, S.D., Koohmaraie, M., Riley, D.G., Chase, C.C. Jr, Johnson, D.D. and Smith, T.P. 2006. Effects of calpastatin and micro-calpain markers in beef cattle on tenderness traits. *Journal of Animal Science* **84**: 520-525.
- Castillo, J., Le Bail, P.Y., Paboeuf, G., Navarro, I., Weil, C., Fauconneau, B. and Gutiérrez, J. 2002. IGF-I binding in primary culture of muscle cells of rainbow trout: changes during in vitro development. *American Journal of Physiology* **283**: R647-R652.
- Castillo, J., Codina, M., Martínez, M.L., Navarro, I. and Gutiérrez, J. 2004. Metabolic and mitogenic effects of IGF-I and insulin on muscle cells of rainbow trout. *American Journal of Physiology* **286**: R935-R941.
- Chauvigné, F., Cauty, C., Rallièrre, C. and Rescan, P.Y. 2005. Muscle fiber differentiation in fish embryos as shown by in situ hybridization of a large repertoire of muscle-specific transcripts. *Developmental Dynamics* **233**: 659-666.

- Chen, J.F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L. and Wang, D.Z. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics* **38**: 228-233.
- Chen, W.S., Xu, P.Z., Gottlob, K., Chen, M.L., Sokol, K., Shiyanova, T., Roninson, I., Weng, W., Suzuki, R., Tobe, K., Kadowaki, T. and Hay, N. 2001. Growth retardation and increased apoptosis in mice with homozygous disruption of the Akt1 gene. *Genes and Development* **15**: 2203-2208.
- Chin, E.R., Olson, E.N., Richardson, J.A., Yang, Q., Humphries, C., Shelton, J.M., Wu, H., Zhu, W., Bassel-Duby, R. and Williams, R.S. 1998. A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes and Development* **15**: 2499-2509.
- Christoffels, A., Koh, E.G., 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**: 1146-1151.
- Ciobanu, D.C., Bastiaansen, J.W., Lonergan, S.M., Thomsen, H., Dekkers, J.C., Plastow, G.S. and Rothschild, M.F. 2004. New alleles in calpastatin gene are associated with meat quality traits in pigs. *Journal of Animal Science* **82**: 2829-2839.
- Clark, J.M. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research* **16**: 9677-9686.
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.M., Eyche, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. and Georges, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* **38**: 813-818.
- Clydesdale, F.M. 1993. Color as a factor in food choice. *Critical Reviews in Food Science and Nutrition* **33**: 83-101.
- Coe, M. J. 1966. The biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica* **23**: 146-177.

- 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. *The Journal of Experimental Biology* **207**: 4239-4248.
- Cortés, F., Daggett, D., Bryson-Richardson, R.J., Neyt, C., Maule, J., Gautier, P., Hollway, G.E., Keenan, D. and Currie PD. 2003. Cadherin-mediated differential cell adhesion controls slow muscle cell migration in the developing zebrafish myotome. *Developmental Cell* **5**: 865-876.
- Cossins, A.R. and Crawford, D.L. 2005. Fish as models for environmental genomics. *Nature Reviews Genetics* **6**: 324-333.
- Cossu, G. and Tajbakhsh, S. 2007. Oriented cell divisions and muscle satellite cell heterogeneity. *Cell* **129**: 859-861.
- Coutelle, O., Blagden, C.S., Hampson, R.J., 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**: 136-150.
- Croall, D.E. and DeMartino, G.N. 1991. Calcium-activated neutral protease (calpain) system: structure, function, and regulation. *Physiological Reviews* **71**: 813-847.
- Crow, K.D., Wagner, G.P. and SMBE Tri-National Young Investigators. 2006. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. What is the role of genome duplication in the evolution of complexity and diversity? *Molecular Biology and Evolution* **23**: 887-892.
- Currie, P.D. and Ingham, P.W. 1996. Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature* **382**: 452-455.
- Currie, P.D. and Ingham, P.W. 2001. Induction of embryonic skeletal muscle cells in the zebrafish. In *Fish Physiology Series Volume 18: Muscle Development and Growth* (Johnston I.A., ed) pp 1-17. Academic Press, San Diego.

- Daggett, D.F., Domingo, C.R., Currie, P.D. and Amacher, S.L. 2007. Control of morphogenetic cell movements in the early zebrafish myotome. *Developmental Biology* **309**: 169-179.
- Dal-Pra, S., Fürthauer, M., Van-Celst, J., Thisse, B. and Thisse, C. 2006. Noggin1 and follistatin-like2 function redundantly to chordin to antagonize BMP activity. *Developmental Biology* **298**: 514-526.
- Darr, K.C. and Schultz, E. 1987. Exercise-induced satellite cell activation in growing and mature skeletal muscle. *Journal of Applied Physiology* **63**: 1816–1821.
- Davis, R.L., Weintraub, H. and Lassar, A.B. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.
- DeChiara, T.M., Efstratiadis, A. and Robertson, E.J. 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**: 78-80.
- Dedieu, S., Mazeres, G., Dourdin, N., Cottin, P., Brustis, J.J. 2003. Transactivation of *capn2* by myogenic regulatory factors during myogenesis. *Journal of Molecular Biology* **326**: 453-465.
- 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**: 432-737.
- Delsuc, F., Brinkmann, H., Chourrout, D. and Philippe, H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**: 965-968.
- de Souza, F.S., Bumachny, V.F., Low, M.J., and Rubinstein, M. 2005. Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. *Molecular Biology and Evolution* **22**: 2417-2427.
- Devoto, S.H., Melançon, E., Eisen, J.S. and Westerfield, M. 1996. Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development* **122**: 3371-3380.

- Devoto, S.H., Stoiber, W., Hammond, C.L., Steinbacher, P., Haslett, J.R., Barresi, M.J., Patterson, S.E., Adiarte, E.G. and Hughes, S.M. 2006. Generality of vertebrate developmental patterns: evidence for a dermomyotome in fish. *Evolution and Development* **8**: 101-110.
- Donoghue, P.C. and Purnell, M.A. 2005. Genome duplication, extinction and vertebrate evolution. *Trends in Ecology and Evolution* **20**: 312-319.
- Dooley, K. and Zon, L.I. 2000. Zebrafish: a model system for the study of human disease. *Current Opinion in Genetics and Development* **10**: 252-256.
- Draghici, S., Khatri, P., Eklund, A.C. and Szallasi, Z. 2006. Reliability and reproducibility issues in DNA microarray measurements. *Trends in Genetics* **22**: 101-109.
- Du, S.J., Devoto, S., 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 Science* **139**: 145-156.
- Dunn, S.E., Burns, J.L. and Michel, R.N. 1999. Calcineurin is required for skeletal muscle hypertrophy. *The Journal of Biological Chemistry* **274**: 21908-21912.
- Dunn, S.E., Chin, E.R. and Michel, R.N. 2000. Matching of calcineurin activity to upstream effectors is critical for skeletal muscle fiber growth. *The Journal of Cell Biology* **151**: 663-672.
- Dunn, S.E., Simard, A.R., Prud'homme, R.A. and Michel, R.N. 2002. Calcineurin and skeletal muscle growth. *Nature Cell Biology* **4**: E46.
- Duston, J. and Saunders, R.L., 1995. Advancing smolting to autumn in age 0+ Atlantic salmon by photoperiod, and long-term performance in sea water. *Aquaculture* **135**: 295-309.
- Edmonson, D.G. and Olson, E.N. 1989. A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes and Development* **3**: 628-640.
- Edmonson, D.G. and Olson, E.N. 1993. Helix-loop-helix proteins as regulators of muscle-specific transcription. *The Journal of Biological Chemistry* **268**: 755-788.

Edmondson, D.G., Lyons, G.E., Martin, J.F. and Olson, E.N. 1994. Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* **120**: 1251-1263.

Ekker, M., Wegner, J., Akimenko, M.A. and Westerfield, M. 1992. Coordinate embryonic expression of three zebrafish engrailed genes. *Development* **116**: 1001-1010.

Esch, F.S., Shimasaki, S., Mercado, M., Cooksey, K., Ling, N., Ying, S., Ueno, N. and Guillemin, R. 1987. Structural characterization of follistatin: a novel follicle-stimulating hormone release-inhibiting polypeptide from the gonad. *Molecular Endocrinology* **1**: 849-855.

Evans, B.J., Kelley, D.B., Tinsley, R.C., Melnick, D.J. and Cannatella, D.C. 2004. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution* **33**: 197-213.

