THE FATTY ACID-BINDING PROTEIN (*fabp*) GENES OF SPOTTED GREEN PUFFERFISH (*TETRAODON NIGROVIRIDIS*) - COMPARATIVE STRUCTURAL GENOMICS AND TISSUE-SPECIFIC DISTRIBUTION OF THEIR TRANSCRIPTS

by

ARULOLI THIRUMARAN

Submitted in partial fulfilment of the requirements
for the degree of MASTER OF SCIENCE

at

Dalhousie University
Halifax, Nova Scotia
December 2013

© Copyright by Aruloli Thirumaran, 2013

Dedicated to my family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The fatty acid-binding protein (*fabp*) genes belong to the multigene family of intracellular lipid-binding proteins (iLBP). To date, 12 different FABPs have been identified in various vertebrate genomes. Owing to the fish-specific whole genome duplication (FSGD) event, many fishes have duplicated copies of the *fabp* genes. Here, I identified and characterized the *fabp* genes of spotted green pufferfish (*Tetraodon nigroviridis*). Initially, a BLAST search was performed and ten *fabp* genes were identified, out of which, three were retained in the pufferfish genome as duplicated copies. The putative pufferfish Fabp proteins shared greatest sequence identity and similarity with their teleost and tetrapod orthologs. Conserved gene synteny was evident between the pufferfish *fabp* genes and human, zebrafish, three-spined stickleback and medaka *FABP/fabp* genes, providing evidence that the duplicated copies of pufferfish *fabp* genes most likely arose as a result of the FSGD. The differential tissue-specific distribution of pufferfish *fabp* transcripts suggests divergent spatial regulation of duplicated pairs of *fabp* genes.

LIST OF ABBREVIATIONS USED

BLAST	Basic local alignment search tool
bp	Base pair(s)
cDNA	Complementary deoxyribonucleic acid
Fabp	Fish fatty acid-binding protein
fabp	Fish fatty acid-binding protein gene
FABP	Human fatty acid-binding protein
FABP	Human fatty acid-binding protein gene
FSGD	Fish-specific genome duplication
iLBP	Intracellular lipid-binding protein
kb	Kilobase
kDa	Kilodalton
LCN1	Human lipocalin 1 protein
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
PCR	Polymerase chain reaction
rpl13α	Ribosomal protein large subunit 13 alpha gene
rpl7	Ribosomal protein large subunit 7
RT-PCR	Reverse transcription-polymerase chain reaction
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
SEM	Standard error of mean

ACKNOWLEDGEMENTS

I sincerely thank my supervisor, Dr. Jonathan M Wright, for his mentorship and optimism throughout this study. I would like to extend my heartfelt thanks to Dr. Eileen Denovan-Wright, for her constant encouragement and enthusiasm. I thank Drs. Abdulkadir Bayir and Mehtap Bayir for their technical assistance during the study. I would also like to thank my colleagues/friends, Manoj and Anand for their motivational and insightful discussions.

I acknowledge the financial support from NSERC and Dalhousie University, without which none of this work would have been possible. Last, but not the least, I would like to thank the Almighty for blessing me with a beautiful family, whose steadfast belief and unconditional love, made me what I am today.

CHAPTER 1

Introduction

1.1 Fatty acid-binding proteins (FABPs)

The members of the family of intracellular lipid-binding protein (iLBP) are known to play vital roles in the transport of hydrophobic ligands such as long-chain fatty acids and retinoid to fuel various metabolic requirements and to regulate gene expression by activating specific transcription factors (Bernlohr *et al.,* 1997).ILBPs are compact, low molecular mass proteins of 14-16 kDa, having a highly conserved gene structure and protein conformation. ILBP gene structure consists of four exons separated by three introns, a common characteristic of the iLBP gene family (Bernlohr *et al.,* 1997). The protein structure is composed of ten anti-parallel β-sheets and two short α-helices. The β-sheets together form a barrel like structure with a hollow central cavity, in which the ligand binds and the two α-helices at the N-terminal end act like a lid to the β-barrel structure (Veerkamp and Maatman, 1995). This conformation forms a characteristic three-dimensional clamshell-shaped structure (Schaap *et al.,* 2002). Based on ligand affinity, and specificity for different ligands, iLBPs can be divided into three subfamilies, the retinol-binding proteins (RBPs), the cellular retinoic acid-binding proteins (CRABPs) and the fatty acid-binding proteins (FABPs/Fabps) (Hanhoff *et al.,* 2002). The multigene family of iLBPs arose as a result of various duplications and evolutionary diversification of an ancestral iLBP gene (Schleicher *et al.,* 1995). So far, 18 different iLBPs including 12 FABPs, 4 RBPs and 2 CRABPs, have been identified in vertebrates. Schaap *et al.,*(2002)reported that due to the absence of iLBPs in plants and fungi, the ancestral

iLBP gene arose after the divergence of animals from plants and fungi approximately 1000-1200 million years ago.

The FABPs have a broad specificity in binding to various ligands such as long-chain fatty acids, bile salts and eicosanoids (Smathers and Petersen, 2011). Although the exact function(s) of FABPs is poorly defined, experimental evidence suggests their role in trafficking fatty acids for lipid storage, cellular signaling, peroxisomal or mitochondrial oxidation and enzymatic interactions (Furuhashi and Hotamisligil, 2008). They are also reported to be involved in the regulation of receptor-mediated gene expression by directly binding to fatty acids and other hypolipidemic drugs (Storch and Corsico, 2008).

When FABPs were first identified, they were named based on the initial tissue of isolation, e.g., intestinal FABP (I-FABP), liver-type FABP (L-FABP). However, due to overlapping patterns of tissue-specific distribution of transcripts and their proteins i.e. different *fabp*s expressed in the same tissue, here I use the nomenclature proposed by Hertzel and Bernlohr (2000), where they used Arabic numerals in the chronological order based on the initial site of isolation, i.e., FABP1 (liver), FABP2 (intestine), FABP3 (heart), FABP4 (adipocyte), FABP5 (epidermal), FABP6 (ileal), FABP7 (brain), FABP8 (myelin), FABP9 (Testis), FABP10 (liver-basic).

1.2 Teleost Fabp/*fabp*

Although mammalian FABPs have been well studied for four decades (Ockner *et al.,* 1972), less is known about the teleost Fabps. The teleost Fabp was initially identified from the heart tissue of ocean pout, *Macrozoarces americanus*and sea raven, *Hemitripterus americanus* (Stewart and Driedzic, 1988). Subsequently, teleost

Fabp/*fabp*s were identified in the nurse shark (*Ginglymostoma cirratum*) liver (Medzihradszky *et al.,* 1992), aerobic muscle of striped bass, *Morone saxatilis* (Londraville and Sidell, 1996), H-FABP of rainbow trout, *Oncorhynchus mykiss* (Ando *et al.,* 1998) and Antarctic teleost fishes, *Chaenocephalus aceratus, Cryodraco antarcticus, Gobionotothen gibberifrons* and *Notothenia coriiceps* (Vayda *et al.,* 1998). More recently, *fabp* genes have been studied extensively in zebrafish, *Danio rerio* (Alves-Costa *et al.,* 2008; Denovan-Wright *et al.,* 2000a, 2000b; Karanth *et al.,* 2008;Karanth *et al.,* 2009; Liu *et al.,* 2003a, 2003b, 2004, 2007; Sharma *et al.,* 2004, 2006; Venkatachalam *et al.,* 2009), medaka, *Oryzias latipes* (Parmar *et al.,* 2012a, 2012b; Parmar and Wright, 2013a, 2013b), three-spined stickle back, *Gasterosteus aculeatus* (Parmar *et al.,* 2012a, 2012b; Parmar and Wright, 2013a, 2013b) and Atlantic salmon, *Salmo salar* (Lai *et al*., 2009; Lai *et al*., 2012).

Studies on zebrafish *fabp* started with the discovery of intestinal fatty acid-binding protein gene (*fabp2*) (Pierce *et al.,*2000), followed by the identification and tissue-specific expression patterns of other paralogous zebrafish *fabp*s (Alves-Costa *et al.,* 2008; Denovan-Wright *et al.,* 2000a, 2000b; Karanth *et al.,* 2008;Karanth *et al.,* 2009; Liu *et al.,* 2003a, 2003b, 2004, 2007; Sharma *et al.,* 2004, 2006; Venkatachalam *et al.,* 2009). To date, twelve *fabp*s have been reported in zebrafish, where eight genes (four pairs) are retained in the genome as duplicated copies, owing to the teleost fish-specific whole genome duplication (FSGD) event. The *fabp1b.2* gene however, was identified as a tandem duplicate of *fabp1b*, presumably due to unequal crossing over during meiosis, an event that most likely arose subsequent to the FSGD (Karanth *et al.,* 2009). Identification and structural characterization of medaka and three-spined stickleback *fabp*

genes have shown that ten *fabp* genes (five pairs) were retained in the medaka and three-spined stickleback genomes as duplicated copies (Parmar *et al.,* 2012a, 2012b; Parmar and Wright, 2013a, 2013b).

1.3 *Tetraodon nigroviridis* (Marion de Procé, 1822)

Tetraodon nigroviridis, commonly called the spotted green pufferfish (hereafter referred to simply as pufferfish), lives in fresh/brackish waters of Southeast Asia. Tetraodon (meaning "four teeth" in Latin) has two fused upper and lower teeth, forming a beak like structure. Interest in Tetraodon as model organism for genomics arose after Hinegardner in 1968, discovered its remarkably small genome of about 380 million base pairs compared to the 1.6 billion base pairs of the zebrafish genome. The compact pufferfish genome is mostly due to the reduction in the size of intronic and intergenic regions, while the structure of coding regions is well conserved with that of other teleost fishes and vertebrates in general. It is also important to note that transposable elements which constitute about 45% of the human genome represent only 3.8% of the pufferfish genome and 2.7% of *Takifugu sp.* (Crollius *et al.,* 2000; Dasilva *et al*., 2002). Another factor that may, in part, explain the compact pufferfish genome is the resistance of the genome to retrotransposon-induced insertions, as proposed after a comparative study of diodons and tetraodontid pufferfish (Neafsey and Palumbi, 2003).

A comparative analysis based on the distribution of duplicated genes in the pufferfish genome with their human orthologs provides evidence for a FSGD (Jaillon *et al.,* 2004). Based on these findings, Jaillon *et al.,* (2004), proposed an ancient karyotype of the teleost fish, consisting of 12 chromosomes, where they confirmed the occurrence of a possible FSGD event after their divergence from the ancestral tetrapod lineage.

These characteristics make the pufferfish a suitable reference genome for the study of comparative and evolutionary genomics. A complete dataset of pufferfish genome is available at the National Center for Biotechnology Research (NCBI) database. To date, no studies have described members of the iLBP multigene family in the pufferfish genome, with the exception of a recent report by Parmar and Wright (*in press*) on *fabp7a*, *fabp7b* and *fabp10* of pufferfish. Here, I report all the other uncharacterized *fabp* genes of pufferfish.

1.4 Objectives of the study

The focus of this study was to answer the following questions:

1. Are the putative *fabp*s identified in the pufferfish genome orthologs of their mammalian and teleost counterparts?
2. Based on their tissue distribution patterns, could I provide insight into the evolutionary processes that lead to the retention or loss of the duplicated pufferfish *fabp* genes?
3. How are the duplicated *fabp* genes retained in the pufferfish genome compared to duplicated *fabp*s in other teleost genomes?

CHAPTER 2

Materials and Methods

2.1 Bioinformatics and computational study

2.1.1 Identification and structural elucidation of pufferfish *fabp* genes

The putative gene sequence of each pufferfish *fabp* gene was retrieved by performing separate BLAST (http://useast.ensembl.org/Multi/blastview) searches with corresponding orthologous Fabp protein sequence of zebrafish or medaka as a query (Table 1) to the Ensembl genome sequence database (www.ensembl.org). To determine the gene structure the nucleotide sequence database of NCBI was searched for the longest cDNA sequence encoded by each pufferfish *fabp* gene by performing a BLAST using the Ensembl-derived genomic sequence as a query. The location of a putative TATA box, exons and introns splice junctions, presence of polyadenylation signals were determined for all the *fabp* genes of pufferfish based on their location on other orthologous *fabp* genes in zebrafish (Alves-Costa *et al.,* 2008; Denovan-Wright *et al.,* 2000a, 2000b; Karanth *et al.,* 2008;Karanth *et al.,* 2009; Liu *et al.,* 2003a, 2003b, 2004, 2007; Sharma *et al.,* 2004, 2006; Venkatachalam *et al.,* 2009), medaka and three-spined stickleback (Parmar *et al.,* 2012a, 2012b; Parmar and Wright, 2013a, 2013b).

2.1.2 Multiple sequence alignment

Sequence alignment was performed using CLUSTALW (Thompson *et al.,* 1994) application available at BioEdit version 7.2.0 software (http://www.mbio.ncsu.edu/bioedit/page2.html). The protein sequence of each pufferfish Fabp retrieved from Ensembl database was aligned with Fabp protein

Table 1: Fatty acid-binding protein NCBI query sequence IDs of zebrafish used for the identification of pufferfish *fabp* genes

Pufferfish *fabp*s	Query protein	Query sequence ID
fabp1	Fabp1	NP_001019822
fabp2	Fabp2	NP_571506
fabp3	Fabp3	NP_694493
fabp6a	Fabp6	NP_001002076
fabp6b	Fabp6	NP_001002076
fabp7a	Fabp7a	NP_571680
fabp7b	Fabp7b	NP_999979
fabp10	Fabp10a	NP_694492
fabp11a	Fabp11a	NP_001004682
fabp11b	Fabp11b	NP_001018394

sequences from human, rat, mouse, zebrafish, medaka and three-spined stickleback. The percentage of sequence similarity and identity was determined using the pairwise alignment of BLOSUM62 matrix (Gromiha, 2010).

2.1.3 Phylogenetic analysis and conserved gene synteny

The phylogenetic relationship of the pufferfish Fabp protein sequences were determined using MEGA5 (Tamura *et al.,* 2011), a computational genetic analysis tool. A bootstrapped neighbor-joining was constructed with Fabp/FABP sequences of human, rat, mouse, zebrafish, medaka and three-spined stickleback using the Poisson correction distance model based on amino acid substitution per site. A maximum-likelihood tree was constructed using WAG substitution model based on gamma distribution for rate heterogeneity.

Conserved gene synteny was determined based on some regions of the ancestral chromosome that remain unaltered in the course of evolution (Woods *et al.,* 2005).Genes that are present along with the specific pufferfish *fabp* genes in specific chromosome were identified in medaka, zebrafish, three-spined stickleback and human genome and the conserved gene synteny was determined manually using the region overview option of Ensembl database to locate the co-localized genes.

2.2 Husbandry of fish (*Tetraodon nigroviridis*)

Adult pufferfish were brought from a commercial supplier, Pets Unlimited, Halifax, Nova Scotia, Canada. All the animal care conditions and experimental protocols were reviewed by the Animal Care Committee of Dalhousie University, in accordance with the recommendations of Canadian Council on Animal Care.

2.3 Tissue dissection and RNA isolation

Fish were anesthetized using 0.2% (w/v) Tricaine methanesulfonate (FinquelMS-222, Argent Chemical Laboratories, Redmond, WA, USA). Liver, intestine, muscle, brain, heart, eye, swim bladder, gills, kidney, skin, ovary and testis were dissected from three male and three female fish. Tissues were immediately homogenized in TRIzol reagent (Invitrogen, Carlsbad, California, USA) and total RNA was isolated from each tissue based on the protocol suggested by the supplier. The concentration of the isolated RNA was measured using a spectrophotometer. The quality of the RNA was determined by calculating the ratio of absorbance at 260 and 280 nm (A_{260}/A_{280}), where the ratio of pure RNA is 2. About 2 µg of total RNA with A_{260}/A_{280} ratio of greater than 1.8 was used for conversion of mRNA to cDNA.

2.4 Reverse transcription (RT), primer optimization and polymerase chain reaction (PCR)

Complimentary DNA (cDNA) was synthesized using the Omniscript Reverse Transcription Kit (Qiagen, Mississauga, Canada). For primer optimization, cDNA derived from a whole pufferfish total RNA was used. A pufferfish was anesthetized, frozen in liquid nitrogen, homogenized and RNA was isolated as described above. All the primers used in the study (Table 2) were designed using NCBI primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). To avoid amplification of products from contaminating genomic DNA or heterogeneous nuclear RNA, either the forward or the reverse primer was designed to span an exon-exon junction. Primers were purchased from Integrated DNA Technologies Inc. (Coralville, Iowa, USA) and were prepared according to the manufacturer's instructions to a final stock concentration of 100 µM.

Table 2: Primers used for quantitative PCR (qPCR) amplification of pufferfish *fabp* genes and normalization genes, and primer sets used for sequencing cDNA.

Gene	Forward primer	Reverse primer	Tm (°C)*
RT-qPCR primers			
fabp1	AGGCCATTGGTGTTCCTG	TGTCCTCCAGACGAAAGACC	56.2
fabp2	CAACATGGTCAAGAGGAAGCTGG	GCTCCACGAACCCGATAGCTC	59.5
fabp3	GGAACCTCGTCAAAAGCGA	CGGTAACAACGGACTTAACAT	54.7
fabp6a	TACGACGACTTCTGCAAACTGC	CATGAAGACGGTGGCCTTGAAC	60.3
fabp6b	TCGTGATGGACGATGGGAGG	GAAGGTCACACCCTTCTCTCC	56.8
fabp11a	TGAAAGCAATTGGTGTTGGGT	CGAACGGCTCGTTGAGTTTG	56.2
fabp11b	AAGACGATCGGTGTGAACTTTGC	GTCATCCGACCATCTGCTGT	57.4
rpl13α	CCCTACGACAAGAGGAAGCG	GGTACTTCCAGCCAACCTCA	58.6
rpl7	CGAGAAAAAGGCCCGCAAGG	GCTTGACGAACACTCCGTTG	61.2
Sequencing primers			
fabp1 (genomic)	GCTTTGCATACGCCAGAGTG	GGCCAGTAGTGGTGGTGATT	57.0
fabp2	CAGCACTCGGGAGATGAGAT	TTGTAGCTCTGGACGAGTTC	55.2
fabp3	GGAACCTCGTCAAAAGCGA	GAGCGTGAGCGTCAGAGTGA	57.3
fabp6a	GAGACCCAGGAGGGTTACGAC	GTTCTCTTCAGCACCACTGAGC	58.5
fabp6b	GAGTTTCTGGATGCCATCG	GTGCAATTCATGATCAGCTTG	54.0
fabp11a	TGAAAGCAATTGGTGTTGGGT	GAGGAGACCTTCCTCACACCC	57.3
fabp11b	AGGAAGTCTGCGTGTGTGAG	TGGCATGATGGAAGACTGGA	56.6

*Annealing temperature

For primer optimization, PCR was performed with cDNA prepared from RNA of a whole fish using the Platinum PCR Super mix (Invitrogen, Carlsbad, California, USA). RT-PCR reactions contained 0.4 µM forward and reverse primers and 100 ng cDNA template in a final volume of 25 µl. Optimization of annealing temperature was performed by using six different temperatures for each of *fabp* gene primers (data not shown). The RT-PCR conditions consisted of an initial denaturation at 94.0ºC for 2 min, followed by denaturation at 94.0 ºC for 30 sec and 35 cycles, annealing of primers at an optimum temperature (differs for each *fabp* gene) for 30 sec, extension at 72.0 ºC for 1 min and a final extension at 72.0 ºC for 5 min. The amplified PCR products were visualized on a 1% agarose gel electrophoresis to confirm the expected product size. All the RT-PCR products were sequenced by Macrogen corp., (Rockville, MD, USA) to confirm the product size and sequence. The gene structure of each pufferfish *fabp*s were confirmed by comparing the sequence of the PCR product with the gene sequence data available at Ensembl and NCBI.

