THE CHARACTERIZATION OF LACTATE DEHYDROGENASE GENES IN RAINBOW SMELT (*Osmerus mordax*)

by

Xuezheng (Jenny) Ma
B.Sc. (Honours), Saint Mary's University, 2007

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Summer 2009

APPROVAL

Name: **Xuezheng (Jenny) Ma**

Degree: **Master of Science**

Title of Thesis: **The Characterization of Lactate Dehydrogenase genes in rainbow smelt (*Osmerus mordax*)**

Examining Committee:

Chair: **Dr. David L. Baillie**
Professor, Department of Molecular Biology and Biochemistry

_____

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

_____

Dr. Jack N. Chen
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

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

_____

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

Date Defended/Approved: August-17-2009

Declaration of
Partial Copyright Licence

ABSTRACT

Lactate Dehydrogenase isozymes (LDH-A, LDH-B and LDH-C) represent the classical example of a multi-gene system derived by successive gene duplications. By investigating the genes encoding the LDH isozymes in rainbow smelt, a diploid out-group of the tetraploid salmonids, I sought to gain insight into the effect of a whole genome duplication superimposed upon more ancient gene duplications. I isolated rainbow smelt BAC clones containing the LDH-A, LDH-B and LDH-C genes, made shotgun libraries of three representative BACs and annotated the sequences. I characterized the smelt LDH genes with respect to structure, tissue expression and genome organization. This information was used for comparative genomic analyses with the LDH genes from Atlantic salmon. There was no evidence for positive selection, an expectation of neo-functionalization, but different rates of amino acid substitutions between and within lineages were evident in the LDH-A and LDH-B salmonid duplicates. LDH-B1 and LDH-B2 in salmonids have experienced sub-functionalization.

Keywords:

Gene duplication; Genome duplication; Lactate dehydrogenase; Rainbow smelt; Salmonids

DEDICATION

To my parents: Lixin Ma and Chunrong Wang, grandparents: Shaoquan Wang and Shuqin Zhang.

ACKNOWLEDGEMENTS

I would sincerely like to thank my senior supervisor Dr. William Davidson, for giving me the opportunity to work on this project, and for his patience, encouragement, understanding help and financial support. I also would like to extend my thanks to all past and present members in Davidson Lab, for their help and support during my Master project. Finally, I would like to bring my special thanks to my grandparents and parents, for their never-ending love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Gene duplication

Charles Darwin (1872) first proposed the remarkable theory of evolution by natural selection. He stated that, “from the strong principle of inheritance, any selected variety will tend to propagate its new and modified form” (Darwin 1859). It was suggested that evolution is the accumulation of genetic changes within the genome and that natural selection drives the degree of the genetic changes (Ohno 1970a). In 1970, Susumu Ohno published the book “Evolution by Gene Duplication”. He stated that gene duplication is “natural selection merely modified, while redundancy created” and proposed that the cumulative allelic mutations arising from existing gene loci under the pressure of natural selection are extremely conservative and cannot provide new genes with novel functions. However, evolution requires the creation of new genes with new functions to allow organisms to adapt to changing environments. In order to escape from the pressure of natural selection, the redundant gene loci derived by duplication accumulate formerly forbidden mutations, which can change the active site of a protein and develop proteins with novel functions. Therefore, gene duplication has a major role in evolution (Ohno 1970a). In Ohno’s theory, he concluded the two major factors driving the evolution of gene duplication are tandem gene duplications and entire genome duplications. Since the theory of gene duplication

proposed by Ohno, the evidence and investigations based on genetic and genomic projects have confirmed his speculations.

1.1.1 Early gene duplication research

Gene duplication was first proposed by Haldane and Muller who suggested that the duplicated gene is derived by divergent mutations that finally drive the production of a new gene. The early stage for studying gene duplication mostly focused on the observation of the organism, speciation and chromosome morphology. In the 1910s, Calvin Bridges addressed the idea that morphology varies according to the karyotype, which may be related to the gene duplication events. Muller proposed that the duplication of chromosomal regions produced the redundant gene loci, which give rise to the divergent mutations. Furthermore, Bridges stated that gene duplication could lead to morphological variations and speciation, and he concluded that the phenotypic differentiation of size of the eyes (Bar and Bar-double) in fruit flies was derived from the tandem duplication of a region of the polytene chromosome (see Graur and Li 2000 for review). The data accumulated from cytological observations, chromosomal analysis and whole genome sequencing are helping to define the mechanism and the significance of gene duplications.

1.2 Mechanisms of gene duplication

Today, gene duplication can be classified into tandem duplication, duplicative transposition and polyploidy or whole genome duplication. Genome

sequencing projects duplications, which has given rise to several models of the molecular level (see Hastings et al. 2009 for review).

1.2.1 Tandem duplication

Tandem duplication refers to the duplicated chromosome segments being next to each other. One example of tandem gene duplication is represented by the genes encoding ribosomal RNA (rRNA). Eukaryotic organisms need four different types of rRNA (5S, 5.8S, 18S and 28S) for translation. Each of these rRNA genes has a large number of copies. These tandem gene repeats are separated as either a locus encoding 5.8S, 18S and 28S rRNA, or encoding 5S rRNA. For example, the fruit fly (*Drosophila melanogaster*) contains 130-250 tandem duplicated copies of 18S and 28S rRNA genes; the African clawed frog (*Xenopis laevis*) has 500-760 complete sets of 5.8S, 18S and 28S rRNA that are tandemly arrayed; and humans have approximately 300 tandem gene copies of these rRNA genes (Graur and Li 2000). Another similar example to support tandem duplication is transfer RNA (tRNA). Each individual cell needs to produce many copies of tRNA for the translation of a messenger RNA (mRNA). For instance, the genome of the fruit fly has 13 duplicated groups of tRNA genes (Ohno 1970a). These great quantities of repetitive genes may be undergoing concerted evolution to maintain their structures (Zimmer et al. 1980).

Other examples of tandem gene duplication also indicate the divergence of gene loci and functions. Some gene copies become gene families such as hemoglobin, immunoglobulin and homeobox. For instance, hemoglobin is a tetrameric protein that carries the oxygen in the blood. The human hemoglobin

gene family, which is encoded by different genes, contains two α chains on chromosome 16 and two β chains on chromosome 11. The α and β family diverged from a common globin gene ancestor approximately 450-500 million years ago. In human, the α family has three functional genes and two pseudogenes; the β family has five functional genes and one pseudogene. The hemoglobin proteins are composed of different combinations of α and β chains and these genes are expressed at different developmental stages (Gregory 2005).

1.2.1.1 Unequal crossing-over and unequal exchange

Two main factors causing tandem duplications are the unequal crossing-over between homologous chromosomes at meiosis, and unequal exchange between two sister chromatids of the same chromosome at mitosis. The predominant mechanism for tandem duplication is unequal crossing-over. During the prophase of first meiosis, homologous chromosomes do not have a correct and equal amount of genetic exchange. This unequal exchange results in an uneven duplication of a gene locus on one chromatid and a deletion on the other chromatid (Ohno 1970a). Unequal exchange between two sister chromatids occurs on the same chromosome at metaphase during mitosis. The two chromatids of the same chromosomes are identical. However, the unequal exchange on the two chromatids of the same chromosome gives rise to one chromatid containing duplicated genes and the other chromatid having a deletion of that gene (Ohno 1970a).

1.2.2 Drawbacks of gene duplication

The two mechanisms of tandem duplication mentioned above provide a force for vertebrate evolution, but Ohno (1970a) indicated three main shortcomings resulting from tandem duplication. The first drawback is the unstable presence of tandemly duplicated segments of DNA, which produce further unequal exchange and unequal crossing-over. The second is that the duplicated structural genes change the gene dosage ratio with respect to other genes that are not duplicated. Finally, and most importantly, if the tandem duplication of a gene excludes the regulatory region that controls the gene, there are few opportunities to make the duplicated gene functional (Ohno 1970a).

1.2.3 Retrotransposition

Retrotransposition is the result of an RNA-based gene duplication at the stage of transcription. Because the mRNA is reversed transcribed into complementary DNA (cDNA) and randomly inserted into the genome, most duplicated genes generated by retrotransposition become junk DNA or pseudogenes. The special characteristics of duplicative retrotransposition are a lack of introns and regulatory regions, the presence of a poly (A) tract and flanking direct repeats. The expression of duplicated genes derived by retrotransposition may be caused by where the cDNA is inserted into the genome. In some cases, the cDNA insertion may interrupt the structure of a gene with the removal of stop codons and the creation of a new chaemeric protein (Brosius 1991). Moreover, because the regulatory region of a gene is not transcribed, most duplicated genes caused by retrotransposition lack the

regulatory region for transcription and become pseudogenes. Therefore, retropseudogenes have been described as junk genes and dead ends of evolution. In fact, many retropseudogenes are not detectable because the retropseudogenes may be divergent from their ancestral gene and fused with the sequences of other genes in the genome (Kaessmann et al. 2009).

1.2.4 Polyploidy

Large segmental gene duplications and doubling of entire chromosomes are also remarkable forces for making gene complexity, diversification and novel functions. Polyploidy usually occurs when an error occurs during meiosis and adds one or more additional chromosomal sets to the original chromosomes (Gregory 2005). Polyploidy has been investigated in plant genomes for a long time. Since Kuwada (1911) made the hypothesis of an ancient genome duplication in maize (*Zea mays*), several studies indicated that most the major crops, such as wheat, oats, cotton, tobacco, potato and coffee, are polyploids. Ohno (1970a) proposed that tandem duplication and polyploidy can complement each other to drive evolution. In most plants and animals, the two main types of polyploidization are autopolyploidy and allopolyploidy.

1.2.4.1 Autopolyploidy

Autopolyploidy is doubling the number of each set of chromosomes within one species (reviewed by Ohno 1970a). In many cases, autopolyploidy occurs when pairs of homologous chromosomes cannot be separated into different gametes in meiosis such that unreduced diploid gametes are formed rather than

haploid ones. Instead of a pair of bivalent homologous chromosomes, polyploids with more than two copies of homologous chromosomes produce multivalent chromosomes during the prophase stage of meiosis. Consequently, the abnormal chromosome pairing will produce triploids, tetraploids or polyploids. For instance, potatos, bananas and apples are triploid plants. In vertebrates, South American frogs (*Odontophrynus americanus*) are tetrapoids, and all the fish belonging to the family *Salmonidae* are autotetraploids (Ohno 1970a).

1.2.4.2 Allopolyploidy

Allopolyploidy is derived from the fusion of distinct chromosome sets by interspecific hybridization. Allopolyploidy may provide viable species if the parental genomes are very similar; otherwise, the organism produced by distinct species becomes sterile due to the non-pairing of chromosomes during meiosis (Gregory 2005). However, the hybridization between different genomes can create an important evolutionary force and can lead to a selective advantage in agricultural breeding and ecological adaptation (Spring 2003; Rieseberg et al. 2003). Some allopolyploid plants often provide novel phenotypes, such as pest resistance, drought tolerance, organ size and flowering time, which are not present in their ancestral diploid species.

1.2.4.3 Aneuploidy

From Ohno's conclusion (1970a), polysomy is another mechanism other than autopolyploidy and allopolyploidy for the contribution of genome duplication. Polysomy results from nondisjunction, that is a failure of homologous

chromosome separation during meiosis. The result of this incorrect separation leads to aneuploidy, which is the situation for gaining or losing an extra chromosome of the original set. Polysomy is usually deleterious. In many cases, this abnormal situation causes lethality, infertility or genetic disorders (reviewed by Trask 2002). For example, Down syndrome is caused by the presence of three copies of human chromosome 21 and the Klinefelter's syndrome is a condition caused by gaining an extra X chromosome and becoming a 47 XXY male.

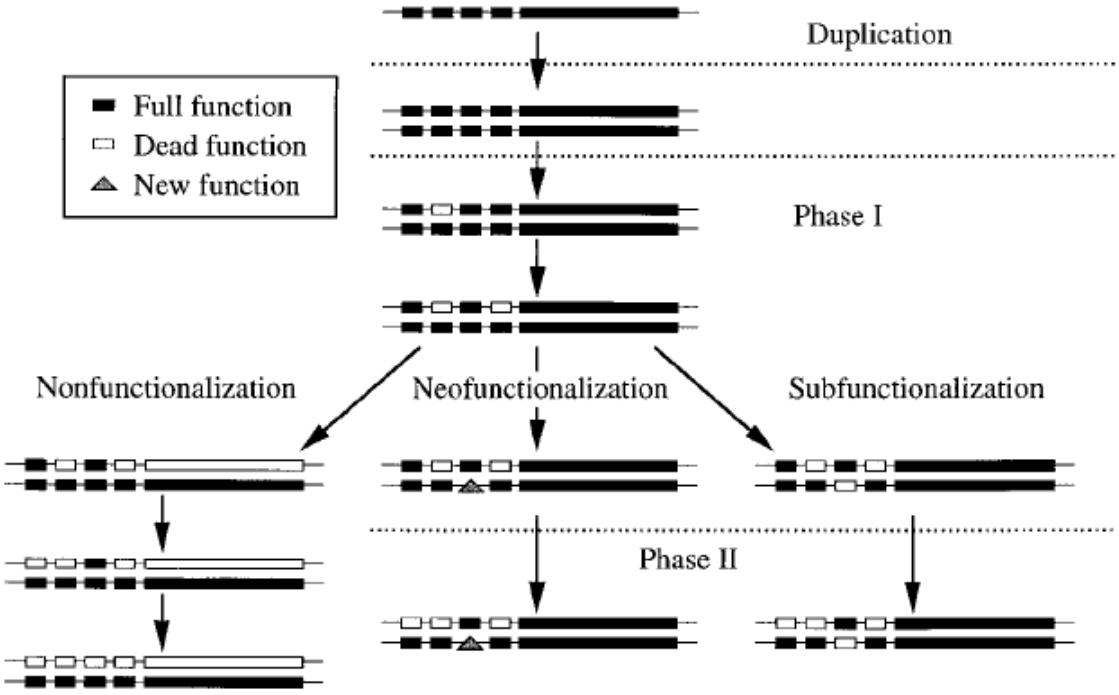
1.2.5 Shortcomings of polyploidy

Genome duplication by polyploidy is undeniably an important contribution to gene evolutionary diversification and functional divergence. One limitation of this polyploid genome duplication is the potential change in gene dosage ratio between regulator and regulated structural genes. For example, the *lac* operon of *E. coli* contains one dose of repressor and one dose of inducer in its haploid type. However, when haploids increase the dosage ratio between regulators and the structural genes to 2:2 (diploids), the absence of the inducer lactose will decrease the inactivation of the repressor and less β-galactosidase will be produced by *lac*Z (Ohno 1970a). A higher level of inducer will be required to reach the equivalent synthesis of the structural genes. Therefore, the dosage ratio resulting from polyploidy between regulator and the regulated genes may affect the level of gene expression.

1.3 The consequences of gene duplication

Gene duplication is an indispensable factor to improve the complexity and development of organisms. Observations on the duplication of single genes, chromosomal segments and entire genomes provide insight into the fate of duplicated genes. The duplication-degeneration-complementation (DDC) model identifies three different fates for duplicated genes (Force et al. 1999) (Figure 1.1): (1) Nonfunctionalization, one of the duplicated genes becomes silenced or a non-functional pseudogene by degenerative mutations; (2) Neofunctionalization, one of the redundant genes gains a novel function and is favored by natural selection; (3) Subfunctionalization, the duplicated genes have complementary expression patterns as a result of degenerative mutations in the regulatory regions.

Figure 1.1 The duplication-degeneration-complementation (DDC) model showing three potential fates of duplicate gene pairs with multiple regulatory regions (Force et al. 1999)



1.3.1 Nonfunctionalization

In the process of nonfunctionalization, after a gene or genome duplication occurs, one copy of a pair of duplicated genes loses its function and becomes a silenced pseudogene while the other one still keeps the ancestral function (Force et al. 1999). Mutations can destroy the function of protein-coding genes, and most of them are deleterious. A duplicated gene can carry and accumulate the deleterious mutations, and then become silenced or nonfunctional (Guar and Li 2000). An analysis of the fate of duplicate genes compared the rates of nucleotide substitution at replacement and silent sites using genomic data from nine eukaryotic species to study whether the different phases of evolutionary divergence affect the duplicated genes. The results of this study observed that most gene duplicates have a high rate of silencing rather than preservation. At the high rate of gene duplication, 400 - 500 redundant genes per haploid genome are expected to duplicate at least once per million years, and most of these will subsequently lose their function becoming pseudogenes (Lynch and Conery 2000). Under natural selection, these mutated genes will either be removed from the population or be retained at low frequency (Grauner and Li 2000). For example, human and mice have the same number of olfactory receptors (~1000), but the percentage of pseudogenes in human is more than 60% whereas in mice it is 20%. This is probably due to a greater selection with respect to the sense of smell in rodents compared to humans (Zhang 2003).

1.3.2 Neofunctionalization

Neofunctionalization is defined as one copy of the duplicated genes acquiring a new beneficial function while the other copy retains the original function (Force et al. 1999). Ohno stated that the new function of the duplicated gene arises from an existing gene with accumulation of mutations that change the active site of the old gene product, because he believed that “nothing in evolution is created *de novo*” (Ohno 1970a). Neofunctionalization plays an important role in gene diversity, species divergence and evolution. An example is the ribonuclease (RNAse1) gene in leaf-eating colobine monkeys such as douc langur. Because the leaf-eating monkeys digest leaves with the aid of symbiotic bacteria, to be able to digest RNA that is released from the bacteria in the foregut, the leaf-eating monkeys have a specialized RNAse1b. This enzyme comes from a duplication of RNAse1 and has accumulated several amino acid substitutions that allow it to function in the acidic environment of the foregut. Therefore, colobine monkeys have two RNAse1 genes. RNAse1a digests double stranded RNA and RNAse1b can digest the bacterial RNA in their acidic foregut. The new function of RNAse1b gene by the duplication of an RNAse1 gene and subsequent mutations suggests that fitness of the monkeys is improved by gaining more nutrition from their food and is driven by adaptive selection (Zhang et al. 2002). Another example of evolution of a new and adaptive function in duplicated genes is the opsin involved in color vision in primates. There are three opsin genes expressed in red, green and blue photoreceptor cells, respectively in the vision system of monkeys. The blue opsin is an autosomal gene, while the red and green opsins are X-linked genes. The duplication occurred after the blue

opsin and the ancestor of red and green opsin divergence. The red and green opsins have 96% amino acid identity, but only 43% with blue opsin (Yokoyama and Yokoyama 1989). The close linkage and high similarity suggested that the red and green opsins diverged by tandem gene duplication. Most New World monkeys have one blue autosomal opsin and one X-linked opsin gene (red or green). Therefore, New World monkeys have dichromatic vision. Nevertheless, Howler monkeys, a group of New World monkeys, have trichromatic vision (Jacobs et al. 1996). Howler monkeys have one blue autosomal opsin and two X-linked opsin genes. This novel function may be advantageous to Howler monkeys which now have the ability to distinguish more colors so that they can expand their range of food.

1.3.3 Subfunctionalization

Subfunctionalization describes the process whereby the two copies of the duplicated genes undergo degenerative mutations and change their expression patterns from the ancestral gene (Force et al. 1999). However, complementary degenerative mutations in different regulatory regions of the duplicated genes can control the preservation of both copies of the duplicated genes and lead to complementary expression patterns (Lynch et al. 1999). Unlike the classical model that indicates that nonfunctionalization and neofunctionalization are the main fates of duplicated genes, the DDC model suggests that the preservation of duplicate genes is due to the fixation of complementary degenerative mutations in promoter regions rather than by the fixation of new beneficial mutations in coding regions (Lynch et al.1999). One example which investigated

subfunctionalization is the zebrafish (*Danio rerio*) engrailed genes. Zebrafish has four engrailed genes: eng1, eng1b, eng2 and eng3. Two pairs of engrailed genes, eng1/eng1b and eng2/eng3, were produced by a whole genome duplication. The engrailed-1 gene family provides a good example of subfunctionalization. From linkage analysis and syntenic comparisons the engrailed-1 gene family members in zebrafish, eng1/eng1b were found to be syntenic with En1 of mammals. En1 is therefore an outgroup of eng1 and eng1b and can be used to infer the ancestral expression domains of eng1 and eng1b. Zebrafish has different expression patterns for the engrailed1 genes; eng1 expression is in the pectoral appendage bud and the eng1b is in a specific set of hindbrain and spinal neurons. However, in mice and chickens, En1 is expressed in all of these tissues. This observation predicted that the eng1 and eng1b in zebrafish are derived from the duplication of an En1 like gene and have been retained due to subfunctionalization (Lynch et al. 1999).

1.4 Evidence for genome duplication in vertebrates (2R/3R/4R)

From the early studies of genome size and isozyme patterns, Ohno hypothesized that two rounds of genome duplication occurred in the early vertebrate evolution timeline (Ohno 1970a). The first round occurred before the cephalochordates and vertebrates diverged, and the second round was predicted to have taken place in the jawless fish or amphibian lineage (Ohno 1970a). Hox genes provide a good example to illustrate the two rounds of the genome duplication hypothesis. The cephalochordates including amphioxus (*Branchiostoma lanceolatum*), only have a single Hox cluster whereas the lobe-

finned fishes, amphibians, reptiles, birds and mammals have four Hox clusters (Holland and Garcia Fernandez 1996; Holland 1997; Larhammer et al. 2002); and a recent study showed that the human Hox gene family was quadruplicated (Lemon and McGinnis 2006). However, the refined, debated and controversial views of Ohno's hypothesis in past decades force the development and understanding of the evidence for and against the 2R hypothesis. Holland et al. (1994) proposed that the first round genome duplication occurred after the divergence of cephalochordates, and the second one after the divergence of jawless fish. A prediction of the 2R hypothesis states that hypothetical paralogs, A-D, derived from two rounds of genome duplication should have the topology (AB)(CD), and similar divergence times. However, 70.9% of human four-member gene families and clusters showed topologies A(BCD), which is inconsistent with two rounds of genome duplication in vertebrates (Friedman and Hughes 2001).

Another round of genome duplication (3R) was proposed have occurred at the base of teleost fishes (Talyor et al. 2001). The hypothesis suggested that an additional genome duplication event occurred in the ray-finned fish lineage before the divergence of most teleosts (Amore et al. 1998). In this case, there should be a "1-4-8 rule", meaning that for every gene observed in an organism that is an outgroup to vertebrates there should be four genes in tetrapods and eight in teleosts. However, 7 as opposed to 8 Hox gene clusters were identified in zebrafish, medaka (*Oryzias latipes*), and pufferfish (*Sphoeroides nephelus* and *Takifugu rubripes*) (Amore et al. 1998, 2004; Málaga-Trillo and Meyer 2001; Prohaska and Stadler 2004). Several studies suggested that the complementary

pattern of duplicated Hox genes shows the evidence of a post-duplication Hox gene loss in teleosts. The example of Hox gene clusters may indicate that the genome duplication occurred in ray-finned fish before the divergence of zebrafish, medaka and pufferfish (Taylor and Raes 2004). However, it has been suggested that there are not enough studies to confirm the fish specific 3R genome duplication hypothesis (Vandepoele et al. 2004). The generally accepted view is that two rounds of Hox chromosome duplications occurred before the divergence of ray-finned fish and lobe-finned fish, and the additional whole genome duplication took place in the ancestor of the ray-finned fish (Amore et al. 1998).

The salmonids underwent an additional whole genome duplication event (4R) for the following reasons. Salmonids were found to have higher DNA contents and chromosome numbers than other members of teleosts (Ohno 1970a). The genome of the common ancestor of salmonids was doubled by autopolyploidy between 25 and 100 million years ago (MYA) (Ohno 1970a; Allendorf and Thorgaard 1984). The presence of multivalents during meiosis and a high occurrence of duplicated enzyme loci suggested that the salmonids are autotetraploids (Allendorf and Thorgaard 1984). Hox genes provide further evidence for the 4R duplication in salmonid fish. Fourteen Hox clusters were observed in the Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). These observed Hox clusters from salmonids are consistent with the 1-4-7-14 rule of genome duplication (Moghadam et al. 2005a; Moghadam et al. 2005b).

All salmonid fish are considered to be autotetraploids, which have progressed toward diploidization in different degrees. The diploidization in salmonids is driven by Robertsonian fusions, which may be related to the selection for new metacentric chromosomes rather than acrocentic chromosomes (Ohno et al.1969; Ohno 1970b). The salmonid fish are thought to have originated from a diploid ancestor with 48 acrocentric chromosomes and the derived tetraploids would have had 96 acrocentric chromosomes (Ohno 1970a). Because of the Robertsonian fusions of acrocentric chromosomes made many changes in the karyotypes of salmonid fish have occurred (Phillips and Rab 2001).

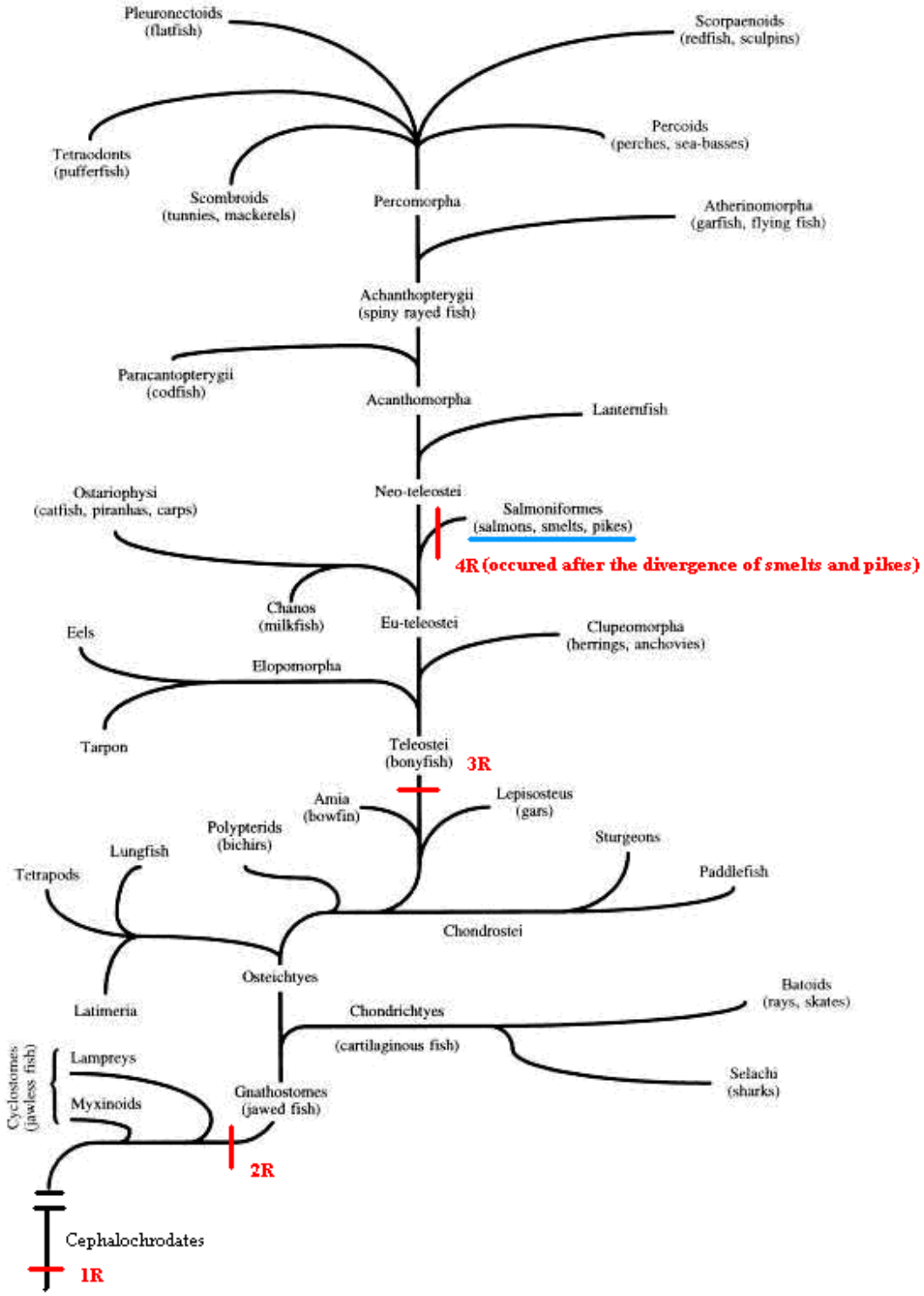
1.5 Evolution of fish

1.5.1 Fish species evolution

The evolution of fishes provides evidence to support the extensive polyploidy among the vertebrates. The general classification of fishes includes jawless fishes, cartilaginous fishes, lungfishes, chondrosteans and teleosts (Gregory 2005) (Figure1.2). Jawless fishes (lampreys and hagfishes) represent the class *Agnatha* in the development of early vertebrates. Cartilaginous fishes include sharks, rays and skates. Bony fishes appeared in the middle Devonian period (400 MYA) and diverged into two distinct groups: lobe-finned fish (*Sarcopterygii*) and ray-finned fish (*Actinopterygii*). Lobe-finned fishes (*Sarcopterygii*) include lungfishes and the coelacanth. Chondrostean fishes form a group of ray-finned fish (*Actinopterygii*). It has been suggested that sturgeons

and American paddlefishes have a polyploid origin (Ohno et al. 1969; Dingerkus and Howell 1976).

Figure 1.2 **The phylogenetic tree of fishes (Huss 1995).** The red markers indicate the 1R, 2R, 3R and 4R genome duplications. The blue underline indicates the species of Salmoniformes including Atlantic salmon and rainbow smelt.



1.5.2 Teleost gene and genome duplication

Most of the modern ray-finned fish (*Actinopterygii*) are bony fish and provide evidence of more recent gene and genome duplications. For example, Ohno believed that polyploidy is very important in the fish family *Cyprinidae* (Ohno 1970a). There is evidence that goldfish and carp from the family *Cyprinidae* are tetraploid species compared to other diploid members in that family. These species have 104 chromosomes whereas two barb species in the family *Cyprinidae*, *Barbus tetrazona* and *Barbus jasciatus,* have been identified as diploids with chromosome numbers of 50 and 52 (Ohno 1970a).