FAO (Food and Agriculture Organisation of the United Nations) culture species information programme: *Salmo salar* (Linnaeus, 1758). [http://www.fao.org/fi/website/FIRetrieveAction.do?dom=culturespecies&xml=Salmo\\_salar.xml](http://www.fao.org/fi/website/FIRetrieveAction.do?dom=culturespecies&xml=Salmo_salar.xml).

FAO (Food and Agriculture Organisation of the United Nations). 2003. 'Aquaculture: not just an export industry'. <http://www.fao.org/english/newsroom/focus/2003/aquaculture.htm>.

FAO. (Food and Agriculture Organisation of the United Nations). 2006. 'The state of world fisheries and aquaculture'. [http://www.fao.org/sof/sofia/index\\_en.htm](http://www.fao.org/sof/sofia/index_en.htm).

Fares, M.A., Byrne, K.P. and Wolfe, K.H. 2006. Rate asymmetry after genome duplication causes substantial long-branch attraction artifacts in the phylogeny of *Saccharomyces* species. *Molecular Biology and Evolution* **23**: 245-253.

Felsenfeld, A.L., Curry, M., and Kimmel, C.B. 1991. The fub-1 mutation blocks initial myofibril formation in zebrafish muscle pioneer cells. *Developmental Biology* **148**: 23-30.

Feng, X., Adiarte, E.G., and Devoto, S.H. 2006. Hedgehog acts directly on the zebrafish dermomyotome to promote myogenic differentiation. *Developmental Biology* **300**: 736-746.

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.

Fernandes, J.M.O., Kinghorn, J.R. and Johnston, I.A. 2007. Differential regulation of multiple alternatively spliced transcripts of MyoD. *Gene* **391**: 178-185.

Fernandes, J.M., MacKenzie, M.G., Kinghorn, J.R. and Johnston, I.A. 2007. FoxK1 splice variants show developmental stage-specific plasticity of expression with temperature in the tiger pufferfish. *The Journal of Experimental Biology* **210**: 3461-3472.

Fernandes, J.M.O., Macqueen, D.J., Lee, H-T. and Johnston, I.A. 2007. Genomic, evolutionary and expression analyses of *cee*, an ancient gene involved in normal growth and development *Genomics: In press*. DOI: 10.1016/j.ygeno.2007.10.017.

Finn, R.N. and Kristoffersen, B.A. 2007. Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts. *PLoS One* **2**: e169.

Fleischer, T.C., Weaver, C.M., McAfee, K.J., Jennings, J.L. and Link, A.J. 2006. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes and Development* **20**: 1294-1307.

Fletcher, G., Hew, C.H. and Davies, P.L. 2001. Antifreeze proteins of teleost fishes. *Annual Review of Physiology*. **63**: 359-390.

Florini, J.R., Ewton, D.Z. and Coolican, S.A. 1996. Growth hormone and the insulin-like growth factor system in myogenesis. *Endocrine Reviews* **17**: 481-517.

Flynt, A.S., Li, N., Thatcher, E.J., Solnica-Krezel, L. and Patton, J.G. 2007. Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. *Nature Genetics* **39**: 259-263.

- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yam, Y.L. and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **252**: 1531-1545.
- Gabillard, J.C., Rescan, P.Y., Fauconneau, B., Weil, C. and Le Bail, P.Y. 2003. Effect of temperature on gene expression of the Gh/Igf system during embryonic development in rainbow trout (*Oncorhynchus mykiss*). *The Journal of Experimental Zoology* **298A**: 134-142.
- Gabillard, J.C., Yao, K., Vandeputte, M., Gutierrez, J., Le Bail, P.Y. 2006. Differential expression of two GH receptor mRNAs following temperature change in rainbow trout (*Oncorhynchus mykiss*). *The Journal of Endocrinology* **190**: 29-37.
- Galloway, T.F., Bardal, T., Kvam, S.N., Dahle, S.W. Nesse, G., Randøl, M., Kjørsvik, E. and Andersen, Ø. 2006. Somite formation and expression of MyoD, myogenin and myosin in Atlantic halibut (*Hippoglossus hippoglossus* L.) embryos incubated at different temperatures: transient asymmetric expression of MyoD. *The Journal of Experimental Biology* **209**: 2432-2441.
- Garcia-Fernández, J. and Holland, P.W.H. 1996. Amphioxus Hox genes: insights into evolution and development. *International Journal of Developmental Biology* **Supplement 1** 71S-72S.
- Garikipati, D.K., Gahr, S.A. and Rodgers, B.D. 2006. Identification, characterization, and quantitative expression analysis of rainbow trout myostatin-1a and myostatin-1b genes. *The Journal of Endocrinology* **198**: 879-888.
- Garikipati, D.K., Gahr, S.A., Roalson, E.H. and Rodgers, B.D. 2007. Characterization of rainbow trout myostatin-2 genes (rtMSTN-2a and -2b): genomic organization, differential expression, and pseudogenization. *Endocrinology* **148**: 2106-2115.
- Garry, D. J., Yang, Q., Bassel-Duby, R. and Williams, R.S. 1997. Persistent expression of MNF identifies myogenic stem cells in postnatal muscles. *Developmental Biology* **188**: 280-294.
- 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. *Proceedings of the National Academy of Sciences, USA* **97**: 5416-5421.

Geer, L.Y., Domrachev, M., Lipman, D.J. and Bryant, S.H. 2002. CDART: protein homology by domain architecture. *Genome Research* **12**: 1619-1623.

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**: 436-450.

Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Høyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., Okamoto, N., Ferguson, M.M., Holm, L.E. and Guyomard, R. 2006. A linkage map for brown trout (*Salmo trutta*): chromosome homologies and comparative genome organization with other salmonid fish. *Genetics* **172**: 2405-2419.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M. Gotte, D., Güldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kötter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W. and Johnston, M. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387-391.

Glass, D.J. 2003. Signalling pathways that mediate skeletal muscle hypertrophy and atrophy. *Nature Cell Biology* **5**: 87-90.

Glass, D.J. 2005. Skeletal muscle hypertrophy and atrophy signaling pathways. *The International Journal of Biochemistry and Cell Biology* **37**: 1974-1984.

Goll, D.E., Thompson, V.F., Li, H., Wei, W. and Cong, J. 2003. The calpain system. *Physiological Reviews* **83**: 731-801.

- Gorodilov, Y.N. 1996. Description of the early ontogeny of the Atlantic salmon, *Salmo salar*, with a novel system of interval (state) identification. *Environmental Biology of Fishes* **47**: 109-127.
- Gossett, L.A., Kelvin, D.J., Sternberg, E.A., and Olson, E.N. 1989. A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Molecular and Cellular Biology* **9**: 5022-5033.
- Gotensparre, S.M. 2004. Characterisation of myogenic genes in skeletal muscle of Atlantic salmon (*Salmo salar* L.). PhD thesis. University of St Andrews.
- Gotensparre, S.M., Anderssen, E. Wargelius, A., Hansen, T. and Johnston, I.A. 2006. Insight into the complex genetic network of tetraploid Atlantic salmon (*Salmo salar* L.): Description of multiple novel Pax-7 splice variants. *Gene* **373**: 8-15.
- Gracey, A.Y., Fraser, E.J., Li, W., Fang, Y., Taylor, R.R., Rogers, J., Brass, A. and Cossins, A.R. 2004. Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proceedings of the National Academy of Sciences, USA* **101**: 16970-16975.
- Greer-Walker, M. and Pull, G.A. 1975. A survey of red and white muscle in marine fish. *Journal of Fish Biology* **7**: 295-300.
- Gregory, D.J., Waldbieser, G.C. and Bosworth, B.G. 2004. Cloning and characterization of myogenic regulatory genes in three ictalurid species. *Animal Genetics* **53**: 425-430.
- Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Ménissier, F., Massabanda, J., Fries, R., Hanset, R., Georges, M. 1997. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature Genetics* **17**: 71-74.
- Groves, J.A., Hammond, C.L. and Hughes, S.M. 2005. Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. *Development* **132**: 4211-4222.