2.5 Quantitative PCR (RT-qPCR)

To determine the tissue distribution and steady-state mRNA levels of pufferfish *fabp* genes, quantitative PCR was performed. The PCR product mentioned above was subjected to a copy number calculation, where the spectrophotometric measurement of copy number of the double stranded DNA was determined in each tissue as outlined by Bustin *et al.,* 2005. A rotor-gene (RG-3000) thermal cycler system (Corbett Research, Sydney, NSW, Australia) and the QuantiTech SYBR Green PCR kit (Qiagen, Mississauga, Canada) were used for the qPCR analysis. The qPCR was run at a final volume of 20 µl with 100 µM each of forward primer, reverse primer and 100 ng cDNA

template. The qPCR conditions consisted of an initial denaturation for 15 min at 95.0 ºC, followed by 40 cycles of denaturation at 95.0 ºC for 20 sec, primer annealing at optimum temperature specific for each *fabp* gene (Table 2)for 30 sec and elongation at 72.0 ºC for 30 sec.

Initially, standards were prepared by serial dilution of PCR product amplified from cDNA of a whole pufferfish. A qPCR was later performed for each *fabp* gene of pufferfish with the standards prepared as template. An appropriate standard curve was constructed with amplification efficiencies ranging from 90 to 100 %. To perform an absolute quantification analysis of a particular *fabp* gene transcript, the selected standards were amplified along with the unknown cDNA of the above mentioned tissue samples for each *fabp* gene. Also, every qPCR run included a negative control, in which the cDNA template was omitted. The steady-state levels of two reference genes, *rpl13α* and *rpl7*, which are constitutively expressed in all the tissues, were also determined. The ratio of copy number of the each pufferfish *fabp* transcript and copy number of the each reference gene transcript was calculated to determine the mean normalized relative copy number for each pufferfish *fabp* gene.

CHAPTER 3

Results and Discussion

3.1 Identification of pufferfish *fabp* genes and their gene structures

By searching the *Tetraodon nigroviridis* genome sequence in the Ensembl database using the orthologous Fabp sequences of zebrafish, medaka and three-spined stickleback as BLAST queries, the *fabp* genes of pufferfish were identified. A BLAST search of the NCBI database was performed with the same query sequences to identify *fabp* cDNA nucleotide sequences (Table 1) of pufferfish. By this approach, I identified *fabp1, fabp2, fabp3, fabp6a, fabp6b, fabp7a, fabp7b, fabp10, fabp11a and fabp11b* (Figure 1) in the genome of the pufferfish. The organization of introns and exons was determined in each *fabp* gene based on sequence alignments with the longest cDNA sequence identified in the NCBI database. All the pufferfish *fabp* genes have a conserved gene structure of four exons separated by three introns, and each gene codes for a polypeptide of 125-135 amino acid residues (Table 3). The alignment between the genomic sequence of the pufferfish *fabp1* and a cDNA sequence revealed many undetermined nucleotides in the genomic and cDNA sequence of pufferfish *fabp1*, therefore, a DNA fragment containing exon 1 and parts of intron 1 was PCR amplified and sequenced to clarify the sequence and exon-intron organization of pufferfish *fabp1* (supplementary material). Hence the putative transcription start site could not be located for *fabp1*. A putative TATA box, 3ˈ untranslated region, transcription initiation site, polyadenylation signal and the exon-intron junction, which conforms to the GT/AG rule (Breathnach and Chambon, 1981), were identified in each pufferfish *fabp* gene, except for *fabp6b*, where a putative TATA box could not be found.

Figure 1: Nucleotide sequence of pufferfish *fabp* genes.

A) *fabp1*, **B)** *fabp2*, **C)** *fabp3*, **D)** *fabp6a*, **E)** *fabp6b*, **F)** *fabp7a* [reproduced from Parmar and Wright, *in press*], **G)** *fabp7b* [reproduced from Parmar and Wright, *in press*] **H)** *fabp10* [reproduced from Parmar and Wright, *in press*], **I)** *fabp11a*, **J)** *fabp 11b* were identified. Exons are shown in capital letters, where the coding region is in bold font. The corresponding amino acid sequences are indicated below the exons. Introns are represented in lowercase letters. The number scale on the right indicates the position of the nucleotide and +1 indicates the putative transcription initiation site. A putative TATA box in the 5′ upstream region of each gene is indicated by bold letter and underlining. A putative polyadenylation signal is shown in bold letter and underlining in the 3′ untranslated region of each gene. Primers used for RT-qPCR analysis are highlighted in green and yellow corresponding to the forward and the reverse primers, respectively. Primers used to confirm the sequence structure for *fabp1* is highlighted in blue and pink corresponding to the forward and reverse primers, respectively.

Figure1A

fabp1

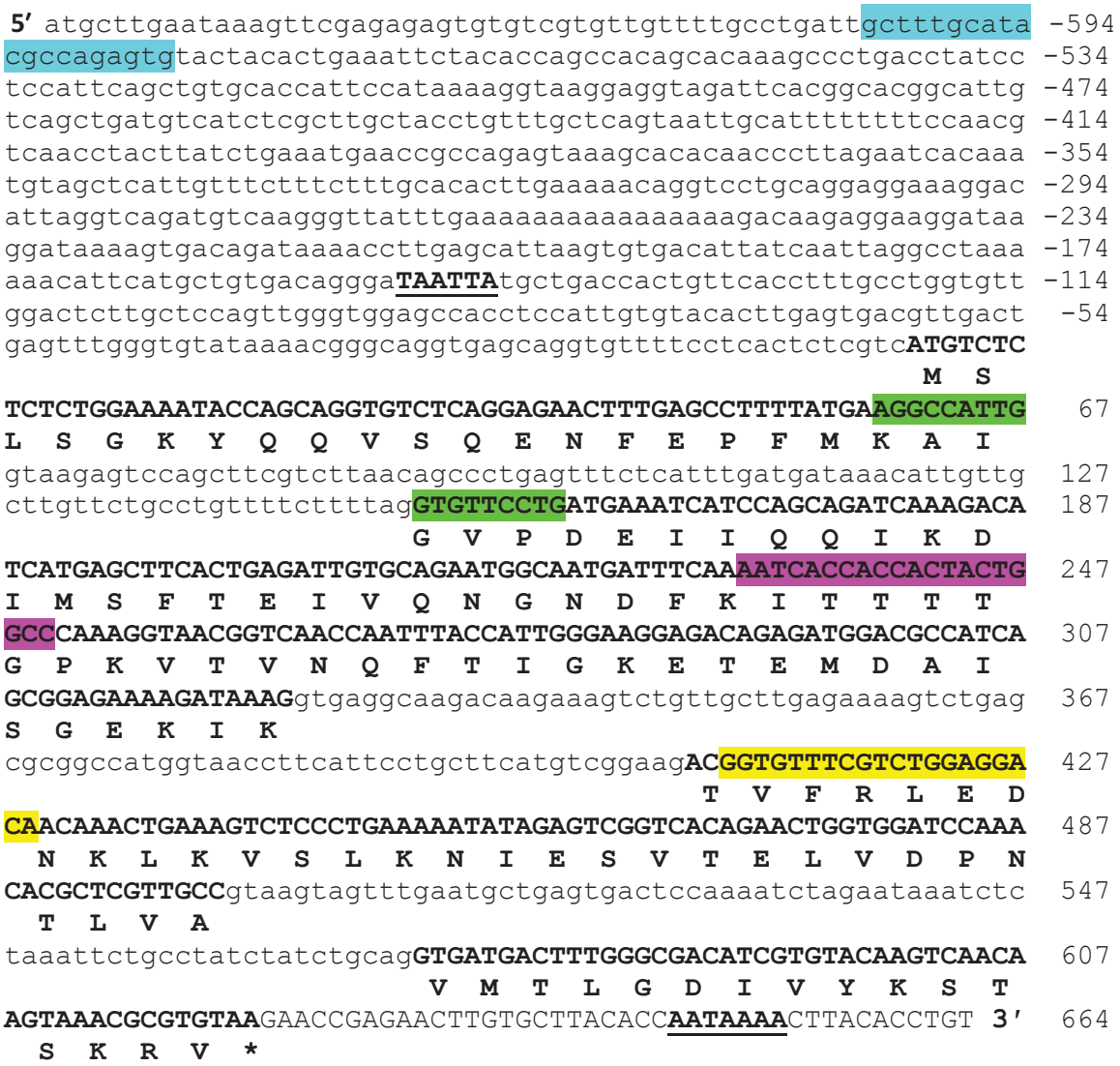


Figure 1B

fabp2

```
5'ccgatcctcggcccggcggacccttggagattgagaggttctgatagggcttctaaag  -74
gccacatcagcactcgggagatgagataaagaggtgggctgagcTTTAAAAGCAGCCGCG  -16
              +1
gcgggagagggagacACTCCTCCCTGCAGTCGTCCCGTCGGCCCTCTCGGAACCATGACT   45
                                                      M  T
TTCAACGGAACCTGGAAAGTTGATCGCAATGACAACTATGAGAAGTTCATGGAACAGATG  105
 F  N  G  T  W  K  V  D  R  N  D  N  Y  E  K  F  M  E  Q  M
Ggtgagtgactgggacggctgcagacgggcgaggcccatgcagcagctgcatctgcagcg  165
tgaaggtctcagcccgtctcctggagctccgcggacctgctgcatcacccagttccctct  225
ttcctttctctctgtgcagGAATCAACATGGTCAAGAGGAAGCTGGCTGCTCACGACAAC  285
                     G  I  N  M  V  K  R  K  L  A  A  H  D  N
CTGAAAATAACCATCCAACAGGATGGAAACACCTTCCACATCAAGGAGAGCAGCAATTTC  345
 L  K  I  T  I  Q  Q  D  G  N  T  F  H  I  K  E  S  S  N  F
CGCAACCTGGAAATGGACTTCACCCTGGGCGTCACCTTCGACTACAGCCTGGCAGACGGA  405
 R  N  L  E  M  D  F  T  L  G  V  T  F  D  Y  S  L  A  D  G
ACAGAGCTATCGgtgagtaaacgtggccggaaacctgaaaccgcctcgttggtcgcagtc  465
  T  E  L  S
ggaagcggtaccaccatgagcaagcgtcgcttgtggcgctgcagGGTTCATGGAGCATGG  525
                                             G  S  W  S  M
AGGGGGAGACCATGAAGGGGGTGTTCGTCCGGAAGGACAACGGAAAACAACTGACAACCA  585
E  G  E  T  M  K  G  V  F  V  R  K  D  N  G  K  Q  L  T  T
TCAGAACTCTCGAAGGCGGCGAACTCGTCCAGgtaaacgcaacagtcagacatatgcagg  645
I  R  T  L  E  G  G  E  L  V  Q
aatccggctgtgaacgtgcacgttttcatactaaaatacagaaactgcagcgtgaggtc  705
gcggatgtaacacaccactctggccttctgtcctcagAGCTACAACTACGAAGGCGTGGA  765
                                        S  Y  N  Y  E  G  V  D
CGCTAAGAGGATTTTCAAGAGGAGCTAGAGCCAAGCGGAGCCTGGGAGCGCCCGCCTGGC  825
  A  K  R  I  F  K  R  S  *
CCCTCGTTACTGCCTGTACAGTCAGCACACCCCTTTTTCATAAAGCAATAAAGCCCTTTG  885
TCAAAGCGTG 3'                                                  895
```

Figure 1C

fabp3

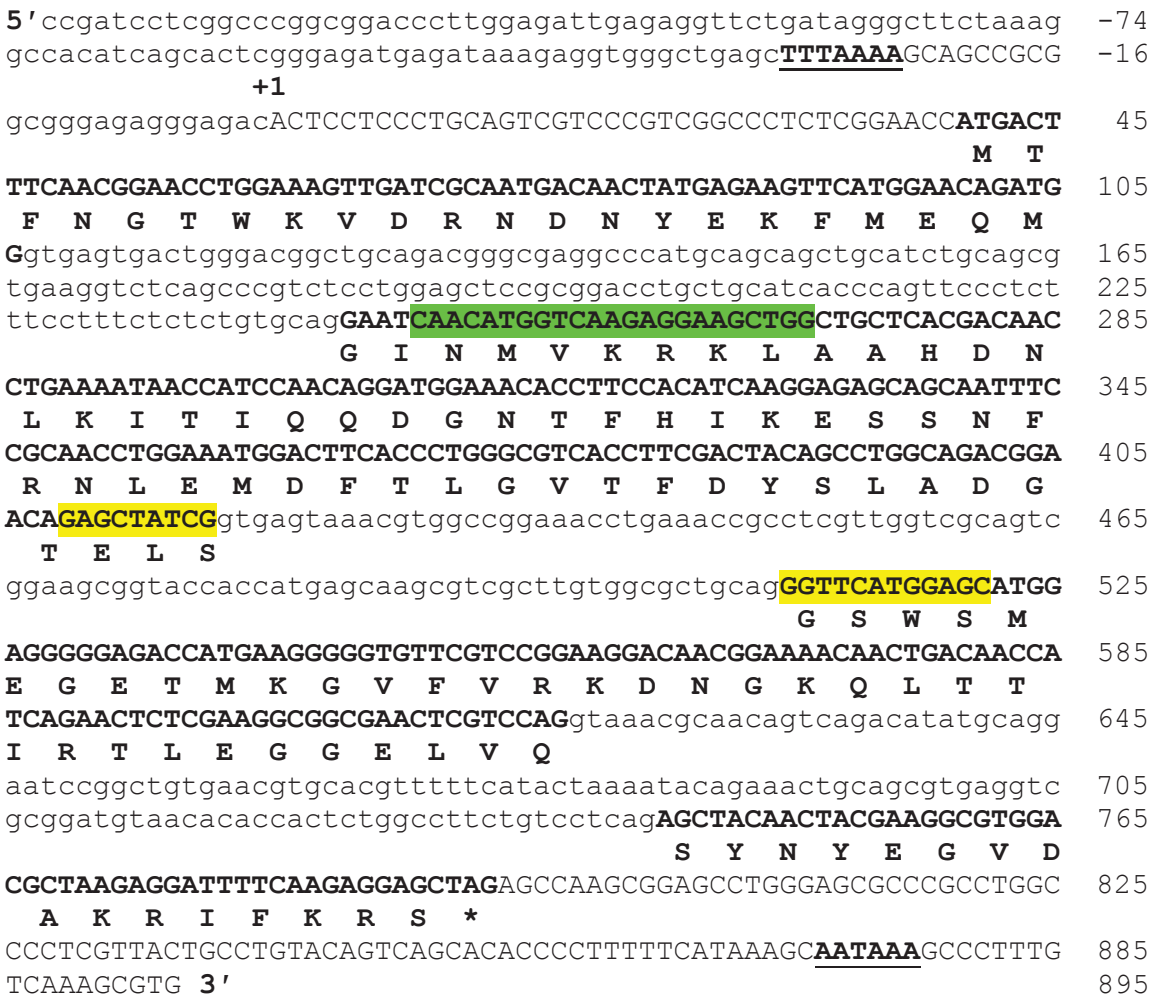


Figure 1D

fabp6a

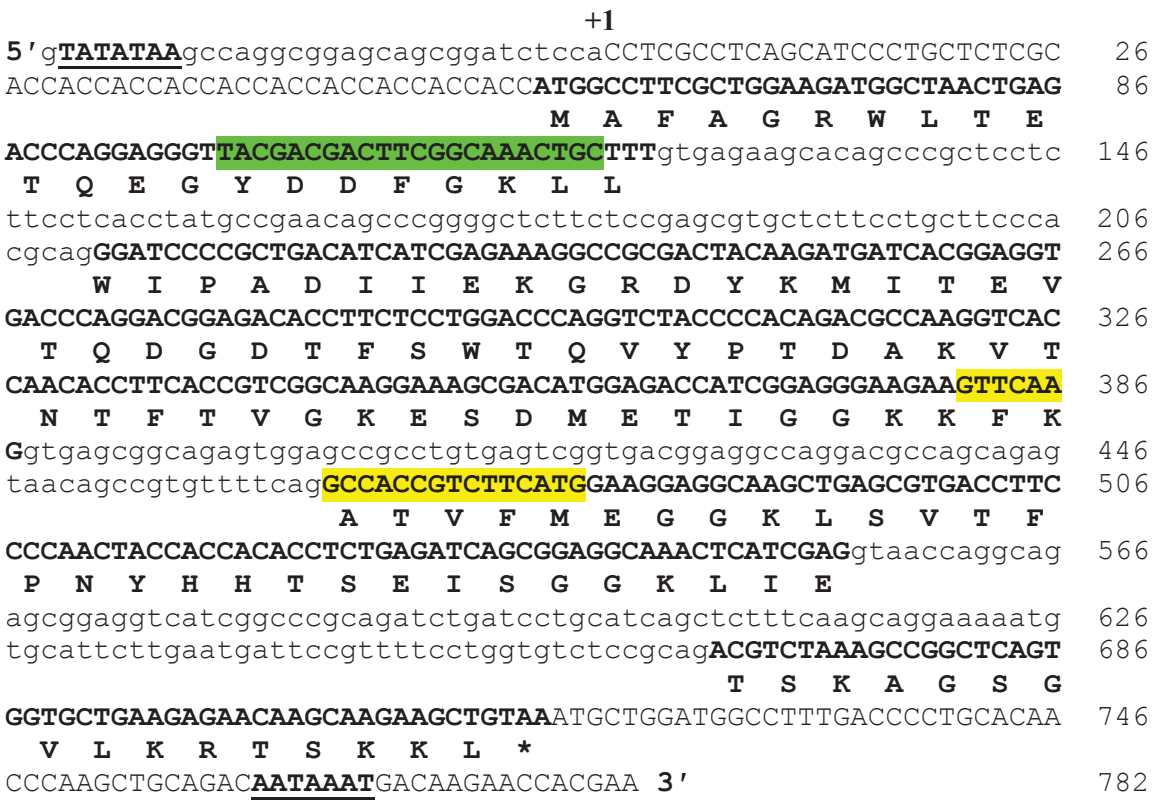


Figure 1E

fabp6b

```
5'ttccagggtgagagcgagcaaacctggaaacagcagcccggcgggtcgatgcggaaca  -99
tcgacgcctcatttctgcgcaacagaaggaagcagcggatgtgggcgtcagcctgagcca  -51
                                                    +1
ctcctttcttcttggcaaaaagtcgttccttgtctcgttctctGTTCTTTTCCACCATGA   10
                                                          M
CTTTCACTGGGAAATACGAGCTGGAGACTCAGGAAAACTACTGTGAGTTTCTGGATGCCA   70
T  F  T  G  K  Y  E  L  E  T  Q  E  N  Y  C  E  F  L  D  A
TCGgtacacacacactttcttcttctcctacgtgtcacaaaaggatggatgtgatttcct  130
I
ttcgctgcggcagGACTTCTCAGCGCCAAGACGGACCACAAAGTGATAACGGAGGTGAGC  190
             G  L  L  S  A  K  T  D  H  K  V  I  T  E  V  S
CAGGACGGGAACGACTTCACCTGGACTCAAAGCATTCCCCACTGGACGTGGACAAACACC  250
 Q  D  G  N  D  F  T  W  T  Q  S  I  P  H  W  T  W  T  N  T
TTCAGCGTCGGTCGGGAATGTGAGCTGGTGACCATGAAGGGCACCAAGTTCAAGgtgagc  310
 F  S  V  G  R  E  C  E  L  V  T  M  K  G  T  K  F  K
aggcggaaagaccagccagcatccctgatgtctggaggcttccatccctttccgtgcag  370
tcggggttgaacgtgtgtctctgacccgcagGCGCCTGTCGTGATGGACGATGGGAGGAT  430
                                A  P  V  V  M  D  D  G  R  I
CTCAGTCCAGTTTCCCCAGTACCATTTCACAGCCGAGGTCAGCGGGGACAAGCTGATCAT  490
  S  V  Q  F  P  Q  Y  H  F  T  A  E  V  S  G  D  K  L  I  M
Ggtaaaagaaccagtcggagcgctgaggtgggattttagtgggtttaacgtcacgtaaaa  550
gctccccttttctacttcgaaccctccagAATTGCACAATTCCAGGAGAGAAGGGTGTGA  610
                             N  C  T  I  P  G  E  K  G  V
CCTTCAGGAGAACCAGCAGGAGGGTCTGAGGCCGCAAGACTCAGCATGTCCTCATGGAAC  670
T  F  R  R  T  S  R  R  V  *
GCCGGGACATTTGGGAGAGTAGAAAAGTTCGCCTCTTCGTTGGGTTTGTGATAGTTAATA  730
AAGATTTAAAAGACCAACTGTGACCTTGCAGCGGTTGTGAAATAAAACGTCACCTGTGAC  790
AACTAGGAAA 3'                                                  800
```

