

**Evolution and Characterization of the Fatty Acid-
Binding Proteins (fabps) in Atlantic salmon
(*Salmo salar*)**

by

Yuk Yin (Yvonne) Lai

B.Sc. (Hons.), Simon Fraser University, 2007

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the

Department of Molecular Biology and Biochemistry
Faculty of Science

© Yuk Yin (Yvonne) Lai 2012

SIMON FRASER UNIVERSITY

Spring 2012

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Yuk Yin (Yvonne) Lai
Degree: Doctor of Philosophy
Title of Thesis: *Evolution and Characterization of the Fatty Acid-Binding Proteins (fabps) in Atlantic salmon (Salmo salar)*

Examining Committee:

Chair: Dr. Nicholas Harden
Associate Professor, Department of Molecular Biology and Biochemistry

Dr. William Davidson
Senior Supervisor
Professor, Department of Molecular Biology and Biochemistry

Dr. Christopher Beh
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

Dr. Norbert Haunerland
Supervisor
Professor, Department of Biological Sciences

Dr. Felix Breden
Internal Examiner
Departmental Chair, Professor, Department of Biological Sciences

Dr. John Taylor
External Examiner
Associate Professor, Department of Biology
University of Victoria

Date Defended/Approved: April 20th, 2012

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website (www.lib.sfu.ca) at <http://summit/sfu.ca> and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

Abstract

It is suggested that gene or genome duplication is the driving force in evolution that leads to speciation. Two models, the classical model and the duplication-degeneration-complementation (DDC) model, have been proposed on the fates of gene duplicates resulting from either a gene or a genome duplication event. The classical model suggests that one of the gene duplicates might result in loss of function (non-functionalization) or gain of a new function (neo-functionalization) depending on whether the accumulated mutations over the years are deleterious or beneficial to the organism. In the DDC model, it is proposed that each of the gene duplicates might accumulate different deleterious mutations in the regulatory region of the gene, such that these genes partition the ancestral gene function (sub-functionalization). Combinations of the phylogenetic analysis of many gene families support that salmonids have undergone two additional whole genome duplications compared to the mammals, one occurred in the common ancestors of teleosts and another happened in the common ancestor of salmonids approximately 25-120 million years ago. In this thesis, the evolution of the fatty acid-binding protein (*fabp*) family in fish and salmonids was examined. I have characterized eighteen unique *fabp* genes in Atlantic salmon. These include the seven *fabp* sub-families described previously in fish. Phylogenetic analyses and conservation of synteny support the two whole genome duplication events in the common ancestors of teleosts and salmonids and indicate when gene losses occurred.

Genetic mapping of *fabp* gene duplicates to homeologous chromosomes in Atlantic salmon also support that they arose by the 4R genome duplication. I also searched for the signatures of neo-functionalization and sub-functionalization by calculating dN/dS ratios, examining the nature of amino acid substitutions and expression patterns, and suggested the fates of *fabp* gene duplicates in Atlantic salmon. Overall, the findings of this project provide insight into the evolutionary processes at play in salmonid genomes.

Keywords: Genome duplication; Intracellular lipid binding protein (ilbp); Fatty acid binding protein (fabp); Salmonids, Fate of duplicated genes.

Acknowledgements

First of all, I thank my senior supervisor, Dr. William Davidson, for his patience and guidance in the past few years. I also would like to thank my two committee members, Dr. Christopher Beh and Dr. Norbert Haunerland, for their input advices for this project. Moreover, I thank past and present members of Davidson's lab for their advices and support. This project would not be carried out as smoothly as it is without the help of our collaborators: Dr. Ben Koop and Dr. Kris von Schalburg from University of Victoria, Dr. Jonathan Wright and Ananda B. Venkatachalam from Dalhousie University and Carlo Biagi and Terri Sutherland from Department of Fishery and Ocean.

I want to thank my family, Sam, Susan and Lucina, and my grandparents for their endless love and support. Lastly, thank you to my friends, Allen, Deidre, Jeanie, Mabel, Nina and Yuka for their support along this journey.

Table of Contents

Approval.....	ii
Abstract	iii
Acknowledgements	v
Table of Contents	vi
List of Figures	x
List of Tables.....	xiii
Chapter 1: Introduction.....	1
1.1 Preamble.....	1
1.2 Background literature to gene and genome duplications	2
1.2.1 Examples of gene duplications.....	2
1.2.2 Evidence of genome duplication and model organisms for its study	5
1.3 Mechanisms of gene and genome duplication.....	10
1.3.1 Gene duplication – Tandem duplication	10
1.3.2 Genome duplication – Allotetraploidization and Autotetraploidization	10
1.4 Genome duplication events in the vertebrates.....	13
1.4.1 2R genome duplication.....	13
1.4.2 3R genome duplication.....	14
1.4.3 4R genome duplication.....	15
1.5 Proposed Fates of Duplicated Genes.....	19
1.5.1 Redundancy	19
1.5.2 Non-functionalization.....	19
1.5.3 Neo-functionalization.....	20
1.5.4 Sub-functionalization	21
1.6 Advantages and disadvantages of gene/genome duplication	25
1.6.1 Advantages.....	25
1.6.2 Disadvantages	26
1.7 Fatty acid-binding proteins (FABPs)	27
1.8 Purpose of the thesis.....	32
2: Materials and methods	34
2.1 Bioinformatics search for <i>fabp</i> sequences in teleosts and tetrapods	37
2.2 Bioinformatics search for salmonid <i>fabp</i> EST sequences	37
2.3 Design of Atlantic salmon <i>fabp</i> gene-specific primers and 40-mer probes	38
2.4 Identification of <i>fabp</i> containing Atlantic salmon BAC clones	43
2.4.1 Screening Atlantic salmon CHORI-214 BAC library.....	43
2.4.2 Polymerase Chain Reaction (PCR) test for specificity of <i>fabp</i> gene-specific primers.....	43
2.4.3 Hot-start Colony PCR on hybridization positive BACs.....	44

2.5	Shotgun library production.....	45
2.5.1	BAC DNA isolation	45
2.5.2	Sonication and end-repairing reaction and cloning	45
2.5.3	Insert size check and sequencing check for bacterial genomic DNA contamination	46
2.6	Sequencing read assembly and gene annotation	47
2.7	Genetic mapping of <i>fabp</i> genes via linkage analysis.....	47
2.7.1	Testing informativeness of the microsatellite primers on mapping parents	47
2.7.2	Genotyping markers on mapping Br5 and Br6 families.....	48
2.8	Minimum tiling paths	52
2.9	Reverse transcription PCR (RTPCR).....	52
2.10	Quantitative PCR.....	54
2.11	Bioinformatics search for 5' putative transcription factor binding sites and dN/dS analysis	54
3:	Characterization of Atlantic salmon <i>fabps</i>.....	56
3.1	Identification of Atlantic salmon <i>fabp</i> transcript and protein sequences	56
3.2	Identification of teleost and tetrapod <i>fabp</i> genomic, transcript and protein sequences.....	59
3.3	Genomic sequences and chromosomal locations of Atlantic salmon <i>fabps</i>	64
3.3.1	Checking the specificity of Atlantic salmon <i>fabp</i> gene-specific primers	64
3.3.2	Identification of BACs with <i>fabps</i> via Atlantic salmon BAC library screening	67
3.3.3	Genomic sequences of <i>fabps</i> via sequencing reads from BAC shotgun libraries.....	75
3.3.4	Genomic sequences of <i>fabps</i> via sequencing reads from Atlantic salmon genome sequencing project	79
3.3.5	Genetic mapping of Atlantic salmon <i>fabps</i>	81
4:	Evolution and expression profiling of the Atlantic salmon <i>fabp2</i> sub-family	87
4.1	Introduction	87
4.2	Identification of salmonids, northern pike and rainbow smelt <i>fabp2</i> transcript sequences.....	88
4.3	Evolution of Atlantic salmon <i>fabp2</i> sub-family	90
4.4	Conserved synteny of <i>fabp2</i> loci in teleosts	95
4.5	Expression profiling of Atlantic salmon <i>fabp2</i> sub-family	100
4.6	Evidence of neo-functionalization.....	102
4.7	Conclusion.....	105
5:	Evolution and expression profiling of the Atlantic salmon <i>fabp7</i> sub-family	106
5.1	Introduction	106
5.2	Identification of salmonids, northern pike and rainbow smelt <i>fabp7</i> transcript sequences.....	107
5.3	Identification of BACs with <i>fabp7</i> via <i>fabp3</i> BAC library screening	109
5.4	Evolution of Atlantic salmon <i>fabp7</i> sub-family	112
5.5	Conserved synteny of <i>fabp7a</i> loci in teleosts	115
5.6	Expression profiling of Atlantic salmon <i>fabp7</i>	119
5.7	Putative 5' regulatory elements.....	121

5.8	Evidence of neo-functionalization.....	127
5.9	Conclusion.....	129
6: Evolution and expression profiling of the Atlantic salmon <i>fabp3</i> sub-family		130
6.1	Introduction	130
6.2	Identification of salmonid, northern pike and rainbow smelt <i>fabp3</i> transcript sequences.....	131
6.3	Evolution of Atlantic salmon <i>fabp3</i> sub-family	133
6.4	Conservation of synteny at <i>fabp3a</i> loci in teleosts.....	138
6.5	Expression profiling of Atlantic salmon	141
6.6	Evidence of neo-functionalization.....	145
6.7	Conclusion.....	148
7: Evolution and expression profiling of the Atlantic salmon <i>fabp1</i> sub-family		149
7.1	Introduction	149
7.2	Identification of salmonids, northern pike and rainbow smelt <i>fabp1</i> transcript sequences.....	149
7.3	Gene annotation of <i>fabp1</i> genes in some teleosts.....	152
7.4	Evolution of Atlantic salmon <i>fabp1</i> sub-family	152
7.5	Expression profiling of Atlantic salmon <i>fabp1</i> sub-family	156
7.6	Conclusion.....	158
8: Evolution and expression profiling of the Atlantic salmon <i>fabp6</i> sub-family		159
8.1	Introduction	159
8.2	Identification of salmonids, northern pike and rainbow smelt <i>fabp6</i> transcript sequences.....	159
8.3	Gene annotation of <i>fabp6</i> genes in some teleosts.....	161
8.4	Evolution of Atlantic salmon <i>fabp6</i> sub-family	161
8.5	Conserved synteny of <i>fabp6a</i> loci in teleosts.....	166
8.6	Expression profiling of Atlantic salmon <i>fabp6s</i>	169
8.7	Conclusion.....	171
9: Evolution and expression profiling of the Atlantic salmon <i>fabp10</i> sub-family		173
9.1	Introduction	173
9.2	Identification of salmonids, northern pike and rainbow smelt <i>fabp10</i> transcript sequences.....	173
9.3	Gene annotation of fugu <i>fabp10</i> gene	175
9.4	Evolution of Atlantic salmon <i>fabp10</i> sub-family	175
9.5	Expression profiling of Atlantic salmon <i>fabp10s</i>	179
9.6	Conclusion.....	181
10: Evolution and Expression profiling of Atlantic salmon <i>fabp11</i> sub-family		182
10.1	Introduction	182
10.2	Identification of salmonid, northern pike and rainbow smelt <i>fabp11</i> transcript sequences.....	184
10.3	Gene annotation of tetraodon <i>fabp11b</i>	186
10.4	Evolution of Atlantic salmon <i>fabp11</i> sub-family	186

10.5 Expression profiling of Atlantic salmon <i>fabp11s</i>	191
10.6 Conclusion.....	193
11: Future Work.....	194
12: Conclusion.....	196
Reference List	198

List of Figures

Figure 1.1. Phylogenetic positions of some proposed polyploidy events during eukaryote evolution.....	8
Figure 1.2. Phylogenetic relationships among the sequenced fungal genomes.....	9
Figure 1.3. Gene duplication by unequal crossing-over between two homologous chromosomes during meiosis is schematically illustrated.	12
Figure 1.4. Phylogenetic tree of major vertebrate groups and superimposed Fugu gene duplication events.....	18
Figure 1.5. Five potential fates of duplicate gene pairs with multiple regulatory regions.	24
Figure 1.6. Scheme for the evolution of the ILBP family (Schaap et al., 2002).	31
Figure 2.1. Scheme of the flowchart of the project.	36
Figure 3.1. ClustalW alignment of Atlantic salmon <i>fabp2</i> unique full-length ESTs.....	65
Figure 3.2. A phosphor screen image of Atlantic salmon genomic BAC library filter twelve that had been hybridized with <i>fabp6</i> and overgo (positive control) oligonucleotide probes.	69
Figure 3.3. A 2% agarose gel image of PCR testing on <i>fabp6</i> hybridization-positive BAC clones with <i>fabp6aI</i> , <i>fabp6aII</i> and <i>fabp6b</i> gene-specific primers.	70
Figure 3.4. A screenshot of ASalbase showing the DNA fingerprinting contig 2577. BAC clone S0487M09 (*), the representative BAC clone of Atlantic salmon <i>fabp6b</i> , belongs to this contig.	72
Figure 3.5. An 1% agarose gel image reflects the quality of the shearing progress on the isolated BAC S0938E16 DNA.....	76
Figure 3.6. Size fractionalization of BAC S0938E16.....	77
Figure 3.7. An agarose gel image of testing the clone insert sizes of the BAC S0938E16 shotgun library.	78
Figure 3.8. An image of the polyacrylamide gel illustrating the genotyping with the <i>fabp2b</i> linked variable marker, Ssa10082BSFU, on the Atlantic salmon Br6 mapping family.	83
Figure 3.9. Agarose gel image illustrates the PCR result that was used for generating a MTP of DNA fingerprinting contig 798 to link the variable microsatellite marker to the <i>fabp7bI</i> representative BAC clone, S0232I03.	84
Figure 3.10. A screenshot from ASalBase on DNA fingerprinting contig 798, where the representative BAC clone for <i>fabp7bI</i> is S0232I02.	86
Figure 4.1. Maximum parsimony phylogenetic tree of FABP2 built from the amino acid sequence alignment shown in Figure 4.2.	92

Figure 4.2. Alignment of the FABP2 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.....	93
Figure 4.3. Comparative genomic analysis of <i>fabp2a</i> loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.	98
Figure 4.4. Comparative genomic analysis of <i>fabp2b</i> loci in Atlantic salmon, stickleback, medaka and fugu.	99
Figure 4.5. Evolutionary history of Atlantic salmon (<i>Ssa</i>) and rainbow trout (<i>Omy</i>) FABP2aI and FABP2aII and chinook salmon FABP2a using the zebrafish (<i>Dre</i>) FABP2 as an outgroup (see also Figure 4.2).	104
Figure 5.1. A 2% agarose gel image of PCR testing on <i>fabp3</i> hybridization-positive BAC clones with <i>fabp7bI</i> and <i>fabp7bII</i> gene-specific primers.	110
Figure 5.2. Alignment of the FABP7 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.....	113
Figure 5.3. Maximum likelihood phylogenetic tree of <i>fabp7</i> built from the nucleotide coding sequence alignment.	114
Figure 5.4. Comparative genomic analysis of the <i>fabp7a</i> loci in zebrafish, stickleback, medaka, fugu and tetraodon.	118
Figure 5.5. Expression patterns of <i>fabp7bI</i> and <i>fabp7bII</i> in Atlantic salmon smolt tissues and of <i>fabp7b</i> in the northern pike tissues.....	120
Figure 5.6. Putative transcription factor binding motifs in the 2000 bp regions upstream of the start codons of teleost <i>fabp7b</i> genes.	123
Figure 5.7. Sequences of the putative transcription binding motifs identified in four or more of the 2000 bp upstream regions of teleost <i>fabp7b</i> genes as shown in Figure 5.6.	124
Figure 5.8. Putative POU transcription binding sites in the 2000bp upstream regions of teleost <i>fabp7b</i> genes.	126
Figure 5.9. Evolutionary history of rainbow smelt (<i>Omo</i>) and Northern pike (<i>Elu</i>) FABP7b, grayling (<i>Tth</i>), Atlantic salmon (<i>Ssa</i>) and rainbow trout (<i>Omy</i>) FABP7bI and FABP7bII using the zebrafish (<i>Dre</i>) FABP7b as an outgroup (see Section 3.2.5).	128
Figure 6.1. Alignment of the vertebrate FABP3 protein sequences.	135
Figure 6.2. Chromatograms showing the sequence reads of the 5' end (a) and exon three (b) of the stickleback <i>fabp3</i> coding sequences.	136
Figure 6.3. Maximum parsimony phylogenetic tree illustrating the evolution of the FABP3 sub-family.	137
Figure 6.4. Comparative genomic analysis of <i>fabp3a</i> loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.	140
Figure 6.5. Expression profiles of <i>fabp3aI</i> and <i>fabp3aII</i> in Atlantic salmon smolt tissues.....	143
Figure 6.6. qPCR analyses on Atlantic salmon <i>fabp3aI</i> and <i>fabp3aII</i> expression in adult tissues.	144
Figure 6.7. Evolutionary history of Atlantic salmon and rainbow trout FABP3aII and FABP3aII using the zebrafish FABP3 as the outgroup (see Figure 6.3).....	147

Figure 7.1. Alignment of the FABP1 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts, chicken, mouse, rat and human.	154
Figure 7.2. Maximum likelihood phylogenetic tree of <i>fabp1</i> built from the nucleotide coding sequence alignment.	155
Figure 7.3. Expression patterns of <i>fabp1a</i> and <i>fabp1b</i> in Atlantic salmon smolt tissues.	157
Figure 8.1. Alignment of the FABP6 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.	164
Figure 8.2. Maximum likelihood phylogenetic tree of <i>fabp6</i> built from the nucleotide coding sequence alignment.	165
Figure 8.3. Comparative genomic analysis of <i>fabp6a</i> loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.	168
Figure 8.4. Expression patterns of <i>fabp6aI</i> , <i>fabp6aII</i> and <i>fabp6b</i> in Atlantic salmon adult tissues.	170
Figure 9.1. Alignment of the FABP10 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and chicken.	177
Figure 9.2. Maximum likelihood phylogenetic tree of <i>fabp10</i> built from the nucleotide coding sequence alignment.	178
Figure 9.3. Expression patterns of <i>fabp10aI</i> , <i>fabp10aII</i> and <i>fabp10b</i> in Atlantic salmon smolt tissues.	180
Figure 10.1. Alignment of the FABP11 protein sequences from Atlantic salmon, rainbow trout, zebrafish, stickleback, medaka, fugu and tetraodon with the FABP4 amino acid sequences from chicken, mouse, rat and human.	188
Figure 10.2. Maximum likelihood phylogenetic tree of <i>fabp11</i> built from the nucleotide coding sequence alignment.	190
Figure 10.3. Expression patterns of <i>fabp11aI</i> , <i>fabp11aII</i> and <i>fabp11b</i> in Atlantic salmon adult tissues.	192

List of Tables

Table 2.1. Oligonucleotide probes used for Atlantic salmon BAC library screening.	40
Table 2.2. Atlantic salmon <i>fabp</i> gene-specific PCR primers used for PCR testing on hybridization positive BAC clones and RT-PCR. <i>Ef1α</i> was used as the endogeneous reference gene for RT-PCR and qPCR.....	41
Table 2.3. Primers not including M13 region that were used for genotyping the Atlantic salmon mapping family for positioning <i>fabp</i> locations in the genetic map. * denotes the RFLP primers used for genotyping that was previously done in the lab.	49
Table 3.1. Atlantic salmon full-length <i>fabp</i> EST sequences and their GenBank accession numbers as of February 16 th , 2012.	58
Table 3.2. Ensembl accession numbers in version 64 of different <i>fabps</i> in the five genome-sequenced teleost species that were examined in this thesis.	60
Table 3.3. Ensembl accession numbers in version 64 of different <i>fabps</i> in four tetrapod species that were examined in this thesis.	63
Table 3.4. Summary of all the <i>fabps</i> identified in Atlantic salmon in this thesis.....	73
Table 3.5. Summary of <i>fabps</i> and their corresponding CIGENE scaffolds. Note that CIGENE scaffold 151259024 only covers exons one and two.	80
Table 4.1. Rainbow trout and chinook salmon full-length <i>fabp</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.....	89
Table 5.1. Rainbow trout, grayling, northern pike and rainbow smelt full-length <i>fabp7</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.....	108
Table 6.1. Rainbow trout full-length <i>fabp3</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.....	132
Table 7.1. Rainbow trout, sockeye salmon and rainbow smelt full-length <i>fabp1</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.	151
Table 8.1. Rainbow trout and northern pike full-length <i>fabp6</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.....	160
Table 9.1. Rainbow trout, sockeye salmon and rainbow smelt full-length <i>fabp10</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.	174
Table 10.1. Rainbow trout, sockeye salmon and rainbow smelt full-length <i>fabp11</i> EST sequences and their GenBank accession numbers as of February 20 th , 2012.	185

Chapter 1: Introduction

I have no special talents, I am only passionately curious. ~Albert Einstein

1.1 Preamble

It is been proposed that gene or genome duplications are the driving force of evolution. As will be illustrated below, there are many examples that provide evidence for gene duplications and indeed whole genome duplication (WGD). The salmonid fishes, which include Atlantic salmon and rainbow trout, descended from a common ancestor which underwent a WGD. This group of fish have been recognized as a good model to investigate evolutionary processes that possibly result from the WGD. There was also a WGD in the common ancestor of all teleost fish, which include salmonids. This means that for every one gene that is observed in tetrapods, such as human, it is predicted that there will be two equivalents in the teleosts and four in the salmonids. Several studies have examined the evolution of duplicated genes in salmonids, such as Hox (Moghadam et al., 2005), growth hormone (von Schalburg et al., 2008), haemoglobin genes (Quinn et al., 2010), myostatin genes (Ostbye et al., 2007), but these have not included other teleosts and tetrapods. My thesis was designed to overcome this shortcoming by investigating a multi-gene family that has been well characterized in a mammal (human) and at least one teleost (zebrafish) and which is amenable in Atlantic salmon. I chose the fatty acid-binding protein (*fabp*) gene family because it arose by a series of gene duplications and there are ten *fabp* genes well annotated in the human

genome. This gene family gives multiple opportunities to investigate the fates of duplicated genes following two rounds of WGD. In this thesis, I describe the historical development of the field of evolutionary biology as it relates to gene and genome duplication. I then build on this foundation and put the evolution of vertebrate *fabp* genes into this context. In particular, the possible fates of *fabp* gene duplicates will be described.

1.2 Background literature to gene and genome duplications

Many documents related to gene and/or genome duplication observations date back to the early 1910s (Bridges, 1935; Kuwada, 1911), and some of them are described in Susumo Ohno's book, *Evolution by gene duplication* (Ohno, 1970). After the DNA sequencing era began in the 1980s and sequencing technologies became more advanced and cheaper in cost, these led to the start of many genome-sequencing projects in the early 1990s. As a result, genomic data of different organisms became available and many of them provided support for gene and/or genome duplications. Below I will discuss some examples of gene and/or genome duplications. As one will see, examples of gene and/or genome duplication can be found in organisms ranging from the kingdoms of fungi, plantae, animalia and protista (Zhou et al., 2010).

1.2.1 Examples of gene duplications

Lactate dehydrogenase (LDH) is the classic example of a gene family that evolved by a series of gene duplications. Markert and Moller (1959) proposed the term isozymes to describe proteins that use the same substrate and coenzyme to carry out the same reaction but that have different kinetic properties (Markert and Moller, 1959). They

used LDH as their prime example. LDH, an enzyme that is involved in the glycolytic pathway, interconverts pyruvate and lactate with the help of NADH and NAD⁺ in the limited oxygen condition and vice versa when oxygen is available. Because it plays an important role in a well-conserved metabolic pathway, LDH is also well-conserved from most of the unicellular bacteria to multi-cellular organisms in the kingdoms plantae and animalia (Markert et al., 1975).

All LDH are isozymes consisting of four subunits, which can be homomers or heteromers of A and B subunits, A₄ (LDH-A), A₃B, A₂B₂, AB₃ and B₄ (LDH-B), or homomers of the C subunit, C₄ (LDH-C) (Markert and Moller, 1959). Protein sequence similarity is the main evidence suggesting that LDH-A, LDH-B and LDH-C isozymes arose by at least two gene duplications (Taylor et al., 1973). In fact, protein sequences of the A and B subunits are similar such that heteromeric isozymes of these subunits can form *in vitro*. Moreover, similar or identical active site amino acid residues are found in all LDH isozymes (Rehse and Davidson, 1986; Taylor et al., 1973; Tsoi and Li, 1994), with the result that these LDH isozymes catalyze the same reaction in a cell even though they have different kinetics. For example: LDH-A (A₄) has low affinity for pyruvate and is expressed in anaerobic tissues in human, such as fetal and skeletal muscle tissues, whereas LDH-B (B₄) has a higher affinity for pyruvate and is expressed in oxygen-rich tissues, such as heart. Hence, lactate dehydrogenase isozymes are examples of genes that resulted from a series of gene duplications.

Another protein family that arose by gene duplication is fructose biphosphate aldolase. In the early 1960s, it was found that there are three distinct fructose biphosphate aldolase subunits: A, B and C (Penhoet et al., 1966). Like the case of LDH,

all fructose biphosphate aldolase isozymes catalyze the same reaction, which is the conversion of fructose-1,6-phosphate to glyceraldehyde-3-phosphate and dihydroxyacetone phosphate and vice versa (Penhoet et al., 1966). Similarly, any two of the fructose biphosphate aldolase A, B and C subunits can bind to each other *in vitro* to form three homomeric isozymes and three heteromeric isozymes even though there are differences in catalytic properties and primary structures between the isozymes (Penhoet et al., 1966). Therefore, members of the fructose biphosphate aldolase protein family resulted from gene duplications.

Although it might appear that gene duplications only occurred in metabolic-related proteins, there are examples of gene duplications in other functional genes, such as myoglobin and haemoglobin. Both myoglobin and haemoglobin carry oxygen to different parts of the body. The x-ray diffraction patterns of both proteins show that myoglobin and hemoglobin folds around the heme group in nearly the same way (Kendrew et al., 1960; Perutz et al., 1960); the heme attaches to two histidine residues that are well apart from each other in either protein. This suggests that myoglobin and haemoglobin have similar protein sequences. Differences between myoglobin and haemoglobin are that myoglobin is a monomeric protein whereas haemoglobin is a tetramer protein, and a haemoglobin subunit is shorter in length compared to a myoglobin protein. Myoglobins are found in muscle (oxygen-limited area) whereas hemoglobins are found in the blood (oxygen-rich area). Interestingly, monomeric haemoglobin can still be found in hagfish and the lamprey. In fact, lamprey haemoglobin size (156 amino acid residues) is similar to the myoglobin amino acid length in sperm whale (153 amino acid residues), which suggests that the loss of a few amino acid residues occurred after the

development of a jaw by vertebrates (Ohno, 1970). Due to the similarity in the protein structure and function, it is known that myoglobin and haemoglobin are related to one another by gene duplication with a series of subsequent gene duplications leading to the haemoglobin subunit diversity.

1.2.2 Evidence of genome duplication and model organisms for its study

Before the sequencing era had begun in the early 1980s, cytogenetic studies revealed the homeologous chromosomes in species that had undergone a WGD, while studies on nuclear DNA content provided evidence for the doubling of the DNA content in the post-WGD species compared to the pre-WGD species (Dingerkus and Howell, 1976; McCollum, 1958; Thiebaud and Fischberg, 1977; Uyeno and Smith, 1972). Furthermore, many sequencing projects on different species from different kingdoms began in the early 1990s. Upon gene annotation, data from these sequencing projects revealed the number of duplicate genes in these genomes (Figure 1.1), providing more support of WGDs.

Saccharomyces cerevisiae is one of the species in the kingdom of eukaryotes that has undergone a WGD. *S. cerevisiae* genome sequencing data were available for analysis in 1996 (Goffeau et al., 1996). It is known that a WGD occurred in the common ancestor of several yeast species within the *Saccharomyces* complex group (Figure 1.2). Studies had shown that chromosome numbers are doubled in “post-WGD” yeast species relative to the ancestral “pre-WGD” yeast (Kellis et al., 2004; Wolfe, 2006). For example, not only are many unique genes in *Kluyveromyces waltii*, a pre-WGD yeast species, seen as duplicates in *S. cerevisiae*, but *K. waltii* also has the gene order as the predicted pre-WGD genomic organization based on the duplicated *S. cerevisiae* genome.

In fact, the number of duplicate genes are about 1/5 of the protein coding genes found in *S. cerevisiae* (~1100 genes); that is 551 duplicated “ohnolog” pairs in the genome (Wolfe, 2006). Ohnologs refers to genes that are related to another by genome duplication (Wolfe, 2001). In addition, homeologous chromosomes in *S. cerevisiae* were identified (Kellis et al., 2004). Hence, *S. cerevisiae* is a well-accepted post-WGD yeast species and many studies used this as a model to study the molecular evolution of genome duplication.

The *Arabidopsis thaliana* genome sequencing project began in 1996 and a few years later its genomic data revealed that it is another degenerate tetraploid species (Arabidopsis Genome Initiative, 2000). The lineages of *Arabidopsis* and rice diverged ~200 million years ago (MYA) (Wolfe et al., 1989) and van Dodeweerd et al. (1999) showed that three regions of the genome of *Arabidopsis* were related to one another and to one region in the rice genome, suggesting that there are multiple duplication events in the *Arabidopsis* lineage (van Dodeweerd et al., 1999). In fact, homeologous genes are 20% to 47% of the genes within the duplicated regions in the *Arabidopsis* genome (Arabidopsis Genome Initiative, 2000).

Xenopus laevis, an African clawed frog, was also thought to have arose from a degenerate tetraploid, as a cytogenetic study first revealed that members of the genus *Xenopus* of the family of Pipidae have chromosome numbers $2N = 20, 36, 40, 72$ and 108 , suggesting that many polyploidization events had occurred in this genus (Kobel and Du Pasquier, 1986). As a post-WGD species, *Xenopus laevis* has a chromosome number of 36 ($2N = 36$) and approximately 3 Gbp genome size, which is twice the amount as in the pre-WGD species, *Xenopus tropicalis*, ($2N = 20$, ~1.5 Gb) (Bisbee et al., 1977;

Hirsch et al., 2002). In addition, there are more than 2200 unique genes that are found as single copies in *X. tropicalis* but which have two co-orthologs in *X. laevis* (Hellsten et al., 2007). Orthologs define as genes that are related to one another by speciation. Examples of the duplicated genes found in *X. Laevis* are hairy 2 genes (Murato et al., 2007) and matrix metalloproteinase-9 genes (Hasebe et al., 2007). Nevertheless, Hellsten et al. estimated that 50-75% gene pairs have subsequently lost one of the gene copies in the *X. laevis* genome after the WGD (Hellsten et al., 2007).

Figure 1.1. Phylogenetic positions of some proposed polyploidy events during eukaryote evolution.

Filled circles mark lineages in which genome duplication has been inferred; open diamonds mark two hexaploid lineages of plants. The question marks shown the positions of the two rounds of genome duplication proposed under the 2R hypothesis. (Spring, 1997; Wolfe, 2001)

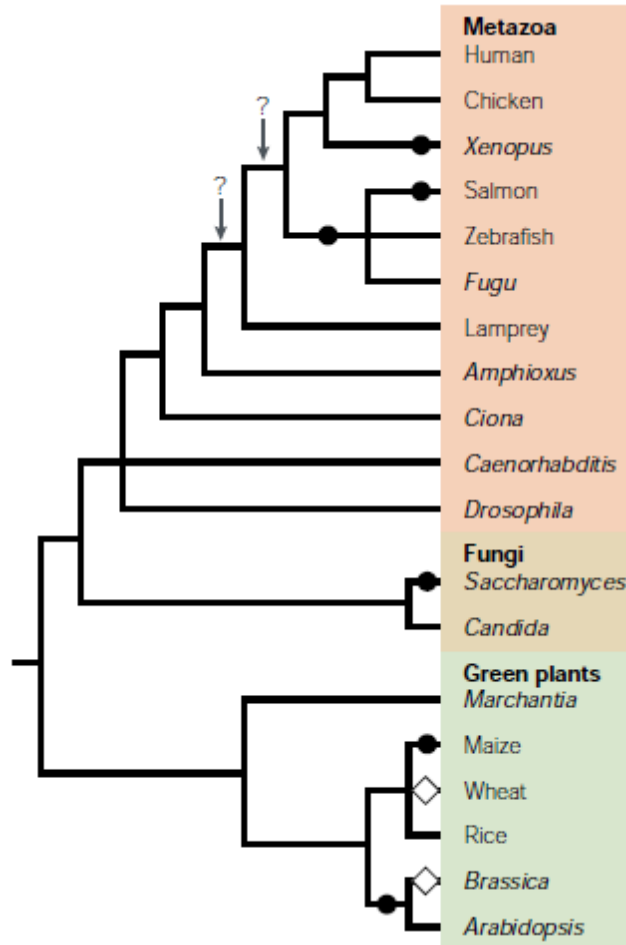
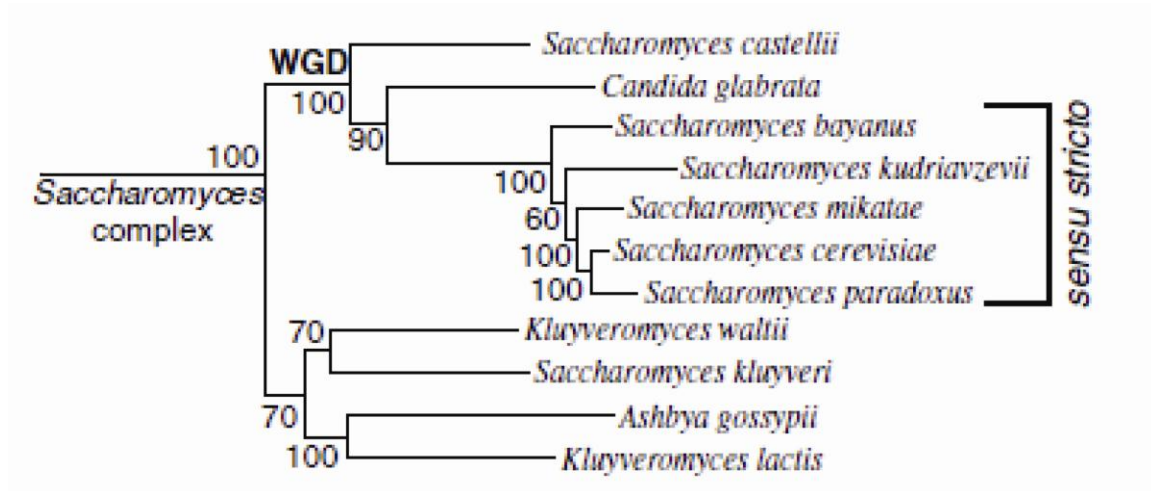


Figure 1.2. Phylogenetic relationships among the sequenced fungal genomes.

The tree is a maximum-likelihood phylogeny reconstructed using the concatenated sequences of 153 genes that are universally present in the 42 genomes shown. Bootstrap percentages are shown for all nodes. The clade, 'Saccharomyces complex', includes the group of species that share the whole-genome duplication (WGD). This figure is modified from (Fitzpatrick et al., 2006).



1.3 Mechanisms of gene and genome duplication

1.3.1 Gene duplication – Tandem duplication

One of the gene duplication mechanisms is tandem duplication. Tandem duplication results in gene duplicates that reside beside one another in the same transcriptional direction. It can be due to unequal crossing-over between two homologous chromosomes that occurs during meiosis as well as between two chromatids of a chromosome in the spermatogonia or oogonia (Figure 1.3). If it is an unequal crossing-over between the two sister chromatids of the same chromosome, the resulting tandem duplicated genes will have the same allele (Ohno, 1970). However, if the crossing-over occurs between the two homologous parental chromosomes, the tandem-duplicated genes might have different allele sets. Ribosomal RNA, transfer RNA and histone genes are examples of tandem duplications. In particular, the 5S rRNA gene cluster spans a 2.2 kb repeated region in the human genome (Stults et al., 2008). A tRNA gene cluster consists of two tRNA (GAAPhe) and two tRNA (UUULys) (Doran et al., 1987). There are 16 histone genes residing in a gene cluster at the D6S105 locus in the human genome (Albig and Doenecke, 1997).

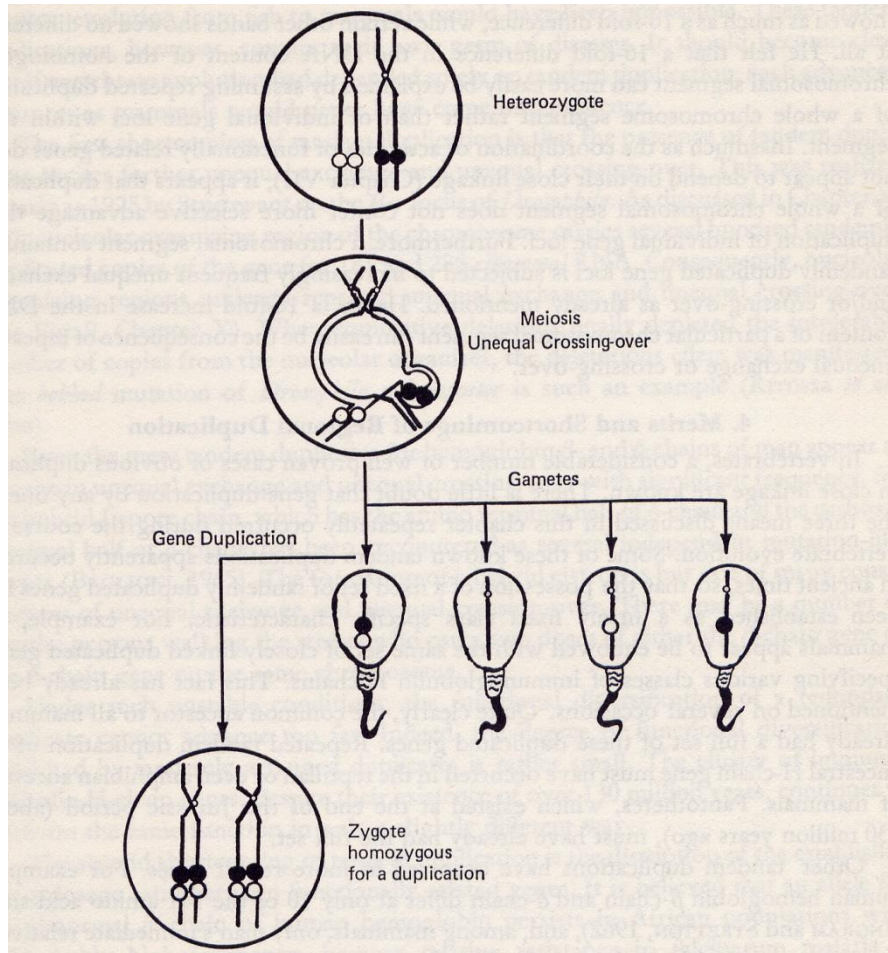
1.3.2 Genome duplication – Allotetraploidization and Autotetraploidization

WGD can occur via allotetraploidization or autotetraploidization (Ohno, 1970). In both cases, they are due to non-disjunctions that occur during the segregation of the duplicated spermatogonia and oogonia in meiosis, such that the gametes end up having 2N instead of N. Allotetraploidization occurs when 2N gametes from two different species (interspecies) are fertilized and fused and the resulting offspring would be an allotetraploid (4N). *Barbus barbus*, a cyprinid fish species, is an example of an

allotetraploid (Wolf et al., 1969). Autotetraploidization, on the other hand, occurs when two intra-species diploid gametes ($2N$) fuse together and form a tetraploid offspring ($4N$). A difference between allotetraploid and autotetraploid is that one might observe two independent sets of bivalents in the meiosis of the allotetraploid species, whereas quadrivalents would be observed in the meiosis of a fresh autotetraploid species. In fact, many south American frog species from the family *Ceratophryidae* arose by autotetraploidization (Saez and Brum, 1960). Members of this family have chromosome numbers ranging from 22 to 104. *Odontophrynus cultripes* of this family has a diploid chromosome number of 22, with eleven bivalents found during meiosis. However, *O. Americanus*, another species of this family, forms eleven quadrivalents in meiosis (Becak et al., 1966). Salmonids are examples of autotetraploids (Ohno, 1970). Both allotetraploidization and autotetraploidization double the genetic material.

Figure 1.3. Gene duplication by unequal crossing-over between two homologous chromosomes during meiosis is schematically illustrated.

If this occurs in a heterozygote, two alleles of the same gene locus become two independent gene loci in extremely close linkage. The small circle denotes the centromere of the chromosomes. The white and the black large circles represent the two alleles of the same locus (top). Unequal crossing-over during meiosis of a heterozygote places two alleles on the same chromatid (second row). One of the 4 gametes produced by such a germ cell carries this duplication (third row). This figure was taken from Ohno (1970).



1.4 Genome duplication events in the vertebrates

1.4.1 2R genome duplication

It has been estimated that vertebrates have approximately four times as many genes as *Drosophila* (Miklos and Rubin, 1996). In addition, the existence of many duplicated gene loci in the tetrapods led to the proposal that there was a genome duplication event early in the evolution of vertebrates (Ohno, 1970). Later, the evolution of the *Hox* gene clusters elucidated that there is a single *Hox* gene cluster in compared to four clusters in tetrapods; and Ohno refined his proposal and suggested that two duplications occurred in the early evolution of vertebrates, one on either side of the jawless fish divergence, which are approximately 500 and 430 MYA, respectively (Ohno, 1998). In other words, the 1-2-4 hypothesis suggests that for every gene copy presenting in the protostome or early deuterostome, there are at least two gene copies in the jawless fish and four gene copies in the tetrapods. Hence, this elucidates the fourfold increase of genes in the jawed vertebrates compared to protostome or early deuterostome (Figure 1.4) (Holland et al., 1994; Holland and Garcia-Fernandez, 1996). Beside *Hox*, there are many more developmental regulator genes supporting this hypothesis, such as *Cdx* (Gamer and Wright, 1993), *MyoD* (Atchley et al., 1994), *btd/SP* (Kingsley and Winoto, 1992). In addition, conserved synteny blocks can be found in the homeologous chromosomes (Abi-Rached et al., 2002; Horton et al., 2003; Lundin et al., 2003; Lynch and Conery, 2000; Meyer and Schartl, 1999; Pebusque et al., 1998; Postlethwait et al., 2004). Hence, many scientists believe that two genome duplications happened in the early evolution of vertebrates.

However, there are studies that provide evidence against the 2R genome duplication (Hughes, 1999; Martin, 2001). According to the 2R hypothesis, the topology of phylogenetic tree illustrating four gene duplicates A, B, C and D, should show the duplication of two ancestral genes (AB) and (CD), which come from the first genome duplication of the ancestral gene, A'. Many gene families follow this phylogenetic topology, but studies have demonstrated that the phylogeny of many developmental protein families (A-D) exhibits a topology of (A) (BCD) instead (Hughes, 1999; Martin, 2001). Moreover, many genes do not follow the 1-2-4 pattern, such as globins and lysozyme, tubulins and actins, manganese superoxide dismutase, large subunits of RNA polymerase, olfactory receptor, immunoglobulin gene families (Sidow, 1996) and Wnt-5 (Sidow, 1992).

1.4.2 3R genome duplication

It was suggested that there was an additional genome duplication in the evolution of vertebrates such that fish have more genes than human in general (Amores et al., 1998; Aparicio et al., 2002; Meyer and Schartl, 1999; Panopoulou et al., 2003; Wittbrodt et al., 1998). There are three main pieces of evidence supporting this hypothesis: 1) More than four *Hox* gene clusters are found in teleost fishes, including zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), African cichlid (*Oreochromis niloticus*), the green pufferfish (*Tetraodon nigroviridis*), fugu pufferfish (*Takifugu rubripes*) and stickleback (*Gasterosteus aculeatus*), compared to four gene clusters in human, and many genes and gene clusters other than *Hox* cluster have two copies in the teleost fishes for every one copy in the tetrapods (Meyer and Schartl, 1999); 2) paralogs (genes that are related to one another by gene or genome duplication) are found in the conserved syntenic blocks

(Jaillon et al., 2004; Taylor et al., 2003); and 3) they reside on different chromosomes (Meyer and Schartl, 1999). Since many gene duplicates are found in the teleost fishes, this 3R genome duplication event appeared to occur in the common ancestor of teleosts before the divergence of most teleost species (Hoegg et al., 2004). The sequences of duplicated genes in fugu and tetraodon were used to estimate the timing of this genome duplication event to be approximately 320 to 425 MYA, which is the time at the base of the radiation of all teleost fish (Figure 1.4) (Vandepoele et al., 2004). Hence, the timing is consistent and can explain the origins of the ohnologs found in many teleost fishes. Hoegg et al. (2004) has been suggested that the 3R genome duplication might correlate to the diversity of the extant teleost fishes; that is ~23,637 species in 425 families (Nelson, 1994). However, when comparing gene copies of the orthologues in fish and mouse, it was observed that different fish lineages have different numbers of duplicated genes in their genomes. This observation suggests that differences in the number of gene duplicates in the fish lineages are the products of lineage-specific duplication events rather than a genome duplication event at the base of the teleost radiation (Robinson-Rechavi et al., 2001a; Robinson-Rechavi et al., 2001b).

1.4.3 4R genome duplication

The salmonid-specific WGD (4R) was first proposed by Svardson (Svårdson, 1945). By observing the chromosome numbers in salmonid species seemed to fall into multiples of ten, and the numbers of bi-armed chromosomes and the multivalent chromosomes in meiotic preparations from several species, Svårdson suggested that the common ancestor of the salmonids possessed ten chromosomes and all the extant salmonids are the result of polyploidization events (Svårdson, 1945). However, the

hypothesis was rejected because even though the genomes of Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*) differ by at least ten chromosomes, they have the same DNA content and total chromosome length (Rees, 1964).

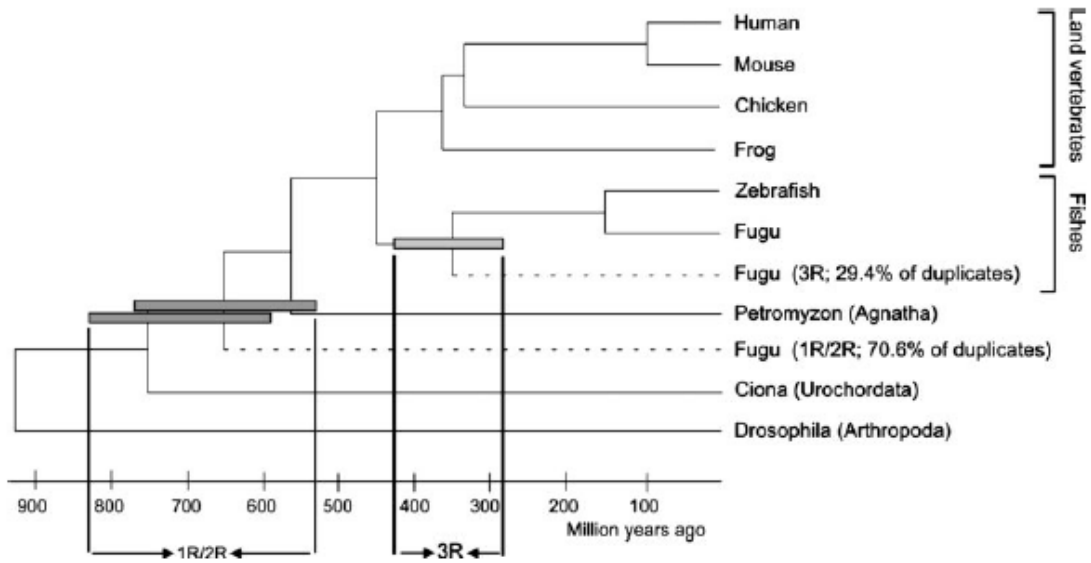
Later, Ohno proposed that salmonids are tetraploid derivatives of clupeoid fish, and suggested that all extant salmonidae species, such as Atlantic salmon and rainbow trout, had experienced a genome duplication (Figure 1.5) (Ohno, 1970). Based on the fossil record of the salmonids, the timing of the 4R genome duplication is estimated to be 25-100 MYA (Norden, 1961). There are several lines of evidence supporting the salmonid-specific genome duplication: 1) The doubling of DNA content and chromosome arms relative to the closest diploid clupeoid fishes, such as smelt, herring and anchovy (Mank and Avise, 2006); 2) The number of duplicated genes annotated by the salmonid EST database (Koop et al., 2008; Leong et al., 2010); 3) The homeologous segments of chromosomes in salmon (Danzmann et al., 2008; Gharbi et al., 2006; Nichols et al., 2003; Phillips et al., 2009; Woram et al., 2004) and; 4) the residual tetrasomal recombination seen in meiosis in salmonids (Allendorf and Danzmann, 1997; Ohno, 1970). The multivalent chromosomes that were observed in meiotic preparations seen in some salmonidae species suggest that these genomes are still undergoing rediploidization and have not reached a stable diploid state (Allendorf and Danzmann, 1997).

It has been proposed that the diploid ancestor of salmonids had about 48 chromosomes ($2N = 48$) or 44-52 with predominantly acrocentric or subtelocentric chromosomes (Gold et al., 1979). The duplicated salmonid common ancestor should have about 96 chromosomes in theory, but the independent genomic rearrangements

occurring in different isolated salmonid populations result in the variety of chromosome numbers seen in the extant salmonids and cause the differences from the expected chromosome number. These genomic rearrangements can be insertions, deletions, inversions, translocations and Robertsonian fusions, which involve the fusion of two acrocentric chromosomes creating a new metacentric chromosome (Ohno, 1970). Eventually these genomic rearrangements lead to re-diploidization, that is the formation of two separate bivalents instead of one quadrivalent (Shaver, 1963). In other words, the duplicated genome reaches a stable diploid state. This is the situation that is seen in the genomes of the closely-related trout and salmon. In fact, many gene families have been used as models to study the re-diploidization process in the salmonids, including LDH (Ohno, 1968) and NAD-dependent malate dehydrogenase (MDH) (Bailey et al., 1969).

Figure 1.4. Phylogenetic tree of major vertebrate groups and superimposed Fugu gene duplication events.

This figure was taken from Vandepoele et al. (2004).



1.5 Proposed Fates of Duplicated Genes

In the classical model, Ohno suggested that there are three possible fates of duplicated genes: redundancy, non-functionalization and neo-functionalization (Figure 1.5) (Ohno, 1970). Later, Force et al. proposed that duplicated genes can be sub-functionalized as a way to retain both gene duplicates in the WGD (Force et al., 1999). The following will describe possible fates of duplicated genes.

1.5.1 Redundancy

Redundancy occurs when both gene duplicates retain the ancestral gene function and remain in the genome. Examples of redundancy are the 5S, 18S and 28S ribosomal and the transfer RNAs gene clusters in the vertebrates. It is suggested that an increased amount of these ribosomal and transfer RNA transcripts by an increase in the number of these genes allows the species to fulfil its protein production demand more easily (Ohno, 1970).

1.5.2 Non-functionalization

Non-functionalization occurs when one of the gene duplicates becomes non-functional due to accumulating deleterious mutations (Force et al., 1999). This can be caused by deleterious mutations in the promoter, such that no transcript is made and a pseudogene is present in the genome (Figure 1.6). Pseudogene defines when a gene does not provide a functional protein. It can also be caused by genomic rearrangement occurring within the coding regions of the gene, such that the gene structure is no longer intact and no complete functional transcript is made. Thirdly, mutations could occur in the coding regions such that the active site of the protein product becomes non-

functional. All three scenarios (mutations that stop normal transcription, translation and or protein function) could lead to non-functionization. In fact, most of the gene duplicates in a duplicated genome are expected to be non-functional and become pseudogenes (Force et al., 1999) because deleterious mutations occur much more frequently than beneficial mutations (Lynch and Conery, 2000; Lynch and Force, 2000). Some pseudogenes can be found in the haemoglobin gene cluster, which resulted from a series of tandem duplications of the haemoglobin gene as mentioned in section 1.2.1.

1.5.3 Neo-functionalization

Neo-functionalization occurs when one of the gene duplicates gains a new function(s) that the ancestral gene did not possess (Ohno, 1970). New functions can be in terms of a new substrate, or binding partner, or protein function, or tissue expression sites and expression time. This can occur when mutations occur in regulatory elements or the binding and active sites of the protein such that now it takes a different substrate or a co-factor to carry out a new cellular function that the ancestral protein does not exhibit. Another possible outcome of the neo-functionalized gene duplicate is a change in expression (e.g. cell type) or at a lifestage that is different the ancestral gene.

As mentioned above, neo-functionalized gene duplicates are rare because most of the mutations in a genome are likely to be deleterious rather than beneficial. However, one set of neo-functionized gene duplicates that are described by Ohno is trypsin and chymotrypsin (Ohno, 1970). Both trypsin and chymotrypsin cleave the peptide bond at the carboxyl side of amino acids, but trypsin targets the basic amino acids lysine and arginine whereas chymotrypsin focuses on cleavage after the aromatic amino acids phenylalanine and tyrosine. The protein lengths of trypsin and chymotrypsin differ by

seven amino acids. With the similarity in protein function acting on different amino acid substrates, the active sites of trypsin and chymotrypsin are similar but distinctly different from each other (Neurath et al., 1967). The only difference in the active sites of trypsin and chymotrypsin is the two of the eleven sites that surround the active serine. Another neo-functionalization example is the anti-freeze glycoproteins (AFGP) in Antarctic notothenioid fishes (Taylor and Raes, 2004). Initially, a trypsinogen-like protein was duplicated and one of the gene duplicates had its threonine-alanine-alanine coding elements repeated presumably via DNA slippage during replication (Chen et al., 1997; Cheng and Chen, 1999). The exons coding for protease-specific sequences were then lost, which gave rise to the current form of AFGP. Hence, AFGP, a protein that prevents freezing of the body fluids, possesses a function that the ancestral trypsinogen-like protein, which is a protease, did not have.