- Gros, J., Manceau, M., Thome, V. and Marcelle, C. 2005. A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* **435**: 954-958.
- Grunwald, D.J. and Eisen, J.S. 2002. Headwaters of the zebrafish -- emergence of a new model vertebrate. *Nature Reviews Genetics* **3**: 717-724.
- Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W. and Li, W.H. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**: 63-66.
- Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696-704.
- Guindon, S., Lethiec, F., Duroux, P. and Gascuel, O. 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* **33**: W557-W559.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.
- Hall, T.E. and Johnston, I.A. 2003. Temperature and developmental plasticity during embryogenesis in the Atlantic cod *Gadus morhua* L. *Marine Biology* **142**: 833-840.
- Hall, T.E., Cole, N.J. and Johnston, I.A. 2003. Temperature and the expression of seven muscle-specific genes during embryogenesis in the Atlantic cod *Gadus morhua* L. *The Journal of Experimental Biology* **206**: 3187-3200.
- Hall, N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *The Journal of Experimental Biology* **210**: 1518-1525.
- Hamade, A., Deries, M., Begemann, G., Bally-Cuif, L., Genêt, C., Sabatier, F., Bonniou, A. and Cousin, X. 2006. Retinoic acid activates myogenesis in vivo through Fgf8 signalling. *Developmental Biology* **289**: 127-140.

- Hammond, C.L., Hinitz, Y., Osborn, D.P.S., Minchin, J.E.N., Tettamanti, G. and Hughes, S.M. 2007. Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. *Developmental Biology* **302**: 504-521.
- Hansen, T., Stefansson, S.O. and Taranger, G.L. 1992. Growth and sexual maturation in Atlantic salmon, *Salmo salar* L., reared in sea cages at two different light regimes. *Aquaculture and Fisheries Management* **23**: 275-280.
- Hansen, L.P., Jonsson, N. and Jonsson, B. 1993. Oceanic migration in homing Atlantic salmon. *Animal Behaviour* **45**: 927-941.
- 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**: 501-506.
- Hatta, K., BreMiller, R.A., Westerfield, M. and Kimmel, C.B. 1991. Diversity of expression of engrailed-like antigens in zebrafish. *Development* **112**: 821-832.
- Hawke, T.J. and Garry, D.J. 2001. Myogenic satellite cells: physiology to molecular biology. *Journal of Applied Physiology* **91**: 534-551.
- Hawke, T.J., Jiang, H and Garry, D.J. 2002. Absence of p21CIP rescues myogenic proliferative and regenerative capacity in Foxk1 null mice. *The Journal of Biological Chemistry* **278**: 4015-4020.
- He, X. and Zhang, J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**: 1157-1164.
- Hedges, S.B., Chen, H., Kumar, S., Wang, D.Y., Thompson, A.S and Watanabe, H. 2001. A genomic timescale for the origin of eukaryotes. *BMC Evolutionary Biology* **1**: 4.
- Hedges, S.B. 2002. The origin and evolution of model organisms, *Nature Reviews Genetics* **3**: 838-849.

- Henry, C.A. and Amacher, S.L. 2004. Zebrafish slow muscle cell migration induces a wave of fast muscle morphogenesis. *Developmental Cell* **7**: 917-923.
- Henthorn, P., Kiledjian, M. and Kadesch, T. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. *Science* **247**: 467-470.
- Heszele, M.F. and Price, S.R. 2004. Insulin-like growth factor I: the yin and yang of muscle atrophy. *Endocrinology* **145**: 4803-4805.
- Hill, J., Davies, M.V., Pearson, M.V., Wang, J.H., Hewick, R.M., Wolfman, N.M. and Qiu, Y. 2002. The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. *The Journal of Biological Chemistry* **277**: 40735-40741.
- Hill, J.J., Qui, Y., Hewick, R.M. and Wolfman, N.M. 2003. Regulation of Myostatin in vivo by growth and differentiation factor-associated serum protein-1: A novel protein with protease inhibitor and follistatin domains. *Molecular Endocrinology* **17**: 1144-1154.
- Himits, Y., Osborn, D.P., Carvajal, J.J., Rigby, P.W. and Hughes, S.M. 2007. Mrf4 (myf6) is dynamically expressed in differentiated zebrafish skeletal muscle. *Gene Expression Patterns* **7**: 738-745.
- Hochachka, P.W. and Somero, G.N. 2002. *Biochemical adaptation: mechanism and process in physiological evolution*. Oxford. Oxford University Press.
- Hoegg, S., Brinkmann, H., Taylor, J.S. and Meyer, A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *Journal of Molecular Evolution* **59**: 190-203.
- Holder, M. and Lewis, P.O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics* **4**: 275-284.
- Holland, P.W., Garcia-Fernández, J., Williams, N.A. and Sidow, A. 1994. Gene duplications and the origins of vertebrate development. *Development* 1994, Suppl: pp 125-133.

- Hollway, G.E., Bryson-Richardson, R.J., Berger, S., Cole, N.J., Hall, T.E. and Currie, P.D. 2007. Whole-somite rotation generates muscle progenitor cell compartments in the developing zebrafish embryo. *Developmental Cell* **12**: 207-219.
- Hu, J.S., Olson, E.N. and Kingston, R.E. 1992. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Molecular and Cellular Biology* **12**: 1031-1042.
- Hu, Z., Potthoff, B., Hollenberg, C.P. and Ramezani-Rad, M. 2006. Mdy2, a ubiquitin-like (UBL)-domain protein, is required for efficient mating in *Saccharomyces cerevisiae*. *Journal of Cell Science* **119**: 326-338.
- Hudson, R.C.L. 1969. Polyneuronal innervation of the fast fibres of the marine teleost *Cottus scorpius* L. *The Journal of Experimental Biology* **50**: 47-67.
- Hudson, R.C.L. 1973. On the functions of the white muscles in teleosts at intermediate swimming speeds. *The Journal of Experimental Biology* **58**: 509-522.
- Hughes, A.L. 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *Journal of Molecular Evolution* **48**: 565-576.
- Hughes, S.M., Taylor, J.M., Tapscott, S.J., Gurley, C.M., Carter, W.J. and Peterson, C.A. 1993. Selective accumulation of MyoD and myogenin mRNAs in fast and slow adult skeletal muscle is controlled by innervation and hormones. *Development* **118**: 1137-1147.
- Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. and O'Shea, E.K. 2003. Global analysis of protein localization in budding yeast. *Nature* **425**: 686-691.
- Hurling, R., Rodell, J.B. and Hunt, H.D. 1996. Fibre diameter and fish texture. *Journal of Texture Studies* **27**: 679-685.
- Iezzi, S., Di Padova, M., Serra, C., Caretti, G., Simone, C., Maklan, E., Minetti, G., Zhao, P., Hoffman, E.P., Puri, P.L. and Sartorelli, V. 2004. Deacetylase inhibitors increase muscle cell size

by promoting myoblast recruitment and fusion through induction of follistatin. *Developmental Cell* **6**: 673-684.

Ingham, H.W. and Kim, H.R. 2005. Hedgehog signalling and the specification of muscle cell identity in the Zebrafish embryo. *Experimental Cell Research* **306**: 336-342.

Inoue, J.G., Miya, M., Venkatesh, B. and Nishida, M. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* **349**: 227-235.

Ishibashi, J., Perry, R.L., Asakura, A. and Rudnicki, M.A. 2005. MyoD induces myogenic differentiation through cooperation of its NH<sub>2</sub>- and COOH-terminal regions. *The Journal of Cell Biology* **171**: 471-482.

Ito, M., Ishikawa, M., Suzuki, S., Takamatsu, N. and Shiba, T. 1995. A rainbow trout SRY-type gene expressed in pituitary glands. *FEBS Letters* **370**: 37-40.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences, USA* **98**: 4569-4574.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., 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., Biémont, 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., Guigó, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quétier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J. and Roest Crolius, H. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**: 946-957.

Jalabert, B. 2005. Particularities of reproduction and oogenesis in teleost fish compared to mammals. *Reproduction Nutrition Development* **45**: 261-279.

Jobling, M. 1994. *Fish Bioenergetics*. London. Chapman and Hall.

Johnston, I.A. 1981. Structure and function of fish muscles. *Symposia of the Zoological Society of London* **48**: 71-113.

Johnston, I.A. 1993. Temperature influences muscle differentiation and the relative timing of organogenesis in herring (*Clupea harengus*) larvae. *Marine Biology* **116**: 363-379.

Johnston, I.A. 1999. Muscle development and growth: potential implications for flesh quality in fish. *Aquaculture* **177**: 99-115.

Johnston, I.A. 2001a. Implications of muscle growth patterns for the colour and texture of fish flesh. In *Farmed Fish Quality*. (Kestin, S.C. and Warriss, P.D., eds) pp 13-30. Blackwell Science, Oxford.