Figure 1F

fabp7a

```
5'gccaagaataacatgattaaatgttacgattaagagcaatgacaacacgtacgctgat  121
aagcaaacggcgaccactcgaacccaggaatggcgtttaacggtcaccccgtcttccggc   61
cgaggcgctagccccgccccccgtggctgTATTAAtctgggggatggagacgcgggttcc    1
+1
ACTGCGGGGAGACCCGAGGCTGCGCGACGGCTGCTGACCCGGAGCCGAAGCTGCTGCAGA   60
CATGGTCGACGCTTTCTGCGCCACATGGAAGCTGGTCGACAGCCAGAACTTTGATGACTA  120
  M  V  D  A  F  C  A  T  W  K  L  V  D  S  Q  N  F  D  D  Y
CATGAAGGCGCTAGgtgagcaagttcacctctttgaattatgtgttgtctttagtcaatt  180
  M  K  A  L
Gcatttttatggctgcaaaattagggtgcacaattagtattgttcttttgttttgttttg  240
ttttcttatagttagattttaaatgttgatttaatgcatcatgaatggtactgacctctg  300
ttttatttttagGTGTTGGCTTTGCAACAAGGCAAGTCGGCAATGTCACCAAACCAACTGT  360
          G  V  G  F  A  T  R  Q  V  G  N  V  T  K  P  T  V
AGCCATTAGCAAGGATGGAGACAAAGTGGTGATAAAGACCATGAGCACCTTCAGGAACAC  420
  A  I  S  K  D  G  D  K  V  V  I  K  T  M  S  T  F  R  N  T
GGAGATCTCCGCCAAGCTGGGAGAGGAGTTTGATGAGACCACGCCTGATGATCGACATGT  480
  E  I  S  A  K  L  G  E  E  F  D  E  T  T  P  D  D  R  H  V
CAAAgtgagctgaataaatgtagaaataaaatatataactctgacatttgctccatccgc  540
  K
ttaaaatggcgtgaaattctcggtgtgtgttcttgcagTCAACTTTCTCTATGGAGGGAG  600
                                        S  T  F  S  M  E  G
ACAAGCTGATGCAGGTGCAGAAGTGGAACGGCAAGGAGACCAAATTTGTCAGAGAAATCA  660
D  K  L  M  Q  V  Q  K  W  N  G  K  E  T  K  F  V  R  E  I
AGGATGGAAAGATGGTGATGgtaagaagtgcaaagaaaaaaaatggtgcacagttgaagg  720
K  D  G  K  M  V  M
atgctgttcaaaccatattatctcactatatttacagtgtgcagtaacttatatttccaa  780
atgaaaacaatctgttttgctgcacatgtacaacattttatcgttttcccttctcttttg  840
acagACTTTGACTTTCGAAGGGGTCACGGCGGTTCGCACGTATGAGAAAGCCTAATCTCA  900
     T  L  T  F  E  G  V  T  A  V  R  T  Y  E  K  A  *
GCAAGTCATTCCAGCACTTCAACAATATCTTCAACGTTGTTTTACAATATGCAGTGGATC  960
TTTTTGTCAGATCTTTTTGAATTAAACGCCTAAAGTGTGACTATGTGTGAACAAAATTGT 1020
TACAAAAAAATAAAGTGCTTCAGTAATGTTTAAATGTG 3'                     1058
```

Figure 1G

fabp7b

```
5'aacgagctcatctagccgtctcccagccctctgcgggtcagtaggaccatgacccaat  121
ttagggatccaataaaaacctgaggaaaaatctcctgagccaatgggacctatggctccg   61
cccacccacctgcatactaatgaggaggagaaTAATAAcacactgggggggaaacttgagg    1
+1
GATCGCAGCTCGTCCAGCGTGCTCTCCAGATTACTCTGCTTCTAATCAACCGTCACCATG    60
                                                          M
GTCGACGCCTTCTGTGCAACCTGGAAACTGGTGGAGAGCGAGAACTTTGATGATTACATG   120
 V  D  A  F  C  A  T  W  K  L  V  E  S  E  N  F  D  D  Y  M
AAAGCGCTCGgtcagtgatgtgcttatgttgtttgctttgatctgttgttttaggtccga   180
  K  A  L
gtgctgaaatgtgctttatttgccatgtctcccagGCGTGGGTTTCGCCACCCGACAGGT   240
                                    G  V  G  F  A  T  R  Q  V
CGGTAACGTGACCAAACCGACTGTGATCATCAGCCAGGAGGGCGACAAGGTGGCGATCCG   300
  G  N  V  T  K  P  T  V  I  I  S  Q  E  G  D  K  V  A  I  R
CACACAGAGCACCTTCAAAAACACAGAGATCTCCTTCAAGCTGGGAGAGGAGTTCGATGA   360
  T  Q  S  T  F  K  N  T  E  I  S  F  K  L  G  E  E  F  D  E
AACCACCGCCGACGACAGGAACTGTAAAgtatgacgaaacgtcttcgatgttcccactaa   420
  T  T  A  D  D  R  N  C  K
accctaatgggaatctgttttgttcaggatgaattagtcggtagtttcgctaaggtcacc   480
atgtgtcttgtctctcagTCCACAGTGACTCTGGACGGAGACAAGTTGGTCCACGTCCAG   540
                     S  T  V  T  L  D  G  D  K  L  V  H  V  Q
AAGTGGGATGGCAAAGAGACCAAGTTTGTCAGAGAGATCAAGGATGGCAAACTGGTCATG   600
 K  W  D  G  K  E  T  K  F  V  R  E  I  K  D  G  K  L  V  M
gtaggattcaattaaaaagttgatctaaattatgacactgataaaattcggaagcattaa   660
aaatggaatcttggaatctggtctgtttttaaagtgcagattctatactctaatgaagag   720
atcattacctgattattatgatggggactccgctgtaaacaatgagctctaggtgtgatt   780
tctaggcagcagcaaaatcaacaagtgggctctttttcactcttccctttttcttctgcgt   840
caacacacccctcagatattgtgtctgtcgccatagtaacctttgtctctcctctcctgg   900
cagGACCTGACCTTTGAAGACGTCCACGCAGTGCGCACCTACGAGAAGGCATGAAGGGCC   960
    D  L  T  F  E  D  V  H  A  V  R  T  Y  E  K  A  *
CGACCTTTGCACTCCAGGAGGTTTCCTGCAGGCTGCACGTTTTTGTAAAAATAAAACATT  1020
TTTATACAAACTGTTTCTGTCATGATGCATCGTCTTTGCAGCCCCGCGTGAACGATTAAA  1080
AAAGTGAACAGGCTAAA 3'                                          1097
```

Figure 1H

fabp10

```
5'tgacaaatatcaccttacacttggtcgaaggcaccatttggccaaacgggagtacgtg  121
aacctgaggttgttcaaacaacagtgattacctcactgggatctattactcattagacag   61
agagctgggggccccgacctgtcctcTATAAGAAccaggacccctctgatcacatccttc    1
+1
ACATTGTGTTGAGCTTCACACAGCTGTCACAGCCTCCAGCCCATCTTGGTTAAGGAGAAG   60
CCAGATCTAAGAAGCATGGACTTCAACGGGACATGGCAGGTTTATGCCCAGGAGAACTAC  120
                M  D  F  N  G  T  W  Q  V  Y  A  Q  E  N  Y
GAGGAGTTCCTCAAGGCCATGGgtgagaagtaacaccaataccaggctgtcctattttta  180
 E  E  F  L  K  A  M
gttttctcaattaaataaacgtgaacatgtttctttgtaccagAACTGCCCGAGGATGTC  240
                                           E  L  P  E  D  V
ATCAAGATGGCCAAAGACATTAAGCCGATTACTGAGATCAAGCAAAATGGCAAAGACTTT  300
 I  K  M  A  K  D  I  K  P  I  T  E  I  K  Q  N  G  K  D  F
GTTGTTACCTCCAAAACGCCCGGGAAGACTTTGACCAACTCCTTTACCATTGGCAAAGAG  360
 V  V  T  S  K  T  P  G  K  T  L  T  N  S  F  T  I  G  K  E
GCTGATATAACCACCATGGACGGCAAGAAGCTCAAGgttcgctggcccctccggggtgat  420
 A  D  I  T  T  M  D  G  K  K  L  K
tctgctttggtcacgttctcattttgatacgatcctcttgtgtcttatcttcacaaacac  480
acagTGCACTGTCACTATGGAGGGGAGCAAACTGGTCTGCAAAACGGAGAAGTTCTGTCA  540
     C  T  V  T  M  E  G  S  K  L  V  C  K  T  E  K  F  C  H
CACACAGGAGCTCAAAGGGGGAGAGATGATTGAGgtacggtggacatctttatatcttaa  600
  T  Q  E  L  K  G  G  E  M  I  E
gatactggttcgttgaaatgtttttgaaagaccacacagacggacgctgacaatcgactg  660
attcattttcagACTTTGACCAAGGGCTCAACAACTTTTGTCAGAAAGAGTAAGAAGATG  720
               T  L  T  K  G  S  T  T  F  V  R  K  S  K  K  M
AACTAAGACAAAGGAAAAACAAGGTTGAATAAAATAAATAAAAGTGTTTCCCGACAGGCG  780
 N  *
TGGTTTGAGTTAATTCTTTACAGTTCACCTATCGGTCTGGGCTAAAGGTCAAAGCTGCGA  840
GGACGAGGCCGTGCCGATGCTATTGCCTAAGGCTATCAATAAAATGATGAGCTGTAATTA  900
TCAACTGAGGAAAAAAAAACACAGACAAGAAGTGGGGGGAAAGGGAAGGGATTTATTTTT  960
CTTTTTTTTTAAATTTAGACATGAATTTTAGAGTGCATTTGCCTTTTTTTTAATAAAAAC 1020
ACCAATTTTAACAAATCTTGAAA 3'                                    1043
```

Figure 1I

fabp11a

```
                                                                +1
5'tgggcgcgaggccgccgcttTAAAAggcagcaggttcccacagcacagAGACCGCGTC    10
CACGCACTCCCCACGGGCTGACCAGGGTCGACGCGCCGGGCTGCAAACATGGTTGAGAAG    70
                                                    M  V  E  K
TTTGTTGGGACGTGGAAAATGATTTCCAGCGAGAACTTTGACGATTACTTGAAAGCAATT   130
 F  V  G  T  W  K  M  I  S  S  E  N  F  D  D  Y  L  K  A  I
Ggtacgagagcaaaaagagggataacgggatgagtatcactgaggcaacagtgatttata   190
gtaccacagtaggaaaacgcatttagaaacattttctgcttcctgatgattgtttattag   250
atttatccatgctttacctttattctgcagGTGTTGGGTTCGCGACCCGGCAGGTGGGCA   310
                                 G  V  G  F  A  T  R  Q  V  G
ATCGGACCAAGCCCAATCTGGTGTTCACCGTGGACGACCAGGGGACGGTGTGCATGAGGA   370
N  R  T  K  P  N  L  V  F  T  V  D  D  Q  G  T  V  C  M  R
CTCAGAGCACCTTCAAAAACACCGAGATCAAGTTCAAACTCAACGAGCCGTTCGACGAGA   430
T  Q  S  T  F  K  N  T  E  I  K  F  K  L  N  E  P  F  D  E
CCACCGCCGACGACAGGAAGACGAGGgttagtgcgggccgtgaacgcatcacggcccgtc   490
T  T  A  D  D  R  K  T  R
cgcggcgtttgggcgcgcggggctctcacggcttttctgtccccgcagACCGTGGTGAC   550
                                                T  V  V  T
TCTGGAGAACGGCAAGCTTGTGCAGAAACAGAACTGGGACGGCAAAGAAACCTGCATCGA   610
  L  E  N  G  K  L  V  Q  K  Q  N  W  D  G  K  E  T  C  I  E
AAGGGAGATCTCGGACGGGAAGTTGATAGCGgtgagacgttcctgctttccccacttta   670
  R  E  I  S  D  G  K  L  I  A
ttgagtctagttgtgcgtccggttttagtggctcttaaactcattaatgcagtctgcga   730
atgtgtttgtgacagaaaacatgaaactgaacccctaaaatgctaattaataacagcagc   790
agcagttattaccgcttacgggtgaaatccgcatatttgcatgtattgattgatcgatca   850
gtctgtttttatctgtgtttttttcttccccccgcagAAATGCATCATGGGAGATGTGGTT   910
                                        K  C  I  M  G  D  V  V
GCTGTCAGGACATATGTGAGAGAGGCATGAGGAGGGTGTGAGGAAGGTCTCCTCTCTGGA   970
 A  V  R  T  Y  V  R  E  A  *
CAGGCTCAGTTCAATGAGCGTCGGATACCAAACTGTTCTTTCCTGTCTTCCTCTTTTCTT  1030
TACTGTGTTTCTACAGCTGCAGTCCTGTACTCCAGACCCCACACCAGCAACCTTCTCCTT  1090
TTTTTTGATGATTGGAATAAAAATGAATTTATCCACATGG 3'                   1130
```

Figure 1J

fabp11b

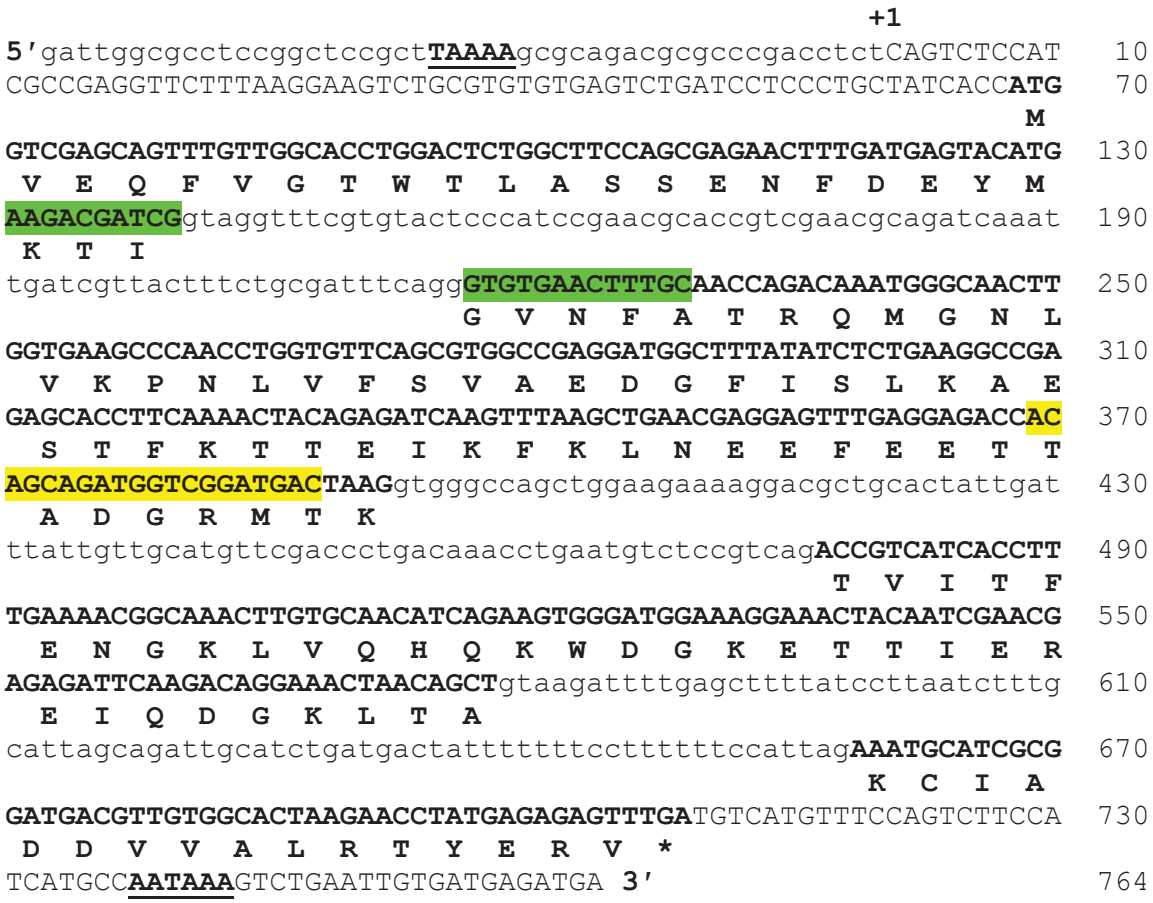


Table 3: Fatty acid-binding protein genes of pufferfish with their corresponding Ensembl gene ID, GenBank accession numbers of their cDNA sequences, location and number of amino acid.

Gene	Ensembl ID	GenBank ID	Location	Number of amino acids
fabp1	ENSTNIG00000016680	CR720131.2	Chromosome 12: 9,154,751-9,155,415	127
fabp2	ENSTNIG00000012360	CR734529.2	unassigned_random: 14,883,076-14,884,115	132
fabp3	ENSTNIG00000016979	CR637284.2	Chromosome 21: 1,218,101-1,220,308	133
fabp6a	ENSTNIG00000017612	CR733593.2	Unassigned_random: 20,198,312-20,199,077	127
fabp6b	ENSTNIG00000004652	CR699217.2	Chromosome 1: 1,118,490-1,119,289	125
fabp7a	ENSTNIG00000017217	CR705358.2	Chromosome 10: 8,694,895-8,695,952	132
fabp7b	ENSTNIG00000003393	CR683239.2	Chromosome 14: 9,433,632-9,434,728	132
fabp10	ENSTNIG00000002424	CR639897.2	Unassigned_random: 52,169,316-52,170,358	127
fabp11a	ENSTNIG00000017781	CR733066.2	Chromosome 21: 1,566,398-1,567,527	134
fabp11b	ENSTNIG00000018400	CR671793.2	Chromosome 8: 7,093,697-7,094,465	134

However, all *Fabp* transcripts were detected in the RT-PCR assay, using primers specific for each pufferfish *fabp* gene indicating that all the *fabp* genes identified are functional and the presence of a possible TATA-less promoter in the *fabp6b*.

3.2 The sequence identities and similarities of pufferfish *fabp* genes with teleost fish and human *fabp*/*FABP* genes

The sequence identity and similarity showed that pufferfish Fabps and other Fabps from other vertebrates shared 45-95% sequence identity (Figure 2). A default cut-off <40% identity was set while aligning orthologous sequences (Thompson *et al*., 1994). Meanwhile, the percentage identity decreased significantly (15%-40%) when aligned with other paralogs of pufferfish Fabps. However, the pufferfish fish Fabps shared 50-95% sequence identity with other teleost fish Fabps, suggesting that the gene coding for these proteins are orthologous and most likely arose from a common ancestral gene.