There is considerable evidence to support the hypothesis that there were genome duplications in teleosts. For example, there are 14 copies of Hox gene clusters in Atlantic salmon and rainbow trout compared to 7 copies in zebrafish, medaka and pufferfish, 4 copies in mammals and one copy in most invertebrates (Prohaska and Stadler 2004; Postlethwait et al. 2000; Moghadam et al. 2005a). In addition, Jaillon et al. (2004) investigated the syntenic map between the freshwater pufferfish (*Tetraodon nigroviridis*) and human. The test revealed 76% orthologues between pufferfish and human with ~80% of the orthologues following the 2:1 ratio between pufferfish and human. That is, two chromosomal regions in pufferfish match one in a human chromosome. This is called “double conserved synteny”. The distribution of gene duplication in pufferfish chromosomes reveals the ancient genome duplication in the ray-finned fish lineage and suggests that the mechanism of eukaryotic genome duplication involves massive gene loss and local gene shuffling.

As previous studies revealed, the evidence of a whole genome duplication is predicted to be found in more than 20,000 species of living teleost fish, and to have occurred close to the origin of the divergence of teleosts (Hoegg et al. 2004; Taylor et al. 2003). Today, five teleost genomes have been sequenced and used to study duplication events. They are zebrafish, stickleback (*Gasterosteus aculeatus*), medaka, tetraodon and takifugu (*Takifugu rubripes*) (Hubbard et al. 2009). As more teleost genomes are sequenced, the sequenced genome data will provide a rich source to study the gene duplication and morphological and genetic evolution, thereby resolving the mechanism and consequences of whole genome duplications in teleosts. However, another family of teleosts, *Salmonidae*, provides an excellent example of autotetraploidization, and a good model system for studying more recent gene and genome duplication events.

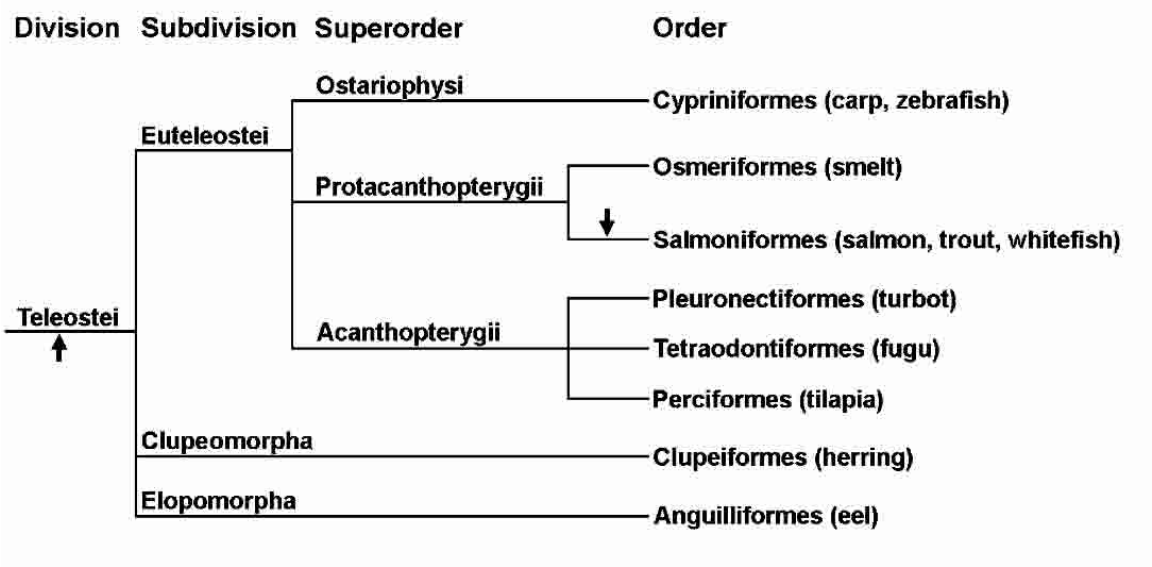
1.5.3 Salmonidae and Osmeridae

1.5.3.1 Introduction to Salmonidea and Osmeridae

The *Salmonidae* family is native to the northern hemisphere and represents a separate evolutionary lineage from other teleosts. The members in this family have been studied broadly due to their commercial importance. In the past 20 years, extensive scientific research has been carried out on salmonids in the fields of ecology, behaviour, physiology and genetics (Thorgaard et al. 2002).The salmonid fish are classified into 9 genera with approximately 68 species and 3 subfamilies (Nelson 2006). The three subfamilies include *Coregoninae* (whitefishes and ciscoes), *Thymallinae* (graylings) and *Salmoninae* (lenok, huchen, trout, charr and salmon). The phylogenetic tree shows that the

diploid species in the Osmeriformes separated before the genome duplication in the ancestor of the salmonids (Rise et al. 2004) (Figure 1.3).

Figure 1.3 Phylogenetic tree of teleosts. Phylogenetic tree, based on morphological characters, showing evolutionary relationships among teleosts and other fish orders with genome projects (Nelson 1994). The arrows denote the genome duplication events.



1.5.3.2 Gene and genome duplication in salmonids and rainbow smelt

The salmonid genome duplication is the most recent in a series of genome duplications in teleosts (Koop and Davidson 2008). The salmonids appear to have evolved by autotetraploidization from a common ancestor between 25-100 MYA (Ohno 1970a; Allendorf and Thorgaard 1984). There is a significant amount of evidence to support the autotetraploidization in the salmonids. First, the genome size of salmonids is 3.2 pg, which is more than double that of the *Osmeridea* (0.69 pg) (Gregory 2005) (http://www.genomesize.com/). All sebsites mentioned in this thesis can be found in Appendix 1. Second, multivalent chromosomes are formed in meiosis and there is tetrasomic inheritance found in salmonid species (Allendorf and Danzmann 1997). Third, the karyotypes of salmonids reveal 100 chromosome arms compared to that is seen in their osmerid relatives (50-56) (Mank and Avise 2006). Finally, the high incidence of the duplicated enzyme loci has been investigated in salmonids including Hox, MHC and growth hormone genes (Moghadam et al. 2005a; Hoegg and Meyer 2005; McKay et al. 2004; Shiina et al. 2005). Considering the species number, the recent genome duplication event and the rich resources of biological data available for salmonid fish, they are excellent model organisms for studying evolutionary genomics, comparative genomics, fates of gene duplication and genetic architecture, toxicology, ecology, comparative immunology, diseases, physiology and nutrition (Thorgaard et al. 2002; Koop and Davidson 2008).

1.5.3.3 The fate of duplicated gene loci in salmonids and rainbow smelt

The evolution of the duplicated gene loci in diploidized autotetraploid salmonids can be described in three stages. First, the autotetraploid salmonid has four doses of every gene (tetrasomy); second, the tetrasomy will be changed into two independent pairs (disomy) by diploidization in meiosis; finally, each pair of the duplicated gene loci will become functionally divergent with variable degrees of expression and tissue specific patterns. However, many of the duplicated gene loci that underwent the diploidization after tetraploidization might be silenced or lost without the advantage of positive selection (Ohno 1970a). For studying the fate of gene loci duplication in salmonids and the diploidization process, it is necessary to study one of the diploid relatives. I have chosen the diploid species rainbow smelt (*Osmerus mordax*) as an outgroup for comparisons with salmonids, in particular the Atlantic salmon (Ohno 1970a).

1.6 Isozymes and gene duplication

1.6.1 Molecular basis of isozymes

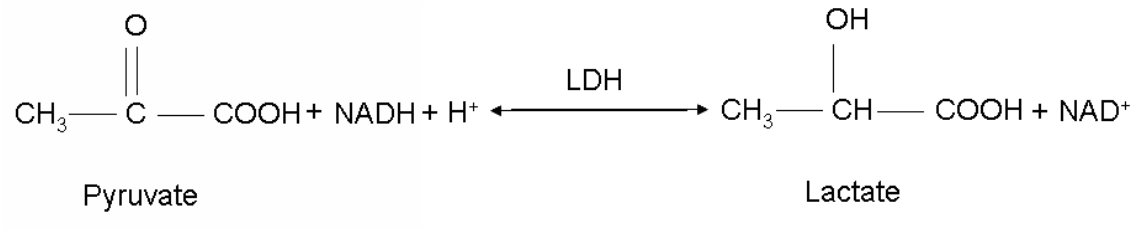
Fifty-two years ago, Hunter and Markert (1957) first discovered enzyme heterogeneity. They showed that the esterase-active proteins could be separated into different bands (zymograms) by starch gel electrophoresis followed by histochemical staining. The zymograms revealed the enzyme heterogeneity and the diverse substrate specificity of esterases in specific tissues of mouse. Two years later, Markert and Møller (1959) first defined the term “isozyme”, that is, enzymes having different molecular forms but catalyzing the same chemical reaction. Based on the technique of starch gel electrophoresis, Markert and

Møller developed a tetrazolium staining method to study the isozymes. Lactate dehydrogenase (LDH) was the one of first examples of isozymes. They suggested that LDH in many organisms has five distinct forms, which change during the different stages of tissue development (Markert and Møller 1959).

1.6.2 LDH function

LDH is a cytosolic enzyme that catalyzes the interconversion of pyruvate and lactate using the coenzyme nicotinamide adenine dinucleotide (NAD) (Figure1.4). There are two LDH families based on the stereochemical forms of lactate (D or L). L-LDHs belong to L-specific NAD-dependent dehydrogenases and D-LDHs belong to the D-isomer specific 2-hydroxy acid dehydrogenases and the FAD-binding oxidoreductase/transferase type 4 family D-LDH (Cristescu et al. 2008). Although the L-LDHs and D-LDHs have similar functions they are not related evolutionarily (Kochhar et al. 1992; Vinals et al. 1993). The L-LDH enzyme family has been extensively studied with respect to structure, function, kinetics and evolution in vertebrates. All the LDHs discussed in this thesis belong to the L-LDH family.

Figure 1.4 The interconversion of pyruvate and lactate by LDH using the coenzyme nicotinamide adenine dinucleotide (NAD)



1.6.3 LDH gene control

LDH functions as a tetramer (Appella and Markert 1961). It may be a homotetramer composed of four identical subunits or a heterotetramer with different protein subunits. An experiment showed that the tetramer LDH containing two different protein subunits, A and B, leads to five isozymes by the random tetrameric association: LDH-5=A_4, LDH-4=A_3B_1, LDH-3=A_2B_2, LDH-2=A_1B_3 and LDH-1=B_4 (Markert 1963). The measurement of total amino acid composition confirmed that two proteins in LDH-1 (B_4) and LDH-5 (A_4) were different and it was concluded that the A and B protein subunits are encoded by different genes (Markert 1963). Therefore, the homotetramer A_4 is encoded by gene locus A (LDH-A), and B_4 is encoded by gene locus B (LDH-B) in vertebrates. However, LDH isozymes in fishes do not have restricted numbers of isozymes and only a few species showed five tetramers composed of A and B subunits (Markert et al. 1975). Markert (1968) suggested that the three heterotetramers (A_3B_1, A_2B_2, A_1B_3) may be encoded by various allelic genes either from locus A or B. Moreover, several mutant alleles were found at the A and B loci in human LDH genes (Boyer et al. 1963; Nance et al. 1963). A mutant allele was observed at the B locus in mouse by breeding experiments, which indicated that the genetic variance was inherited as an autosomal codominant gene (Shaw and Barto 1963).

Beyond the five tetrameric LDH isozymes composed of A and B subunits, Blanco and Zinkham (1963) first discovered a sixth LDH band (X-band) in the sperm of many mammals by starch gel electrophoresis. This X-band is

composed of a third homotetramer subunit, C_4, distinct from the A and B subunits. The subunit C was defined as LDH-X (or LDH-C) (Zinkham 1968). C subunits encoded by gene loci other than those for A and B have been investigated in many mammals, birds, amphibians and fishes (Markert et al. 1975).

1.6.4 Kinetics and tissue specificity of LDH

Many investigations indicated that isozymes have different kinetic properties in addition to tissue specificity. For example, the kinetics of LDHs from various cell types in mouse and human revealed that the K_m and V_{max} depended on the tissue in which they are expressed and they appear to be related to the heterogeneity of cellular metabolism. The kinetic study was based on lactate as the substrate. The K_m for LDH from skeletal muscle fibres was 10.4 - 12.5 mM; in hepatocytes it was in range of 14.3 - 16.7 mM and in cardiac muscle fibres it was 13.4 mM. The V_{max} was 59-68 μ moles hydrogen equivalents/cm^3 cytoplasm/min units in skeletal and cardiac muscle; 102-110 units in hepatocytes; 29 units for parotid gland cells and 62-65 for gastric pariental cells and oocytes (Nakae and Stoward 1994). The LDH kinetic properties characterized from different tissues are consistent with an earlier study (Cahn et al. 1962). The anaerobic tissue (e.g. skeletal muscle) specific LDH has a lower K_m for lactate compared to what is found in aerobic tissue (e.g. liver and heart). The change in kinetic parameters in isozymes can be considered as “partial neofunctionalization”. The studies of tissue specificity of LDH patterns and their encoded genes will provide the information to understand the biological significance of isozymes.

LDH-A is commonly found in anaerobic tissue such as skeletal muscle and LDH-B is mostly expressed in aerobic tissue such as heart muscle, liver and brain in most vertebrates (Cahn et al. 1962). The third LDH subunit, which gave rise to an X-band by starch gel electrophoresis, was first found in primary spermatocytes of mammals and pigeons (Blanco and Zinkham 1963; Zinkham et al. 1969). Zinkham and his colleagues indicated that the LDH-B and LDH-C gene loci in mammals and pigeons are homologous as their products have similar amino acid compositions (Zinkham et al. 1969). Several studies revealed that there is an LDH-C that is active in eyes in most fish. A retinal-specific LDH-C was observed, which is synthesized in the ellipsoid region of the photoreceptor cells (Whitt 1970; Whitt and Booth 1970). However, a LDH-C with a different net charge pattern in fish from the orders of Cypriniformes and Gadiformes has a liver specific expression. For example, a liver specific LDH-C with a cathodal mobility was observed in Atlantic cod and hornyhead chub (Whitt et al. 1975). From a study of expression patterns of LDH-C in different species of fish, it was observed that there is either an anodal mobility LDH predominant in eyes or a cathodal mobility LDH predominant in liver. Support for the hypothesis that the retinal specific LDH-C is closely related to the LDH-B rather than LDH-A comes from studies involving immunochemistry, kinetics and physical properties. These results suggested that the LDH-C gene arose by a LDH-B gene duplication event (Whitt 1969; Sensabaugh and Kaplan 1972). The change in LDH-C tissue specificity is considered an example of subfunctionalizaiton of one of the products of the LDH-B gene duplicates.

1.6.5 Evolution of LDH

It was proposed that the vertebrate gene loci encoding LDH were derived by gene duplications from a single ancestral LDH gene and the accumulation of mutations (Markert et al. 1975). This hypothesis of the evolutionary origin and divergence of vertebrate LDHs described an original ancestral LDH gene giving rise to the LDH-A and LDH-B gene loci by gene duplication, and then a second duplication occurred involving the LDH-B gene from which the LDH-C was derived (Whitt et al. 1975; Markert et al. 1975). Unfortunately, how the LDH gene duplications are related to the whole genome duplication that has been predicted at the base of vertebrate evolution is unkown. However, given that all species after the proposed 2R duplication (see Figure 1.2) have LDH-A and LDH-B whereas lamprey has a single LDH, it is tempting to speculate that the 2R whole genome duplication resulted in LDH-A and LDH-B.

1.6.5.1 Support for the Markert et al. (1975) LDH evolution model

Several results support the hypothesis that the LDH-A and LDH-B were derived from an ancestral LDH gene; for example, 1) the association of hybrids of A and B subunits make functional tetramers both in vivo and in vitro when the homopolymers are from distant vertebrates (Markert 1963), and 2) the identical amino acid sequence of the dodecapeptide at the active site of A and B subunits (Taylor et al. 1973).

The LDH-C gene is obviously related to the LDH-A and LDH-B genes, but the nature of the relationship among them has been controversial. Several physical, kinetic, amino acid composition and immunochemical results suggested

that the LDH-C subunit is closer to the LDH-B subunit than the LDH-A subunit in vertebrates (Markert et al. 1975). The result from an immunochemical study stated that the anti-B antibodies only precipitated LDH-B and LDH-C but not LDH-A in sea trout (*Cynoscion regalis*). The anti-A antibodies precipitated LDH-A but not other LDH isozymes. The immunochemical cross-reactions indicated that a higher similarity existed between the LDH-B and LDH-C than between either of these with LDH-A (Holmes 1969). The comparison of amino acid compositional relatedness showed a difference between LDH-A and LDH-B proteins (Markert 1963). Zinkham and his colleagues indicated that the LDH-B and LDH-C gene loci are closely linked in pigeon, and these two loci coded for proteins with similar amino acid compositions in both mammals and birds. These results suggested that the LDH-B and LDH-C genes separated before the divergence of mammals and pigeons (Zinkham et al. 1969).

The prediction of an early LDH gene duplication in fishes follows the evolution fish species from ancestral Agnatha to the advanced teleosts (Whitt et al. 1975) (Figure1.5). For example, sea lamprey (*Petromyzon marinus*) only has a single LDH gene. A phylogenetic analysis suggested that the lamprey LDH is closer to the LDH-A of other vertebrates and the single LDH locus is the result of the loss of LDH-B before the LDH-C divergence (Stock and Whitt 1992). The single LDH in sea lamprey is consistent with the evolution of fish, as sea lamprey is a representative of an early branch in vertebrate evolution. The occurrence of two LDH genes was investigated in cartilaginous fish (sharks, rays and skates) and both an LDH-A and an LDH-B were found. Bony fishes of the Osteichthyes

class are a more advanced level of fish evolution and all teleosts have three LDH genes: LDH-A, LDH-B and LDH-C. The LDH-C genes of bony fishes are much more like LDH-B than LDH-A. The kinetic, physical and immunochemical evidence suggested that the teleost LDH-C gene is derived from LDH-B by a single gene duplication (Markert et al. 1975). A phylogenetic analysis using cDNA sequences revealed that LDH-C was derived from LDH-A by tandem duplication and that LDH-B was separated from the group containing LDH-A and LDH-C in mammals, whereas LDH-C genes was derived from independent tandem gene duplications from LDH-B genes after the LDH-A duplication in pigeons, frogs and fishes (Li et al. 2002; Mannen et al. 1997) (Figure 1.6).

Figure 1.5 The evolution of LDH gene in fish as proposed by Whitt et al. (1975)

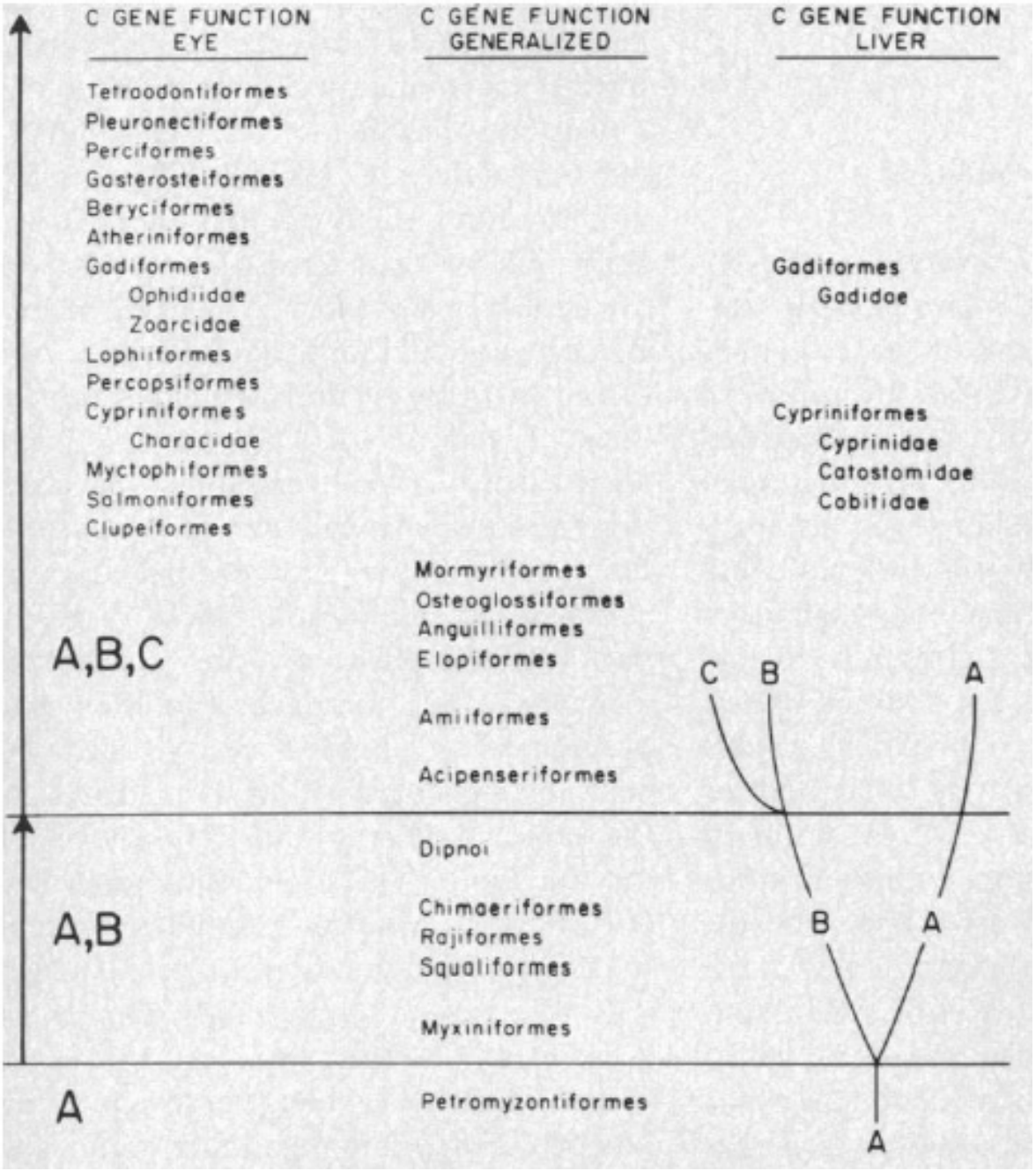
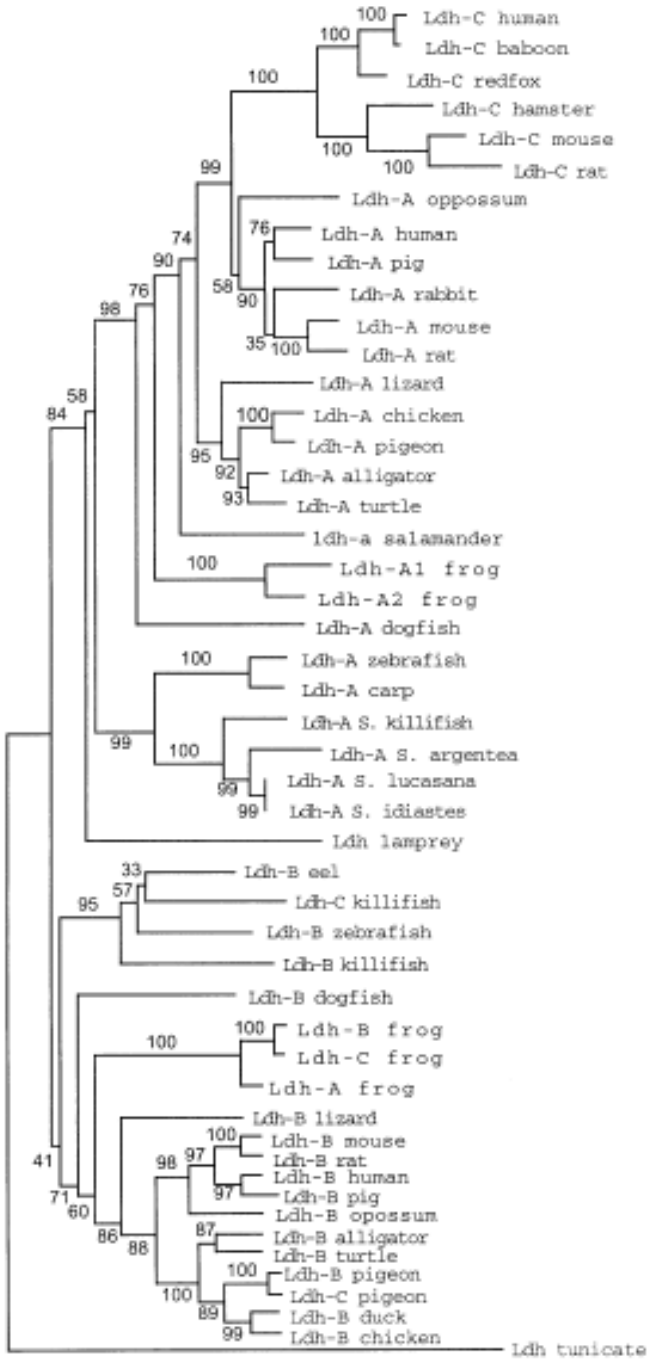


Figure 1.6 The Neighbor-Joining (NJ) tree of LDH from nucleotide sequences. Tunicate LDH gene was suggested to be an outgroup for vertebrate LDH genes (Li et al. 2002).



1.6.5.2 Alternative hypotheses for the evolution of vertebrate LDH genes

Li et al. (1983) proposed that the LDH-C in mammals is the ancestral gene rather than LDH-A. This hypothesis was based on pairwise comparisons of LDH amino acid sequences from dogfish LDH-A, chicken LDH-A and LDH-B, pig LDH-A and LDH-B, and mouse and rat LDH-C isozymes. Another analysis of LDH evolutionary relationships used amino acid compositions of the LDH-C from Atlantic cod and LDHs that had been sequenced. It suggested that the first gene duplication occurred on LDH-C and then a further gene duplication produced the LDH-A and LDH-B genes (Rehse and Davidson 1986). Moreover, several investigations reported that mammalian LDH-C arose before the divergence of LDH-A and LDH-B in vertebrates. However, these studies showed that LDH-B and LDH-C are most closely related in killifish and frog (Tsuiji et al. 1994; Tsoi and Li 1994) (Figure1.7 and 1.8).

Figure 1.7 **The maximum parsimony evolutionary tree of LDH subunits from amino acid sequences**. The numbers on the branches are nucleotide substitutions required to amino acid replacement. Bootstrap shows as asterisks (99-100%) or plus signs (80-89%). The diamond indicated gene duplication events (Tsuji et al. 1994).