1.5.4 Sub-functionalization

With non-functionalization and neo-functionalization, one would expect a low number of preserved gene duplicates in a duplicated genome. However, the observed number of duplicated genes preserved in duplicated genomes are greater than expected (Lynch and Force, 2000). Hence, Force et al. proposed the duplication-degenerate complimentary model, which suggests that sub-functionalization is another fate of duplicated genes (Force et al., 1999). Sub-functionalization occurs when accumulated deleterious mutations occur in both gene duplicates such that the duplicated genes partition the ancestral gene function in terms of substrates, binding partners, expression sites or timing or level or protein function. For instance, the presence of regulatory elements A and B result in expression in tissue A and tissue B, respectively. The

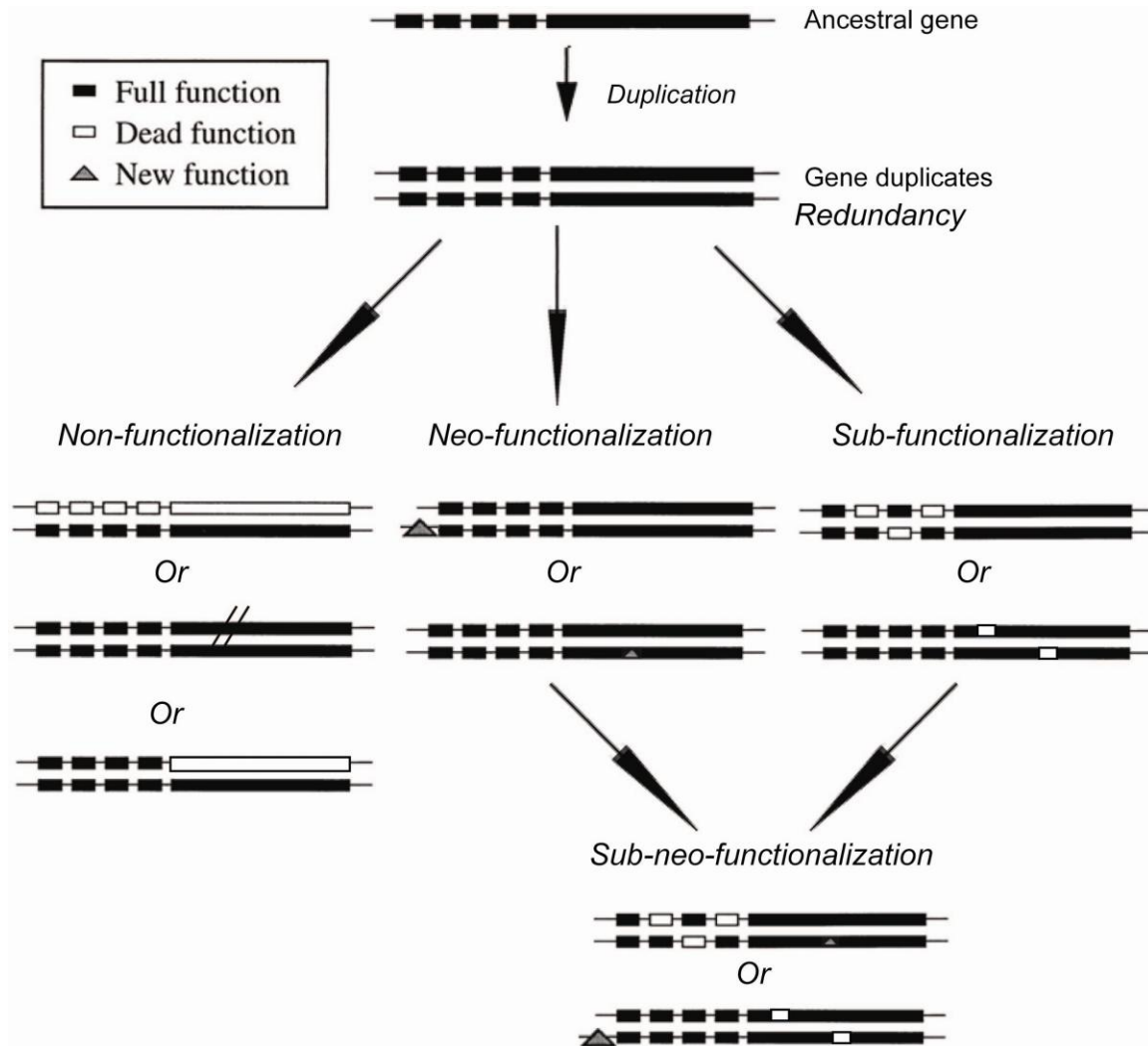
ancestral gene possesses both regulatory elements A and B; hence, the gene is expressed in both tissues, A and B. However, due to the accumulated deleterious mutations in gene duplicates, such that gene duplicate #1 only contains regulatory element A whereas gene duplicate #2 only possesses regulatory element B. Therefore, both gene duplicates are now partitioning the ancestral gene expression pattern. Similar phenomena can occur at the regulatory region and the active site(s) of the ancestral gene structure, such that partitioning of gene expression level or protein function or substrate or binding partner binding can be observed in a particular tissue at a particular life stage of the species. Because the complementary degenerative mutations in different regulatory elements or active site(s) can facilitate the preservation of both gene duplicates, it maximizes the opportunities for the evolution of new gene functions.

An example of sub-functionalized genes that Force et al. described is the *engrailed* gene in the zebrafish (Force et al., 1999). Tetrapods possess two engrailed genes, *Eng1* and *Eng2*, and there are four engrailed genes, *Eng1*, *Eng1b*, *Eng2* and *Eng3*, in zebrafish. Phylogenetic analysis and conservation of synteny suggest that zebrafish *Eng1* and *Eng1b* were originated from tetrapod *Eng1*, whereas zebrafish *Eng2* and *Eng3* are gene duplicates of tetrapod *Eng1*. Without knowing the closest teleost pre-WGD species at that time, tetrapod *Eng1* expression pattern was used to represent the expression pattern of *Eng1* prior to the gene or genome duplication event. Mouse and chicken *Eng1s* are expressed in both pectoral appendage bud and specific neurons in hindbrain and spinal cord. Indeed, zebrafish *Eng1* is expressed in pectoral appendage bud and *Eng1b* is expressed in specific neurons in hindbrain and spinal cord.

It is even possible that the duplicated genes are sub-neo-functionalized (He and Zhang, 2005; Marcussen et al., 2010). Here, I will take the example I used in describing sub-functionalization. After gene duplicates #1 and #2 partition the ancestral gene expression pattern, more mutations occurred in one of the gene duplicates, such that gene duplicate #1 not only contains regulatory element A and is expressed in tissue A but also interacts with a new substrate relative to the ancestral gene. The difference between sub-functionalization, neo-functionalization and sub-neo-functionalization is heavily dependent on the ancestral gene function. Hence, comprehending all functions of the ancestral gene is critical to determine the fate of duplicated genes.

Figure 1.5. Five potential fates of duplicate gene pairs with multiple regulatory regions.

The small boxes denote regulatory elements with unique functions, and the large boxes denote transcribed regions. Solid boxes denote intact regions of a gene, while open boxes denote null mutations, triangles denote the evolution of a new function and double tilt lines denote translocation. The figure is modified from Force et al. (1999).



1.6 Advantages and disadvantages of gene/genome duplication

1.6.1 Advantages

Natural selection should select the changes that give an overall benefit to the survival of organism. Hence, what are the advantages for a species experiencing a gene or genome duplication? One of the advantages is that the gene duplicates can provide more amount of the same product to fulfill the metabolic requirement (Ohno, 1970). Example of this type of advantage is the copy number of human amylase (Perry et al., 2007). The amount of carbohydrate content in one's main diet correlates with the copy number of amylase in an individual. Hence, an increase of amylase copy number allows the individual to fulfill his/her metabolic demand more easily.

A second possible advantage is that the gene-duplicated species can absolutely obtain benefits from two different alleles of the same gene by making the two alleles of the ancestral gene into gene duplicates via tandem duplication (Ohno, 1970). However, this phenomenon can only be advantageous when the gene does not interfere with gene dosage with its interact partners. An example of this is the esterase in catostomid fish of the Colorado River system (Koehn and Rasmussen, 1967). The Colorado River is a long river that extends from the cold Northern mountains to the hot Southern desert. There are two allelic forms of esterase, A and B, in the catostomid fish; esterase A works best at 5°C and esterase B functions well under 20°C. Hence, the catostomid fish population in the Northern mountains and in the Southern desert would be homozygous A form and B form of esterase, respectively. Because the water temperature in the intermediate part of the river fluctuates widely with season, fish would have a better-fit if they are heterozygous in this subpopulation. However, if heterozygous

fish are in the subpopulation in the intermediate part of the river, only 50% of the subpopulation can become heterozygotes and the remaining will not survive in each generation, so natural selection would favour a duplication of the esterase locus in the intermediate subpopulation. Indeed, the intermediate stream fish obtain two gene copies of esterase, which are tandem-duplicated and encode for the A and B forms.

A third possible advantage is that the generation of isozymes allows them to function more efficiently in different conditions, such that the overall cellular processes can operate more efficiently. Metabolic isozyme gene families are examples of this, such as LDH, fructose biphosphate aldolase, pyruvate kinases and phosphoglucomutase (Markert et al., 1975). Members of each of these isozyme gene families carry out the same reaction but with different kinetics, such that the same metabolic reaction can operate at different cellular or tissue condition most efficiently.

Lastly, a WGD event is believed to be an opportunity to provide more genetic material to allow species to survive and adapt to different niches and this would eventually lead to a species radiation. This is supported by the relatively few species in the sister group of the salmonids, Esocidae (pike and mudminnows), compared with the approximately 70 salmonid species (Ramsden et al., 2003).

1.6.2 Disadvantages

A disadvantage of genome duplication is that it disturbs many well-established cellular mechanisms in the ancestral (pre-WGD) species. Sex-determination is one of the affected cellular pathways (Ohno, 1970). For example, the pre-WGD species has established gene “A” as the sex-determination gene, there would be two copies of the

gene “A” in the genome after a WGD event, so the post-WGD species must establish a new sex-determination gene. Other cellular pathways, such as transcription and signalling pathways, may also be affected by super-repression, which is when there is a secondary repressor to turn off a pathway that has already been activated by a primary activator (Ohno, 1970). Hence, there are both benefits and risks involved in a WGD.

Another possible disadvantage of a WGD is the presence of pseudogenes. As mentioned above due to the probability of accumulating mutations that are deleterious, one of each gene duplicates set resulting from either gene or genome duplication is likely to become non-functional and a pseudogene (Lynch and Conery, 2000). Although it has been suggested that pseudogenes and “junk DNA” are neutral or in fact beneficial (Muro et al., 2011; Veitia and Bottani, 2009), the function of the non-transcribed pseudogenes in a cell is unknown (Muro et al., 2011). However, it would seem to be a waste of resources for a species to undergo a gene or genome duplication without specific benefit to the survival of an organism. Hence, the presence of pseudogenes might be a disadvantage of a WGD.

1.7 Fatty acid-binding proteins (FABPs)

According to the salmon gene nomenclature “policy”, I will describe the genes encoding the fatty acid-binding proteins in lower case and in italics, whereas I will describe the proteins in capital letters. Fatty acid-binding proteins (FABPs), cellular retinoid-binding proteins (CRBPs) and cellular retinoic-acid binding proteins (CRABPs) are part of the intracellular lipid-binding proteins (ILBPs) superfamily (Schaap et al., 2002). It is known that *ilbps* arose by a series of gene duplications in the common ancestor of vertebrates (Schaap et al., 2002). As of today, twelve *fabp* sub-families, two

crabp and two *crbp* sub-families have been documented from different species, but not all gene sub-families occur in any particular vertebrate (Figure 1.6). These proteins were named after the tissue they were isolated initially; however, this causes naming confusion due to their overlapping expression profiles. Hence, another naming nomenclature, which is based on the chronological order of these proteins being discovered, was proposed (Hertzel and Bernlohr, 2000). There are several recent reviews focusing on the *fabp* divergence and function in chicken and human (Hughes and Piontkivska, 2011; Smathers and Petersen, 2011; Storch and McDermott, 2009).

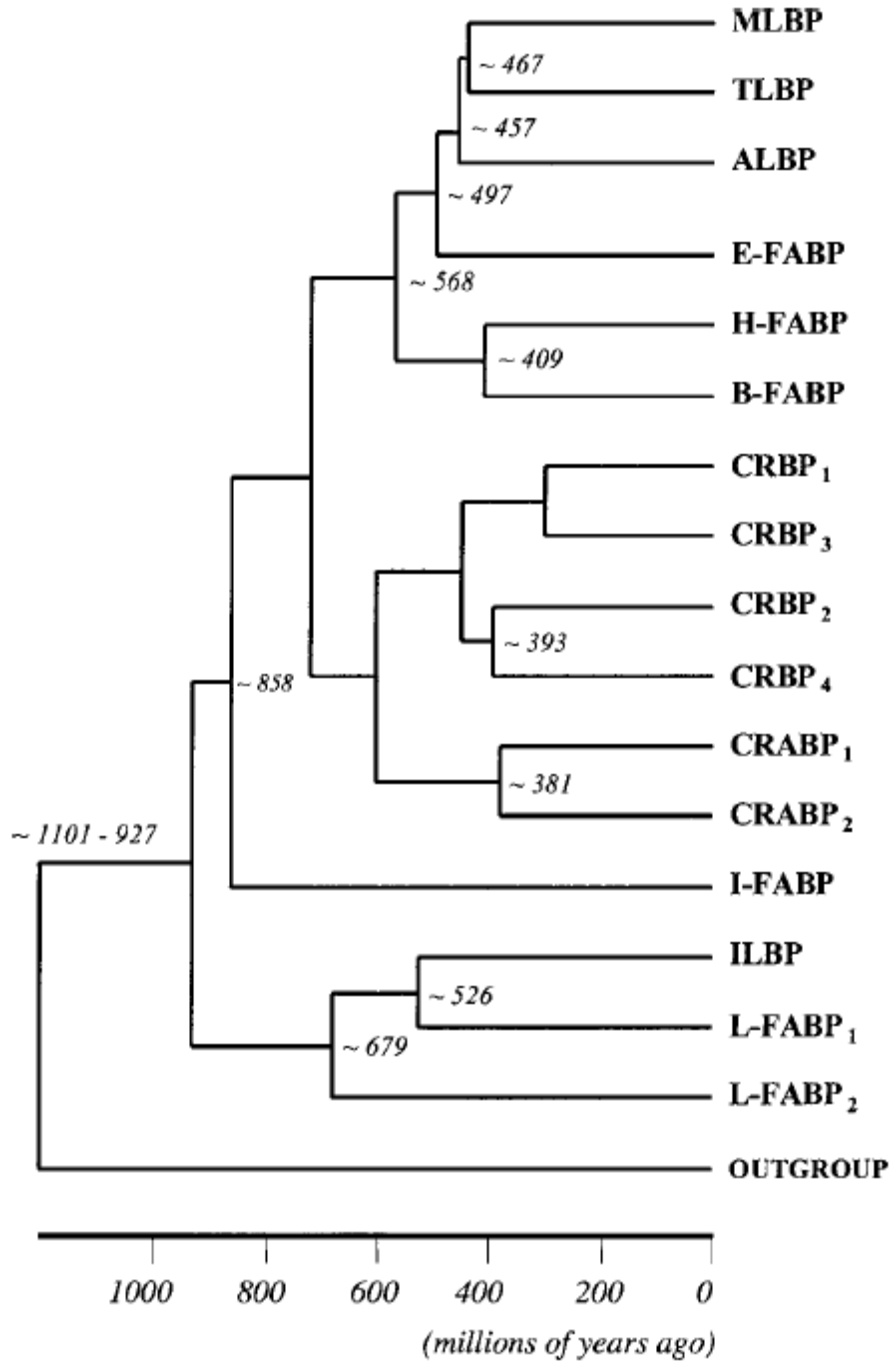
All FABPs bind to one molecule of fatty acid with the exception of the liver-type FABP (L-FABP), which binds two molecules (Bernlohr et al., 1997; Haunerland and Spener, 2004). FABPs are 14-15 kDa proteins of 127-133 amino acids, that are found in the intracellular fluid (Coe and Bernlohr, 1998; Veerkamp and Maatman, 1995; Zimmerman and Veerkamp, 2002). Even though all FABPs have a distinct conserved protein structural feature, a β -barrel consisting of ten strands of anti-parallel β -sheets with a helix-loop-helix domain at its cap (Chmurzynska, 2006), they have different binding affinities for different fatty acids (Hanhoff et al., 2002). This difference of fatty acid binding is due to the extremely wide difference of coding sequence identity, ranging from 15% to 70%, between different *fabp* sub-families, which leads to the differences in protein sequences that affect volume capacity in the binding pocket and binding affinity for a particular fatty acid (Chmurzynska, 2006; Veerkamp and Maatman, 1995). For example, the fatty acid has an extended, slightly bent conformation in FABP2 and FABP4, but a U-shaped conformation in FABP3 and FABP8 (Veerkamp, 1995).

In fact, the function of each individual *fabp* sub-family is unknown because of their overlapping expression in a particular tissue and knock-out mice might not have any physical effect (Furuhashi and Hotamisligil, 2008). Nevertheless fatty acids are very insoluble in aqueous medium, so it is known that FABPs play a role in influencing the flux of the fatty acids by diffusion (Luxon and Weisiger, 1993). FABPs control the fatty acid content of a membrane (Chmurzynska, 2006) and transfer fatty acids to lipid droplets for storage (Furuhashi and Hotamisligil, 2008). FABPs also transport fatty acids to mitochondria or peroxisomes for oxidation to obtain energy source. Hence, there is a correlation between the FABP content and the fatty acid oxidation capacity, as well as the amount of fatty acid exposure from the diet (Chmurzynska, 2006; Giuseppe, 1999; Haunerland and Spener, 2004; Veerkamp and van Moerkerk, 1993; Veerkamp, 1995). Since fatty acids can be metabolized into eicosanoids, which provide a large variety of lipid mediators that may function as pro- and anti-inflammatory mediators (Dickinson Zimmer et al., 2004; Ek et al., 1997; Funk, 2001; Serhan, 2007), FABPs are linked to both metabolic and inflammatory pathways (Chmurzynska, 2006; Coe and Bernlohr, 1998; Haunerland and Spener, 2004; Makowski and Hotamisligil, 2005; Zimmerman and Veerkamp, 2002). In addition, FABPs play a role in fatty acid-dependent gene regulation by shuttling specific fatty acids to a fatty acid-binding nuclear receptor (Clarke and Jump, 1993). For example, studies have shown that the liver FABP interacts with peroxisome proliferator-activated receptor (PPAR) to regulate the fatty acid metabolism (Desvergne and Wahli, 1999; Hostetler et al., 2009) and preadipocyte differentiations (Gregoire et al., 1998). Interestingly, the transcriptional levels of *fabps* may be modified by these PPARs. FABPs also transport fatty acids for the incorporation of absorbed fatty acids into

triacylglycerols and phospholipids and secretion in chylomicrons (Veerkamp, 1995), interact with other cytosolic enzymes to regulate their activities, control various cellular processes by modulation of the free fatty acid concentration (Veerkamp, 1995), transport lipids to the endoplasmic reticulum (ER) for signalling and trafficking and membrane synthesis and to extracellular space as signals in an autocrine or paracrine manner (Furuhashi and Hotamisligil, 2008). Since FABPs are conserved from invertebrates, such as *Drosophila* and *C. elegans*, to tetrapods, this suggests FABPs are important in many cellular processes.

Figure 1.6. Scheme for the evolution of the ILBP family (Schaap et al., 2002).

The tree was rooted by including an outgroup (Von Ebner's gland protein) in the phylogenetic analysis. Gene duplication times are shown in millions of years ago. Note that fish FABP11 and tetrapod FABP12 were not included in this figure because they were not characterized at that time.



1.8 Purpose of the thesis

Several studies have examined the evolution of duplicated genes in salmonids, such as Hox (Moghadam et al., 2005), growth hormone (von Schalburg et al., 2008), haemoglobin genes (Quinn et al., 2010), myostatin genes (Ostbye et al., 2007), and they support the 4R salmonid-specific WGD. The central hypothesis of my thesis is that the 3R and the 4R WGD occurred in which for every gene in tetrapods, there are two gene duplicates in teleosts and four gene copies in salmonids. I set out to test this hypothesis by searching for evidence in the *fabp* family in tetrapods and teleosts, whose genomes have been sequenced, and EST databases for salmonids. If this hypothesis did not hold for a *fabp* gene sub-family, I hypothesized that one or more gene losses would explain the observation. The null hypothesis is that when a gene is duplicated, there is no change in the expression pattern or rate of evolution along the lineages leading to the paralogs. I tested this using phylogenetic analysis, quantitative PCR, reverse-transcribed PCR, and indicators of signatures of selection.

This gene family was chosen because it arose by a series of gene duplications and there are ten *fabp* genes well annotated in the human genome, which provide multiple opportunities to investigate the fates of duplicated genes. As mentioned in section 1.6, FABP is a small, conserved and abundant protein in different vertebrate lineages. Given the great number of EST sequences available in the Atlantic salmon EST database, I predicted to be able to capture most of the transcribed *fabps* in this species. Moreover, the *fabp* gene family has been well characterized in zebrafish (Alves-Costa et al., 2008; Denovan-Wright et al., 2000a; Denovan-Wright et al., 2000b; Karanth et al., 2008; Karanth et al., 2009; Liu et al., 2003a; Liu et al., 2003b; Liu et al., 2004; Liu et al., 2007;

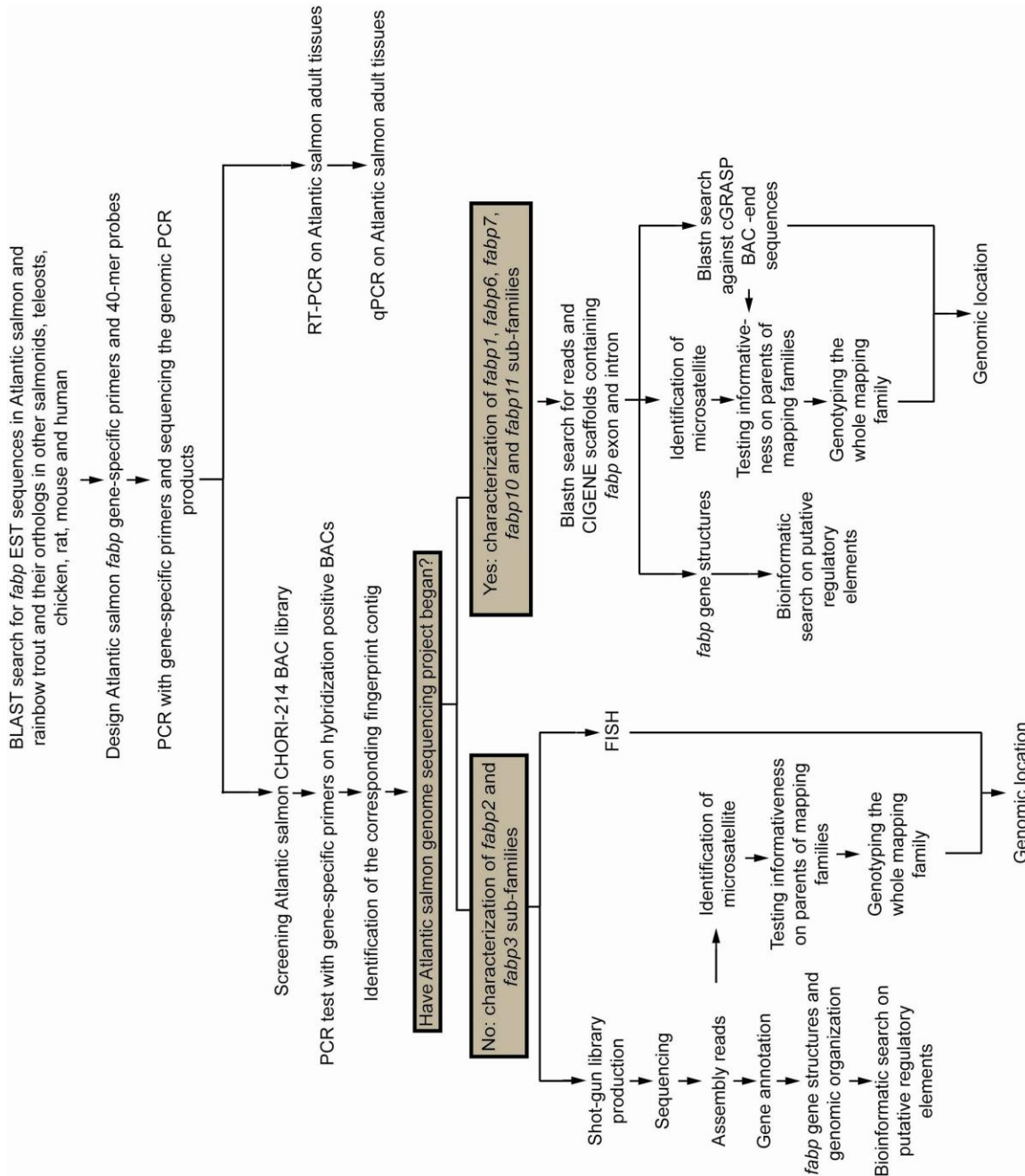
Pierce et al., 2000; Sharma et al., 2004; Sharma et al., 2006; Venkatachalam et al., 2009), which allowed me to carry out a comparative genomic analysis and compare the expression patterns among the *fabp* orthologs and to test if there is no change in expression pattern or rate of evolution along the lineages leading to these paralogs. This project was designed to provide insights into the evolution of the *fabp* gene family in teleosts and salmonids, in particular the possible fates of *fabp* gene duplicates.

2: Materials and methods

The project was carried out as indicated in Figure 2.1. First, rainbow smelt (*Osmerus mordax*), northern pike (*Esox lucius*) and unique salmonid *fabp* EST sequences, including the ones from Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*) and grayling (*Thymallus thymallus*), were found from the cGRASP EST databases (http://lucy.ceh.uvic.ca/contigs/cbr_contig_viewer.py). Their orthologous sequences in five sequenced teleost species, including zebrafish (*Danio rerio*), three-spined stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes*), tetraodon (*Tetraodon nigroviridis*) and fugu (*Takifugu rubripes*), and in chicken (*Gallus gallus*), rat (*Rattus norvegicus*), mouse (*Mus musculus*) and human (*Homo sapiens*), were retrieved from Ensembl versions 52 to 64. Since *fabps* are well conserved in gene structure, the exon-intron boundaries of each Atlantic salmon *fabp* were predicted by aligning the salmonid *fabp* unique coding sequence with their orthologous sequences. Next, gene-specific primers and 40-mer probes were designed from Atlantic salmon *fabp* coding sequences for PCR, screening the BAC library, RT-PCR and qPCR purposes. The specificity of each gene-specific primer set was first tested by sequencing the single Atlantic salmon genomic DNA PCR product. These *fabp* gene-specific primers were then used to test the corresponding hybridization-positive BACs to check if the BAC clones did indeed contain an insert with a particular *fabp* gene. Because the characterization of *fabp2s* and *fabp3s* had started before the Atlantic salmon genome sequencing project began, shotgun library production of the representative BAC clones

was carried out to obtain the genomic organization of these two *fabp*-subfamilies. For other *fabp* gene structures, including *fabp1s*, *fabp6s*, *fabp7s*, *fabp10s* and *fabp11s*, the gene organizations were retrieved from assembled scaffolds of the sequencing reads from the Atlantic salmon genome project (Davidson et al., 2010). All *fabps* were mapped to the Atlantic salmon genetic map via linkage analysis using informative microsatellites markers. Additionally, *fabp2s* were physically positioned on Atlantic salmon chromosomes by fluorescent *in situ* hybridization (FISH). End-point RT-PCR was carried out to provide preliminary *fabp* expression profiles while quantitative RT-PCR was performed to obtain more detailed expression profiles of some *fabp* sub-families.

Figure 2.1. Scheme of the flowchart of the project.



2.1 Bioinformatics search for *fabp* sequences in teleosts and tetrapods

Given that the zebrafish *fabps* have been well characterized, I first retrieved the coding sequences and protein sequences of *fabp1s*, *fabp2*, *fabp3*, *fabp6*, *fabp7s*, *fabp10s* and *fabp11s* from Ensembl zebrafish database version 52 (<http://uswest.ensembl.org/index.html>). Using Ensembl BLAT (<http://uswest.ensembl.org/Multi/blastview>) and the zebrafish *fabp* coding sequences as queries, I searched for the *fabp* coding sequences and peptide sequences in medaka, stickleback, tetraodon, fugu, chicken, rat, mouse and human. The sensitivity of the search was near-exact matches. In addition, I searched for the coding and protein sequences of *fabp4*, *fabp5*, *fabp8*, *fabp9* and *fabp12* in chicken, rat, mouse and human in Ensembl versions 52 to 64 even though they are only found in tetrapods. Obtaining these sequences provides insights of the evolution of *fabp11* as *fabp4*, *fabp5*, *fabp8*, *fabp9* and *fabp12* are decedents of the same progenitor gene as *fabp11* in fish (Karanth et al., 2008; Liu et al., 2008).

2.2 Bioinformatics search for salmonid *fabp* EST sequences

To search for *fabp* EST contig sequences in the salmonids, I used the zebrafish *fabp* coding sequences as queries for BLASTn searches against the cGRASP salmonid EST All 100/99 databases. The EST databases that I searched included rainbow smelt, northern pike, graylings, rainbow trout and Atlantic salmon. All contigs from these databases were built based on a minimum of 100 bp overlap with at least 99% identity. I only considered the EST contigs with full-length coding sequences as hits because I was only looking for transcribed *fabps*. In cases where more than one full-length EST contig of a particular *fabp* sub-family was found in a particular species, I compared the

consensus sequences of these contigs to one another using the NCBI bl2seq program (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). To account for possible sequencing errors and allelic variation, 98% sequence identity was used as the cut-off to define the unique *fabp* cluster in each species. The FABP amino acid sequences were obtained from the predicted open reading frames of these unique EST cluster consensus sequences using the ExPasy translate tool (<http://ca.expasy.org/tools/dna.html>).

Alignments of unique vertebrate *fabp* coding sequences and the corresponding protein sequences of each *fabp* sub-family were made using Clustal W (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>), MUSCLE (<http://www.ebi.ac.uk/Tools/muscle/index.html>), MAFFT (<http://www.ebi.ac.uk/Tools/mafft/index.html>) and T-coffee (<http://www.ebi.ac.uk/Tools/t-coffee/index.html>). All methods gave the same alignments for the nucleotide and amino acid sequences. The coding sequence alignments were the inputs to MEGA5 (Kumar et al., 2008; Tamura et al., 2007; Tamura et al., 2011) to build phylogenetic trees using the nucleotide substitutions type on a maximum composite likelihood model with 500 bootstrap replications.

2.3 Design of Atlantic salmon *fabp* gene-specific primers and 40-mer probes

Prior to designing the Atlantic salmon *fabp* gene-specific primers and the oligonucleotide (40-mer) probes for each *fabp* sub-family, I predicted the exon-exon boundaries of the Atlantic salmon full-length *fabp* coding sequences by aligning the Atlantic salmon *fabp* coding sequences with their ortholog(s) in zebrafish using CLUSTALW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). This was to ensure that the

oligonucleotide probes or the *fabp* gene-specific primers would be able to bind properly to the genes of interest. I manually designed the *fabp* gene-specific primers with the following criteria: 1) the primer was to be at least 20 bases long; 2) the primer has three unique bases at the 3' end of at least one of the primer set; and 3) the melting temperatures (T_m) of the forward and reverse primers of a primer set have to be within 2°C of one another. The T_m of each primer is calculated using Oligo Calc (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). For the oligonucleotide probes for each *fabp* sub-family, I designed them from a conserved exonic region of the gene duplicates within the sub-family. The list of *fabp*-subfamily oligonucleotide probes and gene-specific primers that were used in this thesis are given in Table 1 and 2, respectively.

Table 2.1. Oligonucleotide probes used for Atlantic salmon BAC library screening.

<i>fabp</i> -subfamily	Target <i>fabp</i> member	Sequence (5'-3')
<i>fabp1</i>	<i>fabp1a</i> and <i>fabp1b</i>	ATCCAGRAGGGCAARGACATCAAGAGCWTSCTCKGAGMTTG
<i>fabp2</i>	<i>fabp2a</i> and <i>fabp2all</i>	AAAGGTACATTCCACCAGAAAGGACAAATGGAAAAGGTGCTGA
	<i>fabp2b</i>	AAGGGTAAATTCACCAGGAAAGACAAACAGTAAGGTGCTGA
<i>fabp3</i>	<i>fabp3a</i> and <i>fabp3all</i>	TTCACGTGCARAAGTGGGACGGCAAGGAGACCACCTCTGGT
	<i>fabp6a</i> and <i>fabp6all</i>	ACTGAGGTGGTGCAGAACGGCAATGAGTTCTCATGGTCCC
<i>fabp10</i>	<i>fabp6b</i>	ACAGAAGTGCTTCAGGAGGGGACGATTTTACCTGGTCAC
	<i>fabp10a</i> , <i>fabp10all</i> and <i>fabp10b</i>	TCACCAACTCMTTCACYATMGGSAARGAGGCMGAMATCAC
<i>fabp11</i>	<i>fabp11a</i> , <i>fabp11all</i> and <i>fabp11b</i>	ARAYGGMAAMCTKGTGCAAAAACAGAMBTGGGRCGGCAA

Table 2.2. Atlantic salmon *fabp* gene-specific PCR primers used for PCR testing on hybridization positive BAC clones and RT-PCR. *Ef1 α* was used as the endogenous reference gene for RT-PCR and qPCR.

<i>fabp</i> subfamily	Gene Name	Forward primer (5'-3')	Reverse primer (5'-3')
<i>fabp1</i>	<i>fabp1a</i>	CACGGGCACCAAGGTCAATAGT	GCGTTTGCTTATCCCTCTTGTAGAC
	<i>fabp1b</i>	CACGGGAACAAGATCCCTCAC	TCGTTTGCTGGTCTTCTTGTATGA
<i>fabp2</i>	<i>fabp2aI</i>	CTTGGAAGTAGACCACACAGT	AACAATGGCTCTGGTAGTTG
	<i>fabp2aII</i>	CTTGGAAGTAGACCAGCAGC	AACAATGGCTCTGGTAGTTT
	<i>fabp2b</i>	GGAAAGTTGAAACGCAACGAG	TACTAGCTCTCCTCCCACCA
<i>fabp3</i>	<i>fabp3aI</i>	TCGAGGTAGCTGGAGACACA	GAACCACTCTTACCACCGTCTAT
	<i>fabp3aII</i>	TTGAGGTAGCAGGAGACATG	ACCATCTTACCACCGTCCAC
	<i>fabp6aI</i>	GGATACGAAAGAAATCTGCAAGCTCG	CTGGTCAATTACAAAACCTTGTGGTGAC
<i>fabp6</i>	<i>fabp6aII</i>	GGATACGATGAATCTGTCAAGCTCC	CTGGTCAATGACAAAACCTTGTGCTGAT
	<i>fabp6b</i>	AATTACGATGAGTTTCTGGAGGCAA	CTTTCCGATGGTGAATTTGTTAGTCAA
	<i>fabp7bI</i>	GGTCGACAGTGACAACCTTTGATGAG	GTTTCTGTCACTGCAAGTGGCTTCA
<i>fabp7</i>	<i>fabp7bII</i>	GGTCGACAGTGAAAACCTTCGATGAC	GTCATCTGCAAGTGGCCTCG
	<i>fabp10aI</i>	ATGTGCAACACAGATAAATTCACACA	AAATAGCCAAAGGACTGCCCC
<i>fabp10</i>	<i>fabp10aII</i>	GCAAGTCCGTCAACCAACTCCTT	CCTTGAGCTCCTGGACGGCTG
	<i>fabp10b</i>	ATTTAATGGAAAATGGAAAATCCACAG	GAATGAGTTGGTGACAGTACG
<i>fabp11</i>	<i>fabp11aI</i>	ACCGCACCAAGCCCCAGTTTG	CAGTCACCTCTCTCTATTGTTGA
	<i>fabp11aII</i>	CGCACCAAGCCCAGTCTC	GTCACCTCTCTCTCTATTGTCGT
	<i>fabp11b</i>	GAAATCCCCAGAGCACTTTCAAAACACTAC	CCGCTCGAGCGTGGTAGTC

2.4 Identification of *fabp* containing Atlantic salmon BAC clones

2.4.1 Screening Atlantic salmon CHORI-214 BAC library

Atlantic salmon CHORI-214 BAC library filters were first pre-hybridized with the pre-hybridization buffer (5 x SSC, 5 x Denhardt's reagent and 0.5% SDS) for three hours at 65°C. Meanwhile, the *fabp* oligonucleotide probes and the overgo probe (an oligonucleotide probe designed from *Arabidopsis thaliana* and was used as a positive control) were labelled with ³²P-γATP using T4 polynucleotide kinase (Invitrogen, Burlington, Ontario, Canada) according to the manufacturer's protocol. The labelled probes were added to the BAC library filters and hybridization was carried out overnight. The filters were washed three times with the washing buffer (0.1% SDS and 1 x SSC) at 50°C the next day, with each time being one hour. The filters were wrapped in Saran™ wrap before they were placed inside the phosphor screen cassette and exposed overnight. The images of the filters were visualized by scanning the phosphor screen in the Typhoon imaging system. According to the position of the signals and their locations on the filters, the identity of the hybridization-positive BAC clones could be identified. The DNA fingerprinting contigs to which the BAC clones belong can be retrieved from ASalBase (<http://asalbase.org/sal-bin/index>).

2.4.2 Polymerase Chain Reaction (PCR) test for specificity of *fabp* gene-specific primers

The specificity of each *fabp* gene-specific primer set was tested by hot-start PCR using the primers and Atlantic salmon genomic DNA (50 ng/μl) and sequencing the PCR products. A 25 μl PCR reaction consisted of 0.2 mM dNTPs, 0.5 μM forward and reverse primers and 0.04 U of Taq polymerase (QIAGEN, Mississauga, Ontario, Canada).

Gradient PCR cycle was used as follows: 95°C for 5 minutes followed by, 35 cycles of 95°C for 45 seconds, T_m for 45 seconds and 72°C for 2 minutes, and then 72°C for 10 minutes. The melting temperatures (T_m) of gradient PCR cycle ranged from 50°C to 75°C. Three µl of each PCR reaction was loaded in a 2% agarose gel for visualization before the rest of the PCR products was purified using the QIAquick PCR purification kit (QIAGEN, Mississauga, Ontario, Canada).

The genomic PCR products were sequenced using Amersham Biosciences DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare, Baie d'Urfe, Quebec, Canada). The sequencing reactions and the cleanup of the sequencing reactions were carried out following the manufacturer's instructions. The sequencing samples were sequenced using an ABI PRISM 377 DNA sequencer following the instruction from the manufacturer. The analyzed sequences were then aligned with the Atlantic salmon *fabp* EST contig sequences using CLUSTALW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) to check the amplification specificity of these *fabp* gene-specific primers.

2.4.3 Hot-start Colony PCR on hybridization positive BACs

A 10 µL colony PCR reaction consisted of 1 µL of 10 fold diluted BAC clone glycerol stock, 0.2 mM dNTPs, 0.5 µM forward and reverse primers and 0.4 U of Taq polymerase (QIAGEN, Mississauga, Ontario, Canada). The hot-start PCR condition was the same as for gradient PCR except that the optimal melting temperature varied depending on which *fabp* gene-specific primers were used. Only BAC clones that were positive by hybridization and the particular *fabp* gene-specific PCR were considered to contain that corresponding gene.

2.5 Shotgun library production

2.5.1 BAC DNA isolation

In brief, 2.5 µl of the BAC clone glycerol stock was added to 5 ml of 20 µg/ml LB/chloramphenicol broth as the starter culture, and shaken at 37°C for eight hours prior to add the 5 ml of starter culture into a 500 ml of 20 µg/ml LB/chloramphenicol broth to be shaken for 16 hours. BAC clone DNA was isolated using a QIAgen Large Construct kit according to the manufacturer's protocol, and was dissolved in prewarmed 65°C Buffer EB and left for re-hydration overnight. The quality and quantity of the BAC clone DNA was measured by the Nanodrop Spectrophotometer.

2.5.2 Sonication and end-repairing reaction and cloning

A minimum of 5 µg of BAC clone DNA is required before proceeding to make a shot-gun library. Initially, the BAC clone DNA was sonicated to fragments of 3-5 kb, followed by an end-repairing treatment with End-It™ DNA End-Repair kit (Epicentre, Madison, WI, USA). The end-repairing reactions were carried out as indicated by the manufacturer's protocol. The end-repaired DNA was purified from a 1% agarose gel using QIAquick Gel Extraction Kit (QIAgen, Mississauga, ON, Canada) and was ligated to a previously SmaI-digested and dephosphorylated pUC19 vector using T4 ligase (Invitrogen, Burlington, ON, Canada). The recombinant DNA was transformed into *E. coli* XL1-Blue Supercompetent cells (Stratagene, La Jolla, CA, USA), which were plated onto 200 µg/mL Ampicillin LB plates.

2.5.3 Insert size check and sequencing check for bacterial genomic DNA contamination

Sixty-four recombinant white colonies were picked from each BAC library in 200 µg/mL ampicillin/2xYT broth in an autoclaved 96 well block, and the culture was carried out overnight. Hybrid recombinant DNA were isolated according to the following. Bacterial cell pellets were resuspended in 200 µL chilled solution I (50 µg/mL RNase A, 1 M glucose, 0.5 M Tris Cl (pH 8) and 0.2M EDTA (pH8)), lysed in 200 µl of prewarmed at 37°C solution II (0.1 M NaOH and 1% SDS) and neutralized with 200 µl of chilled solution III (3 M potassium acetate containing 0.115% glacial acetic acid). The supernatant and the cell lysis material were separated by centrifugation, and the hybrid recombinant DNA was precipitated with isopropanol. The DNA pellets were washed with 70% ethanol before they were re-hydrated with 30 µl of nuclease-free water overnight. PvuII restriction digestions were carried out using 400 ng of plasmid, and the insert DNA was visualized in a 1.5% agarose gel.

In addition, plasmid DNA was sequenced using the Amersham Biosciences DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare, Baie d'Urfe, Quebec, Canada) with the T7 and Sp6 sequencing primers to test if there was any bacterial genomic DNA contamination during the cloning process. The sequencing procedure was carried out as indicated in Section 2.4.2. Approximately another 3,000 colonies containing hybrid recombinant plasmids for each BAC were sent to the Michael Smith Genome Sciences Centre (Vancouver, BC) for paired end Sanger sequencing only if the amount of bacterial genomic contamination accounted for less than 10% of the picked colonies for each shot-gun library.

2.6 Sequencing read assembly and gene annotation

The shot-gun sequences of each BAC were assembled using Phred and Phrep (Ewing and Green, 1998; Ewing et al., 1998) and the assembled contigs were visualized using Consed (Gordon et al., 1998). Even though there are variations between reads due to sequencing errors, only the consensus contig sequences were annotated using the consortium for Genomics Research on Atlantic Salmon Project (GRASP) annotation pipeline (<http://grasp.mbb.sfu.ca/bacannotations/GRASPbac.html>).

2.7 Genetic mapping of *fabp* genes via linkage analysis

2.7.1 Testing informativeness of the microsatellite primers on mapping parents

Putative variable microsatellites (1-4 nucleotide repeats) were identified from either BAC end sequences, the assembled shotgun library consensus sequence or the assembled consensus sequences of the assembled CIGENE scaffolds using FastPCR (Kalendar et al., 2009). All microsatellite primers were designed such that the forward primers contained the universal M13 sequence (5'-TGTAACAACGACGGCCAGT-3') at their 5' ends. To test if the microsatellite primers are variable in the Br5, Br6 and NB Atlantic salmon mapping parents, touch-down PCR was carried out. A 10 µl of PCR reaction consisted of 0.2 µM of forward primer, 0.5 µM of reverse and M13 fluorescence-labelled forward primers, 0.2 mM dNTPs, 0.05 U of Taq and 7.5 ng of genomic DNA. The touch-down PCR cycle used is described below: 94°C for two minutes, 94°C for 30 seconds, 30 seconds at the melting temperature which began at 60°C and decreased 0.5°C after each round of the cycle till it reached 50°C, 72°C for 30 seconds the extension time. Once the melting temperature reached at 50°C, there would be another 14 cycles with annealing temperature at 50°C and 72°C for 10 minutes for a final extension. The PCR

products were separated in a 7% polyacrylamide denaturing gel and visualized using the ABI PRISM 377 DNA sequencer.

2.7.2 Genotyping markers on mapping Br5 and Br6 families

When a microsatellite was found to be variable in a mapping family, the genotyping of that particular mapping family was carried out. In brief, a 6 μ l of PCR reaction consisted of 0.3 μ M of forward primer, 0.8 μ M of reverse and M13 fluorescence-labelled forward primers, 0.08 mM dNTPs, 0.04 U of Taq and 7.5 ng of genomic DNA. The touch-down PCR cycle described in Section 2.7.1 was used.

Table 2.3. Primers not including M13 region that were used for genotyping the Atlantic salmon mapping family for positioning *fabp* locations in the genetic map. * denotes the RFLP primers used for genotyping that was previously done in the lab.

<i>fabp</i> subfamily	Gene Name	Microsatellite Name	Forward primer (5'-3')	Reverse primer (5'-3')
<i>fabp1</i>	<i>fabp1a</i>	Ssa0313BSFU	CTAGCCATCGTTCACACATA	ATGACTAGCAATGTCGGAGT
	<i>fabp1b</i>	Ssa1469BSFU	TACACAGAGCTAGACGCAGA	CCCCTAAGTGGGTCTTTAAT
<i>fabp2</i>	<i>fabp2aI</i>	Ssa10060BSFU	TTTTAAGATGGGACAATGCT	CATCTGGGTCTTACCCTTCAT
	<i>fabp2aII</i>	Ssa10056BSFU	AGGAGTGACATTAGAGCGGAG	ACTCTCCCAACTCACTAGCA
<i>fabp3</i>	<i>fabp2b</i>	Ssa10082BSFU	CCACTTGTTTGAAATTTTCCTTG	CAGAAAGGCCCTGGAGCATA
	<i>fabp3aI</i>	Ssa10087BSFU	GTGTATTTGTTTTTCCCTGC	TCAAACTCAAACCACAGACA
<i>fabp6</i>	<i>fabp3aII</i>	Ssa10088BSFU	GGAGAGCAAGA AACTTTGATG	CCCATGCTTTACCACCTTTAC
	<i>fabp6aI</i>	Ssa10088BSFU	TGATGAGCAATCGGTTTTTACC	TCCCAAAAATTTCACACACCA
<i>fabp7</i>	<i>fabp6aII</i>	Ssa10154BSFU	GAGAAAGGCCGTAATGTCAAACG	AGCATGTTCCACATCTCCTG
	<i>fabp6b</i>	Ssa10142BSFU	AGATCAGTGGGGACAAAGCTG	GGAGAAAAGAACACGGAGAGA
<i>fabp10</i>	<i>fabp7bI</i>	Ssa0907BSFU	GTCCTCAATTGGGTTGAGAA	AACACGAGTAAATGAGGCAT
	<i>fabp7bII</i>	Ssa0890BSFU	CACATTTCACTGGACATCAA	GCTTTGTGATTGGAAAAGGTA
<i>fabp10b</i>	<i>fabp10aI</i>	Ssa0015SSFU	*ATTTCATTGGTGACCCGTA AAC	*ATCTGTCTGGAAAACCGTGTC
	<i>fabp10aII</i>	Ssa2795BSFU	CTGCTTCATTGAGCTTACCT	CTCCTCCTCCCTCCTCTC
	<i>fabp10b</i>	Ssa10143BSFU	CTGATGAACACAGGGCAATG	TAAATGCCATGTCCAGCTCA

<i>fabp</i> subfamily	Gene Name	Microsatellite Name	Forward primer (5'-3')	Reverse primer (5'-3')
<i>fabp11</i>	<i>fabp11aIII</i>	Ssa10156BSFU	TGCAGGCATGCATCTAAAAA	AGGCAGCTACCTGAGTGACA
	<i>fabp11b</i>	Ssa10155BSFU	AATGGGAAAGGCACCATTTA	TGGGAGTCAGTTGTGTCTGC

2.8 Minimum tiling paths

Some of the *fabp* representative BACs belong to DNA fingerprinting contigs had already been mapped to the Atlantic salmon genetic map (Moen et al., 2004; Moen et al., 2008). With the visualization in ASalBase (<http://asalbase.org/sal-bin/index>), I identified BAC-end sequenced BACs that can link the *fabp* representative BACs to the BACs containing the informative microsatellites (minimum tiling paths). To do so, both end sequences of each BAC clones (T7 and Sp6 end sequences) were first masked with all known Atlantic salmon repetitive elements using the cGRASP repeat masker software (http://lucy.ceh.uvic.ca/repeatmasker/cbr_repeatmasker.py) before primers were designed using Primer 3 (<http://frodo.wi.mit.edu/primer3/>). The hot-start colony PCR that is described in Section 2.4.3 was used to test if the positioning of the BACs within the minimum tiling path is indeed as illustrated in ASalBase and the orientation of the insert of each BAC relative to the DNA fingerprinting contig was determined.

2.9 Reverse transcription PCR (RTPCR)

Fourteen tissues were isolated from a male Atlantic salmon smolt with the exception of ovary, which was isolated from a female Atlantic salmon smolt. These tissues are: brain, eye, spleen, heart, gill, ovary, stomach, pyloric caecum, posterior kidney, head kidney, intestine, liver, muscle and testis. In addition, fifteen tissues were isolated from five Atlantic salmon adult individuals. The adult tissues include the same organs that I collected in the smolt samples with the addition of swimbladder. Both smolt and adult Atlantic salmon total mRNAs were isolated using Trizol (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's protocol. In brief, 100 mg of tissue was homogenized thoroughly in 1 mL Trizol using a bead homogenizer and/or

passing the lysate through a 21G needle. The total RNA was separated from the rest of the crude lysate by the phenol: chloroform separation method and was precipitated with isopropanol. The RNA pellet was washed with 70% RNase-free ethanol and was resuspended in 175 μ L RNase-free water in a 60 °C incubator for 10 minutes. The total RNA solution was treated with a DNase digestion using QIAgen RNase-free DNase Set (QIAgen, Mississauga, Ontario, Canada) and purified using QIAgen RNeasy MinElute kit (QIAgen, Mississauga, Ontario, Canada). The total RNA was eluted from the column with 14 μ L of RNase-free water. cDNA synthesis was carried out using M-MuLV Reverse Transcriptase (NEB, Pickering, Ontario, Canada) following the manufacturer's instruction. A 20 μ L of reverse transcription (RT) reaction consisted of 4 μ M oligo dT, 0.5 mM dNTPs, 1 μ g of RNA, 10 U Rnase OUT (Invitrogen, Burlington, Ontario, Canada) and 200 U M-MuLV reverse transcriptase. Hot start PCR was carried out for the PCR portion of RT-PCR. A 10 μ L reaction consisted of 0.2 mM dNTPs, 0.5 μ M of *fabp* gene-specific forward and reverse primers, 5 ng of synthesized single-stranded cDNA or 0.5 μ L of BAC clone glycerol stock or 25 ng of Atlantic salmon genomic DNA. EF1 α has been suggested to be the best reference gene for examining the expression profiling at different life stages and tissues in Atlantic salmon (Olsvik et al., 2005); hence, it is the reference gene for the RT-PCR and quantitative PCR (qPCR) analysis. The forward primer sequence: 5'-CCCCTCCAGGACGTTTACAAA-3' and reverse primer sequence: 5'-CACACGGCCCCACAGGTACA-3'. The PCR products were visualized in a 2% agarose gel.

2.10 Quantitative PCR

qPCR was carried to obtain a more precise measure of the expression level of *fabp* genes. ABI 7900HT Fast Real-Time PCR system and PerfeCTa® SYBR® Green SuperMix, ROX™ (Quanta Biosciences, Gaithersburg, MD, USA), which contained dNTPs, buffer and Taq polymerase, were used. A 25 µL of qPCR reaction consisted of 0.3 µM of forward and reverse primers and 1 µL of template. Six dilutions (5 ng, 2.5 ng, 1.25 ng, 0.625 ng, 0.3125 ng) were used, and each dilution of each gene was carried out in triplicate to account for the variation from pipetting errors between each reaction. Elongation factor 1-alpha (*EF1α*) was chosen as the endogenous control gene as Olsvik et al. suggested that this gene is a good reference gene for qPCR expression analysis in Atlantic salmon (Olsvik et al., 2005). The qPCR cycle used is as follows: 95°C for 3 minutes, followed by 40 cycles of 95°C for 15 seconds, 65°C for 30 seconds, 72°C for 15 seconds. Efficiency ranged from 93% to 104% for the fifteen tissues for *EF1α*. The CT values for the equivalent amount of cDNA ranged from 20 to 22.