Johnston, I.A. 2001b. Genetic and environmental determinants of muscle growth patterns. In *Fish Physiology Series Volume 18: Muscle Development and Growth* (Johnston, I.A., ed) pp 141-186. Academic Press, San Diego.

Johnston, I.A. 2006. Environment and plasticity of myogenesis in teleost fish. *The Journal of Experimental Biology* **209**: 2249-2264.

Johnston, I.A. and Moon, T.W. 1981. Fine structure and metabolism of multiply innervated fast muscle fibres in teleost fish. *Cell and Tissue research*. **219**: 93-109.

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

Johnston, I.A. and Hall, T.E. 2004. Mechanisms of muscle development and responses to temperature change in fish larvae. *American Fisheries Society Symposium* **40**: 85-116.

Johnston, I.A., Davison, W. and Goldspink, G. 1977. Energy metabolism of carp swimming muscles. *Journal of Comparative Physiology* **114**: 203-216.

Johnston, I.A., Bernard, L.B. and Maloiy, G.M.O. 1983. Aquatic and aerial respiration rates, muscle capillary supply and mitochondrial volume density in the air-breathing catfish (*Clarias mossambicus*), acclimated to either air saturated or hypoxic water. *The Journal of Experimental Biology* **105**: 317-338.

Johnston, I.A., Vieira, V.L.A. and Abercromby, M. 1995. Temperature and myogenesis in embryos of the Atlantic herring (*Clupea harengus*). *The Journal of Experimental Biology* **198**: 1389-1403

Johnston, I.A., Cole, N.J. Vieira, V.L.A. and Davidson, I. 1997. Temperature and developmental plasticity of muscle phenotype in herring larvae. *The Journal of Experimental Biology* **200**: 849-868.

Johnston, I.A., Strugnell, G., McCracken, M.L. and Johnstone, R.R. 1999. Muscle growth and development in normal-sex-ratio and all-female diploid and triploid Atlantic salmon. *The Journal of Experimental Biology* **202**: 1991-2016.

Johnston, I.A., McLay, H.A., Abercromby, M. and Robins, D. 2000a. Phenotypic plasticity of early myogenesis and satellite cell numbers in Atlantic salmon spanning in upland and lowland tributaries of a river system. *The Journal of Experimental Biology* **203**: 2539-2552.

Johnston, I.A., Alderson, D., Sandham, C., Dingwall, A., Mitchell, D., Selkirk, C., Nickell, B., Baker, R., Roberston, B., Whyte, A. and Springate, J. 2000b. Muscle fibre density in relation to the colour and texture of smoked Atlantic salmon (*Salmo salar* L.). *Aquaculture* **189**: 335-349.

Johnston, I.A., Manthri, S., Robertson, R. Campbell, P., Mitchell, D. and Alderson, R. 2000c. Family and population differences in muscle fibre recruitment in farmed Atlantic salmon (*Salmo salar*). *Basic Applied Myology* **10**: 291-296.

Johnston, I.A., Vieira, V.L.A. and Temple, G.K. 2001. Functional consequences and population differences in the response of neuromuscular development to embryonic temperature regime in Atlantic herring (*Clupea harengus* L.). *Marine Ecology Progress Series* **213**: 285-300.

Johnston, I.A., Manthri, S., Alderson, B., Campbell, P., Mitchell, D., Whyte, D., Dingwall, A., Nickell, D., Selkirk, C. and Robertson, B. 2002. Effects of dietary protein level on muscle cellularity and flesh quality in Atlantic salmon with particular reference to gaping. *Aquaculture* **210**: 259-283.

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

Johnston, I.A., Hall, T.E. and Fernández, D.A. 2003b. Genes regulating the growth of myotomal muscle in teleost fish. In *Aquatic Genomics: Steps Toward a Great Future*. (Shimizu, N., Aoki, T., Hirono, I. and Takashima, F., eds) pp153-166. Springer-Verlag, Tokyo.

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

Johnston, I.A., Abercromby, M., Vieira, V.L., Sigursteindóttir, R.J., Kristjánsson, B.J., Sibthorpe, D. and Skúlason, S. 2004. Rapid evolution of muscle fibre number in post-glacial populations of Arctic charr *Salvelinus alpinus*. *The Journal of Experimental Biology* **207**: 4343-4360.

Johnston, I.A., Abercromby, M. and Andersen, Ø. 2005. Loss of muscle fibres in a landlocked dwarf salmon population. *Biology Letters* **1**: 419-422.

Kablar, B., Krastel, K., Ying, C., Asakura, A., Tapscott, S.J. and Rudnicki, M.A. 1997. MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. *Development* **124**: 4729-4738.

- Kablar, B., Krastel, K., Tajbakhsh, S. and Rudnicki, M.A. 2003. Myf5 and MyoD activation define independent myogenic compartments during embryonic development. *Developmental Biology* **258**: 307-318.
- Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P. and Ahringer, J. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231-237.
- Kambadur, R., Sharma, M., Smith, T.P. and Bass, J.J. 1997. Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Research* **7**: 910-916.
- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin-I, T., Takeda, H., Morishita, S. and Kohara, Y. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**: 714-719.
- Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S. 2004. Mrf4 determines skeletal muscle identity in Myf5:MyoD double-mutant mice. *Nature* **431**: 466-467.
- Kassar-Duchossoy, L., Giacone, E., Gayraud-Morel, B., Jory, A., Gomès, D. and Tajbakhsh, S. 2005. Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes and Development* **19**: 1426-1431.
- Kaushal, S., Schneider, J.W., Nadal-Ginard, B. and Mahdavi, V. 1994. Activation of the myogenic lineage by MEF2A, a factor that induces and cooperates with MyoD. *Science* **266**: 1236-1240.
- 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 and Development* **7**: 390-400.

- Kim, H.K., Lee, Y.S., Sivaprasad, U., Malhotra, A., Dutta, A. 2006. Muscle-specific microRNA miR-206 promotes muscle differentiation. *The Journal of Cell Biology* **174**: 677-687.
- Kim, J., Wang, Z. Heymsfield, S.B., Baumgartner, R.N. and Gallagher, D. 2002. Total-body skeletal muscle mass: estimation by a new dual-energy X-ray absorptiometry method. *The American Journal of Clinical Nutrition* **76**: 378-383.
- Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E., Seidman, J.G. 2007. Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**: 1481-1484.
- 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**: 253-310.
- Klüver, N., Kondo, M., Herpin, A., Mitani, H. and Schartl M. 2005. Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. *Development Genes and Evolution* **215**: 297-305.
- Knapp, J.R., Davie, J.K., Myer, A., Meadows, E., Olson, E.N. and Klein WH. 2006. Loss of myogenin in postnatal life leads to normal skeletal muscle but reduced body size. *Development* **133**: 601-10
- Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* **5**: 288-298.
- Koller, B.H. and Smithies, O. 1992. Altering genes in animals by gene targeting. *Annual Reviews in Immunology* **10**: 705-730.
- Koohmaraie, M.1996. Biochemical factors regulating the toughening and tenderization processes of meat. *Meat Science* **43**: S193-S201.
- Korber, B. 2000. HIV Signature and Sequence Variation Analysis. In *Computational Analysis of HIV Molecular Sequences* (Rodrigo A.G. and Learn, G.H., eds) pp 55-72. Kluwer Academic Publishers, Dordrecht.

- Kottelat, M., Britz, R., Tan, H.H. and Witte, K.E. 2006. *Paedocypris*, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. *Proceedings of the Royal Society* **273B**: 895–899.
- Koumans, J.T.M., Akster, H.A., Booms, G.H.R. Lemmens, C.J.J. and Osse, J.W.M. 1991. Numbers of myosatellite cells in white axial muscle of growing fish: *Cyprinus carpio* L. (teleostei). *American Journal of Anatomy* **192**: 418-424.
- Koumans, J.T.M. and Akster, H.A. 1995. Review: Myogenic cells in development and growth of fish. *Comparative Biochemistry and Physiology* **110B**: 3-20.
- Kumar, S., Tamura, K. and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetic analysis and sequence alignment. *Briefings in Bioinformatics* **5**: 150-163.
- Kuang, S., Kuroda, K., Le Grand, F. and Rudnicki, M.A. 2007. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**: 999-1010.
- Lai, K.M., Gonzalez, M., Poueymirou, W.T., Kline, W.O., Na, E., Zlotchenko, E., Stitt, T.N., Economides, A.N., Yancopoulos, G.D. and Glass, D.J. 2004. Conditional activation of akt in adult skeletal muscle induces rapid hypertrophy. *Molecular and Cellular Biology* **24**: 9295-9304.
- Lang, D., Powell, S.K., Plummer, R.S., Young, K.P. and Ruggeri, B.A. 2007. PAX genes: roles in development, pathophysiology, and cancer. *Biochemical Pharmacology* **73**: 1-14.
- Langley, B., Thomas, M., Bishop, A., Sharma, M., Gilmour, S. and Kambadur, R. 2002. Myostatin inhibits myoblast differentiation by down-regulating MyoD expression. *The Journal of Biological Chemistry* **277**: 49831-49840.
- Le Comber, S.C. and Smith, C. 2004. Polyploidy in fishes: patterns and processes. *Biological Journal of the Linnean Society* **82**: 431-442.
- Lee, H-T. 2007. Modelling muscle fibre recruitment in the model teleost *Danio rerio*. PhD Thesis. University of St Andrews. *In preparation*.