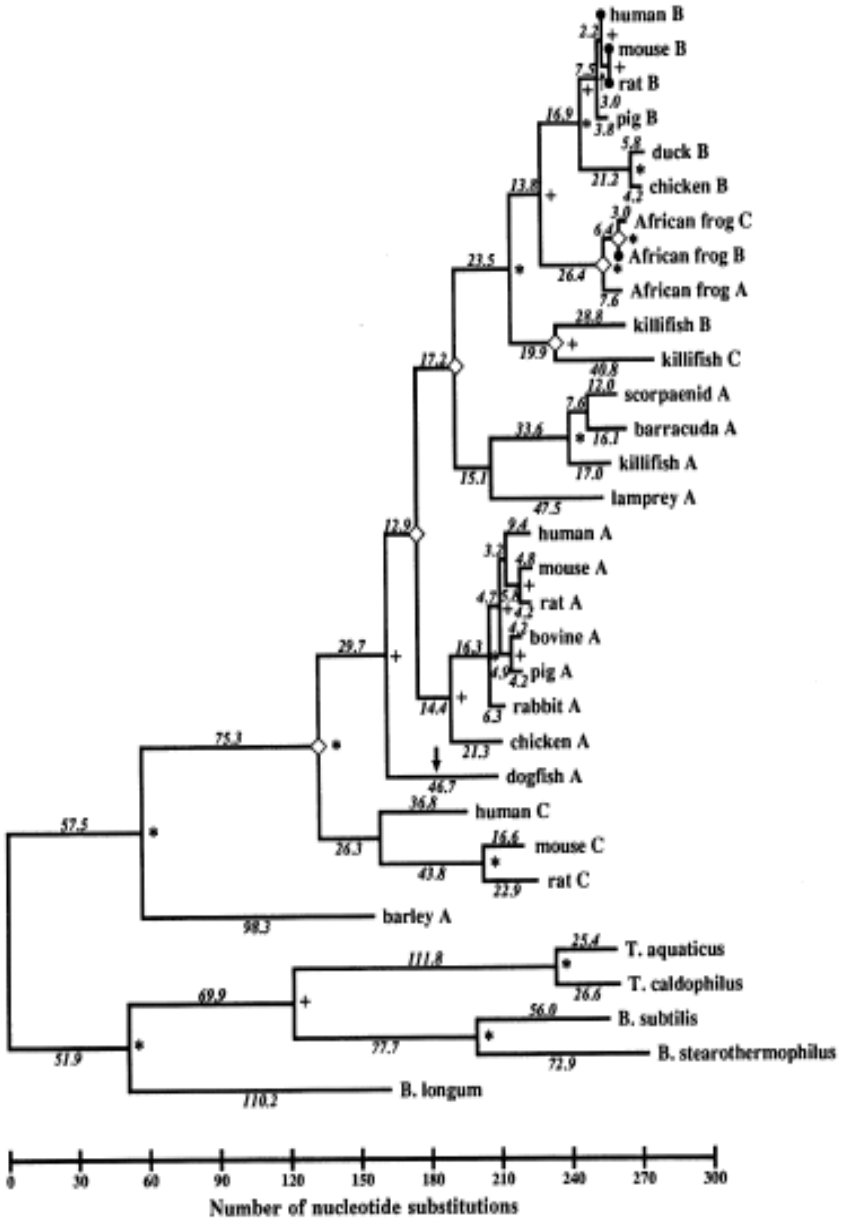
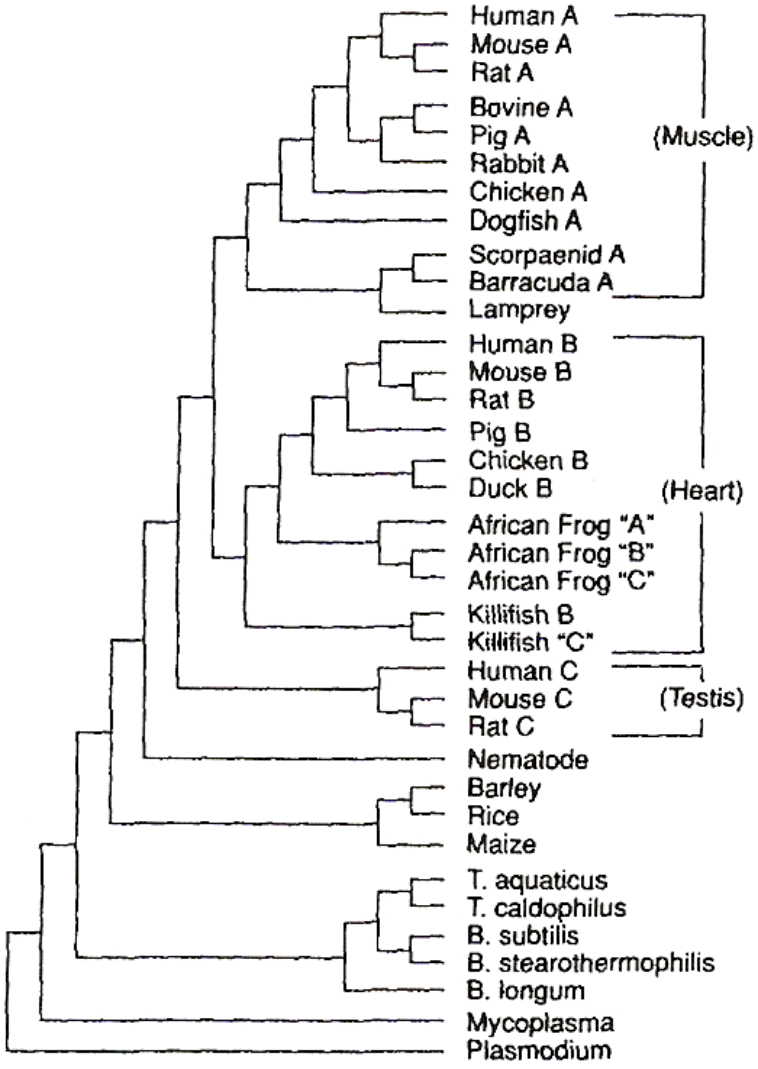


Figure 1.8 The UPGMA Evolutionary tree of LDH from amino acid sequences (Tsoi and Li 1994)



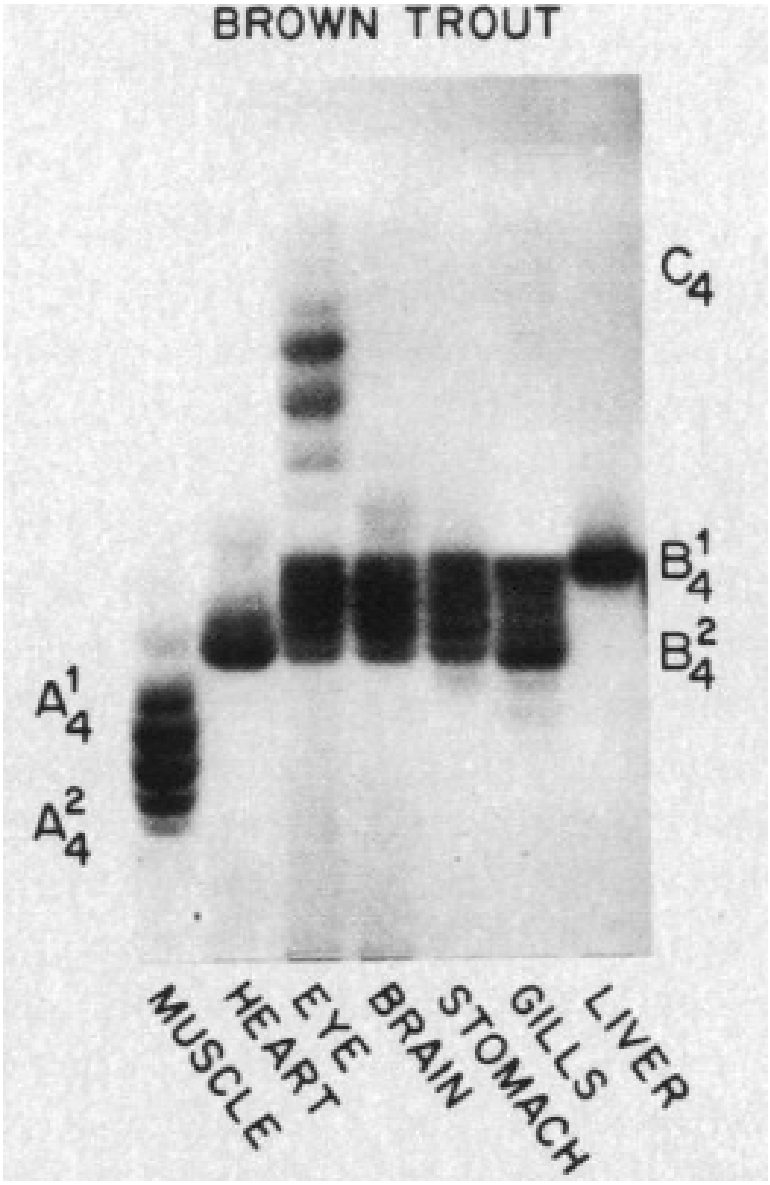
1.7 LDH gene duplication in Salmonids

1.7.1 LDH in salmonids

In salmonids, both LDH-A and LDH-B genes have been duplicated. An examination of skeletal muscle tissue indicated that A_4 LDH (also called M_4 LDH, LDH-A or LDH-5) in salmonids was homologous to the higher vertebrate A_4 LDH (Bailey and Wilson 1968). A further experiment reported that there were two LDH-A subunits, which are catalytically equivalent and have arisen by a gene duplication in salmonids (Lim and Bailey 1977). In addition, a biochemical and genetic study of B_4 LDH (also called H_4 LDH, LDH-B or LDH-1) showed that there are two LDH-B subunits, which were produced by the salmonid genome duplication. The two LDH-Bs from salmonids were immunologically related to the H subunit of higher vertebrates (Bailey and Wilson 1968). Immunochemical tests of LDH indicated that the duplicated A subunits encoded by LDH-A genes are A_1 and A_2, and these two subunits are expressed equally in skeletal muscle (Holmes and Markert 1969). However, the expression of LDH-B is different in certain tissues. The regulatory mechanism distinguishes two B subunits (B_1 and B_2) expressed in different tissues by starch gel electrophoresis analysis (Markert et al. 1975) (Figure1.9). In brown trout, B_1 subunit predominated in liver and B_2 in heart. In brain, both B_1 and B_2 were expressed equally (Markert et al. 1975). Moreover, the A and B subunits do not interact and make the A-B subunits containing tetramers in salmonid fish (Markert et al. 1975). The pattern of LDH expression in different tissues from brown trout showed that an extra homoteteramer band (C_4) was only expressed in eyes and that it is distinct from

the A and B loci (Markert et al. 1975) (Figure1.9). Markert and his colleagues believed that two LDH-C genes were to be expected considering the genome duplication in salmonids. However, there is no evidence for duplicated LDH-C genes in salmonids (Markert et al. 1975). The prediction of the duplicated LDH-C gene suggests that the duplicated LDH-C gene may have been silenced or lost by nonfunctionalization during the evolution of salmonids (Markert et al. 1975).

Figure 1.9 **Expression of LDH genes in brown trout and brook trout.** The duplicated A gene (A1 and A2) are equally expressed in muscle; however, duplicated B genes (B1 and B2) are differently regulated (in liver and heart) (taken from Markert et al. 1975).



Recent work in the Davidson lab identified five LDH genes corresponding to two LDH-As, two LDH-Bs and one LDH-C in the Atlantic salmon EST database. BAC clones containing each of the five salmon LDHs have been identified and sequenced (Lubieniecki et al. in preparation). RT-PCR gene expression patterns show that LDH-A_1 and LDH-A_2 have a strong muscle expression. LDH-B_1 has a high level expression in liver and a low level expression in heart. LDH-B_2 has an opposite expression compared to LDH-B_1 in heart. LDH-B_2 has a lower level expression in liver but a higher level expression in heart. LDH-C is expressed in both brain and eye tissues. A comparison of the amino acid sequences of the Atlantic salmon and rainbow trout LDHs with one another and LDHs from other vertebrates indicates that the LDH-C is derived from an LDH-B as has been observed in other teleosts (Lubieniecki et al. in preparation).

1.7.2 LDH in rainbow smelt

The diploid rainbow smelt serves as a outgroup reference species to investigate the fate of LDH gene duplicates in salmonids. Rainbow smelt has A, B and C genes for LDH (Markert et al. 1975). The subunits encoded by LDH genes in rainbow smelt do not interact, and they are seen as homotetramer A_4, B_4 and C_4 (Whitt et al. 1975).

1.7.3 Genomic resources for rainbow smelt

The genome size of rainbow smelt is 0.69 pg (Gregory 2005) (http://www.genomesize.com/). A Bacterial Artificial Chromosome (BAC) library

(CHORI-74) was prepared by the Children's Hospital Oakland Research Institute (CHORI) Oakland, CA, USA. The BAC library contains 52,410 clones with an average clone insert size of 146 kb, giving an 11-fold coverage of the rainbow smelt genome (Schalburg et al. 2008). The rainbow smelt Expressed Sequence Tags (EST) clustering database at the University of Victoria provides 36758 expressed sequence tags and 16063 transcripts, which joined into 9044 contigs except singletons transcripts from EST consensus sequences (http://web.uvic.ca/grasp/).

1.8 Purpose of thesis

The purpose of this project is to characterize the LDH genes from rainbow smelt and to compare them to the five LDH genes from Atlantic salmon. First, I will be able to provide a reference to study the effect of the salmonid whole genome duplication event superimposed upon more ancient gene duplications by investigating the genes encoding the LDH isozymes in rainbow smelt. Second, the comparison of the LDH gene family in Atlantic salmon and smelt will help us understand how duplicated genes are maintained and evolve under the duplication-divergence-complementation model. Finally, the results of this thesis will contribute to the characterization of the genome duplication event in salmonids, which is an excellent model system for studying the importance of genome duplications in evolution.

CHAPTER 2: MATERIALS AND METHODS

2.1 Rainbow smelt LDH probes and design of gene specific primers design

2.1.1 PCR protocol

The oligonucleotide probes and primers were designed to have at least a 50% GC content and an annealing temperature of 65°C. The probes for each of the three LDH genes were designed as 40-mers to increase their specificity. The reverse primers were designed as 20-mers. In addition, the 40-mer probes of LDH-A and LDH-B were used as 5' forward primers to amplify the specific LDH gene by PCR. However, the LDH-C 40-mer probe was designed separately from its gene specific primers.

The PCR conditions for checking the primer specificity of each LDH gene were established using a T3 Thermocycler (Biometra). The PCR protocol was carried out as following steps: 95°C for 5 min; 35 cycles of 95°C for 45 sec, Tm for 45 sec and 72°C for 2 min; and a final extension at 72°C for 10 min. Each 10 µL of PCR mix was composed of 0.5 µL of 10 mM forward primer and 0.5 µL of 10 mM reverse primer; 1 µL of 2 mM dNTPs; 1 µL of 10 x PCR buffer (Invitrogen); 0.15 µL of 5 U/µL Taq polymerase (Qiagen); 6.35 µL dH_2O and 0.5 µL of 100 ng/µL template. PCR products were electrophoresed on a 1% agarose gel containing 1 X TAE and ethidium bromide (0.5 µg/mL), and visualized using a UV trans-illuminator (Alpha Innotech).

2.1.2 LDH-A

The LDH-A gene specific oligonucloetide probe and reverse primer were designed based on predicted exon 2 and exon 3 of the rainbow smelt EST sequence from contig 4546. The 40-mer forward probe (primer) based on exon 2 was

5'-GTGTGATGAGCTGGCCCTGGTTGACGTGATGGTGGACAAG-3'. The 20-mer reverse primer based on exon 3 was 5'-ACTTGACGATGTTGGGGATG-3'. The annealing temperature for the primers was 65°C.

2.1.3 LDH-B

Because the rainbow smelt LDH-B gene is not available in the EST database, the 40-mer oligonucleotide forward probe (primer) and 20-mer reverse primer were designed based on a salmonid LDH-B1 specific region, which was selected from the alignment of the LDH-B and LDH-C coding sequences from Atlantic salmon and rainbow trout using ClustalX (Larkin et al. 2007). The probe based on exon 5 of salmonid LDH-B1 was

5'-TCAGCGTAGCTGGAGTCAACCTGCAGAAGCTGAACCCAGAG-3'

and the reverse primer based on exon 6 was

5'-TGAGATCAGCCACACTCAGG-3'. The annealing temperature of LDH-B specific primers was 65°C.

2.1.4 LDH-C

The alignment of Atlantic salmon and rainbow trout LDH-B and LDH-C sequences indicated that they were highly conserved. The LDH-C specific

primers were designed based on salmonid LDH-C specific regions of the coding sequence. The 20-mer forward primer based on exon 2 was

5'-CACGGCAGCCTCTTCCTTAAAACAC-3' and the reverse primer based on exon 3 was

5'-CTGGGTTGGAGACCACGATGATGA-3'. The annealing temperature for the primers was 65°C.

The sequence of the PCR product with rainbow smelt genomic DNA as template could provide phylogenetic evidence to confirm the specific amplification of LDH-C. The PCR protocol and reaction were exactly as described in Section 2.1.1. The single band (350 bp) PCR product was cut out from a 1% SYBR Safe (Invitrogen) gel and purified by Ultrafree-DA column (Millipore). The purified DNA was subcloned using the pSTBlue-1 Acceptor Vector (Novagen) and transformed into Novablue Singles Competent Cells using the manufacturer's instructions in the pSTBlue Acceptor Vector Kit from Novagen. In order to confirm the positive insert, three white colonies were taken and each of these colonies was grown in 3 mL LB broth with 3 µL ampicillin (20 mg/mL) overnight. The plasmid DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen). The positive insert DNA was checked by digestion of plasmid DNA using the restriction enzyme FastDigest EcoRI (Fermentas). 400 ng of plasmid DNA, 1 µL of 10 x FastDigest buffer and 0.5 µL of FastDigest EcoRI were mixed in 10 µL. The digestion mix was kept in 37°C for 5 min and then electrophoresed on a 1% agarose gel with 1 X TAE and ethidium bromide (0.5 µg/mL). One selected plasmid with a positive insert was chosen for sequencing. Each

sequence reaction mixture consisted of 1 µL of Amersham DYEnamic ET terminator cycle sequencing kit master mix, 1 µL of DYEnamic ET Terminator dilution buffer, 2.5 µL of isolated DNA (400 ng) and 0.5 µL of 2 µM primer (R-20mer or U-20mer), whose sequences were designed from the flanking regions of the AccepTor Vector insert site. The U-20mer primer was

5'-GGTGACACTATAGAATACAG-3' and the R-20mer primer was

5'-ATGACCATGATTACGCCAAG-3'. The sequencing reaction was set up in a T3 Thermocycler (Biometra), and the protocol comprised the following steps: an initial denaturation at 95°C for 1 min; 35 cycles of 95°C for 20 sec, 50°C for 15 sec, and 60°C for 2 min; and a final elongation step at 60°C for 10 min. After the sequencing reaction, a sequencing cleanup was carried out. 50 µL 95% EtOH and 2 µL sodium acetate/EDTA buffer (1.5 M sodium acetate, 250 mM EDTA) were added to each sequencing reaction and mixed well. Each mix was centrifuged at 2700 x g for 30 min at 4°C. After removal of the supernatants, 150 µL of 70% EtOH was added and centrifuged at 2700 x g for 10 min at 4°C. The pellets were kept and air dried for 10 min. Finally, the pellet was dissolved in 2 µL formamide loading dye, and the sequencing analysis was carried out on an ABI Prism 377 DNA Sequencer (Applied Biosystems).

The sequence was trimmed of vector, and uploaded to NCBI Megablast (NCBI) to determine if it is similar to a partial LDH gene (Altschul et al. 1990). In order to confirm the sequence is a partial LDH-C sequence, the alignment and phylogenetic tree of this sequence and LDH-A, LDH-B and LDH-C coding sequences from Atlantic salmon and rainbow trout, as well as LDH-A and LDH-B

coding sequences of rainbow smelt were constructed using the MEGA4 package (Tamura et al. 2007).

After confirming the PCR amplification was specific for rainbow smelt LDH-C, a 40-mer probe was designed based on the partial sequence of LDH-C, 5'- CTTCAAACACATCATTCCCCAGATAGTGAGGTACAGCCCC-3'.

2.2 Rainbow smelt LDH BAC library screening

2.2.1 BAC Library

A rainbow smelt bacterial artificial chromosome (BAC) library (CHORI-74) was prepared by Children's Hospital Oakland Research Institute (CHORI) Oakland, CA, USA. The BAC library contains 52,410 clones with an average clone insert size of 146 kb, giving an 11-fold coverage of the rainbow smelt genome. The library has been set up on to three 22 x 22 cm nylon high-density filters for screening by probe hybridization. Each hybridization membrane contains 36,864 BAC clones, which represent 18,432 independent clones that have been spotted in duplicate (Schalburg et al. 2008).

2.2.2 Probe labelling

The design of each LDH-A, LDH-B and LDH-C gene specific probe was described in Section 2.1.2. The reference probe for the filter hybridization was a *Caenorhabditis briggsae* 40-mer overgo probe,

5'- GTTGCCAAATTCCGAGATCTTGGCGACGAAGCCACATGAT-3'.

Each LDH probe with a reference probe was hybridized simultaneously on one set of rainbow smelt BAC library filters.

The total volume for each labelling reaction mix was 5 µL, which contained 0.5 µL of 10 µM probe, 1 µL of 5 X Forward Reaction Buffer (Invitrogen), 0.5 µL of 10 U/µL of T4 polynucleotide kinase (Invitrogen), 1 µL of ^{32}P-γ ATP (0.37 MBq/µL) and 2 µL dH_2O. In order to position the ^{32}P on the 5' end of the probe, each probing reaction was incubated at 37°C for 1 hour.

2.2.3 Pre-hybridization

Each set of the CHORI-74 rainbow smelt BAC library contains three hybridization filters, which were pre-hybridized in a Roller-Blot Hybridizer HB-3D oven using a hybridization tube at 65°C for 2 hours. The 100 mL buffer of each pre-hybridization tube consists of 25 mL 20 X SSC (pH 7.0), 5 mL 10% SDS, 10 mL 50 X Denhardt's solution (5 g of bovine serum albumin, 5 g of Ficoll 400, 5 g of polyvinyl pyrrolidine and 500 mL of dH_2O) and 60 mL dH_2O.

2.2.4 Hybridization

After the pre-hybridization, to each tube was added one reaction mix of the oligonucleotide probe and one reaction of reference probe simultaneously. The hybridization tube was kept rotating at 65°C for 18 hours.

2.2.5 Washes

In order to remove the unhybridized labelled probe, two one hour washes were set up at 50°C. The buffer in each hybridization tube was composed of 20 mL 20 X SSC (pH 7.0), 4 mL 10% SDS and 376 mL dH_2O.

The BAC filters were removed from the hybridization tube and wrapped in Saran wrap. The filters were exposed to storage Phosphor screens (Molecular

Dynamics) for 20 hours. Then, the Phosphor screens were scanned using a Typhoon 9410 Phosphor Imager.

2.2.6 Positive selection

After the hybridization signals were detected, the hybridization positive BAC clones were picked from the rainbow smelt BAC library and amplified by PCR with LDH gene specific primers. One PCR positive BAC clone was selected from each LDH gene specific amplification. Each LDH gene specific BAC clone was grown in 5 mL LB broth with 2.5 µL of 25 mg/mL chloramphenicol with shaking at 250 RPM for 16 hours at 37°C. For future use, 700 µL culture from each BAC clone was mixed with 300 µL 50% glycerol and kept as a stock in a -80°C freezer.

2.3 Shotgun Library

2.3.1 BAC DNA Preparation

To obtain a single colony of the BAC clone, the BAC stock was streaked on a LB agar plate with chloramphenicol (25 mg/mL) and incubated at 37°C for 16 hours. Ten single colonies were picked and each colony was put into 5 mL LB broth with 2.5 µL 25 mg/mL chloramphenicol and shaken at 250 RPM for 16 hours at 37°C. The ten selected colonies were tested by PCR with LDH gene specific primers followed by agarose gel electrophoresis and by imaging on a UV trans-illuminator (Alpha Innotech).

A PCR positive colony was selected and prepared for QIAGEN Large-Construct start culture. The single colony was inoculated in 5 mL LB broth with

25 mg/mL chloramphenicol and grown at 37°C for 8 hours. This starter culture was diluted and grown in 500 mL LB broth with 250 µL chloramphenicol (25 mg/mL) at 250 RPM shaking for 16 hours at 37°C. The bacterial cells were harvested by centrifugation at 6000 x g for 15 min at 4°C. After the supernatant was removed, the cell pellet was used to extract BAC DNA. The BAC DNA extraction exactly followed the QIAGEN Large-Construct Kit Protocol.

2.3.2 DNA shearing

The concentration of the BAC DNA was determined using a NanoDrop DN-1000 Spectrophotoimager. 10 µg of BAC DNA was separated into four 0.6 mL tubes and sonicated with an Ultrasonic Processor for 2 sec, 4 sec, 6 sec and 8 sec, which resulted in 2-5 Kb fragments. 2 µL of each sonicated DNA reaction was checked for fragment size by electrophoresis separately.

2.3.3 End-repairing

The 2-5 Kb sonicated DNA fragments were mixed together and concentrated to 52 µL. An end-repair reaction was set up using the Epicentre End-it DNA End-repair Kit. Each reaction mix had a total of 80 µL and contained 52 µL of sheared DNA, 8 µL of 10 X buffer, 8 µL of 10 mM ATP, 8 µL of 2.5 mM dNTP and 4 µL End-repair Enzyme Mix. The end-repairing mix was incubated at 25°C for 45 min.

2.3.4 Gel extraction

The sonicated, end-repaired BAC DNA was loaded on a 1% TAE agarose gel with SYBR Safe (Invitrogen) and run for 50 min at 200 V. The DNA with the

size of 2-5 Kb was cut out and purified using a QIAGEN Gel Extraction Kit. The concentration of the BAC DNA was determined using a NanoDrop DN-1000 Spectrophotoimager.

2.3.5 Ligation

The BAC DNA was ligated into SmaI digested, phosphatase treated pUC19 vector and transformed into competent cells. The ligation reaction mix of 14.5 µL of 100 ng BAC DNA and dH_2O, 0.5 µL pUC19 (50 ng/µL), 1 µL T4 DNA Ligase (Invitrogen) and 4 µL 5 X Reaction buffer was incubated at 14°C for 20 hours. In order to quantify the efficiency of the transformation, a positive control and a negative control were set up. The positive control was composed of the same amount of reagents as the BAC ligation reaction mix but the100 ng of positive BAC was from Atlantic salmon LDH-A1 BAC (S0052D13). The negative control also had the same ligation reaction mix but had 14.5 µL dH_2O instead of BAC DNA.

2.3.6 Transformation

2.5 µL of each ligation reaction was used to transform 100 µL of Stratagene XL-1 Blue Competent Cells. The competent cells were thawed on ice for 5 min and 1.7 µL beta-mercaptoethanol was added. The competent cell mix was kept on ice for 10 min with gently swirling every 2 min. 2.5 µL of BAC DNA ligation was added into the competent cell mix and kept on ice for 30 min. The cell mix was put into a 42°C water bath for a heat shock of 45 sec and then kept on ice for 2 min. 900 µL of SOC medium was added to the cell mix and incubated

for 1 hour with 250 RPM at 37°C. 250 µL of transformation mix was spread on an ampicillin LB agar plate, which contained 375µL of 20 mg/mL X-gal and 225 µL of 200 mg/mL UltraPure IPTG (Invitrogen). The agar plates were incubated at 37°C for 20 hours.

2.3.7 Insert check by digestion

To quantify the positive performance of the transformation, 64 colonies were picked from the transformation plates and each was grown in 1200 µL LB broth with 12 µL of ampicillin (20 mg/mL). Plasmid DNA was isolated from each culture. The concentration was estimated using a NanoDrop DN-1000 Spectrophotoimager. The positive insert DNA was checked by digestion using the restriction enzyme FastDigest PvuII (Fermentas). The 10 µL digestion mix contained 500 ng of plasmid DNA, 1 µL of 10 x FastDigest buffer and 0.5 µL of FastDigest PvuII. The digestion mix was kept in 37°C for 5 min and electrophoresed on a 1% agarose gel with 1 X TAE and ethidium bromide (0.5 µg/mL), and visualized using a UV trans-illuminator (Alpha Innotech).

2.3.8 Sequencing check library quality

Sequencing reactions were set up to check the library quality. Each sequence reaction mix consisted of 1 µL of Amersham DYEnamic ET terminator cycle sequencing kit master mix, 1 µL DYEnamic ET Terminator dilution buffer, 2.5 µL isolated DNA (100 ng) and 0.5 µL of 2 µM primer (M13F or M13R), whose sequences were M13F 5'- GTAAAACGACGGCCAGT-3' ; M13R 5'-GGAAACAGCTATGACCATG-3'. The sequencing reaction was set up in a T3

Thermocycler (Biometra), and the sequencing protocol and cleanup were the same as described in Section 2.1.3. The sequencing analysis was carried out using an ABI Prism 377 DNA Sequencer (Applied Biosystems).

The vector trimmed sequences were compared with the NCBI non-redundant nucleotide database using BLASTN (Altschul et al. 1990). The result provided evidence to support the good quality of the BAC library if the results had less than 3% *E. coli* sequence hits.

2.3.9 Re-transformation

The method of re-transformation is exactly same as Section 2.3.6 but with double the amount of reagents. The competent cells were thawed on ice for 5 min and 3.4 µL beta-mercaptoethanol was added. The competent cell mix was kept on ice for 10 min with gently swirling every 2 min. 5 µL of BAC DNA ligation was added into the competent cell mix and kept on ice for 30 min. The cell mix was put into a 42°C water bath for a heat shock of 45 sec and then kept on ice for 2 min. 1800 µL of SOC medium was added to the cell mix and incubated for 1 hour with 250 RPM at 37°C. 500 µL of transformation mix was spread on an ampicillin LB agar plate, which contained 750 µL of 20 mg/mL X-gal and 450 µL of 200 mg/mL UltraPure IPTG (Invitrogen). The agar plates were incubated at 37°C for 20 hours.

2.3.10 Large scale sequencing of BAC library

Approximately 2300 positive colonies from each BAC shotgun library transformation were picked. Each clone was grown in a 60 µL mix composed of

20 µL 2 X YT (Yeast Exact Tryptone) medium, 37.5 µL of 50% glycerol and 2.5 µL of ampicillin (20 mg/mL) at 37°C for 20 hours. The selected colonies were sent to the Michael Smith Genome Sciences Centre for sequencing.

2.4 LDH BACs assembling and annotation

The BAC sequences were assembled using the Phred, Phrap and Consed software package (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). The BAC sequence annotation was carried out using the Genomics Research on All Salmonids Project (GRASP) annotation pipeline, which was created by William Chow (http://grasp.mbb.sfu.ca/).

2.5 Rainbow smelt LDH gene expression

Rainbow smelt tissues were collected by colleagues at the Ocean Sciences Centre, Memorial University, St. John's. Nine tissues in total were collected: brain, eye, gill, muscle, heart, liver, head kidney, spleen and gonad, and flash frozen in liquid nitrogen.

2.5.1 Total RNA extraction

100 mg of each tissue was cut up and added to 500 µL Trizol in a 1.5 mL microcentrifuge tube before being homogenized using a pestle. The tissue and Trizol mix had an additional 500 µL Trizol added and the mixture was slowly passed through a 1 mL syringe with a 27 G needle to shear the genomic DNA. The homogenized samples were left at room temperature for 5 min, and then 200 µL of chloroform was added. The mixed samples were shaken vigorously and left at room temperature for 3 min. The mixture was vortexed vigorously for 20 sec

and left for 3 min, then centrifuged at 12,000 x g for 15 min at 4°C. The upper layer was transferred into a new 1.5 mL tube, and to it was added 500 µL 2-propanol. The solution was mixed by inverting. The mixture was allowed to stand at room temperature for 10 min, and then centrifuged at 12,000 x g for 10 min at 4°C. After removing the supernatant, the pellet was dissolved in 1 mL of 75% EtOH in DEPC (Diethypyrocarbonate) treated water by vortexing, and then the mixture was centrifuged at 7500 x g for 5 min at 4°C. The supernatant was removed with a pipette, and then the pellet was kept and air dried for 10 min. When the pellet became transparent, it was dissolved in 87.5 µL RNase free water. In order to remove the DNA within the total RNA solution, DNase I (QIAGEN RNase-free DNase Set) was applied to make a digestion mix composed of 87.5 µL of total RNA solution, 10 µL Buffer RDD and 2.5 µL DNase I stock solution. The mix was incubated at room temperature for 10 min. The RNA cleanup followed the Qiagen MinElute Cleanup Handbook: RNA Cleanup and Concentration protocol. 350 µL of Buffer RTL was added to the RNA solution, and then 250 µL of 100 % ethanol was added with mixing using a pipette. The sample was transferred to an RNeasy MinElute spin column and placed in a 2 ml collection tube, and centrifuged at room temperature for 15 sec at 9000 x g. After discarding the flow-through, the RNeasy MinElute spin column was placed in a new 2 mL collection tube. 500 µL of Buffer RPE was added to the spin column and centrifuged at room temperature for 15 sec at 9000 x g. Then the flow-through was removed. To wash the spin column membrane, 500 µL of 80% EtOH was added to the RNeasy MinElute column and centrifuged at room

temperature for 2 min at 9000 x g. The RNeasy MinElute spin column was removed and placed in a new 2 mL collection tube. To remove the extra EtOH from the spin column, the RNeasy MinElute spin column was centrifuged with the lid open at room temperature for 5 min at 9000 x g. The RNeasy MinElute spin column was placed in a new 1.5 mL microcentrifuge tube and 14 µL RNase-free water was added. Then it was centrifuged at room temperature for 1 min at 9000 x g to elute the RNA. The concentration of the total RNA was tested using a NanoDrop DN-1000 Spectrophotoimager. To check the quality of the total RNA, 2 µL of eluted RNA was electrophoresed on a 1% agarose gel with 1 X TAE and ethidium bromide (0.5 µg/mL), and visualized using a UV trans-illuminator (Alpha Innotech).

2.5.2 RT-PCR

The First-Strand cDNA was synthesized using the Invitrogen SuperScript III kit. The first step of the RT-PCR reaction was composed of 2 µL of random hexamers (IDT ReadyMade Primer 100 ng/µL), 1 µL of 10 mM dNTPs and 10 µL of extracted total RNA (1 µg) with RNase free H_2O. The mixed reagents were placed in a T3 Thermocycler (Biometra) at 65°C for 5 min and then kept on ice for 1 min. The second step of the RT-PCR reaction was carried out in a total of 20 µL for each reaction, and all reagents were from the Invitrogen SuperScript III kit. The sample from the first step was added along with 4 µL of 5 x First Strand Buffer, 1 µL of DTT (Dithiothreitol) (0.1 M), 1 µL of RNase OUT (Invitrogen RNaseOUT Recombinant RNase inhibitor 40 U/µL) and 1 µL of SuperScript III. The second step reaction was incubated in a T3 Thermocycler (Biometra) at

25°C for 5 min and then 50°C for 1 hour. An inactivation step was applied by heating at 70°C for 15 min in a T3 Thermocycler (Biometra). The cDNA was stored at -20°C until used for studying tissue expression.