2.11 Bioinformatics search for 5' putative transcription factor binding sites and dN/dS analysis

Putative 5' conserved cis-regulatory elements between zebrafish *fabp7b* and Atlantic salmon *fabp7bI* and *fabp7bII* were identified using rVISTA (Loots et al., 2002) (<http://genome.lbl.gov/vista/rvista/submit.shtml>). POU transcription factor binding sites were identified using TFSEARCH with a threshold of 84.5 (Heinemeyer et al., 1998) (<http://www.cbrc.jp/research/db/TFSEARCH.html>) and the DNA motifs in the 5' upstream region of the teleost *fabp7b* genes were defined using TOMTOM (Gupta et al., 2007) (<http://meme.sdsc.edu/meme/cgi-bin/tomtom.cgi>) and MEME (Bailey and Elkan,

1994) (<http://meme.sdsc.edu/meme/intro.html>). For TOMTOM, the IUPAC motifs were used, with JASPAR and UniPROBE as the databases. The comparison function used Pearson correlation coefficient with an e-value less than 10. The normalized non-synonymous (dN) to synonymous (dS) ratios were calculated using SNAP in the HIV database (Korber, 2000) (www.hiv.lanl.gov).

3: Characterization of Atlantic salmon *fabps*

3.1 Identification of Atlantic salmon *fabp* transcript and protein sequences

Since *fabps* have been well characterized in zebrafish (Alves-Costa et al., 2008; Denovan-Wright et al., 2000a; Denovan-Wright et al., 2000b; Karanth et al., 2008; Karanth et al., 2009; Karanth et al., 2009; Liu et al., 2003a; Liu et al., 2003b; Liu et al., 2004; Liu et al., 2007; Pierce et al., 2000; Sharma et al., 2004; Sharma et al., 2006; Venkatachalam et al., 2009), I used these nucleotide coding sequences as queries to determine their orthologs in Atlantic salmon, rainbow trout, sockeye salmon, chinook salmon, brook trout, grayling, northern pike, lake whitefish and rainbow smelt in the cGRASP EST database (http://lucy.ceh.uvic.ca/contigs/cbr_contig_viewer.py). The *fabp* nucleotide coding sequences and the amino acid sequences that I retrieved from the zebrafish database are from Ensembl versions 52-64 with an exception that the zebrafish *fabp10b* coding sequences was retrieved from NCBI. The cGRASP EST accession numbers change whenever the database is upgraded and it is impossible to retrieve the sequence webpages again unless one uses the cGRASP BLASTn search with the EST sequences as queries in the upgraded database. Hence, I used the unique full-length *fabp* EST sequences and BLASTn searched to the NCBI EST database to obtain their corresponding GenBank accession numbers. Table 3.1 lists the GenBank accession numbers of the unique full-length Atlantic salmon *fabp* ESTs found in the cGRASP EST database. The GenBank accession numbers of the unique full-length rainbow trout,

sockeye salmon, chinook salmon, brook trout, grayling, northern pike, lake whitefish and rainbow smelt *fabp* ESTs found in the cGRASP EST database are stated in Chapters 4-10. Unfortunately, there were not many full-length *fabp* ESTs in sockeye salmon, chinook salmon, brook trout, grayling, northern pike, lake whitefish and rainbow smelt. This might be due to the combination of the number of available ESTs of these species in the database, the number of tissues that were used for making the EST libraries and the number of transcribed *fabps* in these fish species. In fact, as of November 16th, 2011, the number of ESTs in Atlantic salmon, rainbow trout, sockeye salmon, chinook salmon, grayling, brown trout, lake whitefish, northern pike and rainbow smelt databases are 434384, 275413, 12056, 14535, 10975, 10051, 10842, 32908 and 36758, respectively (Koop et al., 2008; Leong et al., 2010).

Table 3.1. Atlantic salmon full-length *fabp* EST sequences and their GenBank accession numbers as of February 16th, 2012.

The gene names were based on their positions in the phylogenetic trees (see Chapters 4 – 10).

	Gene Name	Atlantic salmon <i>fabp</i> EST sequences
<i>fabp1</i>	<i>fabp1a</i>	CK890396.1
	<i>fabp1b</i>	EG888201.1
<i>fabp2</i>	<i>fabp2aI</i>	EG790658.1
	<i>fabp2aII</i>	CB503762.1
	<i>fabp2b</i>	CB509140.1
<i>fabp3</i>	<i>fabp3aI</i>	GO063695.1
	<i>fabp3aII</i>	EG804768.1
<i>fabp6</i>	<i>fabp6aI</i>	CB505758.1
	<i>fabp6aII</i>	CB504199.1
	<i>fabp6b</i>	EG845337.1
<i>fabp7</i>	<i>fabp7bI</i>	EG797585.1
	<i>fabp7bII</i>	EG828698.1
<i>fabp10</i>	<i>fabp10aI</i>	BG935343.1
	<i>fabp10aII</i>	CA037797.1
	<i>fabp10b</i>	CA037369.1
<i>fabp11</i>	<i>fabp11aI</i>	EG792265.1
	<i>fabp11aII</i>	EG835123.1
	<i>fabp11b</i>	EG779590.1

3.2 Identification of teleost and tetrapod *fabp* genomic, transcript and protein sequences

In order to understand the evolution of the *fabp* gene family and how Atlantic salmon *fabps* are related to one another and to their orthologs in other teleosts and in tetrapods, I searched for the *fabp* genomic, transcript and protein sequences in other teleosts and in tetrapods. As previously, I used the zebrafish *fabp* nucleotide coding sequences as queries to data-mine their orthologs in stickleback, medaka, fugu, tetraodon, chicken, mouse, rat and human. The *fabps* in the four sequenced teleosts (stickleback, medaka, tetraodon and fugu) and in the tetrapods were identified using the BLAST/BLAT search in Ensembl with the zebrafish *fabp* coding sequences as queries. Table 3.2 lists the Ensembl accession numbers of the *fabps* that were found in the zebrafish, stickleback, medaka, fugu, tetraodon, chicken, mouse, rat and human genomes. However, some of these *fabp* genes, such as stickleback *fabp3*, fugu *fabp1*, stickleback *fabp6s*, medaka *fabp6b*, tetraodon *fabp6a* and *fabp11b*, required re-annotation. This was carried out using GenScan prediction (<http://genes.mit.edu/GENSCAN.html>) or using the EST sequences retrieved from the The Gene Index Project (TIGR) database version 9 (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>) (see Chapters 6 - 10) (Quackenbush et al., 2001). The inferred FABP amino acid sequences were obtained using ExPASy translate tool (<http://web.expasy.org/translate/>).

Table 3.2. Ensembl accession numbers in version 64 of different *fabps* in the five genome-sequenced teleost species that were examined in this thesis.

* denotes genes that I re-annotated with EST sequences retrieved from TIGR. ^ represents genes that were re-annotated by Genscan. # denotes gene that I annotated by using a EST sequence retrieved from NCBI and blastn searched against the genomic sequence of the genome. + represents with gene that is in a region where a completely different gene has been annotated. The gene names were based on their positions in the phylogenetic trees (see Chapters 4 – 10).

<i>fabp</i> -subfamily	Zebrafish (<i>Danio rerio</i>)	Stickleback (<i>Gasterosteus aculeatus</i>)	Medaka (<i>Oryzias latipes</i>)	Fugu (<i>Takifugu rubripes</i>)	Tetraodon (<i>Tetraodon nigroviridis</i>)
<i>fabp1</i>	ENS DARG00000019357 (<i>fabp1a</i>)	ENS GACG00000006574 (<i>fabp1b</i>)	ENS ORLG00000003820 (<i>fabp1b</i>)	*ENS TRUG000000015614 (<i>fabp1b</i>)	
	ENS DARG000000059227 (<i>fabp1b.1</i>)				
	ENS DARG000000095866 (<i>fabp1b.2</i>)				
<i>fabp2</i>	ENS DARG00000006427 (<i>fabp2a</i>)	ENS GACG00000017790 (<i>fabp2a</i>)	ENS ORLG00000019478 (<i>fabp2a</i>)	ENS TRUG00000014256 (<i>fabp2a</i>)	ENS TNIG00000012360 (<i>fabp2a</i>)
		ENS GACG00000019474 (<i>fabp2b</i>)	ENS ORLG00000004146 (<i>fabp2b</i>)	ENS TRUG00000000040 (<i>fabp2b</i>)	
<i>fabp3</i>	ENS DARG00000023290	^ENS GACG0000000055	ENS ORLG00000011573	ENS TRUG00000013298	ENS TNIG00000016979
<i>fabp6</i>	ENS DARG00000044566 (<i>fabp6a</i>)	^ENS GACG00000018140 (<i>fabp6a</i>)	ENS ORLG00000005456 (<i>fabp6a</i>)	ENS TRUG00000001091 (<i>fabp6a</i>)	^ENS TNIG00000004652 (<i>fabp6a</i>)
		^ENS GACG00000014354 (<i>fabp6b</i>)	*ENS ORLG00000012622 (<i>fabp6b</i>)	ENS TRUG00000004369 (<i>fabp6b</i>)	ENS TNIG00000017612 (<i>fabp6b</i>)
<i>fabp7</i>	ENS DARG00000007697 (<i>fabp7a</i>)	ENS GACG00000010166 (<i>fabp7a</i>)	ENS ORLG00000013475 (<i>fabp7a</i>)	ENS TRUG00000014810 (<i>fabp7a</i>)	ENS TNIG00000017217 (<i>fabp7a</i>)
	ENS DARG00000034650 (<i>fabp7b</i>)	+ENS GACG00000012080 (<i>fabp7b</i>)	ENS ORLG00000010773 (<i>fabp7b</i>)	ENS TRUG00000011334 (<i>fabp7b</i>)	ENS TNIG00000003393 (<i>fabp7b</i>)
<i>fabp10</i>	ENS DARG000000083439 (<i>fabp10a</i>)	ENS GACG00000013639 (<i>fabp10a</i>)	ENS ORLG00000014794 (<i>fabp10a</i>)	# (<i>fabp10a</i>)	ENS TNIG00000002424 (<i>fabp10a</i>)

<i>fabp</i> -subfamily	Zebrafish (<i>Danio rerio</i>)	Stickleback (<i>Gasterosteus aculeatus</i>)	Medaka (<i>Oryzias latipes</i>)	Fugu (<i>Takifugu rubripes</i>)	Tetraodon (<i>Tetraodon nigroviridis</i>)
<i>fabp10</i>	ENSDARG00000091101 (<i>fabp10b</i>)	ENSGACG00000002234 (<i>fabp10b</i>)	ENSORLG00000007702 (<i>fabp10b</i>)		
<i>fabp11</i>	ENSDART00000021798 (<i>fabp11a</i>)	ENSGACG00000003457 (<i>fabp11a</i>)	ENSORLG00000010830 (<i>fabp11a</i>)	ENSTRUG00000014661 (<i>fabp11a</i>)	ENSTNIG00000017781 (<i>fabp11a</i>)
	ENSDARG00000002311 (<i>fabp11b</i>)	ENSGACG00000008725 (<i>fabp11b</i>)	ENSORLG00000008282 (<i>fabp11b</i>)	ENSTRUG00000013523 (<i>fabp11b</i>)	*ENSTNIG00000018400 (<i>fabp11b</i>)

Table 3.3. Ensembl accession numbers in version 64 of different *fabps* in four tetrapod species that were examined in this thesis.

<i>fabp</i> -subfamily	Chicken (<i>Gallus gallus</i>)	Mouse (<i>Mus musculus</i>)	Rat (<i>Rattus norvegicus</i>)	Human (<i>Homo sapiens</i>)
<i>fabp1</i>	ENSGALG00000015937	ENSMUSG000000054422	ENSRNOG00000006675	ENSG000000163586
<i>fabp2</i>	ENSGALG00000011986	ENSMUSG00000023057	ENSRNOG00000024947	ENSG000000145384
<i>fabp3</i>	ENSGALG00000000620	ENSMUSG00000028773	ENSRNOG00000012879	ENSG000000121769
<i>fabp4</i>	ENSGALG00000015767	ENSMUSG000000062515	ENSRNOG00000010805	ENSG000000170323
<i>fabp5</i>	ENSGALG00000015764	ENSMUSG00000027533	ENSRNOG00000021811	ENSG000000164687
<i>fabp6</i>	ENSGALG00000001445	ENSMUSG00000020405	ENSRNOG00000003902	ENSG000000170231
<i>fabp7</i>	ENSGALG00000014866	ENSMUSG00000019874	ENSRNOG00000000814	ENSG000000164434
<i>fabp8</i>	ENSGALG00000019459	ENSMUSG000000052468	ENSRNOG00000022707	ENSG000000147588
<i>fabp9</i>	Not found	ENSMUSG000000027528	ENSRNOG000000011082	ENSG000000205186
<i>fabp10</i>	ENSGALG00000004141	Not found	Not found	Not found
<i>fabp12</i>	Not found	ENSMUSG000000027530	ENSRNOG000000038825	ENSG000000197416

3.3 Genomic sequences and chromosomal locations of Atlantic salmon *fabps*

3.3.1 Checking the specificity of Atlantic salmon *fabp* gene-specific primers

Prior to obtaining the genomic sequences of Atlantic salmon *fabps*, I designed gene-specific primers from each unique full-length EST sequences. To design gene-specific primers for genes of each *fabp* sub-family, I initially aligned the Atlantic salmon unique full-length EST sequences and the corresponding zebrafish coding sequences, and identified gene-specific regions and exon-exon boundaries within the Atlantic salmon *fabp* EST sequences. Here, I will demonstrate the primer designing procedure for members of the *fabp2* sub-family. Figure 3.1 illustrates where the Atlantic salmon *fabp2* gene-specific primers were designed from in the alignment of the *fabp2* ESTs. To test the specificity of these primers, I sequenced the individual single genomic amplicons of the *fabp* gene-specific primer sets. I then used these sequencing reads as queries to search for their identities via BLASTn searches in NCBI. The amplification specificity testing was carried out for all *fabp* gene-specific primers that are listed in Table 2.2. Therefore, these primers could be used to identify which *fabp* genes the hybridization-positive BAC clones contain and to carry out the expression studies.

Figure 3.1. ClustalW alignment of Atlantic salmon *fabp2* unique full-length ESTs.