To infer the evolutionary relationship between the pufferfish Fabps and Fabps from other vertebrates, neighbor-joining and maximum-likelihood phylogenetic trees were constructed using MEGA5 software (Figure 3). Human lipocalin 1 (LCN1) protein was used as a designated outgroup. Both trees generated identical topologies indicating similar phylogenetic relationship, where each of the ten Fabps of pufferfish clustered in distinct clades with their orthologs of other teleost and tetrapod Fabps. Based on the robust bootstrap values >80%, the *fabp* genes of pufferfish appear to be orthologs of other teleost and tetrapod *fabp* genes.

Figure 2: Amino acid sequence alignments of various orthologous and paralogous sequences showing percentage identity and similarity to pufferfish Fabp.

Amino acid sequence of pufferfish (Pf) **A)** Fabp1, **B)** Fabp2, **C)** Fabp3, **D)** Fabp6a, **E)** Fabp6b, **F)** Fabp11a and **G)** Fabp11b aligned with Fabp/FABP of zebrafish (Zf), Medaka (Me), three-spined stickleback (St), fugu (Fu), human (Hu), mouse (Mo) and rat (Rt). The percentage of sequence identity and similarity are show at the end of each sequence. The dots in the alignment represent sequence identity, whereas the dashes represent missing amino acids.

Figure 2A

Fabp1

```
                    10        20        30        40        50        60
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp1        M--SLSGKYQQVSQENFEPFMKAIGVPDEIIQQIKDIMSFTEIVQ-NGNDFKITTTTGPK
St Fabp1        .--.FT....LET.....A....V.L..DL..KG...K.IS..EE-..DN..V.V...S.
Me Fabp1        .--.FT....LE.S..Y.A....V.IS..N..LT.GLK.I..MEE-..DN.RV.V...SH
Hu FABP1        .--.F.....LQ......A......L.E.L..KG...KGVS....-..KH..F.I.A.S.
Mo FABP1        .--NF.....LQ.............L.EDL..KG...KGVS...H-E.KKI.L.I.Y...
Rt FABP1        .--NF.....VQ...........M.L.EDL..KG...KGVS...H-E.KKV.L.I.Y.S.
Zf Fabp1b.1     .--.FT....LE...G.VE....V.L..DM.EKG...K.VS..EE-...Q..V.V...S.
Zf Fabp1a       .--AFT....LE.H....A....V....DEVEKG...K.IS..H.-D.K...V.V.A.T.
Zf Fabp1b.2     .--PF...FELEKY.GW.E......-HQ.LLKKAN.GKTQF..QE-....YCL..RS.G.
Pf Fabp11a      .ADAFI.TWNLKESK..DEY..G...GFATR.VANMTKPT.I.SK-E.DV.TLK.VSTF.
Pf Fabp3        .VDKFV.TWKMTTSD..DEY......GFATR.VGNRTKPNLVVCVDEQGLICMKSQSTF.
Zf Fabp2        .--TFN.TWKVDRN..Y.K..EQM..NMVKRKLAAHDNLKITLE.-T.DK.NVKEVSTFR

                    70        80        90       100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp1        VTVNQGTIGKETEMDAISGEKIKTVFRLED-NKLKVSLKN----IESVTELVDPNTLVAV
St Fabp1        .L..TFK..Q.A.IET.T...V.A.VKRDG-...T.C..G----M......L.K..I.NT
Me Fabp1        .II.SF...Q....LSVT.......VKK.G-T..MM...R----ADI......G....TT
Hu FABP1        .IQ.EF.V.E.C.LETMT...V...VQ..GD...VTTF..----.K.....-NGDIITNT
Mo FABP1        .VR.EF.L.E.C.LETMT...V.A.VK..GD..MVTTF.G----.K.....-NGD.ITNT
Rt FABP1        .IH.EF.L.E.C.LETMT...V.A.VKM.GD..MVTTF.G----.K....F-NGD.ITNT
Zf Fabp1b.1     .LT.SF...Q.ADIETLT...V...VHKDG-Q...AA.G.----.TLTW..L.E...LIT
Zf Fabp1a       .ILYSF.V.E.C.LETFT.DRA...VQMDG-...TAFV.G----.......-.GD.ISNT
Zf Fabp1b.2     .LN.SF...QD..IQML..D.V...VHKDG-Q...AA.G.----.TLTW..L.E...LIT
Pf Fabp11a      S.EINFKL.E.FDETTADDR.V.S.IT.DGGKL.H.QKWD--GKETTLLRE.SD.N.TLT
Pf Fabp3        T.EIKFKLNEPF.ETTADDR.TT..MTI.NGKLVQKQTWD--GKESTIERE.SDGK.I.K
Zf Fabp2        TLEINF.L.VTFDYSLAD.TELTGSWVI.GDTLKGTFTRKDNGKVLTTVRTIVNGE..QS

                   130
                ....|....|....|..
Pf Fabp1        MTLGD-IVYKSTSKRV-
St Fabp1        I...S-....R....M-
Me Fabp1        .....-....T.Y..I-
Hu FABP1        .....-..F.RI...I-
Mo FABP1        .....-....RV...I-
Rt FABP1        .....-....RV...I-
Zf Fabp1b.1     LSTA.N.A..RF....V
Zf Fabp1a       LSFNG-....RI.R.IS
Zf Fabp1b.2     LSTA.N.A..RF....V
Pf Fabp11a      L....-..STRHYVKAE
Pf Fabp3        CKM..-V.AVR.YVKEA
Zf Fabp2        YSYDG-VEA.RIF..A-
```

	Identity (%)	Similarity (%)
Pf Fabp1	100	100
St Fabp1	60	77
Me Fabp1	58	74
Hu FABP1	58	74
Mo FABP1	55	73
Rt FABP1	51	72
Zf Fabp1b.1	51	70
Zf Fabp1a	48	70
Zf Fabp1b.2	48	63
Pf Fabp11a	23	35
Pf Fabp3	22	38
Zf Fabp2	16	35

Figure 2B

Fabp2

```
                  10        20        30        40        50        60
          ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp2  M--TFNGTWKVDRNDNYEKFMEQMGINMVKRKLAAHDNLKITIQQDGNTFHIKESSN-FR
St Fabp2a .--..D....I......D.............................N..K.AV.....-..
Me Fabp2a .--..........SE..........V.V................E.T.DK..V.....-..
Zf Fabp2  .--...........E..........V.................LE.T.DK.NV..V.T-..
St Fabp2b .--..............D.......V.IM.....E.....V..E.T.DQ.Q.....T-..
Me Fabp2b .--A......A..SE..D..L.RI.V.VM.....E........E.T.DN..V....T-..
Hu FABP2  .--A.DS......SE..D....K..V.I............L..T.E..K.TV....A-..
Mo FABP2  .--A.D........E.......K....VM....G......L..T....K.TV.....-..
Rt FABP2  .--A.D........E.......K....V.....G......L..T.E..K.TV.....-..
Zf Fabp6  .--A...K.ETESQEG..P.CKLI..PDDVIAKGRDFK.VTE.V.N.DD.TWTQYYPNNH
Pf Fabp10 .--D.....Q.YAQE...E.LKA.ELPEDVI.M.KDIKPITE.K.N.KD.VVTSKTP-GK
Pf Fabp11a .VEK.V....MISSE.FDDY.KAI.VGFAT.QVGNRTKPNLVFTV.DQGTVCMRTQST.K

                  70        80        90       100       110       120
          ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp2  NLEMDFTLGVTFDYSLADGTELSGSWSMEGESMKGVFVRKDNGKQLTTIRTLEGGELVQS
St Fabp2a ...IN....D..E.............N...DML..I.N..........T.IVQRD..I..
Me Fabp2a .I.I........E...........A.AL..DVL..T.T......A...T.VIQ.D.MI..
Zf Fabp2  T..IN.................T...VI..DTL..T.T......V...V..IVN......
St Fabp2b TKDI......P..........V..T.E...DLL..K.T....S.V...T.A.V.......
Me Fabp2b TK.F......K..........V..T.E...DML..L.N....N.L...T.SIVN.....T
Hu FABP2  .I.VV.E.....N.N.......R.T..L..NKLI.K.K.T...NE.N.V.EII.D....T
Mo FABP2  .IDVV.E...N.P.........T.A.TI..NKLI.K.T.V....E.IAV.EVS.N..I.T
Rt FABP2  .IDVV.E...D.A.I.GNCCHFT.T.T...NKLV.K.K.V....E.IAV.EIS.N..I.T
Zf Fabp6  VVTNK.IV.KES.METVG.KKFK.IV....GKLTIS.PK------YQQTTEIS..K..ET
Pf Fabp10 T.TNS..I.KEA.ITTM..KK.KCTVT...SKLVCKTEK------FCHTQE.K...MIET
Pf Fabp11a .T.IK.K.NEP..ETT..DRKTRTVVTL.NGKLVQKQNWD--..ETCIE.EISD.K.IAK

                  130       140          Identity (%)    Similarity (%)
          ....|....|....|....|
Pf Fabp2  YNYEGVDAKRIFKRS-----              100             100
St Fabp2a .............KC-----               80              92
Me Fabp2a ...D..........G-----               77              92
Zf Fabp2  .S.D..E.......A-----               76              91
St Fabp2b .............KQ-----               75              86
Me Fabp2b .............KH-----               69              85
Hu FABP2  .V....E......KD-----               66              83
Mo FABP2  .T....E...F..KE-----               66              82
Rt FABP2  .T....E......KE-----               62              78
Zf Fabp6  STAS.AQGTAVLV.TSKKV-               22              36
Pf Fabp10 LTK----GSTT.V.KSKKMN               22              43
Pf Fabp11a CIMGD.V.V.TYV.EA----              21              40
```

Figure 2C

Fabp3

```
                  10        20        30        40        50        60
          ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp3  MAEAFAGTWNLVKSEKFDEYMKELGVGLAMRKMGNLAKPTLSITIEGDKVTLKNSSTFKT
Fu Fabp3  .....V.....IS.....D........V.L..I.GM...STI.S.D....V..T.....N
Me Fabp3  .....V....MKE.........A....F.T..V...T...TI.SMD..R..V.TQ..I.N
St Fabp3  KVN..V.K.E.KE.V...D...K....F.T.Q....T...TI.SV...V..V.TQ.SL.N
Zf Fabp3  ..D..I.....KE.KN......GI...F.T.QVA.MT...TI.SK...VF...TV....S
Rt FABP3  ..D..V...K..D.KN..D...S....F.T.QVASMT...TI.EKN..TI.I.TH....N
Hu FABP3  .VD..L...K..D.KN..D...S....F.T.QVASMT...TI.EKN..IL...TH....N
Mo FABP3  ..D..V...K..D.KN..D...S....F.T.QVASMT...TI.EKN..TI.I.TQ....N
Zf Fabp2  .--T.N...KVDRN.NYEKF.EQM..NMVK..LAAHDNLKITLEQT...FNV.EV...R.
Pf Fabp10 .--D.N...QVYAQ.NYE.FL.AMELPEDVI..AKDI..ITE.KQN.KDFVVTSKTPG..
Pf Fabp1  .--SLS.KYQQ.SQ.N.EPF..AI..PDEIIQQIKDIMSFTE.VQN.NDFKITTTTGP.V

                  70        80        90       100       110       120
          ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp3  TEVSFKLGEEFDESTADGRNVKSVVTVED--GKLVHVQKWDGKETSLVREVEG-NNLTLT
Fu Fabp3  ..I.....................T.NIV.--..M..A....D...T.....ND-KS....
Me Fabp3  ..LT.........T...D.K...I..L..--.....................N.-......
St Fabp3  ..L....D.....T...E.K...F.K.S.D-...................S...-.A....
Zf Fabp3  ..IN.........T...D.K....I.LDG--...L...........T.L...SD-......
Rt FABP3  ..I..Q..V....V...D.K......LDG--............Q..T.T..LSD-GK.I..
Hu FABP3  ..I.....V....T...D.K...I..LDG--.....L.....Q..T....LID-GK.I..
Mo FABP3  ..IN.Q..I....V...D.K...L..LDG--...I.....N.Q..T.T..LVD-GK.I..
Zf Fabp2  L.IN.T..VT..Y.L...TELTGSWVI.GDTL.GTFTR.DN..VLTT..TIVN-GE.VQS
Pf Fabp10 LTN..TI.K.A.IT.M..KKL.CT..M.G------SKLVCKTEKFCHTQ.LK.-GEMIE.
Pf Fabp1  .VNQGTI.K.TEMDAIS.EKI.T.FRL..------NKL.VSL.NIES.T.LVDP.T.VAV

                  130                Identity (%)      Similarity (%)
          ....|....|....|.
Pf Fabp3  LTLGNVVCTRHYEKVQ              100               100
Fu Fabp3  ....K.........AE                75                86
Me Fabp3  ..M.D.....R...AE                75                87
St Fabp3  ..MDD..S....V.AE                67                78
Zf Fabp3  ....DI.S....V.AE                66                82
Rt FABP3  ..H....S..T...EA                62                74
Hu FABP3  ..H.TA....T...EA                61                74
Mo FABP3  ..H.S..S..T...EA                58                74
Zf Fabp2  YSYDG.EAK.IFKRA-                28                50
Pf Fabp10 ..K.STTFV.KSK.MN                26                45
Pf Fabp1  M...DI.YKSTSKR.-                22                41
```

Figure 1D

Fabp6a

```
                  10        20        30        40        50        60
            ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp6a   M--AFAGRWLTETQEGYDDFCKLLGIPADIIEKGRDYKMITEVTQDGDTFSWTQVYPTDA
Fu Fabp6a   .--...T...E..N.....E........D.......N..L........E......I...N.
St Fabp6a   .--....K.E..S.....A........N.V........L.........N...C.I...N.
Me Fabp6a   .--T.....EI.S.....E...VV.....V........AV.....N..E.T.......-I
Zf Fabp6    .--..N.K.E..S....EP....I...D.V.A....F.LV..IV.N..D.T...Y..NNH
Hu FABP6    .--..T.KFEM.SEKN..E.M.....SS.V...A.NF.IV...Q...QD.T.S.H.SGGH
Mo FABP6    .--..S.KYEF.SEKN..E.M.R..L.G.V..R..NF.I....Q...QD.T.S.S.SGGN
Rt FABP6    .--..T.KYEF.SEKN..E.M.R..L.E.V..R..NF.I....Q...EN.T.S.S.SGGN
Pf Fabp6b   .--T.T.KYEL....N.EE.LDAI.TA.GLLSAKT.H.V....S...ND.T...SI.H-W
Pf Fabp7a   .VD..CAT.KLVDSQNF..YM.A..VGFATRQV.NVT.PTVAISK...KVVIKT-MS.FR
Pf Fabp2    .--T.N.T.KVDRNDN.EK.MEQM..NMVKRKLAAHDNLKITIQ...N..HIKE-SSNFR

                  70        80        90       100       110       120
            ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp6a   KVTNTFTVGKESDMETIGGKKFKATVFMEGGKLSVTFPNYHHTSEIS------GGKLIET
Fu Fabp6a   .......I...C..............HL.....................------.......
St Fabp6a   R...N..I...C..............H......TTA....Q..T...------.......
Me Fabp6a   ....K..I....E..SL.........NL....MT.N....TQ.N...------.......
Zf Fabp6    V...K.I.........V......GI.S......TIS..K.QQ.T...------....V..
Hu FABP6    TM..K.......NIQ.M...T.....Q......V.N.....Q....V------.D..V.V
Mo FABP6    IMS.K..I...CE.Q.M.........K.....VVAE.....Q...VV------.D..V.I
Rt FABP6    IMS.K..I...CE.Q.M.........K.....VVAD.....Q...VV------.D..V.I
Pf Fabp6b   TW....S..R.CELV.MK.T....P.V.DD.RI..Q..Q..F.A.V.------.D...MN
Pf Fabp7a   NTEISAKL.E.F.ET.PDDRHV.S.FS...D..MQVQKWNGKETKFVR--EIKD..MVM.
Pf Fabp2    NLEMD..L.VTF.YSLAD.TELSGSWS...ESMKGV.VRKDNGKQLTTIRTLE..E.VQS

                  130       140         Identity (%)      Similarity (%)
            ....|....|....|....|
Pf Fabp6a   SKA----GSVVLKRTSKKL-            100                100
Fu Fabp6a   ...----.........R..-             88                 95
St Fabp6a   .T.----A.........I-              81                 89
Me Fabp6a   C..----.....RKV...I-             74                 88
Zf Fabp6    .T.SGAQ.TA..V.....V-             64                 80
Hu FABP6    .TI----.G.TYE.V..R.A             57                 75
Mo FABP6    .TI----.D.TYE.V..R.A             54                 75
Rt FABP6    .TI----.D.TYE.V..RVA             53                 75
Pf Fabp6b   CTIPG-EKG.TFR...RRV-             44                 64
Pf Fabp7a   LTF----EG.TAV..YE.A-             25                 37
Pf Fabp2    YNY----EG.DA..IF.RS-             21                 37
```

Figure 2E

Fabp6b

```
                10        20        30        40        50        60
           ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp6b  M--TFTGKYELETQENYEEFLDAIGT-AAGLLSAKTDHKVITEVSQDGNDFTWTQSIP-H
Fu Fabp6b  .--..................E....QIT...G...........V.N........T..-N
St Fabp6b  .--A.A......S........A..------...N...........L...SG........-N
Me Fabp6b  .--K.A......S.S......E..-----D.VN...E...V...V.............-N
Rt FABP6   .--A......F.SEK..D..MKRL.-LPEDVIERGRNF.I....Q...EN...S..YSGG
Mo FABP6   .--A.S....F.SEK..D..MKRL.-LPGDVIERGRNF.I....Q...Q....S..YSGG
Pf Fabp6a  .--A.A.RWLT....G.DD.CKLL.-IP.DIIEKGR.Y.M....T...DT.S...VY.TD
Hu FABP6   .--A....F.M.SEK..D..MKLL.-ISSDVIEKARNF.IV...Q...Q....S.HYSGG
Zf Fabp6   .--A.N..W.T.S..G..P.CKL..-IPDDVIAKGR.F.LV..IV.N.D......YY.NN
Pf Fabp2   .--..N.TWKVDRND...K.MEQM.-INMVKRKLAAHDNLKITIQ....T.HIKE.SN-F
Pf Fabp7b  .VDA.CATWK.VES..FDDYMK.L.-VGFATRQVGNVT.PTVII..E.DKVAIRTQST-F

                70        80        90       100       110       120
           ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp6b  WTWTNTFSVGRECELVTMKGTKFKAPVVMDDGRISVQFPQYHFTAEVSGDKLIMNCTIPG
Fu Fabp6b  ...S......Q.........IQ.....R.EN.K.............N.........T..
St Fabp6b  ...S.K.T..QD...T....V..T.A.T.EG.K..IP.........LIE...V.I.LT..
Me Fabp6b  .S...K.T..Q....L...NI....L.TLEN.KL.IP....L....IID...V.T.VT..
Rt FABP6   NIMS.K.TI.K...MQ..G.K....T.K.EG.KVVAD..N..Q.S..V....VEIS..--
Mo FABP6   NIMS.K.TI.K...MQ..G.K....T.K.EG.KVVAE..N..Q.S..V....VEIS..--
Pf Fabp6a  AKV....T..K.SDME.IG.K....T.F.EG.KL..T..N..H.S.I..G...ETSKA--
Hu FABP6   H.M..K.T..K.SNIQ..G.KT...T.Q.EG.KLV.N..N..Q.S.IV....VEVS..--
Zf Fabp6   HVV..K.I..K.SDME.VG.K...GI.S.EG.KLTIS..K.QQ.T.I..G..VETS.AS.
Pf Fabp2   RNLEMD.TL.VTFDYSLAD..ELSGSWS.EGESMKGV.VRKDNGKQLTTIRTLEGGELVQ
Pf Fabp7b  KNTEIS.KL.E.FDET.ADDRNC.ST.TL.GD--KLVHV.KWDGK.TKFVRE.KDGKLVM

                130              Identity (%)     Similarity (%)
           ....|....|....|..
Pf Fabp6b  ---EKGVTFRRTSRRV-        100               100
Fu Fabp6b  ---......K.I...L-         83                93
St Fabp6b  ---..S.........I-         71                82
Me Fabp6b  ---........VNK.I-         65                82
Rt FABP6   ----GD..YE.V.K..A         46                64
Mo FABP6   ----GD..YE.V.K.LA         45                65
Pf Fabp6a  ----GS.VLK...KKL-         44                64
Hu FABP6   ----G...YE.V.K.LA         44                65
Zf Fabp6   A--QGTAVLV...KK.-         41                61
Pf Fabp2   SYNYE..DAK.IFK.S-         19                38
Pf Fabp7b  DLTFED.HAV..YEKA-         18                36
```