2.5.3 Tissue expression

The cDNA from each tissue of rainbow smelt was amplified by β–actin primers and each pair of LDH specific primers. The primers for the reference control, β–actin gene, were designed from the rainbow smelt EST database 100/99 (http://web.uvic.ca/grasp/). The pairs of LDH-A and LDH-B primers for cDNA amplification were same as the one for screening BAC library. The pair of LDH-C gene specific primers was designed based on the LDH-C coding sequences. The primer sequences and Tm for rainbow smelt LDH tissue expression are listed in Table 3.8 of Section 3.6. The PCR reactions for β–actin and each LDH gene amplification of rainbow smelt were exactly the same as the protocol described in Section 2.1. The PCR products were electrophoresed on a 1% agarose gel with 1 X TAE and ethidium bromide (0.5 μg/mL), and visualized using a UV trans-illuminator (Alpha Innotech).

2.6 LDH gene structure

2.6.1 LDH cDNA sequences

In order to identify the exon-intron boundaries from the genomic DNA sequences of each the LDH genes, LDH-A, LDH-B and LDH-C specific primers were designed based on the predicted coding sequences of each annotated BAC sequence respective to each LDH gene. The cDNA from rainbow smelt brain was amplified by each pair of LDH-A, LDH-B and LDH-C specific primers. The primer sequences are listed in Table 3.6 of Section 3.4.1. The PCR reaction and protocol were exactly the same as Section 2.1.1. PCR products were electrophoresed on a 1% agarose gel containing 1 X TAE and 1% SYBR Safe (Invitrogen) and purified using Ultrafree-DA columns (Millipore). The purified DNA was subcloned using the pSTBlue-1 Acceptor Vector (Novagen). The method of transformation into Novablue Single Competent Cells and the insert sequencing were the same as described in Section 2.1.3.

2.6.2 LDH exon-intron boundary identification

After sequencing the PCR amplified cDNA products corresponding to LDH-A, LDH-B and LDH-C, alignments between the BAC sequences and cDNAs for each LDH gene were made using the NCBI Splign software (http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi?textpage=overview&level=form) (Kapustin et al. 2008). The exon-intron boundaries of LDH-A, LDH-B and LDH-C were identified, while the coding sequences for each LDH gene were confirmed from their cDNAs.

2.7 Phylogenetic analysis of rainbow smelt and Atlantic salmon LDHs

The rainbow smelt coding sequences for LDH-A, LDH-B and LDH-C were translated into amino acid sequences using ExPASy Translation Tool (http://www.expasy.ch/tools/dna.html). A phylogenetic analysis was carried out based on the LDH coding sequences and amino acid sequences from rainbow smelt, Atlantic salmon, rainbow trout and an outgroup tunicate (*Ciona intestinalis*). The LDH coding sequences and amino acid sequences of Atlantic salmon, rainbow trout and tunicate were from previous study (Lubieniecki et al. in preparation). The coding sequences and amino acid sequences were analyzed separately to make Neighbor-Joining and Minimum Evolution phylogenetic trees created using the MEGA4 package (Tamura et al. 2007). The different types of phylogenetic trees are reviewed in Graur and Li (2000).

The evolutionary rates for LDH-A and LDH-B between salmonids and rainbow smelt were examined separately. The evolutionary rate was based on the number of amino acids substitutions between each pair of sequences under the same evolutionary time and this was done manually. The LDH-A group contained the amino acid sequences from Atlantic salmon, LDH-A1 and LDH-A2, rainbow trout, LDH-A1 and LDH-A2, and rainbow smelt LDH-A. The LDH-B group contained the amino acid sequences from Atlantic salmon, LDH-B1 and LDH-B2, rainbow trout, LDH-B1 and LDH-B2, and rainbow smelt LDH-B.

In order to understand the evolution of LDH genes in salmonids and smelt, the ratio of nonsynonymous (d_N) and synonymous (d_S) nucleotide changes was

estimated using http://www.datamonkey.org/dataupload.php (Pond and Frost 2005a) and PAL2NAL (http://www.bork.embl.de/pal2nal/) (Suyama et al. 2006). Datamonkey is a webserver to test the signature of positive or negative selection from site by site of coding nucleotide sequence alignments using state-of-the-art statistical models (Pond and Frost 2005a). PAL2NAL is a web tool using Phylogenetic Analysis by Maximum likelihood (PALM) to test d_N and d_S value (Suyama et al. 2006).The d_N/d_S ratios were used to determine the type of selection in the nucleotide coding sequences for every translated amino acid at each codon (Pond and Frost 2005b).

CHAPTER 3: RESULTS

3.1 Overview and purpose

Ohno (1970a) was among the first to propose that gene duplication provides the raw material for the evolution of genes with novel functions. He recognized that the most likely fate of one copy of a duplicated gene is nonfunctionalization (pseudogenization); that is, the deletion of the gene or accumulation of deleterious mutations that prevent the expression of the gene product or result in the production of a faulty gene product. The observation that duplicate genes persist in genomes at a higher frequency than expected, suggested that either neofunctionalization or subfunctionalization must play a role in maintaining both copies of the duplicated gene. In the neofunctionalization model, one of the copies accumulates mutations that alter the function of the gene product. This altered function could be a change in enzyme specificity or more simply a change in kinetic parameters while catalyzing the same reaction as in the case of isozymes. The DDC model of Force et al. (1999) suggested that there would be a selective pressure to maintain both copies of the duplicated genes if the ancestral functions, including the expression pattern, were partitioned between the duplicates. This could arise through complementary deleterious mutations in promoter regions yielding different tissue expression patterns (e.g., as found in isozymes). These three potential fates of gene duplicates are not mutually exclusive, and other possibilities exist. For example,

neofunctionalization could produce not merely a new protein product, but also a totally different expression pattern (e.g., as has been observed for RNAse1 in leaf-eating monkeys and lysozyme in ruminants). To examine the fate of a pair of duplicated genes after a whole genome duplication event, it is necessary to have a well characterized copy of the gene from a representative of diploid out-group.

As indicated in the Introduction, I have chosen to study the fate of salmonid duplicated LDH genes that result from autotetraploidization event in the common ancestor of species such as Atlantic salmon and rainbow trout. Work in the Davidson lab has characterized the structure, expression patterns and genome organization of the duplicated Atlantic salmon LDH-A (LDH-A1 and LDH-A2) and LDH-B (LDH-B1 and LDH-B2) genes as well as the single copy LDH-C. Therefore, I set out to isolate and characterize the LDH genes (LDH-A, LDH-B and LDH-C) from rainbow smelt, whose genome is considered a good representative of the diploid ancestral state that preceded the salmonid whole genome duplication. The purpose was to use the information I produced to gain insight on how duplicated genes evolve.

There were two parts to my thesis. The first was to isolate rainbow smelt BACs that contain the LDH genes, to sequence them and then to annotate the genomic sequences such that the structure and genome organization of each of the LDH genes could be determined. This involved: (1) screening the rainbow smelt EST database for LDH transcripts (Section 3.1); (2) selecting specific oligonucleotide probes for each of the LDH genes and screening the rainbow smelt BAC library (Section 3.2); (3) preparing shotgun libraries of three BACs

that each contain a different LDH isozyme gene and assembling and annotating the BAC sequences (Section 3.3); and (4) determining the gene structures and the inferred amino acid sequences of rainbow smelt LDH-A, LDH-B and LDH-C (Section 3.4). The second part of my thesis involved an evolutionary comparison of the rainbow smelt LDH genes with those from Atlantic salmon. I also used the sequences of rainbow trout ESTs for LDH-A1, LDH-A2, LDH-B1, LDH-B2 and LDH-C and the LDH gene sequences from teleosts, whose genomes have been sequenced, when appropriate. First, I searched for evidence of positive selection as a signature of neofunctionalization in one or other of the salmonid duplicates (Section 3.5). Having examined the patterns of amino acid substitution in the different types of LDH genes, I determined the tissue expression patterns of the three rainbow smelt LDH genes and then attempted to relate this information to the subfunctionalization of the salmonid LDH-B1 and LDH-B2 genes (Section 3.6). It has been suggested that significant genome reorganization must occur immediately after an autotetraploidization event as the two pairs of identical homeologous chromosomes change such that the homeologues no longer interact and a stable diploid state is re-established. Therefore, I searched for conservation of synteny and changes such as inversions and deletions in the genomic regions containing LDH genes in rainbow smelt and the corresponding regions of Atlantic salmon (Section 3.7).

3.1 Data mining rainbow smelt EST database for LDH transcripts

I wanted to obtain sequences of rainbow smelt LDH genes or transcripts so that I could design oligonucleotide probes that could be used to screen the rainbow smelt BAC library specifically for LDH-A, LDH-B and LDH-C. Initially, the rainbow smelt EST clustering from the University of Victoria database (http://web.uvic.ca/grasp/) was searched for rainbow smelt LDH open reading frames (ORFs) from EST consensus sequences. The rainbow smelt LDH EST database only provided a partial-length EST sequence of LDH-A, while no information on LDH-B and LDH-C was available for rainbow smelt. The design of an LDH-A gene specific probe to screen the rainbow smelt BAC library was based on the alignment of the partial EST sequence and the LDH-A1 and LDH-A2 coding sequences from Atlantic salmon (see Section 3.2.1.1.).

3.2 Screening the rainbow smelt BAC library for LDH genes

As it was not possible to use the rainbow smelt EST database resource to design LDH-B and LDH-C specific probes, they had to be based on Atlantic salmon coding sequences. Five LDH gene containing BACs were characterized in the Atlantic salmon BAC library, and they are LDH-A1 = S0052D13; LDH-A2 = S0069I14; LDH-B1 = S0225J21; LDH-B2 = S0276I15; LDH-C = S0116D13. Each of the LDH BACs from Atlantic salmon was sequenced and annotated, and the gene structure and exon-intron boundary of each LDH from Atlantic salmon has been determined. Both LDH-A1 and LDH-A2 genes encode 332 amino acid proteins and have 7 exons; while LDH-B1, LDH-B2 and LDH-C genes also have 7 exons that produce proteins with 334 amino acids (Lubieniecki et al. in

preparation). Five LDH EST sequence contigs of rainbow trout from the University of Victoria EST clustering database (http://web.uvic.ca/grasp/) provided additional LDH gene information for the identification of rainbow smelt LDH genes. Each LDH gene specific 40-mer oligonucleotide probe was designed to identify one of the expected LDH genes in the rainbow smelt BAC library. Each pair of LDH gene specific primers was designed for checking the gene specificity of BAC DNA and the tissue expression in rainbow smelt.

3.2.1 Design of probes for rainbow smelt LDH

3.2.1.1 LDH-A from EST clustering of University of Victoria

LDH-A was searched for in the 100/99 rainbow smelt EST clustering database from the University of Victoria (http://web.uvic.ca/grasp/). The contig 4546, which had the smallest E-value and longest ORF, was considered as a candidate LDH-A transcript. The BLAST results against the Swissprot database of the EST sequence selected from the contig 4546 confirmed that the EST sequence is LDH-A specific with an E-value = 1e-136, bit score = 1242 and identity = 86%. In order to design a probe within one exon from the contig 4546 EST sequence, it was necessary to confirm the location of exon-intron boundaries. The alignment of the contig 4546 EST sequence of rainbow smelt and the coding sequences of the LDH-A1 and LDH-A2 genes from Atlantic salmon revealed the LDH-A conservative regions and the predicted positions of exon-intron boundaries for the rainbow smelt LDH-A gene (Figure 3.1). An LDH-A 40-mer oligonucleotide probe was designed based on predicted exon 2 of the rainbow smelt LDH-A EST sequence. The 40-mer probe was used as the 5’

forward primer and the 3' reverse 20-mer primer was designed based on predicted exon 3 (Table 3.1). Rainbow smelt genomic DNA was amplified using the LDH-A specific primers. The specific PCR product of LDH-A amplification was 350 bp at an optimum Tm of 65°C.

3.2.1.2 LDH-B probe and primer design

The design of the rainbow smelt LDH-B probe and gene specific primers was based on the alignment of coding sequences of LDH-B and LDH-C from Atlantic salmon and full-length EST sequences of LDH-B and LDH-C from rainbow trout (Figure 3.2). Because the coding sequences of LDH-B and LDH-C are highly similar, the regions that distinguished between LDH-B and LDH-C were identified from both Atlantic salmon and rainbow trout. The 40-mer LDH-B probe was designed based on the Atlantic salmon LDH-B specific coding sequence in exon 5. The 5' 40-mer probe was used as the forward gene specific primer and the 3' reverse primer was designed based on a conserved region of exon 6 of Atlantic salmon (Table 3.2). When rainbow smelt genomic DNA was amplified using the LDH-B specific primers, a PCR product of 250 bp was obtained at a Tm of 65°C.

Figure 3.1 LDH-A alignment of Atlantic salmon and rainbow smelt EST nucleotide coding sequences. The full-length nucleotide coding sequences of LDH-A1 and LDH-A2 are from Atlantic salmon BAC sequences and the partial LDH-A sequence is from rainbow smelt EST database. Each exon boundary indicates as red bar. LDH-A specific probe and reverse primer were designed on exon 2 and 3. AT: Atlantic salmon; RS: rainbow smelt.

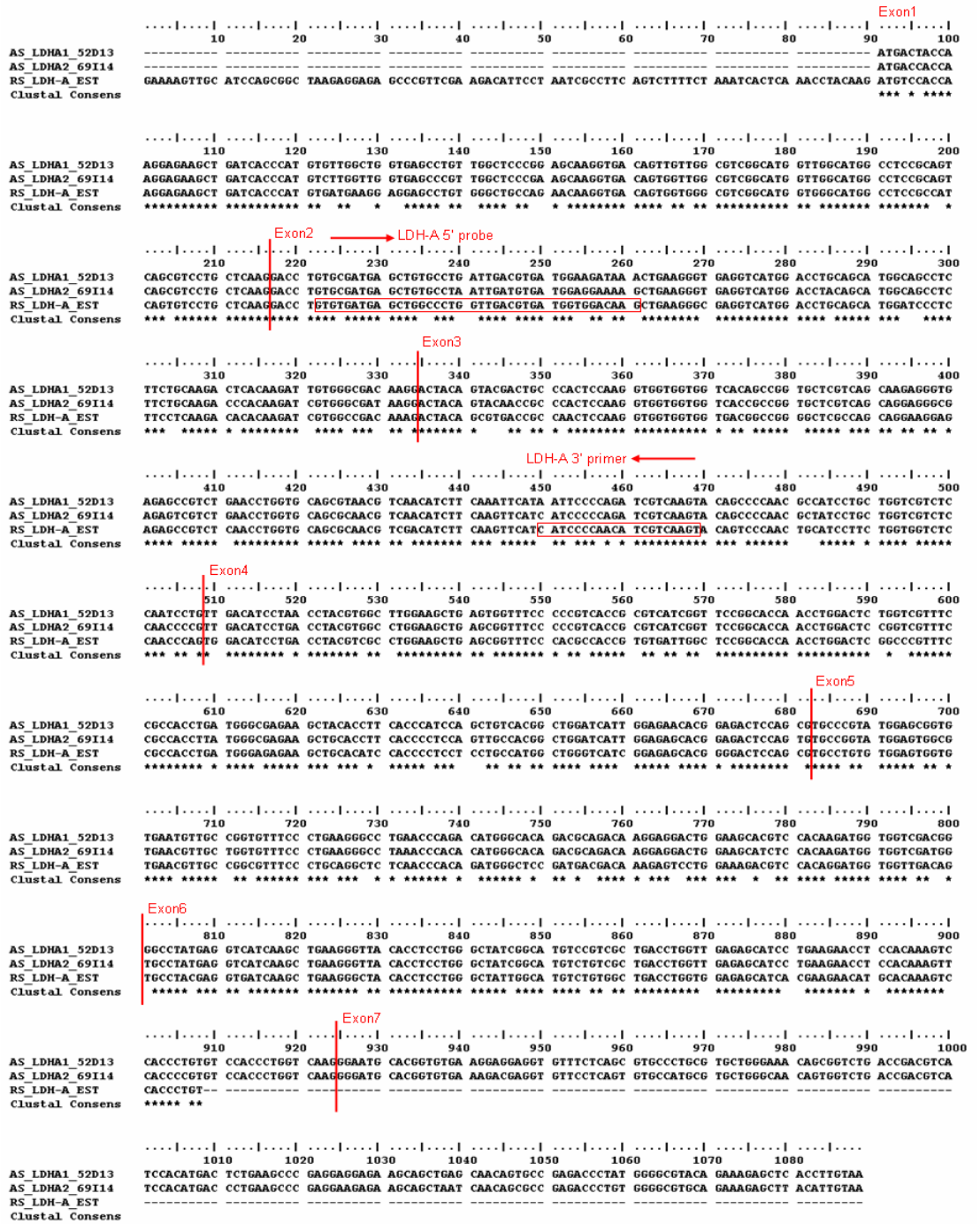


Table 3.1 The probe and forward and reverse LDH-A specific primers.

	LDH-A
Probe	5'-GTGTGATGAGCTGGCCCTGGTTGACGTGATGGTGGACAAG-3'
Forward primer	5'-GTGTGATGAGCTGGCCCTGGTTGACGTGATGGTGGACAAG-3'
Reverse primer	5'-ACTTGACGATGTTGGGGATG-3'

Figure 3.2 The alignment of LDH-B and LDH-C full-length nucleotide coding sequences from Atlantic salmon and rainbow trout EST. The LDH-B1, LDH-B2 and LDH-C full-length nucleotide coding sequences are from Atlantic salmon BAC sequences and those from rainbow trout are from EST database. Each exon boundary indicates as red bar. LDH-B specific probe and reverse primer were designed on exon 5 and 6. AT: Atlantic salmon; RT: rainbow trout.

Table 3.2 The probe and forward and reverse LDH-B specific primers

	LDH-B
Probe	5'-TCAGCGTAGCTGGAGTCAACCTGCAGAAGCTGAACCCAGAG-3'
Forward primer	5'-TCAGCGTAGCTGGAGTCAACCTGCAGAAGCTGAACCCAGAG-3'
Reverse primer	5'-TGAGATCAGCCACACTCAGG-3'

3.2.1.3 LDH-C probe and primer design

The LDH-C specific probe and primers were designed based on the alignment of LDH-B coding sequences from rainbow smelt BAC DNA, the LDH-C coding sequence from Atlantic salmon BAC DNA and the LDH-C full-length EST sequence from rainbow trout (Figure 3.3). The coding sequences of LDH-B and LDH-C have 81% identity in Atlantic salmon and 80% in rainbow trout. A 40-mer LDH-C specific probe which has at least 3 unique bases different from LDH-B could not be identified within one exon of the LDH-C coding sequence in the LDH-B and LDH-C alignment in Atlantic salmon. Therefore, in order to identify LDH-C and distinguish it from LDH-B in rainbow smelt, 6 sets of primers containing at least one LDH-C unique base at the 3' end were designed based on each exon of the Atlantic salmon LDH-C coding sequence. The rainbow smelt genomic DNA PCR product amplified by these 6 pairs of primers indicated that only primers from exon 2 and exon 3 gave a single specific 350 bp amplification product at Tm 65°C (Figure 3.3).

To test that the PCR amplification product is from the LDH-C gene in rainbow smelt, the 350 bp PCR product was sequenced. The BLASTN results of the PCR sequence showed that the best hit is carp (*Cyprimus carpio*) LDH-B with an E-value = 5e-53, score = 215 and 81% identity, and another strong hit is killifish (*Fundulus heteroclitus*) LDH-C with an E-value = 1e-48, bit score = 201 and 80% identity. Because the PCR product is only a partial sequence, I felt that more evidence had to be obtained. The smelt partial sequence contains one intron; therefore, tBLASTX from NCBI was used to predict the location of the

intron in the smelt partial sequence (query) based on the translated nucleotide sequences (subject) of fish species. The best tBLASTX annotated hit was zebrafish LDH-B with an E-value = 1e-33, bit score = 130 bits, and 7 pieces of alignments with different translated frame between query and subject sequences. The two continuous pieces of alignments with 90% identity show the translated nucleotide fragments are frame +3 of the query sequence and frame +1 of the subject sequence. In order to identify the correct query translational frame, I compared the query translated frame +3 with the Atlantic salmon LDH-C amino acid sequence. The identity between the smelt partial sequence with frame +3 and Atlantic salmon amino acid sequence is 94.4%. This confirms that the two continuous pieces of translated frame +3 are correct and it also located the intron position in the smelt partial sequence. The smelt partial sequence with the intron removed, was aligned with the LDH-B and LDH-C coding sequences from Atlantic salmon and the LDH-B and LDH-C EST sequences from rainbow trout, and this comparison was used to make a Neighbor-Joining tree (see section 2.7) (Figure 3.4). The smelt partial sequence (smelt EX2-3 seq) grouped with LDH-B and LDH-C genes of Atlantic salmon, rainbow trout and rainbow smelt. Because this partial sequence is only ~220 bp, which was much shorter than other sequences, I predicted the partial sequence was either a LDH-C type or a LDH-B type in rainbow smelt. Moreover, the bootstrap on the branch node before LDH-B and LDH-C divergence was 54; and bootstrap before the divergence of Atlantic salmon and rainbow smelt LDH-B was 19. The low values of bootstraps in the Neighbor-Joining tree did not give me a lot of confidence to support the

phylogenetic relationships in the pattern of the tree. However, a 40-mer LDH-C specific probe was designed based on this partial sequence and used to screen the rainbow smelt BAC library. The sequence of the probe and its reverse primer are shown in Table 3.3.

3.2.2 LDH hybridization

The 40-mer oligonicleotide probes of LDH-A, LDH-B and LDH-C were hybridized separately to the rainbow smelt BAC library (CHORI-74). The positive hybridizations for each LDH probe are shown in Figures 3.5, 3.6 and 3.7.

3.2.3 PCR verification of positive BAC hybridization

The selected positive BACs from each LDH hybridization were PCR amplified with the LDH specific primers (Figure 3.8). The PCR products provide confirmation of the presence of a specific LDH gene in each positive BAC and distinguish among the LDH-A, LDH-B and LDH-C genes (Table 3.4). Because a rainbow smelt BAC library contig map has not been built, one positive BAC with PCR positive confirmation corresponding to each LDH was randomly selected. They are LDH-A = O0109H14; LDH-B = O0079M15; LDH-C = O0113P10. To confirm the LDH gene in each BAC, three sets of PCR were carried out using the three LDH gene specific primer pairs and with three LDH BAC DNA as templates (Figure 3.9).

Figure 3.3 The alignment of full-length nucleotide coding sequences of rainbow smelt LDH-B, LDH-C from Atlantic salmon and rainbow trout. The rainbow smelt LDH-B and Atlantic salmon LDH-C is from full-length nucleotide coding sequences of BAC sequences. The rainbow trout LDH-C is from full-length EST sequence. Each exon boundary indicates as red bar. LDH-C specific probe and reverse primer were designed on exon 2 and 3. RS: rainbow smelt; AT: Atlantic salmon; RT: rainbow trout.

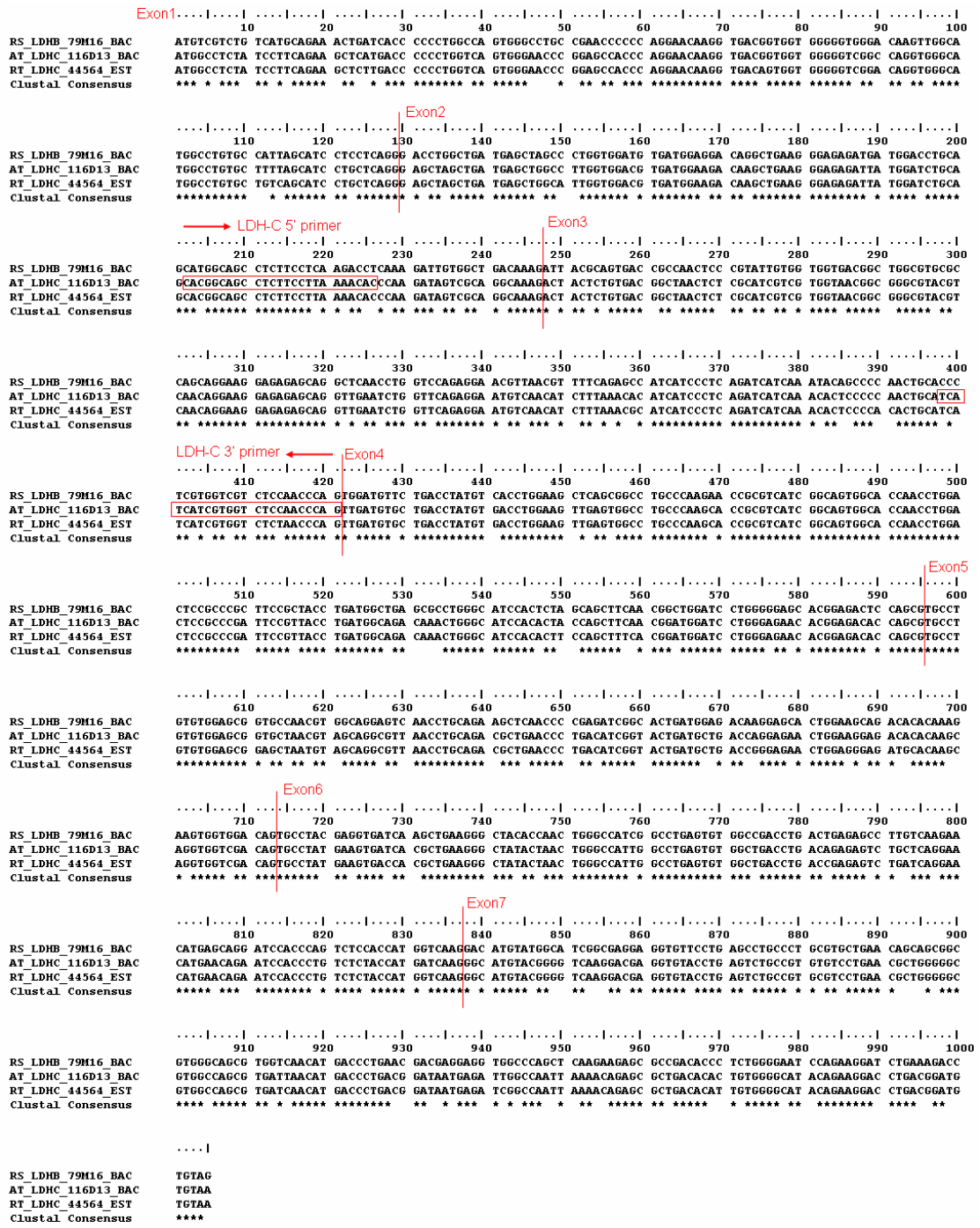


Figure 3.4 **The Neighbor-Joining tree of all LDH coding sequences from Atlantic salmon,** rainbow trout, rainbow smelt and partial sequence of rainbow smelt. Outgroup is from tunicate LDH amino acid sequence (Ciona intestinalis). AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt. The pink dot indicates the partial sequence of rainbow smelt.

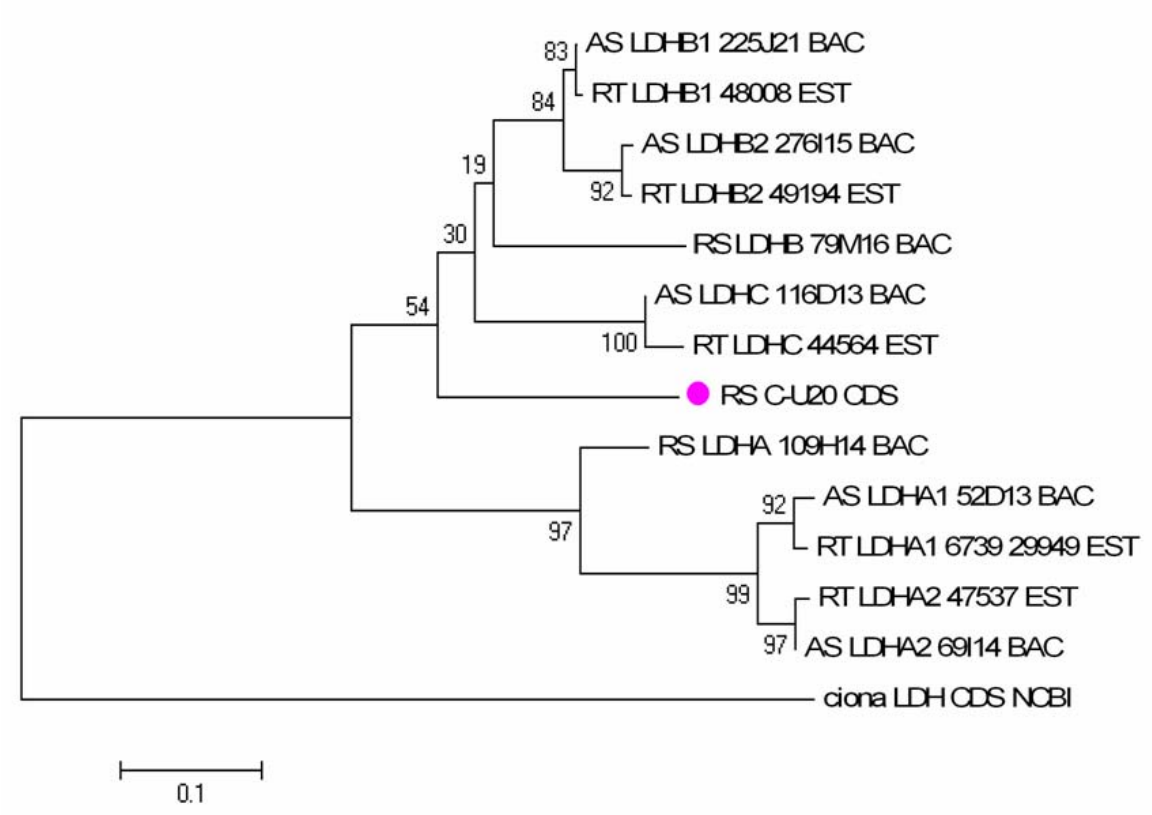
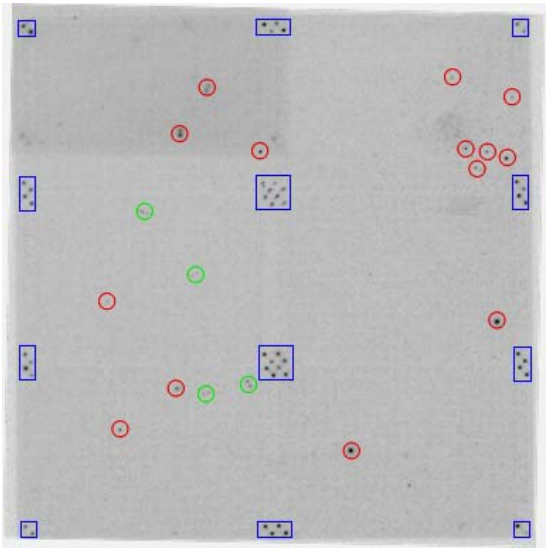


Table 3.3 The probe and forward and reverse LDH-C specific primers

	LDH-C
Probe	5'-CTTCAAACACATCATTCCCCAGATAGTGAGGTACAGCCCC-3'
Forward primer	5'-CACGGCAGCCTCTTCCTTAAAACAC-3'
Reverse primer	5'-CTGGGTTGGAGACCACGATGATGA-3'

Figure 3.5 CHORI-74 rainbow smelt BAC library filters hybridized with LDH-A probe. The blue rectangles indicate the hybridization by overgo reference probe (*C. briggsea*). The duplicated spots circled by green represent positive hybridization with LDH-A probe. The red circles show the noise, which are single spots.