Lines indicate where the 5'UTR (untranslated region) and exon one, the exon-exon boundaries, exon four, the stop codon and the 3' UTR are based on alignment with the zebrafish *fabp2* (ENSDART00000027973). The lime and the teal highlighted regions are where the *fabp2* forward and reverse primers would align, respectively.


```

                                     5'UTR
Ssa_fabp2aI      ----GCACG-----AGGGCCAGTCCAACCCTTCCAGCAACCATGACCTACAATGGC 47
Ssa_fabp2aII    GGGGGGACTCCTTAGAGTAGAGCCAGTCCAACCCTTCCAGCAACCATGACCTACAACGGC 60
Ssa_fabp2b      ----GCAC-----GAGGTGCTC--ACC-----GTAGACATGGCATTCAACGGT 37
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      ACTTGGAAAGTAGACCACAGT GAGAACTATGAGAAATTCATGGAGCAAATGSGTGTCAAC 107
Ssa_fabp2aII    ACTTGGAAAGTAGACCAGCGAG GAGAACTATGAGAAATTCATGGAGCAGATGSGTGTCAAC 120
Ssa_fabp2b      ACCTGGAAAGTTGAACGCAACGAGAACTATGACAAGTTCATGGAGCAGATGGCATCAAT 97
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      ATGGTCAAGAGGAAGCTGGCTACTCAGATAACCTCAAGATCACCCCTTGAACAAACTGGA 167
Ssa_fabp2aII    ATGGTCAAGAGGAAGCTGGCCGCTCAGATAACCTCAAGATCACCCCTTGAACAAACTGGA 180
Ssa_fabp2b      GTGATGAAGAGGAAGTTAGCTGAACATGACAACTCAAGGTATCATCGAGCAGACAGGA 157
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      GACAAGTTTGTCTGTAAGGAGGCCAGTACTTTCCGCACGCTGGATATGGAATTTACCCTG 227
Ssa_fabp2aII    GACAATTTTGTCTGTAAGGAGGCCAGTCTTTCCGCACGCTGGATATGGAATTTACCCTG 240
Ssa_fabp2b      GACAAGTTCCACATCAAGGAATCCAGCACCTTCGGTAACAAAGACATCGACTTCACCCTG 217
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      GGAGTCACCTTTGAGTATGCCCTTGCAGATGGGACAATGCTATCAGTTCATGGGGCATG 287
Ssa_fabp2aII    GGAGTCACCTTTGAATATGCTCTTGCAGATGGGACAATGCTATCAGTTCATGGGGCATG 300
Ssa_fabp2b      GGCCTACAGTTTGACTATGCACTGGCCGACGGCACCGAAGTCTCAGTGCCTGGGAGATG 277
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      GAAGGAGACATGATGAAAGGTACATTCACCAGAAAGGACAAATGGAAAGGTGCTGACAACT 347
Ssa_fabp2aII    GAAGGAGACATGATGAAAGGTACATTCACCAGAAAGGACAAATGGAAAGGTGCTGACAACT 360
Ssa_fabp2b      GATGGAGATGTGTTGAAGGGTAAATTCACCAGAAAGGACAAACAGTAAGGTGCTGACCACC 337
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      ACCAGAGCCATTGTTGGAGATGAACTTGTACAGAGCTACAGCTACGATGGAGTCGAAGCC 407
Ssa_fabp2aII    ACCAGAGCCATTGTTGGAGAGGAACTTGTACAGAGCTACAGCTATGATGGAGTCGAAGCC 420
Ssa_fabp2b      ACTCGGGCCCCTGGTGGGAGGAGAGCTAGTACAGCTATAACTACGATGGAGTGGATGCC 397
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Stop codon and 3'UTR
Ssa_fabp2aI      AAGAGAATTTTCAAGAGGGGTTAGAGGAAACC--GATCAAC--CAATC-CA---ACCAC 459
Ssa_fabp2aII    AAGAGAATTTTCAAGAGGGGTTAGAGGAAACC--GATCAA--CAATC-CA---ACCAC 472
Ssa_fabp2b      AAGAGGATTTTCAAGAAGGCTTAAAGGGATTTCATGATGAATTTCTATCTCACATACATT 457
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      TAG-----ACC--AAGAG-----GCGTTACCTTTAA-T-CCAGAATGATAC 496
Ssa_fabp2aII    AAG-----ACC--AAGAA-----GAGCTACGTTCAAAT-CCAGAATTATAC 510
Ssa_fabp2b      ACGTTTTTCTAACCTTAAGAATATGTGGCAAGACTTTGGCTCAGGTGCCAAATTTAAAG 517
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      CATTGTGACACTGAACATTTTACAA--CCTT---CCCTAAGCCATTGGACCACTTCAGGT 551
Ssa_fabp2aII    CATT-----TTACAA--CCTTAAGCCTTAAGCCACTGGACCACTTTAGGT 553
Ssa_fabp2b      TATAAGTCAA-----AATTGTCAGTGCCTT--CCCCTCATTCTCTAAACC-CCATGTAC 568
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      CTTTG-----TGACGCCAACA--TAATTATTAGCTGTACAAAT-GTACTATTTGTAATA 603
Ssa_fabp2aII    CTTTG-----TAATGCCAACA--TCATTTTATAGCTGTACAAAGAT-GCACTGTTTGTAAATA 605
Ssa_fabp2b      CCCCRAAAAATAGCTCCAATAGTTCACTATTGGATG---AAGACCGTATCACCCCTCCATA 625
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      TAAATATATAAAGCAGG-----TT-----TGATTGTC-----ACTTCAAGTT 640
Ssa_fabp2aII    GAAATGTGTAAAGCAGG-----TT-----CGATTGTC-----ACTTCAAGTT 642
Ssa_fabp2b      --AATTCATAAAGCAAGGAACATTATGTTGGCTGATATACTCCAGAGTGACTACAACCT 683
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Ssa_fabp2aI      TTGATATTTTAAAGAAGTAAACATGTTTTTGAATTATT 677
Ssa_fabp2aII    TTGATATTTTAAAGAAGTAAACATGTTTTTGAATT--- 676
Ssa_fabp2b      TTGAACAATCAAGAT-TAACTCTCCACTGAATCCAG 719
                 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

3.3.2 Identification of BACs with *fabps* via Atlantic salmon BAC library screening

To characterize the Atlantic salmon *fabp* gene structures and to identify the location of these genes in the genetic map, I screened the Atlantic salmon genomic BAC library to identify BAC clones containing the *fabp* of interest. Here I will demonstrate the identification of BACs with *fabp6s*. Figure 3.3 is a phosphor screen image of Atlantic salmon BAC library filter 12 after being hybridized with the *fabp6* oligonucleotide probes. The identity of the hybridization-positive BAC clones could be retrieved by the locations and the patterns of the doublet dots, which have similar radioactive signals. Based on the image in Figure 3.3, the *fabp6* hybridization-positive BAC clones from filter 12 are BACs S0533M19, S0556D15, S0556O10, S0565D06, S0570O21 and S0575A11.

Next, the hybridization-positive BAC clones were tested by PCR amplification with the gene-specific primer sets of the same *fabp* sub-family as the oligonucleotide probes were used for the hybridization. Therefore, in this demonstration example, I used the *fabp6* gene-specific primers to test the *fabp6* hybridization-positive BAC clones. Figure 3.4 is a gel electrophoresis image showing the PCR reactions with the three *fabp6* gene-specific primers and the *fabp6* hybridization-positive BAC clones as the templates. Based on Figure 3.4, BAC clones S0565D06 and S0874K14 are PCR-positive for *fabp6aI*, BAC clones S0361I22, S0513F01 and S0556O10 are PCR-positive for *fabp6aII* and BAC 134A09, S0487M09 and S0556D15 contain the *fabp6b* gene. After screening the BAC library for BAC clones containing *fabp* genes, I retrieved the information of the DNA fingerprinting contig where these hybridization and PCR-positive BAC clones belong to from ASalBase (www.asalbase.org). The information includes the locations of

these contigs if they had previously been mapped to the Atlantic salmon genetic map, and the BAC-end sequences of other BAC clones that are within the same contigs as the *fabp* hybridization and PCR-positive BACs. In fact, these BAC-end sequences of other BAC clones might be useful in some cases for identifying putative markers for genetic mapping of some other *fabps* (see Section 3.3.5). Figure 3.5 is a screenshot of ASalBase showing the DNA fingerprinting contig 2577, the contig where *fabp6b* resides. Finally, one of the hybridization and PCR-positive BACs for a particular *fabp* was chosen to be the representative BAC for that gene. The criteria are: 1) its location relative to the DNA fingerprinting contig (it preferably locates in the middle of the contig) and 2) its BAC ends had been previously sequenced. Therefore, BAC clone S0487M09, was chosen to be the representative BAC for *fabp6b*. Table 3.3 summarizes the representative BAC clones for each *fabp* members in Atlantic salmon for this thesis.

Unfortunately, I could not find any BAC clones that are both hybridization and PCR-positive for any of the *fabp11s*. This may reflect the BAC library not completely covering the whole Atlantic salmon genome because some parts of the genome were not clonable into BAC vector. Hence, the characterization of *fabp11* sub-family was based on the available sequence reads from the Atlantic salmon genome sequencing project (Davidson et al., 2010) (see Section 3.3.4).

Figure 3.2. A phosphor screen image of Atlantic salmon genomic BAC library filter twelve that had been hybridized with *fabp6* and *overgo* (positive control) oligonucleotide probes.

Blue and black circles highlight the positive control with the *overgo* probe and the actual hybridization-positive doublet dots, respectively. The yellow circle is an example of a pair of false-positive doublet dots.

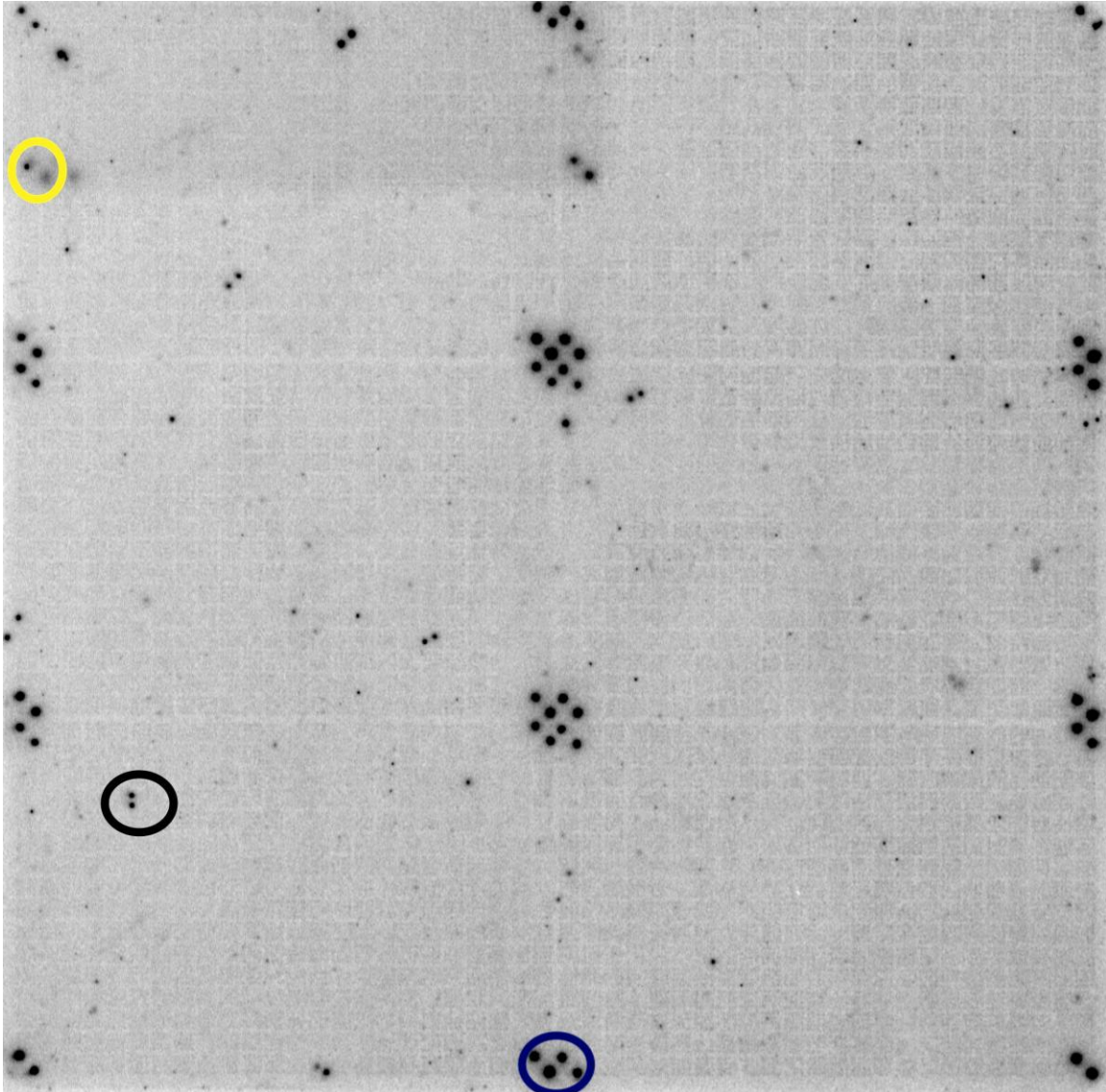


Figure 3.3. A 2% agarose gel image of PCR testing on *fabp6* hybridization-positive BAC clones with *fabp6aI*, *fabp6aII* and *fabp6b* gene-specific primers.

Lanes 1, 26, 27, 52 and 62 were loaded with 2.5 μ L Bioline HyperLadderII. Lanes 2 to 21 were PCR reactions with *fabp6aI* gene-specific primers. Lanes 22 to 25, 28 to 41 are PCR reactions with *fabp6aII* gene-specific primers; and lanes 42 to 51, 53 to 61 were PCR reactions with *fabp6b* gene-specific primers. Template used for lanes 2, 22 and 42 was BAC S0361I23. Template used for lanes 3, 23, 43 was BAC S0134A09. Template used for lanes 4, 24, 44 was BAC S0185F12. Template used for lanes 5, 25, 45 was S0275H13. Template used for lanes 6, 27, 46 was BAC S0426J17. Template used for lanes 7, 28, 47 was BAC S0465N05. Template used for lanes 8, 29, 48 was BAC S0487M09. Template used for lanes 9, 30, 49 was BAC S0513F01. Template used for lanes 10, 31, 50 was BAC S0533M19. Template used in lanes 11, 32 and 51 was BAC S0556D15. Template used in lanes 12, 33 and 53 was BAC S0556O10. Template used in lanes in 13, 34 and 54 was BAC S0565D06. Template used in lanes 14, 35 and 55 was BAC S0570O21. Template used in lanes in 15, 36 and 56 was BAC S0575A11. Template used in lanes in 16, 37 and 57 was BAC S0871O11. Template used in lanes 17, 38 and 58 was BAC S0874K14. Template used in lanes in 18, 39 and 59 was BAC S0896E03. Lanes 19, 40 and 60 were negative controls and lanes 20, 41 and 61 were positive controls with Atlantic salmon genomic DNA as template for *fabp6aI*, *fabp6aII* and *fabp6b* PCR reactions, respectively.

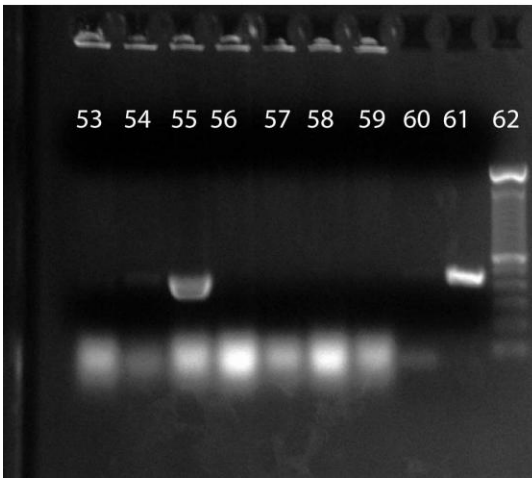
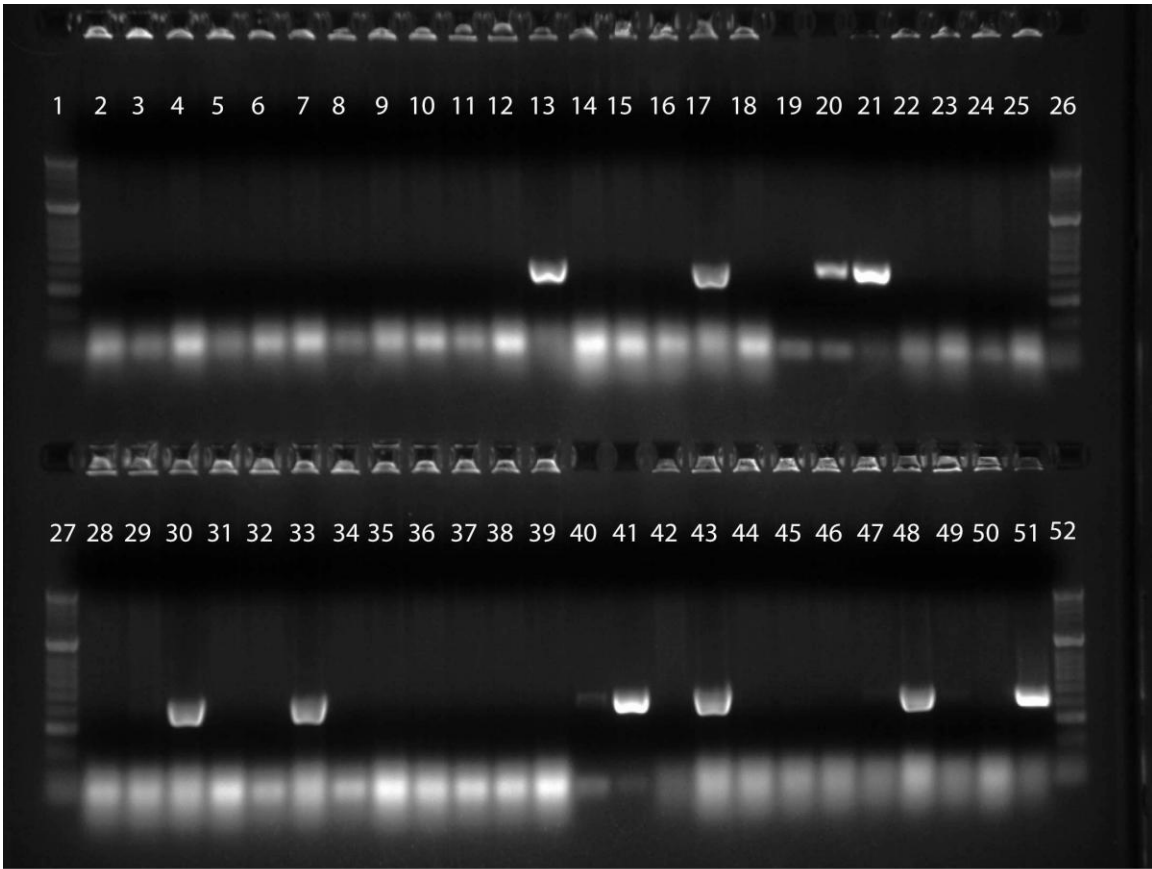


Figure 3.4. A screenshot of ASalbase showing the DNA fingerprinting contig 2577. BAC clone S0487M09 (*), the representative BAC clone of Atlantic salmon *fabp6b*, belongs to this contig.

This screenshot was taken on July 12th, 2010. The insert sizes of the BAC clones are represented by the lengths of the rectangles. The red rectangles appearing in some of the BAC clones denote that the BAC-end sequences of the corresponding BACs are available. Since, this DNA fingerprinting contig has not been mapped to the Atlantic salmon genetic map, one method I used was to search for putative microsatellites from the BAC-end sequences in order to identify the chromosomal location of *fabp6b* (see Section 3.35).

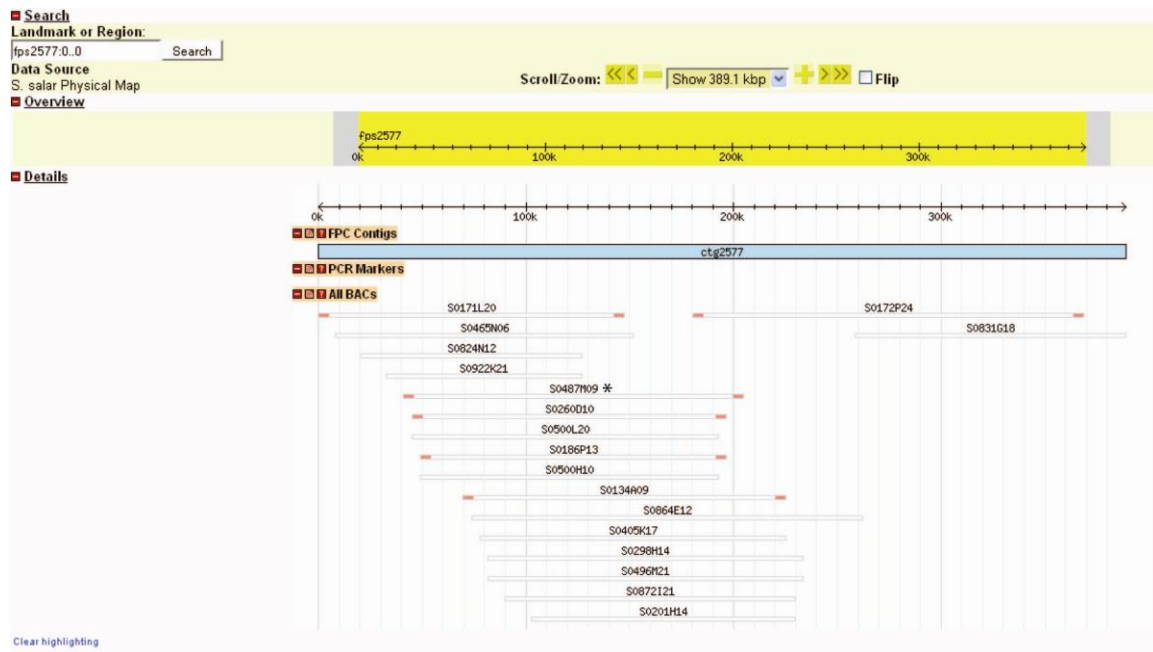


Table 3.4. Summary of all the *fabps* identified in Atlantic salmon in this thesis.

<i>fabp</i> -subfamily	<i>fabp</i> oligonucleotide probe used	<i>fabp</i> gene	Representative BAC clones	Corresponding DNA fingerprinting config	Informative microsatellite	Linkage group	Chromosome	GenBank Accession number
<i>fabp1</i>	<i>fabp1a</i> and <i>fabp1b</i>	<i>fabp1a</i>	S0277M02	310	Ssa0313BSFU	9	Ssa11	
		<i>fabp1b</i>	S0250M23	900	Ssa1469BSFU	7	Ssa24	
<i>fabp2</i>	<i>fabp2aI</i> and <i>fabp2aII</i>	<i>fabp2aI</i>	S0065J23	988	Ssa10060BSFU	19	Ssa8	EU880417
		<i>fabp2aII</i>	S0252E04	339	Ssa10056BSFU	28	Ssa4	EU880418
		<i>fabp2b</i>	S0938E16	6545	Ssa10082BSFU	24	Ssa7	EU880418
<i>fabp3</i>	<i>fabp3aI</i> and <i>fabp3aII</i>	<i>fabp3aI</i>	S0212F03	Singleton	Ssa10087BSFU	15	Ssa27	GU985445
		<i>fabp3aII</i>	S0812G14	Singleton	Ssa10088BSFU	3	Ssa14	GU985446
<i>fabp6</i>	<i>fabp6aI</i> and <i>fabp6aII</i>	<i>fabp6aI</i>	S0874K14	6110	Ssa10145BSFU	28	Ssa4	
		<i>fabp6aII</i>	S0269H18	1625	Ssa10154BSFU	5	Ssa13	
		<i>fabp6b</i>	S0487M09	2577	Ssa10142BSFU	10	Ssa9	
<i>fabp7</i>	<i>fabp3aI</i> and <i>fabp3aII</i>	<i>fabp7bI</i>	S0232I02	798	Ssa0907BSFU	8	Ssa15	JF421127
		<i>fabp7bII</i>	S0025A21	756	Ssa0890BSFU	4	Ssa6	JF421128
<i>fabp10</i>	<i>fabp10aI</i> , <i>fabp10aII</i> and	<i>fabp10aI</i>	S0131N18	2169	Ssa0015SSFU	1	Ssa2	
		<i>fabp10aII</i>	S0016L02	7333	Ssa2795BSFU	12	Ssa5	

<i>fabp</i> -subfamily	<i>fabp</i> oligonucleotide probe used	<i>fabp</i> gene	Representative BAC clones	Corresponding DNA fingerprinting contig	Informative microsatellite	Linkage group	Chromosome	GenBank Accession number
<i>fabp10</i>	<i>fabp10b</i>	<i>fabp10b</i>	S0183M09	881	Ssa10143BSFU3		Ssa14	
<i>fabp11</i>	<i>fabp11as</i> and <i>fabp11b</i>	<i>fabp11aI</i>						
		<i>fabp11aII</i>			Ssa10156BSFU8		Ssa15	
		<i>fabp11b</i>			Ssa10155BSFU1 and 5			

3.3.3 Genomic sequences of *fabps* via sequencing reads from BAC shotgun libraries

As mentioned in Chapter 2, the Atlantic salmon genome sequencing project had not begun when I started characterizing the *fabp2* and *fabp3* gene sub-families, so I made shotgun libraries of the representative BAC clones to obtain the genomic sequences of these genes. Here I will describe the BAC shotgun library production process with the *fabp2b* representative BAC clone, S0938E16 (Lai et al., 2009). Initially, I randomly sheared the isolated BAC DNA using sonication. Figure 3.6 illustrates the shearing quality as the duration of sonication increased. The smearing in the sheared DNA sample lanes reflects different sizes of DNA fragments. As the sonication duration increased, the smearing shifted more downwards towards to the bottom of the gel, indicating that the overall DNA fragment sizes were decreasing. The goal of shearing was to randomly break the isolated BAC DNA into 2 kb - 5 kb fragments so that they could be sub-cloned into the vector. Figure 3.7 is an image of size fractionalization showing the 2 kb – 5 kb end-repaired BAC S0938E16 DNA fragments, which were excised from the agarose gel. After I sub-cloned the end-repaired DNA fragments, I randomly picked 64 colonies and isolated the recombinant plasmid DNA. I performed a PvuII restriction digestion on them to check the insert sizes (Figure 3.8) and used the recombinant plasmid DNA as templates for sequencing reactions to test the identity of the inserts. The purpose of this test was to ensure that there was not a significant amount of bacterial genomic DNA as insert in the shotgun library clones before the shotgun library (~2300 colonies) was sent to the Michael Smith Genome Sciences Centre for sequencing. The GenBank accession numbers of the assembled consensus sequences of the *fabp2* and *fabp3* BAC shotgun libraries are given in Table 3.3.

Figure 3.5. An 1% agarose gel image reflects the quality of the shearing progress on the isolated BAC S0938E16 DNA.

The gel was ran at 200V for 15 min. Lane 1 was loaded with 1 kb DNA ladder. Lane 2, 3, 4 and 5 were loaded with 2 μ L of BAC S0938E16 DNA that had been sheared by sonication for 2 sec, 4 sec, 6 sec and 8 sec, respectively.

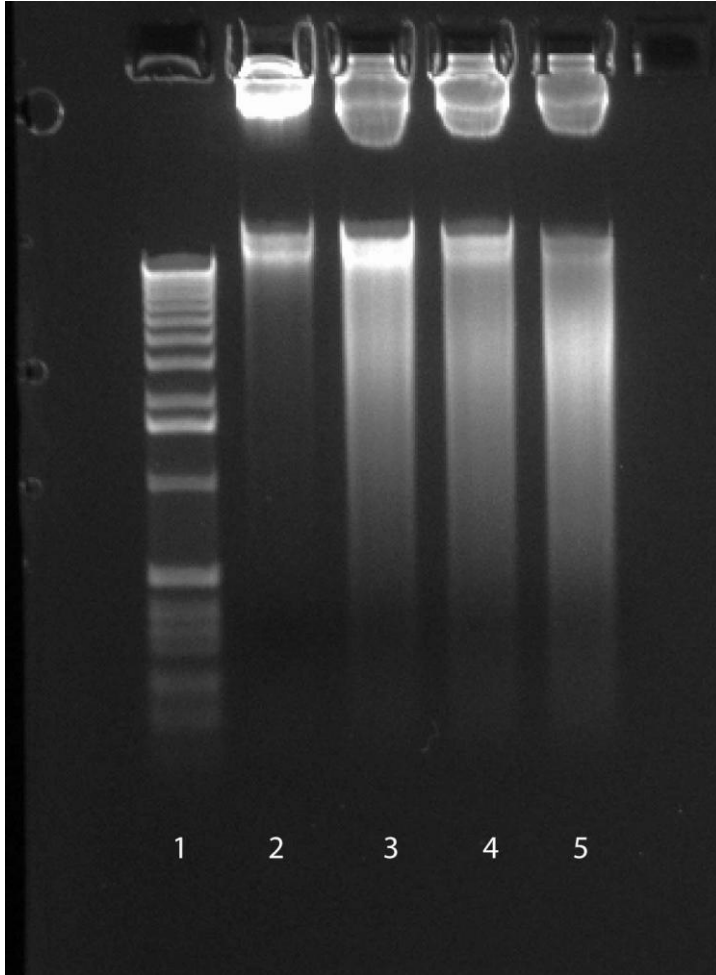


Figure 3.6. Size fractionalization of BAC S0938E16.

Lanes 1 and 3 were loaded with 1 kb DNA ladder. Lane 2 was loaded with the sheared, end-repaired BAC S0938E16 DNA. The black box in the smear of lane 2 illustrates the removal of the piece of agarose gel that contained DNA fragments ranging from 2 kb to 5 kb.

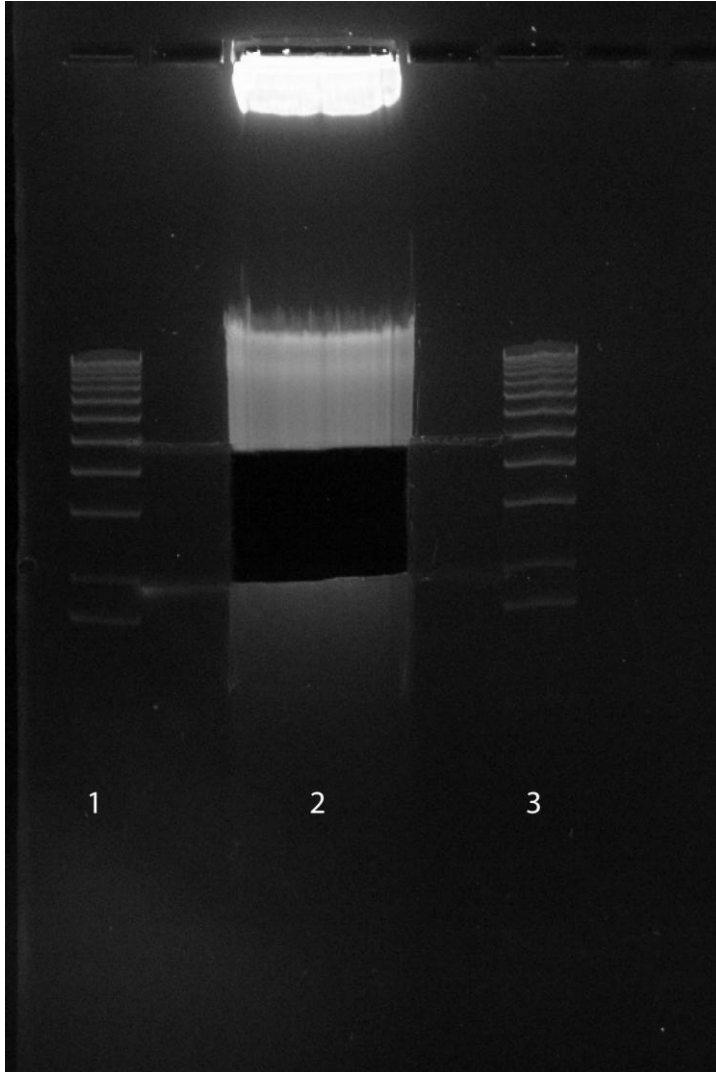
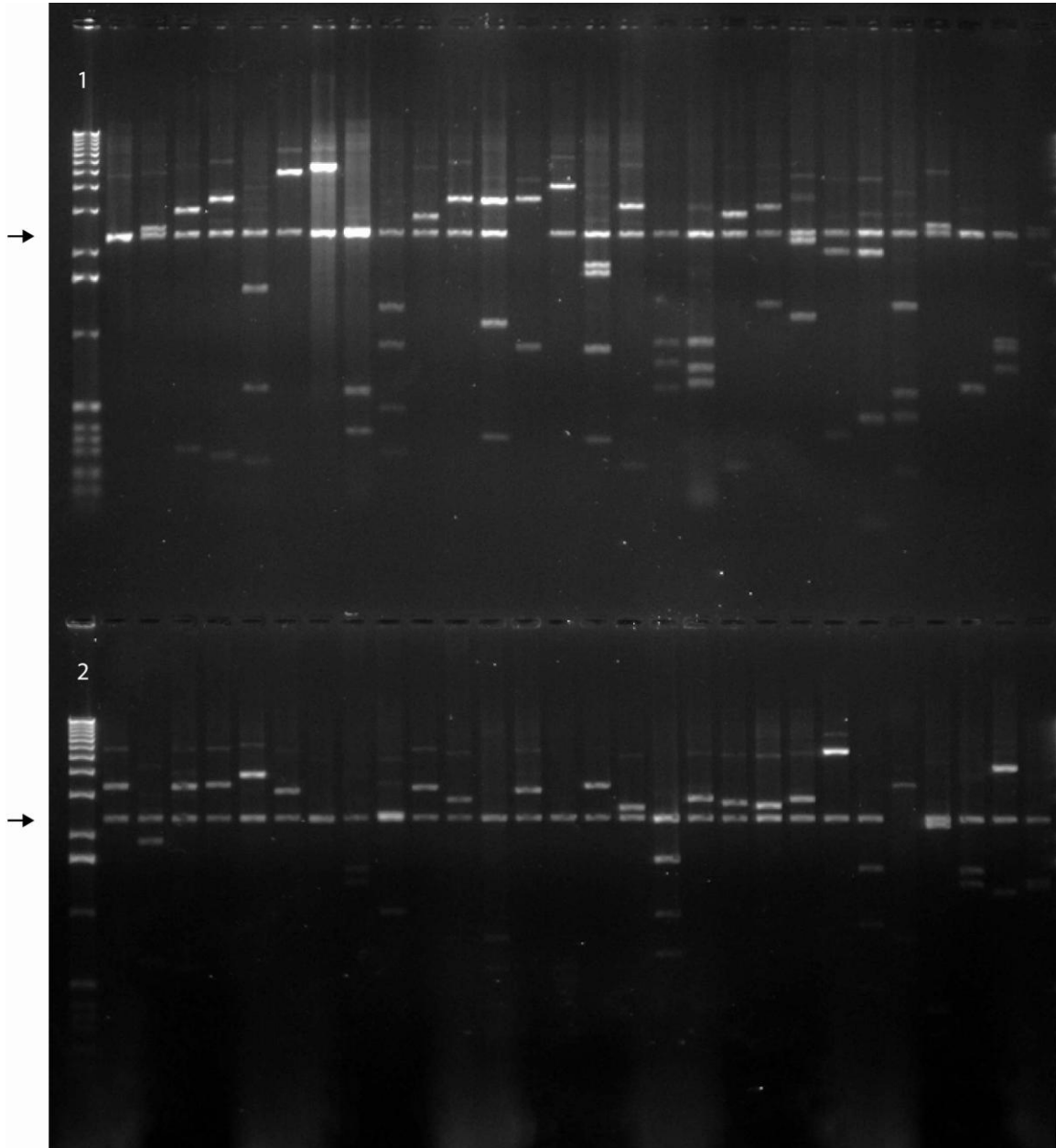


Figure 3.7. An agarose gel image of testing the clone insert sizes of the BAC S0938E16 shotgun library.

Lanes 1 and 2 were loaded with 5 μ L of 1 kb DNA ladder. The arrows indicate where the linearized vector DNA fragment were on the agarose gel. Band The other band sizes in each sample lanes should sum up to 2 to 5 kb size of DNA fragment as I only inserted DNA fragments ranging from 2 to 5 kb into the vector.



3.3.4 Genomic sequences of *fabps* via sequencing reads from Atlantic salmon genome sequencing project

As mentioned in Chapter 2, the characterization of *fabp1*, *fabp6*, *fabp7*, *fabp10* and *fabp11* gene sub-families was based on the sequencing reads released from the Atlantic salmon genome sequencing project (Davidson et al., 2010). The CIGENE scaffolds were assembled using sequence data from the Atlantic salmon genome sequencing project, which corresponded to a three fold coverage of the whole genome, and were available in the private ASalBase BLASTn search around March 2011. To identify CIGENE scaffolds containing *fabps*, I used Atlantic salmon *fabp* coding sequences as queries and searched for the highest scaffold hits. Table 3.4 lists the scaffold hits of each Atlantic salmon *fabp*. Regardless of the consensus sequences of the assembled BAC shotgun libraries or the *fabp* CIGENE scaffold sequences, they were all subjected to annotation using the cGRASP annotation pipeline. In addition, I was able to investigate the *fabp* locations in the genome for the genes that had not been mapped in the Atlantic salmon genetic map using the BAC shotgun library assembled consensus sequences or the *fabp* CIGENE scaffold sequences.

Table 3.5. Summary of *fabps* and their corresponding CIGENE scaffolds. Note that CIGENE scaffold 151259024 only covers exons one and two.

<i>fabp</i> sub-family	Gene Name	CIGENE scaffold
<i>fabp1</i>	<i>fabp1a</i>	scf15149518
	<i>fabp1b</i>	scf15135926
<i>fabp6</i>	<i>fabp6aI</i>	scf15148348
	<i>fabp6aII</i>	scf15259024 (only covered exons one and two)
	<i>fabp6b</i>	scf15123132
<i>fabp7</i>	<i>fabp7bI</i>	scf15194177
	<i>fabp7bII</i>	scf15188937
<i>fabp10</i>	<i>fabp10aI</i>	scf15191969
	<i>fabp10aII</i>	scf15194866
	<i>fabp10b</i>	scf15193889
<i>fabp11</i>	<i>fabp11aI</i>	scf15101812
	<i>fabp11aII</i>	scf15138436
	<i>fabp11b</i>	scf15169353

3.3.5 Genetic mapping of Atlantic salmon *fabps*

Given the available information of the integration of the Atlantic salmon genetic and physical maps (Phillips et al., 2009), all *fabps* except for *fabp11aI* were mapped to the genetic map by genotyping and linkage analysis (Table 3.3). The putative microsatellites were identified within: 1) the consensus assembled sequences of the BAC shotgun libraries (*fabp2* and *fabp3* sub-families); 2) the BAC-end sequences of BACs within the same DNA fingerprinting contigs as the *fabp* representative BACs (*fabp6aII* and *fabp10aII*); 3) the sequences of the *fabp* CIGENE scaffolds (*fabp6aI*, *fabp6b* and *fabp10b*). However, there was no putative microsatellite found in the *fabp11aII* and *fabp11b* CIGENE scaffolds. Therefore, I used the sequences of these scaffolds as queries and BLASTn searched against the BAC-end sequences in ASalBase. I then retrieved the DNA fingerprint contigs where the top hit BAC clones are and searched for putative microsatellites within the BAC-end sequences of these contigs. Unfortunately, no available putative microsatellite for *fabp11aI* was found in any of the four methods that are mentioned above as of November 2011; hence, its genetic and physical location in the genome is still unknown. Figure 3.9 illustrates an image of the polyacrylamide gel for the genotyping of Ssa10082BSFU in the Br6 Atlantic salmon mapping family. This variable microsatellite was identified from the consensus sequence of the assembled BAC S0938E16 shotgun library sequencing reads (see Table 3.3). Therefore, it is linked to *fabp2b*. The sire and the dam have genotypes 1-2 and 2-3, respectively. As a result, the possible offspring genotypes, 1-2, 1-3, 2-2 and 2-3, appeared in the 46 progenies. This set of microsatellite genotype data was compared with other previously characterized microsatellite data of this mapping family using LINKFMFEX (Danzmann) and its genetic location was identified to be on linkage group 24, which corresponds to

chromosome 7 based on the integration of Atlantic salmon genetic and physical maps (Phillips et al., 2009).

In the case of Atlantic salmon *fabp1s* and *fabp7bs*, the representative BACs of these genes belong to DNA fingerprinting contigs that already had variable microsatellites previously mapped on to the Atlantic salmon genetic map. Therefore, I generated minimum tiling paths (MTPs) to link the BAC clones with these variable microsatellites to the corresponding *fabp* representative BACs. Figure 3.10 is the agarose gel image of PCR amplifications that were used to generate the MTP for the DNA fingerprinting contig 798, and Figure 3.11 is a screenshot of this MTP for this contig on ASalBase.

Figure 3.8. An image of the polyacrylamide gel illustrating the genotyping with the *fabp2b* linked variable marker, Ssa10082BSFU, on the Atlantic salmon Br6 mapping family.

This family has 48 individuals. The sire and the dam are in lanes 47 and 48, respectively, and the remainder are the progeny.

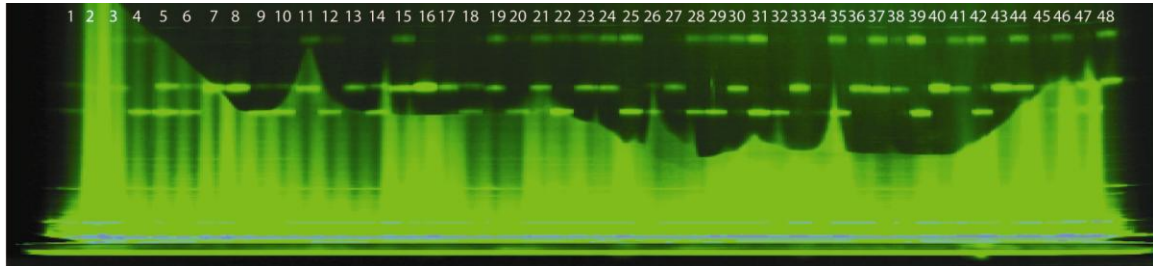


Figure 3.9. Agarose gel image illustrates the PCR result that was used for generating a MTP of DNA fingerprinting contig 798 to link the variable microsatellite marker to the *fabp7b1* representative BAC clone, S0232I03.

Microsatellite Ssa0907BSFU was found in the Sp6 BAC end sequence of BAC S0152B01. Different sets of PCR reactions used different primers for amplification. PCR1 to PCR23 used primers designed from S0152B01 Sp6, S0152B01 T7, S0173P12 T7, S0173P12 Sp6, S0269H08 T7, S0269H08 Sp6, S0238M22 T7, S0238M22 Sp6, S0037L07 T7, S0037L07 Sp6, S0235O17 T7, S0272E19 T7, S0272E19 Sp6, S0011I12 T7, S0011I12 Sp6, S0052J03 T7, S0052J03 Sp6, S0201P23 T7, S0201P23 Sp6, S0172N08 T7, S0172N08 Sp6, S0232I02 T7 and S0232I02 Sp6 BAC end sequences, respectively.

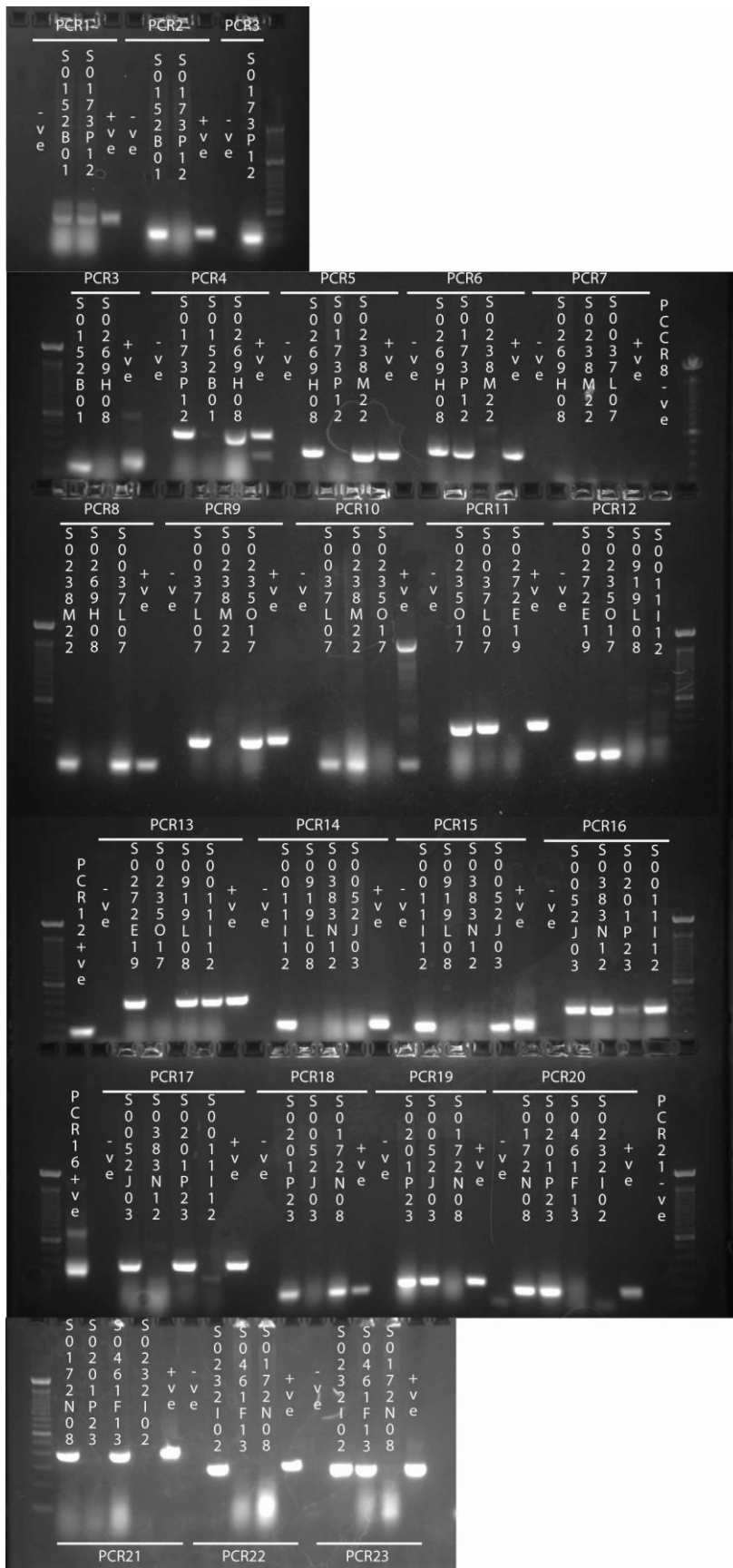
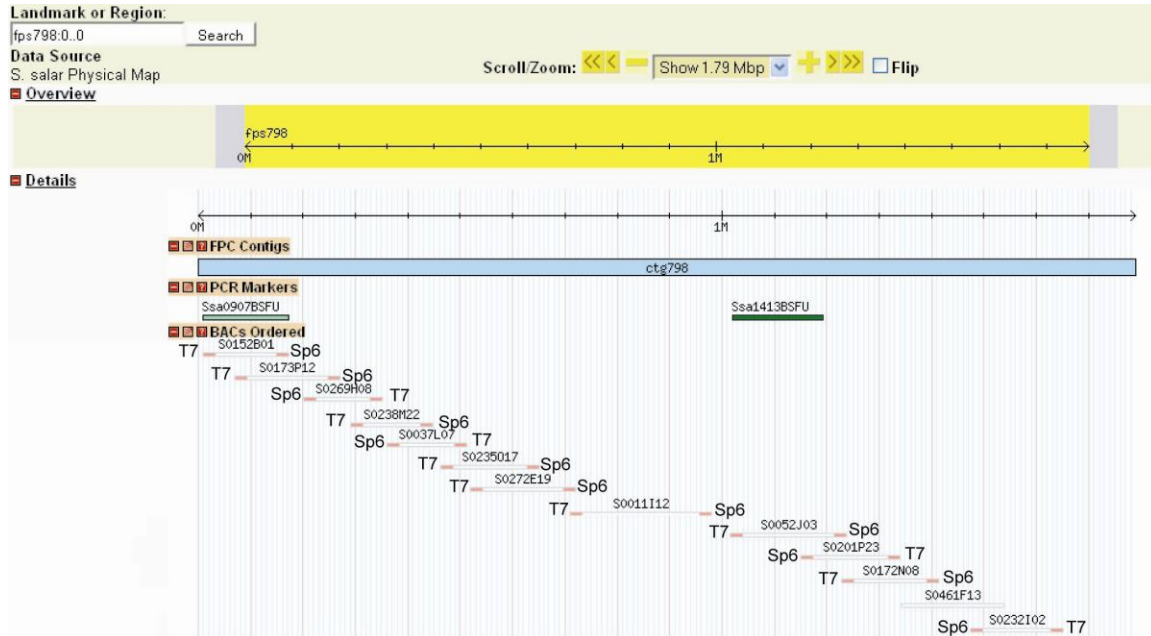


Figure 3.10. A screenshot from ASalBase on DNA fingerprinting contig 798, where the representative BAC clone for *fabp7bI* is S0232I02.

An informative microsatellite marker Ssa0907BSFU was found in the Sp6 BAC end sequence of S0152B01. The red ends of each BAC clone indicate that there are available BAC sequence information for these clones. I masked the repetitive regions of these sequences and designed primers from each BAC end to generate a MTP that covers the entire region of DNA fingerprinting contig 798. I confirmed that Ssa0907BSFU is linked to *fabp7bI*.



4: Evolution and expression profiling of the Atlantic salmon *fabp2* sub-family

4.1 Introduction

FABP2 is the intestinal-type fatty acid-binding protein. It is found in the epithelium of the small intestine and is most abundant at the distal segment of the small intestine, but FABP1 and FABP6 are also found in the proximal and distal regions of small intestine in mice, respectively (Hauerland and Spener, 2004). *Fabp2* knock-out mice were viable, fertile, had normal fat absorption and serum glucose levels, but had a higher plasma insulin level (Furuhashi and Hotamisligil, 2008; Vassileva et al., 2000). Hence, it has been suggested that FABP1 and FABP6 also participate in the fatty acid uptake in the intestine without FABP2 and without increasing the total amount of FABP in the intestine (Hauerland and Spener, 2004). Nevertheless, male FABP2 knock-out mice had a larger liver and higher triglyceride levels and gained more weight regardless of the fat content in the diet (Furuhashi and Hotamisligil, 2008), whereas loss of weight, no difference in plasma triglyceride levels, and smaller liver on a high fat diet were observed in female FABP2 knock-out mice (Furuhashi and Hotamisligil, 2008). Hence, there are differences in the effect of FABP2 knock-out in both genders. Moreover, the human FABP2 polymorphism, A54 and T54, is suggested to be associated with insulin sensitivity and the rate of lipid metabolism (Baier et al., 1995), but this suggestion is still controversial (Chmurzynska, 2006; Hauerland and Spener, 2004).

4.2 Identification of salmonids, northern pike and rainbow smelt *fabp2* transcript sequences

Using the zebrafish *fabp2* nucleotide coding sequence as the query, I searched and found three and two unique *fabp2* full-length EST sequences in rainbow trout and chinook salmon, respectively (Table 4.1). As mentioned in Chapter 3.1, the absence of *fabp2* transcripts in fish species databases other than Atlantic salmon, rainbow trout and chinook salmon might be due to the number of available EST sequences in each fish species database and the type of tissues used to make in the EST libraries. Because the low number of chinook salmon EST sequences might not provide the completed assessment of the number of *fabp2* transcribed genes in this species, I did not use the chinook salmon *fabp2* sequences for further analysis. Nevertheless, the identity percentage of the two chinook unique full-length *fabp2* coding sequences to one another is 72%. This is less than the 96% identity of rainbow trout *fabp2aI* and *fabp2aII*, which can be attributed to a recent *fabp2a* gene duplication or the 4R WGD. Hence, it suggests that the two chinook salmon *fabp2* transcripts are related by a more ancient *fabp2* gene duplication or the 3R WGD.

Table 4.1. Rainbow trout and chinook salmon full-length *fabp* EST sequences and their GenBank accession numbers as of February 20th, 2012.

The gene name of rainbow trout *fabp2s* are based on their positions in the phylogenetics tree (see Figure 4.1).

	Rainbow trout (Omy)	Chinook salmon (Ots)
<i>fabp2aI</i>	BX882498.2	
<i>fabp2aII</i>	CT572560.1	
<i>fabp2b</i>	CT572560.1	EL555018.1
<i>fabp2a</i>		EL553586.1

4.3 Evolution of Atlantic salmon *fabp2* sub-family

A maximum parsimony phylogenetic tree (Figure 4.1) was built from the coding sequences of these vertebrate proteins, and it reveals that there are two *fabp2* clades in teleosts, which I named *fabp2a* and *fabp2b*. Assuming that the 3R genome duplication event occurred in the common ancestor of teleosts, I predicted that all extant teleosts would have two copies of *fabp2*, *fabp2a* and *fabp2b*. The phylogenetic tree is consistent with a whole genome duplication event (3R) occurring in the common ancestor of teleosts (Christoffels et al., 2004; Jaillon et al., 2004; Meyer and Van de Peer, 2005; Vandepoele et al., 2004). Stickleback, medaka and fugu possess two *fabp2*s with a gene belongs to each *fabp2* clade, whereas zebrafish and tetraodon only contain a single *fabp2* in their genomes and they belong to the *fabp2a* clade (Sharma et al., 2004). Hence, it suggests that there were independent *fabp2b* gene losses in the zebrafish and tetraodon lineages. In addition, both Atlantic salmon and rainbow trout contain three *fabp2*s in their genomes. According to the phylogenetic tree, two of the three *fabp2*s in each salmonid species are part of the *fabp2a* clade, where as the remaining *fabp2* belongs to the *fabp2b* group. Although it would be more parsimonious to postulate that a single gene duplication in the ancestor of Atlantic salmon and rainbow trout gave rise to *fabp2aI* and *fabp2aII* genes, I suggest that the 4R WGD in the common ancestor of salmonids resulted in *fabp2aI* and *fabp2aII* genes, and that a subsequent loss of one of the *fabp2b* genes occurred in the common ancestor of Atlantic salmon and rainbow trout after the salmonid-specific WGD event (Allendorf and Thorgaard, 1984; Danzmann et al., 2008).

The inferred protein sequences of the five sequenced fish species (zebrafish, stickleback, medaka, fugu and tetraodon) and mouse were obtained by using ExPASy

translate tool to translate the *fabp2* coding sequences that were retrieved from Ensembl release version 52. I compared these protein sequences with those from FABP2aI, FABP2aII and FABP2b from Atlantic salmon and rainbow trout. The clustalW alignment shows that the protein size is well conserved (Figure 4.2). There were 56 of the 132 amino acid residues invariant in these vertebrate proteins.

Figure 4.1. Maximum parsimony phylogenetic tree of FABP2 built from the amino acid sequence alignment shown in Figure 4.2.

The white diamond indicates the 3R duplication event, which gave rise to the *fabp2a* and *fabp2b* genes in teleosts. The tree did not yield the expected configuration for the salmonid FABP2A sequences, which is shown in Figure 4.5. As indicated in Figure 4.5, this is due to an increase in the number of amino acid substitution in the Omy FABP2AII.

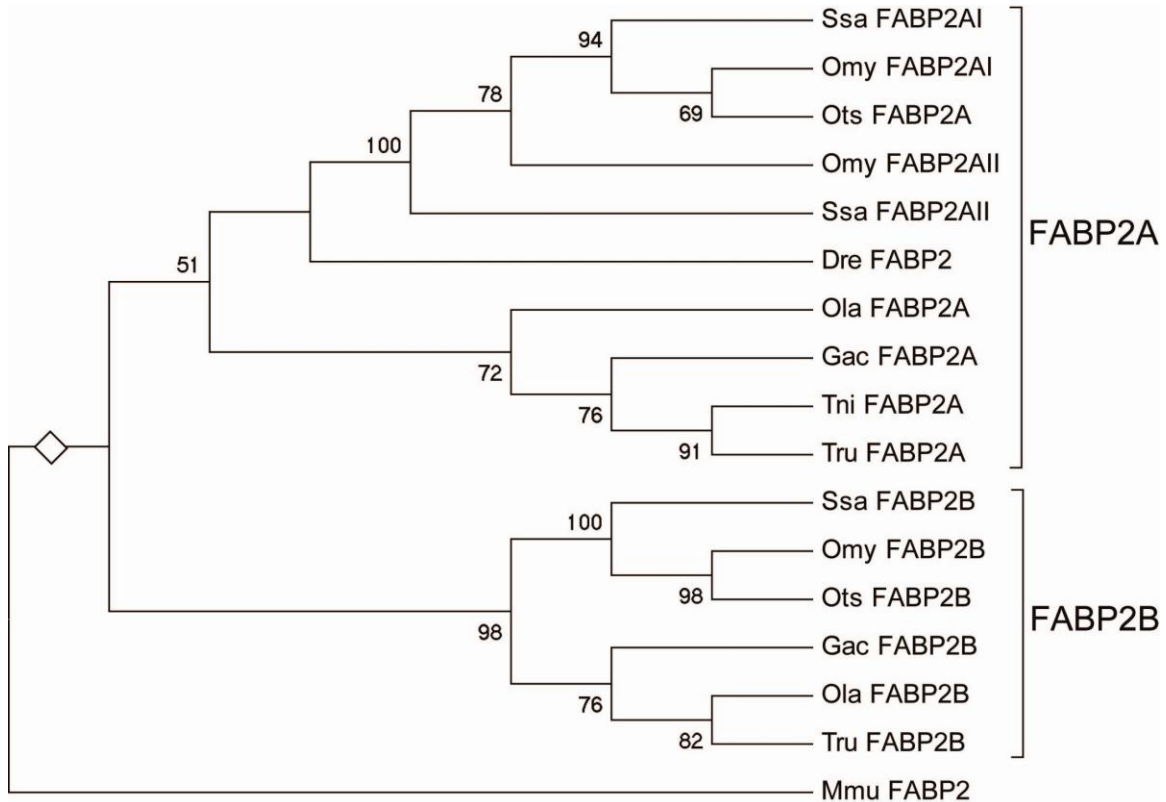


Figure 4.2. Alignment of the FABP2 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.

The species abbreviations are: *Salmo salar* (Ssa), *Oncorhynchus mykiss* (Omy), *Oncorhynchus tshawytscha* (Ots), *Danio rerio* (Dre), *Tetraodon nigroviridis* (Tni), *Takifugu rubripes* (Tru), *Gasterosteus aculeatus* (Gac), *Oryzias latipes* (Ola) and *Mus musculus* (Mmu). Protein size and great number of amino acid residues are well conserved in the FABP2s from different vertebrate species.

Ssa_FABP2aI MTYNGTWKVDRESENIEKFMQMGVNMVKKRLAAHDNLKITLEQTDGDKFVVKEASSFRITLD 60
 Omy_FABP2aI MTYNGTWKVDRESENIEKFMQMGVNMVKKRLAAHDNLKITLEQTDGDKFVVKEASSFRITLD 60
 Ots_FABP2a MTYNGTWKVDRESENIEKFMQMGVNMVKKRLAAHDNLKITLEQTDGDKFVVKEASSFRITLD 60
 Ssa_FABP2aII MTYNGTWKVDHRESENIEKFMQMGVNMVKKRLATHDNLKITLEQTDGDKFVVKEASTFRITLD 60
 Omy_FABP2aII MTYNGTWNVDRESENIEKFMQMGVNMVKKRLAAHDNLKIIIEQTDGDKFVVKEASTFRITLD 60
 Dre_FABP2 MTFNGTWKVDRENENIEKFMQMGVNMVKKRLAAHDNLKITLEQTDGDKFVNVKEVSTFRITLE 60
 Tni_FABP2 MTFNGTWKVDRENENIEKFMQMGVNMVKKRLAAHDNLKIIIEQDGNFTFHIEKSSNFRNLE 60
 Tru_FABP2a MTFNGTWKVERNDNYDKFMQMGVNMVKKRLAAHDNLKMTIEQSGDTFFHVKEHSNFRNLD 60
 Gac_FABP2a MTFDGTWKIDRNDNYDKFMQMGVNMVKKRLAAHDNLKIIIEQNGNKFVAVKESNFRNLE 60
 Ola_FABP2a MTFNGTWKVDRESENIEKFMQMGVNMVKKRLAAHDNLKIIIEQTDGDKFHVKESSNFRNIE 60
 Ssa_FABP2b MAFNGTWKVERNENYDKFMQMGVNMVKKRLAEHDNLKVIIEQTDGDKFHIKESSTFRNKD 60
 Omy_FABP2b MAFNGTWKVEHNENYDKFMQMGVNMVKKRLAEHDNLKVIIEQTDGDKFHIKESSTFRNKD 60
 Ots_FABP2b MAFNGTWKVEHNENYDKFMQMGVNMVKKRLAEHDNLKVIIEQTDGDKFHIKESSTFRNKD 60
 Tru_FABP2b MAFNGTWKAERNENYDKFMQMGVNMVKKRLAEHDNLKVIIEQTDGDKFHIKESSTFRNKD 60
 Gac_FABP2b MAFNGTWKVDRENENYDKFMQMGVNMVKKRLAEHDNLKVIIEQTDGDFQIKESSTFRNKD 60
 Ola_FABP2b MAFNGTWKADRSENENYDKFLERIGVNMVKKRLAEHDNLKIIIEQTDGNFHVKESSSTFRITKE 60
 Mmu_FABP2 MAFDGTWKVDRENENIEKFMQMGVNMVKKRLGAHDNLKLTITQDGNKFTVKESSNFRNID 60
 * :****: :.:****:****:****: :****: . *****: : * * : * :** *,** . :

Ssa_FABP2aI MEFTLGVTFEYALADGTMLSGSWGMEGDMMKGTFTFRKDNKGVLKTTTRAIVGEELVQSYSY 120
 Omy_FABP2aI MEFTLGVTFEYALADGTMLSGTWGMEGDMMKGTFTFRKDNKGVLTTTTRAIVGEELVQSYSY 120
 Ots_FABP2a LEFTLGVTFEYALADGTMLSGSWGMEGDMMKGTFTFRKDNGRVLTTRAIVGEELVQSYSY 120
 Ssa_FABP2aII MEFTLGVTFEYALADGTMLSGSWGMEGDMMKGTFTFRKDNKGVLTTTTRAIVGDDELVQSYSY 120
 Omy_FABP2aII LEFTLGVTFEYALADGTMLSGSWGMEGDMMKGTFTFRKDNKGVLTTTTRAIVGEELVQSYSY 120
 Dre_FABP2 INFTLGVTFDYSLADGTELTGSWVIEGDTLKGFTFRKDNKGVLTTVRTIVNGELVQSYSY 120
 Tni_FABP2 MDFTLGVTFDYSLADGTELTGSWSMEGESMKGVFVRKDNKGQLTTIRTLEGGELVQSYNY 120
 Tru_FABP2a IDFTLGVTFEYSLADGTDVSGSWTIEGDMMKGVFVRKDNKGQLTTTRVQGDDELIVQSYSY 120
 Gac_FABP2a INFTLGDTFEYSLADGTELTGSWNEGDMMLKGFNFRKDNKGQLTTTRIVQRDELIVQSYNY 120
 Ola_FABP2a IDFTLGVTFEYSLADGTELTSGAWALEGDVVLKGTFTFRKDNKGALTTTRVIVQGDDELIVQSYNY 120
 Ssa_FABP2b IDFTLGVTQFDYALADGTEVSGAWEMDGDVVLKGFTRKDNKSKVLTTRALVGGELVQSYNY 120
 Omy_FABP2b IDFTLGVTQFDYALADGTEVSGTWEMDGDVVLKGFTRKDNKSKVLTTRAVVGGELVQSYNY 120
 Ots_FABP2b IDFTLGVTQFDYALADGTEVSGTWEMDGDVVLKGFTRKDNKSKVLTTRAVVGGELVQSYNY 120
 Tru_FABP2b IDFTLGVTQFDYALADGTEVSGTWEMDGDVVLKGFTRKDNKSKVLTTRSLVDGELVQSYNY 120
 Gac_FABP2b IDFTLGVTQFDYALADGTEVSGTWEMDGDVVLKGFTRKDNKSKVLTTRALVGGELVQSYNY 120
 Ola_FABP2b FDFTLGVTQFDYALADGTEVSGTWEMDGDVVLKGFTRKDNKSKVLTTRSLVDGELVQSYNY 120
 Mmu_FABP2 VVFEELGVNFPYSLADGTELTGAWTLEGNKLGKFTFRVDNKGKELIIVREVSGNELIQTYTY 120
 . * * * * :***** :****:****: : * * * **.: * : * : * :****:

Ssa_FABP2aI DGVEAKRIFKRG 132
 Omy_FABP2aI DGVEAKRIFKRG 132
 Ots_FABP2a DGVEAKRIFKRG 132
 Ssa_FABP2aII DGVEAKRIFKRG 132
 Omy_FABP2aII DGVEAQRFFKRG 132
 Dre_FABP2 DGVEAKRIFKRA 132
 Tni_FABP2 EGVDAKRIFKRS 132
 Tru_FABP2a DGVDKRIFKRS 132
 Gac_FABP2a EGVDAKRIFKRC 132
 Ola_FABP2a DGVDKRIFKRG 132
 Ssa_FABP2b DGVDKRIFKKA 132
 Omy_FABP2b DGVDKRIFKKA 132
 Ots_FABP2b DGVDKRIFKKA 132
 Tru_FABP2b EGVDAKRIFKKQ 132
 Gac_FABP2b EGVDAKRIFKKQ 132
 Ola_FABP2b EGVDAKRIFKKH 132
 Mmu_FABP2 EGVDAKRFFKKE 132
 :****:****:

4.4 Conserved synteny of *fabp2* loci in teleosts

I first compared the flanking regions of the Atlantic salmon *fabp2aI* and *fabp2aII* genes since this is where I predicted to see the most conservation of synteny because the phylogenetic analysis (see Section 4.2) suggested that *fabp2aI* and *fabp2aII* genes arose as a result of the more recent salmonid-specific WGD 25-120 MYA (Allendorf and Thorgaard, 1984). Three genes, ubiquitin specific peptidase53 (*usp53*), myozenin-2 (*myoz-2*) and similar to myopodin (*synpo2*), are conserved with respect to the 3' side of *fabp2aI* and *fabp2aII* (Figure 4.3) and orientation with respect to *fabp2as*. Guanylate cyclase beta subunit (*gucy-β*) is present on the 5' side of *fabp2aI* and *fabp2aII*; however, the transcriptional orientation is reversed in the two loci. In addition, SET domain bifurcated 2 protein (*setdb2*) is present in one of the flanking genes of *fabp2aI*, but it appears to be absent in the flanking region of *fabp2aII*. None of the flanking genes of *fabp2aI* and *fabp2aII* were observed in the vicinity of *fabp2b*.

When I compared the *fabp2a* genomic regions in Atlantic salmon, zebrafish, medaka, stickleback, tetraodon and fugu (Figure 4.3), I found that five genes are conserved in terms of position and orientation relative to the *fabp2a* gene in most lineages. The five genes are: guanylate cyclase alpha and beta subunits (*gucy-α* and *gucy-β*), *usp53* (absent from the medaka *fabp2a* locus), *myoz-2* and *synpo2*. The presence of these genes near the *fabp2a* loci in different lineages suggests that these genes were present in the common teleost ancestor after the 3R duplication and all extant teleosts would have this cluster of genes in their genomes. Therefore, I predicted to find evidence for a *usp53* pseudogene or unannotated gene somewhere between *fabp2a* and *myoz-2* in the medaka genome. Indeed, using NCBI bl2seq megablast, four portions of this medaka

genomic region have somewhat similar matches (76-84% identity) with two of the putative *usp53* exons near Atlantic salmon *fabp2aII*. This suggests that either there is an as yet unannotated functional *usp53* gene on medaka scaffold 461 between *fabp2a* and *myoz-2* or else it is present in a remnant form as a pseudogene. Since the phylogenetic tree suggests that the zebrafish *fabp2* is most closely-related to the two Atlantic salmon *fabp2as*, I predicted to find the greatest conservation of synteny in these two species. However, only *gucy- α* and *gucy- β* are conserved in the *fabp2a* flanking regions in Atlantic salmon and zebrafish, reflecting many genomic rearrangements that have occurred in both lineages since they diverged. The observation that tetraodon and fugu pufferfish have the most conservation of synteny at the *fabp2a* locus probably reflects the relatively recent separation of these species. The presence of *setdb2* at the *fabp2aI* locus and its absence in the *fabp2aII* flanking region reflects independent genomic rearrangements that occurred in the duplicated Atlantic salmon *fabp2a* loci, *fabp2aI* and *fabp2aII*, since the 4R duplication event.

Similar comparative genomic analyses were carried out at the *fabp2b* loci in Atlantic salmon and medaka, as well as stickleback and fugu in whose genomes it was not annotated in the Ensembl release version 57. I found that two genes, *dctn6* and *rbpms*, are conserved in terms of position and orientation relative to *fabp2b* in Atlantic salmon, medaka, fugu and stickleback (Figure 4.4). In addition, the doublesex-mab3 related transcription factor (*dmrt*) and zeta-sarcoglycan (*sgcz*) are conserved in the *fabp2b* flanking regions of medaka, fugu and stickleback. Again, all extant teleosts would obtain this cluster of genes, *sgcz-dctn6-rbpms-fabp2b-dmrt*, in their genomes. I looked for evidence for the presence of *fabp2b* pseudogene or an unannotated gene in zebrafish by

identifying the location of any *fabp2b* neighboring genes in its genome. Only *sgcz* is found on chromosome 1 and the other *fabp2b* flanking genes are either absent or have not been annotated yet in the zebrafish genome. By using NCBI bl2seq megablast and Atlantic salmon *fabp2b* coding sequence as reference, I identified a sequence similar to a portion of *fabp2b* exon 4 located approximately 1.15 Mb on the 5' side of *sgcz*. This observation is indicative of an as yet unannotated *fabp2b* functional gene or a pseudogene in the zebrafish genome. A similar search for a possible *fabp2b* pseudogene or unannotated gene was also carried out in the tetraodon genome. Both *dctn6* and *dmrt* could be found in the tetraodon genome on scaffold 7577, but the genomic sequence that is flanked by these two genes did not align with any region of Atlantic salmon *fabp2b* coding sequence. This suggests that the syntenic block of *rbpms* and *fabp2b* was translocated elsewhere in the tetraodon genome, was deleted or has decayed to the point that these sequences are no longer recognizable even as pseudogenes.

The lack of shared synteny and the striking difference in genomic organization between the *fabp2a* and the *fabp2b* loci in teleosts reflects the genomic rearrangements that have occurred in the *fabp6a* and *fabp6b* gene lineages since they were produced by the 3R WGD at the base of the teleost radiation approximately 320-370 MYA (Christoffels et al., 2004; Jaillon et al., 2004; Meyer and Van de Peer, 2005; Vandepoele et al., 2004).

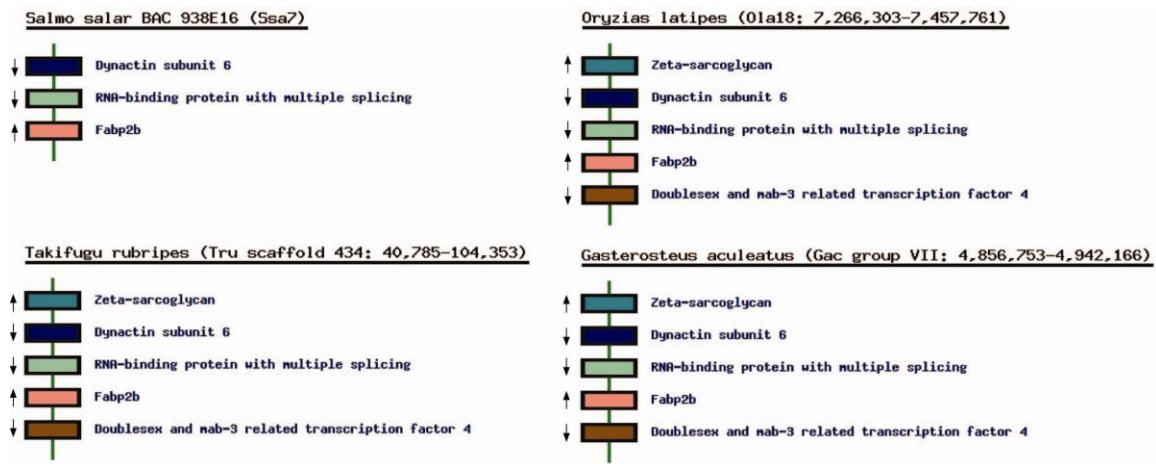
Figure 4.3. Comparative genomic analysis of *fabp2a* loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.

The arrows indicate where the gene orientations are.



Figure 4.4. Comparative genomic analysis of *fabp2b* loci in Atlantic salmon, stickleback, medaka and fugu.

The arrows indicate where the gene orientations are.



4.5 Expression profiling of Atlantic salmon *fabp2* sub-family

Using RT-PCR with *fabp2* gene-specific primers, I examined the expression profiles of Atlantic salmon *fabp2aI*, *fabp2aII* and *fabp2b* in twelve smolt tissues. Among the three genes, *fabp2b* is the most broadly expressed gene (Table 4.2). Its transcript was detected in all the tissues examined except for stomach. This broad tissue expression profile is similar to what has been observed for zebrafish *fabp2* (Sharma et al., 2004). Both Atlantic salmon *fabp2aI* and *fabp2aII* are expressed in intestine and pyloric caecum, but the *fabp2aII* transcript was also detected in eye, gonad, muscle, skin and thyroid. Although *fabp2aI* and *fabp2aII* exhibit reduced tissue expression compared to *fabp2b* and differ in their expression profiles relative to one another, these expression patterns overlap rather than complement one another. This could reflect some loss of regulatory elements in the *fabp2aI* and *fabp2aII* genes which may be leadin to sub-functionalization or even pseudogenization.

Table 4.2. Summary of RT-PCR analysis of the expression profiles of Atlantic salmon *fabp2aI*, *fabp2aII*, *fabp2b* and zebrafish *fabp2* (*taken from Sharma et al. (Sharma et al., 2004)).

√ indicates that an RT-PCR product was observed, - indicates no products was seen and N/T indicates that the particular tissue was not tested.

Tissue	<i>fabp2aI</i>	<i>fabp2aII</i>	<i>fabp2b</i>	*Zebrafish <i>fabp2</i>
Brain	-	-	√	√
Eye	-	√	√	N/T
Gonad	-	√	√	√
Heart	-	-	√	√
Intestine	√	√	√	√
Liver	-	-	√	√
Muscle	-	√	√	√
Pituitary Gland	-	-	√	N/T
Pyloric caecum	√	√	√	N/T
Skin	-	√	√	√
Stomach	-	-	-	N/T
Thyroid	-	√	√	N/T

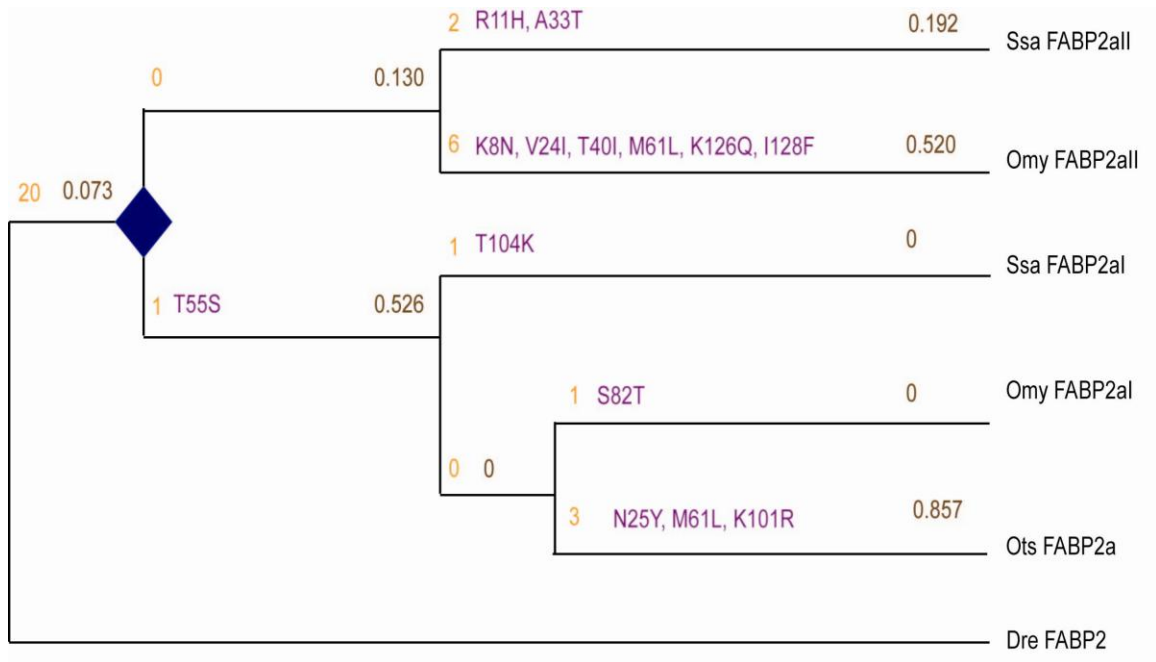
4.6 Evidence of neo-functionalization

Although the expression pattern of *fabp2* genes suggested Atlantic salmon *fabp2a* genes had experienced sub-functionalization or even pseudogenization, I wondered if there was any evidence of neo-functionalization in these gene products. Using the coding and protein sequences of zebrafish FABP2 as an outgroup, I examined the evolutionary history of the Atlantic salmon and rainbow trout FABP2aI and FABP2aII and Chinook salmon FABP2a (Figure 4.5). The non-synonymous substitution at non-synonymous site (dN) to synonymous substitution at synonymous site (dS) ratios indicating that these proteins have experienced negative or purifying selection. However, I was intrigued by the T55S substitution that occurred in the FABP2aI lineage after the duplication of the salmonid *fabp2a* genes, but before the speciation of Atlantic salmon, rainbow trout and chinook salmon. A study on human FABP2 revealed that two alleles, A55 and T55, vary in their binding affinities with respect to different lipids (Zhang et al., 2003). This implies that amino acid residue 55 plays a critical role in FABP2 function. The finding that a T55S substitution in FABP2aI occurred in the common ancestor of the Atlantic salmon and rainbow trout suggests that the salmonid FABP2aI and FABP2aII may have different lipid binding affinities, and this presumed neo-functionalization may have provided a selective pressure to maintain these duplicates. Moreover since the Atlantic salmon / rainbow trout speciation event occurred, there has been an increase in the rate of amino acid substitutions in FABP2aII in both lineages, particularly that of rainbow trout. I speculate that these changes are enabling FABP2aII to be more suitable for its protein function in different tissues in both species, such that Atlantic salmon and rainbow trout *fabp2aII* and

chinook salmon *fabp2a* can maintain their broad expression like it is in the zebrafish linkage (see Section 4.4).

Figure 4.5. Evolutionary history of Atlantic salmon (Ssa) and rainbow trout (Omy) FABP2AI and FABP2aII and chinook salmon FABP2a using the zebrafish (Dre) FABP2 as an outgroup (see also Figure 4.2).

The blue diamond indicates the 4R salmonid specific genome duplication event that gave rise to the FABP2aI and FABP2aII genes in Atlantic salmon and rainbow trout. The number of amino acid substitutions is given in orange (e.g., there are 20 amino acid sequence differences between the zebrafish FABP2 and the inferred sequence of the FABP2a in the common ancestor of the salmonids just prior to the salmonid genome duplication). The amino acid substitutions that occurred after the salmonid genome duplication in the common ancestor of Atlantic salmon and rainbow trout or in each of these species are shown on the branches of the tree. This is followed by the non-synonymous substitution at non-synonymous site (dN) to synonymous substitution at synonymous site (dS) ratios.



4.7 Conclusion

Three Atlantic salmon *fabp2* genes, *fabp2aI*, *fabp2aII* and *fabp2b*, were characterized and they are the product of two WGD events. The 3R genome duplication that occurred at the base of the teleost radiation gave rise to *fabp2a* and *fabp2b*; and this produced the two copies of *fabp2* genes seen in some but not all extant teleosts. The 4R genome duplication that occurred in the common ancestor of salmonids then gave rise to *fabp2aI*, *fabp2aII*, *fabp2bI* and *fabp2bII*. This was followed by a loss of *fabp2bI* or *fabp2bII*, which resulted in the three *fabp2* genes observed in Atlantic salmon and rainbow trout. I anticipated there are two *fabp2a* genes in chinook salmon. The fact that only a single *fabp2a* transcript found in this species might be due to the number of available ESTs in the Chinook salmon EST database and not a *fabp2a* gene loss in this lineage. Expression profiles provide evidence for sub-functionalization in Atlantic salmon *fabp2aI* and *fabp2aII* genes. In addition, a preliminary analysis of amino acid substitution patterns leads me to speculate that salmonid *fabp2a* genes may be subject to neo-functionalization because the T55S substitution in the lineage leading to FABP2AI in Atlantic salmon, rainbow trout and chinook salmon occurs at a residue known to be critically important for binding different fatty acids (Zhang et al., 2003).

5: Evolution and expression profiling of the Atlantic salmon *fabp7* sub-family

5.1 Introduction

FABP7 is the brain-type fatty acid-binding protein and it is highly selective for very long-chain n-3 polyunsaturated fatty acids, such as docosahexaenoic acid (Balendiran et al., 2000), which is an important component for the development of the brain (Guesnet and Alessandri, 2011). As a result, the *fabp7* transcript is highly abundant in radial glia cells of the developing brain, but is weakly expressed in mature glia of the white matter and not the grey matter of neurons (Furuhashi and Hotamisligil, 2008). FABP7 affects the correct migration of developing neurons into cortical layers (Feng, 1994) and is overexpressed in patients with Down's syndrome (Sanchez-Font et al., 2003) and schizophrenia (Watanabe et al., 2007). *Fabp7* is also expressed in the mammary gland, and a high level of this protein in the cell can inhibit tumor growth in mice with breast cancer (Hohoff and Spener, 1998; Shi et al., 1997), suggesting its role as a tumor suppressor gene. FABP7-deficient mice do not have any apparent macroscopic differences but they have altered emotional behavioural responses, attenuated neurogenesis and decreased prepulse inhibition, which is a typical behaviour in schizophrenia (Owada et al., 2006; Watanabe et al., 2007). Due to its abundance in the brain, *fabp7* has been proposed as a biomarker for the detection of brain injury (Pelsers et al., 2005).

5.2 Identification of salmonids, northern pike and rainbow smelt *fabp7* transcript sequences

Using the zebrafish *fabp7a* and *fabp7b* nucleotide coding sequences as the queries, I searched and found unique *fabp7* full-length EST sequences in rainbow trout, grayling, northern pike and rainbow smelt in the cGRASP EST database (Table 5.1). The sequence identity percentages between the two rainbow trout *fabp7* nucleotide coding sequences is 96%, which is the same as the sequence identity percentages between the two grayling *fabp7* sequences. However, the two rainbow smelt *fabp7* coding sequences are 77% identical. Based on the sequence identities, it suggests that the two rainbow smelt *fabp7* transcripts are related by a more ancient *fabp7* gene duplication or the 3R WGD, whereas the two *fabp7*s in grayling and rainbow trout are related by a recent gene duplication or the 4R WGD.

Table 5.1. Rainbow trout, grayling, northern pike and rainbow smelt full-length *fabp7* EST sequences and their GenBank accession numbers as of February 20th, 2012.

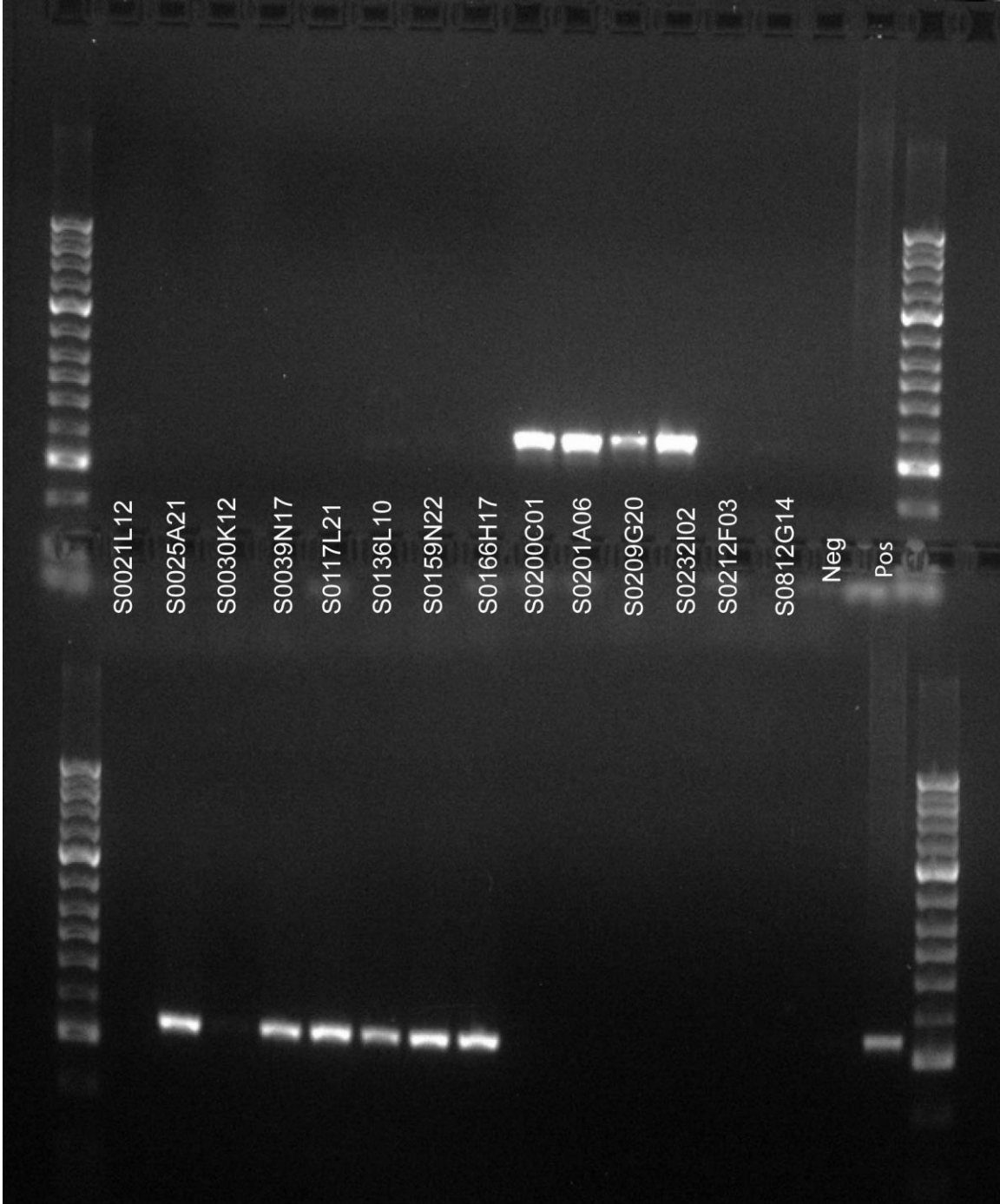
	Rainbow trout (Omy)	Grayling (Tth)	Northern pike (Elu)	Rainbow smelt (Omy)
<i>fabp7a</i>				EL525820.1
<i>fabp7b</i>			GH244102.1	EL527123.1
<i>fabp7bI</i>	BX082627.2	FF841289.1		
<i>fabp7bII</i>	CA369721.1	FF841199.1		

5.3 Identification of BACs with *fabp7* via *fabp3* BAC library screening

As mentioned in Section 3.3.3, the identification of BAC clones containing the *fabp7s* was different from other *fabp* sub-families; that is, these BAC clones were identified during the BAC library screening with the *fabp3* oligonucleotide probe when I searched for BAC clones containing the *fabp3* genes. Most of the *fabp3* hybridization-positive BAC clones belong to the DNA fingerprinting contigs 756 and 798; however, there was no PCR amplification product when these BAC clones were used as template with the two *fabp3* gene-specific primers. Given the fact that *fabp7* is the closest *fabp* sub-family to *fabp3* (Schaap et al., 2002), I hypothesized that DNA fingerprint contigs 756 and 798 may contain *fabp7s* instead. Indeed, there were PCR amplification products when the reactions contained the *fabp3* hybridization-positive BAC clones from these two contigs as template and the two *fabp7* gene-specific primers. In fact, BAC clones from contig 798 were *fabp7bI* PCR-positive and BACs from contig 756 were PCR positive with *fabp7bII* gene-specific primers (Figure 5.1). As the result, based on the criteria mentioned in Section 3.3.3, BAC clones S0232I02 and S0025A21 were chosen to be the representative BACs for *fabp7bI* and *fabp7bII* (Table 3.3).

Figure 5.1. A 2% agarose gel image of PCR testing on *fabp3* hybridization-positive BAC clones with *fabp7bI* and *fabp7bII* gene-specific primers.

The first and last lanes of the top and bottom gel were loaded with 2.5 μ L Bioline HyperLadderII. The top gel was loaded PCR reactions with *fabp7bI* gene-specific primers, whereas the bottom gel was loaded PCR reactions with *fabp7bII* gene-specific primers. The BAC clones that were subjected for this PCR test were some of the *fabp3* hybridization-positive BACs. According to ASalBase, BAC clones S0021L12, S0030K12 and S0209G20 belong to DNA fingerprinting contigs 3120, 1164 and 363, respectively. BAC clones S0025A21, S0039N17, S0117L21, S0159N22 and S0166H17 are from contig 756, whereas BAC S0200C01, S0209G20 and 232I02 belong to contig 798. BAC S0136L10, S0212F03 and S0812G14 are singletons. In fact, BACs S0212F03 and S0812G14 are the representative clones for *fabp3aI* and *fabp3aII*, respectively (Table 3.3). “Neg” and “Pos” are the abbreviations for negative and positive controls (Atlantic salmon genomic DNA), respectively.



5.4 Evolution of Atlantic salmon *fabp7* sub-family

Human, rat, mouse, chicken, zebrafish, stickleback, medaka, fugu and tetraodon *fabp7* transcripts and protein sequences were retrieved from Ensembl release version 64. I used these tetrapod and teleost FABP7 protein sequences and compared the inferred amino acid sequences with their orthologs in rainbow smelt, northern pike, grayling, Atlantic salmon and rainbow trout using a clustalW alignment (Figure 5.2). FABP7 sub-family is the most conserved in terms of numbers of conserved amino acid residues and protein size compared to any other FABP sub-families that I studied in this project. There are 65 invariant out of 132 amino acids in these FABP7 vertebrate proteins. A phylogenetic tree based on the coding sequences of these FABP7 vertebrate proteins reveals that there was a *fabp7* gene duplication that gave rise to two *fabp7* clades, which Liu et al. named *fabp7a* and *fabp7b* (Figure 5.3) (Liu et al., 2004). This supports the hypothesis that there was a *fabp7* gene duplication as a result of the 3R WGD in the common ancestor of teleosts. Two copies of *fabp7s* are found in the grayling, Atlantic salmon and rainbow trout EST databases, and all of these genes fall into the *fabp7b* clade. The fact that rainbow smelt possesses a *fabp7a* and a *fabp7b* and that northern pike only has a *fabp7b* in its genome suggests that there was a *fabp7a* gene loss in the common ancestor of northern pike, grayling and the other salmonids after rainbow smelt had diverged from it. It does not appear to be due to the number of ESTs in the northern pike EST database because the grayling EST database has a similar number of ESTs as the northern pike EST database (see Section 3.1). Without additional information, such as flanking sequences, it is not possible to determine if the salmonid *fabp7bI* and *fabp7bII* genes arose via a WGD or simply a gene duplication in their common ancestor.

Figure 5.2. Alignment of the FABP7 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.

The species abbreviations are the same as they are in Figure 4.1 with the addition of *Thymallus thymallus* (Tth), *Esox lucis* (Elu), *Osmerus mordax* (Omo), *Gallus gallus* (Gga), *Rattus norvegicus* (Rno) and *Homo sapiens* (Hsa). Protein size and many amino acid residues are conserved in the FABP7s from different vertebrate species.