- Lee, S.J. 2004. Regulation of muscle mass by myostatin. *Annual Reviews in Cell and Developmental Biology* **20**: 61-86.
- Lee, S.J. 2007a. Sprinting without myostatin: a genetic determinant of athletic prowess. *Trends in Genetics* **23**: 475-477.
- Lee, S.J. 2007b. Quadrupling muscle mass in mice by targeting TGF- $\beta$  signalling pathways. *PLoS ONE* **8**: e789.
- Lee, S.J. and McPherron, A.C. 2001. Regulation of myostatin activity and muscle growth. *Proceedings of the National Academy of Sciences, USA* **98**: 9306-9311.
- Lewis, K.E., Currie, P.D., Schauerte, H., Haffter, P. and Ingham, P.W. 1999. Control of muscle cell-type specification in the zebrafish embryo by hedgehog signalling. *Development Biology* **216**: 469-480.
- Li, Y., Li, J., Zhu, J., Sun, B., Branca, M., Tang, Y., Foster, W., Xiao, X. and Huard J. 2007. Decorin gene transfer promotes muscle cell differentiation and muscle regeneration. *Molecular Therapy* **15**: 1616-1622.
- 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 Cellular Biology* **11**: 267-280.
- Liu, J.P., Baker, J., Perkins, A.S., Robertson, E.J. and Efstratiadis, A. 1993. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell* **75**: 59-72.
- Liu, N., Williams, A.H., Kim, Y., McAnally, J., Bezprozvannaya, S., Sutherland, L.B., Richardson, J.A., Bassel-Duby, R. and Olson, E.N. 2007. An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133. *Proceedings of the National Academy of Sciences, USA* **104**: 20844-20849.
- Lynch, M. and Force, A. 2000. Gene duplication and the origin of interspecific genomic incompatibility. *American Naturalist* **156**: 590-605.

Maccatrozzo, L., Bargelloni, L., Cardazzo, B., Rizzo, G. and Patarnello, T. 2001. A novel second myostatin gene is present in teleost fish. *FEBS Letters* **509**: 36-40.

Mackenzie, M.G. 2006. Characterisation of genes regulating muscle development and growth in two model puffer fish species (*Takifugu rubripes* and *Tetraodon nigroviridis*). PhD thesis. University of St Andrews.

Macqueen, D.J. and Johnston, I.A. 2006. A novel salmonid myoD gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization. *FEBS Letters* **580**: 4996-5002.

Macqueen, D.J., Robb, D. and Johnston, I.A. 2007. Temperature influences the co-ordinated expression of myogenic regulatory factors during embryonic myogenesis in Atlantic salmon (*Salmo salar* L.). *The Journal of Experimental Biology* **210**: 2781-2794.

Maissey, J. G. 1996. *Discovering fossil fishes*. Henry Holt and Company, New York.

Makalowski, W. 2001. Are We Polyploids? A Brief History of One Hypothesis. *Genome Research* **5**: 667-670.

Mal, A. and Harter, M.L. 2003. MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. *Proceedings of the National Academy of Sciences, USA* **100**: 1735-1739.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. and Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

- Martin, A. 2001. Is Tetralogy True? Lack of Support for the One-to-Four Rule *Molecular Biology and Evolution* **18**: 89-93.
- 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**: 360-363.
- Mauro, A. 1961. Satellite cell of skeletal muscle fibres. *The Journal of Biophysical and Biochemical Cytology* **9**: 493-498.
- Maves, L., Waskiewicz, A.J., Paul, B., Cao, Y., Tyler, A., Moens, C.B. and Tapscott, S.J. 2007. Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* **134**: 3371-3382.
- McKinnon, J.S. and Rundle, H.D. 2002. Speciation in nature: the threespine stickleback model systems. *Trends in Ecology and Evolution* **17**: 480-488.
- McPherron, A.C., Lawler, A.M., and Lee, S.J. 1997. Regulation of skeletal muscle mass in mice by a new TGF-beta 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. *Proceedings of the National Academy of Sciences of the USA* **94**: 12457-12461.
- Méndez, E., Planas, J.V., Castillo, J., Navarro, I. and Gutiérrez, J. 2001. Identification of a type II insulin-like growth factor receptor in fish embryos. *Endocrinology* **142**: 1090-1097.
- Meyer, A and Schartl, M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology* **11**: 699-704.
- Meyer, A. and Van de Peer, Y.P. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**: 937-945.
- Michel, R.N., Dunn, S.E. and Chin, E.R. 2004. Calcineurin and skeletal muscle growth. *Proceedings of the Nutrition Society* **63**: 341-349.

- Michie, I. 2001. Causes of downgrading in the salmon farming industry. In *Farmed Fish Quality*. (Kestin, S.C. and Warriss, P.D., eds) pp 129-136. Blackwell Science, Oxford.
- Minetti, G.C., Colussi, C., Adami, R., Serra, C., Mozzetta, C., Parente, V., Fortuni, S., Straino, S., Sampaolesi, M., Di Padova, M., Illi, B., Gallinari, P., Steinkühler, C., Capogrossi, M.C., Sartorelli, V., Bottinelli, R., Gaetano, C. and Puri, P.L. 2006. Functional and morphological recovery of dystrophic muscles in mice treated with deacetylase inhibitors. *Nature Medicine* **12**: 1147-1150.
- Miura, T., Kishioka, Y., Wakamatsu, J., Hattori, A., Hennebry, A., Berry, C.J., Sharma, M., Kambadur, R. and Nishimura, T. 2006. Decorin binds myostatin and modulates its activity to muscle cells. *Biochemical and Biophysical research Communications* **340**: 675-680.
- Moghadam, H.K., Ferguson, M.M., and Danzmann, R.G. 2005. Evolution of Hox clusters in Salmonidae: A comparative analysis between Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). *Journal of Molecular Evolution* **61**: 636-649.
- Moghadam, H.K., Ferguson, M.M., Rexroad, C.E. 3rd, Coulibaly, I. and Danzmann, R.G. 2007. Genomic organization of the IGF1, IGF2, MYF5, MYF6 and GRF/PACAP genes across Salmoninae genera. *Animal Genetics* **38**: 527-532.
- 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**: 1125-1136.
- Morin-Kensicki, E.M. and Eisen, J.S. 1997. Sclerotome development and peripheral nervous system segmentation in embryonic zebrafish. *Development* **124**: 159-167.
- Murre, C., McCaw, P.S. and Baltimore, D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**: 777-783.
- Musarò, A., McCullagh, K.J., Naya, F.J., Olson, E.N. and Rosenthal N. 1999. IGF-1 induces skeletal myocyte hypertrophy through calcineurin in association with GATA-2 and NF-ATc1. *Nature* **400**: 581-585.