Figure 2F

Fabp11a

```
                  10        20        30        40        50        60
            ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp11a  MVEKFVGTWKMISSENFDDYMKAIGVGFATRQVGNRTKPNLVFTVDDQGTVCMRTQSTFK
Fu Fabp11a  ........................................VS.........KS.....
Me Fabp11a  .A......................................VSIE....IS.KS.....
St Fabp11a  ............A..........................IV..E.D.V..LKS.....
Zf Fabp11a  ..D........TT.D...E.....................VC..E..LI..KS.....
Pf Fabp11b  ...Q.....TLA......E...T...N.....M..LV......S.AED.FISLKAE....
Me Fabp11b  ...Q.A...TLA......E....V.M......M..I......IQ.GAD.LIS.KAE....
St Fabp11b  ..DQ.....TL.A.....E........L....M..LA.....IG..EA.LIS.KSET...
Zf Fabp11b  ...QY..K...T..D...E....V.AS..S..MA.LA..S.LIA..E..VIT.KAVT...
Pf Fabp7b   ..DA.CA...LVE..........L...........V...TVIISQEGD-K.AI.......
Pf Fabp7a   ..DA.CA...LVD.Q........L...........V...TVAISK.GD-K.VIK.M...R
Pf Fabp2    .--T.N....VDRND.YEKF.EQM.INMVK.KLAAHDNLKITIQQ.GN-.FHIKES.N.R

                  70        80        90       100       110       120
            ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp11a  NTEIKFKLNEPFDETTADDRKTRTV--VTLENGKLVQKQNWDGKETCIEREISDGKLIAK
Fu Fabp11a  ............E.M..........--..............................
Me Fabp11a  T..V........E.........T..--I..D........S......S............
St Fabp11a  T...R....V..E.........T..--.S..G.......S......N.....T.......
Zf Fabp11a  T...........E.........T..--M.I.........T.....ST....V........
Pf Fabp11b  T.........E.E.....G.M.KADGRM.K.......H.K......T.....Q....T..
Me Fabp11b  T.....R.G.E.....V.G.NVKST--F..D.............TSTL....QE......
St Fabp11b  T..F......EC......G.T.K.L--...KD.....Q.T....T.TL....QG.NMS..
Zf Fabp11b  TL......D.E..........VK.T--MS.AD...I...T.E..T.I.....Q.T.M...
Pf Fabp7b   ....S...G.E.........NCKST--...DGD...HV.K......KFV...K....VMD
Pf Fabp7a   ....SA..G.E.....P...HVKST--FSM.GD..M.V.K.N....KFV...K...MVMT
Pf Fabp2    .L.MD.T.GVT..YSL..GTELSGSWSMEG.SM.G.FVRKDN..QLTTI.TLEG.E.VQS

                  130          Identity (%)     Similarity (%)
            ....|....|....|.
Pf Fabp11a  CIMGDVVAVRTYVREA        100              100
Fu Fabp11a  ..I..........K..         94               98
Me Fabp11a  ..............K..         86               95
St Fabp11a  ......I......K..         84               94
Zf Fabp11a  .K...........K..         82               93
Pf Fabp11b  ..AD....L...E.V-         67               77
Me Fabp11b  ...D........EKAT         66               79
St Fabp11b  .V.D........AKK.         65               82
Zf Fabp11b  .T.D....I...EKD.         61               79
Pf Fabp7b   LTFE..H.....EKA-         61               71
Pf Fabp7a   LTFEG.T.....EKA-         53               67
Pf Fabp2    YNYEG.D.K.IFK.S-         24               44
```

Figure 2G

Fabp11b

```
                    10        20        30        40        50        60
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp11b      MVEQFVGTWTLASSENFDEYMKTIGVNFATRQMGNLVKPNLVFSVAEDGFISLKAESTFK
Fu Fabp11b      .......N...V..D...D.................M........MGD.....M.S.....
Me Fabp11b      .....A.................AV.MG........IT.....IQ.GA..L..M.......
St Fabp11b      ..D........IA.........A...GL........A.....IG.D.A.L..M.S.T...
Fu Fabp11a      ...K.....KMI......D...A...G.....V..RT.....V..DDQ.TVCM.SQ....
Pf Fabp11a      ...K.....KMI......D...A...G.....V..RT......T.DDQ.TVCMRTQ....
Me Fabp11a      .A.K.....KMI......D...A...G.....V..RT.....V.IEDQ.T..M.SQ....
Zf Fabp11a      ..DK.....KMTT.D.......A...G.....V..RT.....VC.D.Q.L.CM.SQ....
St Fabp11a      ...K.....KMIA.....D...A...G.....V..RT....IVT.ED..VVC..SQ....
Zf Fabp11b      ....Y..K.KMT..D.......AV.AS..S...A..A..S.LIA.D.Q.V.TM..VT...
Pf Fabp6a       .--A.A.R.LTETQ.GY.DFC.LL.IPADIIEK.RDY.MITEVTQDG.T.SWTQVYP.DA
Pf Fabp10       .--D.N...QVYAQ..YE.FL.AMELPEDVIK.AKDI..ITEIKQNGKD.VVTSKTP-G.
Pf Fabp6b       .--T.T.KYE.ETQ..YE.FLDA..TAAGLLSAKTDH.VITEV.QDGND.TWTQSIP-HW

                    70        80        90       100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pf Fabp11b      TTEIKFKLNEEFEETTADGRMTKADGRMTKENGKLVQHQKWDGKETTIEREIQDGKLTAK
Fu Fabp11b      ........T...........N..T--VV.F..............................
Me Fabp11b      ......R.G...D...V...NV.S--TF.LD......K.N....TS.L.....E...I..
St Fabp11b      ...F.......CD.......T..T--LV.LKD.....Q.T....T..L.....G.NMS..
Fu Fabp11a      N.........P...M...D.K.RT--VV.L.......K.N......C.....S....I..
Pf Fabp11a      N.........P.D.....D.K.RT--VV.L.......K.N......C.....S....I..
Me Fabp11a      ...V......P.......D.K.TT--VI.LD......K.S......S.....S....I..
Zf Fabp11a      ..........P.......D.K.TT--V..I.......K.T.....S.....VS....I..
St Fabp11a      ....R....VP.......D.K.TT--VVSL.G.....K.S......N.....T....I..
Zf Fabp11b      .L......D...D.....D.KV.T--T.SLAD...I.K.T.E..T.I.......T.MI..
Pf Fabp6a       KVTNT.TVGK.SDME.IG.KKF..--TVFM.G...SVTF----PNYHHTS..SG...IET
Pf Fabp10       .LTNS.TIGK.ADI..M..KKL.C--TV.M.GS...CKT----EKFCHTQ.LKG.EMIET
Pf Fabp6b       .WTNT.SVGR.C.LV.MK.TKF..--PVVMDD.RISVQF----PQYHFTA.VSGD..IMN

                   130                 Identity (%)      Similarity (%)
                ....|....|....|....
Pf Fabp11b      CIAD---DVVALRTYERV-          100               100
Fu Fabp11b      ....---..........-             86                92
Me Fabp11b      ..M.---....V....KAT           72                81
St Fabp11b      .VM.---....V...AKKA           69                80
Fu Fabp11a      ..IG---....V...VKEA           67                77
Pf Fabp11a      ..MG---....V...V.EA           67                77
Me Fabp11a      ..MG---....V...VKEA           67                78
Zf Fabp11a      .KMG---....V...VKEA           67                77
St Fabp11a      ..MG---..I.V...VKEA           64                77
Zf Fabp11b      .TM.---....I....KDA           61                76
Pf Fabp6a       SK.G---S..LK..SKKL-           23                36
Pf Fabp10       LTKG---STTFV.KSKKMN           21                41
Pf Fabp6b       .TIPGEKG.TFR..SR..-           21                33
```

Figure 3: Phylogenetic analysis of pufferfish Fabp with other vertebrates.

A) A neighbour-joining tree showing the phylogenetic relationship between Fabp/FABPs of pufferfish (Pf), zebrafish (Zf), medaka (Me), three-spined stickleback (St), tilapia (Ti), human (Hu) and chicken (Ch). The bootstrap values are placed at the nodes and indicates the percentage number of times, out of 1000 bootstrap replications. The polypeptide sequence used in the phylogenetic analysis and their corresponding Ensembl/NCBI accession numbers are Pf Fabp1 (ENSTNIG00000016680), Me Fabp1 (ENSORLG00000003820), St Fabp1 (ENSGACG00000006574), Hu FABP1 (NP_001434), Pf Fabp 2 (ENSTNIG00000012360), Zf Fabp2 (NP_571506), St Fabp2a (ENSGACG00000017790), St Fabp2b (ENSGACG00000019474), Me Fabp2a (ENSORLG00000019478), Me Fabp2b (ENSORLG00000004146), Hu FABP2 (NP_000125), Pf Fabp3 (ENSTNIG00000016979), Me Fabp3 (ENSORLG00000011573), Ti Fabp3 (ENSONIG00000002309), Hu FABP3 (NP_004093), Hu FABP4 (NP_001433), Hu FABP5 (NP_001435), Pf Fabp6a (ENSTNIG00000017612), Pf Fabp6b (ENSTNIG00000004652), Me Fabp6a (ENSORLG00000005456), Me Fabp6b (ENSORLG000000012622), St Fabp6a (ENSGACG00000014354), St Fabp6b (ENSGACG00000018140), Hu FABP6 (NP_001436), Pf Fabp7a (ENSTNIG00000017217), Pf Fabp7b (ENSTNIG00000003393), Me Fabp7a (ENSORLG00000013475), Me Fabp7b (ENSORLG00000010773), St Fabp7a (ENSGACG00000010166), St Fabp7b (ENSGACT00000016034), Hu FABP7 (NP_001437), Ch FABP7 (NP_990639), Hu FABP8 (NP_002668), Hu FABP9 (NP_001073995), Pf Fabp10 (ENSTNIG00000002424), Me Fabp10a (ENSORLG00000014794), Me Fabp10b (ENSORLG00000007702), St Fabp10a (ENSGACG00000013639), St Fabp10b (ENSGACG00000002234), Ch FABP10 (P80226), Pf Fabp11a (ENSTNIG00000017781), Pf Fabp11b (ENSTNIG00000018400), Me Fabp11a (ENSORLG00000010830), Me Fabp11b (ENSORLG00000008282), St Fabp11a (ENSGACG00000003457), St Fabp11b (ENSGACG00000008725), Zf Fabp 11a (NP_001004682), Zf Fabp11b (NP_001018394) and Hu FABP12 (NP_001098751). The human lipocalin 1 protein (Hu LCN1, NP_002288) was used as outgroup. The scale bar indicates 0.2 substitutions per site. **B)** A maximum-likelihood tree showing the

evolutionary relationship of vertebrate Fabps, using the same amino acid sequence as mentioned above. The bootstrap values shown at the nodes indicate percentage based on 1000 replicates. The scale bar indicates 0.2 substitutions per site.

Figure 3A

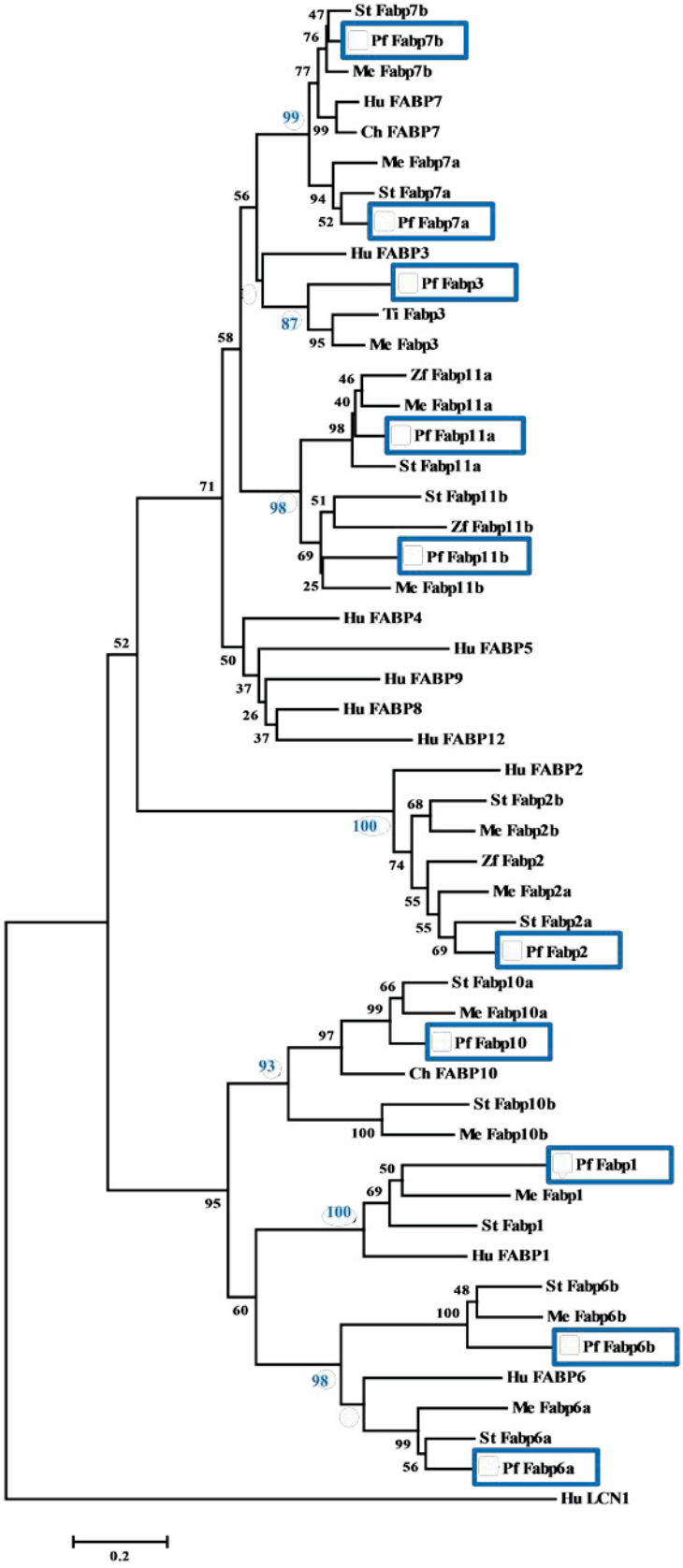
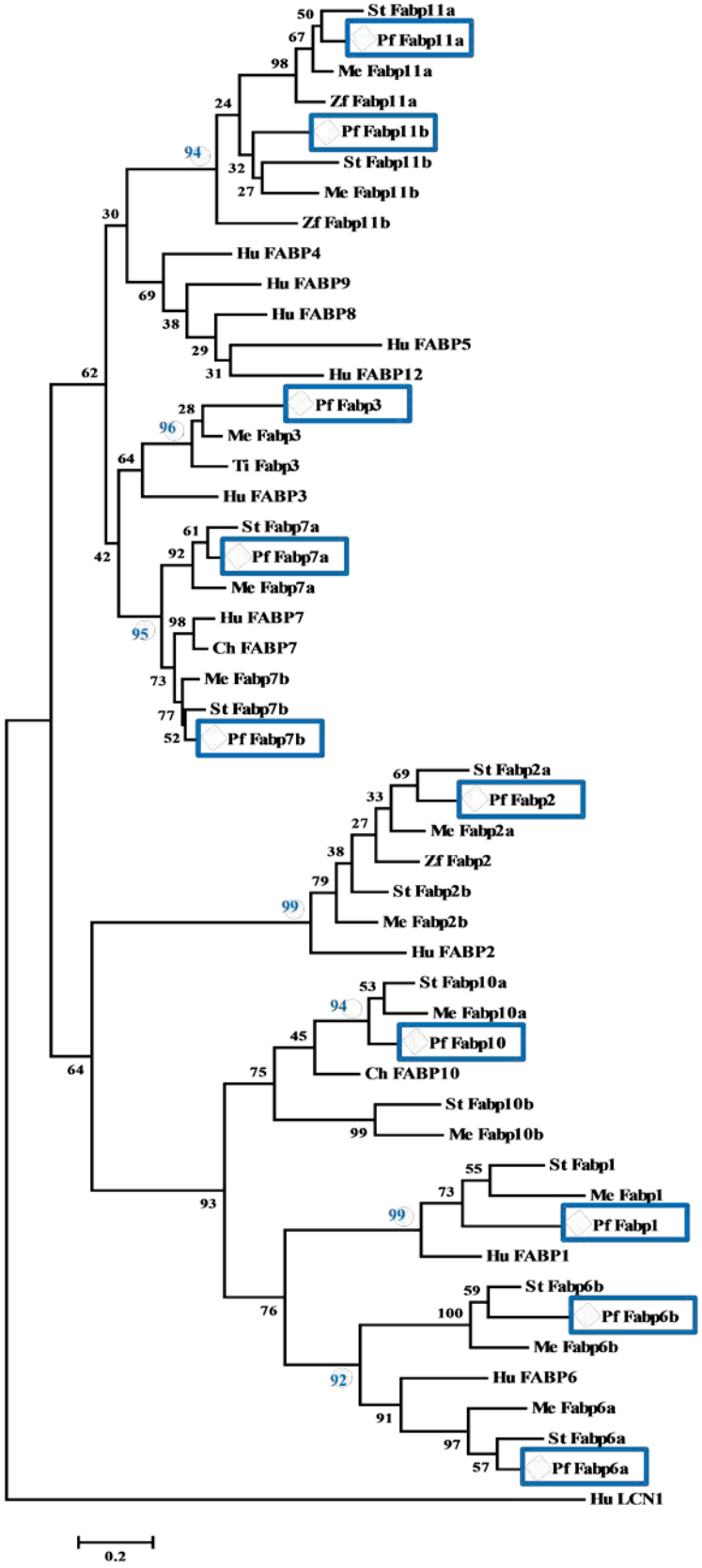

St Fabp7b
Pf Fabp7b
Me Fabp7b
Hu FABP7
Ch FABP7
Me Fabp7a
St Fabp7a
Pf Fabp7a
Hu FABP3
Pf Fabp3
Ti Fabp3
Me Fabp3
Zf Fabp11a
Me Fabp11a
Pf Fabp11a
St Fabp11a
St Fabp11b
Zf Fabp11b
Pf Fabp11b
Me Fabp11b
Hu FABP4
Hu FABP5
Hu FABP9
Hu FABP8
Hu FABP12
Hu FABP2
St Fabp2b
Me Fabp2b
Zf Fabp2
Me Fabp2a
St Fabp2a
Pf Fabp2
St Fabp10a
Me Fabp10a
Pf Fabp10
Ch FABP10
St Fabp10b
Me Fabp10b
Pf Fabp1
Me Fabp1
St Fabp1
Hu FABP1
St Fabp6b
Me Fabp6b
Pf Fabp6b
Hu FABP6
Me Fabp6a
St Fabp6a
Pf Fabp6a
Hu LCN1
0.2