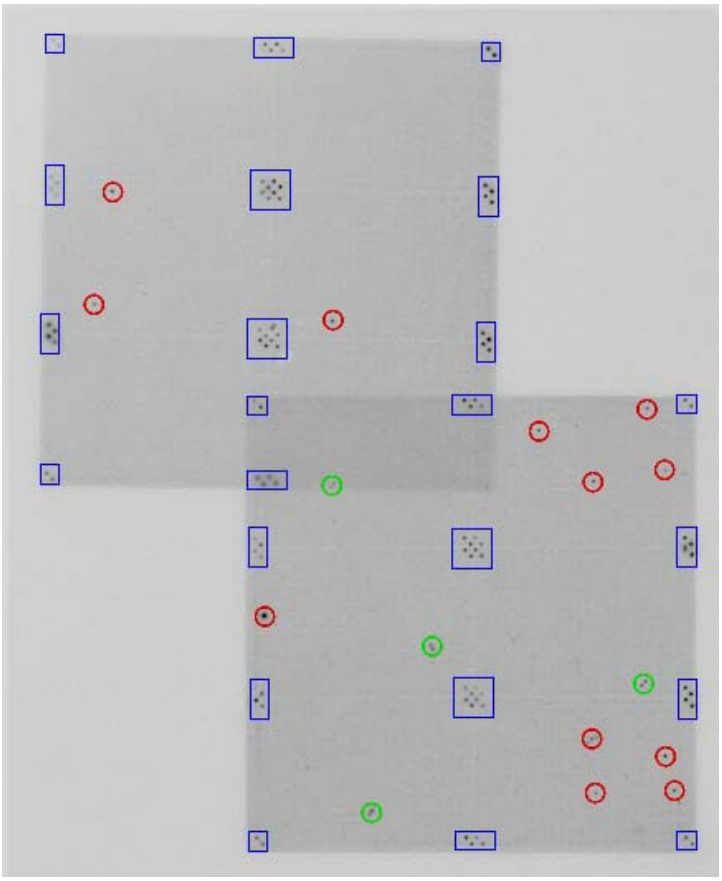
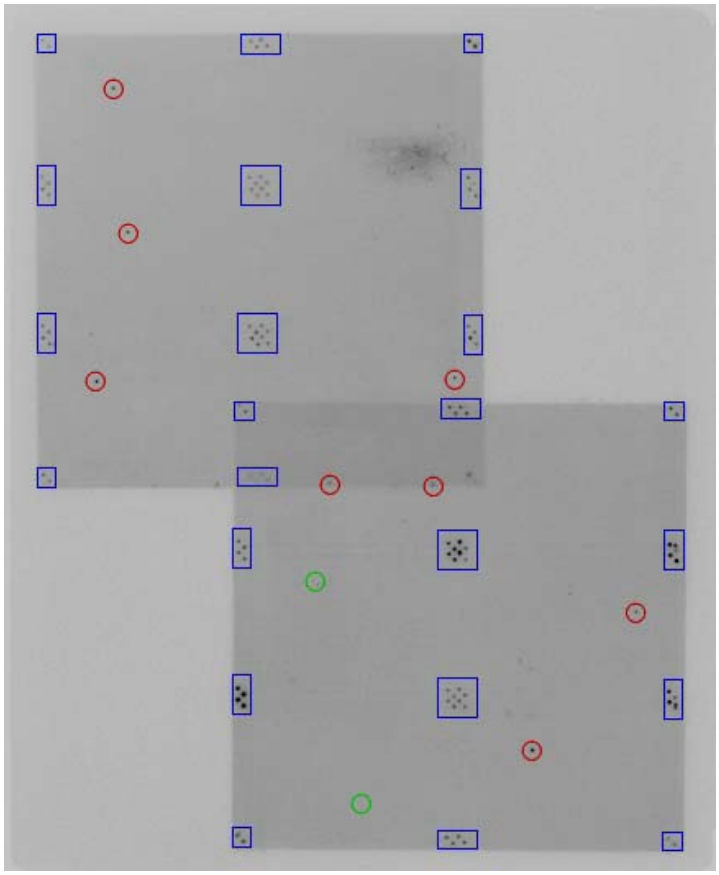


Figure 3.6 CHORI-74 rainbow smelt BAC library filters hybridized with LDH-B probe. The blue rectangles indicate the hybridization by overgo reference probe (*C. briggsea*). The duplicated spots circled by green represent positive hybridization with LDH-B probe. The red circles show the noise, which are single spots.



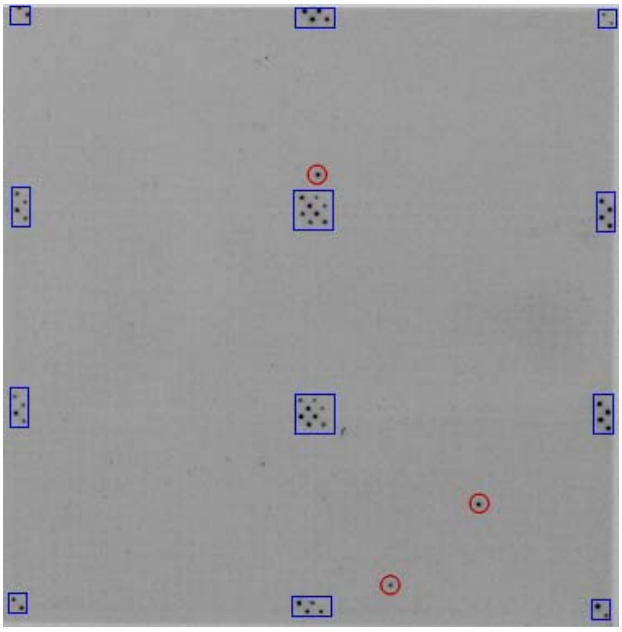
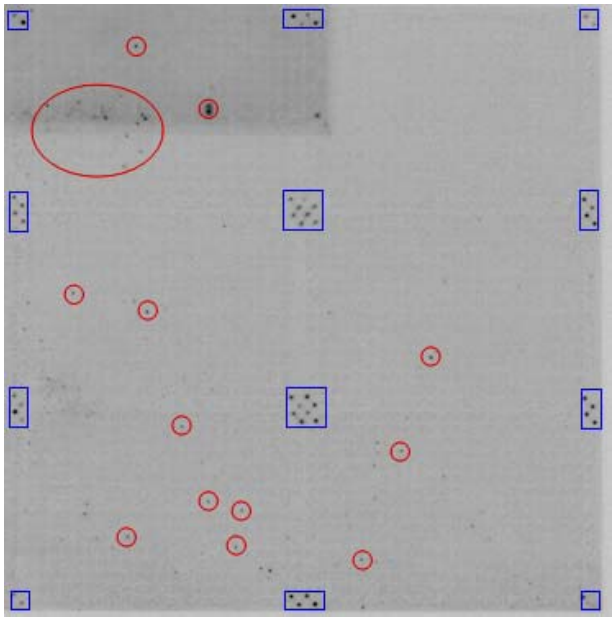


Figure 3.7 CHORI-74 rainbow smelt BAC library filters hybridized with LDH-C probe. The blue rectangles indicate the hybridization by overgo reference probe (*C. briggsea*). The duplicated spots circled by green represent positive hybridization with LDH-C probe. The red circles show the noise, which are single spots.



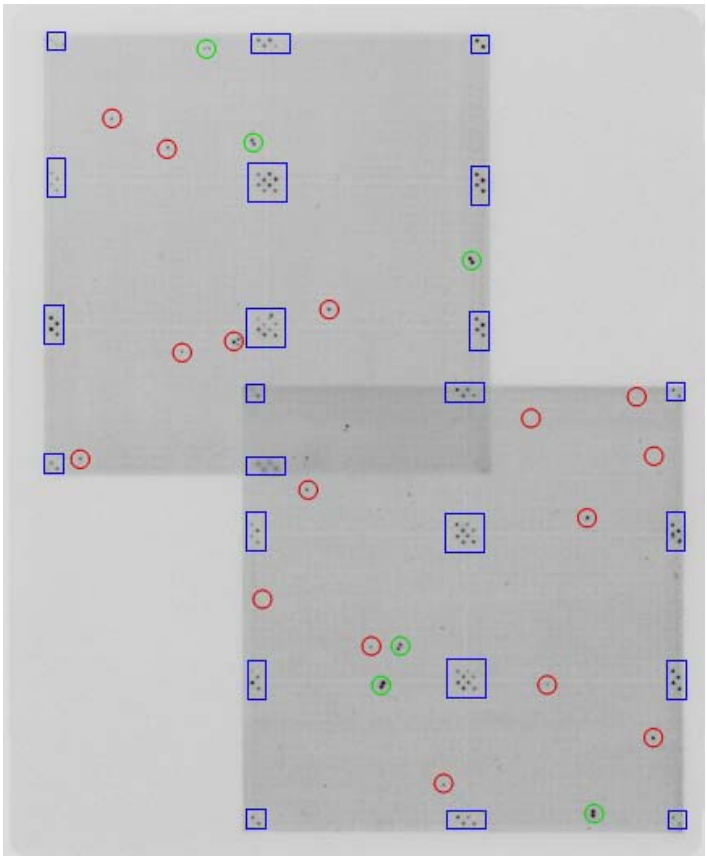
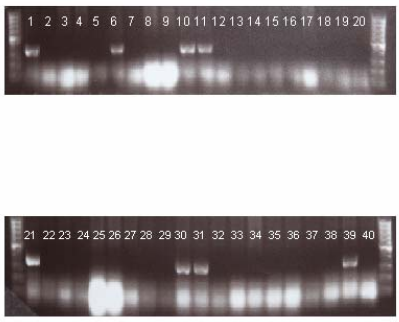


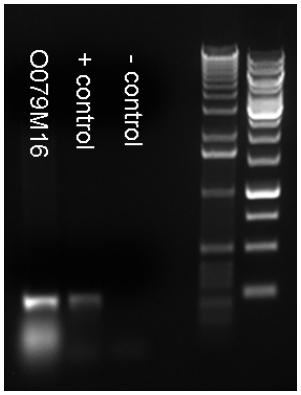
Figure 3.8 The PCR confirmation of positive hybridization BACs with LDH specific primers. (a) Positive BACs for LDH-A (b) Positive BACs for LDH-B (c) Positive BACs for LDH-C. + control: genomic DNA; - control: no template.

(a)



1	20H23	21	109H14
2	23O03	22	112L14
3	28P10	23	113D12
4	29C05	24	113N05
5	30E14	25	113P10
6	33N09	26	115F05
7	41N07	27	118K10
8	46P05	28	119N19
9	53I21	29	120C19
10	55P19	30	120H23
11	56N06	31	120N18
12	72H02	32	123L12
13	79N04	33	125H09
14	82H16	34	126B06
15	95O04	35	128D07
16	100N11	36	130H13
17	104H08	37	133P05
18	108M02	38	136N12
19	109D08	39	138B13
20	109D13	40	144J18

(b)



(c)

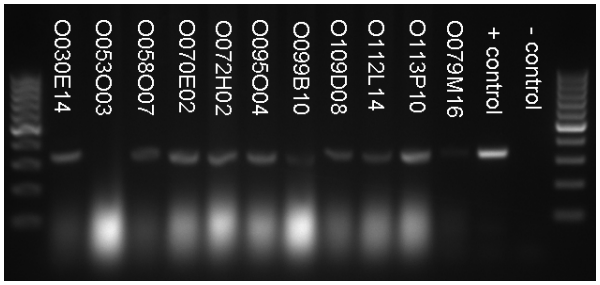
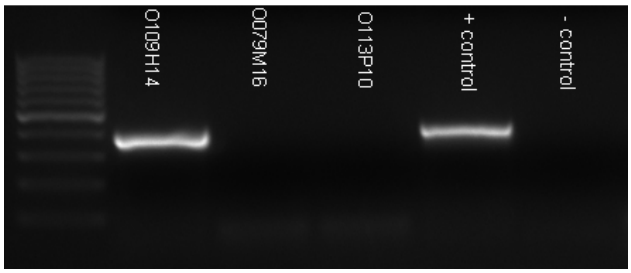


Table 3.4 The PCR positive BAC list from LDH-A, LDH-B and LDH-C hybridization on rainbow smelt BAC library.

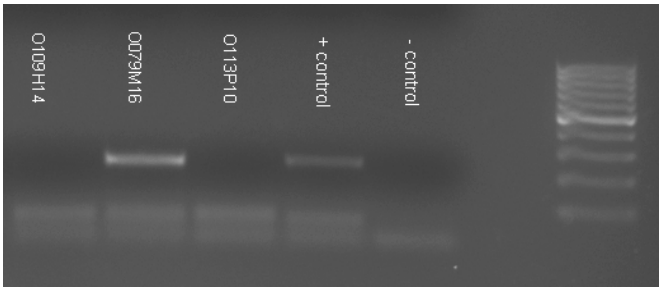
BACs	LDH-A	LDH-B	LDH-C
20H23	√		
33N09	√		
55P19	√		
56N06	√		
109H14	√		
120C19	√		
120H23	√		
136N12	√		
79M16		√	
30E14			√
72H02			√
95O04			√
109D08			√
112L14			√
113P10			√
58O07			√
70E02			√
99B10			√

Figure 3.9 PCR confirmation for each LDH specific primers on rainbow smelt. (a) PCR amplification by LDH-A primers (b) PCR amplification by LDH-B primers (c) PCR amplification by LDH-C primers. O109H14: LDH-A, O079M16: LDH-B, O113P10: LDH-C. + control: genomic DNA; - control: no template.

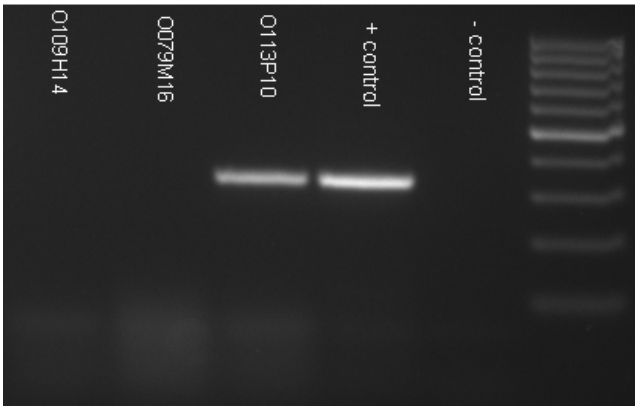
(a)



(b)



(c)



3.3 Shotgun Libraries of LDH

3.3.1 Sequences

Shotgun libraries of BACs containing LDH-A, LDH-B and LDH-C were constructed separately following the method described in Section 2.3. Each LDH BAC library had approximately 2304 clones sent for sequencing at the Michael Smith Genome Sciences Centre.

3.3.2 Assembly

The Phred program was used to read the DNA sequencing trace files and assign a quality value to each called base, and the Phrap program was used to assemble and build contigs from sequences of each BAC (Ewing et al. 1998; Ewing and Green 1998). Another program, Consed, was used to visualize and edit the contig maps (Gordon et al. 1998). The statistical Phrap data of average of sequence length, the size of sequence reads, BAC insert size and BAC coverage for each BAC library are given in Table 3.5.

The initial assembled sequences and contigs visualized using Consed for BAC O0109H14, O0079M16 and O0113P10 are shown in Figures 3.10, 3.11 and 3.13. The assembled contigs from each BAC were ordered and oriented based on the joins of sequencing reads by PCR. The forward primer was designed based on the 5' sequence of each contig and the reverse primer was designed based on the 3' sequence of each contig. PCR with combinations of forward and reverse primers was used to amplify the BAC DNA to make the order and orientation of the contigs in each BAC and to sequence and join the gaps

Table 3.5 **The Phrap statistical data for each rainbow smelt LDH BAC.** The BAC assembly data show average sequence length, total sequence reads, BAC insert size and BAC sequence coverage.

BAC	Average sequence length (bp)	Total sequence reads	BAC insert size (bp)	BAC coverage
O0119H14	867.8	3846	36738	73.9
O0079M16	977.6	4052	96016	44.9
O0113P10	988.2	2790	157151	19.3

between each set of contigs.

Only two contigs of BAC O0079M16 could be joined by this method. In BAC O0079M16, the PCR product was sequenced and used to join the gap between contig 29 and 28 (Figure 3.12). However, none of the gaps was joined in BAC O0109H14 and O0113P10. The reason for the failure of the PCR amplification may be caused by the large unexpected size of gaps between two contigs in the BAC. Another possibility is that unique primers from the end of BAC sequence contigs could not be designed due to presence of repetitive sequences.

3.3.3 Annotation

The consensus sequences from each contig of each BAC were submitted to the consortium for Genomics Research on All Salmonids Project (GRASP) annotation pipeline (http://grasp.mbb.sfu.ca/bacannotations/GRASPbac.html). Each LDH gene was annotated using different data sources such as BLASTX, Uniref, the salmonid and smelt EST databases from the University of Victoria, and CDD.

3.4 LDH gene structures

3.4.1 Exon-intron boundaries

In order to identify the exon-intron boundary of each LDH annotated BAC sequence, cDNA from rainbow smelt brain was amplified by a pair of LDH-A, LDH-B and LDH-C specific primers, respectively, and the resulting PCR products

Figure 3.10 The initial Consed view of the assembly of rainbow smelt LDH-A BAC O0109H14

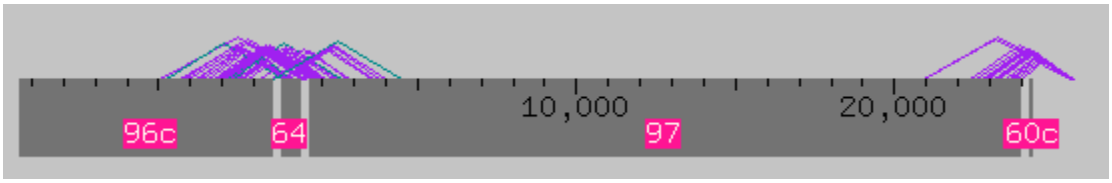


Figure 3.11 The initial Consed view of the assembly of rainbow smelt LDH-B BAC O0079M16

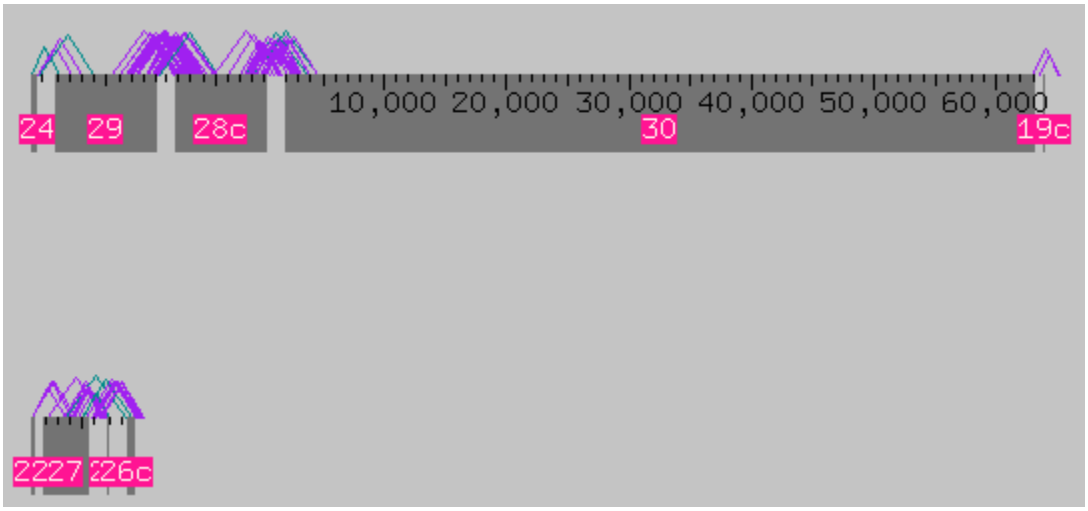


Figure 3.12 The BAC O0079M16 Consed view of re-assemblying. The gap between contig 29 and contig 28 in the initial view was joined by PCR amplification and became contig 31.

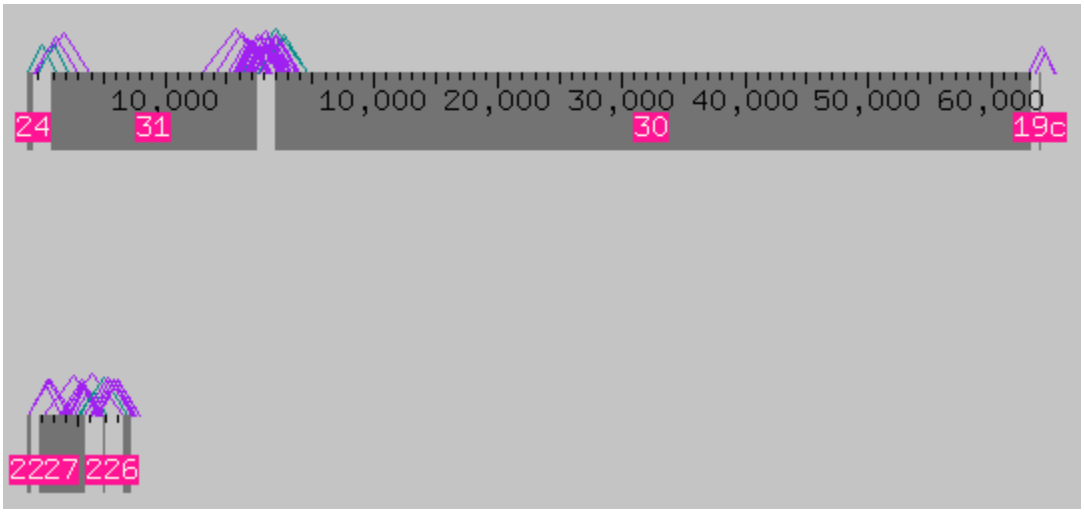
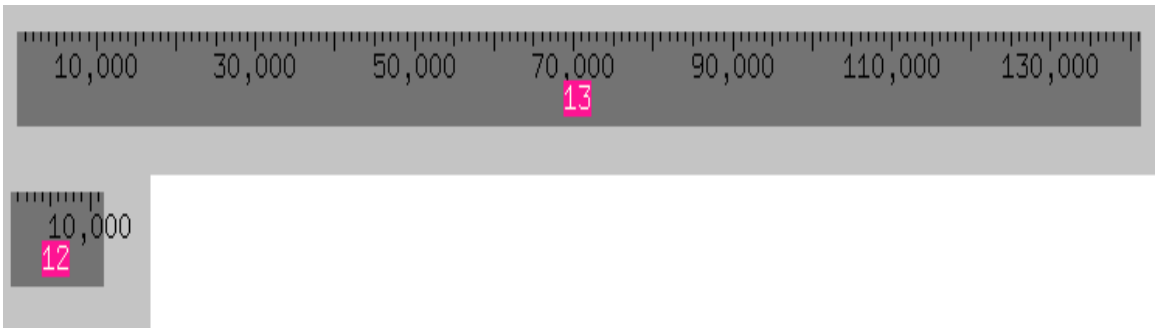


Figure 3.13 The initial Consed view of the assembly rainbow smelt LDH-C BAC O0113P10



were sequenced (Table 3.6). The length of the LDH-A cDNA is 999 bp; LDH-B cDNA is 1005 bp and the LDH-C cDNA is 1005 bp.

To examine the quality of translated nucleotide sequences from each LDH gene, LDH-A, LDH-B and LDH-C cDNA sequences were aligned with their corresponding predicted coding sequences from the BAC by the software Splign (Kapustin et al. 2008). The alignment between the cDNA and predicted coding sequence of LDH-A was 99.8% identity; that of LDH-B has 100% identity and that of LDH-C has 100% identity as well. The information of the percentage of coverage, identity, exon coverage and mismatches and indels compared with genomic DNA is given in Table 3.7.

3.4.2 Coding sequences

The nucleotide alignment of each LDH cDNA sequence and the corresponding genomic DNA indicated the exon-intron boundary for each LDH gene in rainbow smelt. The alignments show that the LDH-A, LDH-B and LDH-C genes in rainbow smelt each has 7 exons. The predicted exon positions of LDH-A, LDH-B and LDH-C are shown in Figures 3.14, 3.15 and 3.16.

3.4.3 Predicted amino acid sequences

After the confirmation of the LDH coding sequences from the cDNA sequences, the coding sequences were translated into amino acid sequences (see Figures 3.14, 3.15 and 3.16) so that this information could be used for an evolutionary study.

Table 3.6 List of primer pairs for PCR amplification of rainbow smelt LDH-A, LDH-B and LDH-C from brain cDNA.

name	primer sequences
smelt_AF_cDNA	5'-ATGTCCACCAAGGAGAAGCTGATCAC-3'
smelt_AR_cDNA	5'-TCACAGGGCGAGCTCCTTCTGCA-3'
smelt_BF_cDNA	5'-ATGTCGTCTGTCATGCAGAAACTG-3'
smelt_BR_cDNA	5'-CTACAGGTCTTTCAGATCCTTCTGGATT-3'
smelt_CF_cDNA	5'-ATGGCCTCAATTCTGCAGAAGCTC-3'
smelt_CR_cDNA	5'-TTACACGTCTTTCAGGTCCTTCTGGATA-3'

Table 3.7 Overview of the cDNA and genomic DNA alignments for LDH-A, LDH-B and LDH-C.

	Covergae (%)	Overall (%)	Exon (%)	Mismatch (bp)
LDH-A	99.8	99.4	99.6	4
LDH-B	100	99.8	99.8	2
LDH-C	100	99.8	99.8	2

Figure 3.14 The exon-intron boundaries in LDH-A of rainbow smelt from Splign. The pink indicates the location of each exon. The red highlights denote the mismatched nucleotides between cDNA and genomic DNA.

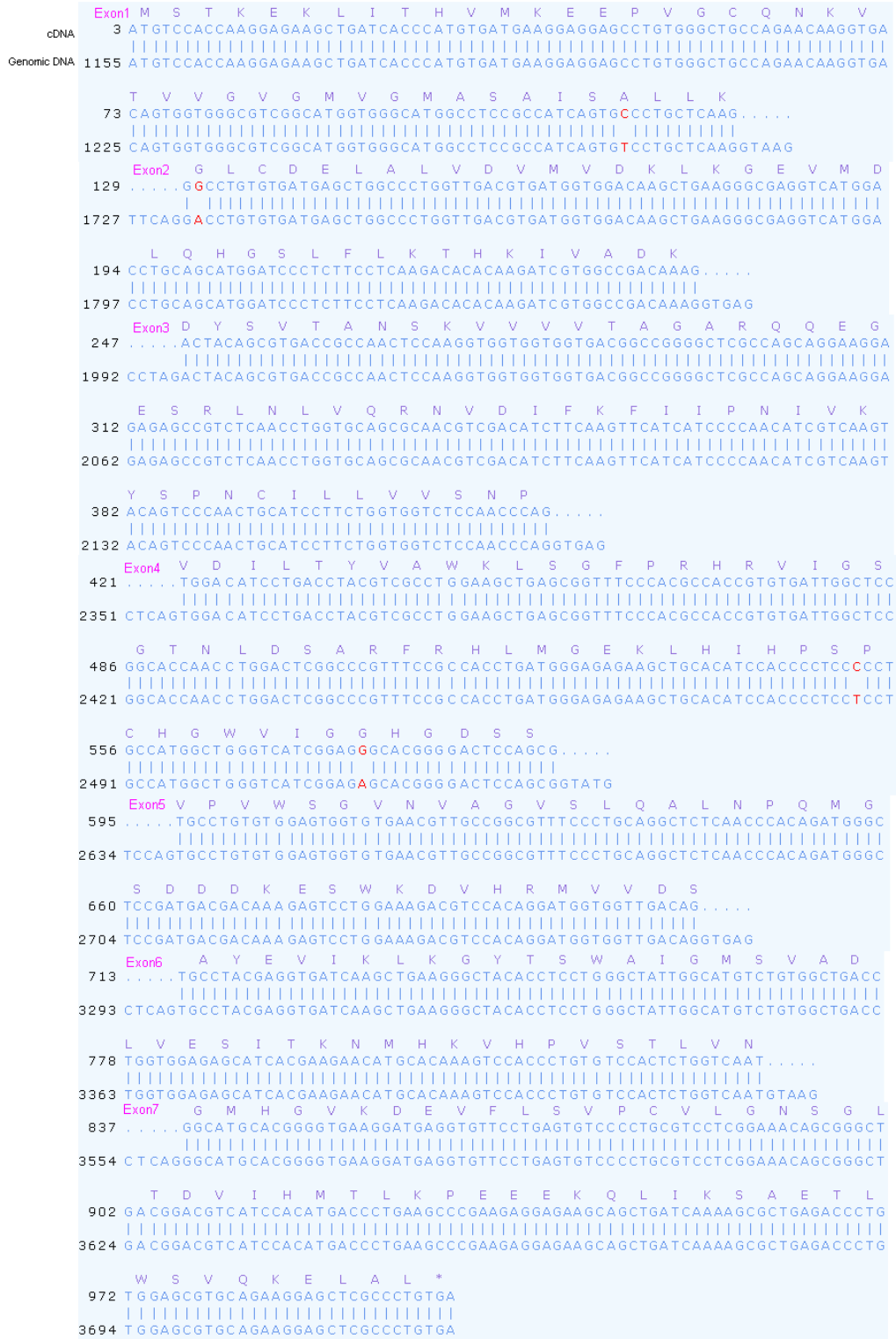


Figure 3.15 The exon-intron boundaries in LDH-B of rainbow smelt from Splign. The pink indicates the location of each exon. The red highlights denote the mismatched nucleotides between cDNA and genomic DNA.