```

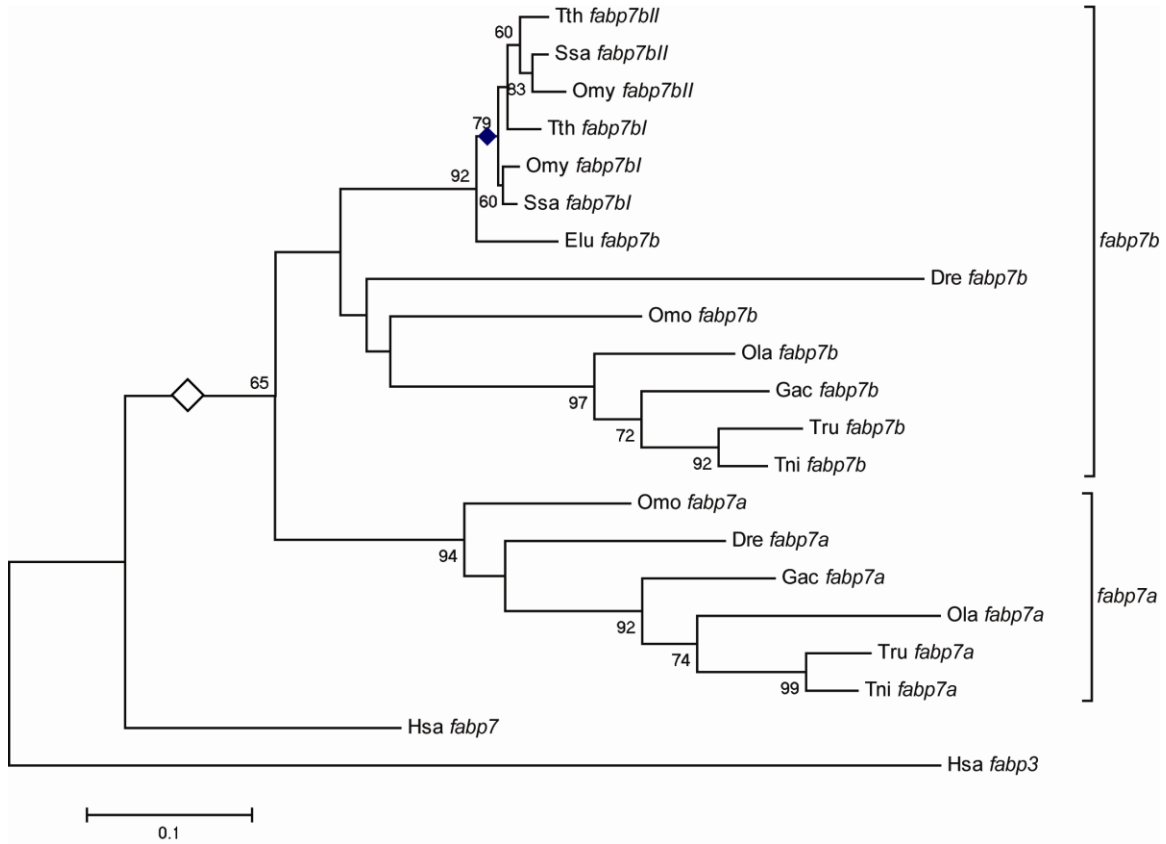
Sea_FABP7b I      MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Omy_FABP7b I      MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tch_FABP7b I      MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Sea_FABP7b II     MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Omy_FABP7b II     MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tth_FABP7b II     MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Elu_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Omo_FABP7b       MVDTFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREL 109
Dre_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Gac_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Ola_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tru_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tni_FABP7b       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Omo_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Dre_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Gac_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Ola_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tru_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Tni_FABP7a       MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Gga_FABP7        MVEAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Mmu_FABP7        MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Rno_FABP7        MVDAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
Hsa_FABP7        MVEAFCATWKLVDSDNFDEVMKALGVGFPATRQVGNVTKPTVIIAKEGDKVVKVQSTFKNTEISFKLGEFDEATADDRNCKSTVSLDGGKLAHVQRWDGKETK FVREI 109
*.:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*.:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Sea_FABP7b I      KDGLVMNTLTFEDIVAVRTYEKA 132
Omy_FABP7b I      KDGLVMNTLTFEDIVAVRTYEKA 132
Tch_FABP7b I      KEGKLVMTLTFEDIVAVRTYEKA 132
Sea_FABP7b II     KDGLVMNTLTFEDIVAVRTYEKA 132
Omy_FABP7b II     KDGLVMNTLTFEDIVAVRTYEKA 132
Tth_FABP7b II     KDGLVMNTLTFEDIVAVRTYEKA 132
Elu_FABP7b       KDGLVMNTLTFEDIVAVRTYEKA 132
Omo_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Dre_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Ola_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Gac_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Tru_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Tni_FABP7b       KDGLVMNMLTFEDIVAVRTYEKA 132
Omo_FABP7a       KDGLVMNTLTFEGVQAVRTYEKA 132
Dre_FABP7a       KDGLVMNTLTFEGVQAVRTYEKA 132
Gac_FABP7a       KDGLVATLTFEGVQAVRTYEKA 132
Ola_FABP7a       KDGLVMNMLTFEGVQAVRTYEKA 132
Tru_FABP7a       KDGLVMNMLTFEGVQAVRTYEKA 132
Tni_FABP7a       KDGLVMNMLTFEGVQAVRTYEKA 132
Gga_FABP7        KDGLVMNMLTFEGVQAVRTYEKA 132
Mmu_FABP7        KDGLVMNMLTFEGVQAVRTYEKA 132
Rno_FABP7        KDGLVMNMLTFEGVQAVRTYEKA 132
Hsa_FABP7        KDGLVMNMLTFEGVQAVRTYEKA 132
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```


Figure 5.3. Maximum likelihood phylogenetic tree of *fabp7* built from the nucleotide coding sequence alignment.

The white diamond indicates the 3R duplication event, which gave rise to the *fabp7a* and *fabp7b* genes in teleosts. Note that the grayling *fabp7bI* gene is not positioned where one might expect it to be, but the bootstrap value for this placement is low. I did not include bootstrap values less than 50 in the figure.



5.5 Conserved synteny of *fabp7a* loci in teleosts

When comparing the *fabp7a* loci in the five sequenced teleosts, I predicted a conserved ancient synteny block that might be present in the common ancestor of teleosts after the 3R genome duplication (Figure 5.4). Five genes are conserved in terms of position and orientation relative to *fabp7a*. They are: thromospondin-type laminin G domain and EAR repeats (*tspear*) (absent in the zebrafish *fabp7a* locus), ubiquitin carboxyl-terminal hydrolase 40 (*usp40*) (absent in the zebrafish *fabp7a* locus), TRAF3 interacting protein 2 (*trap3ip2*) (absent in the zebrafish, stickleback and tetraodon *fabp7a* loci), FYN oncogene related to SRC, FGR, YES (*fyn*) and myristoylated alanine-rich protein kinase C substrate (*marckc*). Given that *usp40* reside on the 5' end of *fabp7a* in four out of five teleosts that I examined, I predicted this relative positioning of these two genes were present in the common ancestor of teleosts. Interestingly, *usp40* locates to the 3' end of *fabp7a* and between *fabp7a* and *trap3ip2* in medaka lineage and this suggests that there was a translocation of *usp40* instead of an inversion because *usp40* and *fabp7a* orientations are not inverted compared to what is observed in the stickleback, fugu and tetraodon genomes. Because of the absence of *usp40* and *tspear* at the zebrafish *fabp7a* locus, I expect that these genes had translocated together elsewhere in the genome. Indeed, *usp40* (ENSDARG00000071197) is located on chromosome 6 along with *tspear* (ENSDARG00000077580). Therefore, it appears that there was a translocation event nearby the *fabp7a* locus in the zebrafish lineage. *Trap3ip2* might appear to be an inserted gene between *fabp7a* and *fyn* in both medaka and fugu lineages, but there is evidence suggesting that there are *trap3ip2* pseudogenes or possibly unannotated genes in the

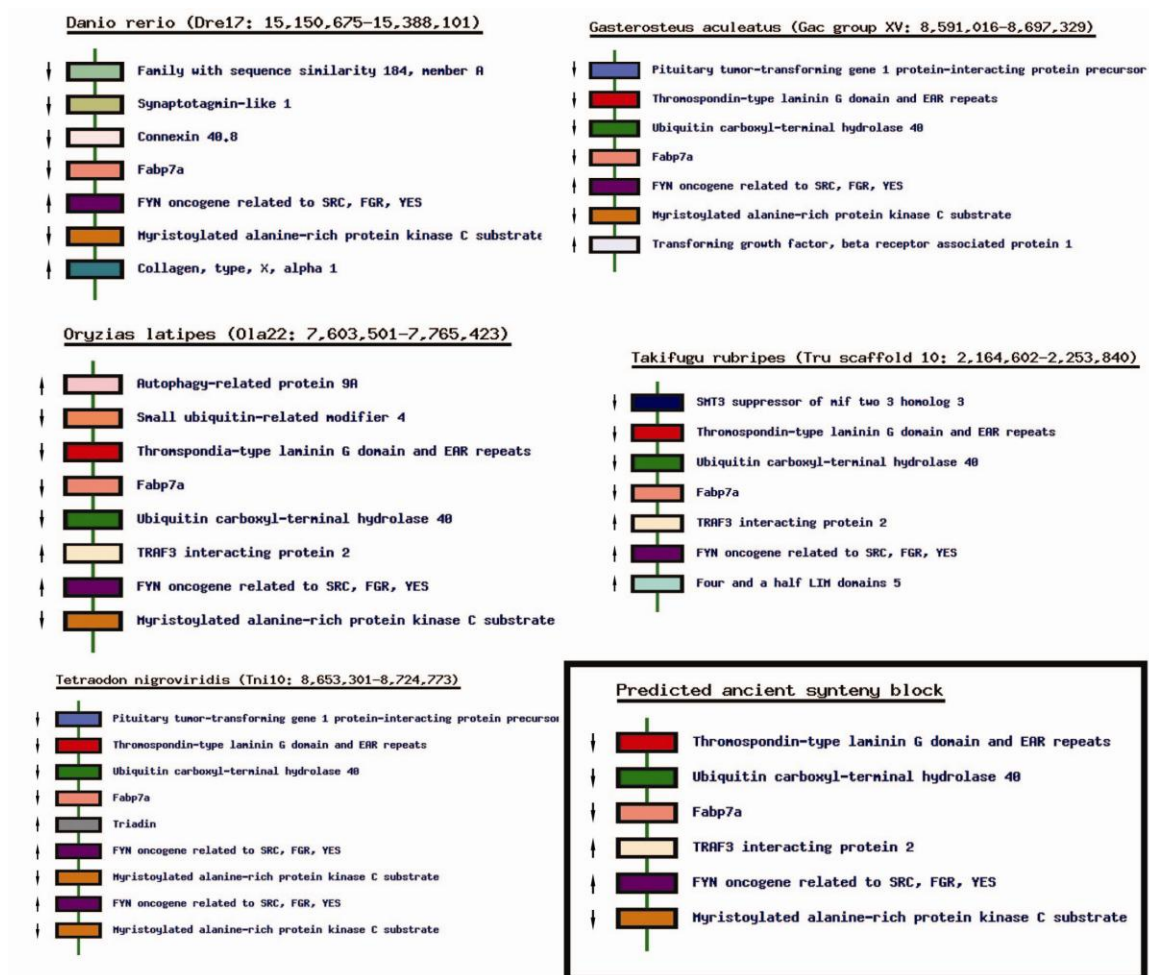
region between *fabp7a* and *fyn* near the *fabp7a* loci in zebrafish, stickleback and tetraodon genomes. Using medaka *trap3ip2* coding sequence (ENSGACG00000010180) as the query for NCBI bl2seq megablast search, there was a short portion of exon two sequence (13 nt in length) aligned with 100% identity to the region between *fabp7a* and *fyn* in zebrafish. In the stickleback genome, there are two portions of medaka *trap3ip2* exons two and six (66 nt and 140 nt in length) aligned with 79% identity to the region between *fabp7a* and *fyn*. Moreover, portions of medaka *trap3ip2* exons four, five and six (47-131 nt in length) aligned with 75-83% identity to the region between *fabp7a* and *fyn* (ENSTNIG00000017221) in the tetraodon genome. I also searched for the presence of a *marckc* pseudogene or unannotated gene in the region between *fyn* and four and a half LIM domains 5 (*FHL5*) in the fugu genome using the stickleback *marckc* coding sequence (ENSGACT00000013531) as query for the NCBI bl2seq megablast search. The full-length stickleback *marckc* coding sequence, which consists of only two exons, aligned with 77-80% identity with the genomic sequence between *fyn* and *FHL5* in the fugu genome, suggesting that there may be a *marckc* pseudogene or an unannotated *marckc* gene in this region of the genome. Overall, it suggests that there may be *trap3ip2* pseudogenes or unannotated genes in that region between *fabp7a* and *fyn* in the zebrafish, stickleback and tetraodon genomes.

Since there were no *fabp7a* transcripts found in the Atlantic salmon EST database, I attempted to find the *fabp7a* pseudogene location using the available three times coverage of Atlantic salmon genome by the CIGENE scaffolds. Given that *fyn* resides to the 3' end of *fabp7a* in the predicted ancient conserved *fabp7a* synteny block, I searched for *fyn* in the Atlantic salmon CIGENE scaffolds in ASalBase by using

zebrafish FYN oncogene related to SRC, FGR, YES a (*fyna*) coding sequence (ENSDART00000150232) from Ensembl release version 64 as the query sequence and annotated the highest hit scaffold. The highest hit scaffold is scaffold scf15100727 and the annotated genes are *fyna*, *trap3ip2*, and DNA polymerase zeta catalytic subunit (*rev3l*). Indeed, I found there are three portions of zebrafish *fabp7a* coding sequences (12-31 nt in length) between *trap3ip2* and *rev3l*, suggesting that a putative *fabp7a* pseudogene exists in this region of the Atlantic salmon genome. Furthermore, I used scaffold scf15100727 as the query sequence to search against the BAC end sequences available in ASalBase to identify the DNA fingerprinting contig corresponding to this region in the genome. The sequence of scaffold scf15100727 matches to the T7 and Sp6-end BAC end sequences of BAC clones S0236B14 and S0129K21, respectively. Both BAC clones belong to the DNA fingerprinting contig 563 and this contig had been mapped to linkage group 4, which corresponding to chromosome 6 according to the integration information of Atlantic salmon genetic and physical maps(Phillips et al., 2009). Hence, the location of a putative *fabp7a* pseudogene in the Atlantic salmon genome was identified.

Figure 5.4. Comparative genomic analysis of the *fabp7a* loci in zebrafish, stickleback, medaka, fugu and tetraodon.

The arrows indicate where the gene orientations are.

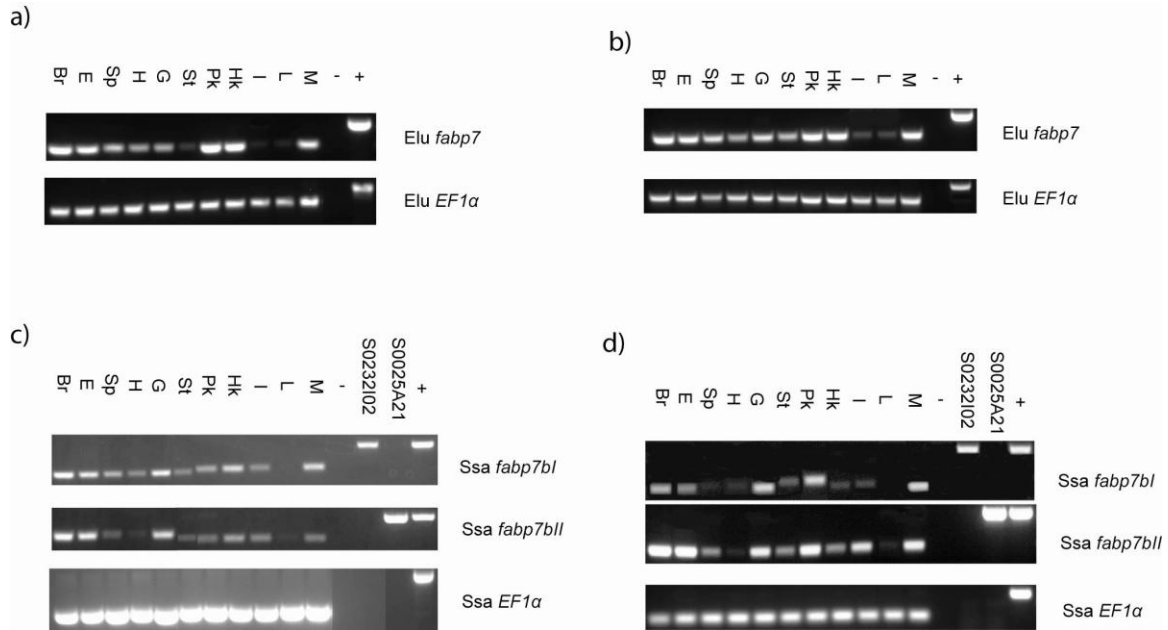


5.6 Expression profiling of Atlantic salmon *fabp7*

To test if sub-functionalization occurred in the *fabp7b* gene duplicates, the expression patterns of *fabp7bI* and *fabp7bII* in Atlantic salmon were examined because the prediction is that if sub-functionalization has occurred the expression patterns would differ from the expression pattern of *fabp7b* in northern pike because the phylogenetic tree shows that northern pike is the pre-WGD species that is most closely-related to salmonids (Figure 5.3) (Ramsden et al., 2003). As illustrated in Figure 5.5, this gene is expressed in a wide variety of northern pike tissues including brain, eye, spleen, heart, gill, stomach, posterior and head kidney, intestine, liver and muscle. This is a more diverse expression pattern compared to the *fabp7b* gene in zebrafish (Liu et al., 2004). Liu et al. showed that the zebrafish *fabp7b* gene is expressed in the liver, intestine, brain and testis but not in ovary, skin, heart, muscle and swim bladder (Liu et al., 2004). It is not possible to tell which of these expression patterns represents an ancestral state, but I speculate that it is the one seen in northern pike. Moreover, I suggest that the loss of the *fabp7a* gene in northern pike provided a selection pressure against loss of expression of the *fabp7b* gene, such as is observed for the *fabp7b* gene in zebrafish (e.g., heart). Taking the northern pike *fabp7b* expression pattern as representing the *fabp7b* gene expression in the common ancestor of the salmonids prior to the 4R WGD, I examined the expression patterns of the Atlantic salmon *fabp7bI* and *fabp7bII* genes for evidence of sub-functionalization (Force et al., 1999). As shown in Figure 5.5b, I see that *fabp7bI* and *fabp7bII* genes have broad tissue expression patterns that are similar to those of the corresponding gene in northern pike, but *fabp7bI* is not expressed in liver whereas *fabp7bII* is.

Figure 5.5. Expression patterns of *fabp7bI* and *fabp7bII* in Atlantic salmon smolt tissues and of *fabp7b* in the northern pike tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill, (G), stomach (St), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M), negative control (-) and northern pike genomic DNA (Figure 5.5a and 5.5b) and Atlantic salmon genomic DNA (Figure 5.5c and 5.5d) (+). Figure 5.5a and 5.5c are RT-PCR with 35 cycles whereas figure 5.5b and 5.5d are RT-PCR with 40 cycles. Differential expression patterns of Atlantic salmon *fabp7bI* and *fabp7bII* were seen in liver.



5.7 Putative 5' regulatory elements

The DDC model suggests that sub-functionalization is one mechanism for retaining duplicated genes by having complementary losses of regulatory elements in the gene duplicates, such that the paralogs exhibit differential expression profiles that together would reflect what is observed in representative of the ancestral gene (Force et al., 1999). I wondered if the differential expression profiles of *fabp7bs* are indeed due to complementary losses of regulatory elements. Hence, I compared the predicted 5' transcription factor binding sites of Atlantic salmon *fabp7bI* and *fabp7bII*, zebrafish, medaka, stickleback, fugu and tetraodon *fabp7bs*. I used TOMTOM (Gupta et al., 2007) and MEME (Bailey and Elkan, 1994) to identify the top ten DNA motifs (Figure 5.6) that are conserved in relative position and orientation in the proximal promoter regions (2000 bp upstream of the start codon) of zebrafish, stickleback, medaka, tetraodon *fabp7b* and Atlantic salmon *fabp7bI* and *fabp7bII* genes. The details of these top ten DNA motifs are shown in Figures 5.7. The only difference in these upstream regions of the Atlantic salmon *fabp7bI* and *fabp7bII* genes appears to be an indel between motif 3 and motif 6 approximately at -1050 to -1200 bp from the start codon. TFSEARCH (Heinemeyer et al., 1998) analysis of this indel sequence identified the presence of CdxA, IRF-1, Evi-1, TATA, Oct-1 and SRY putative transcription factor binding sites (Figure 5.6). Rath et al. suggested that the presence of SRY along with CdxA and GATA-2 may act to repress the expression of senescence marker protein-30 (*smp30*), which is preferentially expressed in hepatocytes (Rath et al., 2008). Therefore, the presence of CdxA and SRY binding motifs, which occur in the Atlantic salmon *fabp7bI* gene but not the *fabp7bII* gene, might explain the loss of *fabp7bI* expression in the liver. If this is the case, it should be noted

that this might be contrary to what is expected from the sub-functionalization model of Force et al., as it could appear that an insertion of DNA is the basis of the sub-functionalization, although the end result is the same (Force et al., 1999).

I also used rVISTA (Loots et al., 2002) to search for conserved putative transcription factor binding sites between Atlantic salmon *fabp7bI* and *fabp7bII*. It appears that many putative transcription factor binding sites that Liu et al. identified in the regulatory regions of the zebrafish *fabp7a* and *fabp7b* genes are conserved (e.g., AP1, BRN2, CAAT, CREB, GATA4, NF1, OCT1, PIT1, POU, PPAR, TATA and TST1), and this might explain the co-expression of both Atlantic salmon genes in a wide variety of tissues (Liu et al., 2003b; Liu et al., 2004). It has been suggested that POU transcription factors are essential for *fabp7b* gene expression in the brain of zebrafish (Liu et al., 2003b) and medaka (Maruyama et al., 2008). Moreover, Maruyama et al. demonstrated that deletion of the POU transcription factor binding sites in the region from -898 to -801 bp of the start codon of the medaka *fabp7b* gene significantly reduced the expression of this gene in the brain (Maruyama et al., 2008). I searched for putative POU transcription factor binding sites in the genomic sequences that are 2000 bp 5' upstream of the start codons of zebrafish, stickleback, medaka and tetraodon *fabp7b* genes and both Atlantic salmon *fabp7bI* and *fabp7bII* genes using TFSEARCH (Figure 5.8). Putative POU transcription factor binding sites were identified at locations equivalent to the key -898 to -801 bp region of the medaka *fabp7b* gene in the *fabp7b* gene of tetraodon and possibly that of zebrafish as well as the Atlantic salmon *fabp7bI* and *fabp7bII* genes. These results are consistent with Atlantic salmon *fabp7bI* and *fabp7bII* genes being expressed in brain (Figure 5.5).

Figure 5.6. Putative transcription factor binding motifs in the 2000 bp regions upstream of the start codons of teleost *fabp7b* genes.

The motif sequences identified using TOMTOM (Gupta et al., 2007) and MEME (Bailey and Elkan, 1994) and the putative transcription factor were identified by TFSEARCH (Heinemeyer et al., 1998).

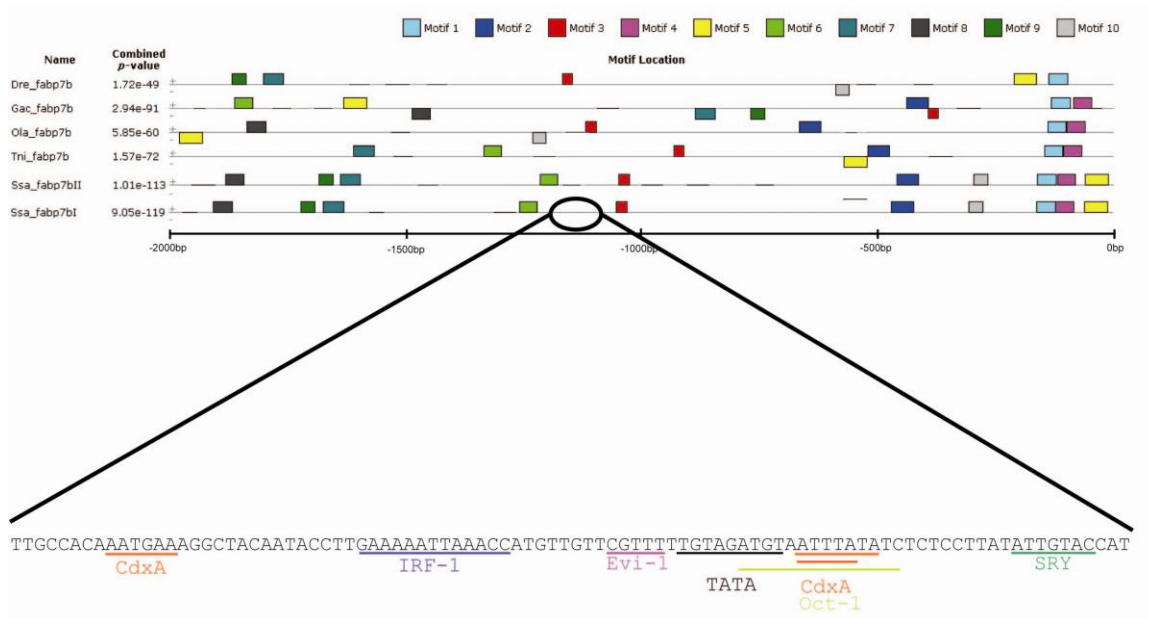


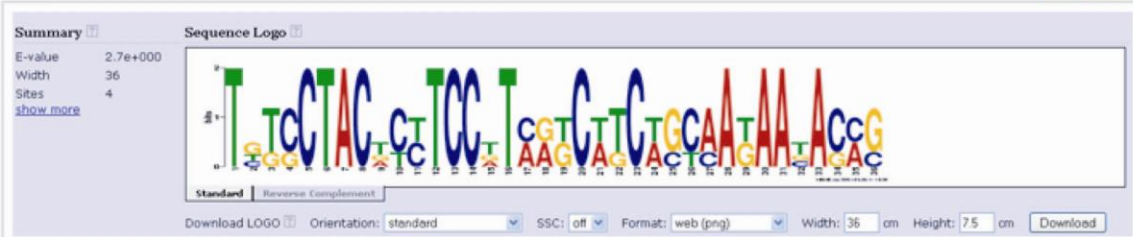
Figure 5.7. Sequences of the putative transcription binding motifs identified in four or more of the 2000 bp upstream regions of teleost *fabp7b* genes as shown in Figure 5.6.

The sequences were identified using TFSEARCH with a threshold of 84.5 (Heinemeyer et al., 1998), and the cis-regulatory elements within the conserved DNA motifs were defined using TOMTOM (Gupta et al., 2007) and MEME (Bailey et al., 2009).



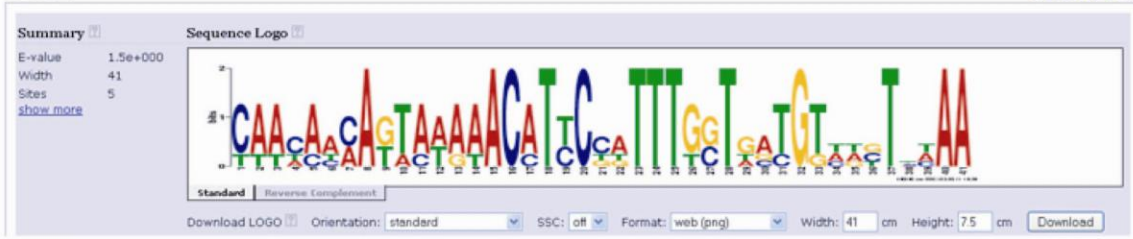
MOTIF 6

[Previous](#) [Next](#) [Top](#)



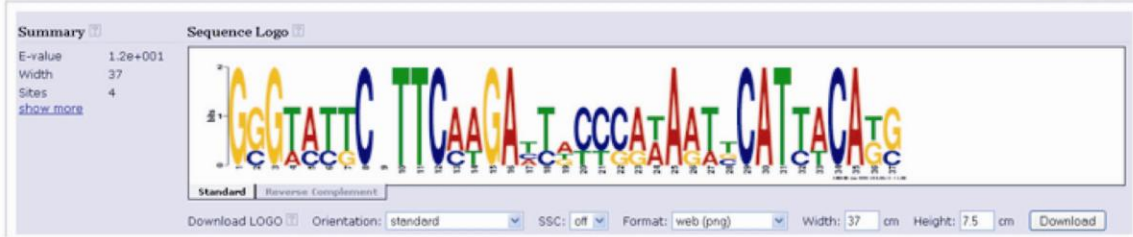
MOTIF 7

[Previous](#) [Next](#) [Top](#)



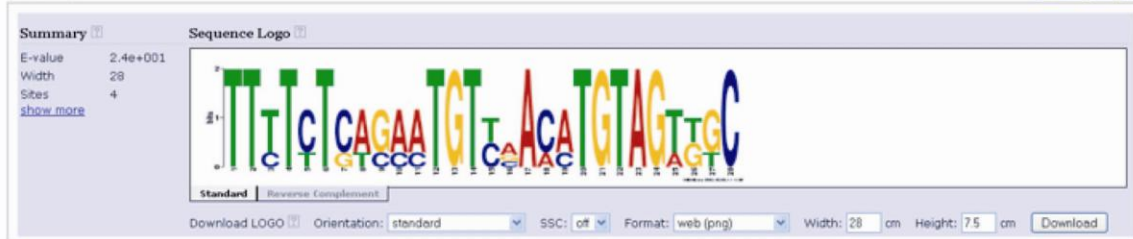
MOTIF 8

[Previous](#) [Next](#) [Top](#)



MOTIF 9

[Previous](#) [Next](#) [Top](#)



MOTIF 10

[Previous](#) [Top](#)

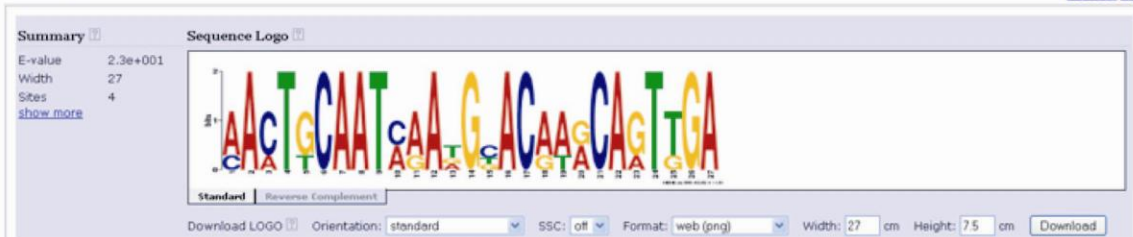
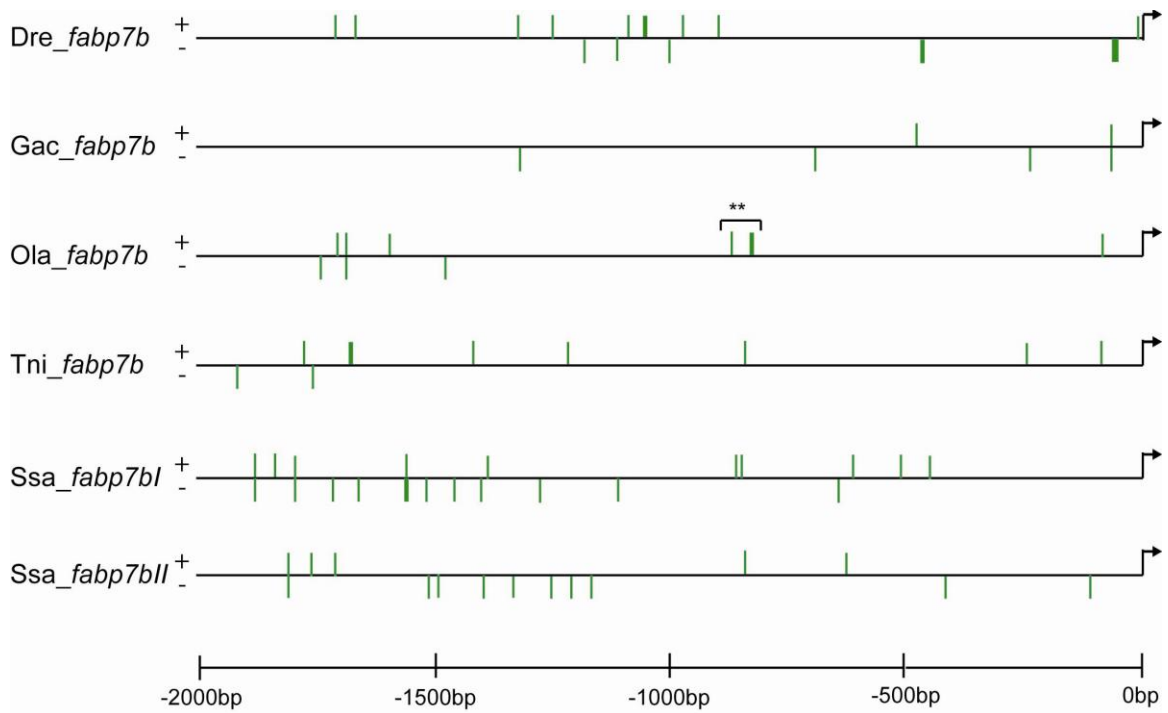


Figure 5.8. Putative POU transcription binding sites in the 2000bp upstream regions of teleost *fabp7b* genes.

The POU sites identified by ** in the medaka *fabp7b* are those identified by Maruyama et al. as being important for expression in brain (Maruyama et al., 2008).

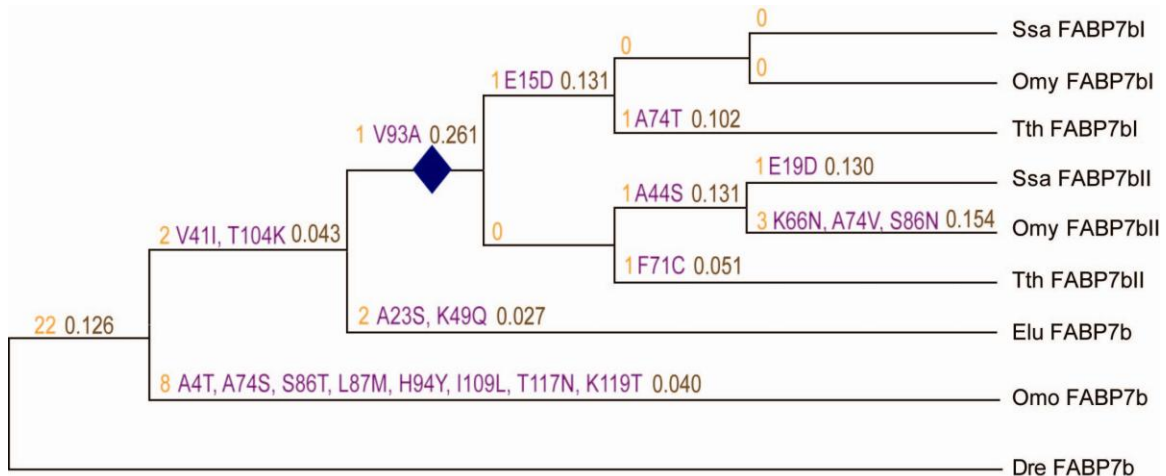


5.8 Evidence of neo-functionalization

Although the RT-PCR analysis suggests that Atlantic salmon *fabp7bI* and *fabp7b* might have sub-functionalized in the liver, I wondered if there was evidence of positive selection in the *fabp7b* genes and neo-functionalization of their gene products in Atlantic salmon, as well as in the rainbow smelt, northern pike, grayling and rainbow trout. The zebrafish *fabp7b* was used as the outgroup of the phylogenetic tree (Figure 5.9). The overall non-synonymous (dN) substitution at non-synonymous site to synonymous (dS) substitution at synonymous site ratios indicate that these proteins in the rainbow smelt, northern pike and the salmonids have experienced purifying selection. There appears to be a trend for fewer amino acid substitutions along the lineages leading to the northern pike and the salmonid FABP7bs (an average of 5) compared to along the lineage leading to rainbow smelt (8). I also examined the nature of the amino acid changes. Balendiran et al. resolved the crystal structures of a complex of the human FABP7 with oleic acid (OA) and a complex of the human FABP7 with DHA, and identified the amino acid residues that contact with the fatty acid ligands (Balendiran et al., 2000). There were no significant amino acid substitutions at those ligand-binding amino acid residues in all the FABP7s of the rainbow smelt, northern pike, grayling, rainbow trout and Atlantic salmon.

Figure 5.9. Evolutionary history of rainbow smelt (Omo) and Northern pike (Elu) FABP7b, grayling (Tth), Atlantic salmon (Ssa) and rainbow trout (Omy) FABP7bI and FABP7bII using the zebrafish (Dre) FABP7b as an outgroup (see Section 3.2.5).

The blue diamond indicates the 4R salmonid specific genome duplication event that gave rise to the FABP7bI and FABP7bII genes in graylings, Atlantic salmon and rainbow trout. The number of amino acid substitutions is given in orange (e.g. there are 22 amino acid sequence differences between the zebrafish FABP7b and the inferred sequence of the FABP7b in the common ancestor of the salmonids and rainbow smelt. The amino acid substitutions that occurred after salmonid genome duplication in the common ancestor of grayling, Atlantic salmon and rainbow trout or in each of these teleost species are shown on the branches of the tree. The list of amino acid substitutions is followed by the dN to dS ratios.



5.9 Conclusion

The 3R WGD that occurred at the base of the teleost radiation gave rise to duplicated brain-type fatty acid binding protein genes, *fabp7a* and *fabp7b*. A loss of the *fabp7a* gene occurred in the common ancestor of northern pike and the salmonids after it had diverged from the rainbow smelt, and then the 4R salmonid-specific WGD subsequently resulted in the *fabp7bI* and *fabp7bII* genes seen in Atlantic salmon, rainbow trout and grayling. This is supported by genetic mapping that placed the Atlantic salmon duplicated *fabp7b* genes on homeologous chromosomes (Lien et al., 2011). Atlantic salmon *fabp7bI* and *fabp7bII* genes are both expressed in several tissues like the *fabp7b* expression pattern in northern pike. However, only Atlantic salmon *fabp7bII*, like its counterpart in northern pike and zebrafish, was expressed in the liver. I compared ~2000 bp upstream of the translational start sites of Atlantic salmon *fabp7b* gene duplicates and found a 62 bp insertion in the *fabp7bI* promoter region, which contains putative CdxA and SRY binding sites. I suggest that this might explain the differential expression of the Atlantic salmon *fabp7b* gene duplicates in liver. There was no evidence of neofunctionalization in the salmonid *fabp7bI* and *fabp7bII* genes based on dN/dS ratios and an examination of amino acid substitutions.

6: Evolution and expression profiling of the Atlantic salmon *fabp3* sub-family

6.1 Introduction

FABP3, the heart-type fatty acid binding protein, is found in the heart and skeletal muscle. It is involved in the uptake of fatty acids and their subsequent transport towards the mitochondrial β -oxidation system. Therefore, exercising, PPAR- α agonists exposure and oscillation of circadian rhythm affect the levels of FABP3 (Coe and Bernlohr, 1998; Furuhashi et al., 2002; Motojima, 2000). *Fabp3* knock-out mice had greater concentrations of free fatty acids in the plasma and were easily exhausted by exercise due to inhibition of fatty acid uptake in the heart and skeletal muscle (Binas et al., 1999). Moreover, FABP3 is abundant in the mammary gland in the course of cell differentiation and formation of ductal structures during lactation (Binas et al., 1992), and has been suggested to be a tumor suppressor as it inhibits the growth of human breast cancer cells (Huynh et al., 1995). However, this proposed function of FABP3 is still controversial because other studies showed that overexpression of *fabp3* did not alter the development pace of mammary gland (Clark et al., 2000; Wallace, 1995). Nevertheless, due to their abundance in these organs, *fabp2* and *fabp3* have been proposed as biomarkers for the detection of intestine and heart injury, respectively (Pelsers et al., 2005).

6.2 Identification of salmonid, northern pike and rainbow smelt *fabp3* transcript sequences

Using the zebrafish *fabp3* nucleotide coding sequence as the query (Ensembl accession: ENSDART00000033724), I searched and only found two unique *fabp3* full-length EST sequences in the rainbow trout database. The sequence identity of the two rainbow trout *fabp3* coding sequences is 94%, whereas the two Atlantic salmon *fabp3* coding sequences are 92% identical to one another. Because of the high percentage sequence identity of the two *fabp3* sequences in Atlantic salmon and rainbow trout, this suggests that these transcribed genes are related to one another by the 4R WGD and speciation of the two species.

Table 6.1. Rainbow trout full-length *fabp3* EST sequences and their GenBank accession numbers as of February 20th, 2012.

	Rainbow trout (Omy)
<i>fabp3aI</i>	CX147347.1
<i>fabp3aII</i>	CA365295.1

6.3 Evolution of Atlantic salmon *fabp3* sub-family

To examine the evolutionary relationships among the Atlantic salmon and rainbow trout *fabp3*s and their orthologs in other teleosts, I first compared the FABP3 protein sequences by a clustalW alignment (Figure 6.1). The FABP3 sub-family is similarly well conserved like the FABP2 sub-family, with 58 invariant amino acid residues in these vertebrate lineages. Nonetheless, stickleback FABP3 has five additional amino acid residues compared to other vertebrate FABP3s, which have 133 amino acid residues. Four of the five additional amino acid residues locate to the N-terminus and the remaining one is an insertion of an aspartate (D94) residue before an invariant glycine in the alignment. I tested by sequencing stickleback genomic PCR products and confirmed that the 5' end of the *fabp3* coding sequence and the codon resulting in the insertion of the aspartate residue are not due to an artefact of the stickleback genome sequencing (Figure 6.2). In fact, the additional four amino acids at the N-terminus of the stickleback *fabp3* result from from a transversion at the second position of what is the equivalent of the start codon in the other vertebrate *fabp3* genes (ATG to AAG; Met to Lys) and the recruitment of a novel start codon from the 5' untranslated region of this gene.

A phylogenetic tree based on the FABP3 protein sequence alignment revealed that there is a single copy of *fabp3* in zebrafish, stickleback, medaka, fugu and tetraodon genomes and two *fabp3*s in both Atlantic salmon and rainbow trout (Figure 6.3). Although it is more parsimonious to state that *fabp3* just experienced a gene duplication in the common ancestor of salmonids, one can interpret the topology of the phylogenetic tree as the 3R WGD event gave rise to two *fabp3* gene duplicates followed by a subsequent gene loss of one of the gene duplicates in the common ancestor of teleosts,

and the salmonid WGD that occurred in the common ancestor of salmonids generated two copies of *fabp3*, which can be found in the extant Atlantic salmon and rainbow trout genomes.

Figure 6.1. Alignment of the vertebrate FABP3 protein sequences.

The species abbreviations are the same as in Figures 4.1 and 5.2.

```

Ssa_FABP3aI      ----MAEAFAGTWNLKDSKNFDEYMKALGVGFATRQVGGMTKPTTIIIEVAGDVTLLKTQS  56
Omy_FABP3aI      ----MAEAFAGTWNLKDSKNFDEYMKALGVGFATRQVGGMTKPTTIIIEVAGDVTLLKTQS  56
Ssa_FABP3aII     ----MAEAFAGTWNLKESKNFDEYMKALGVGFATRQVGSMTKPTTIIIEVAGDMVTLKTQS  56
Omy_FABP3aII     ----MAEAFAGTWNLKESKNFDEYMKALGVGFATRQVGGMTKPTTVIEVAGDVTLLKTQS  56
Dre_FABP3a       ----MADAFIGTWNLKESKNFDEYMKGIGVGFATRQVANMTKPTTIIISKEGDVFTLLKTVS  56
Tni_FABP3a       ----MAEAFAGTWNLVKSEKFDDEYMKELGVGLAMRKMGNLAKPTLSITIEGDKVTLKNSS  56
Tru_FABP3a       ----MAEAFVGTWNLISSSEKFDDYMKELGVGVALRKIGGMAKPSTIIISIDGDKVVLKTS  56
Ola_FABP3a       ----MAEAFVGTWNMKESEKFDDEYMKALGVGFATRQVGNLTkPPTTIIISMDGDRVTVKTS  56
Gac_FABP3a       MAADKVNAFVKGWELKESVKFDDYMKKLGVGAFATRQMGNLTkPPTTIIISVEGDVVTVKTS  60
Hsa_FABP3        ----MVDAFLGTWKLVDKSNFDDYMKSLGVGFATRQVASMTPKPTTIIIEKNGDILTLKTHS  56
Hsa_FABP4        ----MCDAFVGTWKLVSSENFDDYMKVEVGFATRQVAGMAKPNMIIISVNGDVIITIKSES  56
                  : ** * . * : : . * : ** : ** : ** : ** : ** : ** : ** : ** : ** : ** : *
                  : * : ** : ** : ** : ** : ** : ** : ** : ** : ** : ** : ** : ** : **

Ssa_FABP3aI      TFKNTEISFKLGEEFDETTADDRKVKSLITIDG-GKMHVQKWDGKETTlVREVSGNALE  115
Omy_FABP3aI      TFKNTEISFKLGAEFDETTADDRKVKSLITIDG-GKMHVQKWDGKETTlVREVSGNALE  115
Ssa_FABP3aII     TFKNTEINFKLGEEFDETTADDRKVKSLITVDG-GKMHVQKWDGKETTlVREVSGNALE  115
Omy_FABP3aII     TFKNTEINFKLGEEFDETTADDRKVKSLITVEG-GKMIHVQKWDGKETTlVREVSGNALE  115
Dre_FABP3a       TFKSTEINFKLGEEFDETTADDRKVKSVITLDG-GKLLHVQKWDGKETTLLREVSDNNLT  115
Tni_FABP3a       TFKTTEVSFKLGEEFDESTADGRNVKSVVTVED-GKLVHVQKWDGKETSIlVREVVEGNNLT  115
Tru_FABP3a       TFKNTEISFKLGEEFDESTADGRNVKSTVNIVD-GKMHVQKWDGKETTlVREVNDKSLT  115
Ola_FABP3a       TIKNTElTFKLGEEFDETTADDRKVKSlVTLED-GKLVHVQKWDGKETSIlVREVNGNNTL  115
Gac_FABP3a       SLKNTElSFKLDEEFDETTADERKVKsFVKVSDDGKLVHVQKWDGKETSIlVRSVEGNALt  120
Hsa_FABP3        TFKNTEISFKLGVEFDETTADDRKVKSlVTLDG-GKLVHLQKWDGQETTlVRELIDGKLI  115
Hsa_FABP4        TFKNTEISFILGQEFDEVTADDRKVKSTITLDG-GVLVHVQKWDGKSTTIKRKREDDKLV  115
                  : * . ** : . * . * . * . * . * . * . * . * . * . * . * . * . * . *

Ssa_FABP3aI      LTLTLGDVVSTRSYVKAe  133
Omy_FABP3aI      LTLTLGDVVSTRSYVKAe  133
Ssa_FABP3aII     LTLTLGDVICTRSyIkAE  133
Omy_FABP3aII     LTLNLGDVICTRSyIkAE  133
Dre_FABP3a       LTLTLGDIVSTRHyVKAe  133
Tni_FABP3a       LTLTLGNVVCTRHyEKVQ  133
Tru_FABP3a       LTLTLGKVVCTRHyEKAE  133
Ola_FABP3a       LTLTMGDVVCTRrYEkAE  133
Gac_FABP3a       LTLTMDDVVSTRHyVKAe  138
Hsa_FABP3        LTLTHGTAVCTRrYEkEA  133
Hsa_FABP4        VECVMKGVTSTRVYERa- 132
                  :           . ** * :
    
```

Figure 6.2. Chromatograms showing the sequence reads of the 5' end (a) and exon three (b) of the stickleback *fabp3* coding sequences.

a) the transversion occurred at the second position of highlighted codon "AAG", resulting four additional amino acids at the N-terminus of stickleback FABP3. b) The highlighted codon "GAT" is responsible for the inserted aspartate residue in stickleback FABP3 protein sequence.

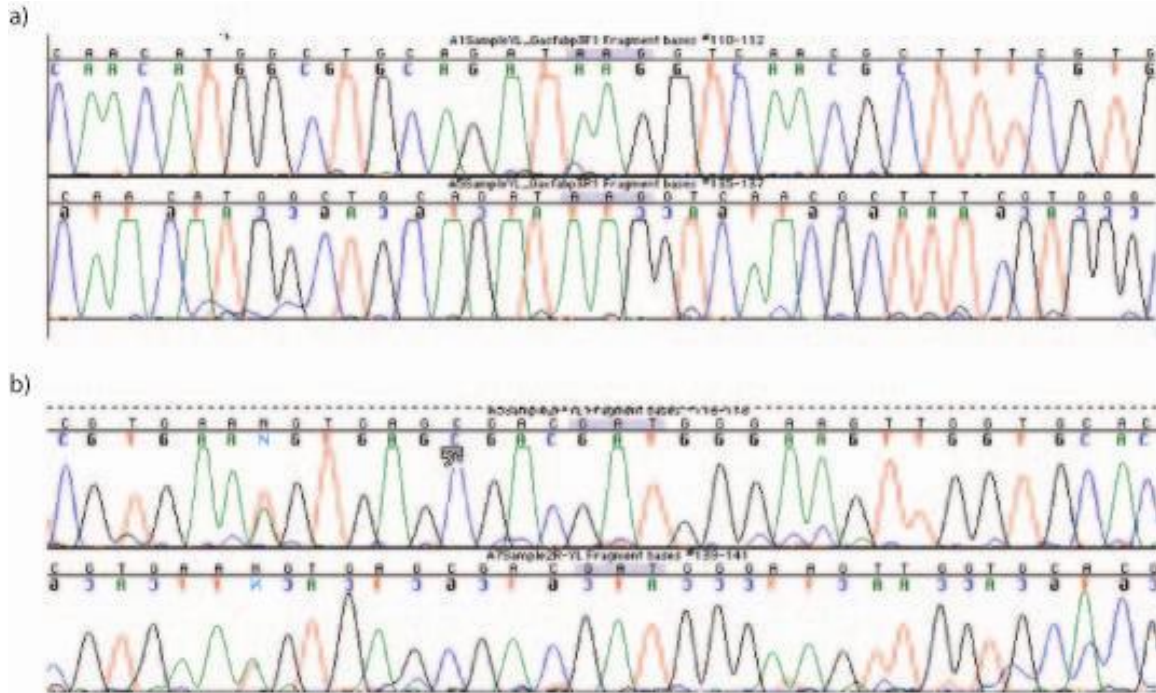
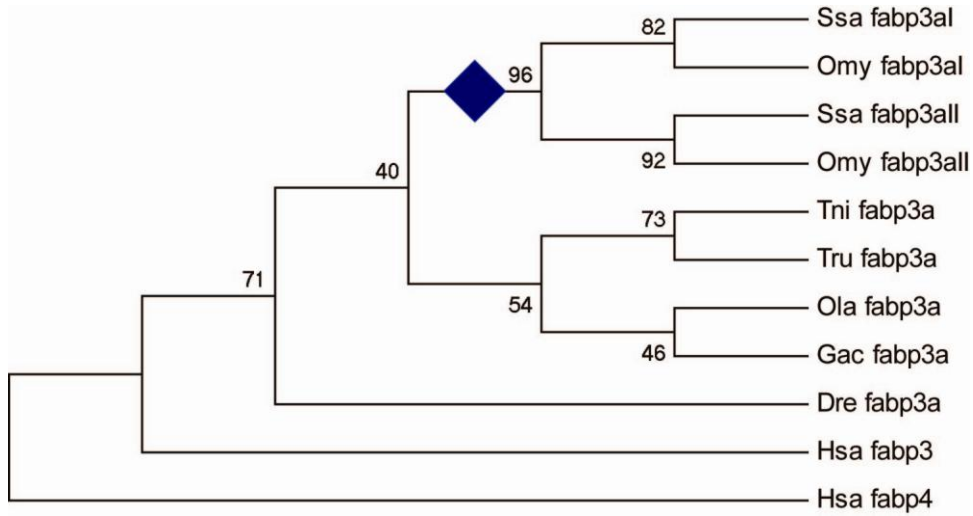


Figure 6.3. Maximum parsimony phylogenetic tree illustrating the evolution of the FABP3 sub-family.