- Musarò, A., McCullagh, K., Paul, A., Houghton, L., Dobrowolny, G., Molinaro, M., Barton, E.R., Sweeney, H.L. and Rosenthal, N. 2001. Localized Igf-1 transgene expression sustains hypertrophy and regeneration in senescent skeletal muscle. *Nature Genetics* **27**: 195-200.
- Myer, A., Olson, E.N. and Klein, W.H. 2001. MyoD cannot compensate for the absence of myogenin during skeletal muscle differentiation in murine embryonic stem cells. *Developmental Biology* **229**: 340-350.
- 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**: 532-535.
- Nakae, J., Kido, Y. and Accili, D. 2001. Distinct and overlapping functions of insulin and IGF-I receptors. *Endocrine Reviews* **22**: 818-835.
- Nakamura, T., Takio, K., Eto, Y., Shibai, H., Titani, K. and Sugino, H. 1990 Activin-binding protein from rat ovary is follistatin. *Science* **247**: 836-838.
- Nakatani, M., Takehara, Y., Sugino, H., Matsumoto, M., Hashimoto, O., Hasegawa, Y., Murakami, T., Uezumi, A., Takeda, S., Noji, S., Sunada, Y. and Tsuchida, K. 2007. Transgenic expression of a myostatin inhibitor derived from follistatin increases skeletal muscle mass and ameliorates dystrophic pathology in mdx mice. *The FASEB Journal* doi: 10.1096/fj.07-8673com.
- Nathanailides, C., Lopez-Albors, O. and Stickland, N.C. 1995. Influence of pre-hatch temperature on the development of muscle cellularity in post-hatch Atlantic salmon (*Salmo salar* L.). *Canadian Journal of Fisheries and Aquatic Sciences* **52**: 675-680.
- Naya, F.J., Mercer, B., Shelton, J., Richardson, J.A., Williams, R.S. and Olson E.N. 2000. Stimulation of slow skeletal muscle fiber gene expression by calcineurin in vivo. *The Journal of Biological Chemistry* **275**: 4545-4548.
- Naylor, R.L., Goldberg, R.J., Primavera, J.H., Kautsky, N., Beveridge, M.C., Clay, J., Folke, C., Lubchenco, J., Mooney, H. and Troell, M. 2000. Effect of aquaculture on world fish supplies. *Nature* **405**: 1017-1024.

- Naylor, R.L., Williams, S.L. and Strong, D.R. 2001. Ecology. Aquaculture--a gateway for exotic species. *Science* **294**: 1655-1656.
- Naylor, R., Hindar, K., Fleming, I.A., Goldberg, R., Williams, S., Volpe, J., Whoriskey, F., Eagle, J., Kelso, D. and Mangel, M. 2005. Fugitive Salmon: Assessing the Risks of Escaped Fish from Net-Pen Aquaculture. *Bioscience* **55**: 427-437.
- Nelson, J. S. 2006. *Fishes of the world*. John Wiley and Sons, Inc. New York.
- Neyt, C., Jagla, K., Thisse, C., Thisse, B., Haines, L. and Currie, P.D. 2000. Evolutionary origins of vertebrate appendicular muscle. *Nature* **408**: 82-86.
- Nickell, D.C. and Bromage N.R. 1998. The effect of timing and duration of feeding astaxanthin on the development and variation of fillet colour and efficiency of pigmentation in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* **169**: 233-246.
- Nornes, S., Mikkola, I., Krauss, S., Delghandi, M., Perander, M. and Johansen, T. 1996. Zebrafish Pax9 encodes two proteins with distinct C-terminal transactivating domains of different potency negatively regulated by adjacent N-terminal sequences. *The Journal of Biological Chemistry* **271**: 26914-26923.
- Notredame, C., Higgins, D. and Heringa, J. 2000. T-Coffee: A novel method for multiple sequence alignments. *Journal of Molecular Biology* **302**: 205-217.
- Nowak, M.A., Boerlijst, M.C., Cooke, J. and Smith, J.M. 1997. Evolution of genetic redundancy. *Nature* **388**, 167-171.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.
- Olson, E.N., Perry, M. and Sculz, R.A. 1995. Regulation of muscle differentiation by the Mef2 family of MADS box transcription factors. *Developmental Biology* **172**: 2-14.
- Østbye, T.K., Galloway, T.F., Nielson, C., Gabested, I., Bardal, T. and Anderson, Ø. 2001. The two myostatin genes of Atlantic salmon (*Salmo salar*) are expressed in a variety of tissues. *European Journal of Biochemistry* **268**: 5249-5257.

- Otto, S.P. and Whitton, J. 2000 Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401-437.
- Parsons, S.A., Wilkins, B.J., Bueno, O.F. and Molkentin, J.D. 2003. Altered skeletal muscle phenotypes in calcineurin Aalpha and Abeta gene-targeted mice. *Molecular and Cellular Biology* **23**: 4331-4343.
- Parsons, S.A., Millay, D.P., Wilkins, B.J., Bueno, O.F., Tsika, G.L., Neilson, J.R., Liberatore, C.M., Yutzey, K.E., Crabtree, G.R., Tsika, R.W. and Molkentin, J.D. 2004. Genetic loss of calcineurin blocks mechanical overload-induced skeletal muscle fiber type switching but not hypertrophy. *The Journal of Biological Chemistry* **279**: 26192-26200.
- Patel, K. 1998. Follistatin. *The International Journal of Biochemistry and Cell Biology* **30**: 1087-1093.
- Pelegri, F. 2003. Maternal factors in zebrafish development. *Developmental Dynamics* **228**: 535-554.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e45.
- Pfaffl, M.W., Horgan, G.W. and Dempfle, L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* **30** e36.
- Phillips, D.J. and de Krestor, D.M. 1998. Follistatin: a multifunctional regulatory protein. *Frontiers in Neuroendocrinology* **19**: 287-322.
- Phillips, R.B. and Rab, P. 2001. Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews* **76**: 1-25.
- Pinheiro, J.C. and Bates, D.M. 2000. *Mixed-effects models in S and S-plus*. Springer. New York.

Pisconti, A., Brunelli, S., Di Padova, M., De Palma, C., Deponti, D., Baesso, S., Sartorelli, V., Cossu, G. and Clementi, E. 2006. Follistatin induction by nitric oxide through cyclic GMP: a tightly regulated signaling pathway that controls myoblast fusion. *The Journal of Cell Biology* **172**: 233-244.

Podrabsky, J.E. and Somero, G.N. 2004. Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*. *The Journal of Experimental Biology* **207**: 2237-2254.

Postlethwait, J., Amores, A., Cresko, W., Singer, A. and Yan, Y.L. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends in Genetics* **20**: 481-490.

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**: 62-66.

Rao, P.K., Kumar, R.M., Farkhondeh, M., Baskerville, S. and Lodish, H.F. 2006. Myogenic factors that regulate expression of muscle-specific microRNAs. *Proceedings of the National Academy of Sciences, USA* **103**: 8721-8276.

Rastogi, S. and Liberles, D.A. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology* **5**:28.

R Development Core Team. 2007. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Reinartz, J., Bruyns, E., Lin, J.Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M. and Woychik, R. 2002. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics and Proteomics* **1**: 95-104.

Relaix, F., Rocancourt, D., Mansouri, A and Buckingham, M. 2005. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**: 948-953.

- Rescan, P.Y. 2001. Regulation and function of myogenic regulatory factors in lower vertebrates. *Comparative Biochemistry and Physiology* **130B**: 1-12.
- Rescan, P.Y. 2005. Muscle growth patterns and regulation during fish ontogeny. *General and Comparative Endocrinology* **142**: 111-116.
- 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* **113B**: 711-715.
- Rhodes, S.J. and Konieczny, S.F. 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. *Genes and Development* **3**: 2050-2061.
- Ríos, R., Carneiro, I., Arce, V.M. and Devesa, J. 2002. Myostatin is an inhibitor of myogenic differentiation. *American Journal of Physiology* **282**: C993-C999.
- Ririe, K.M., Rasmussen, R.P. and Wittwer C.T. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry* **245**: 154-160.
- Roff, D.A. 1992. *The evolution of life histories*. London: Chapman and Hall.
- Rome, L.C., Loughna, P.T. and Goldspink, G. 1984. Muscle fiber activity in carp as a function of swimming speed and muscle temperature. *American Journal of Physiology* **247**: R272-R279.
- Rommel, C., Bodine, S.C., Clarke, B.A., Rossman, R., Nunez, L., Stitt, T.N., Yancopoulos, G.D. and Glass, D.J. 2001. Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. *Nature Cell Biology* **3**: 1009-1013.
- Ronquist, F. and Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.
- Rosenberg, M.I., Georges, S.A., Asawachaicharn, A., Analau, E. and Tapscott S.J. 2006. MyoD inhibits Fstl1 and Utrn expression by inducing transcription of miR-206. *The Journal of Cell Biology* **175**: 77-85.