Figure 3B


St Fabp11a
Pf Fabp11a
Me Fabp11a
Zf Fabp11a
Pf Fabp11b
St Fabp11b
Me Fabp11b
Zf Fabp11b
Hu FABP4
Hu FABP9
Hu FABP8
Hu FABP5
Hu FABP12
Pf Fabp3
Me Fabp3
Ti Fabp3
Hu FABP3
St Fabp7a
Pf Fabp7a
Me Fabp7a
Hu FABP7
Ch FABP7
Me Fabp7b
St Fabp7b
Pf Fabp7b
St Fabp2a
Pf Fabp2
Me Fabp2a
Zf Fabp2
St Fabp2b
Me Fabp2b
Hu FABP2
St Fabp10a
Me Fabp10a
Pf Fabp10
Ch FABP10
St Fabp10b
Me Fabp10b
St Fabp1
Me Fabp1
Pf Fabp1
Hu FABP1
St Fabp6b
Pf Fabp6b
Me Fabp6b
Hu FABP6
Me Fabp6a
St Fabp6a
Pf Fabp6a
Hu LCN1
0.2

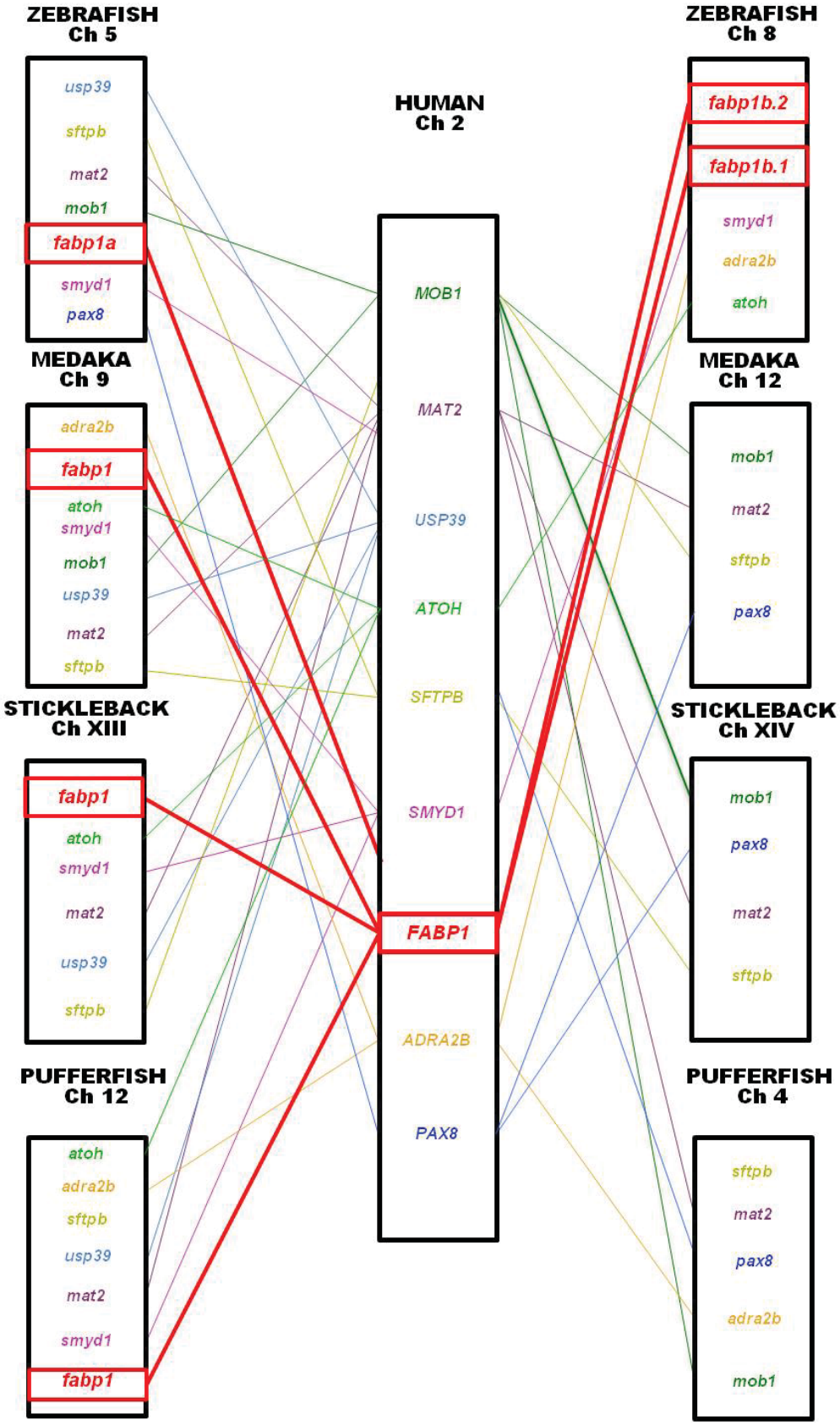
3.3 Conserved gene synteny

Conserved gene synteny was evident between *fabp* genes of pufferfish and the orthologous *Fabp/fabp* genes of human and other teleost fish (Figure 4). Conserved gene synteny of human *FABP1* on chromosome 2 with zebrafish *fabp1a* located on chromosome 5 and *fabp1b* located on zebrafish chromosome 8, suggests that the duplicated copies arose owing to the FSGD event (Sharma *et al*., 2006), followed by a tandem-duplication event, which lead to the retention of *fabp1b.1* and *fabp1b.2* on zebrafish chromosome 8 (Karanth *et al*., 2009). The syntenic genes from pufferfish chromosome 12 (location of *fabp1*) and 4, medaka chromosome 9 (location of *fabp1*) and 12, three-spined stickleback chromosome XII (location of *fabp1*) and XIV exhibit conserved gene synteny with the syntenic genes of human chromosome 2, suggesting a possible loss of duplicated *fabp1* gene from chromosome 4 of pufferfish, chromosome 12 of medaka and chromosome XIV of three spined stickleback (Figure 4A). In the Ensembl database location of pufferfish *fabp2* was designated 'un_random', indicating an unassigned chromosome. The synteny of genes on human chromosome 4(location of *FABP2*) is conserved with the order of genes in medaka scaffold 461(location of *fabp2a*) and chromosome 18 (location of *fabp2b*), three spined stickleback chromosome IX (location of *fabp2a*) and VII (location of *fabp2b*), suggest that these orthologous genes arose from the same ancestral genes as a result of the FSGD event. The conserved gene synteny of zebrafish chromosome 1 (location of *fabp2*) and chromosome 7, and pufferfish unassigned chromosome (location of *fabp2*) and chromosome 18 with human chromosome 4 suggest a possible loss of duplicated *fabp2* gene from zebrafish chromosome 7 and pufferfish chromosome 18 (Figure 4B). Figure 4C displays the

Figure 4: (A) Conserved gene synteny of the pufferfish, zebrafish, medaka and three-spined stickleback *fabp1* with human *FABP1*. The syntenic genes of human *FABP1* located chromosome (ch) 2, exhibit conserved gene synteny with pufferfish ch 12 and 4, three-spined stickleback ch XIII and XIV, medaka ch 9 and 12, and zebrafish ch 5 and 8. **(B)** The syntenic genes of human *FABP2* located on ch 4, exhibit conserved gene synteny with pufferfish unassigned chromosome and ch 18, three-spined stickleback ch IX and VII, medaka scaffold 461 and ch 18, and zebrafish ch 1 and ch 7. **(C)**The syntenic genes of human *FABP3* located on ch 1, exhibit conserved gene synteny with pufferfish ch 21, three-spined stickleback ch X and medaka ch 11. **(D)**The syntenic genes of human *FABP6* located on ch 5, exhibit conserved gene synteny with pufferfish unassigned chromosome and ch 1, three-spined stickleback ch III and IV, medaka ch 17 and ch 10, and zebrafish ch 21 and ch 2. **(E)**The syntenic genes of zebrafish duplicated *fabp11a/ fabp11b* located on ch 19 and ch 16, respectively, exhibit conserved gene synteny with pufferfish ch 21 and ch 8, three-spined stickleback ch X and XX, and medaka ch 11 and ch 18.

fabp1

Figure 4A



fabp2

Figure 4B

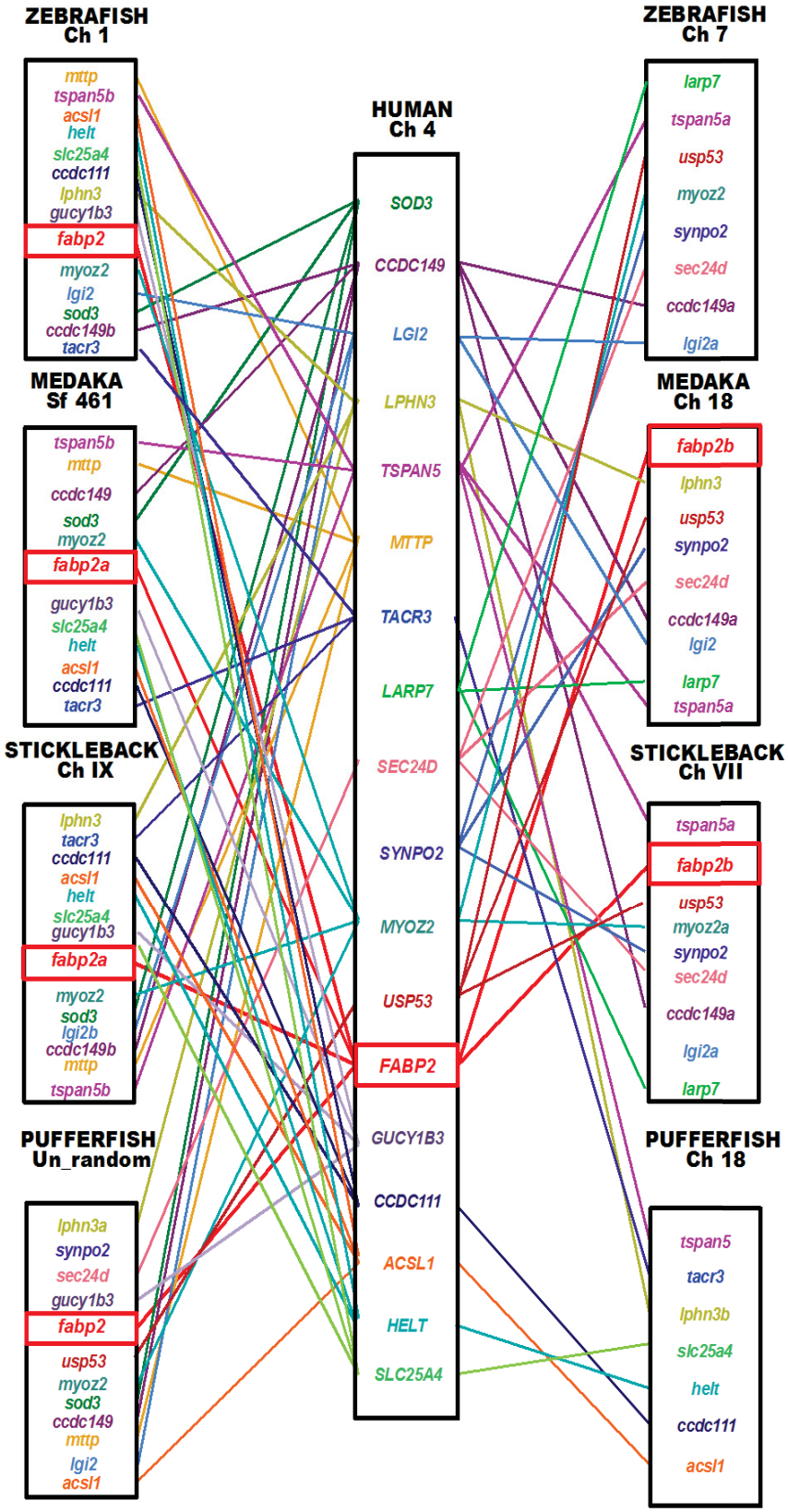
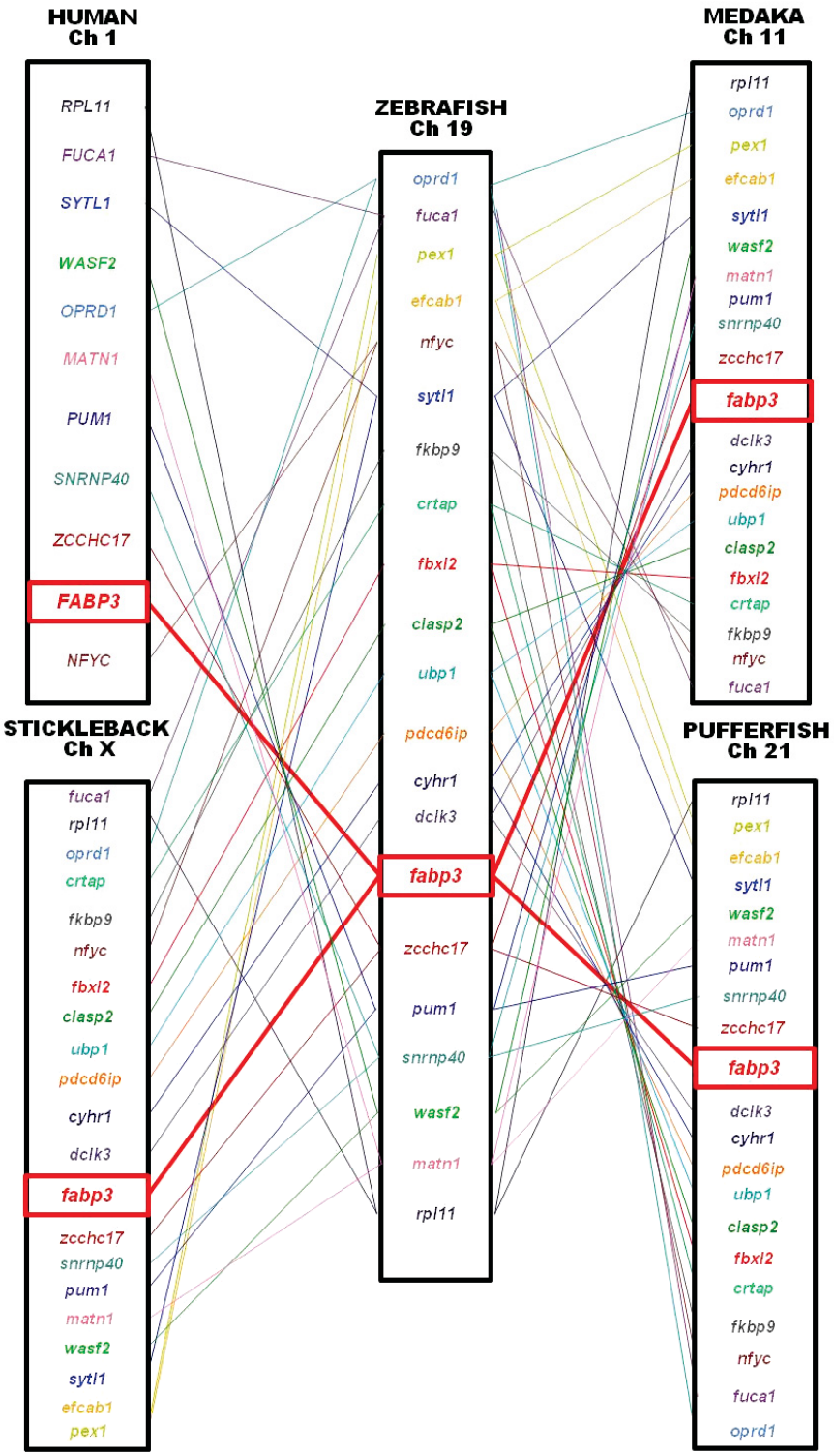