It was built from the protein sequence alignment (Figure 6.1). The blue diamond indicates the 4R WGD event, which gave rise to the FABP3aI and FABP3aII in teleosts.



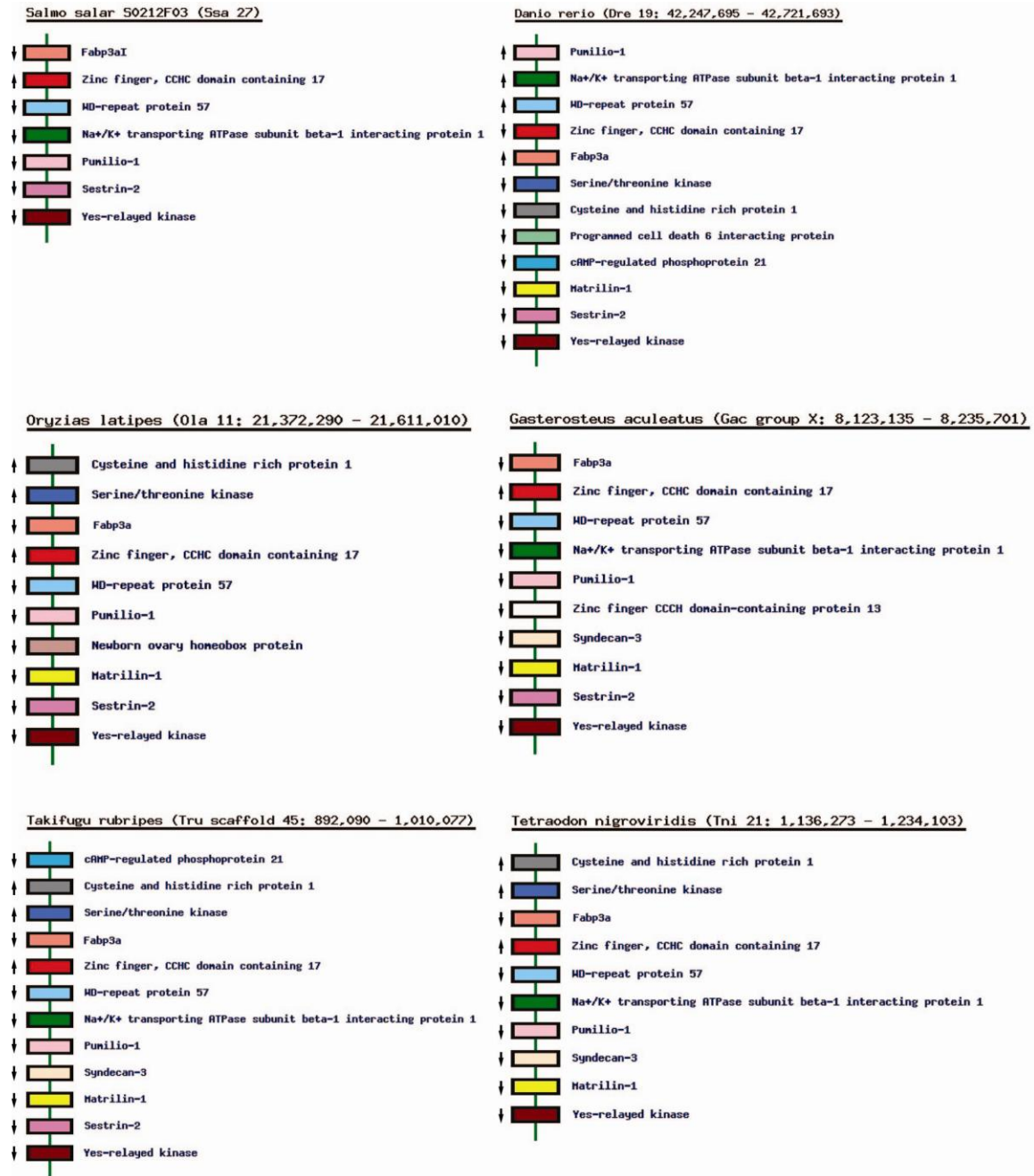
6.4 Conservation of synteny at *fabp3a* loci in teleosts

Unfortunately Atlantic salmon *fabp3aII* appears to be located in a gene-poor or a repetitive region in the genome. Even though the assembled consensus sequence of BAC clone S0812G14 has a length of ~51,000 bp, only *fabp3aII* was annotated from this sequence. Hence, I could not compare the genomic organization of Atlantic salmon *fabp3aII* with the *fabp3aI* locus and the *fabp3* loci in other teleosts. Nevertheless, six genes are conserved and can be found near the Atlantic salmon *fabp3aI* locus and the *fabp3* loci in zebrafish, medaka, stickleback, fugu and tetraodon (Figure 6.4). These genes are: cysteine and histidine-rich protein 1 (*cyhr1*) (absent near stickleback *fabp3* locus), serine/threonine-protein kinase (*dclk3*) (absent near stickleback *fabp3* locus), zinc finger CCHC domain containing 17 (*zcchc17*), WD-repeat protein 57 (*snrnp40*) and sodium/potassium transporting ATPase subunit β -1 interacting protein 1 (*nkain1*) (absent near medaka *fabp3* locus). Since there are some genes that are not present near the *fabp3* loci in some lineages, I expected to find these as possible pseudogenes or unannotated genes in the corresponding genomes. Indeed, using NCBI bl2seq megablast, the exons one, two, three, four, five and six of the stickleback *nkain1* coding sequence (ENSGACT00000007547, Ensembl release verion 54) aligned with 87%-91% identity to the medaka genomic sequence between *snrnp40* and pumilio-1 (*pum1*). Hence, it suggests that there is a *nkain1* pseudogene or unannotated gene in that region of the genome. I also used the medaka *cyhr1* coding sequence (ENSORLT00000014529, Ensembl release verion 54) to search for the presence of parts of *cyhr1* in the stickleback genomic region between *fabp3* and cAMP-regulated phosphoprotein 21 (*arpp-21*). In fact, exons one, five and six aligned with 88%-92% identity with the stickleback genomic

sequence between *fabp3* and *arpp-21*. Again, this suggests that there is a *cyhr1* pseudogene or unannotated gene between stickleback *fabp3* and *arpp-21*. Overall, the genomic organization of *fabp3* loci in teleosts is well-conserved even though they diverged from a common ancestor approximately 320-370 MYA (Christoffels et al., 2004; Jaillon et al., 2004; Meyer and Van de Peer, 2005; Vandepoele et al., 2004)

Figure 6.4. Comparative genomic analysis of *fabp3a* loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.

The arrows indicate where the gene orientations are.



6.5 Expression profiling of Atlantic salmon

To look for signatures of sub-functionalization in terms of expression sites, I first examined the the expression patterns of *fabp3aI* and *fabp3aII* in Atlantic salmon smolt tissues. Out of the fourteen tissues that I examined by RT-PCR, *fabp3aI* transcript was detected in all tissues with the exception of liver (Figure 6.5). Moreover, *fabp3aII* is expressed in brain, eye, spleen, heart, gill, stomach, pyloric caeca, posterior and head kidney, intestine, liver and muscle. In other words, differential expression of *fabp3aI* and *fabp3aII* was observed in ovary and liver; that is, only the *fabp3aI* transcript was detected in ovary and *fabp3aII* is expressed in liver. This broad expression seen in Atlantic salmon *fabp3s* is similar to what was observed in zebrafish and in mammals (Liu et al., 2003a; Veerkamp and Maatman, 1995). In fact, the zebrafish *fabp3* transcript was detected in ovary, heart, skin, intestine, brain, heart, muscle and testis. Based on the phylogenetic analyses (Figure 6.3) and assuming the zebrafish *fabp3* is representative of the ancestral *fabp3* gene in the common ancestor of salmonids, this suggests that Atlantic salmon *fabp3aI* and *fabp3aII* are undergoing sub-functionalization with respect to ovary and liver. I further examined the expression patterns of the two Atlantic salmon *fabp3* gene duplicates in adult tissues using q-PCR analyses, and significant differential expression patterns were observed in liver and muscle (Figure 6.6) even though there was considerable variation in the *fabp3aI* and *fabp3aII* expression levels from different individuals in other tissues examined. Moreover, the finding that both Atlantic salmon *fabp3s* are expressed weaker in heart than they are in skeletal muscle in both RT-PCR and qPCR analysis is interesting. In zebrafish, the *fabp3* gene is expressed in heart and muscle, testis, brain, skin, but to lesser extent than in liver and ovary (Liu et al., 2003a).

However, the relative expression in muscle and heart was not reported. In rat, the *fabp3* gene is expressed in heart and slow twitch muscle to a greater extent than fast twitch muscle (Heuckeroth et al., 1987), which suggests that FABP3 is related to fatty acid metabolism in heart and slow twitch muscle compared to the reliance on glucose for fast twitch muscle. It might be that Atlantic salmon muscle utilizes fatty acids as a source of energy, and it is worthy of further investigation. In summary, the qPCR analyse further confirmed that Atlantic salmon *fabp3aI* and *fabp3aII* are experiencing sub-functionalization as their differential expression patterns were seen in both smolt and adult liver.

Figure 6.5. Expression profiles of *fabp3aI* and *fabp3aII* in Atlantic salmon smolt tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill (G), ovary (O), stomach (St), pyloric caecum (Pc), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M) and testis (T).
Differential expression of *fabp3aI* and *fabp3aII* was observed in liver.

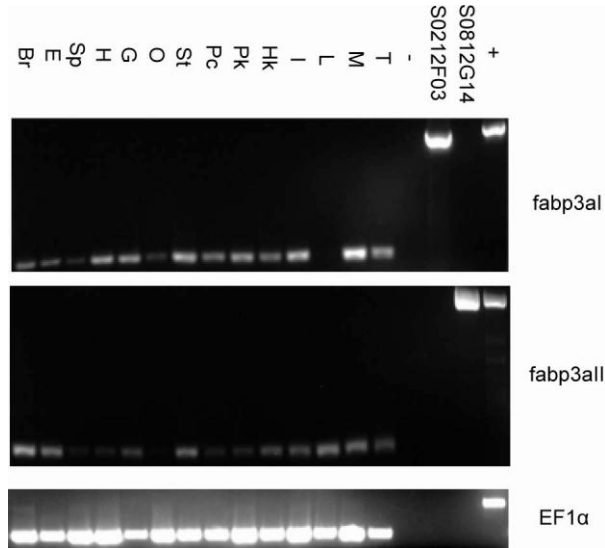
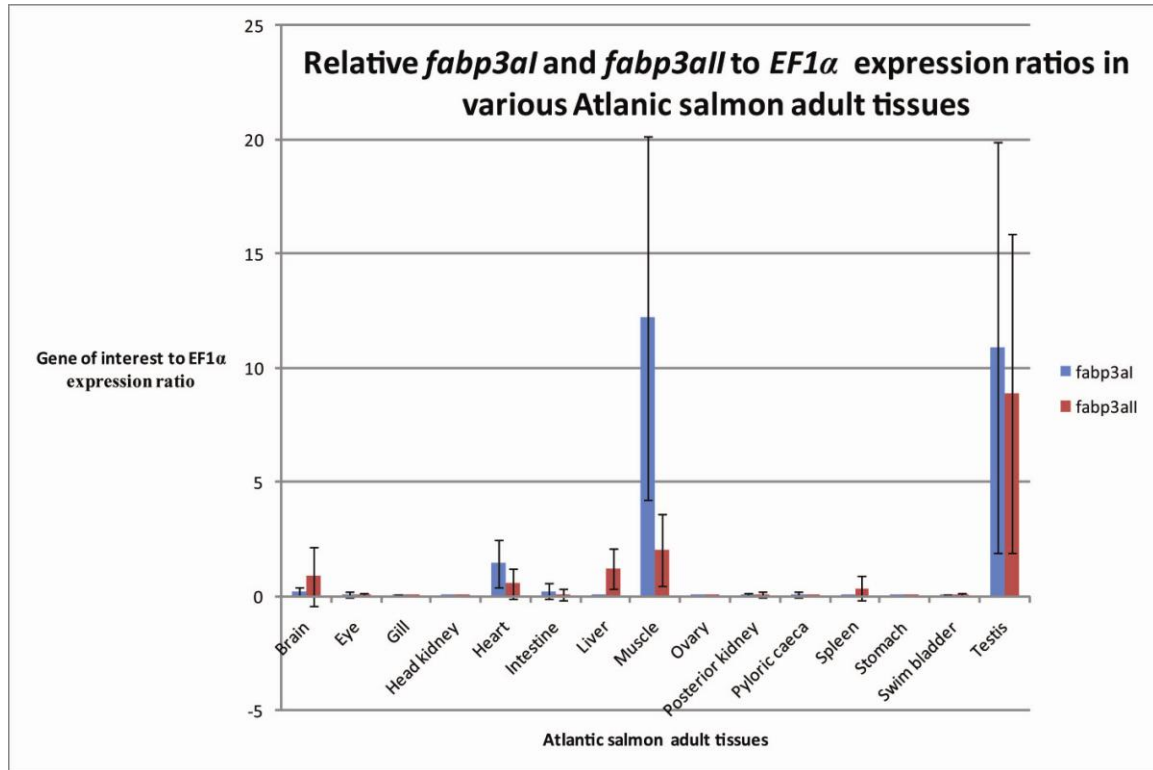


Figure 6.6. qPCR analyses on Atlantic salmon *fabp3aI* and *fabp3aII* expression in adult tissues.

For each adult tissue (n=5), the expression level of *fabp3aI* and *fabp3aII* was compared to a reference gene, elongation 1 factor alpha (*EF1 α*). Paired t-tests suggest that there is significant differential expression of *fabp3aI* and *fabp3aII* in adult liver and muscle ($p < 0.025$).

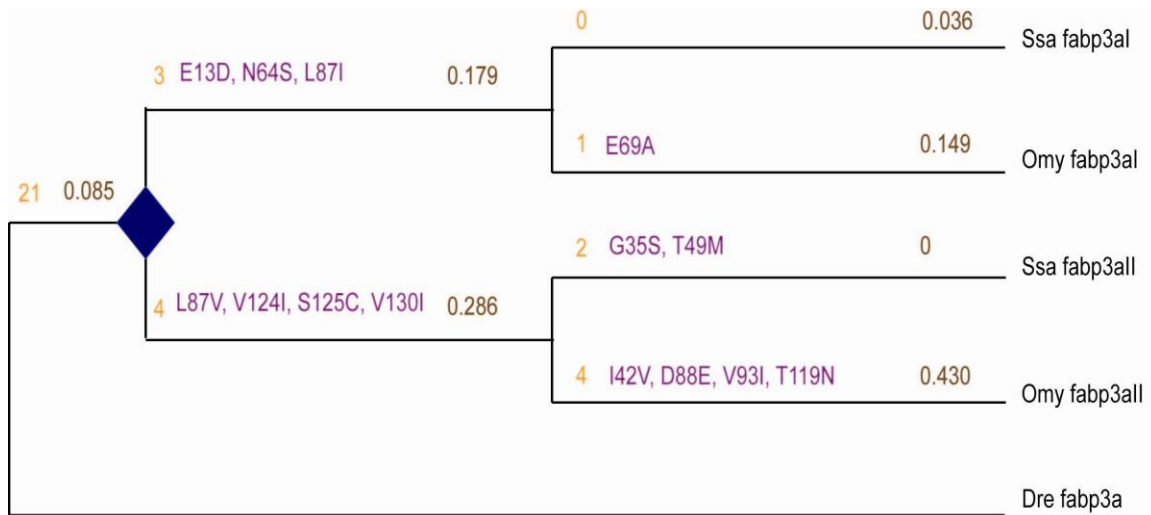


6.6 Evidence of neo-functionalization

With the Atlantic salmon *fabp3aI* and *fabp3aII* genes being differentially expressed in liver, I wondered if they are also experiencing any positive selection, which might be an indication of neo-functionalization if the non-synonymous substitution occurs at residue that affects the protein function. I first looked for signature of positive selection by calculating the overall protein non-synonymous (dN) to synonymous (dS) substitution ratio (Figure 6.7). However, the dN/dS ratios suggest that the evolution of FABP3 from zebrafish to the two salmonids, Atlantic salmon and rainbow trout, are undergoing purifying selection. Another approach I took was to analyse the amino acid substitutions along different lineages, looking for different rates of change or radical substitutions at critical residues. None of the amino acid substitutions shown in Figure 6.7 appear to be radical, with the possible exception of E69A which is seen in rainbow trout FABP3aI, and they do not occur at positions known to be essential for FABP3 function, either structural or with respect to lipid binding (Lucke et al., 2001). In addition, there is no apparent difference in the rates of amino acid substitution in the FABP3aI and FABP3aII lineages in the period after the salmonid-specific gene duplication, but before the speciation event that gave rise to Atlantic salmon and rainbow trout (3 and 4 amino acid changes, respectively). The number of substitutions in Atlantic salmon and rainbow trout FABP3aI and FABP3aII since these species separated varies from 0 to 4. Although there does seem to be an increased rate in the salmonid FABP3aIIs, the small number of changes makes it difficult to make any definitive statement regarding neo-functionalization in these proteins

Figure 6.7. Evolutionary history of Atlantic salmon and rainbow trout FABP3aII and FABP3aII using the zebrafish FABP3 as the outgroup (see Figure 6.3).

The blue diamond indicates the 4R WGD event that gave rise to the FABP3aII and FABP3aII in Atlantic salmon and rainbow trout. The number of amino acid substitutions is given in orange (e.g., there are 21 amino acid substitutions between the zebrafish FABP3 and the inferred FABP3a protein sequence in the common ancestor of salmonids prior to the WGD event. The amino acid substitutions that occurred after the 4R WGD are shown on the branches of the tree as well. This is followed by the protein non-synonymous (dN) to synonymous (dS) substitution ratios (Nei and Gojobori, 1986).



6.7 Conclusion

I characterized two *fabp3* genes in Atlantic salmon. They are named as *fabp3aI* and *fabp3aII* based on their full evolutionary history. The phylogenetic analysis does not support the 3R WGD due to lack of a second *fabp3* gene in the teleosts examined, though one might argue that the 3R WGD at the base of the teleosts first gave rise to *fabp3a* and *fabp3b* following by a loss of *fabp3b* prior to the teleost radiation. Both RT-PCR on smolt tissues and q-PCR analyses on adult tissues showed that Atlantic salmon *fabp3aI* and *fabp3aII* are significantly differentially expressed in liver. There was no evidence to suggesting the Atlantic salmon *fabp3* gene duplicates are undergoing neo-functionalization based on the dN/dS and amino acid substitution analysis. Hence, these genes are experiencing sub-functionalization as a way to maintain their existence in the duplicated Atlantic salmon genome.

7: Evolution and expression profiling of the Atlantic salmon *fabp1* sub-family

7.1 Introduction

The first FABP (FABP1) discovered was the liver-type fatty acid-binding protein. It is abundant in the liver cytoplasm, representing 5% of all the cytosolic proteins in hepatocytes (Haunerland and Spener, 2004). Unlike other FABP sub-families, FABP1 binds to a variety of compounds including fatty acids, acyl-CoA, acylcarnitine, eicosanoids, lysophospholipids, haem, and some steroid hormones and peroxisomal proliferators (Coe and Bernlohr, 1998; Maatman et al., 1994; Veerkamp and van Moerkerk, 1993). In fact, FABP1 can bind to two ligands at the same time via two different binding sites with high and low affinities (Rolf et al., 1995). Moreover, it is important for hepatic lipid metabolism (Martin et al., 2009), and this may be due to the fact that FABP1 interacts with PPAR- α to regulate lipid metabolic genes as a co-activator (Furuhashi and Hotamisligil, 2008). FABP1 knock-out mice have no change in appearance, viability or weight, but do have a slight reduction in fatty acid uptake and serum levels of triacylglycerols and fatty acids (Martin et al., 2003; Newberry et al., 2003).

7.2 Identification of salmonids, northern pike and rainbow smelt *fabp1* transcript sequences

Using the zebrafish *fabp1* nucleotide coding sequences as the queries (Ensembl accession: ENSDART00000006606 and ENSDART00000082286), I searched and found

two, one and one unique *fabp1* full-length EST sequences in rainbow trout, sockeye salmon and rainbow smelt, respectively (Table 7.1). The percentage identity of the two rainbow trout unique full-length *fabp1* coding sequences is 95%. With this high sequence identity, it suggests that the two rainbow trout *fabp1* transcripts are related by the 4R WGD.

Table 7.1. Rainbow trout, sockeye salmon and rainbow smelt full-length *fabp1* EST sequences and their GenBank accession numbers as of February 20th, 2012.

	Rainbow trout (Omy)	Sockeye salmon (Ots)	Rainbow smelt (Omo)
<i>fabp1a</i>			
<i>fabp1b</i>			
<i>fabp1bI</i>	CX039230.1		
<i>fabp1bII</i>	CX034425.1	CD510811.1	CX350756.1

7.3 Gene annotation of *fabp1* genes in some teleosts

As mentioned in Section 3.2 and Table 3.2, I reannotated one of the fugu *fabp1* genes (ENSTRUG00000015614) because I noticed its Ensembl transcript sequence (ENSTRUT00000040056) in version 64 does not end with a stop codon. Hence, I used the zebrafish *fabp1* transcript sequences and BLASTn searched against the EST database in The Gene Index Project (TIGR) version 9 (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>) (Quackenbush et al., 2001) to find fugu *fabp1* transcripts. Indeed, one of the retrieved fugu *fabp1* EST sequences (TIGR Gene ID: NP1061476) aligned with the genomic sequence of Ensembl annotated *fabp1* gene (ENSTRUG00000015614). Therefore, I used this retrieved fugu *fabp1* EST sequence for the remaining of the project and obtained its inferred protein sequence from ExPASy translate tool (<http://web.expasy.org/translate/>). Moreover, I also retrieved an Atlantic cod *fabp1* EST sequence (TC19298) from the Atlantic cod database of TIGR as the phylogenetic tree with this sequence gives a better resolution of the evolution the *fabp1* sub-family (see Section 7.4). The Atlantic cod FABP1 inferred protein sequence was then obtained from the translation of its coding sequence by the ExPASy translate tool.

7.4 Evolution of Atlantic salmon *fabp1* sub-family

I compared the inferred protein sequences of salmonid FABP1s, including rainbow smelt, sockeye salmon, Atlantic salmon and rainbow trout with their orthologs in the species mentioned above (Figure 7.1). The clustalW alignment of the FABP1 protein sequences illustrates that FABP1s range from 125-127 amino acids long. Moreover, there are only 19 conserved amino acid sites. Although there is an annotated *fabp1* gene

(ENSTNIG00000016680) in the tetraodon genome in Ensembl version 64; however, there is no start codon at its 5' end of the predicted exon one. Therefore, based on the genomic sequence available in Ensembl, it suggests that there is a non-functional *fabp1* in the tetraodon genome. It also suggests that the loss of a functional *fabp1* gene in tetraodon is relatively recent because it only misses a start codon that codes for methionine. However, re-sequencing of this genomic region in tetraodon is required to confirm this hypothesis.

A maximum likelihood tree was built from the clustalW alignment of the *fabp1* coding sequences from the species mentioned above (Figure 7.2). Stickleback, medaka and fugu all possess only a single copy of *fabp1* in their genomes, whereas zebrafish contains both *fabp1a* and *fabp1b* in its genome (Sharma et al., 2006). Interestingly, there is a second copy of *fabp1b*, *fabp1b.2*, in the zebrafish genome, Karanth et al. (2009) suggested it to be the result of a tandem duplication of *fabp1b.1* (Karanth et al., 2009). Both Atlantic salmon and rainbow trout possess two copies of *fabp1*; however, the two Atlantic salmon *fabp1s* correspond to *fabp1a* and *fabp1b*, whereas the two rainbow trout *fabp1s* are grouped in the *fabp1b* clade. The presence of two *fabp1b*s in rainbow trout genome suggests that a *fabp1b* gene or a WGD event in rainbow trout. Moreover, Atlantic cod, rainbow smelt and sockeye salmon appear to possess a single copy of *fabp1*, but there might be more *fabp1* genes in these species because of the low number of ESTs available in these species. The lack of a second *fabp1* gene, *fabp1a*, in several fish species may be because that these have not been annotated and this must be resolved before the evolution of this gene sub-family can be determined.

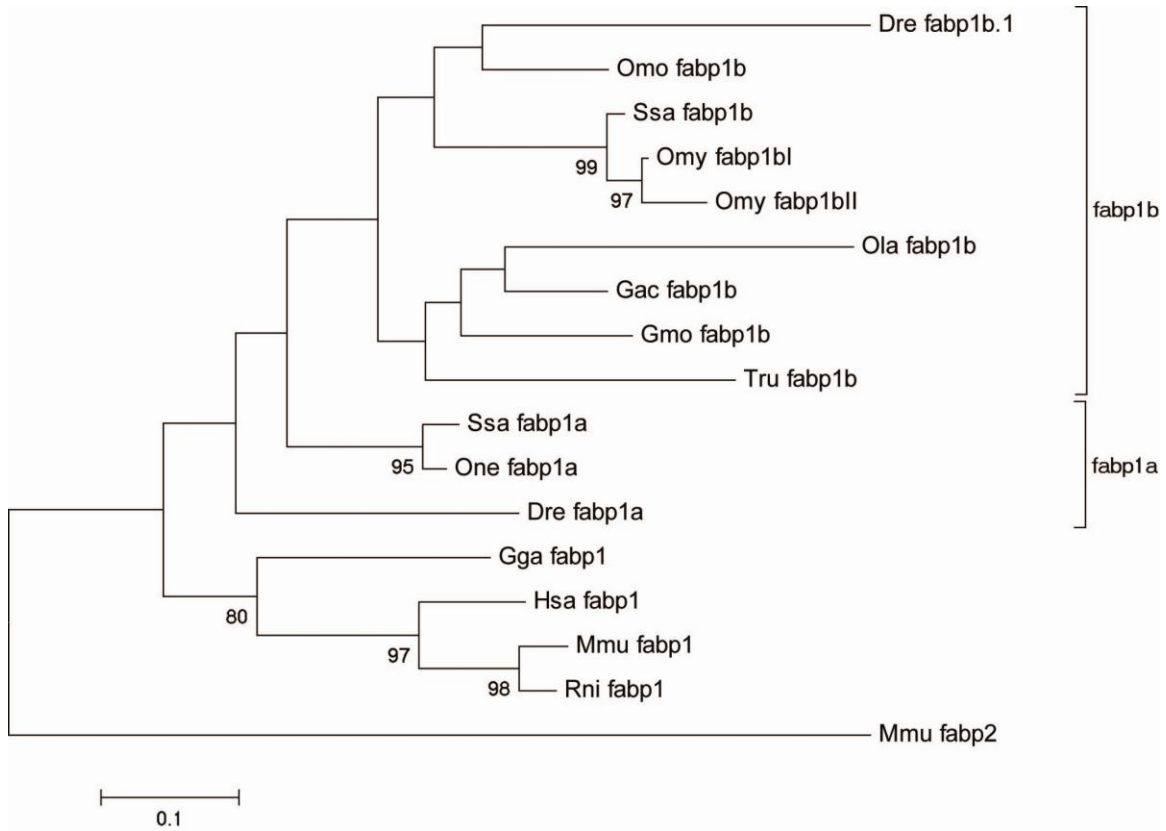
Figure 7.1. Alignment of the FABP1 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts, chicken, mouse, rat and human.

The species abbreviations are the same as in Figures 4.1 and 5.2 with addition of *Gadus morhua* (Gmo). Protein size varies and only a few number of amino acid residues are conserved in the FABP1s from different vertebrate species.

Ssa_FABP1b	MSFSGKYQMETHENFESFMEAI GLPDELI QEGKD IKS ISEI EETGNHFKVTVTGT KILT 60
Omy_FABP1bI	MSFSGKYQMESHDN FESFMEAVGLPDELI QEGKD IKS ISEI EETGDHFKVTVTGT KILT 60
Omy_FABP1bII	MSFSGKYFMESPDR FESFLEAVGLPDELI QEGQDLQSLSEI EETGDPFQVTVPPGT KILT 60
Omo_FABP1b	MSFSGKYQLETHENFEAFMKAI GLPDDLI QKQKD IKS VSEI EENGNSFKVTVTGT KVL C 60
Dre_FABP1b.1	MSFTGKYQLESQEGFVEFMKAVGLPDDMI EKGKD IKS VSEI EENGNGFKVTVTGT SKVL T 60
Dre_FABP1b.2	MPFSGKFELEKYEKGWEEFMKAI GH-QELLK KANDGKT QFEI QENGN DYCLT TRSGGKVL N 59
Gmo_FABP1b	MSFTGKYQLESQENFESFMKAI GLPDELI QKQKD IKS LSEI EENGDNFKVTVTGT KVL V 60
Gac_FABP1b	MSFTGKYQLETQENFEAFMKAVGLPDDLI QKQKD IKS ISEI EENGDNFKVTVTGT SKVL V 60
Ola_FABP1b	MSFTGKYQLESSENFEAFMKAVGI SDENI QLT KGLKS IT EMEENGDNFRVTVTGT SHVI I 60
Tru_FABP1b	MSFSGKYQQVQSENFEFPMKAI GLPDEVI QQVKE LKSTSEI EQNGNDFKIT ITTGPKVTV 60
Ssa_FABP1a	NAFTGKYQLESQENFEFPMKAI GLPDDLI QKQKD IKS LSEI EQNGDHFKVTVTGT KVL V 60
Orn_FABP1a	NAFTGKYQLESQENFEAFMKAI GLPDDLI QKQKD IKS LSEI EQNGDHFKVTVTGT KVMV 60
Dre_FABP1a	NAFTGKYQLESSENFEAFMKAVGV PDD EEVERGKD IKS ISEI HQDGKDFKVTAGT KVI L 60
Gga_FABP1	MSFTGKYELQSHENFEFPMKAL GLPDDQI QKQKD IKS ISEI VQNGNKFKIT ITTGSKVMT 60
Mmu_FABP1	MNFSGKYQLQSQENFEFPMKAI GLPEDLI QKQKD IKS VSEI VHEGKKIKLIT ITYGPKVVR 60
Rno_FABP1	MNFSGKYQVQSQENFEFPMKAMGL PEDLI QKQKD IKS VSEI VHEGKKIKLIT ITYGSKVI H 60
Hsa_FABP1	MSFSGKYQLQSQENFEAFMKAI GLPEELI QKQKD IKS VSEI VQNGKHKFKFIT ITAGSKVI Q 60
	* **::: . : : *:::* : : : : : * : . * . * * : :
Ssa_FABP1b	NSFTIGQETELE SPTGKVN SVVTREG-NKLM AFLNGIE YVTELT DANTLVN TLTLP G-I 118
Omy_FABP1bI	NSFTIGQETELE SPTGKVN SVVTREG-NKLT AI LNGIE YVTELT DANTLVN TMTL SG-M 118
Omy_FABP1bII	NSFTLGQETELE SPTGKVN SVVTREG-NKLT AI LNGIE YVTELT DANTLVN TMTL SG-M 118
Omo_FABP1b	NNFTIGQEAEL ETLTGEKVKSV M RDG-NKLVSLK GIE SI TELVDANT IVN TMTL VAG-I 118
Dre_FABP1b.1	NSFTIGQEA D I E T L T G E K V K T I V N R E G - N K L K V V L N R I T S I T E L V D A N T L V N T L T L G G - L 118
Dre_FABP1b.2	NSFTIGQDTEIQMLSGDKVKT VVHKDG-QK LK AALGNITLTWELLDANTLLITLSTADNI 118
Gmo_FABP1b	NTFTVQGQEADEL TVTGEKIKH AVVKREG-NKLVSLK GIE SVTELV DANT IVN TMTL G G-I 118
Gac_FABP1b	NTFKIGQEA E I T I T G E K V K A V V K R D G - N K L T V C L K G M E S V T E L L D A N T I V N T I T L G S - I 118
Ola_FABP1b	NSFTIGQET E M L S V T G E K I K T V V K K E G - T K L M M S L K R A D I V T E L V D G N T L V T M T L G D - I 118
Tru_FABP1b	NKFTIGKET E M D T I T G E K I K T V F H L D G - N K L K V S L K G I E S V T E L A D P N T I T - - M T L G D - V 116
Ssa_FABP1a	NSFTVQGQEAEL ETLTGEKIKSTVNLV G-NKLMVSLK GIDSVTELDG-DTISNTLSFNG-I 117
Orn_FABP1a	NSFTVQGQEAEL ETLTGEKIKSTVNLV G-NKLMVSLK GVE SVTEFNG-DTIIATMTLGP-I 117
Dre_FABP1a	YSFTVGECELE T F T G D R A K T V V Q M D G - N K L T A F V K G I E S V T E F N G - D T I I A T M T L G P - I 117
Gga_FABP1	NEFTIGECEM ELLTGEKAKCI VNMEGNNKLVANL KGLK SVTE L NG-DTITHTMTKGD-L 118
Mmu_FABP1	NEFTLGECELE TMTGEKVKAVVKLEGDNKMVTT FKG I K S V T E L N G - D T I T N T M T L G D - I 118
Rno_FABP1	NEFTLGECELE TMTGEKVKAVVKMEGDNKMVTT FKG I K S V T E F N G - D T I T N T M T L G D - I 118
Hsa_FABP1	NEFTVGECELE TMTGEKVT VVQL E G D N K L V T T F K N I K S V T E L N G - D I I T N T M T L G D - I 118
	* : * : : : : : : * : * : . * : . : : : : :
Ssa_FABP1b	SYKRTSKRM- 127
Omy_FABP1bI	SYKRTSKRM- 127
Omy_FABP1bII	SYKRTSKRM- 127
Omo_FABP1b	AYKRTSKRM- 127
Dre_FABP1b.1	VYKRISKRVA 128
Dre_FABP1b.2	AYKRFSKRUV 128
Gmo_FABP1b	VYKRISKRV- 127
Gac_FABP1b	VYKRTSKRM- 127
Ola_FABP1b	VYKTYKRI- 127
Tru_FABP1b	VYKRTSKRM- 126
Dre_FABP1a	VYKRISRRI S 127
Ssa_FABP1a	VYKRISKRI- 126
Orn_FABP1a	VYKRISKRI- 126
Gga_FABP1	TYKRISKRI- 127
Mmu_FABP1	VYKRVSKRI- 127
Rno_FABP1	VYKRVSKRI- 127
Hsa_FABP1	VFKRISKRI- 127
	: * : * :

Figure 7.2. Maximum likelihood phylogenetic tree of *fabp1* built from the nucleotide coding sequence alignment.

Only bootstrap values that are greater than 50 are shown.

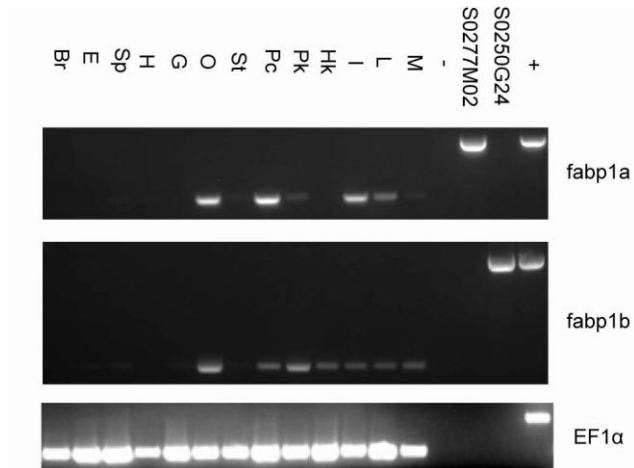


7.5 Expression profiling of Atlantic salmon *fabp1* sub-family

Atlantic salmon *fabp1a* is expressed in eight out of thirteen tissues I examined and the *fabp1b* transcript was detected in eye, spleen, gill, ovary, stomach, pyloric caeca, posterior and head kidney, intestine, liver and muscle (Figure 7.3). The Atlantic salmon *fabp1a* expression profile is broader compared to its ortholog's expression in zebrafish (Sharma et al., 2006). However, both Atlantic salmon and zebrafish *fabp1a* and *fabp1b* are expressed in intestine (Karanth et al., 2009; Sharma et al., 2006). Moreover, both Atlantic salmon and zebrafish *fabp1b* genes are transcribed in liver and ovary. Nevertheless, I observed differential expression patterns of *fabp1a* and *fabp1b* in eye, gill and head kidney whereas there is more differential expression between the zebrafish *fabp1a* and *fabp1b* genes.

Figure 7.3. Expression patterns of *fabp1a* and *fabp1b* in Atlantic salmon smolt tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill, (G), ovary (O), stomach (St), pyloric caecum (Pc), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M), negative control (-) and Atlantic salmon genomic DNA (+). BAC clones S0277M02 and S0250G24 are the representative BACs of Atlantic salmon *fabp1a* and *fabp1b* (Table 3.3). Differential expression patterns of Atlantic salmon *fabp1a* and *fabp1b* were seen in eye, gill and head kidney.



7.6 Conclusion

Given that the 3R WGD occurred in the common ancestor of teleosts gave rise to zebrafish *fabp1a* and *fabp1b* (Sharma et al., 2006), the phylogenetic tree analysis suggests that a *fabp1a* gene loss occurred in the common ancestor of stickleback, medaka, fugu and tetraodon whereas zebrafish was able to maintain both *fabp1* genes. In addition, it is known that zebrafish *fabp1b.1* gave rise to *fabp1b.2* by tandem duplication in this lineage (Karanth et al., 2009). With the *fabp1* coding sequences that are available in the teleosts and salmonids at this point of time, it is unknown what events happened that could result in the *fabp1a* and *fabp1b* genes in Atlantic salmon and the two *fabp1b* genes in rainbow trout. The most parsimonious explanation is that the 4R WGD did not occur and there was only a *fabp1b* gene duplication event occurred in the rainbow trout lineage. However, one may argue that there was the 4R WGD event occurred in the common ancestor of salmonids with one of the *fabp1a* and one of the *fabp1b* gene duplicates lost in Atlantic salmon and two subsequent *fabp1a* gene loss in the rainbow trout lineage. To resolve this ambiguity, more salmonid *fabp1* coding or protein sequences are required in order to gain more details on the evolution of the *fabp1* sub-family.

8: Evolution and expression profiling of the Atlantic salmon *fabp6* sub-family

8.1 Introduction

FABP6, the ileal-type fatty acid-binding protein, is found in the ileum, ovary and adrenal gland (Amano, 1993; Crossman, 1994). Interestingly, FABP6 has a higher affinity for bile acids than fatty acids (Sacchettini et al., 1990), and its amount within the cell is regulated by bile acids through the action of the farnesoid-X-receptor (FXR) (Foucand, 1998; Zaghini, 1999), which is a receptor for bile acids. It is hypothesized that FABP6 functions as a cytosolic receptor for bile acids transported by the sodium dependent action of the ileal bile acid transporter (Kramer et al., 1993).

8.2 Identification of salmonids, northern pike and rainbow smelt *fabp6* transcript sequences

Using the zebrafish *fabp6* nucleotide coding sequence as the query (Ensembl Accession: ENSDART00000065448), I searched and found one unique *fabp6* full-length EST sequence in rainbow trout and in northern pike (Table 8.1).

Table 8.1. Rainbow trout and northern pike full-length *fabp6* EST sequences and their GenBank accession numbers as of February 20th, 2012.

	Rainbow trout (Omy)	Northern pike (Elu)
<i>Fabp6b</i>	CX139611.1	GH259361.1

8.3 Gene annotation of *fabp6* genes in some teleosts

As mentined in Section 3.2 and Table 3.2, I re-annotated the tetraodon *fabp6a*, both stickleback *fabp6s* and medaka *fabp6b*. For the re-annotations of tetraodon *fabp6a* and stickleback *fabp6a* and *fabp6b*, I obtained their genomic sequences from Ensembl version 64 and used Genscan (<http://genes.mit.edu/GENSCAN.html>) to predict the the coding sequences of these genes. As for the reannotation of medaka *fabp6b*, I used the fugu *fabp6b* transcript sequence (ENSTRUT00000010446) as the query and BLASTn searched against the medaka EST database in the Gene Index Project (TIGR) version 9 (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>) (Quackenbush et al., 2001) to search for a medaka *fabp6b* transcript. In fact, there was a medaka *fabp6b* transcript (TC138134) and it aligned with the genomic sequence of the Ensembl *fabp6b* gene (ENSORLG00000012622). Hence, I used these Genscan predicted coding sequences and the TIGR medaka *fabp6b* transcript sequence and translated them into protein sequences using ExPasy translate tool (<http://web.expasy.org/translate/>), and used them for the rest of the project.

8.4 Evolution of Atlantic salmon *fabp6* sub-family

I retrieved the *fabp6* transcripts and protein sequences of the five sequenced teleosts species and a tetrapod species, chicken, from the Ensembl release version 64. I compared these FABP6 protein sequences with the inferred amino acid sequences of Atlantic salmon and rainbow trout FABP6s by a clustalW alignment (Figure 8.1). Even though the protein sizes vary from 125 to 131 amino acids in length, there are 24 invariant residues in these vertebrate FABP6s. This indicates that the FABP6 sub-family is less conserved compared to the FABP2 and FABP3 sub-families in terms of protein

size and amino acid residues, but more conserved in compare to the FABP1 sub-family in terms of numbers of conserved amino acids. A phylogenetic tree based on the alignment of the coding sequences of these vertebrate *fabp6s* reveals that medaka, stickleback, fugu and tetraodon possess two *fabp6s*, which I named *fabp6a* and *fabp6b* (Figure 8.2). This supports the 3R genome duplication event giving rise to the *fabp6as* and *fabp6bs* that are found in these teleost species today. Although both protein sequence alignment and the phylogenetic tree suggest that the zebrafish *fabp6* (ENSDARG00000044566) and two of the three Atlantic salmon *fabp6s* are more closely-related to the other four teleosts (medaka, stickleback, fugu and medaka) *fabp6bs* and hence should be grouped as a *fabp6b* type gene, the comparative genomic analysis of the genome organization of the *fabp6a* loci in these species contradicts this interpretation (see Section 8.5). A similar explanation applies to the clustering of the remaining Atlantic salmon *fabp6* and the rainbow trout *fabp6* genes to the four sequenced teleosts (medaka, stickleback, fugu and tetraodon) *fabp6as* in the protein sequence alignment as well as the phylogenetic tree analysis. I suspect the comparative genomic analysis provides more reliable data concerning the grouping of the *fabp6s* in teleosts because it is unlikely that there was a double cross-over to exchange of the *fabp6* gene duplicates (*fabp6a* and *fabp6b*) between the homeologous chromosomes unless the *fabp6* flanking sequences in both chromosomes are highly similar. Hence, the contradiction of the protein sequence alignment and phylogenetic tree analysis and the comparative genomic analysis may be due to an artefact of two independent gene modifications resulted from natural selection at two different time points of evolution of teleosts. The first *fabp6* gene modification occurred at the ancestral *fabp6b* in the common ancestor of medaka, stickleback, fugu

and tetraodon, whereas the second *fabp6* gene modification occurred to the *fabp6a* in the common ancestor of the zebrafish and Atlantic salmon. Therefore, Atlantic salmon *fabp6b* appears to be more closely-related to stickleback, medaka, fugu, tetraodon *fabp6as*.

Moreover, there are two *fabp6a* type genes, *fabp6aI* and *fabp6aII*, in Atlantic salmon. This supports at least another *fabp6* gene duplication in this species or more probably the existence of the 4R genome duplication. Interestingly, there is only a single transcriptional *fabp6* found in rainbow trout. Hence, the fact that Atlantic salmon possesses three *fabp6s* and rainbow trout contains a single *fabp6* suggests that gene duplicates in different post-WGD species encountered different natural selection pressures resulting in different numbers of gene duplicates retained in the post-WGD species genomes. Overall, the phylogenetic tree may be interpreted as the 3R genome duplication event gave rise to *fabp6a* and *fabp6b* in teleosts following by the salmonid-specific genome duplication (4R) that occurred in the common ancestor of salmonids. Several subsequent gene losses occurred later, including the *fabp6b* gene loss in the zebrafish and in Atlantic salmon lineages and three *fabp6* gene losses (both *fabp6as* and a *fabp6b*) in the rainbow trout genome.

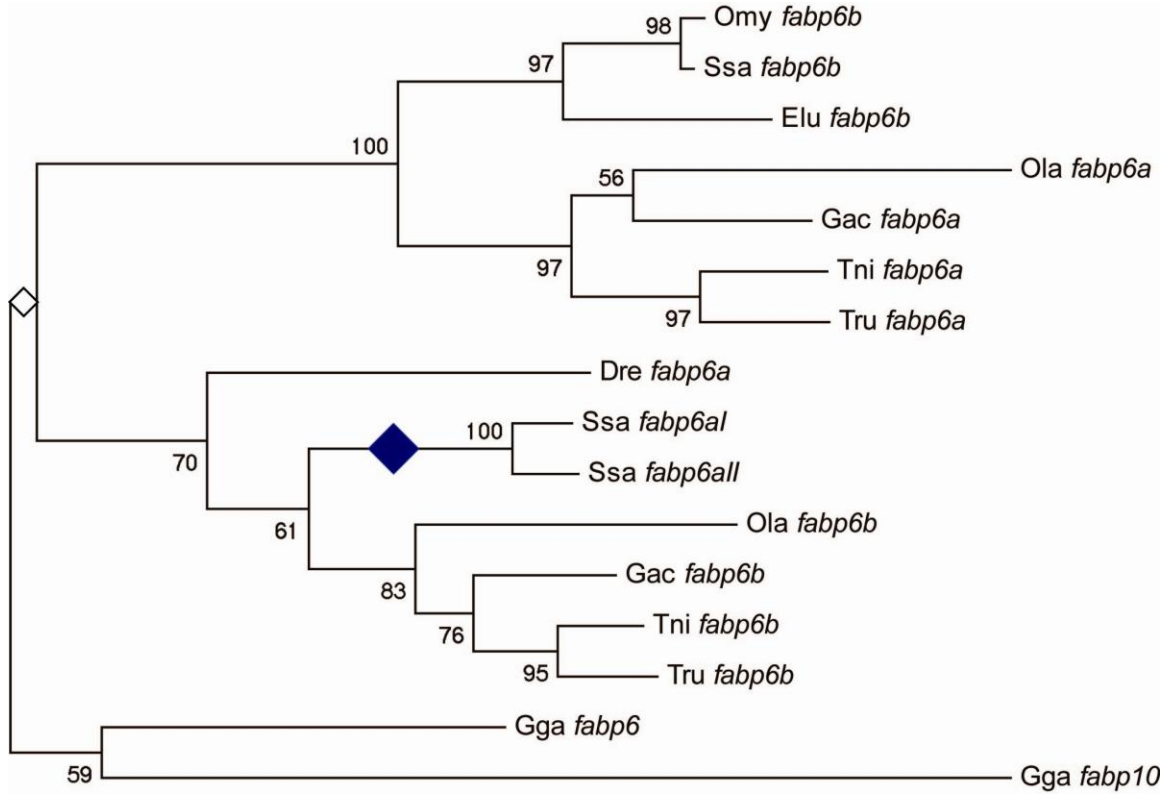
Figure 8.1. Alignment of the FABP6 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and mouse.

The species abbreviations are the same as in Figures 4.1 and 5.2.

Ssa_FABP6aI	MAFAGKWEETQEGYEEFCKLVGIPDD I IEKGRD YKLVTEV VQNGNEF SWSQLYPTNKS	60
Ssa_FABP6aII	MAFAGKWEETQEGYDEFCCKLLGIPDD I IEKGRD YKLVTEV VQNGNEF SWSQLYPTNKT	60
Dre_FABP6a	MAFAGKWEETQEGYEPFCKLLGIPDD V IAKGRD FKLVTET VQNGDDF TWQYYPNNHVV	60
Gac_FABP6a	MAFAGKYELESQENYEEFLAAIGLLNAKT ----DHKVITEV LQDGGDF TWQSIIPN -WTW	55
Ola_FABP6a	MKFAKYELESQENYEEFLAIDLVAKT ----EHKVITEV VQDGDND TWQSIIPN -WSW	55
Tni_FABP6a	MTFTGKYELETQENYEEFLDAIGLLSART ----DHKVITEV SQDGDND TWQSIIPH -WTW	55
Tru_FABP6a	MTFTGKYELETQENYEEFLAIGLLGART ----DHKVITEV VQNGDND TWQTIIPN -WTW	55
Gac_FABP6b	MAFAGKWEETQEGYDAFCCKLLGIPND V IAKGRD YKLVTEV TQDGDND SWSQIYPTNARV	60
Ola_FABP6b	MTFAGRWEIESQEGYDFCKVVGIPAD V IAKGRD YKAVTEV TQNGDND TWQYIYPT -IKV	59
Tni_FABP6b	MAFAGRWLTETQEGYDDFCCKLLGIPAD I IEKGRD YKMITEV TQDGDND SWSQIYPT DAKV	60
Tru_FABP6b	MAFTGRWETENQEGYDFCKLLGIPDD I IEKGRD YKLVTEV TQDGETF SWSQIYPTNARV	60
Elu_FABP6b	MAFSGKYELDCQENYEEFLAALGFSNAKT ----DFKVITEV IQEGDND TWSQIIPN -WTW	55
Ssa_FABP6b	MAFSGTYELESQENYDFLEAIGFANAKT ----DFKVITEV LQEGDND TWSQIIPN -WTL	55
Omy_FABP6b	MAFSGTYELESQENYDFLEAIGFAHAKT ----DFKVITEV LQEGDND TWSQIIPN -WTL	55
Gga_FABP6	MAFTGKYEFESDENYDDVFKIGLIPAD K IEMGRD CKLVTEV VQNGDND TWQHPFGGRIT	60
	* * * : : : . . * : * : : : : * : * : * : * : * : * : * : * : * : * : * : *	
Ssa_FABP6aI	TNKFV IDQECIDMET IGGKKFKATVME GGLSTKFPKYHHT SEISGGKLIETSV ----G	116
Ssa_FABP6aII	SNKFV IDQECIDMET IGGKKFKATVME GGLSTKFPNYHHT SEISGGKLIETSIA ----G	116
Dre_FABP6a	TNKFV VGRKSDMET VGGKKFNGIVSME GGLTISFRKYQQT TEISGGKLVETSTASGAQG	120
Gac_FABP6a	SNKFT VQDCEL TTMGVKFTAAVTME GGLSIPFPQYHFT AELIEDKLVMICTLP G-EK	114
Ola_FABP6a	TNKFV VQDCEL LTMGNKFKALVTE NGKLSIPFPQYHFT AEIIDDKLVMTCTVP G-EK	114
Tni_FABP6a	TNFTSVGRECELVTMGTKFKAFVMD DGRISVQFPQYHFT AEVSGDKLIMNCTIP G-EK	114
Tru_FABP6a	SNTFSVQDCELVTMGIQFKAFVME NGKLSVQFPQYHFT AEVNGDKLIMNCTIP G-EK	114
Gac_FABP6b	TNNFT IGKCIDMET IGGKKFKATVME GGLTIAFPNYHHT TEISGGKLIETSTA ----A	116
Ola_FABP6b	TNKFV IGKSEME SLGGKFKATVME GGLTVNFPNYHHT TEISGGKLIETCKA ----G	115
Tni_FABP6b	TNFTV VGRKSDMET IGGKKFKATVME GGLSVTFPNYHHT SEISGGKLIETSKA ----G	116
Tru_FABP6b	TNFTV IGKCIDMET IGGKKFKATVME GGLSVTFPNYHHT SEISGGKLIETSKA ----G	116
Elu_FABP6b	SNKFT IGKCELEDMMRSKFMATATME EGKISVPFPQYHFT AEISGKLIITCTIP G-EK	114
Ssa_FABP6b	TNKFV IGKCELESMLRSKFTVTVME GGLIIPFPQYHFT AEISGDKLIMLCTTAG-EK	114
Omy_FABP6b	TNKFV IGKCELESMLRSKFTVTVME GGLIIPFPQYHFT AEISGDKLIMLCTTAG-EK	114
Gga_FABP6	TNSFT IDREADMETMGGKKFKATVME GGLVADFPNYHHT AEISGKLVETSTAS ----S	116
	: * . * : : : : : : : : * : : : * : * : * : * : * : * : * : * : * : *	
Ssa_FABP6aI	SVVLKRTSKKI- 127	
Ssa_FABP6aII	SVMLKRTSKKI- 127	
Dre_FABP6a	TAVLVRTSKKV- 131	
Gac_FABP6a	SVTFRTSRRI- 125	
Ola_FABP6a	GVTFRVWKRI- 125	
Tni_FABP6a	GVTFRTRRW- 125	
Tru_FABP6a	GVTFRISRRL- 125	
Gac_FABP6b	SVVLKRTSKKI- 127	
Ola_FABP6b	SVVLRKVSKKI- 126	
Tni_FABP6b	SVVLKRTSKKL- 127	
Tru_FABP6b	SVVLKRTSKKL- 127	
Elu_FABP6b	GVTMKRINKRI- 125	
Ssa_FABP6b	GVTMKRINKRI- 125	
Omy_FABP6b	GVTMKRINKRI- 125	
Gga_FABP6	GVVYKRTSKKIA 128	
	. : : : :	

Figure 8.2. Maximum likelihood phylogenetic tree of *fabp6* built from the nucleotide coding sequence alignment.

The white diamond indicates the 3R duplication event, and the blue diamond indicates the 4R duplication event.