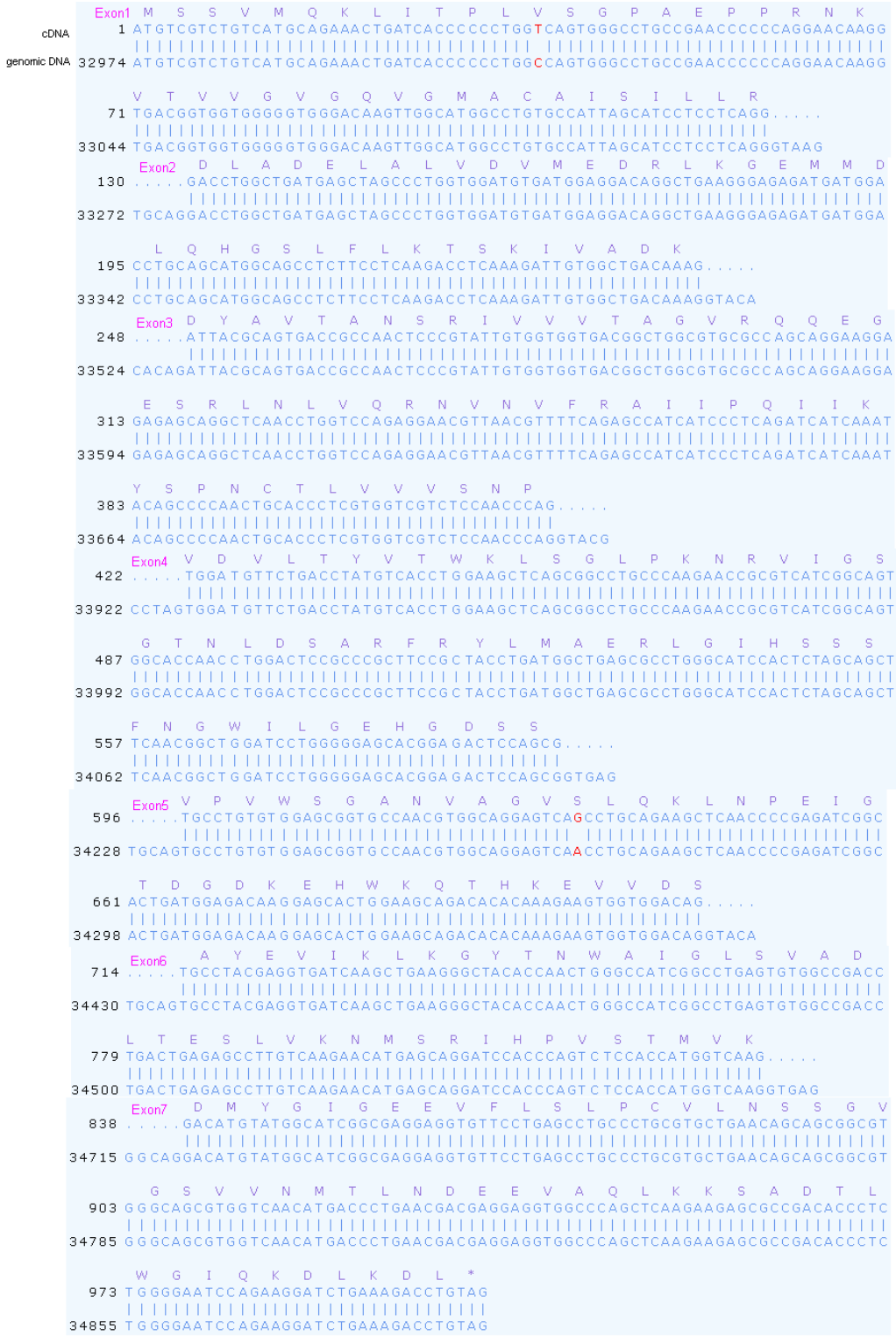
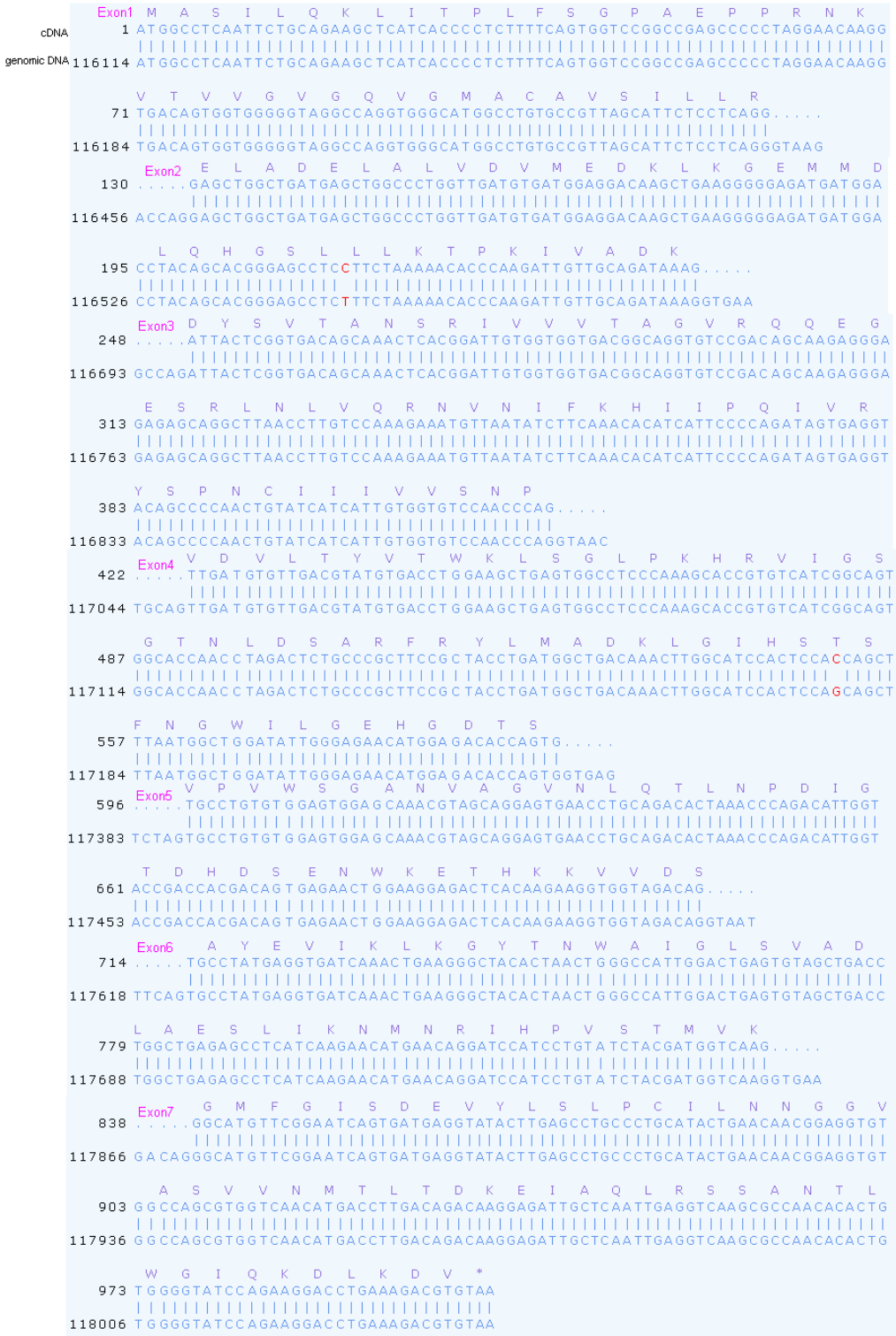


Figure 3.16 The exon-intron boundaries in LDH-C of rainbow smelt from Splign. The pink indicates the location of each exon. The red highlights denote the mismatched nucleotides between cDNA and genomic DNA.



3.5 Evolution of LDH genes in salmonids

In the previous sections I described the isolation and characterization of the three LDH genes in rainbow smelt. The sequences of the LDH-A, LDH-B and LDH-C genes provide information concerning the diploid ancestral state prior to the tetraploidization event in the common ancestor or the salmonids. Here I will use this information to examine the changes in the coding regions of the duplicated LDH-A and LDH-B genes from Atlantic salmon and rainbow trout as well as the single copy LDH-C gene in these species. Note that it has been suggested that there were two copies of LDH-C after the salmonid genome duplication, and that one of these copies was subsequently silenced or lost before the speciation of Atlantic salmon and rainbow trout (Lubieniecki et al. in preparation).

The neofunctionalization hypothesis predicts that after a gene duplication, one of the duplicates continues to fulfil the function of the ancestral gene and is under negative selection, whereas the other duplicate is free to accept mutations that lead to amino acid changes (relaxed selection) and potentially a useful (selectable) novel function. Therefore, one of the predictions of neofunctionalization is that there will be asymmetric amino acid sequence evolution in the paralog lineages. I set out to test this hypthothesis using the rainbow smelt LDH sequences as an outgroup for the Atlantic salmon and rainbow trout LDH paralogues. The main questions that I wished to answer are: (1) is there evidence for positive selection in one of the duplicated genes; (2) are the rates of amino acid substitutions along different lineages the same; (3) do the

rates of amino acid substitutions vary in different parts of a lineage leading to an extant gene product; and (4) are the patterns of amino acid changes the same in the duplicated LDH-A and LDH-B groups and how do they relate to what has happened in the LDH-C group?

3.5.1 Rainbow smelt and salmonids phylogenetic tree

The alignments of the amino acid and corresponding nucleotide sequences from Atlantic salmon, rainbow trout and rainbow smelt for each LDH are shown in Figures 3.17 - 3.22. Note that rainbow trout LDH-A1 is missing 6 amino acids at its C-terminus (18 nucleotides). A Minimum Evolution tree based on the codons corresponding to the first 326 amino acids in LDH-A, and the entire amino acid sequences for LDH-B and LDH-C from rainbow smelt, Atlantic salmon and rainbow trout was constructed using MEGA4 (Tamura et al. 2007) and is shown in Figure. 3.23. Similarly, a Minimum Evolution tree based on the corresponding LDH nucleotide coding sequences from these species was built and is shown in Figure 3.24. Both the amino acid and nucleotide phylogenetic trees confirm that there is a closer relationship between the LDH-Bs and LDH-Cs than either has with an LDH-A, as has been observed previously (Lubieniecki et al. in preparation).

Figure 3.17 Alignment of amino acid sequences for LDH-As from rainbow smelt, Atlantic salmon and rainbow trout.

Figure 3.18 Alignment of nucleotide sequences for LDH-As from rainbow smelt, Atlantic salmon and rainbow trout.

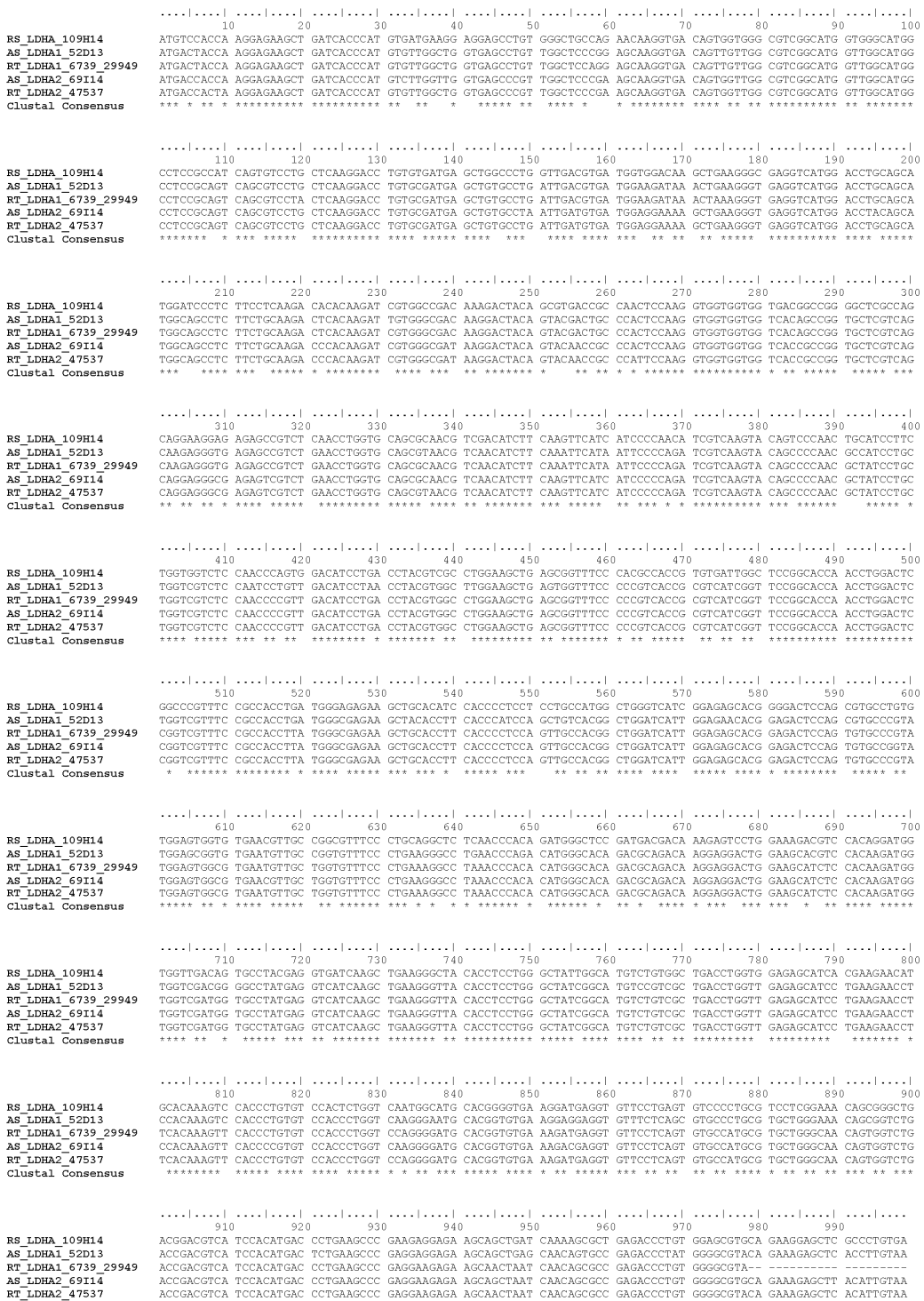


Figure 3.19 Alignment of amino acid sequences for LDH-Bs from rainbow smelt, Atlantic salmon and rainbow trout.

Figure 3.20 Alignment of nucleotide sequences for LDH-Bs from rainbow smelt, Atlantic salmon and rainbow trout.

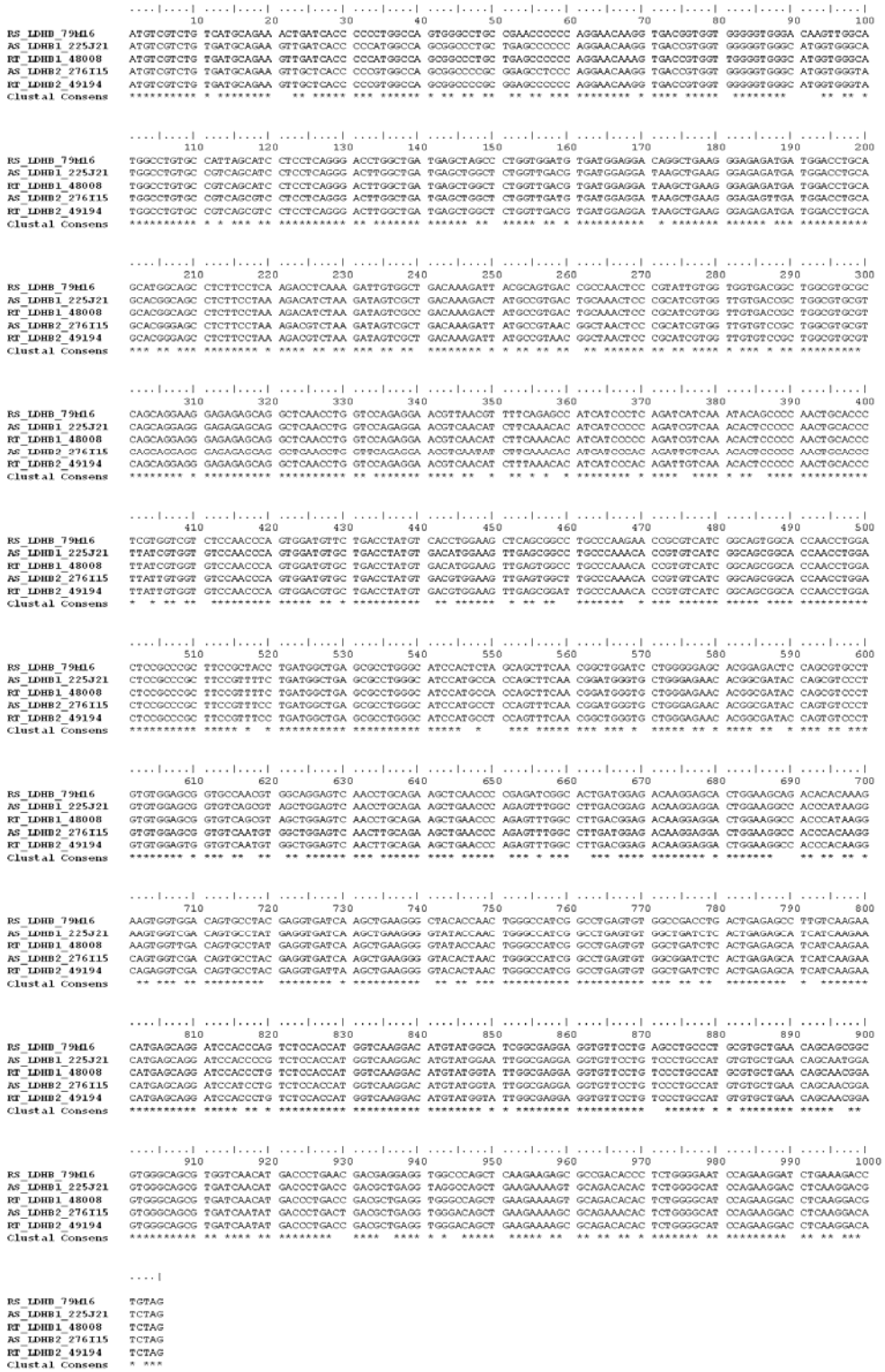


Figure 3.21 Alignment of amino acid sequences for LDH-Cs from rainbow smelt, Atlantic salmon and rainbow trout.

Figure 3.22 Alignment of nucleotide sequences for LDH-Cs from rainbow smelt, Atlantic salmon and rainbow trout.

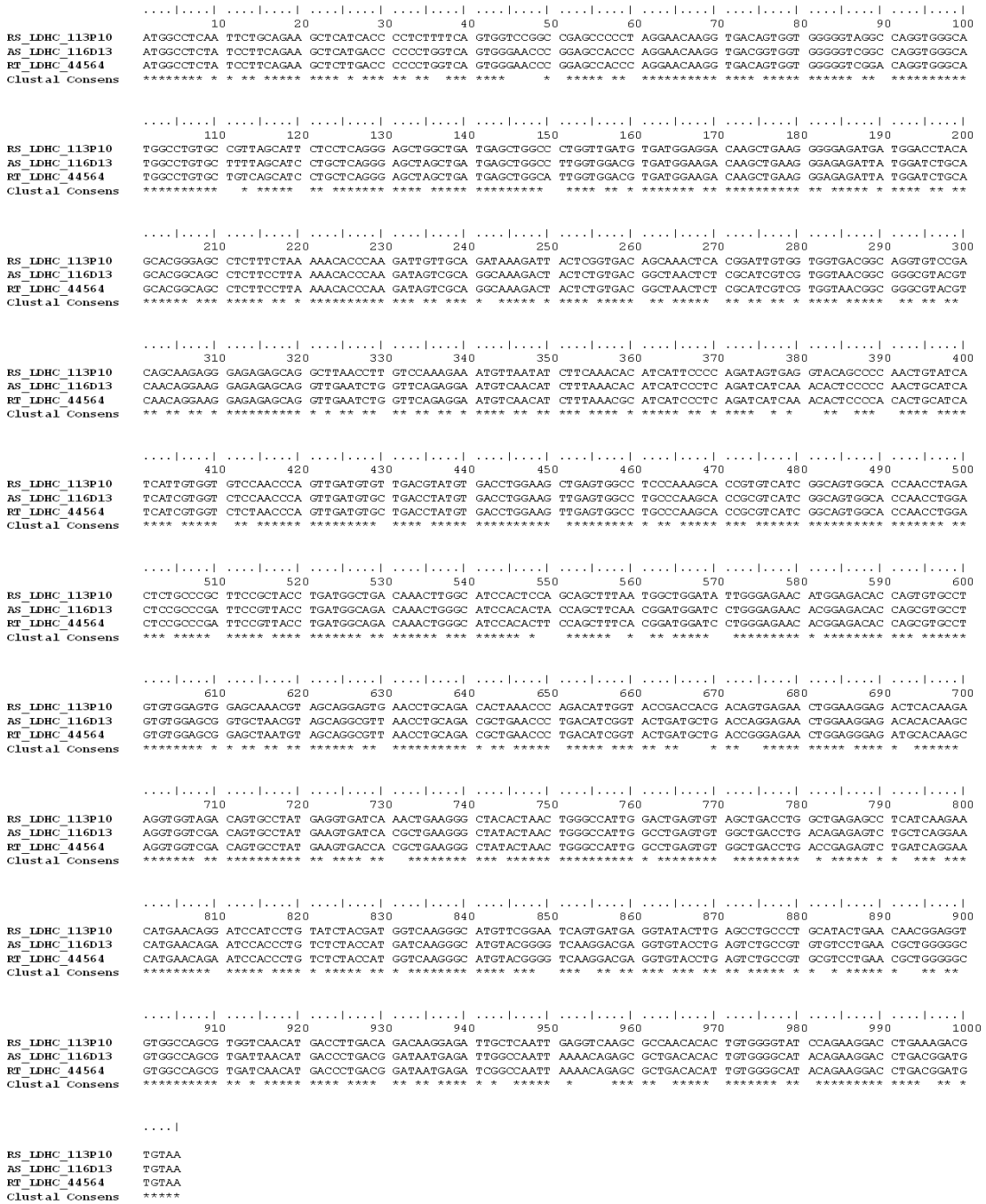


Figure 3.23 The Minimum Evolution tree build of all LDH amino acid coding sequences in rainbow smelt. Outgroup is from tunicate LDH amino acid sequence (*Ciona intestinalis*). AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt. The pink dots indicate the positions of rainbow smelt amino acid coding sequences.

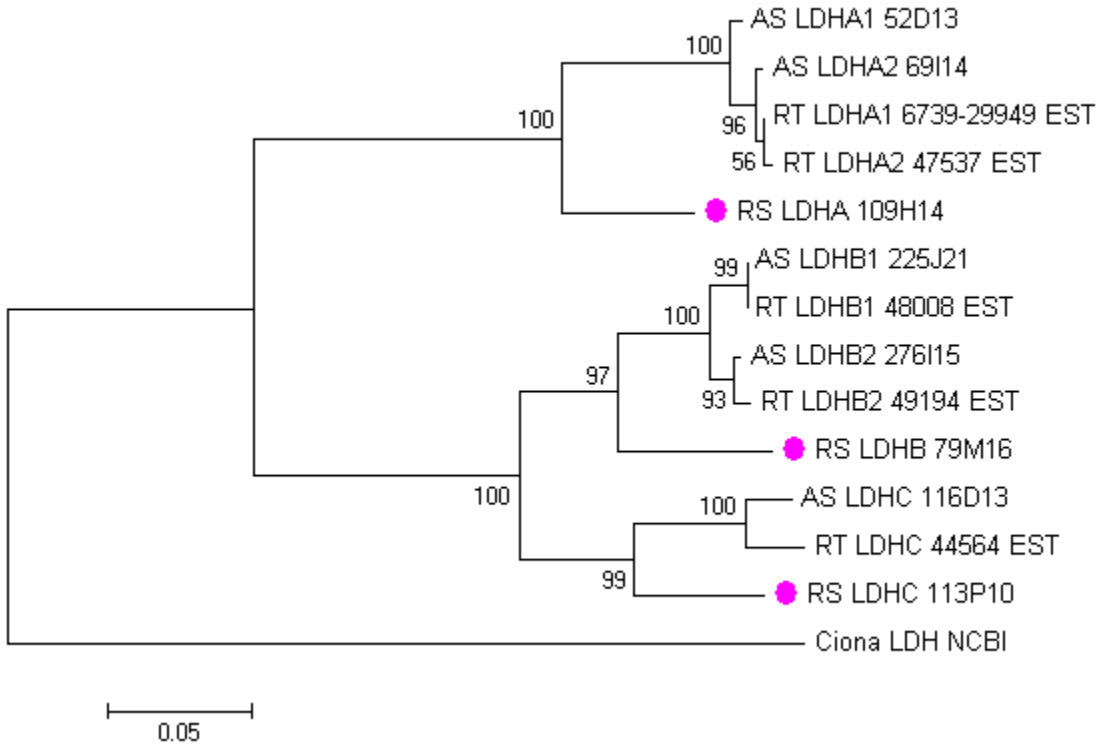
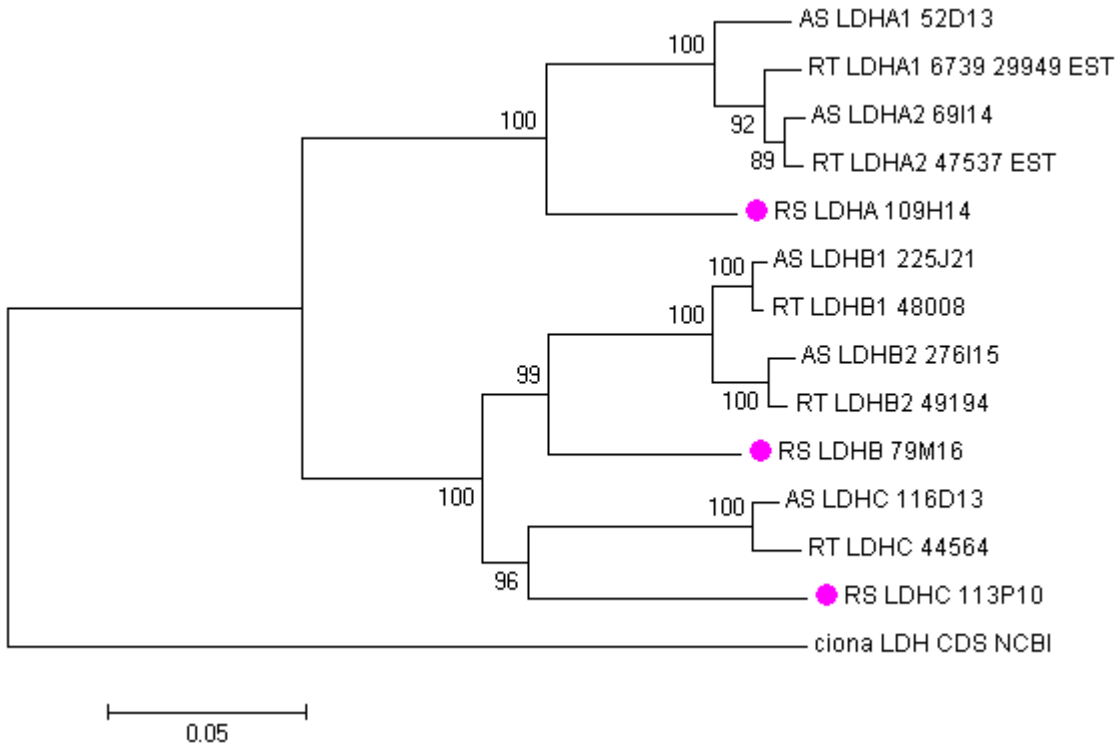


Figure 3.24 The Minimum Evolution tree build of all LDH nucleotide coding sequences in rainbow smelt. Outgroup is from tunicate LDH nucleotide sequence (*Ciona intestinalis*). AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt. The pink dots indicate the positions of rainbow smelt nucleotide coding sequences.



3.5.2 LDH positive selection test

Nucleotide substitutions that do not change the amino acid encoded by a codon are defined as synonymous or silent substitutions, while nucleotide substitutions that do change the amino acid are termed nonsynonymous or replacement substitutions. The ratio of nonsynonymous substitutions over synonymous ones (d_N/d_S) is used as an indicator to determine selection pressures at individual amino acid sites and over the entire coding region (Nei and Gojobori 1986). An excess of nonsynonymous substitutions over synonymous ones ($d_N/d_S > 1$) indicates positive selection. If d_N/d_S equals 1, then the rates of accumulation of nonsynonymous and synonymous substitutions are the same and this is considered an example of neutral selection. If $d_N/d_S < 1$, it indicates that negative or purifying selection has occurred, and this means that certain amino acid substations are incompatible with the function of the protein product.

The d_N/d_S ratios were tested using the web server DATAMONKEY (http://www.datamonkey.org/) (Pond and Frost 2005a) (see section 2.7). Each LDH group contained coding sequences from Atlantic salmon, rainbow trout and rainbow smelt and was tested separately. There was no evidence of positive selection in any LDH group (P-value < 0.01) and all d_N/d_S values were less than one, indicating purifying selection was occurring. The mean d_N/d_S in the LDH-A group is 0.084 with a P-value less than 0.1. The mean d_N/d_S in the LDH-B group is 0.077 (P-value < 0.1). In LDH-C group the mean of d_N/d_S is 0.084 with a P-value less than 0.1. The results of pairwise d_N/d_S for LDH-A, LDH-B and LDH-C

were tested using PAL2NAL (http://www.bork.embl.de/pal2nal/) (Suyama et al. 2006) (see section 2.7). The summary of all the result of pairwise d_N/d_S are shown in Tables 3.8 – 3.10, respectively.