- Rowe, R.W. and Goldspink, G. 1969. Muscle fibre growth in five different muscles in both sexes of mice. *Journal of Anatomy* **104**: 519-530.
- Rowlerson, A. and Veggetti, A. 2001. Cellular mechanisms of post-embryonic muscle growth in aquaculture species. In *Fish Physiology Volume 18: Muscle Development and Growth* (Johnston, I.A. ed) pp103-140. Academic Press, San Diego.
- Rowlerson, A., Radaelli, G., Mascarello, F. and Veggetti, A. 1997. Regeneration of skeletal muscle in two teleost fish: *Sparus aurata* and *Brachydanio rerio*. *Cell and Tissue Research* **289**: 311-322.
- Roy, S., Wolff, C. and Ingham, P.W. 2001. The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo. *Genes and Development* **15**: 1563-1576.
- Rudnicki, M. and Jaenisch, R. 1995. The MyoD family of transcription factors and skeletal myogenesis. *Bioessays* **17**: 203-209.
- Rudnicki, M.A., Braun, T., Hinuma, S. and Jaenisch, R. 1992. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* **71**: 383-390.
- Rudnicki, M.A., Schegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H. and Jaenisch, R. 1993. MyoD or Myf5 is required for the formation of skeletal muscle. *Cell* **75**: 1351-1359.
- Salem, M., Yao, J., Rexroad, C.E., Kenney, P.B., Semmens, K., Killefer, J. and Nath, J. 2005. Characterization of calpastatin gene in fish: its potential role in muscle growth and fillet quality. *Comparative Biochemistry and Physiology* **141B**: 488-497.
- Scales, J.B., Olson, E.N. and Perry, M. 1990. Two distinct *Xenopus* genes with homology to MyoD1 are expressed before somite formation in early embryogenesis. *Molecular and Cellular Biology* **10**: 1516-1524.

- 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**: 167-170.
- Schmidt, K., Glaser, G., Wernig, M. and Rosorius, O. 2003. Sox8 is a specific marker for muscle satellite cells and inhibits myogenesis. *The Journal of Biological Chemistry* **278**: 29769-29775.
- Schuelke, M., Wagner, K.R., Stolz, L.E., Hübner, C., Riebel, T., Kömen, W., Braun, T., Tobin, J.F. and Lee, S.J. 2004. Myostatin mutation associated with gross muscle hypertrophy in a child. *The New England Journal of Medicine* **350**: 2682-2688.
- Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S. and Krogan, N.J. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**: 507-519.
- Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., Rudnicki, M.A. 2000. Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**: 777-786.
- Semsarian, C., Wu, M.J., Ju, Y.K., Marciniak, T., Yeoh, T., Allen, D.G., Harvey, R.P. and Graham, R.M. 1999. Skeletal muscle hypertrophy is mediated by a Ca<sup>2+</sup>-dependent calcineurin signalling pathway. *Nature* **400**: 576-581.
- Seo, H.C., Saetre, B.O., Håvik, B., Ellingsen, S. and Fjose A. 1998. The zebrafish Pax3 and Pax7 homologues are highly conserved, encode multiple isoforms and show dynamic segment-like expression in the developing brain. *Mechanisms of Development* **70**: 49-63.
- Serrano, A.L., Murgia, M., Pallafacchina, G., Calabria, E., Coniglio, P., Lømo, T. and Schiaffino, S. 2001. Calcineurin controls nerve activity-dependent specification of slow skeletal muscle fibers but not muscle growth. *Proceedings of the National Academy of Sciences of the USA* **98**: 13108-13113.
- Service, R.F. 2006. Gene sequencing. The race for the \$1000 genome. *Science* **311**: 1544-1546.

- Shanmugalingam, S. and Wilson, S.W. 1998. Isolation, expression and regulation of a zebrafish paraxis homologue. *Mechanisms of Development* **78**: 85-89.
- Shavlakadze, T., Winn, N., Rosenthal, N. and Grounds, M.D. 2005. Reconciling data from transgenic mice that overexpress IGF-I specifically in skeletal muscle. *Growth Hormone and IGF Research* **15**: 4-18.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728-1732.
- Shklover, J., Etzioni, S., Weisman-Shomer, P., Yafe, A., Bengal, E. and Fry, M. 2007. MyoD uses overlapping but distinct elements to bind E-box and tetraplex structures of regulatory sequences of muscle-specific genes. *Nucleic Acids Research* **35**: 7087-7095.
- Sidis, Y., Schneyer, A.L., Sluss, P.M., Johnson, L.N. and Keutmann, H.T. 2001 Follistatin: essential role for the N-terminal domain in activin binding and neutralization. *The Journal of Biological Chemistry* **276**: 17718-17726.
- Sidow, A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Current Opinion in Genetics and Development* **6**: 715-722.
- Silverman, D. 1986. *Density estimate for statistics and data analysis*. New York. Chapman and Hall.
- Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J. and Plasterk, R.H. 2003. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PloS Biology* **1**: E12.
- Spinner, D.S., Liu, S., Wang, S.W. and Schmidt, J. 2002. Interaction of the myogenic determination factor myogenin with E12 and a DNA target: mechanism and kinetics. *Journal of Molecular Biology* **317**: 431-445.
- Spring, J. 1997. Vertebrate evolution by interspecific hybridization-are we polyploid? *FEBS Letters* **400**: 2-8.

Steinbacher, P., Haslett, J.R., Six, M., Gollmann, H.P., Sanger, A.M. and Stoiber, W. 2006. Phases of myogenic cell activation and possible role of dermomyotome cells in teleost muscle formation. *Developmental Dynamics* **235**: 3132-3143.

Steinbacher, P., Haslett, J.R., Obermayer, A., Marschallinger, J., Bauer, H.C., Sanger, A.M. and Stoiber, W. 2007. MyoD and Myogenin expression during myogenic phases in brown trout: a precocious onset of mosaic hyperplasia is a prerequisite for fast somatic growth. *Developmental Dynamics* **236**: 1106-1114.

Stellabotte, F., Dobbs-McAuliffe, B., Fernandez, D.A., Feng, X. and Devoto, S.H. 2007. Dynamic somite cell rearrangements lead to distinct waves of myotome growth. *Development* **134**: 1253-1257.

Stellabotte, F. and Devoto, S.H. 2007. The teleost dermomyotome. *Developmental Dynamics* **236**: 2432-2443.

Stephens, S.G. 1951. Possible significance of duplication in evolution. *Advances in Genetics*. **4**: 247-265.

Stickland, N.C., White, R.N., Mescall, P.E., Crook, A.R. and Thorpe, J.E. 1988. The effect of temperature on myogenesis in embryonic development of the Atlantic salmon (*Salmo salar* L.). *Anatomy and Embryology* **178**: 253-257.

Stickney, H.L., Barresi, M.J.F. and Devoto, S.H. 2000. Somite development in zebrafish. *Developmental Dynamics* **219**: 287-303.

Stockard, C.R. 1921. Developmental rate and structural expression: an experimental study of twins, 'double monsters' and single abnormalities and the interaction among embryonic organs during their origin and development. *The American Journal of Anatomy* **28**: 115-216.

Stoiber, W., Haslett, J.R., Goldschmid, A. and Sanger, A.M. 1998. Patterns of superficial fibre formation in the European pearlfish (*Rutilus frisii meidingeri*) provide a general template for slow muscle development in teleost fish. *Anatomy and Embryology* **197**: 485-496.

- Sumariwalla, V. M. and Klein, W.H. 2001. Similar myogenic functions for myogenin and MRF4 but not MyoD in differentiated murine embryonic stem cells. *Genesis* **30**: 239-249.
- Suyama, M., Torrents, D. and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609-W612.
- Tan, X. and Du, S.J. 2002. Differential expression of two MyoD genes in fast and slow muscles of gilthead seabream (*Sparus aurata*). *Development Genes and Evolution* **212**: 207-217.
- 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**: 537-545.
- Tan, X., Zhang, Y., Zhang, P.J., Xu, P. and Xu, Y. 2006. Molecular structure and expression patterns of flounder (*Paralichthys olivaceus*) Myf-5, a myogenic regulatory factor. *Comparative Biochemistry and Physiology* **145B**: 204-213.
- Tapscott, S.J. 2005. The circuitry of a master switch: myod and the regulation of skeletal muscle gene transcription. *Development* **132**: 2685-2695.
- Taylor, J.S. and Raes, J. 2004. Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics* **38**: 615-643.
- Taylor, J.S., Van de Peer, Y. and Meyer, A. 2001. Genome duplication, divergent resolution and speciation. *Trends in Genetics* **17**: 299-301.
- Taylor, J.S., Braasch, I., Frickey, T., Meyer, A. and Van de Peer, Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research* **13**: 382-390.
- Temm-Grove, C.J., Wert, D., Thompson, V.F., Allen, R.E. and Goll, D.E. 1999. Microinjection of calpastatin inhibits fusion in myoblasts. *Experimental Cell research* **247**: 293-303.
- 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.). *The Journal of Experimental Biology* **204**: 3629-3637.

Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degraeve, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q. and Thisse, C. 2001. Expression of the zebrafish genome during embryogenesis, gene expression section (2001) (<http://zfin.org>).

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876-4882.

Thompson, T.B., Lerch, T.F., Cook, R.W., Woodruff, T.K. and Jardetzky, T.S. 2005. The structure of the follistatin: activin complex reveals antagonism of both type I and type II receptor binding. *Developmental Cell* **9**: 535-543.

Thorpe, J.E. 1988. Salmon migration. *Science Progress Oxford* **72**: 345–370.

Tidball, J.G. and Spencer, M.J. 2002. Expression of a calpastatin transgene slows muscle wasting and obviates changes in myosin isoform expression during murine muscle disuse. *The Journal of Physiology* **545**: 819-828.

Torrissen, O.J., Christiansen, R. Struksnaes, G. and Estermann, R. 1995. Astaxanthin deposition in the flesh of Atlantic salmon, *Salmo salar* L., in relation to dietary astaxanthin concentration and feeding period. *Aquaculture Nutrition* **1**: 77–84.

Usher, M.L., Stickland, N.C. and Thorpe, J.E. 1994. Muscle development in Atlantic salmon (*Salmo salar* L.) embryos and the effect of temperature on its cellularity. *Journal of Fish Biology* **44**: 953-964.

Van de Peer, Y. 2004. *Tetraodon* genome confirms *Takifugu* findings: most fish are ancient polyploids. *Genome Biology* **5**: 250.

Van de Peer, Y., Frickey, T., Taylor, J.S. and Meyer, A. 2002. Dealing with saturation at the amino acid level: A case study involving anciently duplicated zebrafish genes. *Gene* **295**: 205-211.

- Vandepoele, K., De Vos, W., Taylor, J.S., Meyer, A. and Van de Peer, Y. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proceedings of the National Academy of Sciences, USA* **101**: 1638-1643.
- van Raamsdonk, W., Pool, C.W. and te Kronnie, G. 1978. Differentiation of muscle fibre types in the teleost *Brachydanio rerio*. *Anatomy and Embryology* **153**: 137-155.
- Veggetti, A., Mascarello, F., Scapolo, P.A. and Rowlerson, A. 1990. Hyperplastic and hypertrophic growth of lateral muscle in *Dicentrarchus labrax* (L.). An ultrastructural and morphometric study. *Anatomy and Embryology* **182**: 1-10.
- Veggetti, A., Mascarello, F., Scapolo, P.A., Rowlerson, A. and Carnevali, C. 1992. Muscle growth and myosin isoform transitions during development of a small teleost fish, *Poecilia reticulata* (Peters) (Atheriniformes, Poeciliidae): a histochemical, immunohistochemical, ultrastructural and morphometric study. *Anatomy and Embryology* **187**: 353-361.
- Velculescu, V.E. and Kinzler, K.W. 2007. Gene expression analysis goes digital. *Nature Biotechnology* **25**: 878-880.
- .
- Videler, J. J. 1993. *Fish Swimming*. London, Chapman and Hall.
- Vieira, V.L.A. and Johnston, I.A. 1992. Influence of temperature on muscle fibre development in larvae of the herring (*Clupea harengus* L). *Marine Biology* **112**: 333-341.
- Volff, J.N. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* **94**: 280-294.
- Wagner, A. 1998. The fate of duplicated genes: loss or diversification? *BioEssays* **20**: 785-788.
- Wagner, A. 1999. Redundant gene functions and natural selection. *Journal of Evolutionary Biology* **12**: 1-16.
- Walsh, J.B., 1995. How often do duplicated genes evolve new functions? *Genetics* **110**: 345-364.

- Wardle, C.S., Videler, J.J. and Altringham, J.D. 1995. Tuning in to fish swimming waves: body form, swimming mode and muscle function. *The Journal of Experimental Biology* **198**: 1629-1636.
- Waterman, R.E. 1969. Development of the lateral musculature in the teleost, *Brachydanio rerio*: a fine structural study, *The American Journal of Anatomy* **125**: 457–494.
- Weatherly, A.H. and Gill, H.S. 1985. Dynamics of increase in muscle fibers in teleosts in relation to size and growth. *Experientia* **41**: 353-354.
- Weatherly, 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 an ultimate size. *Journal of Fish Biology* **33**: 851-859.
- Wegner, A. 1999. From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Research* **27**: 1409-1420.
- Weinberg, E.S., Allende, M.L., Kelly, C.S., Abdelhami, D.A., Murakami, T., Anderman, P., Doerre, O.G., Gunwald, D.J and Riggleman, B. 1996. Developmental regulation of zebrafish MyoD in normal sex ratio, no tail and spadetail embryos. *Development* **122**: 271-280.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. and Lassar, A. 1991a. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**: 761-766.
- Weintraub, H., Dwarki, V.J., Verma, I., Davis, R.L., Hollenberg, S., Snider, L., Lassar, A.B. and Tapscott, S.J. 1991b. Muscle-specific transcriptional activation by MyoD. *Genes and Development* **5**: 1377-1386.
- Wheelan, S.J., Church, D.M. and Ostell, J.M. 2001. Spidey: a tool for mRNA-to-genomic alignments. *Genome Research* **11**: 1952-1957.
- Wittbrodt, J., Shima, A. and Schartl, M. 2002. Medaka--a model organism from the far East. *Nature Reviews Genetics* **3**: 53-64.

- Wolfe, K.H. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2**: 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**: 1169-1181.
- Wood, A.W., Duan, C. and Bern, H.A. 2005. Insulin-like growth factor signaling in fish. *International Review of Cytology* **243**: 215-285.
- Wood, T.L. 1995. Gene-targeting and transgenic approaches to IGF and IGF binding protein function. *The American Journal of Physiology* **269**: E613-E622.
- Woods, I.G., Wilson, C., Friedlander, B., Chang, P., Reyes, D.K., Nix, R., Kelly, P.D., Chu, F., Postlethwait, J.H. and Talbot, W.S. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Research* **15**: 1307-1314.
- Xie, S.Q., Mason, P.S., Wilkes, D., Goldspink, G., Fauconnueau, B. and Stickland, N.C. 2001. Lower environmental temperature delays and prolongs myogenic regulatory factor expression and muscle differentiation in rainbow trout (*Oncorhynchus mykiss*) embryos. *Differentiation* **68**: 106-114.
- Xing, F., Tan, X., Zhang, P.J., Ma, J., Zhang, Y., Xu, P. and Xu, Y. 2007. Characterization of amphioxus GDF8/11 gene, an archetype of vertebrate MSTN and GDF11. *Development Genes and Evolution* **217**: 549-554.
- Xu, C., Wu, G., Zohar, Y. and Du, S.J. 2003. Analysis of myostatin gene structure, expression and function in zebrafish. *The Journal of Experimental Biology* **206**: 4067-4079.
- Xu, P., Tan, X., Zhang, Y., Zhang, P.J. and Xu, Y. 2007. Cloning and expression analysis of myogenin from flounder (*Paralichthys olivaceus*) and promoter analysis of muscle-specific expression. *Comparative Biochemistry and Physiology* **147B**: 135-145.
- Yafe, A., Etzioni, S., Weisman-Shomer, P. and Fry, M. 2005. Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Research* **33**: 2887-2900.

Yancopoulos, G.D. and Glass, D.J. 2002. Reply: Calcineurin and skeletal muscle growth. *Nature Cell Biology* **4**: E46 - E47.

Yang, Q., Bassel-Duby, R. and Williams, R.S. 1997. Transient expression of a winged-helix protein, MNF-beta during myogenesis. *Molecular and Cellular Biology* **17**: 5236-5243.

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

Zanotti, S., Saredi, S., Ruggieri, A., Fabbri, M., Blasevich, F., Romaggi, S., Morandi, L. and Mora, M. 2007. Altered extracellular matrix transcript expression and protein modulation in primary Duchenne muscular dystrophy myotubes. *Matrix Biology* **26**: 615-624.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**: 292-298.

Zhu, J., Li, Y., Shen, W., Qiao, C., Ambrosio, F., Lavasani, M., Nozaki, M., Branca, M.F. and Huard, J. 2007. Relationships between transforming growth factor-beta1, myostatin, and decorin: implications for skeletal muscle fibrosis. *The Journal of Biological Chemistry* **282**: 25852-25863.