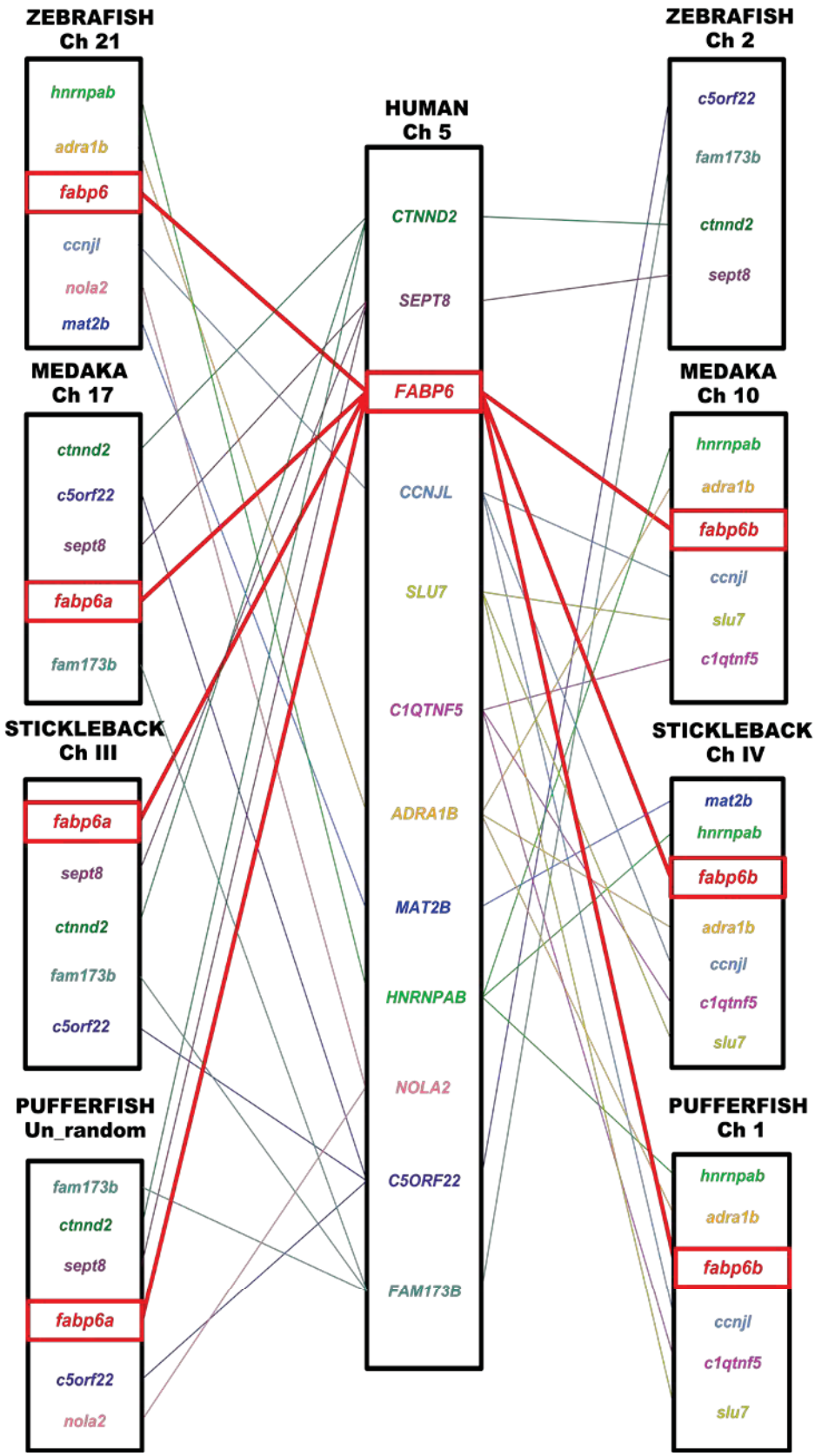
Figure 4C

fabp3



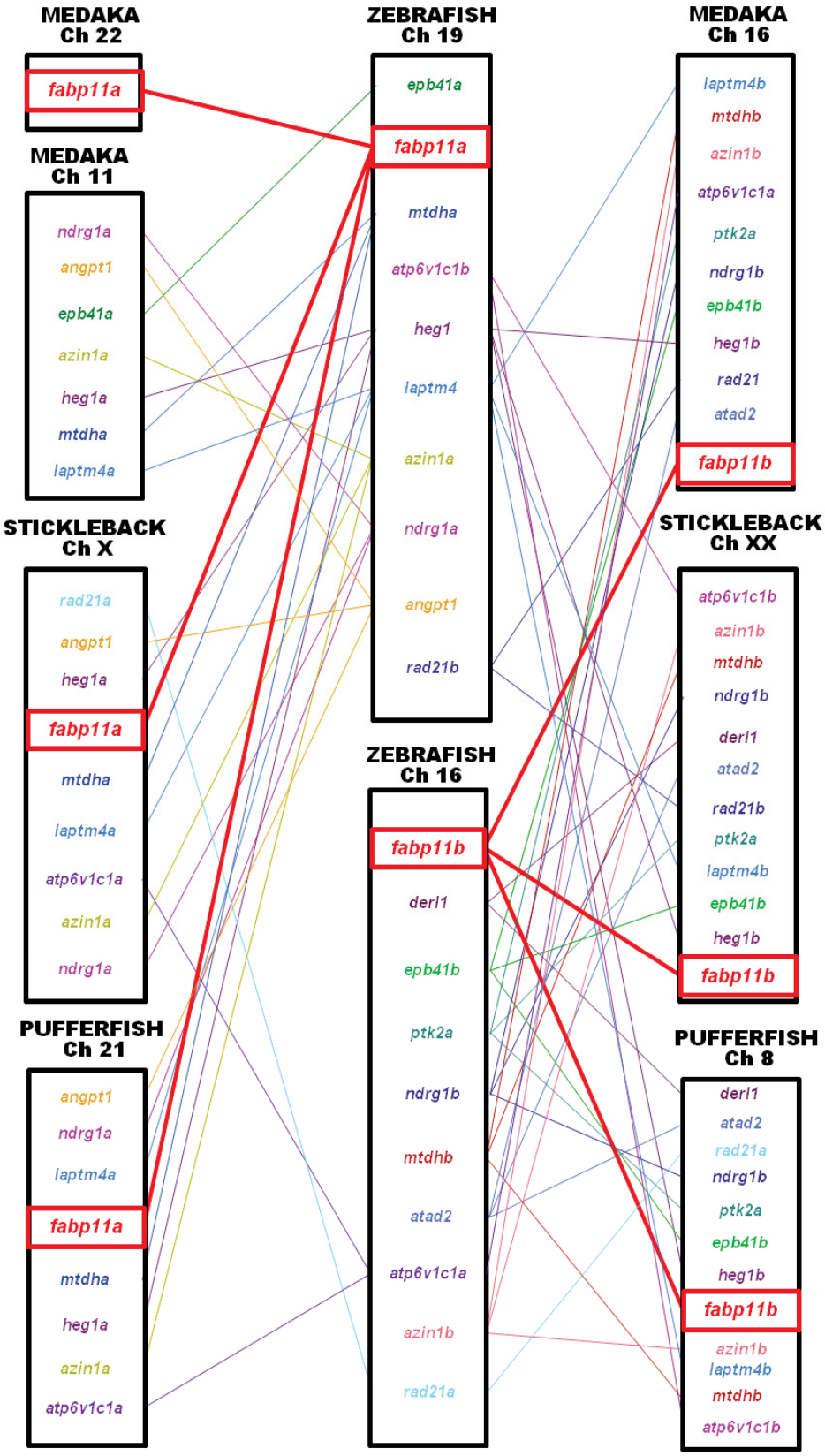
fabp6

Figure 4D



fabp11

Figure 4E



conserved gene synteny of pufferfish chromosome 21 (location of *fabp3*), with human chromosome 1, zebrafish chromosome 19, medaka chromosome 11 and stickleback chromosome X, however, no duplicated *fabp3* copy were identified in any other teleost fish, suggesting a possible loss of the duplicated genes. Similarly, syntenic genes of human *FABP6* on chromosome 5 are conserved with the syntenic genes of *fabp6a* and *fabp6b* from pufferfish on unassigned chromosome and on chromosome 1, three-spined stickleback chromosome III and IV, and medaka chromosome 17 and 10 respectively (Figure 4D), whereas conserved gene synteny on chromosome 21 (location of *fabp6*) and chromosome 2 with human chromosome 5, indicate a possible loss of duplicated *fabp6* copy from chromosome 2. The gene synteny of pufferfish *fabp11* on chromosome 21 (location of *fabp11a*) and 8 (location of *fabp11b*) was found to be conserved with chromosome X (location of *fabp11a*) and XX (location of *fabp11b*) of three-spined stickleback, and chromosome 22 (location of *fabp11a*) and 16 (location of *fabp11b*) of medaka (Figure 4E). There was no synteny observed between chromosome 22 of medaka and other teleost *fabp11a* containing chromosomes, however, a much evident synteny of conserved genes were observed in chromosome 11 of medaka.

3.4 Tissue-specific distribution and steady-state levels of pufferfish *fabp* gene transcripts

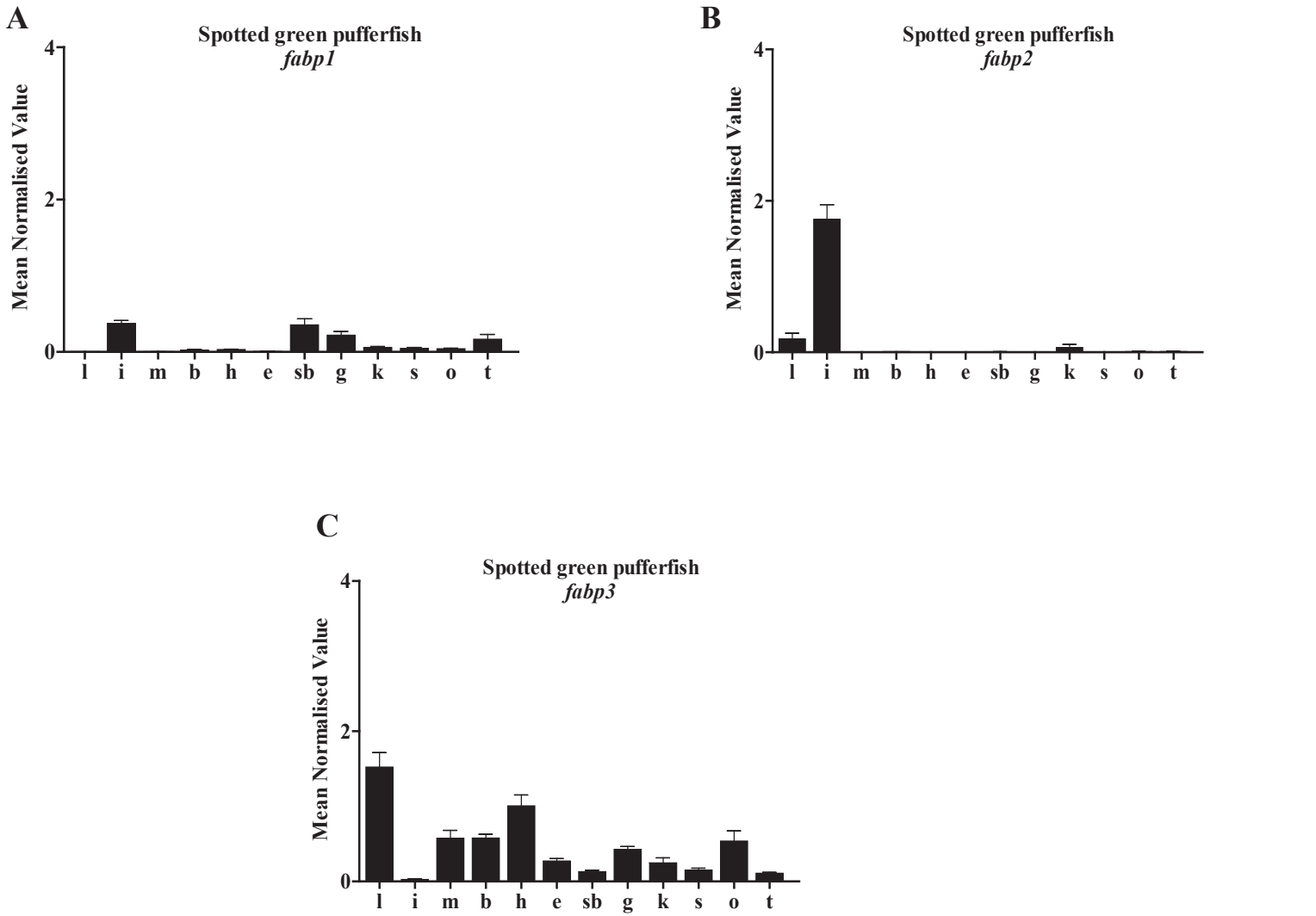
To further investigate the reason behind the retention of pufferfish *fabp* duplicates in their genome, the tissue-specific distribution of *fabp* mRNA transcripts in various tissues was determined. The transcripts of pufferfish *fabp1* were detected in all the tissues assayed except for the liver, muscle and eye (Figure 5A). Although Fabp1 was initially called the liver-type fatty acid binding protein, its mRNA transcripts were not detected in

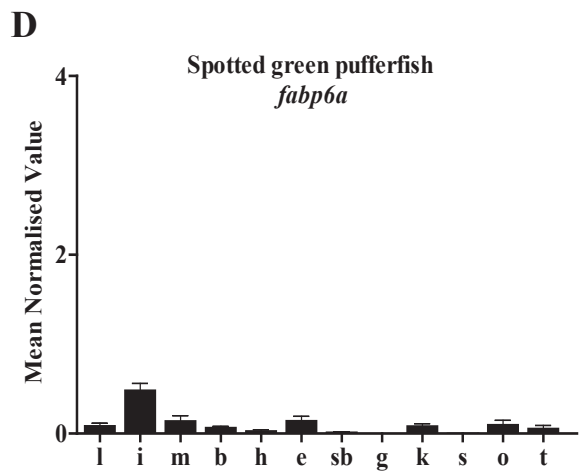
the liver of pufferfish nor zebrafish (Sharma *et al.,* 2006; Karanth *et al.,* 2009), but high transcript levels were evident in the intestine, when compared to other tissues. Unlike pufferfish, duplicated copies of *fabp1* exist in the zebrafish genome owing to the FSGD event, followed by a tandem duplication event, presumably due to unequal crossing-over during meiosis (Sharma *et al.,* 2006 and Karanth *et al.,* 2009). Previous studies have also revealed that the zebrafish genome has undergone extensive translocation leading to considerable inter-chromosomal rearrangements (Woods *et al.,* 2005). The pufferfish genome, however, contains fewer inter-chromosomal rearrangements when compared to other teleost genome, which lead Jaillon *et al.,* to deduce the proto-karyotype of the ancestral bony fish vertebrate consisting of 12 chromosomes, after the divergence from mammals about 450 million years ago. Force *et al.,* (1999) proposed that one of the duplicated gene copies most often assumes redundancy in gene function(s), thereby resulting in functional loss of a gene owing to accumulation of deleterious mutations (nonfunctionalization). This process may account for the loss of one of the duplicated *fabp1* following the FSGD in the pufferfish genome. When nonfunctionalization does not occur, as proposed for zebrafish *fabp1a* and *fabp1b* (Sharma *et al.,* 2006), owing to positive selection and mutation, neofunctionalization may occur, where one copy acquires a new function, while the other duplicated gene copy maintains the function(s) of the ancestral gene. The tissue-specific distribution of pufferfish *fabp1*, when compared with the duplicated copies of zebrafish *fabp1* shows redundancy in function (Figure 5A). The expression levels of pufferfish *fabp1* transcripts were found in the intestine, brain, heart, swim bladder, gills, kidney, skin, ovaries and testis, whereas in the case of

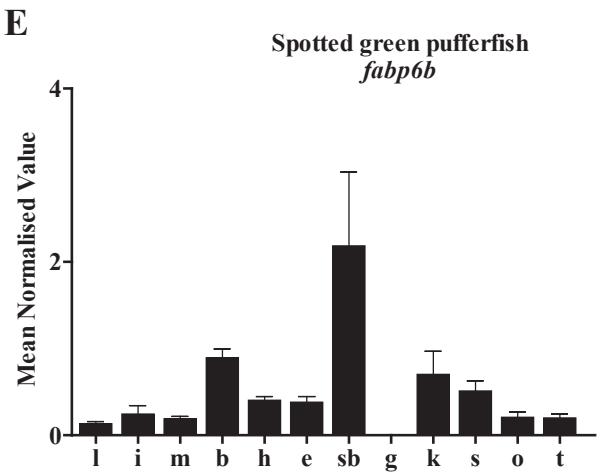
Figure 5: Tissue-specific distribution and steady-state levels of pufferfish *fabp* transcripts in various tissues.

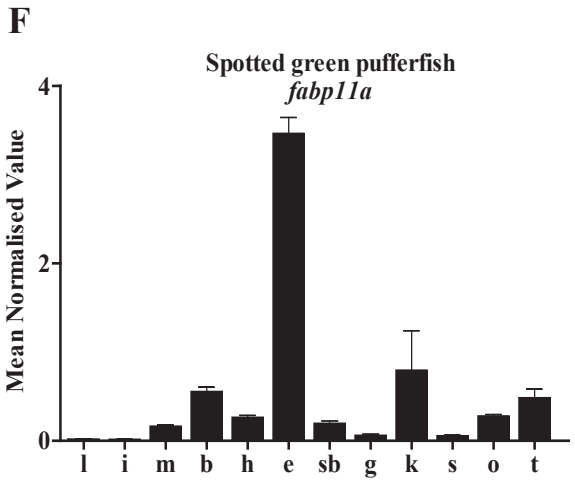
The steady state levels of pufferfish *fabp* transcripts in liver (l), intestine (i), muscle (m), brain (b), heart (h), eye (e), swim bladder (sb), gills (g), kidney (k), skin (s), ovary (o), testis (t). The pufferfish *fabp* transcripts investigated for tissue-specific distribution are **A)** *fabp1*, **B)** *fabp2*, **C)** *fabp3*, **D)** *fabp6a*, **E)** *fabp6b*, **F)** *fabp11a* and **G)** *fabp11b*. The bar graph depicted shows a mean value of three male and three female fishes combined. No significant statistical difference was evident between the tissue-specific steady-state levels of male and female fish. Single black line on each bar represents the standard error of mean.

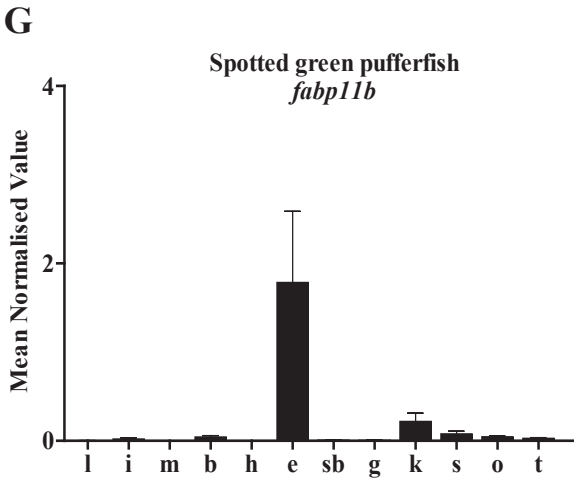
Figure 5


A
Spotted green pufferfish
fabp1
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t
B
Spotted green pufferfish
fabp2
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t
C
Spotted green pufferfish
fabp3
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t


D
Spotted green pufferfish
fabp6a
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t


E
Spotted green pufferfish
fabp6b
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t


F
Spotted green pufferfish
fabp11a
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t


G
Spotted green pufferfish
fabp11b
Mean Normalised Value
4
2
0
l i m b h e sb g k s o t

zebrafish, tissue-specific distribution of *fabp1b.1* transcripts were found in the intestine, liver, heart, ovary and testis, while the tissue-specific distribution of zebrafish *fabp1b.2* transcripts were found in intestine, heart, brain, eye, skin and ovary. This pattern indicates a functional divergence in the tissue-specific distribution, possibly owing to subfunctionalization (Karanth *et al.,* 2009), where the combined functions of both gene copies partition the functions of the ancestral gene. Transcripts of pufferfish *fabp2*were more abundant in the intestine than in the liver and kidney, and absent in muscle, brain, heart, eye, swim bladder, gills, skin, ovaries and testis (Figure 5B). While duplicated copies of medaka and three-spined stickleback *fabp2* are retained in the genome (Parmar *et al.,* 2012), the duplicated copy of zebrafish and pufferfish *fabp2* has been lost from the respective genomes. The tissue distribution of pufferfish *fabp2* transcripts was similar to the distribution of *fabp2* in zebrafish (Parmar *et al*.,2013a). The steady-state level of pufferfish *fabp2* transcripts were low in kidney, compared to the levels of medaka *fabp2a* (Parmar *et al.,* 2012). The tissue-specific distribution of pufferfish *fabp3* mRNA transcripts were distributed in all tissue assayed, except intestine (Figure 5C), with levels most abundant in liver, heart and ovaries. *fabp3* exists as a single copy in all the teleost fishes studied to date (Liu *et al*., 2007; BLAST search in Ensembl genome database), however, *fabp3* of Antarctic teleost fishes are found to be retained as duplicated copies (Vayda *et al.,* 1998).The pufferfish *fabp3* mRNA levels are high in heart and liver tissues, in contrast to lower levels of expression in all other tissues. These results correlate with the findings of Liu *et al.,* in 2003(a), where they first reported *fabp3* of zebrafish. The steady-state levels of pufferfish *fabp6a* mRNA transcripts were higher in intestine, when compared to liver, muscle, brain, heart, eye, kidney, ovaries and testis (Figure 5D),

whereas, *fabp6b*, was abundant in swim bladder, brain and kidney. Swim bladder, brain, heart, kidney and skin showed high steady-state levels of *fabp6b* transcripts when compared to *fabp6a*. This tissue-specific pattern of mRNA distribution may contribute to the retention of duplicated copies in the pufferfish genome after the FSGD event. Parmar and Wright (i*n press*) showed that *fabp7* exists as duplicated copy and *fabp10* as single copy in the pufferfish genome. They identified and discussed the various tissue-specific distribution patterns of the orthologous *fabp7* and *fabp10* transcripts in medaka, zebrafish, three spined stickleback and pufferfish. The *fabp7a* mRNA transcripts are found in all the tissues, except for swim bladder and ovaries, where as *fabp7b* transcripts were found only in intestine, heart and kidney. The tissue-specific distribution of *fabp7b* mRNA transcripts was found only in liver, intestine, brain, gills and kidney (Parmar and Wright, *in press*). The steady-state level of pufferfish *fabp11a* transcripts was abundant in brain, kidney, eye, ovary and testis, when compared to low levels of *fabp11a* mRNA transcripts in muscle, heart, swim bladder, gills and skin. Similar to their orthologs in medaka and zebrafish (Parmar *et al.,* 2012b), both the duplicated copies of pufferfish *fabp11* were dominant in eye. The tissue-specific distribution of *fabp11a* was found to be extensively distributed in most of the tissue samples assayed, when compared to *fabp11b*. This pattern of tissue-specific distribution suggests either a subfunctionalization, where the functions of an ancestral gene is divided between the duplicates or neofunctionalization, where one copy of the sister duplicate acquires a novel function.