The number of synonymous changes (d_S) can be taken as an indirect measure of evolutionary distance between two sequences. The d_S values for comparisons of rainbow smelt LDH-A to salmonid LDH-As ranged from 1.069 – 1.105. This reflects the speciation divergence of the rainbow smelt and the common ancestor of the salmonids. The d_S values between salmonid LDH-A1 and LDH-A2, which correspond to the salmonid genome duplication event, ranged from 0.119 – 0.416 while the speciation of Atlantic salmon and rainbow trout are reflected in the LDH-A1 and LDH-A2 comparisons between the species (d_S values of 0.23 and 0.072). Similar values were observed for the corresponding LDH-B and LDH-C comparisons. These results are consistent with the evolutionary relationships that predict the speciation of rainbow smelt and salmonids occurred before the salmonid genome duplication, which in term occurred prior to the speciation of Atlantic salmon and rainbow trout.

Table 3.8 The pariwise ratio of nonsynonymous over synonymous substitutions (d_N/d_S) for LDH-A among rainbow smelt and salmonids.

LDH-A	AS LDH-A1	AS LDH-A2	RT LDH-A1	RT LDH-A2
AS LDH-A2	0.0193 (dN/dS =0.0080/0.4155)	–	–	–
RT LDH-A1	0.0296 (dN/dS=0.0069/0.2327)	0.0236 (dN/dS=0.0041/0.1719)	–	–
RT LDH-A2	0.0222 (dN/dS =0.0081/0.3642)	0.0369 (dN/dS=0.0027/0.0722)	0.0115 (dN/dS =0.0014/0.1191)	–
RS LDH-A	0.0639 (dN/dS=0.0686/1.0742)	0.0596 (dN/dS=0.0659/1.1054)	0.0629 (dN/dS=0.0668/1.0624)	0.0635 (dN/dS=0.0679/1.0694)

Table 3.9 The pariwise ratio of nonsynonymous over synonymous substitutions (d_N/d_S) for LDH-B among rainbow smelt and salmonids.

LDH-B	AS LDH-B1	AS LDH-B2	RT LDH-B1	RT LDH-B2
AS LDH-B2	0.0552 (dN/dS=0.0130/0.2356)	–	–	–
RT LDH-B1	0.0001 (dN/dS=0.0001/0.0533)	0.0556 (dN/dS=0.0130/0.2341)	–	–
RT LDH-B2	0.0514 (dN/dS=0.0115/0.2246)	0.0486 (dN/dS=0.0038/0.0790)	0.0489 (dN/dS=0.0116/0.2369)	–
RS LDH-B	0.0491 (dN/dS=0.0491/1.0009)	0.0535 (dN/dS=0.0578/1.0800)	0.0483 (dN/dS=0.0492/1.0193)	0.0500 (dN/dS=0.0562/1.1241)

Table 3.10 The pariwise ratio of nonsynonymous over synonymous substitutions (d_N/d_S) for LDH-C among rainbow smelt and salmonids.

LDH-C	AS LDH-C	RT LDH-C
RT LDH-C	0.2226 (dN/dS=0.0160/0.0717)	–
RS LDH-C	0.0449 (dN/dS=0.0558/1.2439)	0.0476 (dN/dS=0.0607/1.2756)

3.5.3 Phylogenetic analysis

Although the d_N/d_S analysis did not reveal any evidence for positive selection, I wondered where and when amino acid substitutions had occurred in the Atlantic salmon and rainbow trout lineages. I also wanted to find out if the rates were the same in the rainbow smelt lineage and the salmonid lineages. In order to study the rates of amino acid substitutions along different lineages I retrieved the LDHs from the genome annotations of zebrafish, tetraodon, takifugu, medaka, killifish, stickleback and dogfish from Ensembl 55 (Hubbard et al. 2009). Since most teleosts from Ensembl 55 have two LDH-Bs and lack a curated annotation of LDH-C, the identification of LDH-C for these teleosts was made by making a phylogenetic tree using MEGA4 (Tamura et al. 2007) (Figure 3.25). The bootstrap value of the of the LDH-C clade is 100, which provides strong evidence that what has been called LDH-B2 in tetraodon, takifugu, stickleback and medaka are actually LDH-Cs. The amino acid sequence alignments of each type of LDH gene were made using ClustalX (Figure 3.26, 3.27 and 3.28).

The number of invariant amino acid sites in each type of LDH is a measure of how much negative selection there is in these proteins. In the LDH-A group, the alignment of amino acid sequences indicates that 207 amino acid sites are invariant over the total length of 332 residues (62%) from all the fish (Figure 3.26). For the LDH-B group there are 209 amino acid residues in LDH-B group are conservative out of a total length of 334 (63%) (Figure 3.27), and for fish LDH-Cs 245 out of 334 amino acid residues are invariant (73%) (Figure

3.28). These numbers reflect what was observed with the d_N/d_S analysis and show that in general LDH is a highly conserved protein.

Figure 3.25 Neighbor-Joining tree build of LDH-B and LDH-C amino acid coding sequences among salmonids, rainbow smelt and other fish. AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt.

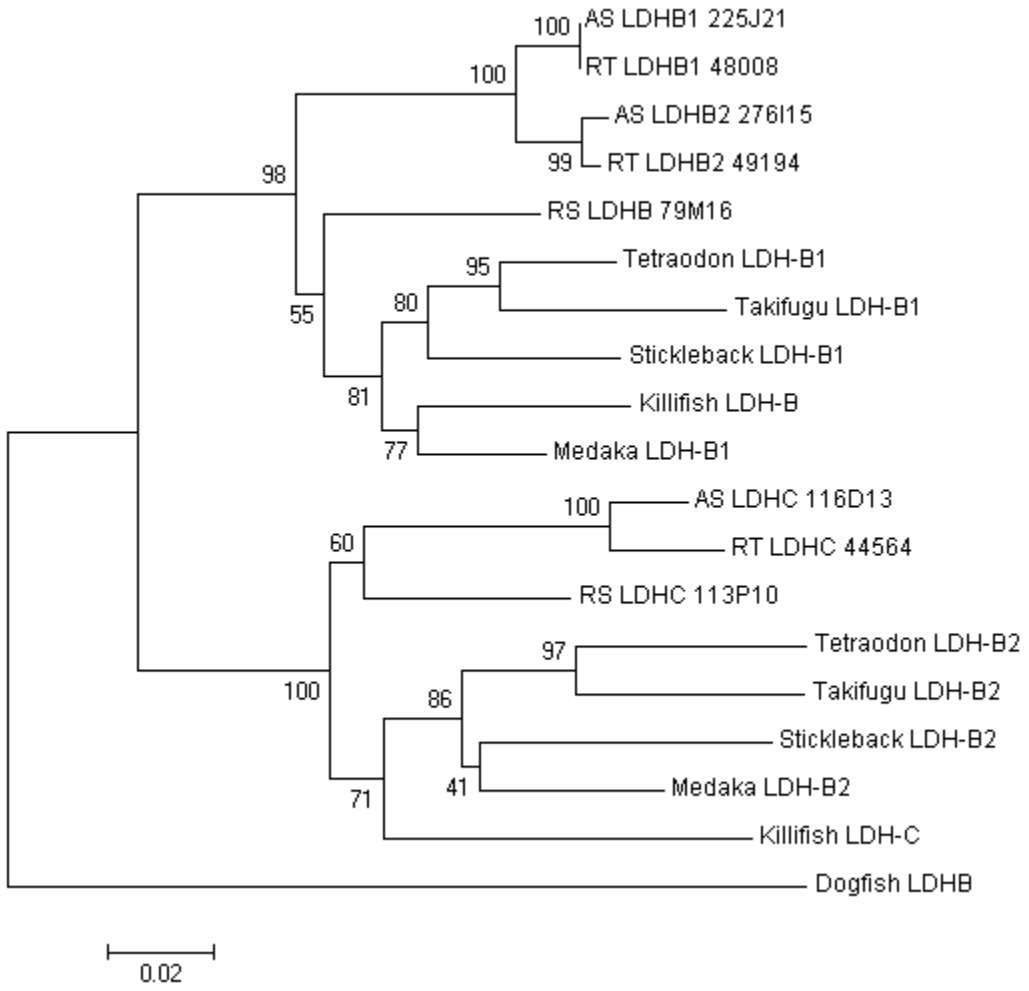


Figure 3.26 Amino acid sequence alignment for LDH-A among samonids, rainbow smelt and other fish. AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt.

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              10        20        30        40        50        60        70        80        90       100
RS_LDHA_109H14      -MSTKEKLIT HVMKEEPVGC QNKVTVVGVG MVGMASAISV LLKDLCDELA LVDVMVDKLK GEVMDLQHGS LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
AS_LDHA1_52D13      -MTTKEKLIT HVLAGEPVGS RSKVTVVGVG MVGMASAVSV LLKDLCDELC LIDVMEDKLK GEVMDLQHGS LFCKTHKIVG DKDYSTTAHS KVVVVTAGAR
RT_LDHA1_6739-29949 -MTTKEKLIT HVLAGEPVGS RSKVTVVGVG MVGMASAVSV LLKDLCDELC LIDVMEDKLK GEVMDLQHGS LFCKTHKIVG DKDYSTTAHS KVVVVTAGAR
AS_LDHA2_69I14      -MTTKEKLIT HVLVGEPVGS RSKVTVVGVG MVGMASAVSV LLKDLCDELC LIDVMEEKLK GEVMDLQHGS LFCKTHKIVG DKDYSTTAHS KVVVVTAGAR
RT_LDHA2_47537      -MTTKEKLIT HVLAGEPVGS RSKVTVVGVG MVGMASAVSV LLKDLCDELC LIDVMEEKLK GEVMDLQHGS LFCKTHKIVG DKDYSTTAHS KVVVVTAGAR
Tetraodon_LDH-A     -MSTKEKLIG HVMKEQPIGC QNKVTVVGVG MVGMASAISI LLKDLCDELA LIDVMEDKLK GEAMDLQHGS LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
Takifugu_LDH-A      -MSTKEKLIS HVMKEQPIGC QNKVTVVGVG MVGMASAISI LLKDLCDELA LVDVMEDKLK GEAMDLQHGS LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
Killifish_LDH-A     -MSTQEKLIS HVMKEEPVGC RNKVTVVGVG MVGMASAISV LLKDLCDELA LVDVMEDKLK GEAMDLQHGA LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
Stickleback_LDH-A   -MSTKEKLIS HVMKEEPVGS ANKVTVVGVG MVGMACAISV LLKDLCDELA LVDVMEDKLK GEIMDLQHGS LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
Zebrafish_LDH-A     MASTKEKLIA HVSKEQPAGP TNKVTVVGVG MVGMAAAVSI LLKDLTDELA LVDVMEDKLK GEAMDLQHGS LFLKTHKIVA DKDYSVTANS KVVVVTAGAR
Dogfish_LDHA        MATLKDKLIG HLATSQEPRS YNKITVVGVG AVGMACAISI LMKDLADEVA LVDVMEDKLK GEMMDLQHGS LFLHTAKIVS GKDYSVSAGS KLVVITAGAR
Clustal Consensus     : ::***  *:   :      .*:******  ****.*:*: *:*** **:. *:*** :*** ** ******: ** :* ***. .****.:* * *:**:*****

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            110       120       130       140       150       160       170       180       190       200
RS_LDHA_109H14      QQEGESRLNL VQRNVDIFKF IIPNIVKYSP NCILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSAR FRHLMGEKLH IHPSSCHGWV IGEHGDSSVP
AS_LDHA1_52D13      QQEGESRLNL VQRNVNIFKF IIPQIVKYSP NAILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSGR FRHLMGEKLH LHPSSCHGWI IGEHGDSSVP
RT_LDHA1_6739-29949 QQEGESRLNL VQRNVNIFKF IIPQIVKYSP NAILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSGR FRHLMGEKLH LHPSSCHGWI IGEHGDSSVP
AS_LDHA2_69I14      QQEGESRLNL VQRNVNIFKF IIPQIVKYSP NAILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSGR FRHLMGEKLH LHPSSCHGWI IGEHGDSSVP
RT_LDHA2_47537      QQEGESRLNL VQRNVNIFKF IIPQIVKYSP NAILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSGR FRHLMGEKLH LHPSSCHGWI IGEHGDSSVP
Tetraodon_LDH-A     QQEGESRLNL VQRNVNIFKF IIPNIVKYSP NCILMVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSAR FRHLMGEKLN LHPSSCHGWI IGEHGDSSVP
Takifugu_LDH-A      QQEGESRLNL VQRNVNIFKF IIPNIVKYSP NCILMVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSAR FRHLMGEKLN LHPSSCHGWI IGEHGDSSVP
Killifish_LDH-A     QQEGESRLNL VQRNVNIFKF IIPNIVKYSP NCILLVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSAR FRHLMGEKFH LHPSSCHGWI VGEHGDSSVA
Stickleback_LDH-A   QQEGESRLNL VQRNVNIFKF IIPNIVKHSP NCIILVVSNP VDILTYVAWK LSGFPRHRVI GSGTNLDSAR FRHLMGEKLN IHPSSCHGWI IGEHGDSSVP
Zebrafish_LDH-A     QQEGESRLNL VQRNVNIFKF IIPNIIKYSP NCILLVVSNP VDILTYVAWK LSGLPRNRVI GSGTNLDSAR FRYLMGEKLG IHPSSCHGWV VGEHGDSSVP
Dogfish_LDHA        QQEGESRLNL VQRNVNIFKF IIPDIVKHSP DCIILVVSNP VDVLTYVAWK LSGLPMHRII GSGCNLDSAR FRYLMGERLG VHSSSCHGWV IGEHGDSSVP
Clustal Consensus   ********** *****:**** ***:*:*:** :.*::***** **:******* ***:* :*:* *** ****.* **:****::  :*.******: :********.

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            210       220       230       240       250       260       270       280       290       300
RS_LDHA_109H14      VWSGVNVAGV SLQALNPQMG SDDDKESWKD VHRMVVDSAY EVIKLKGYTS WAIGMSVADL VESITKNMHK VHPVSTLVNG MHGVKDEVFL SVPCVLGNSG
AS_LDHA1_52D13      VWSGVNVAGV SLKGLNPDMG TDADKEDWKH VHKMVVDGAY EVIKLKGYTS WAIGMSVADL VESILKNLHK VHPVSTLVKG MHGVKEEVFL SVPCVLGNSG
RT_LDHA1_6739-29949 VWSGVNVAGV SLKGLNPHMG TDADKEDWKH LHKMVVDGAY EVIKLKGYTS WAIGMSVADL VESILKNLHK VHPVSTLVQG MHGVKDEVFL SVPCVLGNSG
AS_LDHA2_69I14      VWSGVNVAGV SLKGLNPHMG TDADKEDWKH LHKMVVDGAY EVIKLKGYTS WAIGMSVADL VESILKNLHK VHPVSTLVKG MHGVKDEVFL SVPCVLGNSG
RT_LDHA2_47537      VWSGVNVAGV SLKGLNPHMG TDADKEDWKH LHKMVVDGAY EVIKLKGYTS WAIGMSVADL VESILKNLHK VHPVSTLVQG MHGVKDEVFL SVPCVLGNSG
Tetraodon_LDH-A     VWSGVNVAGV SLQSLNPQMG TESDTKNWKD VHKQVVDGAY EVIKLKGYTS WAIGMSVADL VESITKNLHK VHPVSTLVQG LHGVKDEVFL SIPSVLSNSG
Takifugu_LDH-A      VWSGVNVAGV SLQSLNPKMG TEGDSENWKD VHKKVVDGAY EVIKLKGYTS WAIGMSVADL VESITKNLHK VHPVSTLVQG LHGVKDEVFL SIPSVLGNSG
Killifish_LDH-A     VWSGVNIAGV SLQTLNPNMG ADGDSENWKE LHKKVVDGAY EVIKLKGYTS WAIGMSVADL VESIVKNLHK VHPVSTLVQG MHGVKDEVFL SIPCVLGNSG
Stickleback_LDH-A   VWSGVNVAGV SLQGLNPQMG AEGDSENWKA VHKEVVDGAY EVIRLKGYTS WAIGMSVADL VESILKNLHK VHPVSTLVQG MHGVKDEVFL SVPSVLGNSG
Zebrafish_LDH-A     VWSGVNVAGV SLQALNPDLG TDKDKEDWKS VHKMVVDSAY EVIKLKGYTS WAIGMSVADL CESILKNMHK CHPVSTLVKG MHGVNEEVFL SVPCILGNNG
Dogfish_LDHA        VWSGMNVAGV SLKELHPELG TDKDKENWKK LHKDVVDSAY EVIKLKGYTS WAIGLSVADL AETIMKNLCR VHPVSTMVKD FYGIKNDVFL SLPCVLDNHG
Clustal Consensus   ****:*:*** **: *:*.:* :: *.:.**  :*: ***.** ***:****** ****:*****  *:* **: :  *****:*:. ::*::::*** *:*.:*.* *

                    ....|....| ....|....| ....|....| ....
                            310       320       330
RS_LDHA_109H14      LTDVIHMTLK PEEEKQLIKS AETLWSVQKE LAL-
AS_LDHA1_52D13      LTDVIHMTLK PEEEKQLSNS AETLWGVQKE LTL-
RT_LDHA1_6739-29949 LTDVIHMTLK PEEEKQLINS AETLWGV--- ----
AS_LDHA2_69I14      LTDVIHMTLK PEEEKQLINS AETLWGVQKE LTL-
RT_LDHA2_47537      LTDVIHMTLK PEEEKQLINS AETLWGVQKE LTL-
Tetraodon_LDH-A     LTDVIHMRLK PEEEKQLMKS AETLWDVQKE LTM-
Takifugu_LDH-A      LTDVIHMTLK PEEEKQLVKS AETLWGVQKE LTIL
Killifish_LDH-A     LTDVIHMTLK PEEEKQLVKS AETLWGVQKE LTL-
Stickleback_LDH-A   LTDVIHMTLK PDEEKQLMTS ADTLWGVQKE LVL-
Zebrafish_LDH-A     LTDVVHMTLK PEEEKQLVKS AETLWGVQKE LTL-
Dogfish_LDHA        ISNIVKMKLK PDEEQQLQKS ATTLWDIQKD LKF-
Clustal Consensus   ::::::* ** *:**:** .* * ***.:
```

Figure 3.27 Amino acid sequence alignment for LDH-B among samonids, rainbow smelt and other fish. AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt.

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            10        20        30        40        50        60        70        80        90       100
RS_LDHB_79M16       MSSVMQKLIT PLASGPAEPP RNKVTVVGVG QVGMACAISI LLRDLADELA LVDVMEDRLK GEMMDLQHGS LFLKTSKIVA DKDYAVTANS RIVVVTAGVR
AS_LDHB1_225J21     MSSVMQKLIT PMASGPAEPP RNKVTVVGVG MVGMACAVSI LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTSKIVA DKDYAVTANS RIVVVTAGVR
RT_LDHB1_48008      MSSVMQKLIT PMASGPAEPP RNKVTVVGVG MVGMACAVSI LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTSKIVA DKDYAVTANS RIVVVTAGVR
AS_LDHB2_276I15     MSSVMQKLLT PVASGPAEPP RNKVTVVGVG MVGMACAVSV LLRDLADELA LVDVMEDKLK GELMDLQHGS LFLKTSKIVA DKDYAVTANS RIVVVSAGVR
RT_LDHB2_49194      MSSVMQKLLT PVASGPAEPP RNKVTVVGVG MVGMACAVSV LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTSKIVA DKDYAVTANS RIVVVSAGVR
Tetraodon_LDH-B1    MASILQKLFH PLLSGPPEPP RNKVTVVGVG QVG-ACAVTI LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTPKIVA DKDYSVTANS RVVVVTAGVR
Takifugu_LDH-B1     MSSVLQKLIS PLAGSSSEPP RNKVTVVGVG QVGMACAISI LLRDLADELA LVDVMEDRLK GEMMDLQHGS LFLKTSKIVA GKDYSVTADS RLVVVTAGVR
Killifish_LDH-B     MSSVLQKLIT PLASSSAEPP RNKVTVVGVG QVGMACAVSI LLRDLCDELA LVDVMEDRLK GEMMDLQHGL LFLKTSKVVA DKDYAVTANS RLVVVTAGVR
Stickleback_LDH-B1  MSSVLQKLIT PLASSPAEPP RNKVTVVGVG QVGMACAISI LLRDLCDELA LVDVMEDRLK GEMMDLQHGS LFLKTSKIVA DKDYVVTANS RLVVVTAGVR
Medaka_LDH-B1       MSSVLQKLIT PLASNASETP RNKVTVVGVG QVGMACAVSI LLRDLCDELA LVDVMEDRLK GEMMDLQHGS LFLKTSKIVA DKDYAVTANS RLVVVTAGVR
Dogfish_LDHB        MATVQQKLIT PVSQERSDAS RNKVTVVGVG QVGMACAVSI LLRELTDEIA LVDVLEDKLK GEMMDLLHGS LFLKTPKIMA NKDYGITANS RVVVVTAGAR
Clustal Consensus   *::: ***:  *:    .:.. **********  ** ***::: ***:* **:* ****:**:** **:*** **  *****.*::* .*** :**:* *:***:**.*


                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            110       120       130       140       150       160       170       180       190       200
RS_LDHB_79M16       QQEGESRLNL VQRNVNVFRA IIPQIIKYSP NCTLVVVSNP VDVLTYVTWK LSGLPKNRVI GSGTNLDSAR FRYLMAERLG IHSSSFNGWI LGEHGDSSVP
AS_LDHB1_225J21     QQEGESRLNL VQRNVNIFKH IIPQIVKHSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMAERLG IHATSFNGWV LGEHGDTSVP
RT_LDHB1_48008      QQEGESRLNL VQRNVNIFKH IIPQIVKHSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMAERLG IHATSFNGWV LGEHGDTSVP
AS_LDHB2_276I15     QQEGESRLNL VQRNVNIFKH IIPQIVKHSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMAERLG IHASSFNGWV LGEHGDTSVP
RT_LDHB2_49194      QQEGESRLNL VQRNVNIFKH IIPQIVKHSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMAERLG IHASSFNGWV LGEHGDTSVP
Tetraodon_LDH-B1    QQEGESRLNL VQRNVNIFKH IVPQIVRYSP ECVIIVVSNP VDVLTYVTWK LSGLPKRRVI GSGTNLDSAR FRFLVADKLG LHASSFNGWI LGEHGDTSVP
Takifugu_LDH-B1     QQEGESRLNL VQRNVNVFKS IIPQIVKHSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMAERLG IHASSFNGWV LGEHGDTSVP
Killifish_LDH-B     QQEGESRLNL VQRNVNVFKC IIPQIIKYSP NCTILVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYMMAERLG IHASAFNGWV LGEHGDTSVP
Stickleback_LDH-B1  QQEGESRLNL VQRNVNVFKV IIPQIVKYSP NCTLIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMSERLG IHASSFNGWV LGEHGDTSVP
Medaka_LDH-B1       QQEGESRLNL VQRNVNVFKA IIPQIIKYSP NCTIVVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMAERLG IHASSFNGWV LGEHGDTSVP
Dogfish_LDHB        QQEGESRLNL VQRNVNIFKF IIPQIVKYSP NCTIIVVSNP VDILTYVTWK ISGFPKNRVI GSGCNLDSAR FRYLMAEKLG LHPSSCHGWV LGEHGDSSVP
Clustal Consensus   ********** ******:*:  *:***:::** :*.::***** **:******* :**:**.*** *** ****** **:::::** :*.:: :**: ******:***


                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            210       220       230       240       250       260       270       280       290       300
RS_LDHB_79M16       VWSGANVAGV NLQKLNPEIG TDGDKEHWKQ THKEVVDSAY EVIKLKGYTN WAIGLSVADL TESLVKNMSR IHPVSTMVKD MYGIGEEVFL SLPCVLNSSG
AS_LDHB1_225J21     VWSGVSVAGV NLQKLNPEFG LDGDKEDWKA THKEVVDSAY EVIKLKGYTN WAIGLSVADL TESIIKNMSR IHPVSTMVKD MYGIGEEVFL SLPCVLNSNG
RT_LDHB1_48008      VWSGVSVAGV NLQKLNPEFG LDGDKEDWKA THKEVVDSAY EVIKLKGYTN WAIGLSVADL TESIIKNMSR IHPVSTMVKD MYGIGEEVFL SLPCVLNSNG
AS_LDHB2_276I15     VWSGVNVAGV NLQKLNPEFG LDGDKEDWKA THKAVVDSAY EVIKLKGYTN WAIGLSVADL TESIIKNMSR IHPVSTMVKD MYGIGEEVFL SLPCVLNSNG
RT_LDHB2_49194      VWSGVNVAGV NLQKLNPEFG LDGDKEDWKA THKAEVDSAY EVIKLKGYTN WAIGLSVADL TESIIKNMSR IHPVSTMVKD MYGIGEEVFL SLPCVLNSNG
Tetraodon_LDH-B1    VWSGTNVAGV SLQTLNPDIG TDRDHENWRE THKMVVDSAY EVIRLKGYTN WAIGLSVGDL IESLMKNMSR IHPVSTMVQG MYGIGEEVYL SLPCVLNGRG
Takifugu_LDH-B1     VWSGANVAGV NLQKLNPDIG TEEDKEEWKS THKAVVDSAY EVIKLKGYTN WAIGLSVADL TESIVKNMSR VHPVSTMGQD MYGIDEEVFL SLPCVLNSSG
Killifish_LDH-B     VWSGANVAGV SLQKLNPEIG TDGDKEQWKA THKAVVDSAY EVIKLKGYTN WAIGFSVADL TESIVKNLSR VHPVSTMVKD MFGIGEEVFL SLPCVLNGSG
Stickleback_LDH-B1  VWSGANVAGV NFQSLNPDIG TDGDKEQWAA THKAVVDSAY EVIKLKGYTN WAIGLSVADL IESIVKNMSR VHPVSTMVKD MYGIGEEVFL SLPCVLNSTG
Medaka_LDH-B1       VWSGANVAGV NLQKLNPDIG TDADKEQWKA THKAVVDSAY EVIKLKGYTN WAIGLSVADL TESILKNMSR VHPVSTMVKD MYGIGEEVFL SLPCVLNSSG
Dogfish_LDHB        VWSGVNVAGV GLQQLNPDIG TAQDKENWKD VHKMVVESAY EVIKLKGYTN WAIGLSVAEL TESMVKNLKR VHPVSTMVKG MYGIENEVFL SLPCVLSADG
Clustal Consensus   ****..**** .:* ***::*    *:*.*   .**  *:*** ***:****** ****:**.:*  **::**:.* :****** :. *:** :**:* ******.. *


                    ....|....| ....|....| ....|....| ....
                            310       320       330
RS_LDHB_79M16       VGSVVNMTLN DEEVAQLKKS ADTLWGIQKD LKDL
AS_LDHB1_225J21     VGSVINMTLT DAEVGQLKKS ADTLWGIQKD LKDV
RT_LDHB1_48008      VGSVINMTLT DAEVGQLKKS ADTLWGIQKD LKDV
AS_LDHB2_276I15     VGSVINMTLT DAEVGQLKKS AETLWGIQKD LKDI
RT_LDHB2_49194      VGSVINMTLT DAEVGQLKKS ADTLWGIQKD LKDI
Tetraodon_LDH-B1    VASVINMTLT EDEVSRLQDS ARTLWDIQKD LQNV
Takifugu_LDH-B1     VSSVVNMSLT EEEVSQLRNS ADTLWTIQKD LKDI
Killifish_LDH-B     VGSVVNMTLT DAEVAQLKKS ADTLWGIQKD LKDL
Stickleback_LDH-B1  VSSVVNMTLT NAEVSQLRKS GDTLWGIQKD LKDI
Medaka_LDH-B1       VGSVVNMTLT EGEVAQLRKS ADTLWGIQKD LKDL
Dogfish_LDHB        LIDVINQMLK DDEVAQLRKS AETLWNIQKE LKDL
Clustal Consensus   : .*:*  *. : **.:*:.* . *** ***: *:::
```

Figure 3.28 Amino acid sequence alignment for LDH-C among samonids, rainbow smelt and other fish. AS: Atlantic salmon, RT: rainbow trout; RS: rainbow smelt.