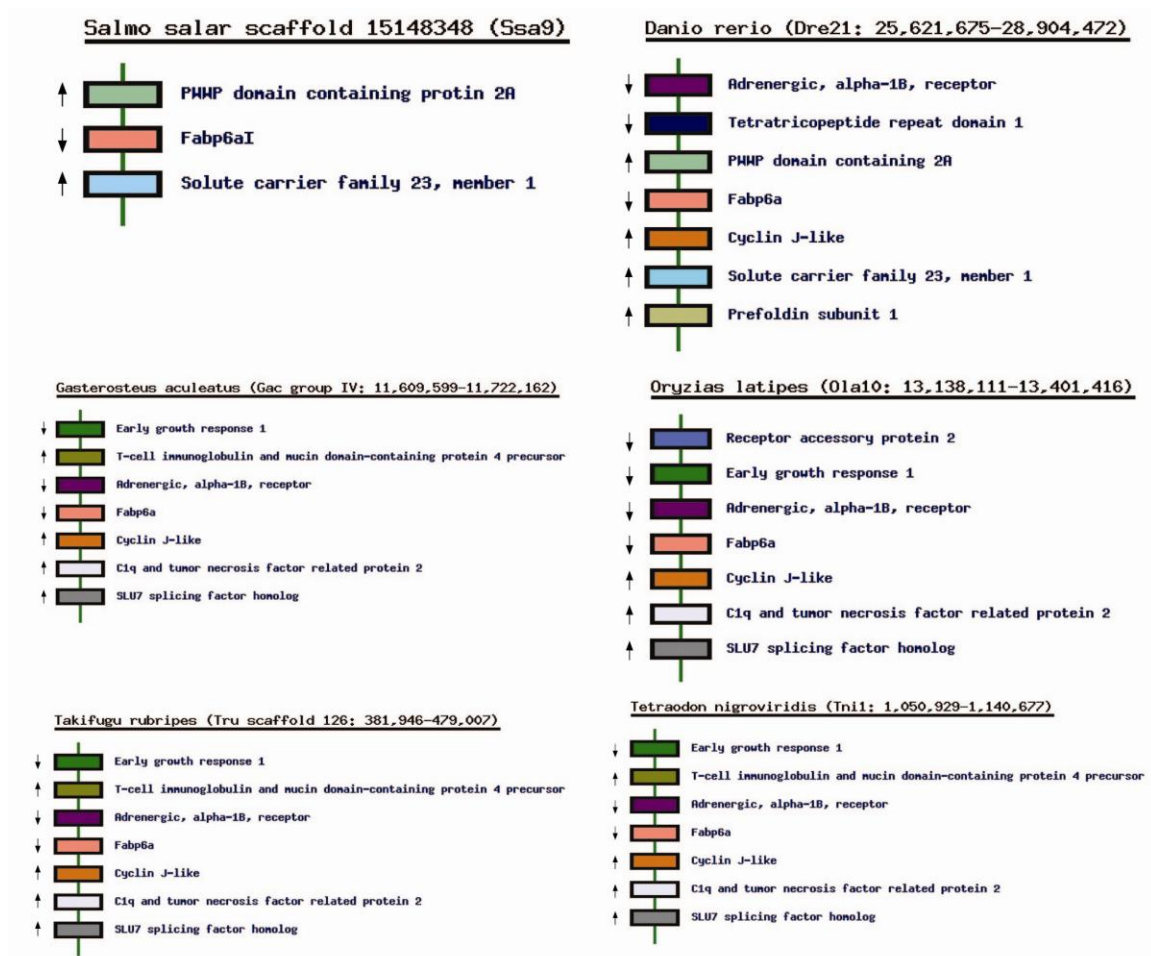
8.5 Conserved synteny of *fabp6a* loci in teleosts

Unfortunately and possibly due to the gaps within or the length of the consensus sequence of the CIGENE scaffolds 15259024 and 15123122, there were no other flanking genes annotated except for the annotation of *fabp6aII* and *fabp6b*, respectively. Hence, I was only able to compare the genomic organization of the Atlantic salmon *fabp6aI* locus with the *fabp6a* loci in zebrafish, stickleback, medaka, fugu and tetraodon. When I compared the Atlantic salmon *fabp6aI* locus with the *fabp6a* loci in the five sequenced teleost species, I could not predict an ancient conserved synteny block containing *fabp6a* in the common ancestor of teleosts (Figure 8.3). Instead, I was able to predict two *fabp6a* conserved synteny blocks with one block for the lineage leading towards zebrafish and Atlantic salmon and the other block for the lineage leading towards stickleback, medaka, fugu and tetraodon. I found six flanking genes in a conserved synteny block containing *fabp6a* as they are conserved in terms of position and orientation relative to the *fabp6a* gene in the lineage leading towards stickleback, medaka, fugu and tetraodon. These genes are: early growth response 1 (*egr1*), T-cell immunoglobulin and mucin domain-containing protein 4 precursor (*timd4*) (absent from medaka *fabp6a* locus), adrenergic, alpha-1B, receptor (*adra1b*), cyclin J-like (*ccnjl*), c1q and tumor necrosis factor related protein 2 (*c1qtnf2*) and SLU7 splicing factor homolog (*slu7*). Although *timd4* is absent between *egr1* and *adra1b* in the medaka genome in Ensembl release version 64, the coding sequence of stickleback *timd4* exons two, four and six were aligned with the medaka genomic sequence of the region between *egr1* and *adra1b*, with 72-78% identity. This suggests either that there is a *timd4* pseudogene or there is an unannotated *timd4* gene in this region of medaka genome. However, I

observed a similar yet different genomic organization at the *fabp6a* locus in the zebrafish genome. *Adra1b* is located at the 5' end of *fabp6a*, but there are two other genes, tetratricopeptide repeat domain 1 (*ttc1*) and PWWP domain containing 2A (*pwwp2a*), inserted between *adra1b* and *fabp6a*. The gene encoding for cyclin J-like (*ccnjl*) still resides beside and at the 3' end of zebrafish *fabp6a* locus as it is in other four sequenced teleost species *fabp6a* loci. However, solute carrier family 23, member 1 (*slc23a1*) and prefoldin subunit 1 (*pfdn1*) are found further downstream of the 3' end of zebrafish *fabp6*. Similar to the genomic organization of zebrafish *fabp6a* loci, *pwwp2a* and *slc23a1* (solute carrier family 23, member 1) are on the 5' and 3' ends of Atlantic salmon *fabp6aI*, respectively. Moreover, *ccnjl* is not annotated on either ends of *fabp6aI* despite the fact that zebrafish *ccnjl* exons two, three, four and five aligned to the Atlantic salmon genomic sequence between *fabp6aI* and *slc23a1* with 68-85% identity. Again, this suggests either that there is a *ccnjl* pseudogene or there is an unannotated *timd4* gene in this region of medaka genome due to the gaps in the CIGENE scaffold. Overall, the comparative genomic analysis of *fabp6a* loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon suggests that a major genomic rearrangement happened around the *fabp6a* locus in one of the two lineages; that is the lineage leading to stickleback, medaka, fugu and tetraodon or the lineage leading to zebrafish and Atlantic salmon.

Figure 8.3. Comparative genomic analysis of *fabp6a* loci in Atlantic salmon, zebrafish, stickleback, medaka, fugu and tetraodon.

The arrows indicate the gene orientations.

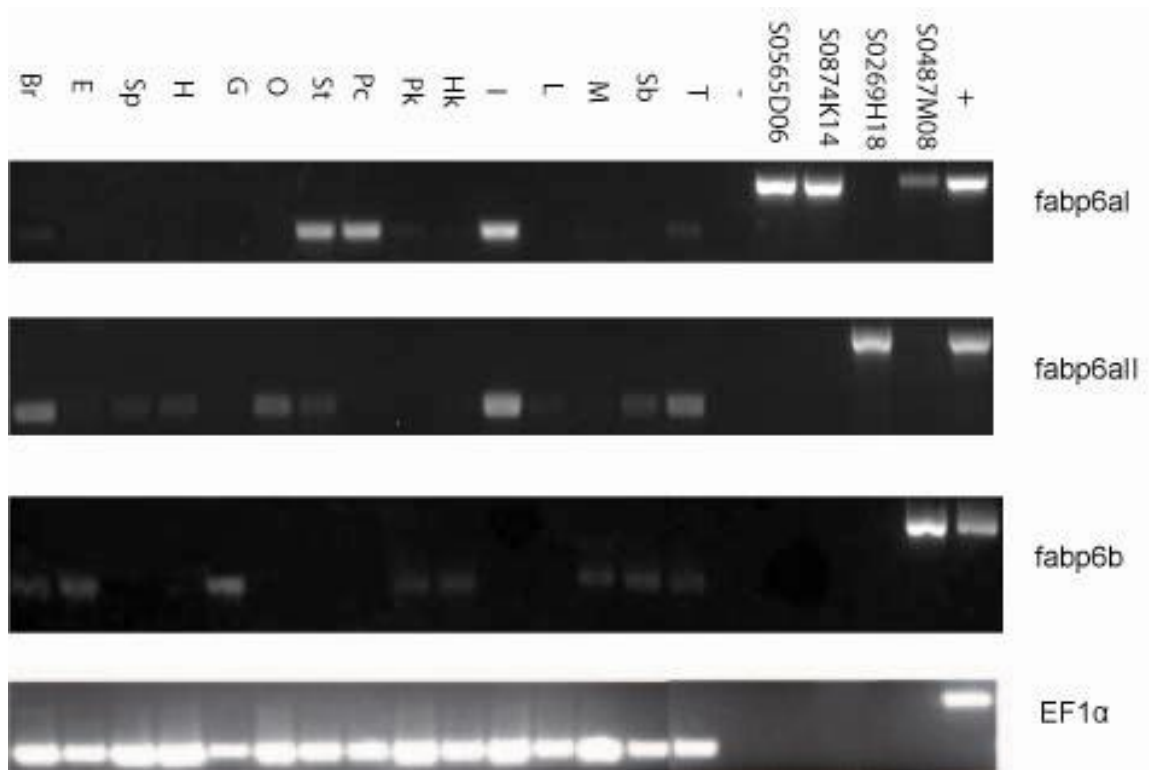


8.6 Expression profiling of Atlantic salmon *fabp6s*

I used RT-PCR with *fabp6* gene-specific primers on Atlantic salmon adult tissues to examine the expression patterns of Atlantic salmon *fabp6aI*, *fabp6aII* and *fabp6b*. All three genes are expressed in the brain (Figure 8.4). In addition, *fabp6b* is expressed in eye, gill, both posterior and head kidney, muscle, swim bladder and testis. The *fabp6aI* transcript was found in six out of fifteen tissues that were examined, whereas *fabp6aII* is expressed in brain, eye, spleen, heart, ovary, stomach, intestine, liver, muscle and swim bladder. No *fabp6a* transcript was found in gill and anterior kidney. Atlantic salmon *fabp6aI* and *fabp6aII* are differentially expressed in eye, spleen, heart, ovary, pyloric caecum, posterior kidney, liver, muscle, swim bladder and testis. Nonetheless, all Atlantic salmon *fabp6s* appear to have broader expression patterns compared to the zebrafish *fabp6*, whose transcript was detected only in liver, heart, intestine, ovary and kidney (Alves-Costa et al., 2008). Since the comparative genomic analysis suggests that zebrafish *fabp6* belongs to the *fabp6a* type and assuming that the zebrafish *fabp6* is representative of the ancestral *fabp6a* gene in the common ancestor of salmonids, it suggests that Atlantic salmon *fabp6aI* and *fabp6aII* are undergoing sub-functionalization in terms of their differential expression in liver, heart, ovary and kidney.

Figure 8.4. Expression patterns of *fabp6aI*, *fabp6aII* and *fabp6b* in Atlantic salmon adult tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill, (G), ovary (O), stomach (St), pyloric caecum (Pc), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M), swim bladder (Sb), testis (T), negative control (-) and Atlantic salmon genomic DNA (+). BAC clones S0565D062 and S0874K14 are the representative BACs of Atlantic salmon *fabp6aI*, whereas BAC S0269H18 and S0487M08 are representative BAC clones for *fabp6aII* and *fabp6b* (Table 3.3). Differential expression patterns of Atlantic salmon *fabp6aI* and *fabp6aII* were eye, spleen, heart, ovary, pyloric caecum, posterior kidney, liver, muscle, swim bladder and testis.



8.7 Conclusion

The combination of phylogenetic tree and comparative genomic analysis suggests that the 3R WGD at the base of the teleost radiation gave rise to *fabp6a* and *fabp6b*. Based on the nucleotide coding sequence alignment, the phylogenetic tree suggests that zebrafish *fabp6* and Atlantic salmon *fabp6aI* and *fabp6aII* are more closely-related to the other teleost (stickleback, medaka, fugu and tetraodon) *fabp6bs* and Atlantic salmon *fabp6b* is more-closely related to the other teleost *fabp6as*. However, this suggestion is contradicted with the conclusion from the comparative genomic analysis; that is, the Atlantic salmon *fabp6aI* and zebrafish *fabp6a* genomic organization contain flanking genes that are present in flanking regions of other teleost *fabp6* genomic loci. The most parsimonious explanation for this phenomenon is that genomic rearrangements at the *fabp6a* loci occurred after the divergence of the common ancestor of stickleback, medaka, fugu and tetraodon from the common ancestor of zebrafish and Atlantic salmon. Moreover, I postulated that a particular selection pressure selected one of the *fabp6* gene duplicates after the 3R genome duplication event to either have the KGR signature (from amino acid residue 31 to 33 according to the amino acid numbering in Atlantic salmon FABP6as) or not. This selection force applied to *fabp6bs* in the lineage leading towards stickleback, medaka, fugu and tetraodon, while it acted on *fabp6as* in the common ancestor of zebrafish and Atlantic salmon. Again, it suggests that these gene modifications are two independent events that occurred after the divergence of the common ancestor of stickleback, medaka, fugu and tetraodon from the common ancestor of zebrafish and Atlantic salmon. Overall, this contradiction between the sequence alignment and phylogenetic tree and the comparative genomic analysis is interesting

because it reveals that *fabp6* genomic rearrangements and gene duplicate modifications might reflect on the natural selection forces acting on these teleosts. Moreover, comparison of the expression patterns of Atlantic salmon *fabp6aI* and *fabp6aII* with that of zebrafish *fabp6* suggests that Atlantic salmon *fabp6a* gene duplicates are undergoing sub-functionalization in terms of their differential expression in liver, heart, ovary and kidney.

9: Evolution and expression profiling of the Atlantic salmon *fabp10* sub-family

9.1 Introduction

FABP10 is the liver basic type fatty acid binding protein because it was found in liver, but it has a basic isoelectric point that is different from FABP1. To date, it has only been found only non-mammalian species (Nichesola et al., 2004; Nolan, 2005). Its coding sequence is more closely-related to *fabp6* than it is to *fabp1*. Hence, it has been proposed to name this as bile liver basic fatty acid binding protein instead, due to FABP10 having a higher affinity towards bile salts than fatty acids (Di Pietro et al., 1997; Di Pietro et al., 2001), but its function with bile salts in the cell is unknown.

9.2 Identification of salmonids, northern pike and rainbow smelt *fabp10* transcript sequences

Using the zebrafish *fabp10* nucleotide coding sequences as the queries (Ensembl Accession: ENSDART00000056095 and GenBank Accession: XM_001335293.2), I searched and found three unique *fabp10* full-length EST sequences in rainbow trout and one *fabp10* EST sequence in sockeye salmon and rainbow smelt (Table 9.1). The sequence identity between rainbow trout *fabp10aI* and *fabp10aII* EST sequences is 94%. Either rainbow trout *fabp10aI* or *fabp10aII* EST sequence is 69% sequence identical to *fabp10b* EST sequences. Hence, it suggests that the rainbow trout *fabp10aI* and *fabp10aII* are related to one another by the 4R WGD, and both are related to *fabp10b* genes by the 3R WGD.

Table 9.1. Rainbow trout, sockeye salmon and rainbow smelt full-length *fabp10* EST sequences and their GenBank accession numbers as of February 20th, 2012.

	Rainbow trout (Omy)	Sockeye salmon (One)	Rainbow smelt (Omo)
<i>fabp10aI</i>	CA359513.1		
<i>fabp10aII</i>	CX144766.1		
<i>fabp10b</i>	BX871583.3		
<i>fabp10a</i>		BX886324.3	CX351101.1

9.3 Gene annotation of fugu *fabp10* gene

As mentioned in Section 3.2 and Table 3.2, I annotated the fugu *fabp10a* because no *fabp10* had been annotated in the fugu genome in Ensembl version 64. Using the zebrafish *fabp10* coding sequences as queries and I BLASTn-searched against the EST database in NCBI to seek for a fugu *fabp10* EST sequence. A fugu *fabp10* EST sequence (GenBank Accession: CA330310.1) was found and the nucleotide coding and protein sequence was identified by the ExPasy translate tool. I then used the fugu *fabp10* nucleotide coding sequence and BLASTn-searched against the medaka genome using Ensembl BLAST/BLAT in version 64. Indeed, the coding sequence aligned with the genomic sequence on scaffold 35 between the myomesin family, member 3 gene (*myom3*) (ENSTRUG00000007897) and proline-rich nuclear receptor coactivator 2 gene (*pnrc2*) (ENSTRUG00000008914). Hence, there was a previously un-annotated *fabp10* gene located in the genomic region between *myom3* and *pnrc2* in fugu scaffold 35.

9.4 Evolution of Atlantic salmon *fabp10* sub-family

By a clustalW alignment, I compared these FABP10 protein sequences with the inferred amino acid sequences of the unique full-length *fabp10* transcripts in Atlantic salmon and rainbow trout EST database (Figure 9.1). Interestingly, these vertebrate FABP10s vary in length, ranging from 126 to 128 amino acids long. There are 27 invariant amino acid residues in these vertebrate FABP10 proteins, suggesting that the FABP10 sub-family is relatively less conserved in terms of protein size and the number of conserved amino acid residues compared to the FABP7 sub-family. A phylogenetic tree was built based on the alignment of the coding sequences of these vertebrate *fabp10* proteins (Figure 9.2). Zebrafish, stickleback and medaka all possess two *fabp10*s, which

Venkatachalam et al. named *fabp10a* and *fabp10b* (Venkatachalam et al., 2009). This suggests that there was at least a *fabp10* gene duplication or possibly the 3R genome duplication that occurred in common ancestor of teleosts. Nevertheless, fugu and tetraodon only contain a single *fabp10* in their genomes, which belongs to the *fabp10a* type group, suggesting that there was a *fabp10b* gene loss at the common ancestor of fugu and tetraodon. Moreover, there are three *fabp10* unique full-length transcripts found in Atlantic salmon and rainbow trout EST databases, two of which are grouped into the *fabp10a* group. The remaining *fabp10* in Atlantic salmon and rainbow trout belong to *fabp10b*. The presence of three *fabp10s* (*fabp10aI*, *fabp10aII* and *fabp10b*) in Atlantic salmon and rainbow trout could be the result of another *fabp10a* gene duplication or possibly the 4R genome duplication that happened in the common ancestor of salmonids followed by a *fabp10b* gene loss in the common ancestor of Atlantic salmon and rainbow trout.

Figure 9.1. Alignment of the FABP10 amino acid sequences from Atlantic salmon, rainbow trout, other teleosts and chicken.

The species abbreviations are the same as in Figures 4.1 and 5.2. Protein size varies and only 27 amino acid residues are conserved in the FABP10s from different vertebrate species.

```

Ssa_FABP10aI      --MAFSGTWQVYAQENYEEFLRAISLPEDEVIKLAKDIKPVTEIQQNGNDFVITSKTPGKS 58
Omy_FABP10aI      --MAFSGTWQVYAQENYEEFLRAISLPEDEVIKLAKDIKPVTEIQQNGNDFVITSKTPGKS 58
Ssa_FABP10aII     --MAFSGTWQVYAQENYEEFLRAISLPEDEVIKLAKDIKPVTEIQQNGNDFVITSKTPGKS 58
Omy_FABP10aII     --MAFSGTWQVYAQENYEEFLRAISLPEDEVIKLAKDIKPITEIQQNGNDFVITSKTPGKS 58
Dre_FABP10a       --MAFSGTWQVYAQENYEEFLRAISLPEDEVIKLAKDVKPVTEIQQNGSDFTITSKTPGKT 58
Gac_FABP10a       --MDFNGTWQVYsqENYEEFLRWMLPEDEVIKMAKGIKPIITEIKQKGNDFVITSKTPGKT 58
Ola_FABP10a       --MDFNGTWQVYsqENYEEFLRALELSEDIKLAQDVKPVTEIKQKGNDFVITSKTPGRT 58
Tru_FABP10a       --MDFNGTWQVYAQENYEEFLKAMELPEDEVIKMAKDIKPVTEIQQNGNDFVITSKTPGKT 58
Tni_FABP10a       --MDFNGTWQVYAQENYEEFLKAMELPEDEVIKMAKDIKPIITEIKQKGNDFVITSKTPGKT 58
Ssa_FABP10b       --MAFNGKWKIHSQENHEEFLKAMAVPEMLIKMIKGVKPTTII EQKGGDDFTIRVETPLRT 58
Omy_FABP10b       --MAFNGKWKIHSQENHEEFLKAMAVPEMLIKMIKGVKPTTII EQKGGDDFTIRVETPLRT 58
Dre_FABP10b       MAVD FNGSWKLYEQENAEFLRAISLPEDEVIKMAKGIKPIITEIKQKGNDFVITSKTPGRT 60
Gac_FABP10b       --MDFSGTWKVIYSEDNLEGLKVVVGAPEMVIKMRKNIKPVVVI EQKNGTDFTYTTKPIYT 58
Ola_FABP10b       --MDFSGTWKVIYSEENLEAFLKVIIGAPEMIVKMRKDVKPVVVI EQKNGKDFVYTTKPMFT 58
Gga_FABP10        --MAFSGTWQVYAQENYEEFLKALALPEDDLKIMARDIKPIVTEIQKGGDDFVITSKTPRQT 58
      : * . * . * : : : * * * : : . * : : : : * . * . * * : * : * . :

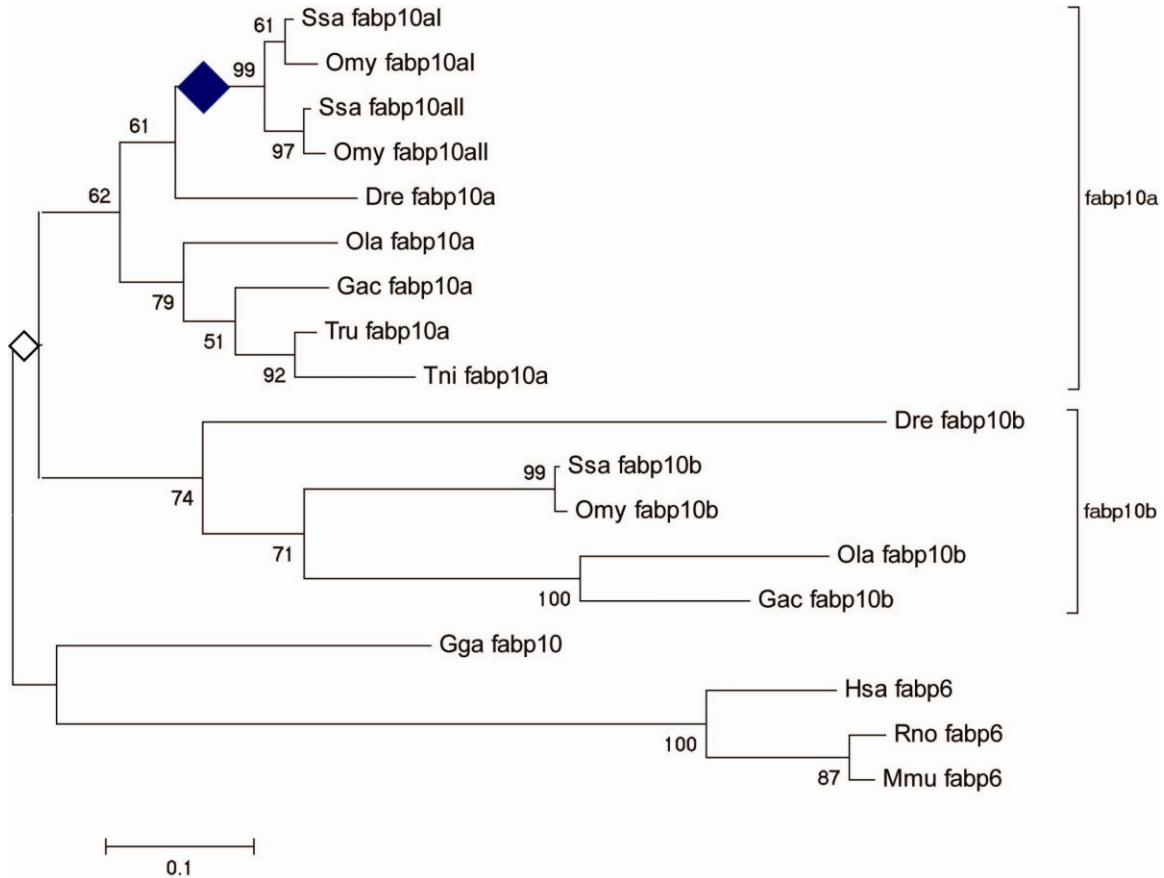
Ssa_FABP10aI      ITNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFTHIQELKGGEMVELTVGSTS 118
Omy_FABP10aI      ITNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFTHIQELKGGEMVELTVGSTS 118
Ssa_FABP10aII     VTNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Omy_FABP10aII     VTNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Dre_FABP10a       VTNSFTIGKEAEITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Gac_FABP10a       VTNTFTVIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Ola_FABP10a       VTNSFTIGKEAEISTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Tru_FABP10a       VTNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Tni_FABP10a       LTNSFTIGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
Ssa_FABP10b       VTNSFTIGKEAEITAM DGRKFKCTARLEDGKLCIDTEKFSHIREIQENMVEITTAAST 118
Omy_FABP10b       VTNSFTIGKEAEITAM DGRKFKCTARLEDGKLCIDTEKFSHIREIQENMVEITTAAST 118
Dre_FABP10b       NTNTFRIGTESEFTT LDGQKINATARLIDGKIVIESEKFTHVRELRDGMLETLTAGHVT 120
Gac_FABP10b       KVHSFSLGKESEM TADGRRFKCTVREENGKLI AET EKFTSVREIQGDDMVEITTAGPET 118
Ola_FABP10b       RVVYFTLGQETEIP SVDGRKVKCII REENGKLI GADKFTSVREIRGDEMTEITTSGSVT 118
Gga_FABP10        VTNSFTLGKEADITTM DGKCLKCTVRLGGKLMCNTDKFSSVQELKGGEMVELTVGSTS 118
      . : * : * * : : : : * * : : . * : : : : * : * : : * * * : * :

Ssa_FABP10aI      LIRRSKRM- 126
Omy_FABP10aI      LIRRSKRL- 126
Ssa_FABP10aII     LIRRSKRL- 126
Omy_FABP10aII     LIRRSKRL- 126
Dre_FABP10a       MIRKSKKI- 126
Gac_FABP10a       LIRRSKRM- 126
Ola_FABP10a       LIRRSKRM- 126
Tru_FABP10a       FVRKSKRMN 127
Tru_FABP10a       FIRKSKRMN 127
Ssa_FABP10b       LTRISKRV- 126
Omy_FABP10b       LTRISKRV- 126
Dre_FABP10b       FIRSRRI- 128
Gac_FABP10b       FISRSRV- 126
Ola_FABP10b       FISSKRA- 126
Gga_FABP10        LIRRSRV- 126
      : * : :

```


Figure 9.2. Maximum likelihood phylogenetic tree of *fabp10* built from the nucleotide coding sequence alignment.

The white diamond indicates the 3R duplication event, which gave rise to the *fabp10a* and *fabp10b* genes in teleosts. The blue diamond indicates the 4R duplication event which resulted the *fabp10aI*s and *fabp10aII*s in Atlantic salmon and rainbow trout. Only bootstrap values that are greater than 50 are shown.

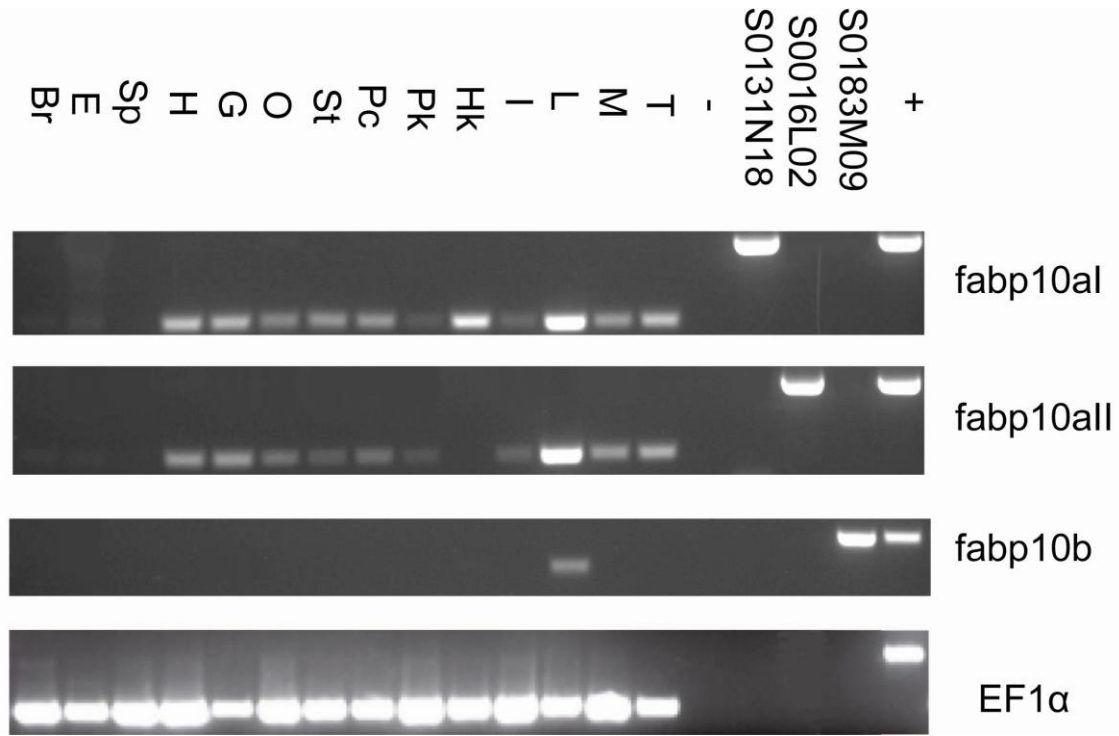


9.5 Expression profiling of Atlantic salmon *fabp10s*

I carried out RT-PCR on smolt tissues to examine the expression patterns of Atlantic salmon *fabp10s* to look for signatures of sub-functionalization (Figure 9.3). Atlantic salmon *fabp10b* transcript was only detected in liver, whereas *fabp10aI* was expressed in heart, gill, ovary, stomach, pyloric caeca, posterior and head kidney, intestine, liver, muscle and testis. Atlantic salmon *fabp10aII* has the same expression pattern as *fabp10aI*; however, its transcript was not detected in head kidney. The broad expression patterns of Atlantic salmon *fabp10a* gene duplicates and the limited expression pattern of the *fabp10b* gene are different from what was observed in zebrafish *fabp10a* and *fabp10b* (Sharma et al., 2006; Venkatachalam et al., 2009). That is, zebrafish *fabp10a* and *fabp10b* transcripts were found in intestine, liver and testis but zebrafish *fabp10b* is also expressed in muscle, brain, heart, eye, gill, ovary, skin, kidney and swim bladder. Assuming the expression patterns of zebrafish *fabp10a* and *fabp10b* are the ones in the common ancestor of salmonids, it appears there were regulatory element losses in the zebrafish *fabp10a* and Atlantic salmon *fabp10b*.

Figure 9.3. Expression patterns of *fabp10aI*, *fabp10aII* and *fabp10b* in Atlantic salmon smolt tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill (G), ovary (O), stomach (St), pyloric caecum (Pc), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M), testis (T), negative control (-) and Atlantic salmon genomic DNA (+). BAC clones S0131N18, S0016L02 and S0183M09 are the representative BACs of Atlantic salmon *fabp10aI*, *fabp10aII* and *fabp10b*, respectively (Table 3.3). Differential expression patterns of Atlantic salmon *fabp10aI* and *fabp10aII* was only observed in head kidney.



9.6 Conclusion

I identified three *fabp10* genes, *fabp10aI*, *fabp10aII* and *fabp10b*, in Atlantic salmon. The phylogenetic analysis suggests that the 3R WGD gave rise to *fabp10a* and *fabp10b* in zebrafish, stickleback and medaka. However, a *fabp10b* gene loss occurred in the common ancestor of the pufferfish, tetraodon and fugu. The phylogenetic tree also suggests that the 4R WGD gave rise to *fabp10aI* and *fabp10aII*, *fabp10bI* and *fabp10bII* in the common ancestor of salmonids with a subsequent gene loss of one of the *fabp10b* gene duplicates in the common ancestor of Atlantic salmon and rainbow trout, such that there are three *fabp10* genes in these two fish lineages. Differential expression of Atlantic salmon *fabp10aI* and *fabp10aII* was observed in anterior kidney. Compared to the expression patterns of zebrafish *fabp10a* and *fabp10b*, Atlantic salmon *fabp10aI* and *fabp10aII* are expressed more broadly compared to the expression pattern of zebrafish *fabp10a*, whereas Atlantic salmon *fabp10b* transcript was found only in liver in contrast to the broad expression pattern seen in zebrafish *fabp10b*. Assuming the expression patterns of zebrafish *fabp10a* and *fabp10b* are the expression patterns of the ancestral *fabp10a* and *fabp10b* in the common ancestor of salmonids before the 4R WGD, it suggests that there were regulatory element losses in the zebrafish *fabp10a* and Atlantic salmon *fabp10b*.

10: Evolution and Expression profiling of Atlantic salmon *fabp11* sub-family

10.1 Introduction

Karanth et al. proposed that *fabp4*, *fabp5*, *fabp8* and *fabp9* in tetrapods are related to *fabp11* in fish by a *fabp4/5/8/9/11/12* ancestral gene in the common ancestor of jawed vertebrates (Karanth et al., 2008). *Fabp11* is only found in fish, whereas *fabp4*, *fabp5*, *fabp8*, *fabp9* are found in species in the tetrapod lineage. In the tetrapod lineage, *fabp4*, *fabp5*, *fabp8* and *fabp9* are found in a cluster that is known to originate from a series of tandem duplications of this *fabp4/5/8/9/11* ancestral gene (Karanth et al., 2008; Liu et al., 2008). Recently, Liu et al. characterized another *fabp*, *fabp12*, in the *fabp4/5/8/9* cluster in the pig genome (Liu et al., 2008) and this gene was subsequently annotated in the rat, mouse and human genomes.

FABP4, FABP5, FABP8 and FABP9 are the adipose-type, skin-type, schwann cell-type and testis-type FABPs, respectively. Due to its association with many cellular processes and human diseases, the function of FABP4 has been heavily studied. *Fabp4* expression is correlated with the availability of free fatty acids from endogenous or exogenous sources, as well as PPAR- γ agonists and insulin (Amri et al., 1991; Haunerland and Spener, 2004; Makowski and Hotamisligil, 2005; Makowski et al., 2005). FABP4 is regulated during differentiation of adipocytes and shuttles fatty acids between the nucleus and cytoplasm continuously (Ayers et al., 2007). FABP4 is also a regulator of the PPAR- γ -liver X receptor- α (LXR- α)-ATP-binding cassette A1 (ABCA1)

pathway, is found in macrophages upon their differentiation from monocytes (Fu et al., 2000; Fu et al., 2002; Kazemi et al., 2005; Makowski et al., 2001; Pelton et al., 1999) and dendritic cells (Rolph et al., 2006), and contributes to foam-cell formation in macrophages (Makowski et al., 2005). The *fabp4* knock-out adipocytes have reduced efficiency of lipolysis (Coe et al., 1999; Scheja et al., 1999; Shen et al., 1999). In FABP4-deficient macrophages, several inflammatory signalling responses were suppressed, such as production of cytokines and production and function of pro-inflammatory enzymes (Makowski et al., 2005). Nevertheless, FABP4-deficient mice appear to increase their protection against asthma (Shum et al., 2006), and FABP4 concentration may be associated with obesity, type 2 diabetes and cardiovascular diseases (Ohlsson et al., 2005; Tso et al., 2007; Xu et al., 2006; Yeung et al., 2007), suggesting its potential functions in different cell types.

As mentioned above, *fabp11* is the fish pro-ortholog of *fabp4*. Indeed, the *fabp11* transcript was enhanced during preadipocyte differentiation (Huang et al., 2010) and its level is affected by exposure to different fatty acids *in vitro*. It is expressed in various Atlantic salmon tissues with being most abundant in the visceral fat tissue, which is consistent with FABP4 high abundance in adipocyte. In the rat, *fabp12* is expressed in ganglion and the inner nuclear layer of the retina and testis. However, *fabp12* has a different expression pattern in the rat testis compared with the expression of *fabp9*, suggesting that *fabp9* and *fabp12* have different roles in spermatogenesis (Liu et al., 2008).

10.2 Identification of salmonid, northern pike and rainbow smelt *fabp11* transcript sequences

Using the zebrafish *fabp11* nucleotide coding sequences as the queries (Ensembl Accession: ENSDART00000021798 and ENSDART00000012718), I searched and found three rainbow trout unique full-length EST sequences and one unique *fabp11* EST sequence in sockeye salmon, brook trout, lake whitefish and northern pike. The sequence identity between rainbow trout *fabp11bI* and *fabp11bII* EST sequences is 97%. Rainbow trout *fabp11a* and *fabp11bI* EST sequences are 71% identical whereas *fabp11a* and *fabp11bII* coding sequences are 70%. Hence, it suggests that rainbow trout *fabp11bI* and *fabp11bII* are related to one another by the 4R WGD, and both are related to *fabp11a* by the 3R WGD.

Table 10.1. Rainbow trout, sockeye salmon and rainbow smelt full-length *fabp11* EST sequences and their GenBank accession numbers as of February 20th, 2012.

* denotes that particular EST sequence does not have a GenBank accession number corresponding to it.

	Rainbow trout (Omy)	Sockeye salmon (One)	Brook trout (Sfo)	Lake Whitefish (Ccl)	Northern pike (Elu)
<i>fabp11a</i>	CF752694.1	CD510724.1		EV366236.1	GH251914.1
<i>fabp11b</i>			EV391512.1		
<i>fabp11aI</i>					
<i>fabp11aII</i>					
<i>fabp11bI</i>	*				
<i>fabp11bII</i>	BX298288.3				

10.3 Gene annotation of tetraodon *fabp11b*

Although there was no annotated tetraodon *fabp11b* in Ensembl version 64, I expected that there would be two copies of *fabp11* in this species because the other four sequenced teleosts (zebrafish, stickleback, medaka and fugu) contain two copies of *fabp11* genes. I used the transcript of fugu *fabp11b* (ENSTRUT00000034626) as the query and BLASTn-searched for a tetraodon *fabp11b* EST sequence in NCBI (<http://www.ncbi.nlm.nih.gov/guide/>). The tetraodon *fabp11b* EST sequence was found (GenBank Accession: CR699906.2) and the nucleotide coding sequence and the protein sequence of this gene was obtained using the ExPasy translate tool. I then used this tetraodon *fabp11b* coding sequence as the query and BLASTn-searched against the tetraodon genome using Ensembl BLAST/BLAT in version 64. The location of the tetraodon *fabp11b* was identified to be the genomic sequence of an unannotated gene (ENSTNIG00000018400) on chromosome 8.

10.4 Evolution of Atlantic salmon *fabp11* sub-family

To understand how the salmonid *fabp11s* are related to their pro-orthologs in tetrapods, I compared the protein sequences of human, rat, mouse, chicken FABP4 and zebrafish, medaka, stickleback, fugu and tetraodon FABP11s with the inferred amino acid sequences of their orthologs in Atlantic salmon and rainbow trout by a clustalW alignment (Figure 10.1). FABP11s in teleosts range in size from 133 to 135 amino acid and there are only 34 invariant amino acid residues in these vertebrate proteins. Hence, it appears that FABP11 sub-family is relatively less conserved in terms of protein size and the number of conserved amino acid residues compared to the FABP7 sub-family. Based on the alignment of the coding sequences coding for the teleosts FABP11 proteins and

the tetrapod FABP4 proteins, a phylogenetic tree was built to reveal the relationship of *fabp11* and *fabp4* sub-families (Figure 10.2). Zebrafish, stickleback, medaka, fugu and tetraodon all possess two *fabp11s* in their genomes, which fall into the *fabp11a* and *fabp11b* groups following nomenclature of the two zebrafish *fabp11s* (Karanth et al., 2008). This suggests that the 3R WGD that occurred in the common ancestor of teleosts resulted in the *fabp11a* and *fabp11b* genes seen in extant teleosts today, however it is not supported by the phylogenetic tree and this may be due to the fact that the tree did not resolve well. Based on the phylogenetic tree, two of the three Atlantic salmon *fabp11* ESTs and one rainbow trout *fabp11* EST sequence belong to the *fabp11a* group; whereas the other two rainbow trout *fabp11* EST sequences and the remaining Atlantic salmon *fabp11* transcript are grouped in the *fabp11b* clade. It is more parsimonious to state that there was a *fabp11a* gene duplication in the Atlantic salmon lineage and a *fabp11b* gene duplication rainbow trout lineage. However, based on the positions of Atlantic salmon *fabp11aI* and *fabp11aII* and rainbow trout *fabp11bI* and *fabp11bII*, it suggests that the two *fabp11a* paralogs in Atlantic salmon and the two *fabp11b* paralogs in rainbow trout are related to one another by the 4R WGD events in the common ancestor of salmonids. Hence, it also suggests that there might be a subsequent *fabp11b* gene loss in the Atlantic salmon and a *fabp11a* gene loss in rainbow trout after the WGD or they are just not annotated in the EST databases.

Figure 10.1. Alignment of the FABP11 protein sequences from Atlantic salmon, rainbow trout, zebrafish, stickleback, medaka, fugu and tetraodon with the FABP4 amino acid sequences from chicken, mouse, rat and human.

The species abbreviations are the same as in Figures 4.1 and 5.2.

```

Sea_FABP11aI      --MVDKRFVGTWKMISSEN FDDYMKALGVGFATR QVGNRTKPSLIINMDDQGMICVKSQST 58
Sea_FABP11aII     --MVDKRFVGTWKMTSSDN FDDYMKALGVGFATR QVGNRTKPSLIINMDDQGMICVKSQST 58
Omy_FABP11a       --MVDKRFVGTWKMISSEN FDDYMKALGVGFATR QVGNRTKPSLIINMDDQGMICVKSQST 58
Dre_FABP11a       --MVDKRFVGTWKMTSSDN FDEYMKALGVGFATR QVGNRTKPNLVVVCVDEQGLICMKSQST 58
Gac_FABP11a       --MVEKRFVGTWKMIASEN FDDYMKALGVGFATR QVGNRTKPNLIVTVVDEDDGVVCLKSQST 58
Ola_FABP11a       --MAEKRFVGTWKMISSEN FDDYMKALGVGFATR QVGNRTKPNLVVSI EDQGTISMKSQST 58
Tru_FABP11a       --MVEKRFVGTWKMISSEN FDDYMKALGVGFATR QVGNRTKPNLVVTVDDQGTIVCMKQST 58
Tni_FABP11a       --MVEKRFVGTWKMISSEN FDDYMKALGVGFATR QVGNRTKPNLVVTVDDQGTIVCMKQST 58
Sea_FABP11b       --MVEAFVGTWKMTSSDN FDEYMKALGVGFATR QMGMMAKPNLQPSI DD-GVISMKSQST 57
Omy_FABP11bI      --MVEAFVGTWKMTSSDN FDEYMKALGVGFATR QMGMMAKPNLVSSI DD-GVISMKSQST 57
Omy_FABP11bII     --MVEAFVGTWKMTSSDN FDEYMKALGVGFATR QMGMMAKPNLLPSI DD-GVISMKSQST 57
Dre_FABP11b       --MVEQFVGTWKMTSSDN FDEYMKAVGAFATR QMGNLAKPSLLIADVDEQGVITMKAVTT 58
Gac_FABP11b       --MVDQFVGTWTLIASEN FDEYMKALGVGLATR QMGNLAKPNLVIGVDEAGLISMKSETT 58
Ola_FABP11b       --MVEQFAGTWTLASSEN FDEYMKAVGAGFATR QMGNITKPNLVIVQVADGLISMKAEST 58
Tru_FABP11b       MDMVEQFVGTWTLVSSDN FDDYMKATIGVNFATR QMGMVVKPNLVVSMGDDGFTISMKSEST 60
Tni_FABP11b       --MVEQFVGTWTLASSEN FDEYMKATIGVNFATR QMGNLVKPSLVVSVADQGVITLKAEST 58
Gga_FABP4         --MCDQFVGTWKLSSSEN FEDYMKELGVGFATR KMGAVAKPNLTISING-DVIT IRSEST 57
Mmu_FABP4         --MCDAFVGTWKLVSSSEN FDDYMKELGVGFATR KVVAGMAKPNMII SVNG-DLVV IRSEST 57
Rno_FABP4         --MCDAFVGTWKLVSSSEN FDDYMKELGVGFATR KVVAGMAKPNLII SVNG-DLVV IRSEST 57
Hsa_FABP4         --MCDAFVGTWKLVSSSEN FDDYMKELGVGFATR KVVAGMAKPNMII SVNG-DVIT IKSEST 57
  * : : * * . : : * * : * * * : * * : * * : * * : * * : * * : * *

```

```

Sea_FABP11aI      FKTVEVKFNLEPFEETT ADDRKMTVTFLENG KLVQKQKWDGKEST IEREVTGDGKLLIA 118
Sea_FABP11aII     FKTVEIKFKLNEPFEETT ADDRKMTVTFLENG NLVQKQKWDGKETT IEREVTGDGKLLIA 118
Omy_FABP11a       FKTVEIKFKLNEPFEETT ADDRKMTVTFLENG NLVQKQKWDGKETI IEREVTGDGKLLIA 118
Dre_FABP11a       FKTT EIKFKLNEPFEETT ADDRKMT TWTIENG KLVQKQTWDGKEST IEREVS-DGKLLIA 117
Gac_FABP11a       FKTT EIKFKLNVPEFEETT ADDRKMT TVVBLEGG KLVQKQSWDGKETN IEREIT-DGKLLIA 117
Ola_FABP11a       FKTT EVKFKLNEPFEETT ADDRKMT TVITLDNG KLVQKQSWDGKETS IEREIS-DGKLLIA 117
Tru_FABP11a       FKNT EIKFKLNEPFEEMT ADDRKTRTVVTFLENG KLVQKQKWDGKETC IEREIS-DGKLLIA 117
Tni_FABP11a       FKNT EIKFKLNEPFEETT ADDRKTRTVVTFLENG KLVQKQKWDGKETC IEREIS-DGKLLIA 117
Sea_FABP11b       FKTT EIKFKLNEPFEEMT ADDRKTKTLMTFENG KLVQKQTWDGKTTT LEREIQ-DGKLLIA 116
Omy_FABP11bI      SKTT EVKFKLNEKFEEMT ADDRKTKTLMTFENG KLVHKQTWDGKTTT LEREIQ-DGKLLIA 116
Omy_FABP11bII     FKTT EAKFKLNEPFEEMT ADDRKTKTLMTFENG KLVQKQTWDGKTTT LEREIQ-DGKLLIA 116
Dre_FABP11b       FKTL EIKFKLDEPFEETT ADDRKVTTMSLADG KLIQKQT WEGKTTI IEREIQ-DTHMIA 117
Gac_FABP11b       FKTT EFKFKLNEBCDETT ADGRITKTLVTLKDG KLVQQQTWDGKTTT LEREIQ-GGNMSA 117
Ola_FABP11b       FKTT EIKFRLGEPFEETT VDGRNVKSTPTLDNG KLVQKQKWDGKST LEREIQ-EGKLLIA 117
Tru_FABP11b       FKTT EIKFKLTPEFEETT ADGRNKTVITFENG KLVQKQKWDGKETT IEREIQ-DGKLLIA 119
Tni_FABP11b       FKTT EIKFKLNEPFEETT ADGRNKTVITFENG KLVQKQKWDGKETT IEREIQ-DGKLLIA 117
Gga_FABP4         FKNT EISFKLGEPEFEETT ADDRKTKNVTILDNG TLKQVQKWDGKETV IKRNV-DGNLLV 116
Mmu_FABP4         FKNT EISFKLGEPEFEETT ADDRKVKSIITLDGG ALVQVQKWDGKSTT IKRFRD-GDKLVV 116
Rno_FABP4         FKNT EISFKLGEPEFEETT PDRKVKSIITLDGG VLVHVQKWDGKSTT IKRFRD-GDKLVV 116
Hsa_FABP4         FKNT EISFKLGEPEFEETT ADDRKVKSTITLDGG VLVHVQKWDGKSTT IKRFRD-DDKLVV 116
  * . * * * : * * * * . . . : * * : * . * * * : * * : * *

```

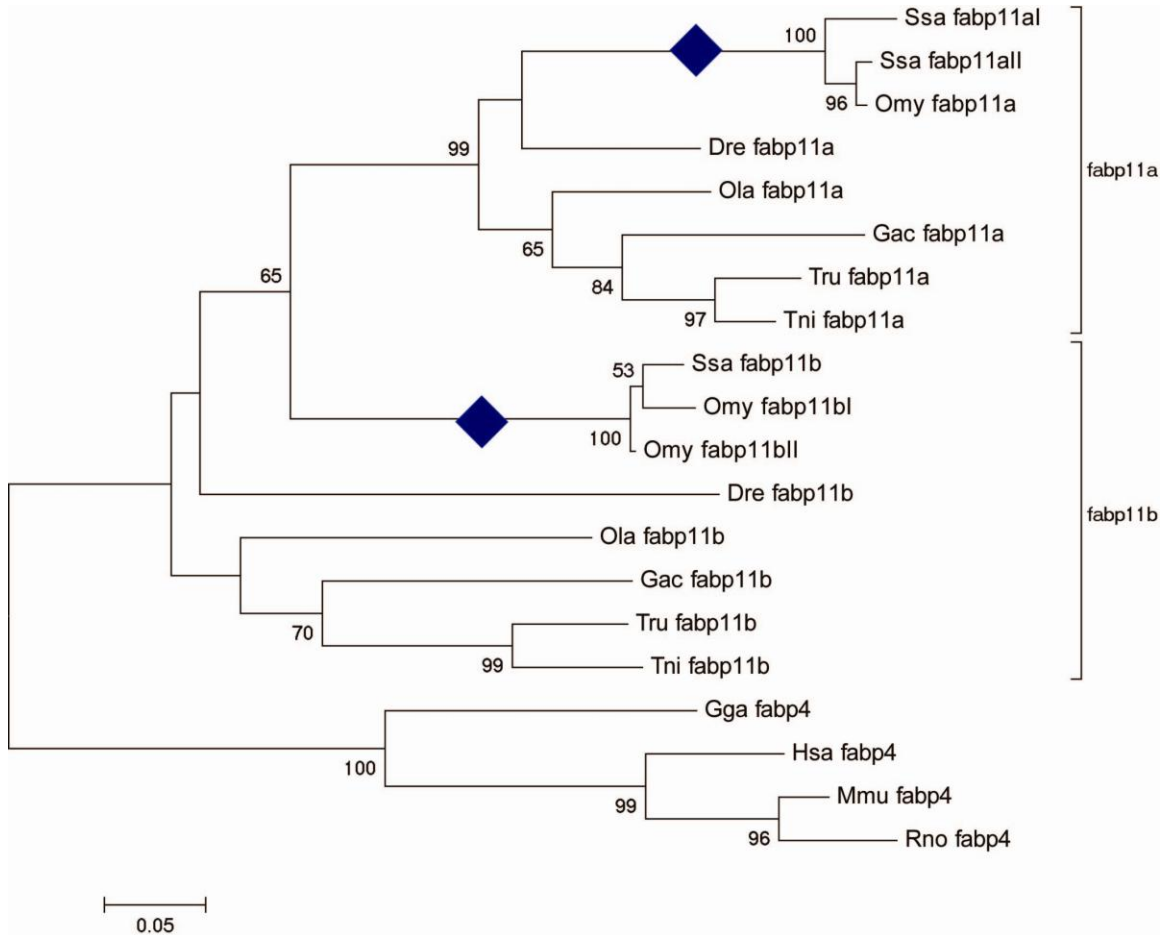
```

Sea_FABP11aI      KCMGDVIAVRYVKEV 135
Sea_FABP11aII     KCMGDVAVRYVKEA 135
Omy_FABP11a       KCMGDVAVRYVKEA 135
Dre_FABP11a       KCMGDVAVRYVKEA 134
Gac_FABP11a       KCMGDVIAVRYVKEA 134
Ola_FABP11a       KCMGDVAVRYVKEA 134
Tru_FABP11a       KCII GDVAVRYVKEA 134
Tni_FABP11a       KCMGDVAVRYVREA 134
Sea_FABP11b       KCMDDVVALRYEKEV 133
Omy_FABP11bI      KCMDDVVALRYEKEM 133
Omy_FABP11bII     KCMDDVVALRYEKEV 133
Dre_FABP11b       KCTMDDVAIRTYEKDA 134
Gac_FABP11b       KCMDDVAVRYVAKA 134
Ola_FABP11b       KCMDDVAVRYVEKAT 134
Tru_FABP11b       KCIA DDVVALRYERV- 133
Tni_FABP11b       KCIA DDVVALRYERV- 133
Gga_FABP4         ECTMDDVSTRVYERA- 132
Mmu_FABP4         ECTMKGVTSTRVYERA- 132
Rno_FABP4         ECTMKGVTSTRVYERA- 132
Hsa_FABP4         ECTMKGVTSTRVYERA- 132
  : * . * : * * :

```

Figure 10.2. Maximum likelihood phylogenetic tree of *fabp11* built from the nucleotide coding sequence alignment.