Susumu Ohno in 1970 proposed that at least two rounds of whole genome duplication in early vertebrate lineage contributed to the increased complexity and genome size of vertebrates. Three whole genome duplication events have so far been

proposed in the evolutionary history of vertebrates. However, studies on timing of the first two-rounds of genome duplication in the vertebrate lineage remains debatable (Dehal and Boore, 2005). The third-round or the FSGD is clearly evident in the evolution of vertebrate lineage (Jaillon *et al.,* 2004; Woods *et al.,* 2005). The comparative analysis of about 6000 pufferfish genes to the human genome showed patterns of 'double conserved synteny', where one chromosomal region in mammal matches with two in the pufferfish genome (Jaillon *et al.,* 2004). The study described here was performed to identify and structurally characterize the *fabp* genes in the pufferfish genome, to determine if the duplicated copies of pufferfish *fabp* genes arose as a consequence of FSGD, approximately 230-400 million years ago in the ray-finned fish lineage and to provide evidence regarding the evolutionary processes that lead to the loss or retention of these duplicated copies after the FSGD. I identified ten different paralogs of pufferfish *fabp* genes, three of which were found to be retained as duplicated copies (*fabp6*, *fabp7* (Parmar and Wright, *in press*) and *fabp11*). Multiple sequence alignment and phylogenetic analysis showed that the pufferfish Fabps are orthologous to vertebrate Fabps/FABPs. Conserved gene syntenies for each pufferfish *fabp* gene suggest that the duplicated genes arose as a result of the FSGD event. A comparative study on the pattern of tissue-specific distribution of duplicated pufferfish *fabp* mRNA transcripts (Table 4) with other teleost orthologs, concluded that, following the FSGD event, a considerable loss of duplicated genes from the pufferfish genome were evident in pufferfish *fabp1*, *fabp2*, *fabp3* and *fabp7* transcripts. Also evident is the differential tissue-specific distribution patterns, which suggests considerable divergence in the *cis*-regulatory elements of pufferfish *fabp6*, *fabp10* and *fabp11* since their duplication.

Table 4: The tissue-specific distribution of the steady-state levels for pufferfish *fabp* transcripts. + / – indicates presence/ absence of transcripts, respectively.

Tissues	Gene transcripts						
	fabp1	*fabp2*	*fabp3*	*fabp6a*	*fabp6b*	*fabp11a*	*fabp11b*
Liver	-	+	+	+	+	-	-
Intestine	+	+	+	+	+	-	-
Muscle	-	-	+	+	+	+	-
Brain	+	-	+	+	+	+	+
Heart	+	-	+	-	+	+	-
Eye	+	-	+	+	+	+	+
Swim bladder	+	-	+	-	+	+	-
Gills	+	-	+	-	-	+	-
Kidney	+	+	+	+	+	+	+
Skin	+	-	+	-	+	+	+
Ovary	+	-	+	+	+	+	+
Testis	+	-	+	+	+	+	-

Further insight into the retention or loss of *fabp* genes in teleost fishes may be gained from the study of tissue-specific distribution of *fabp* gene transcripts in a basal extant species of ray-finned fishes (e.g. spotted gar and bichir), which diverged from the teleost fish lineage before the FSGD event or the ancient fish, coelacanth, a member of the lobe-finned fishes. This type of study might shed some light on the spatial distribution of an ancient *fabp* gene transcript, thereby providing insight into the evolutionary processes that lead to the retention or loss of duplicated *fabp*s in the teleost fish genome. Pre ensembl, a preliminary database, contains partial datasets for spotted gar and coelacanth, while the fully annotated genomic dataset can soon be expected in the Ensembl database.

Bibliography

Alves-Costa, F. A., Denovan-Wright, E. M., Thisse, C., Thisse, B., & 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 Journal*, 275, 3325-3334.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., et al. (1998). Zebrafish *hox* clusters and vertebrate genome evolution. *Science*, 282, 1711-1714.

Ando, S., Xue, X.H., Tibbits, G.F., Haunerland, N.H. (1998). Cloning and sequencing of complementary DNA for fatty acid binding protein from rainbow trout heart. *Computational Biochemistry and Physiology*, 119, 213-217.

Breathnach, R., &Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry*, 50, 349-383.

Bernlohr, D. A., Simpson, M. A., Hertzel, A. V., & Banaszak, L. J. (1997). Intracellular lipid-binding proteins and their genes. *Annual Review of Nutrition*, 17, 277-303.

Bustin, S.A., Benes, V., Nolan, T., & Pfaffl, M.W. (2005). Quantitative real-time RT-PCR – a perspective. *Journal of Molecular Endocrinology*, 34, 597-601.

Dasilva, C., Hadji, H., Ozouf-Costaz, C., Nicaud, S., Jaillon, O.,Weissenbach, J., & Roest Crollius, H. (2002). Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. *PNAS,* 99, 13636–13641.

Dehal, P., &Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate, *PLoS Biology*, 3, e314.

Denovan-Wright, E. M., Pierce, M., & Wright, J. M. (2000a). Nucleotide sequence of cDNA clones coding for a brain-type fatty acid-binding protein and its tissue-specific expression in adult zebrafish (*Danio rerio*). *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1492, 221-226.

Denovan-Wright, E. M., Pierce, M., Sharma, M. K., & Wright, J. M. (2000b). cDNA sequence and tissue-specific expression of a basic liver-type fatty acid-binding protein in adult zebrafish (*Danio rerio*). *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1492, 227-232.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J.H. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151, 1531-1545.

Furuhashi, M., & Hotamisligil, G.S. (2008). Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nature Reviews Drug Discovery*, 7, 489-503.

Gromiha, M.M. (2010). Protein sequence analysis. In: Protein bioinformatics: from sequence to function. Elsevier Inc., New Delhi, India. pp. 29-62.

Hanhoff, T., Lucke, C., & Spener, F. (2002). Insights into binding of fatty acids by fatty-acid binding proteins. *Molecular and Cellular Biochemistry*, 239, 45-54.

Hinegardner, R. (1968). Evolution of cellular DNA content in teleost fishes. *The American Naturalist*, 102, 517-523.

Hertzel, A.V., & Bernlohr, D.A. (2000). The mammalian fatty acid-binding multigene family: molecular and genetic insights into function. *Trends in Endocrinology and Metabolism*, 11, 175-180.

Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E.,Bouneau,

L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biémont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigó, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quétier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J., &Roest Crollius, H.(2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, 946-957.

Karanth, S., Denovan-Wright, E. M., Thisse, C., Thisse, B., & 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 Journal*, 275, 3031-3040.

Karanth, S., Denovan-Wright, E. M., Thisse, C., Thisse, B., & 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.

Lai, Y. Y., Lubieniecki, K. P., Phillips, R. B., Chow, W., Koop, B. F., & 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. *Marine Genomics*, 2,193-200.

Lai, Y. Y., Lubieniecki, K. P., Koop, B. F., & Davidson, W. S. (2012). Characterization of the Atlantic salmon (*Salmo salar*) brain-type fatty acid binding protein (*fabp7*) genes reveals the fate of teleost *fabp7* genes following whole genome duplications. *Gene*, 504, 253-261.

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

Liu, R. Z., Denovan-Wright, E. M., & Wright, J. M. (2003b). Structure, mRNA expression and linkage mapping of the brain-type fatty acid-binding protein gene (*fabp7*) from zebrafish (*Danio rerio*). *European Journal of Biochemistry*, 270, 715-725.

Liu, R. Z., Denovan-Wright, E. M., Degrave, A., Thisse, C., Thisse, B., & 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 (*Daniorerio*). *Gene Expression Patterns*, 4, 379-387.

Liu, R. Z., Saxena, V., Sharma, M. K., Thisse, C., Thisse, B., Denovan-Wright, E. M., & 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 Journal*, 274, 1621-1633.

Londraville, R.L., & Sidell B.L. (1996). Cold acclimation increases fatty acid-binding protein concentration in aerobic muscle of striped bass, *Morone saxatilis*, *The Journal of Experimental Zoology*, 275, 36-44.

Medzihradszky, K.F., Gibson, B.W., Kaur, S., Yu, S., Medzihradszky, D., Burlingame, A.L., &Bass, N.M. (1992). The primary structure of fatty-acid-binding protein

from nurse shark liver.Structural and evolutionary relationship to the mammalian fatty-acid-binding protein family. *European Journal of Biochemistry*, 203, 327-339.

Neafsey, D.E., & Palumbi, S.R. (2003). Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes. *Genome Research*, 13, 821-830.

Ockner, R. K., Manning, J. A., Poppenhausen, R. B., & Ho, W. K. (1972). A binding protein for fatty acids in cytosol of intestinal mucosa, liver, myocardium, and other tissues. *Science*, 177, 56-58.

Ohno, S. (1970). *Evolution by Gene Duplication.* New York (NY): Springer-Verlag.

Parmar, M. B., Venkatachalam, A. B., & Wright, J. M. (2012a). Comparative genomics and evolutionary diversification of the duplicated *fabp6a* and *fabp6b* genes in medaka and three-spined stickleback. *Comparative Biochemisry and Physiology. Part D: Genomics and proteomics*, 7, 311-321.

Parmar, M. B., Venkatachalam, A. B., & Wright, J. M. (2012b). The evolutionary relationship of the transcriptionally active *fabp11a* (intronless) and *fabp11b* genes of medaka with *fabp11* genes of other teleost fishes. *FEBS Journal*, 279, 2310-2321.

Parmar, M. B., & Wright, J. M. (2013a). Comparative evolutionary genomics of medaka and three-spined stickleback *fabp2a* and *fabp2b* genes with *fabp2* of zebrafish. *Genome*, 56, 27-37.

Parmar, M. B., & Wright, J. M. (2013b). Comparative genomic organisation and tissue-specific transcription of the duplicated *fabp7* and *fabp10* genes in teleost fishes. *In*

press.

Pierce, M., Wang, Y. M., Denovan-Wright, E. M., & 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*). *Biochimica et Biophysica Acta-Gene Structure and Expression*, 1490, 175-183.

Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W., &Weissenbach, J. (2000). Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genetics*, 25, 235-238.

Schaap, F. G., van der Vusse, G. J., & Glatz, J. F. C. (2002). Evolution of the family of intracellular lipid-binding proteins in vertebrates. *Molecular and Cellular Biochemistry*, 239, 69-77.

Schleicher, C. H., Córdoba, O. L., Santomé, J. A., & Dell'Angelica, E. C. (1995). Molecular evolution of the multigene family of intracellular lipid-binding proteins. *Biochemistryand Molecular Biology International*, 36, 1117-1125.

Sharma, M. K., Denovan-Wright, E. M., Degrave, A., Thisse, C., Thisse, B., & 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*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 138, 391-398.

Sharma, M. K., Liu, R. Z., Thisse, C., Thisse, B., Denovan-Wright, E. M., & 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 Journal*, 273, 3216-3229.

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

Stewart, J.M., &Driedzic, W.R. (1998). Fatty acid binding proteins in teleost fish. *Canadian Journal of Zoology*, 66, 2671-2675.

Storch, J., & Corsico, B. (2008). The emerging functions and mechanisms of mammalian fatty acid-binding proteins. *Annual Review of Nutrition*, 28, 73-95.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, 1596-1599.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680.

Vayda, M.E., Londraville, R.L., Cashon, R.E., Costello, L., & Sidell, B.D. (1998). Two distinct types of fatty acid-binding proteins are expressed in heart ventricle of Antarctic teleost fishes. *Journal of Biochemistry*, 330, 375-382.

Veerkamp, J. H., & Maatman, R. G. (1995). Cytoplasmic fatty acid-binding proteins: their structure and genes. *Progress in Lipid Research*, 34, 17-52.

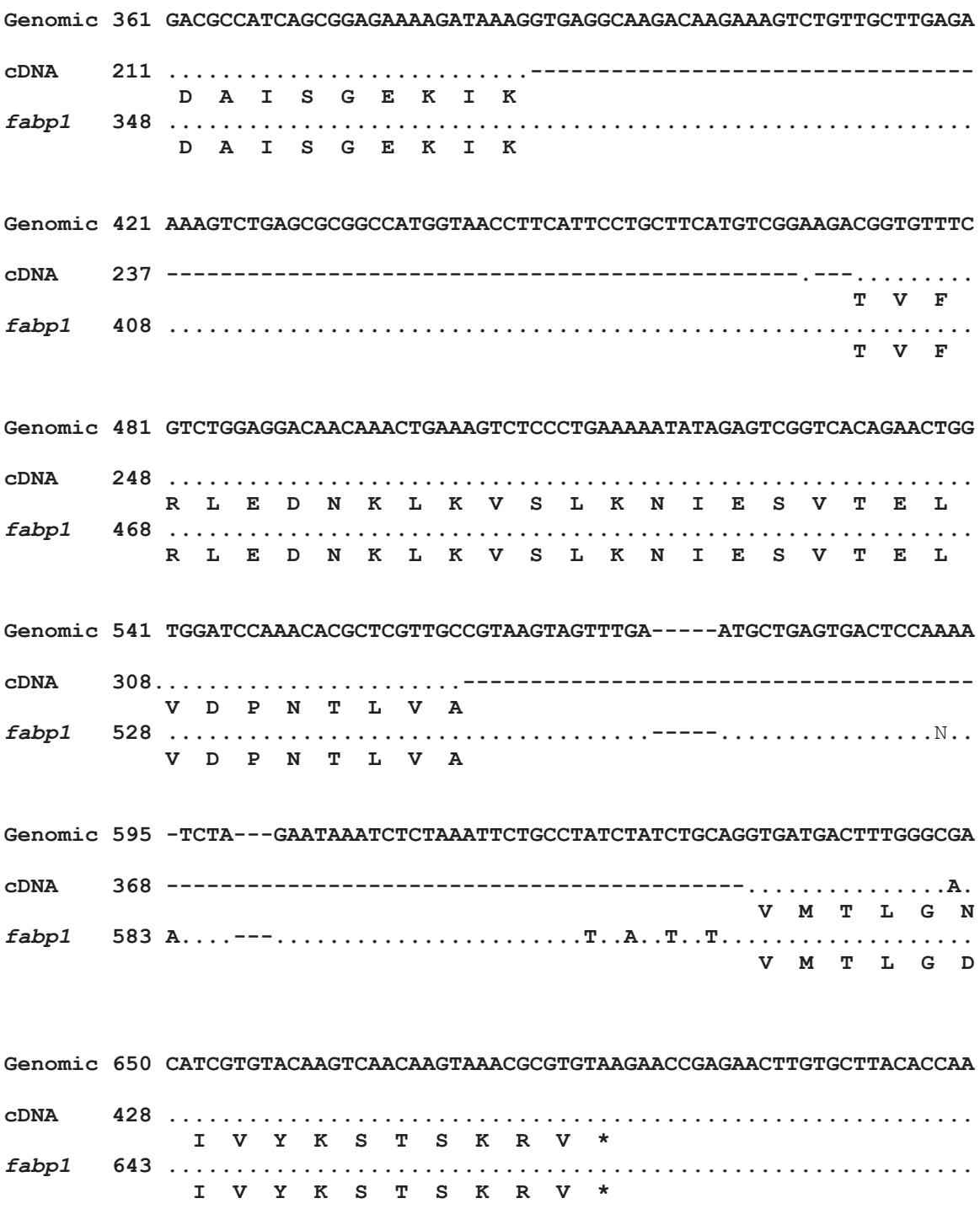
Venkatachalam, A. B., Thisse, C., Thisse, B., & Wright, J. M. (2009). Differential tissue-specific distribution of transcripts for the duplicated fatty acid-binding protein 10 (*fabp10*) genes in embryos, larvae and adult zebrafish (*Danio rerio*). *FEBS Journal*, 276, 6787-6797.

Woods, I.G., Wilson, C., Friedlander, B., Chang, P., Reyes, D.K., Nix, R., Kelly, P.D., Chu, F., Postlethwait, J.H., &Talbot, W.S. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Research*, 14, 1307–1314.

Appendix

Sequence alignment of Ensembl-derived genomic DNA, NCBI-derived cDNA and PCR amplified genomic of pufferfish *fabp1*. The amino acid sequences are shown below in bold. The dots represent identity and the dashes represent missing nucleotides. The star symbol indicates the stop codon.

```
Genomic 1    TTTGGGTGTATAAAACGGGCAGGTGAGCAGGTGTTTTCCTCTACTCTTCGTCATCACTGT

cDNA    1    ------------------------------------------------------------

fabp1   1    ........................................-.....-......----..
                                                                    M

Genomic 61   CTCTCTCATCTGGAAAATACCAGCAGGTGTCATTCCAGGTAGAACTTTGAGCCTTTTATG

cDNA    1    ------------------------------------------------------------

fabp1   55   .......---.....................---T....-.....................
             S  L  S     G  K  Y  Q  Q  V  S     Q   E  N  F  E  P  F  M

Genomic 121  AAGGCCATTGTGAAGAGTCCAGCTTCGTCTTAACAGCCCTGAGTTTCTCATTTGATGATA

cDNA    1    ----------------------.CTC...CTGG.AAATACCAGCAGGTG.C.C.G..-G

fabp1   108  ............................................................
              K  A  I

Genomic 181  AACATTGTTGCTTGTTCTGCCTGTTTTCTTTTAGGTGTTCCTGATGAAATCATCCAGCAG

cDNA    37   ...T...A-..C.T..A..AAG.CCA.T-----..........................
                              K  A  I       G  V  P  D  E  I  I  Q  Q
fabp1   168  ............................................................
                                         G  V  P  D  E  I  I  Q  Q

Genomic 241  ATCAAAGACATCATGAGCTTCACTGAGATTGTGCAGAATGGCAATGATTTCAAAATCACC

cDNA    91   ............................................................
              I  K  D  I  M  S  F  T  E  I  V  Q  N  G  N  D  F  K  I  T
fabp1   228  ............................................................
              I  K  D  I  M  S  F  T  E  I  V  Q  N  G  N  D  F  K  I  T

Genomic 301  ACCACTACTGGCCCAAAGGTAACGGTCAACCAATTTACCATTGGGAAGGAGACAGAGATG

cDNA    151  ............................................................
              T  T  T  G  P  K  V  T  V  N  Q  F  T  I  G  K  E  T  E  M
fabp1   288  ............................................................
              T  T  T  G  P  K  V  T  V  N  Q  F  T  I  G  K  E  T  E  M
```


Genomic 361 GACGCCATCAGCGGAGAAAAGATAAAGGTGAGGCAAGACAAGAAAGTCTGTTGCTTGAGA
cDNA 211 .....----------
D A I S G E K I K
fabp1 348 .....
D A I S G E K I K
Genomic 421 AAAGTCTGAGCGCGGCCATGGTAACCTTCATTCCTGCTTCATGTCGGAAGACGGTGTTTC
cDNA 237 ----------.---.....
T V F
fabp1 408 .....
T V F
Genomic 481 GTCTGGAGGACAACAAACTGAAAGTCTCCCTGAAAAATATAGAGTCGGTCACAGAACTGG
cDNA 248 .....
R L E D N K L K V S L K N I E S V T E L
fabp1 468 .....
R L E D N K L K V S L K N I E S V T E L
Genomic 541 TGGATCCAAACACGCTCGTTGCCGTAAGTAGTTTGA-----ATGCTGAGTGACTCCAAAA
cDNA 308.....----------
V D P N T L V A
fabp1 528 .....-----.....N..
V D P N T L V A
Genomic 595 -TCTA---GAATAAATCTCTAAATTCTGCCTATCTATCTGCAGGTGATGACTTTGGGCGA
cDNA 368 ----------.....A.
V M T L G N
fabp1 583 A....---.....T..A..T..T.....
V M T L G D
Genomic 650 CATCGTGTACAAGTCAACAAGTAAACGCGTGTAAGAACCGAGAACTTGTGCTTACACCAA
cDNA 428 .....
I V Y K S T S K R V *
fabp1 643 .....
I V Y K S T S K R V *