Tetrapod *fabp4s* were used as the outgroups. The blue diamonds indicate where the 4R WGD event The blue diamond indicates the 4R WGD event, which gave rise to the *fabp11aI* and *fabp11aII* in Atlantic salmon and *fabp11bI* and *fabp11bII* in rainbow trout. Only bootstrap values that are greater than 50 are shown.

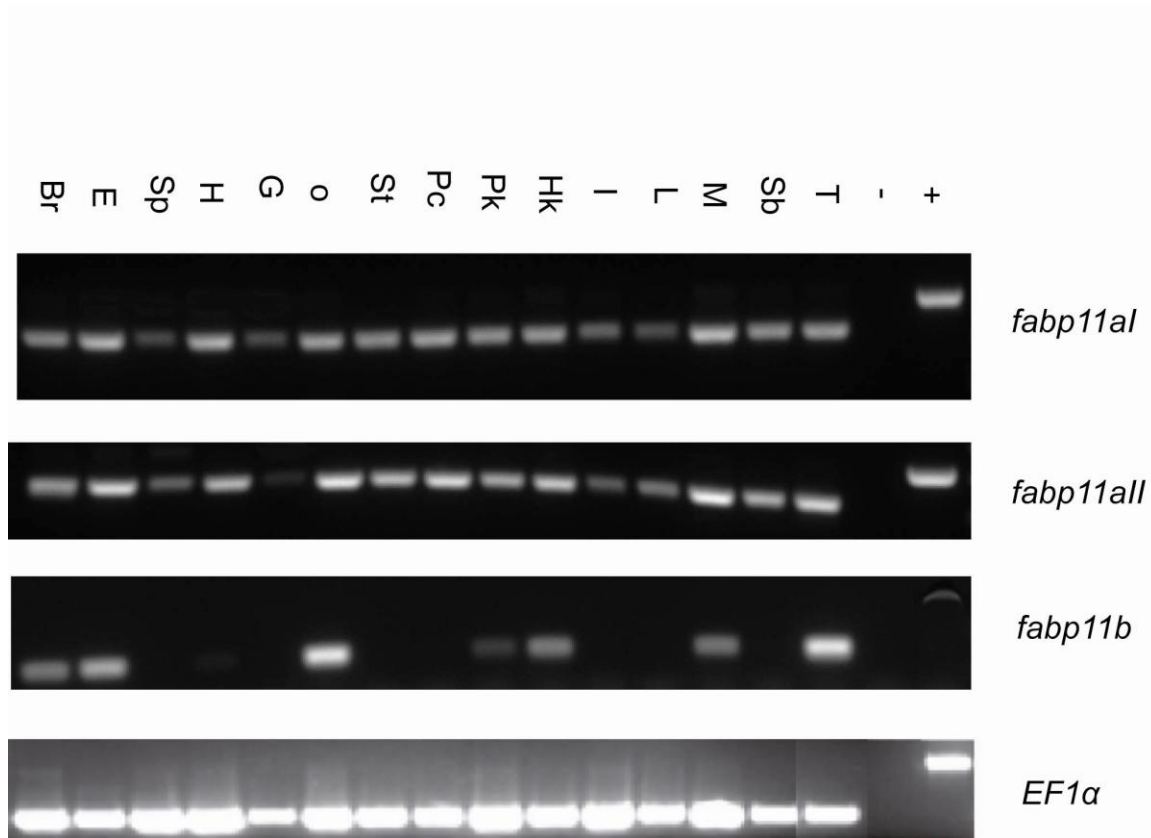


10.5 Expression profiling of Atlantic salmon *fabp11s*

RT-PCR was carried out as a way to look for signatures of sub-functionalization of Atlantic salmon *fabp11* genes in terms of tissue expression. Both *fabp11aI* and *fabp11aII* transcripts were detected in all the adult tissue I examined; that is, brain, eye, spleen, heart, gill, ovary, stomach, pyloric caecum, posterior and head kidney, intestine, liver, muscle, swim bladder and testis (Figure 10.3). This suggests that these genes are broadly expressed, similar to the broad expression pattern of zebrafish *fabp11a* (Liu et al., 2007). However, Atlantic salmon *fabp11b* is only expressed eight of the fifteen tissues. While both zebrafish *fabp11a* and *fabp11b* are expressed in brain, heart and ovary (Karanth et al., 2008; Liu et al., 2007), I observed the co-expression of all three Atlantic salmon *fabp11* gene duplicates in these tissues as well. Overall, there was no evidence of sub-functionalization and it suggests that Atlantic salmon *fabp11aI* and *fabp11aII* are retained in the duplicated genome by redundancy, which reflects the important of FABP11As in cellular processes of Atlantic salmon.

Figure 10.3. Expression patterns of *fabp11aI*, *fabp11aII* and *fabp11b* in Atlantic salmon adult tissues.

The tissue abbreviations are as follows: brain (Br), eye (E), spleen (Sp), heart (H), gill, (G), ovary (O), stomach (St), pyloric caecum (Pc), posterior kidney (Pk), head kidney (Hk), intestine (I), liver (L), muscle (M), swim bladder (Sb), testis (T), negative control (-) and Atlantic salmon genomic DNA (+). No differential expression patterns of Atlantic salmon *fabp11aI* and *fabp11aII* was observed.



10.6 Conclusion

I identified three *fabp11* genes, *fabp11aI*, *fabp11aII* and *fabp11b*, in Atlantic salmon. The phylogenetic analysis does not clearly support that the 3R WGD, which gave rise to *fabp11a* and *fabp11b* in zebrafish, stickleback, medaka, fugu and tetraodon. I also found three unique rainbow trout *fabp11* EST sequences, which I named *fabp11a*, *fabp11bI* and *fabp11bII* based on their positions on the phylogenetic tree. The phylogenetic tree also supports the 4R WGD that gave rise to *fabp11aI* and *fabp11aII* in Atlantic salmon and *fabp11bI* and *fabp11bII* in rainbow trout. The fact that only a single *fabp11a* in rainbow trout and a single *fabp11b* in Atlantic salmon suggests that there were a subsequent *fabp11b* gene loss in the Atlantic salmon and a *fabp11a* gene loss in rainbow trout after the 4R WGD or they are just not presented in the EST databases. More *fabp11* transcripts from other salmonids, such as grayling, might provide a better resolution of the phylogenetic tree and give a more accurate insight of the evolution of this *fabp* sub-family. No differential expression of Atlantic salmon *fabp11aI* and *fabp11aII* was observed in any of the adult tissues examined. Like the co-expression patterns of zebrafish *fabp11a* and *fabp11b*, Atlantic salmon *fabp11aI*, *fabp11aII* and *fabp11b* also maintain co-expression in brain, heart and ovary. Hence, there was evidence of redundancy of Atlantic salmon *fabp11aI* and *fabp11aII* in terms of expression patterns.

11: Future Work

Although I have characterized all the *fabps* whose transcripts are in the salmonid EST databases, it would be ideal to use the predicted ancestral flanking genes from the comparative genomic analysis (see Section 4.4) to identify the genomic regions where possible *fabp* pseudogenes are predicted to be found when the Atlantic salmon has been completed and assembled. In other words, I would expect to find *fabp1a*, *fabp1b*, *fabp2b*, *fabp7a*, *fabp6b*, *fabp10b* and *fabp11b* pseudogenes in the assembled Atlantic salmon genome. It would be interesting to investigate what the mutations are causative for these genes to have become pseudogenes. Give the gar genome had been sequenced it has been suggested that this species is the pre-WGD species of the 3R WGD, one can compare the genomic organization of *fabp* in this species with my predicted *fabp* synteny blocks that might present in the common ancestor of teleosts. This comparison would test if my predicted *fabp* synteny blocks are correct.

Since my transcription factor binding site analysis was carried out using bioinformatic predictions that are based on the findings of mammals, it would be very useful and provide more accurate predictions if one could first identify the transcription factor binding sites in the assembled Atlantic salmon genome. This can be carried out by chromatin immunoprecipitation (ChIP) and DNA binding assay and should be carried out at different lifestages of Atlantic salmon. Both experiments for identifying the transcription factor binding sites are challenging due to the lack of knowledge of the transcription factor repertoire in Atlantic salmon and the lack of antibodies that recognize

the transcription factors in Atlantic salmon. The transcription factor repertoire can be revealed by the assembled and annotated Atlantic salmon genome. Raising antibodies against each transcription factor might be time-consuming. Moreover, DNA binding assay requires the knowledge of the binding affinity of each transcription factor to the regulatory region of interest and this might take some time. Nonetheless, the result of this investigation can provide possible mechanisms for sub-functionalization.

Finally, it would be ideal to increase the number of ESTs in the northern pike database to obtain and to characterize the *fabp* gene family in this species. As the evolution of *fabp7* sub-family illustrates that northern pike is the pre-WGD species that is most closely-related to the salmonids, the availability of all unique *fabp* transcript sequences in this species would provide a greater detail of the evolution of the *fabp* sub-families and the fate of *fabp* genes in Atlantic salmon.

12: Conclusion

In summary, I have characterized eighteen members of the *fabp* family in Atlantic salmon genome. Overall, the protein sizes of FABPs in Atlantic salmon are similar to the sizes of their orthologs in other teleosts and mammals. The genetic and physical locations of these Atlantic salmon *fabp* genes support the hypothesis that these genes are related to one another by either 3R or 4R or both WGD events. The phylogenetic analysis of the *fabp* family reveals when gene losses occurred in each gene-subfamily and in different teleost lineages, which may reflect different natural selection pressures acting on the individual species. Analysis of amino acid substitutions of FABP2 evolution from zebrafish to Atlantic salmon and rainbow trout FABP2aIs and FABP2aIIs suggests that FABP2aIIs in both salmonid species might have experienced neo-functionalization. RT-PCR analysis provide expression profiles of *fabps* and reveal differential expression patterns of *fabp3s*, *fabp6as* and *fabp7bs*, suggesting that these genes are undergoing sub-functionalization. Moreover, the presence of putative CdxA and SRY binding sites in the indel sequence from *fabp7bI* regulatory region suggest that gain or loss of repressor binding site might be another mechanism of sub-functionalization. Overall, conservation of synteny was observed in *fabp2* and *fabp3* loci in Atlantic salmon compared to the corresponding loci in the five genome-sequenced teleost species, and I was able to predict some ancient synteny blocks in the common ancestor of the teleosts. The implications of this project are that the evolution of the *fabp* gene sub-families provide more evidence for the 3R and the 4R WGD events, though there are gene losses and genome rearrangement

in the duplicated blocks of genes. Also, most of the fabp gene duplicates have differential expression patterns compared to the ancestral gene expression patterns, which suggest that most of the Atlantic salmon gene duplicates remaining in the duplicated genome by sub-functionalization.

Reference List

- Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., and Inoko, H. (2002). Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* *31*, 100-105.
- Albig, W., and Doenecke, D. (1997). The human histone gene cluster at the D6S105 locus. *Hum. Genet.* *101*, 284-294.
- Allendorf, F.W., and Danzmann, R.G. (1997). Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics* *145*, 1083-1092.
- Allendorf, F.W., and Thorgaard, G.H. (1984). Tetraploidy and the evolution of salmonid fishes. In *Evolutionary Genetics of Fishes*, Turner BJ ed., (NY: Plenum Press) pp. 1-46.
- Alves-Costa, F.A., Denovan-Wright, E.M., Thisse, C., Thisse, B., and Wright, J.M. (2008). Spatio-temporal distribution of fatty acid-binding protein 6 (fabp6) gene transcripts in the developing and adult zebrafish (*Danio rerio*). *FEBS J.* *275*, 3325-3334.
- Amano, O. (1993). Expression and localization of intestinal 15 kDa protein in the rat. *Mol. Cell. Biochem.* *123*, 113-120.
- Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., *et al.* (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* *282*, 1711-1714.
- Amri, E.Z., Ailhaud, G., and Grimaldi, P. (1991). Regulation of adipose cell differentiation. II. Kinetics of induction of the aP2 gene by fatty acids and modulation by dexamethasone. *J. Lipid Res.* *32*, 1457-1463.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., *et al.* (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* *297*, 1301-1310.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* *408*, 796-815.
- Atchley, W.R., Fitch, W.M., and Bronner-Fraser, M. (1994). Molecular evolution of the MyoD family of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 11522-11526.

- Ayers, S.D., Nedrow, K.L., Gillilan, R.E., and Noy, N. (2007). Continuous nucleocytoplasmic shuttling underlies transcriptional activation of PPARgamma by FABP4. *Biochemistry* 46, 6744-6752.
- Baier, L.J., Sacchettini, J.C., Knowler, W.C., Eads, J., Paolisso, G., Tataranni, P.A., Mochizuki, H., Bennett, P.H., Bogardus, C., and Prochazka, M. (1995). An amino acid substitution in the human intestinal fatty acid binding protein is associated with increased fatty acid binding, increased fat oxidation, and insulin resistance. *J. Clin. Invest.* 95, 1281-1287.
- Bailey, G.S., Cocks, G.T., and Wilson, A.C. (1969). Gene duplication in fishes: malate dehydrogenases of salmon and trout. *Biochem. Biophys. Res. Commun.* 34, 605-612.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202-8.
- Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28-36.
- Balendiran, G.K., Schnutgen, F., Scapin, G., Borchers, T., Xhong, N., Lim, K., Godbout, R., Spener, F., and Sacchettini, J.C. (2000). Crystal structure and thermodynamic analysis of human brain fatty acid-binding protein. *J. Biol. Chem.* 275, 27045-27054.
- Becak, M.L., Becak, W., and Rabello, M.N. (1966). Cytological evidence of constant tetraploidy in the bisexual South American frog *Odontophrynus americanus*. *Chromosoma* 19, 188-193.
- Bernlohr, D.A., Simpson, M.A., Hertzler, A.V., and Banaszak, L.J. (1997). Intracellular lipid-binding proteins and their genes. *Annu. Rev. Nutr.* 17, 277-303.
- Binas, B., Danneberg, H., McWhir, J., Mullins, L., and Clark, A.J. (1999). Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization. *FASEB J.* 13, 805-812.
- Binas, B., Spitzer, E., Zschiesche, W., Erdmann, B., Kurtz, A., Muller, T., Niemann, C., Blenau, W., and Grosse, R. (1992). Hormonal induction of functional differentiation and mammary-derived growth inhibitor expression in cultured mouse mammary gland explants. *In Vitro Cell. Dev. Biol.* 28A, 625-634.
- Bisbee, C.A., Baker, M.A., Wilson, A.C., Haji-Azimi, I., and Fischberg, M. (1977). Albumin phylogeny for clawed frogs (*Xenopus*). *Science* 195, 785-787.
- Bridges, C.B. (1935). Salivary chromosome maps. *The Journal of Heredity* 26, 60-64.

Chen, L., DeVries, A.L., and Cheng, C.C. (1997). Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences* *94*, 3811-3816.

Cheng, C.H., and Chen, L. (1999). Evolution of an antifreeze glycoprotein. *Nature* *401*, 443-444.

Chmurzynska, A. (2006). The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism. *J. Appl. Genet.* *47*, 39-48.

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. *Mol. Biol. Evol.* *21*, 1146-1151.

Clark, A.J., Neil, C., Gusterson, B., McWhir, J., and Binas, B. (2000). Deletion of the gene encoding H-FABP/MDGI has no overt effects in the mammary gland. *Transgenic Res.* *9*, 439-444.

Clarke, S.D., and Jump, D.B. (1993). Regulation of gene transcription by polyunsaturated fatty acids. *Prog. Lipid Res.* *32*, 139-149.

Coe, N.R., and Bernlohr, D.A. (1998). Physiological properties and functions of intracellular fatty acid-binding proteins. *Biochim. Biophys. Acta* *1391*, 287-306.

Coe, N.R., Simpson, M.A., and Bernlohr, D.A. (1999). Targeted disruption of the adipocyte lipid-binding protein (aP2 protein) gene impairs fat cell lipolysis and increases cellular fatty acid levels. *J. Lipid Res.* *40*, 967-972.

Crossman, M.W. (1994). The mouse ileal lipid-binding protein gene: a model for studying axial patterning during gut morphogenesis. *J. Cell Biol.* *126*, 1547-1564.

Danzmann, R.G. LINKFMFEX: LINKage analysis package for outcrossed families with Male or Female Exchange of the mapping parent, ver. 2.3.

Danzmann, R.G., Davidson, E.A., Ferguson, M.M., Gharbi, K., Koop, B.F., Hoyheim, B., Lien, S., Lubieniecki, K.P., Moghadam, H.K., Park, J., Phillips, R.B., and Davidson, W.S. (2008). Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). *BMC Genomics* *9*, 557.

Davidson, W.S., Koop, B.F., Jones, S.J., Iturra, P., Vidal, R., Maass, A., Jonassen, I., Lien, S., and Omholt, S.W. (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* *11*, 403.

Denovan-Wright, E.M., Pierce, M., Sharma, M.K., and Wright, J.M. (2000a). cDNA sequence and tissue-specific expression of a basic liver-type fatty acid binding protein in adult zebrafish (*Danio rerio*). *Biochim. Biophys. Acta* *1492*, 227-232.

Denovan-Wright, E.M., Pierce, M., and Wright, J.M. (2000b). Nucleotide sequence of cDNA clones coding for a brain-type fatty acid binding protein and its tissue-specific expression in adult zebrafish (*Danio rerio*). *Biochim. Biophys. Acta* *1492*, 221-226.

Desvergne, B., and Wahli, W. (1999). Peroxisome Proliferator-Activated Receptors: Nuclear Control of Metabolism. *Endocrine Reviews* *20*, 649-688.

Di Pietro, S.M., Dell'Angelica, E.C., Veerkamp, J.H., Sterin-Speziale, N., and Santome, J.A. (1997). Amino acid sequence, binding properties and evolutionary relationships of the basic liver fatty-acid-binding protein from the catfish *Rhamdia sapo*. *Eur. J. Biochem.* *249*, 510-517.

Di Pietro, S.M., Perduca, M., Santome, J.A., and Monaco, H.L. (2001). Crystallization and preliminary X-ray study of two liver basic fatty acid-binding proteins. *Acta Crystallogr. D Biol. Crystallogr.* *57*, 1903-1905.

Dickinson Zimmer, J.S., Dyckes, D.F., Bernlohr, D.A., and Murphy, R.C. (2004). Fatty acid binding proteins stabilize leukotriene A4. *Journal of Lipid Research* *45*, 2138-2144.

Dingerkus, G., and Howell, W.M. (1976). Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, *Polyodon spathula*. *Science (New York, N.Y.)* *194*, 842-844.

Doran, J.L., Wei, X.P., and Roy, K.L. (1987). Analysis of a human gene cluster coding for tRNA(GAAPhe) and tRNA(UUULys). *Gene* *56*, 231-243.

Ek, B.A., Cistola, D.P., Hamilton, J.A., Kaduce, T.L., and Spector, A.A. (1997). Fatty acid binding proteins reduce 15-lipoxygenase-induced oxygenation of linoleic acid and arachidonic acid. *Biochim. Biophys. Acta* *1346*, 75-85.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* *8*, 186-194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* *8*, 175-185.

Feng, L. (1994). Brain lipid-binding protein (BLBP): a novel signaling system in the developing mammalian CNS. *Neuron* *12*, 895-908.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., and Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* *6*, 99.

- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* *151*, 1531-1545.
- Foucand, L. (1998). Regulation of expression of human intestinal bile acid-binding protein in Caco-2 cells. *Biochem. J.* *330 (Pt 1)*, 261-265.
- Fu, Y., Luo, N., and Lopes-Virella, M.F. (2000). Oxidized LDL induces the expression of ALBP/aP2 mRNA and protein in human THP-1 macrophages. *Journal of Lipid Research* *41*, 2017-2023.
- Fu, Y., Luo, N., Lopes-Virella, M.F., and Garvey, W.T. (2002). The adipocyte lipid binding protein (ALBP/aP2) gene facilitates foam cell formation in human THP-1 macrophages. *Atherosclerosis* *165*, 259-269.
- Funk, C.D. (2001). Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* *294*, 1871-1875.
- Furuhashi, M., and Hotamisligil, G.S. (2008). Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nat. Rev. Drug Discov.* *7*, 489-503.
- Furuhashi, M., Ura, N., Murakami, H., Hyakukoku, M., Yamaguchi, K., Higashiura, K., and Shimamoto, K. (2002). Fenofibrate improves insulin sensitivity in connection with intramuscular lipid content, muscle fatty acid-binding protein, and beta-oxidation in skeletal muscle. *J. Endocrinol.* *174*, 321-329.
- Gamer, L.W., and Wright, C.V. (1993). Murine Cdx-4 bears striking similarities to the *Drosophila* caudal gene in its homeodomain sequence and early expression pattern. *Mech. Dev.* *43*, 71-81.
- Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Hoyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., *et al.* (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics* *172*, 2405-2419.
- Giuseppe, Z. (1999). Muscle fatty acid-binding protein. *Biochimica Et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* *1441*, 94-105.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. *Science* *274*, 546, 563-7.
- Gold, J.R., Karel, W.J., and Strand, M.R., (1979). Chromosome formulae of North American fishes (College Station: Texas A & M University System, Texas Agricultural Experiment Station).

- Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195-202.
- Gregoire, F.M., Smas, C.M., and Sul, H.S. (1998). Understanding Adipocyte Differentiation. *Physiological Reviews* 78, 783-809.
- Guesnet, P., and Alessandri, J.M. (2011). Docosahexaenoic acid (DHA) and the developing central nervous system (CNS) - Implications for dietary recommendations. *Biochimie* 93, 7-12.
- Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8, R24.
- Hanhoff, T., Lucke, C., and Spener, F. (2002). Insights into binding of fatty acids by fatty acid binding proteins. *Mol. Cell. Biochem.* 239, 45-54.
- Hasebe, T., Kajita, M., Fujimoto, K., Yaoita, Y., and Ishizuya-Oka, A. (2007). Expression profiles of the duplicated matrix metalloproteinase-9 genes suggest their different roles in apoptosis of larval intestinal epithelial cells during *Xenopus laevis* metamorphosis. *Dev. Dyn.* 236, 2338-2345.
- Hauerland, N.H., and Spener, F. (2004). Fatty acid-binding proteins--insights from genetic manipulations. *Prog. Lipid Res.* 43, 328-349.
- He, X., and Zhang, J. (2005). Rapid Subfunctionalization Accompanied by Prolonged and Substantial Neofunctionalization in Duplicate Gene Evolution. *Genetics* 169, 1157-1164.
- Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L., and Kolchanov, N.A. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.* 26, 362-367.
- Hellsten, U., Khokha, M.K., Grammer, T.C., Harland, R.M., Richardson, P., and Rokhsar, D.S. (2007). Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biol.* 5, 31.
- Hertzel, A.V., and Bernlohr, D.A. (2000). The mammalian fatty acid-binding protein multigene family: molecular and genetic insights into function. *Trends Endocrinol. Metab.* 11, 175-180.
- Heuckeroth, R.O., Birkenmeier, E.H., Levin, M.S., and Gordon, J.I. (1987). Analysis of the tissue-specific expression, developmental regulation, and linkage relationships of a rodent gene encoding heart fatty acid binding protein. *Journal of Biological Chemistry* 262, 9709-9717.

Hirsch, N., Zimmerman, L.B., and Grainger, R.M. (2002). *Xenopus*, the next generation: *X. tropicalis* genetics and genomics. *Dev. Dyn.* 225, 422-433.

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. *J. Mol. Evol.* 59, 190-203.

Hohoff, C., and Spener, F. (1998). Correspondence re: Y.E. Shi et al., Antitumor activity of the novel human breast cancer growth inhibitor, mammary-derived growth inhibitor-related gene, MRG. *Cancer Res.*, 57: 3084-3091, 1997. *Cancer Res.* 58, 4015-4017.

Holland, P.W., and Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. *Dev. Biol.* 173, 382-395.

Holland, P.W., Garcia-Fernandez, J., Williams, N.A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Dev. Suppl.* 125-133.

Horton, A.C., Mahadevan, N.R., Ruvinsky, I., and Gibson-Brown, J.J. (2003). Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution. *J. Exp. Zool. B. Mol. Dev. Evol.* 299, 41-53.

Hostetler, H.A., McIntosh, A.L., Atshaves, B.P., Storey, S.M., Payne, H.R., Kier, A.B., and Schroeder, F. (2009). L-FABP directly interacts with PPARalpha in cultured primary hepatocytes. *J. Lipid Res.* 50, 1663-1675.

Huang, T.S., Todorčević, M., Ruyter, B., and Torstensen, B.E. (2010). Altered expression of CCAAT/enhancer binding protein and FABP11 genes during adipogenesis in vitro in Atlantic salmon (*Salmo salar*). *Aquacult. Nutr.* 16, 72.

Hughes, A.L. (1999). Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J. Mol. Evol.* 48, 565-576.

Hughes, A.L., and Piontkivska, H. (2011). Evolutionary diversification of the avian fatty acid-binding proteins. *Gene* 490, 1-5.

Huynh, H.T., Larsson, C., Narod, S., and Pollak, M. (1995). Tumor suppressor activity of the gene encoding mammary-derived growth inhibitor. *Cancer Res.* 55, 2225-2231.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., *et al.* (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946-957.

Kalendar, R., Lee, D., and Schulman A.H. (2009). FastPCR Software for PCR Primer and Probe Design and Repeat Search. *Genes, Genomes and Genomics* 3, 1.

- Karanth, S., Denovan-Wright, E.M., Thisse, C., Thisse, B., and Wright, J.M. (2009). Tandem duplication of the *fabp1b* gene and subsequent divergence of the tissue-specific distribution of *fabp1b.1* and *fabp1b.2* transcripts in zebrafish (*Danio rerio*). *Genome* 52, 985-992.
- Karanth, S., Denovan-Wright, E.M., Thisse, C., Thisse, B., and Wright, J.M. (2008). The evolutionary relationship between the duplicated copies of the zebrafish *fabp11* gene and the tetrapod FABP4, FABP5, FABP8 and FABP9 genes. *FEBS J.* 275, 3031-3040.
- Karanth, S., Lall, S.P., Denovan-Wright, E.M., and Wright, J.M. (2009). Differential transcriptional modulation of duplicated fatty acid-binding protein genes by dietary fatty acids in zebrafish (*Danio rerio*): evidence for subfunctionalization or neofunctionalization of duplicated genes. *BMC Evol. Biol.* 9, 219.
- Kazemi, M.R., McDonald, C.M., Shigenaga, J.K., Grunfeld, C., and Feingold, K.R. (2005). Adipocyte Fatty Acid-Binding Protein Expression and Lipid Accumulation Are Increased During Activation of Murine Macrophages by Toll-Like Receptor Agonists. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25, 1220-1224.
- Kellis, M., Birren, B., W., and Lander, E., S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617-624.
- Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 185, 422-427.
- Kingsley, C., and Winoto, A. (1992). Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol. Cell. Biol.* 12, 4251-4261.
- Kobel, H.R., and Du Pasquier, L. (1986). Genetics of polyploid *Xenopus*. *Trends in Genetics* 2, 310-315.
- Koehn, R.K., and Rasmussen, D.I. (1967). Polymorphic and monomorphic serum esterase heterogeneity in catostomid fish populations. *Biochem. Genet.* 1, 131-144.
- Koop, B.F., von Schalburg, K.R., Leong, J., Walker, N., Lieph, R., Cooper, G.A., Robb, A., Beetz-Sargent, M., Holt, R.A., Moore, R., *et al.* (2008). A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics* 9, 545.
- Korber, B. (2000). HIV Signature and Sequence Variation Analysis. In *Computational Analysis of HIV Molecular Sequences*, Rodrigo, A. G., and Learn, G. H. eds., (Dordrecht, Netherlands: Kluwer Academic Publishers) pp. 55-72.
- Kramer, W., Girbig, F., Gutjahr, U., Kowalewski, S., Jouvenal, K., Müller, G., Tripièr, D., and Wess, G. (1993). Intestinal bile acid absorption. Na(+)-dependent bile acid transport

activity in rabbit small intestine correlates with the coexpression of an integral 93-kDa and a peripheral 14-kDa bile acid-binding membrane protein along the duodenum-ileum axis. *Journal of Biological Chemistry* 268, 18035-18046.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9, 299-306.

Kuwada, Y. (1911). Meiosis in the pollen mother cells of *Zea Mays L.* *Bot. Mag.* 25, 163.

Lai, Y.Y., Lubieniecki, K.P., Phillips, R.B., Chow, W., Koop, B.F., and Davidson, W.S. (2009). Genomic organization of Atlantic salmon (*Salmo salar*) fatty acid binding protein (*fabp2*) genes reveals independent loss of duplicate loci in teleosts. *Mar. Genomics* 2, 193-200.

Leong, J.S., Jantzen, S.G., von Schalburg, K.R., Cooper, G.A., Messmer, A.M., Liao, N.Y., Munro, S., Moore, R., Holt, R.A., Jones, S.J., Davidson, W.S., and Koop, B.F. (2010). *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics* 11, 279.

Lien, S., Gidskehaug, L., Moen, T., Hayes, B.J., Berg, P.R., Davidson, W.S., Omholt, S.W., and Kent, M.P. (2011). A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics* 12, 615.

Liu, R.Z., Denovan-Wright, E.M., Degraeve, A., Thisse, C., Thisse, B., and Wright, J.M. (2004). Differential expression of duplicated genes for brain-type fatty acid-binding proteins (*fabp7a* and *fabp7b*) during early development of the CNS in zebrafish (*Danio rerio*). *Gene Expr. Patterns* 4, 379-387.

Liu, R.Z., Denovan-Wright, E.M., and Wright, J.M. (2003a). Structure, linkage mapping and expression of the heart-type fatty acid-binding protein gene (*fabp3*) from zebrafish (*Danio rerio*). *Eur. J. Biochem.* 270, 3223-3234.

Liu, R.Z., Denovan-Wright, E.M., and Wright, J.M. (2003b). Structure, mRNA expression and linkage mapping of the brain-type fatty acid-binding protein gene (*FABP7*) from zebrafish (*Danio rerio*). *Eur. J. Biochem.* 270, 715-725.

Liu, R.Z., Li, X., and Godbout, R. (2008). A novel fatty acid-binding protein (FABP) gene resulting from tandem gene duplication in mammals: transcription in rat retina and testis. *Genomics* 92, 436-445.

Liu, R.Z., Saxena, V., Sharma, M.K., Thisse, C., Thisse, B., Denovan-Wright, E.M., and Wright, J.M. (2007). The *fabp4* gene of zebrafish (*Danio rerio*)--genomic homology with the mammalian FABP4 and divergence from the zebrafish *fabp3* in developmental expression. *FEBS J.* 274, 1621-1633.

Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I., and Rubin, E.M. (2002). rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.* *12*, 832-839.

Lucke, C., Rademacher, M., Zimmerman, A.W., van Moerkerk, H.T., Veerkamp, J.H., and Ruterjans, H. (2001). Spin-system heterogeneities indicate a selected-fit mechanism in fatty acid binding to heart-type fatty acid-binding protein (H-FABP). *Biochem. J.* *354*, 259-266.

Lundin, L.G., Larhammar, D., and Hallbook, F. (2003). Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates. *J. Struct. Funct. Genomics* *3*, 53-63.

Luxon, B.A., and Weisiger, R.A. (1993). Sex differences in intracellular fatty acid transport: role of cytoplasmic binding proteins. *Am. J. Physiol.* *265*, G831-41.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* *290*, 1151-1155.

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* *154*, 459-473.

Maatman, R.G., van Moerkerk, H.T., Nooren, I.M., van Zoelen, E.J., and Veerkamp, J.H. (1994). Expression of human liver fatty acid-binding protein in *Escherichia coli* and comparative analysis of its binding characteristics with muscle fatty acid-binding protein. *Biochim. Biophys. Acta* *1214*, 1-10.

Makowski, L., Boord, J.B., Maeda, K., Babaev, V.R., Uysal, K.T., Morgan, M.A., Parker, R.A., Suttles, J., Fazio, S., Hotamisligil, G.S., and Linton, M.F. (2001). Lack of macrophage fatty-acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis. *Nat. Med.* *7*, 699-705.

Makowski, L., and Hotamisligil, G.S. (2005). The role of fatty acid binding proteins in metabolic syndrome and atherosclerosis. *Curr. Opin. Lipidol.* *16*, 543-548.

Makowski, L., Brittingham, K.C., Reynolds, J.M., Suttles, J., and Hotamisligil, G.S. (2005). The Fatty Acid-binding Protein, aP2, Coordinates Macrophage Cholesterol Trafficking and Inflammatory Activity. *Journal of Biological Chemistry* *280*, 12888-12895.

Mank, J.E., and Avise, J.C. (2006). Phylogenetic conservation of chromosome numbers in Actinopterygian fishes. *Genetica* *127*, 321-327.

Marcussen, T., Oxelman, B., Skog, A., and Jakobsen, K.S. (2010). Evolution of plant RNA polymerase IV/V genes: evidence of subneofunctionalization of duplicated NRPD2/NRPE2-like paralogs in *Viola* (Violaceae). *BMC Evol. Biol.* *10*, 45.

- Markert, C.L., and Moller, F. (1959). Multiple Forms of Enzymes: Tissue, Ontogenetic, and Species Specific Patterns. *Proc. Natl. Acad. Sci. U. S. A.* 45, 753-763.
- Markert, C.L., Shaklee, J.B., and Whitt, G.S. (1975). Evolution of a gene. Multiple genes for LDH isozymes provide a model of the evolution of gene structure, function and regulation. *Science* 189, 102-114.
- Martin, A. (2001). Is tetralogy true? Lack of support for the "one-to-four rule". *Mol. Biol. Evol.* 18, 89-93.
- Martin, G.G., Atshaves, B.P., Huang, H., McIntosh, A.L., Williams, B.J., Pai, P.J., Russell, D.H., Kier, A.B., and Schroeder, F. (2009). Hepatic phenotype of liver fatty acid binding protein gene-ablated mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* 297, G1053-65.
- Martin, G.G., Danneberg, H., Kumar, L.S., Atshaves, B.P., Erol, E., Bader, M., Schroeder, F., and Binas, B. (2003). Decreased liver fatty acid binding capacity and altered liver lipid distribution in mice lacking the liver fatty acid-binding protein gene. *J. Biol. Chem.* 278, 21429-21438.
- Maruyama, K., Kojima, A., Yasuda, T., Suetomi, K., Kubota, Y., Takahashi, S., Ishikawa, Y., and Fujimori, A. (2008). Expression of brain-type fatty acid-binding protein (fabp7) in medaka during development. *J. Exp. Zool. B. Mol. Dev. Evol.* 310, 577-587.
- McCollum, G.D. (1958). Comparative studies of chromosome pairing in natural and induced tetraploid *Dactylis*. *Chromosoma* 9, 571-605.
- 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. *Curr. Opin. Cell Biol.* 11, 699-704.
- Meyer, A., and Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* 27, 937-945.
- Miklos, G.L., and Rubin, G.M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell* 86, 521-529.
- Moen, T., Hayes, B., Baranski, M., Berg, P.R., Kjolglum, S., Koop, B.F., Davidson, W.S., Omholt, S.W., and Lien, S. (2008). A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* 9, 223.
- Moen, T., Hoyheim, B., Munck, H., and Gomez-Raya, L. (2004). A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim. Genet.* 35, 81-92.

- 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*). *J. Mol. Evol.* *61*, 636-649.
- Motojima, K. (2000). Differential effects of PPARalpha activators on induction of ectopic expression of tissue-specific fatty acid binding protein genes in the mouse liver. *Int. J. Biochem. Cell Biol.* *32*, 1085-1092.
- Murato, Y., Nagatomo, K., Yamaguti, M., and Hashimoto, C. (2007). Two alloalleles of *Xenopus laevis hairy2* gene—evolution of duplicated gene function from a developmental perspective. *Dev. Genes Evol.* *217*, 665-673.
- Muro, E.M., Mah, N., and Andrade-Navarro, M.A. (2011). Functional evidence of post-transcriptional regulation by pseudogenes. *Biochimie* *93*, 1916-1921.
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* *3*, 418-426.
- Nelson, J.S. (1994). *Fishes of the world* (New York, NY [u.a.]: Wiley).
- Neurath, H., Walsh, K.A., and Winter, W.P. (1967). Evolution of Structure and Function of Proteases. *Science* *158*, 1638-1644.
- Newberry, E.P., Xie, Y., Kennedy, S., Han, X., Buhman, K.K., Luo, J., Gross, R.W., and Davidson, N.O. (2003). Decreased hepatic triglyceride accumulation and altered fatty acid uptake in mice with deletion of the liver fatty acid-binding protein gene. *J. Biol. Chem.* *278*, 51664-51672.
- Nichesola, D., Perduca, M., Capaldi, S., Carrizo, M.E., Righetti, P.G., and Monaco, H.L. (2004). Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. *Biochemistry* *43*, 14072-14079.
- Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I., Knudsen, K., *et al.* (2003). A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet.* *34*, 102-115.
- Nolan, V. (2005). Chicken Liver Bile Acid-Binding Protein Is in a Compact Partly Folded State at Acidic pH. Its Relevance to the Interaction with Lipid Membranes. *Biochemistry (N. Y.)* *44*, 8486-8493.
- Norden, C.R. (1961). Comparative Osteology of Representative Salmonid Fishes, with Particular Reference to the Grayling (*Thymallus arcticus*) and its Phylogeny. *Journal of the Fisheries Research Board of Canada* *18*, 679-791.

Ohlsson, G., Moreira, J.M., Gromov, P., Sauter, G., and Celis, J.E. (2005). Loss of expression of the adipocyte-type fatty acid-binding protein (A-FABP) is associated with progression of human urothelial carcinomas. *Mol. Cell. Proteomics* 4, 570-581.

Ohno, S. (1970). *Evolution by gene duplication* Springer-Verlag).

Ohno, S. (1998). The notion of the Cambrian pananimalia genome and a genomic difference that separated vertebrates from invertebrates. *Prog. Mol. Subcell. Biol.* 21, 97-117.

Ohno, S. (1968). Evolution from fish to mammals by gene duplication. *Hereditas* 59, 169.

Olsvik, P.A., Lie, K.K., Jordal, A.E., Nilsen, T.O., and Hordvik, I. (2005). Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. *BMC Mol. Biol.* 6, 21.

Ostbye, T.K., Wetten, O.F., Tooming-Klunderud, A., Jakobsen, K.S., Yafe, A., Etzioni, S., Moen, T., and Andersen, O. (2007). Myostatin (MSTN) gene duplications in Atlantic salmon (*Salmo salar*): evidence for different selective pressure on teleost MSTN-1 and -2. *Gene* 403, 159-169.

Owada, Y., Abdelwahab, S.A., Kitanaka, N., Sakagami, H., Takano, H., Sugitani, Y., Sugawara, M., Kawashima, H., Kiso, Y., Mobarakeh, J.I., *et al.* (2006). Altered emotional behavioral responses in mice lacking brain-type fatty acid-binding protein gene. *Eur. J. Neurosci.* 24, 175-187.

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M., and Lehrach, H. (2003). New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.* 13, 1056-1066.

Pebusque, M.J., Coulier, F., Birnbaum, D., and Pontarotti, P. (1998). Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. *Mol. Biol. Evol.* 15, 1145-1159.

Pelsers, M.M.A.L., Hermens, W.T., and Glatz, J.F.C. (2005). Fatty acid-binding proteins as plasma markers of tissue injury. *Clinica Chimica Acta* 352, 15-35.

Pelton, P.D., Zhou, L., Demarest, K.T., and Burris, T.P. (1999). PPAR γ Activation Induces the Expression of the Adipocyte Fatty Acid Binding Protein Gene in Human Monocytes. *Biochem. Biophys. Res. Commun.* 261, 456-458.

Penhoet, E., Rajkumar, T., and Rutter, W.J. (1966). Multiple forms of fructose diphosphate aldolase in mammalian tissues. *Proc. Natl. Acad. Sci. U. S. A.* 56, 1275-1282.

- Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., *et al.* (2007). Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* *39*, 1256-1260.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* *185*, 416-422.
- Phillips, R.B., Keatley, K.A., Morasch, M.R., Ventura, A.B., Lubieniecki, K.P., Koop, B.F., Danzmann, R.G., and Davidson, W.S. (2009). Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* *10*, 46.
- Pierce, M., Wang, Y., Denovan-Wright, E.M., and Wright, J.M. (2000). Nucleotide sequence of a cDNA clone coding for an intestinal-type fatty acid binding protein and its tissue-specific expression in zebrafish (*Danio rerio*). *Biochim. Biophys. Acta* *1490*, 175-183.
- 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 Genet.* *20*, 481-490.
- Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R., and White, J. (2001). The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.* *29*, 159-164.
- Quinn, N.L., Boroevich, K.A., Lubieniecki, K.P., Chow, W., Davidson, E.A., Phillips, R.B., Koop, B.F., and Davidson, W.S. (2010). Genomic organization and evolution of the Atlantic salmon hemoglobin repertoire. *BMC Genomics* *11*, 539.
- Ramsden, S.D., Brinkmann, H., Hawryshyn, C.W., and Taylor, J.S. (2003). Mitogenomics and the sister of Salmonidae. *Trends in Ecology & Evolution* *18*, 607-610.
- Rath, B., Pandey, R.S., Debata, P.R., Maruyama, N., and Supakar, P.C. (2008). Molecular characterization of senescence marker protein-30 gene promoter: identification of repressor elements and functional nuclear factor binding sites. *BMC Mol. Biol.* *9*, 43.
- Rees, H. (1964). The Question of Polyploidy in the Salmonidae. *Chromosoma* *15*, 275-279.
- Rehse, P.H., and Davidson, W.S. (1986). Purification and properties of a C-type isozyme of lactate dehydrogenase from the liver of the atlantic COD (*Gadus morhua*). *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* *84*, 145-150.

- Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., and Laudet, V. (2001a). Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.* *11*, 781-788.
- Robinson-Rechavi, M., Marchand*, O., Escriva, H., and Laudet, V. (2001b). An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Current Biology* *11*, R458-R459.
- Rolf, B., Oudenampsen-Kruger, E., Borchers, T., Faergeman, N.J., Knudsen, J., Lezius, A., and Spener, F. (1995). Analysis of the ligand binding properties of recombinant bovine liver-type fatty acid binding protein. *Biochim. Biophys. Acta* *1259*, 245-253.
- Rolph, M.S., Young, T.R., Shum, B.O.V., Gorgun, C.Z., Schmitz-Peiffer, C., Ramshaw, I.A., Hotamisligil, G.S., and Mackay, C.R. (2006). Regulation of Dendritic Cell Function and T Cell Priming by the Fatty Acid-Binding Protein aP2. *The Journal of Immunology* *177*, 7794-7801.
- Sacchettini, J.C., Hautf, S.M., Van Camp, S.L., Cistola, D.P., and Gordon, J.I. (1990). Developmental and structural studies of an intracellular lipid binding protein expressed in the ileal epithelium. *J. Biol. Chem.* *265*, 19199-19207.
- Saez, F.A., and Brum, N. (1960). Chromosomes of South American amphibians. *Nature* *185*, 945.
- Sanchez-Font, M.F., Bosch-Comas, A., Gonzalez-Duarte, R., and Marfany, G. (2003). Overexpression of FABP7 in Down syndrome fetal brains is associated with PKNOX1 gene-dosage imbalance. *Nucleic Acids Res.* *31*, 2769-2777.
- Schaap, F., van der Vusse, G.J., and Glatz, J.F. (2002). Evolution of the family of intracellular lipid binding proteins in vertebrates. *Molecular and Cellular Biochemistry* *239*, 69-77.
- Scheja, L., Makowski, L., Uysal, K.T., Wiesbrock, S.M., Shimshek, D.R., Meyers, D.S., Morgan, M., Parker, R.A., and Hotamisligil, G.S. (1999). Altered insulin secretion associated with reduced lipolytic efficiency in aP2^{-/-} mice. *Diabetes* *48*, 1987-1994.
- Serhan, C.N. (2007). Resolution phase of inflammation: novel endogenous anti-inflammatory and proresolving lipid mediators and pathways. *Annu. Rev. Immunol.* *25*, 101-137.
- Sharma, M.K., Denovan-Wright, E.M., Degrave, A., Thisse, C., Thisse, B., and Wright, J.M. (2004). Sequence, linkage mapping and early developmental expression of the intestinal-type fatty acid-binding protein gene (*fabp2*) from zebrafish (*Danio rerio*). *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* *138*, 391-398.

Sharma, M.K., Liu, R.Z., Thisse, C., Thisse, B., Denovan-Wright, E.M., and Wright, J.M. (2006). Hierarchical subfunctionalization of *fabp1a*, *fabp1b* and *fabp10* tissue-specific expression may account for retention of these duplicated genes in the zebrafish (*Danio rerio*) genome. *FEBS J.* 273, 3216-3229.

Shaver, D.L. (1963). The Effect of Structural Heterozygosity on the Degree of Preferential Pairing in Allotetraploids of *Zea*. *Genetics* 48, 515-524.

Shen, W.J., Sridhar, K., Bernlohr, D.A., and Kraemer, F.B. (1999). Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5528-5532.

Shi, Y.E., Ni, J., Xiao, G., Liu, Y.E., Fuchs, A., Yu, G., Su, J., Cosgrove, J.M., Xing, L., Zhang, M., *et al.* (1997). Antitumor activity of the novel human breast cancer growth inhibitor, mammary-derived growth inhibitor-related gene, MRG. *Cancer Res.* 57, 3084-3091.

Shum, B.O., Mackay, C.R., Gorgun, C.Z., Frost, M.J., Kumar, R.K., Hotamisligil, G.S., and Rolph, M.S. (2006). The adipocyte fatty acid-binding protein aP2 is required in allergic airway inflammation. *J. Clin. Invest.* 116, 2183-2192.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* 6, 715-722.

Sidow, A. (1992). Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5098-5102.

Smathers, R.L., and Petersen, D.R. (2011). The human fatty acid-binding protein family: evolutionary divergences and functions. *Hum. Genomics* 5, 170-191.

Spring, J. (1997). Vertebrate evolution by interspecific hybridisation--are we polyploid? *FEBS Lett.* 400, 2-8.

Storch, J., and McDermott, L. (2009). Structural and functional analysis of fatty acid-binding proteins. *J. Lipid Res.* 50 *Suppl*, S126-31.

Stults, D.M., Killen, M.W., Pierce, H.H., and Pierce, A.J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res.* 18, 13-18.

Svärdson, G. (1945). *Chromosome studies on Salmonidae* (Stockholm: Ivar Høeggströms).

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596-1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731-2739.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.* 13, 382-390.

Taylor, J.S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615-643.

Taylor, S.S., Oxley, S.S., Allison, W.S., and Kaplan, N.O. (1973). Aminoacid sequence of dogfish M4 lactate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* 70, 1790-1793.

Thiebaud, C.H., and Fischberg, M. (1977). DNA content in the genus *Xenopus*. *Chromosoma* 59, 253-257.

Tso, A.W., Xu, A., Sham, P.C., Wat, N.M., Wang, Y., Fong, C.H., Cheung, B.M., Janus, E.D., and Lam, K.S. (2007). Serum adipocyte fatty acid binding protein as a new biomarker predicting the development of type 2 diabetes: a 10-year prospective study in a Chinese cohort. *Diabetes Care* 30, 2667-2672.

Tsoi, S.C., and Li, S.S. (1994). The nucleotide and deduced amino-acid sequences of a cDNA encoding lactate dehydrogenase from *Caenorhabditis elegans*: the evolutionary relationships of lactate dehydrogenases from mammals, birds, amphibian, fish, nematode, plants, bacteria, mycoplasma, and plasmodium. *Biochem. Biophys. Res. Commun.* 205, 558-564.

Uyeno, T., and Smith, G.R. (1972). Tetraploid origin of the karyotype of catostomid fishes. *Science* 175, 644-646.

van Dodeweerd, A.M., Hall, C.R., Bent, E.G., Johnson, S.J., Bevan, M.W., and Bancroft, I. (1999). Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* 42, 887-892.

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. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1638-1643.

Vassileva, G., Huwyler, L., Poirier, K., Agellon, L.B., and Toth, M.J. (2000). The intestinal fatty acid binding protein is not essential for dietary fat absorption in mice. *The FASEB Journal* 14, 2040-2046.

Veerkamp, J.H. (1995). Fatty acid transport and fatty acid-binding proteins. *Proc. Nutr. Soc.* 54, 23-37.

- Veerkamp, J.H., and Maatman, R.G. (1995). Cytoplasmic fatty acid-binding proteins: their structure and genes. *Prog. Lipid Res.* 34, 17-52.
- Veerkamp, J.H., and van Moerkerk, H.T. (1993). Fatty acid-binding protein and its relation to fatty acid oxidation. *Mol. Cell. Biochem.* 123, 101-106.
- Veitia, R.A., and Bottani, S. (2009). Whole genome duplications and a 'function' for junk DNA? Facts and hypotheses. *PLoS One* 4, e8201.
- Venkatachalam, A.B., Thisse, C., Thisse, B., and Wright, J.M. (2009). Differential tissue-specific distribution of transcripts for the duplicated fatty acid-binding protein 10 (fabp10) genes in embryos, larvae and adult zebrafish (*Danio rerio*). *FEBS J.* 276, 6787-6797.
- von Schalburg, K.R., Yazawa, R., de Boer, J., Lubieniecki, K.P., Goh, B., Straub, C.A., Beetz-Sargent, M.R., Robb, A., Davidson, W.S., Devlin, R.H., and Koop, B.F. (2008). Isolation, characterization and comparison of Atlantic and Chinook salmon growth hormone 1 and 2. *BMC Genomics* 9, 522.
- Wallace, R. (1995). Epithelial proliferation and differentiation in the mammary gland do not correlate with cFABP gene expression during early pregnancy. *Dev. Genet.* 17, 167.
- Watanabe, A., Toyota, T., Owada, Y., Hayashi, T., Iwayama, Y., Matsumata, M., Ishitsuka, Y., Nakaya, A., Maekawa, M., Ohnishi, T., *et al.* (2007). Fabp7 maps to a quantitative trait locus for a schizophrenia endophenotype. *PLoS Biol.* 5, e297.
- Wittbrodt, J., Meyer, A., and Schartl, M. (1998). More genes in fish? *BioEssays* 20, 511-515.
- Wolf, U., Ritter, H., Atkin, N.B., and Ohno, S. (1969). Polyploidization in the fish family Cyprinidae, order Cypriniformes. I. DNA-content and chromosome sets in various species of Cyprinidae. *Humangenetik* 7, 240-244.
- Wolfe, K.H. (2006). Comparative genomics and genome evolution in yeasts. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 403-412.
- Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2, 333-341.
- Wolfe, K.H., Gouy, M., Yang, Y.W., Sharp, P.M., and Li, W.H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6201-6205.
- Woram, R.A., McGowan, C., Stout, J.A., Gharbi, K., Ferguson, M.M., Hoyheim, B., Davidson, E.A., Davidson, W.S., Rexroad, C., and Danzmann, R.G. (2004). A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates

and segregation distortion in hybrid versus pure strain mapping parents. *Genome* 47, 304-315.

Xu, A., Wang, Y., Xu, J.Y., Stejskal, D., Tam, S., Zhang, J., Wat, N.M., Wong, W.K., and Lam, K.S. (2006). Adipocyte fatty acid-binding protein is a plasma biomarker closely associated with obesity and metabolic syndrome. *Clin. Chem.* 52, 405-413.

Yeung, D.C., Xu, A., Cheung, C.W., Wat, N.M., Yau, M.H., Fong, C.H., Chau, M.T., and Lam, K.S. (2007). Serum adipocyte fatty acid-binding protein levels were independently associated with carotid atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* 27, 1796-1802.

Zaghini, I. (1999). Identification of a Bile Acid-responsive Element in the Human Ileal Bile Acid-binding Protein Gene. INVOLVEMENT OF THE FARNESOID X RECEPTOR/9-cis-RETINOIC ACID RECEPTOR HETERODIMER. *The Journal of Biological Chemistry* 274, 29749-29754.

Zhang, F., Lucke, C., Baier, L.J., Sacchettini, J.C., and Hamilton, J.A. (2003). Solution structure of human intestinal fatty acid binding protein with a naturally-occurring single amino acid substitution (A54T) that is associated with altered lipid metabolism. *Biochemistry* 42, 7339-7347.

Zhou, X., Lin, Z., and Ma, H. (2010). Phylogenetic detection of numerous gene duplications shared by animals, fungi and plants. *Genome Biol.* 11, R38.

Zimmerman, A.W., and Veerkamp, J.H. (2002). New insights into the structure and function of fatty acid-binding proteins. *Cell Mol. Life Sci.* 59, 1096-1116.