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            10        20        30        40        50        60        70        80        90        100
RS_LDHC_113P10      MASILQKLIT PLFSGPAEPP RNKVTVVGVG QVGMACAVSI LLRELADELA LVDVMEDKLK GEMMDLQHGS LFLKTPKIVA DKDYSVTANS RIVVVTAGVR
AS_LDHC_116D13      MASILQKLMT PLVSGNPEPP RNKVTVVGVG QVGMACAFSI LLRELADELA LVDVMEDKLK GEIMDLQHGS LFLKTPKIVA GKDYSVTANS RIVVVTAGVR
RT_LDHC_44564       MASILQKLLT PLVSGNPEPP RNKVTVVGVG QVGMACAVSI LLRELADELA LVDVMEDKLK GEIMDLQHGS LFLKTPKIVA GKDYSVTANS RIVVVTAGVR
Tetraodon_LDH-B2    MASILQKLFH PLLSGPPEPP RNKVTVVGVG QVG-ACAVTI LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTPKIVA DKDYSVTANS RVVVVTAGVR
Takifugu_LDH-B2     MASILHKLFH PLFSGPPEPP RNKVTVVGVG QVGMACAVTI LLRDLADELA LVDVMEDKLK GEMMDLQHGS LFLKTPTIVA DKGYSVTANS RIVVVTAGVR
Killifish_LDH-C     MASVLHKLIT PLACSSPEPP RNKVTVVGVG QVGMACAVTI LLRELADELA LVDVVEDKVK GEMMDLQHGS LFLKTPKIVA DKDYSVTSNS RIVVVTAGVR
Stickleback_LDH-B2  MASILQKLLT PLFSGPPEPP RNKVTVVGVG QVGMACAISI LLRELADELA LVDVMEDKLK GEMMDLQHGS LFLKTHKIVA NKDYSVSANS RIVVVTAGVR
Medaka_LDH-B2       MASILQKLIT PLFSGPPEPP RNKVTVVGVG QVGMACAVSI LLKELADELA LVDVVEDKLK GEMMDLQHGS LFLKTPKIVA SKDYSVTANS RIVVVTAGVR
Clustal Consensus   ***:*:**:  ** .. .*** ********** *** ***.:* **::****** ****:***:* **:******* ***** .*** .*.***::** *:********

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            110       120       130       140       150       160       170       180       190       200
RS_LDHC_113P10      QQEGESRLNL VQRNVNIFKH IIPQIVRYSP NCIIIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMADKLG IHSSSFNGWI LGEHGDTSVP
AS_LDHC_116D13      QQEGESRLNL VQRNVNIFKH IIPQIIKHSP NCIIIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMADKLG IHTTSFNGWI LGEHGDTSVP
RT_LDHC_44564       QQEGESRLNL VQRNVNIFKR IIPQIIKHSP HCIIIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRYLMADKLG IHTSSFHGWI LGEHGDTSVP
Tetraodon_LDH-B2    QQEGESRLNL VQRNVNIFKH IVPQIVRYSP ECVIIVVSNP VDVLTYVTWK LSGLPKRRVI GSGTNLDSAR FRFLVADKLG LHASSFNGWI LGEHGDTSVP
Takifugu_LDH-B2     QQDGESRLNL VQRNVNIFKH IVPQIVRYSP ECIIVVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLIAEKLG IHASSFNGWI LGEHGDTSVP
Killifish_LDH-C     QQEGERRLNL DQRNVNIFKH IIPLIVRHSP DCIIIVVSNP VDVLTYVTWK LSGLPMHRVI GSGTNLDSAR FRFLMADKLG IHSSSFNGWI LGEHGDTSVP
Stickleback_LDH-B2  QQEGESRLNL VQRNVNIFKH IVPQIVRYSP DCTIIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMADKLG IHPSSFNGWI LGEHGDTSVP
Medaka_LDH-B2       QQEGESRLNL VQRNVNIFKH IVPQIVRYSP DCTIIVVSNP VDVLTYVTWK LSGLPKHRVI GSGTNLDSAR FRFLMADKLG IHASSFNGWI LGEHGDTSVP
Clustal Consensus   **:** ****  ********: *:* *:::** .* *:***** ********** ***** :*** ********** **:*:*:*** :*.:**:*** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            210       220       230       240       250       260       270       280       290       300
RS_LDHC_113P10      VWSGANVAGV NLQTLNPDIG TDHDSENWKE THKKVVDSAY EVIKLKGYTN WAIGLSVADL AESLIKNMNR IHPVSTMVKG MFGISDEVYL SLPCILNNGG
AS_LDHC_116D13      VWSGANVAGV NLQTLNPDIG TDADQENWKE THKQVVDSAY EVITLKGYTN WAIGLSVADL TESLLRNMNR IHPVSTMIKG MYGVKDEVYL SLPCVLNAGG
RT_LDHC_44564       VWSGANVAGV NLQTLNPDIG TDADRENWRE MHKQVVDSAY EVTTLKGYTN WAIGLSVADL TESLIRNMNR IHPVSTMVKG MYGVKDEVYL SLPCVLNAGG
Tetraodon_LDH-B2    VWSGTNVAGV SLQTLNPDIG TDRDHENWRE THKMVVDSAY EVIRLKGYTN WAIGLSVGDL IESLMKNMSR IHPVSTMVQG MYGIGEEVYL SLPCVLNGRG
Takifugu_LDH-B2     VWSGTNVAGV NLQTLNPNIG TDCDDENWKE THKMVVDSAY EVIRLKGYTN WAIGLSVGDL VESLMKNMNR IHPVSTMVQG MYGINEEVYL SLPCVLYGGG
Killifish_LDH-C     VWSGTNVAGV NLQTLNPNIG TDFDEENWKE THKMVVDSAY EVIKLKGYTN WAIGLSVADL TESLMRNMNR IHPVSTMAKG MYGIGDEVYL SLPCVLNSGG
Stickleback_LDH-B2  VWSGTNVAGV NLQTLNPDIG TDCDEENWTE THKMVVDSAY EVIKLKGYTN WAIGLSVADL TESLIRNMNR IHPVLVCVQG MYGISNEVYL SLPCVLNAGG
Medaka_LDH-B2       VWSGTNVAGV NLQTLNPEIG TDCDKENWQE THKMVVNSAY EVIRLKGYTN WAIGLSVADL IESLMKNLNR IHPVSTMVKG MYGISEEVYL SLPCVLNGGG
Clustal Consensus   ****:***** .******:** ** * *** *  ** **:*** **  ****** *******.**  ***::*:.* **** .  :* *:*: :**** ****:*   *

                    ....|....| ....|....| ....|....| ....
                            310       320       330
RS_LDHC_113P10      VASVVNMTLT DKEIAQLRSS ANTLWGIQKD LKDV
AS_LDHC_116D13      VASVINMTLT DNEIGQLKQS ADTLWGIQKD LTDV
RT_LDHC_44564       VASVINMTLT DNEIGQLKQS ADTLWGIQKD LTDV
Tetraodon_LDH-B2    VASVINMTLT EDEVSRLQDS ARTLWDIQKD LQNV
Takifugu_LDH-B2     VASVINMTLT DDEVTQLQDS ARTLWDIQKD LRDI
Killifish_LDH-C     VGSVVNMTLT DEEVAQLQGS ASTLWDIQKD LRDI
Stickleback_LDH-B2  VASVVNMSLT EDEVARLQAS AGTLWDIQKD LQDV
Medaka_LDH-B2       VASVINMTLT DDEVAQLQAS ANTLWDIQKD LQDI
Clustal Consensus   *.**:**:** :.*: :*: * * ***.**** * ::
```

The alignments of amino acid sequences from Atlantic salmon, rainbow trout, rainbow smelt and other teleosts for LDH-A, LDH-B and LDH-C were separately used to predict the positions of amino acid substitutions. The predicted biological relationships of the LDH duplicates are based on a salmonid genome duplication (♦, in Figures 3.29 and 3.30) and a subsequent speciation (●, in Figures 3.29, 3.30 and 3.31) of Atlantic salmon and rainbow trout. It is assumed that rainbow smelt is more closely related to the salmonids than any of the other fish that form the out-group.

Figure 3.29 The position of amino acid substitutions in LDH-A duplication of salmonids and rainbow smelt. AS: Atlantic salmon, RT: rainbow trout, RS: rainbow smelt. The green diamond denotes the location of genome duplication. The blue solid circle indicates the location of speciation between Atlantic salmon and rainbow trout. Each branch labels as alphabets. The number located below each branch of the tree indicates the number of substitution.

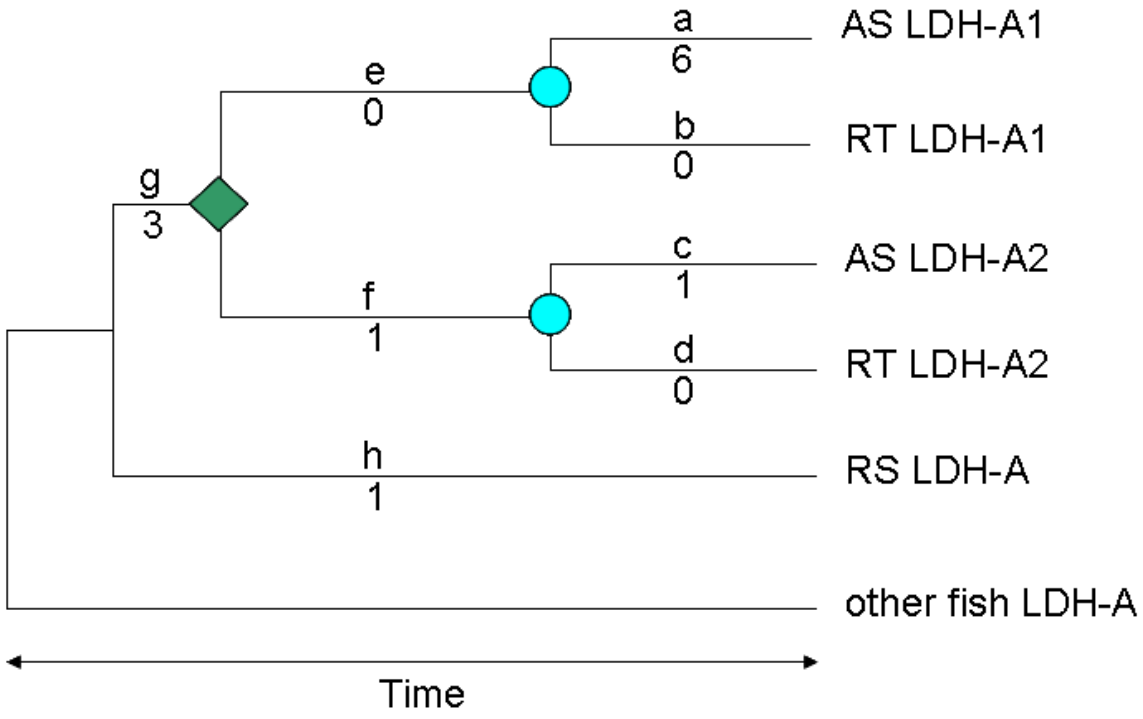


Table 3.11 Amino acid changes in different lineages in LDH-A evolution.

Branch	Position	Amino Acid Change	Type of Change
a	14	A to V	conservative
a	218	H to D	basic to acidic
a	231	L to V	conservative
a	279	Q to K	conservative
a	286	D to E	conservative
a	318	I to S	hydrophobic to hydrophilic
c	279	Q to K	conservative
f	57	D to E	conservative
g	14	K to A	basic to small side chain
g	218	Q to H	conservative
g	231	V to L	conservative
h	279	Q to N	conservative

Figure 3.30 The position of amino acid substitutions in LDH-B duplication of salmonids and rainbow smelt. AS: Atlantic salmon, RT: rainbow trout, RS: rainbow smelt. The green diamond denotes the location of genome duplication. The blue solid circle indicates the location of speciation between Atlantic salmon and rainbow trout. Each branch labels as alphabets. The number located below each branch of the tree indicates the number of substitution.

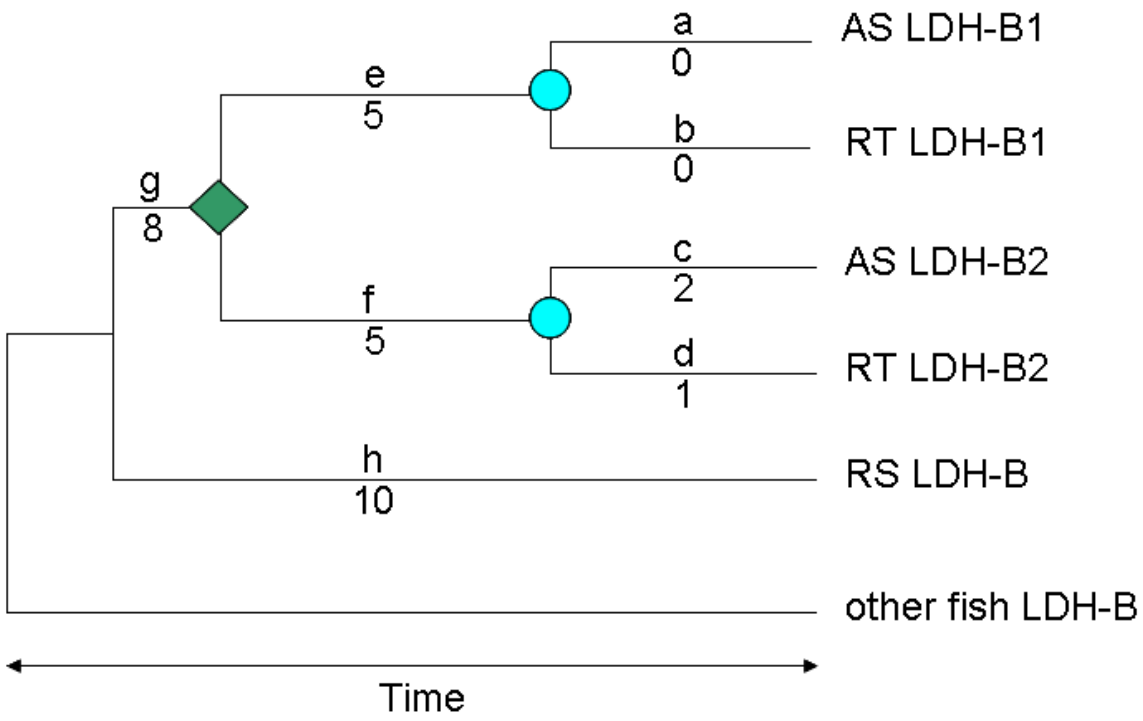


Table 3.12 Amino acid changes in different lineages in LDH-B evolution.

Branch	Position	Amino Acid Change	Type of Change
c	63	M to L	conservative
c	322	D to E	conservative
d	235	V to E	hydrophobic to acidic
e	12	L to M	conservative
e	184	S to T	conservative
e	206	N to S	conservative
e	234	A to E	small side chain to acidic
e	334	L to V	conservative
f	9	I to L	conservative
f	12	L to V	conservative
f	40	I to V	conservative
f	96	T to S	conservative
f	334	L to I	conservative
g	92	L to I	conservative
g	205	A to V	conservative
g	218	D to E	conservative
g	221	T to L	hydrophilic to hydrophobic
g	264	V to I	conservative
g	300	S to N	conservative
g	305	V to I	conservative
g	315	A to G	conservative
h	120	X to A	conservative
h	135	I to V	conservative
h	157	H to N	conservative
h	183	A to S	conservative
h	190	V to I	conservative
h	197	T to S	conservative
h	230	A to Q	conservative
h	234	A to E	small side chain to acidic
h	310	T to N	conservative
h	312	A to E	small side chain to acidic

Figure 3.31 The position of amino acid substitutions in LDH-C duplication of salmonids and rainbow smelt. AS: Atlantic salmon, RT: rainbow trout, RS: rainbow smelt. The blue solid circle indicates the location of speciation between Atlantic salmon and rainbow trout. The number located on each branch of the tree indicates the number of substitution. Each branch labels as alphabets. The number located below each branch of the tree indicates the number of substitution.

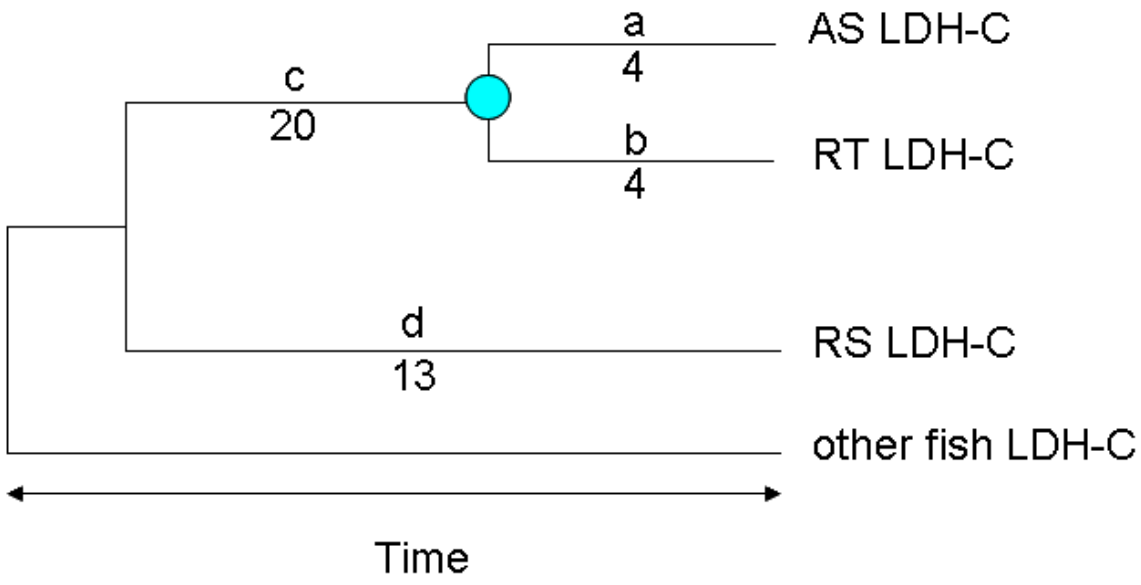


Table 3.13 Amino acid changes in different lineages in LDH-C evolution.

Branch	Position	Amino Acid Change	Type of Change
a	38	V to F	conservative
a	184	S to T	conservative
a	265	I to L	conservative
a	279	V to I	conservative
b	120	H to R	conservative
b	229	K to R	conservative
b	231	T to M	conservative
b	243	I to T	conservative
c	13	F to V	conservative
c	16	P to N	conservative
c	63	M to I	conservative
c	81	D to G	conservative
c	126	V to I	conservative
c	127	R to K	conservative
c	128	Y to H	aromatic to basic
c	183	S to T	conservative
c	223	C to A	conservative
c	234	M to Q	conservative
c	244	K to T	basic to hydrophilic
c	285	I to V	conservative
c	286	S to K	hydrophilic to basic
c	306	V to I	conservative
c	313	D to N	conservative
c	316	A to G	conservative
c	319	Q to K	conservative
c	320	X to Q	conservative
c	323	X to D	conservative
c	333	Q to T	conservative
d	17	P to A	conservative
d	223	C to H	conservative
d	234	M to K	conservative
d	261	T to A	conservative
d	266	R to K	conservative
d	283	Y to F	conservative
d	296	V to I	conservative
d	299	A to N	conservative
d	313	D to K	acidic to basic
d	319	Q to R	conservative
d	320	X to S	conservative
d	323	X to N	conservative
d	333	Q to K	conservative

First I carried out a relative rate test to determine if the number of amino acid substitutions is the same along the rainbow smelt and salmonid lineages under the same time divergence. In the case of LDH-A, there appears to be 1 substitution in the lineage from the common ancestor of the rainbow smelt and salmonids to the present day rainbow smelt, whereas the number of amino acid substitutions along the salmonid LDH-As ranges from 4 – 9. The corresponding values for LDH-B are 10 substitutions along the rainbow smelt lineage and anywhere from 13 – 15 in the salmonids. For LDH-C, there are 13 substitutions leading to rainbow smelt and 24 leading to the salmonids. These results reveal that for all three LDH types, the rate of change is greater along the salmonid lineages than the rainbow smelt lineage. Moreover, the results show that the rate of change differs with respect to LDH type: LDH-C accepting mutations faster than LDH-B, which in turn is evolving faster than LDH-A.

Next, I examined where the changes have occurred in each type of LDH. Although LDH-A appears to change slowly, there is one branch where many changes have occurred. This is seen in the Atlantic salmon LDH-A1 after the speciation that gave rise to the rainbow trout LDH-A1 (Figure 3.26). Starch gel electrophoresis has shown that LDH-A1 and LDH-A2 are equally expressed in salmonid muscle. The faster amino acid substitution rate in Atlantic salmon LDH-A1 compared to Atlantic salmon LDH-A2 or rainbow trout LDH-A1 or LDH-A2 may imply that this gene product is no longer under any selection pressure and is heading for extinction (nonfunctionalization). On the other hand, it may have acquired one or more amino acid changes that provide a novel function

(neofunctionalization) or at least different kinetic parameters that could be considered a partial neofunctionalization. Until direct comparisons of binding constants for substrates and cofactors and catalytic constants are determined, it is not possible to say what this increased rate means for Atlantic salmon LDH-A1.

The locations in the structure of LDH where amino acid changes have occurred are reported in Tables 3.11 – 3.13. The amino acid changes were classified as conservative or radical based on the properties of their side chains. Radical changes include a change in charge, a change in hydrophobicity and a change in size of side chain. Two of the six changes in the Atlantic salmon LDH-A1 lineage can be considered radical: 218 H to D and 318 I to S. However, neither of the residues at these positions has been implicated in the mechanism of action of LDH (Eventoff et al. 1977).

Compared to LDH-As, more amino acid substitutions have occurred in the salmonid LDH-Bs. However, the pattern of amino acid changes is quite different between the LDH-As and LDH-Bs (Figure 3.30). There are very few changes in the branches following the speciation of Atlantic salmon and rainbow trout in LDH-B1 or LDH-B2, but there appears to have been a burst of amino acid substitutions along both the LDH-B1 and LDH-B2 lineages after the salmonid genome duplication. This suggests that there has been a period of positive selection followed by strong negative or purifying selection. Starch gel electrophoresis indicates that there has been subfunctionalization of the salmonid LDH-B1 and LDH-B2 genes with LDH-B1 being expressed mainly in

liver rather than in heart while LDH-B2 is expressed primarily in liver. It is tempting to speculate that the pattern of change seen in the salmonid LDH-Bs is the result of partial neofunctionalization (positive selection) of each enzyme after subfunctionalization as they become better adapted to one aerobic tissue rather than having to be able to cope with the metabolic conditions in both liver and heart. This might be detected as subtle kinetic differentiation between LDH-B1 and LDH-B2. Table 3.12 shows the type of amino acid substitutions that have occurred in each branch. Only one of the five changes leading to LDH-B1 could be considered radical: 234 A to E, while all of the five changes in the branch leading to LDH-B2 are conservative. None of these amino acid residues have been implicated in the mechanism of action of LDH (Eventoff et al. 1977).

The fastest rate of amino acid change was observed in the LDH-Cs (Figure 3.31). LDH-C is predominately expressed in the eye of teleosts (see also Section 3.6). Wistow et al. (1987) observed that a crystallin protein, found in the lens, was the product of the same gene as LDH-B_4 in birds and crocodiles. Perhaps after the duplication that produced LDH-B and LDH-C in teleosts, a change in expression allowed the LDH-C to become a crystallin as well. This would be an example of neofunctionalization through a regulatory mutation rather than subfubnctionalization. Although LDH activity has been retained, the main purpose of the LDH-C protein in the eye may actually be as a lens protein. This would explain the higher rate of evolution of teleost LDH-C compared to LDH-B and LDH-A.

3.6 LDH tissue expression

In the subfunctionalization model, complementary mutations in different regulatory regions of duplicated genes can preserve both copies of the duplicated genes and lead to complementary expressions (Force et al. 1999). The LDH isozymes provide a good example of gene duplication followed by subfunctionalization. In general, the LDH-A isozymes are expressed in tissues with anaerobic metabolism (e.g., skeletal muscle) whereas the LDH-B isozymes are expressed in aerobic tissues such as liver and heart. In the evolution of salmonids, the LDH-B genes have undergone subfunctionalization. LDH-B1 is mostly expressed in liver and LDH-B2 is predominately expressed in heart. The duplicated LDH-A1 and LDH-A2 appear to be equally expressed in skeletal muscle. Rainbow smelt should provide an important reference to study the subfunctionalization model in salmonids. Therefore, I determined the tissue expression patterns of the three rainbow smelt LDHs.

3.6.1 LDH tissue expression

Total RNA was isolated from nine rainbow smelt tissues (brain, eye, gill, muscle, heart, liver, head kidney, spleen and gonad) and converted into cDNA using random hexamers and reverse transcriptase. The cDNA was amplified with primers specific for β–actin and each of the three LDH genes in rainbow smelt (Table 3.14). These primers were designed from exonic sequences to cross introns in genomic DNA. In eukaryotic transcription, the non-coding regions (introns) of the primary transcript are spliced and excised and the coding regions (exons) are joined to form the mature mRNA. The positive control used rainbow

smelt genomic DNA as template to amplify a PCR product. The amplified PCR products with cDNA as template were smaller than those from genomic DNA. These results confirm that introns were spliced from mature mRNA and that the product was not from contaminating genomic DNA. I carried out the analysis using 35, 30 and 25 cycles of PCR. This semi-quantitative analysis allowed me to determine if a particular gene was expressed in one or more of the tissues but it did not give the information required to be able to compare transcript levels accurately. The purpose was to determine in which tissues the LDHs are expressed.

From the gel electrophoresis image (Figure 3.32a), it can be seen that the reference β–actin primers amplified a 400 bp product with an equal intensity from each tissue. In rainbow smelt, LDH-A specific PCR products were produced from all tissues at apparently the same level except for liver (Figure 3.32b). The weak expression in liver and equal expressions in other tissues from LDH-A specific amplification is consistent with the result of LDH starch gel electrophoresis from brown trout (Figure 1.9) (Markert et al. 1975). In Atlantic salmon, LDH-A1 and LDH-A2 are highly expressed in eye and muscle (Lubieniecki et al. in preparation). The rainbow smelt LDH-C gene is predominately expressed in the eye, and weakly expressed in brain, gill, head kidney and gonad (Figure 3.32d). This result is consistent with what has been observed at the protein level in other teleosts (Markert et al. 1975) and at the transcriptional level in Atlantic salmon (Lubieniecki et al. in preparation).

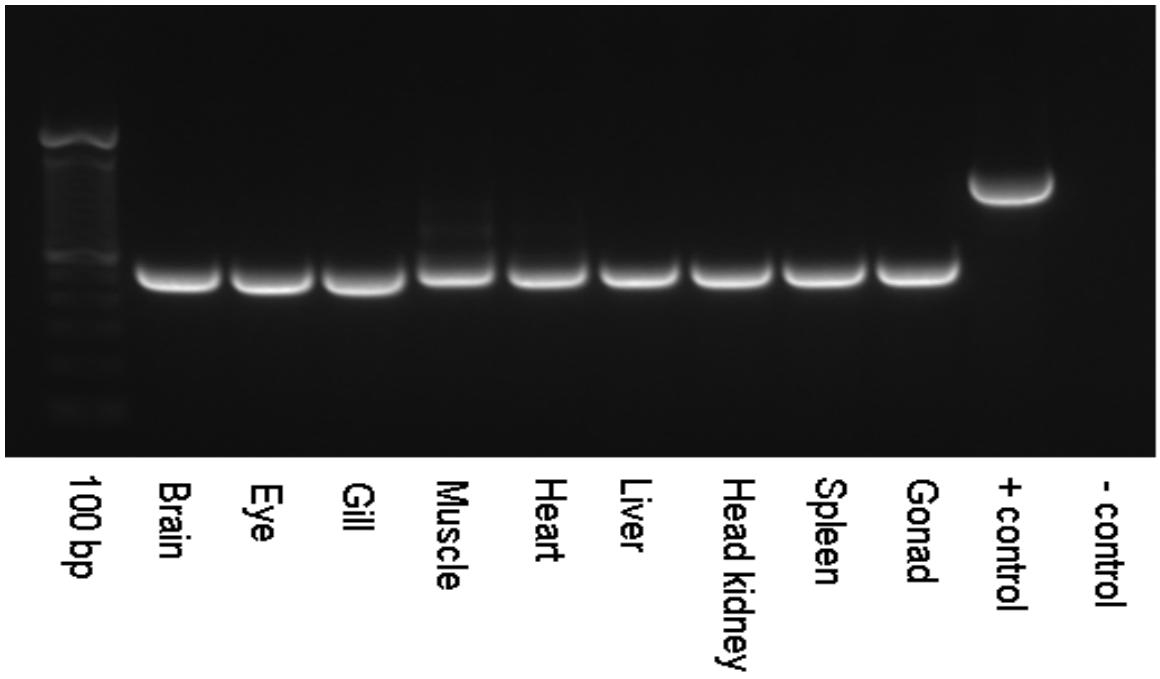
The tissue expression of LDH-B in rainbow smelt shows that it is mostly expressed in heart and brain, moderately expressed in eye, gill, head kidney, spleen and gonad, and weakly expressed in muscle (Figure 3.32c). In Atlantic salmon, LDH-B1 has been shown to have a high expression level in liver, and LDH-B2 expression has been observed in heart (Lubieniecki et al. in preparation). Therefore, I expected to see evidence of expression of the LDH-B in rainbow smelt liver. Indeed, I anticipated that LDH-B would be expressed equally in liver and heart of rainbow smelt. However, my result did not show any LDH-B expression in liver. The quality of the liver RNA and cDNA does not appear to be a problem as the β–actin primers did not show an apparent decrease in PCR product. Further study of LDH-B enzyme activity in different rainbow smelt tissues may give more information to explain the absence of LDH-B expression in liver. Moreover, quantitative PCR (qPCR) would allow me to quantify the amplification at each cycle of the expression study.

Table 3.14 Primer list for rainbow smelt tissue expression

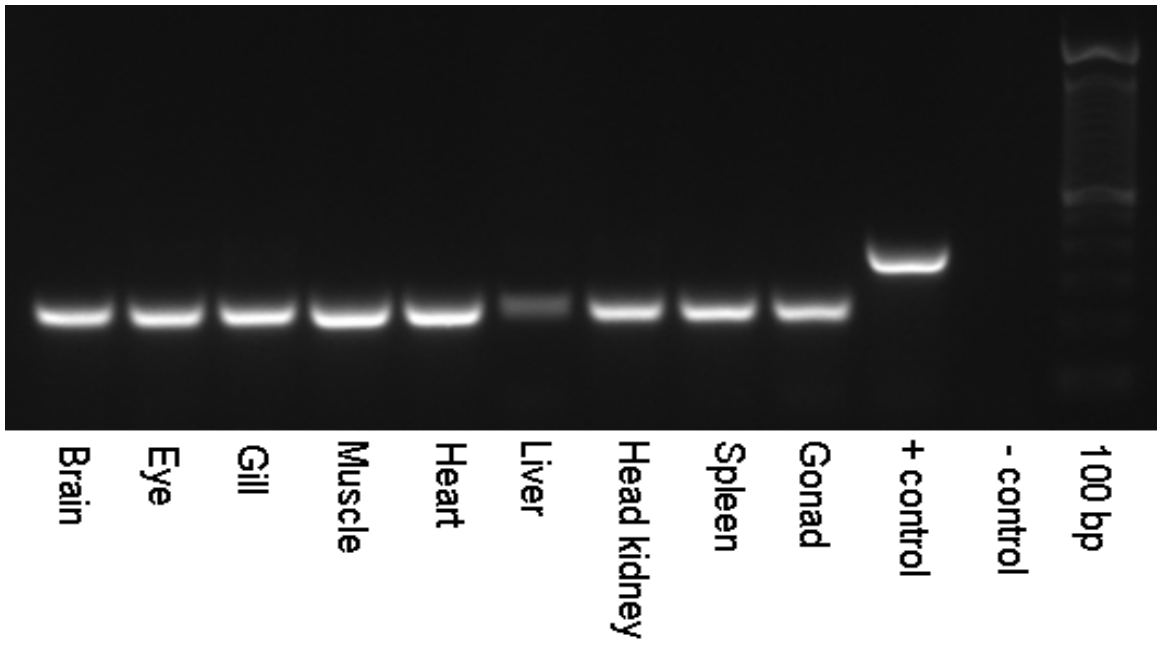
	Primer sequences	Tm (°C)
β–actin_F	5’-CGGATCCGGTATGTGTAAGG-3'	65
β–actin_R	5’-GCTCCGTCAGGATCTTCATC-3'	65
LDH-A_F	5'-GTGTGATGAGCTGGCCCTGGTTGACGTGATGGTGGACAAG-3'	65
LDH-A_R	5'-ACTTGACGATGTTGGGGATG-3'	65
LDH-B_F	5'-TCAGCGTAGCTGGAGTCAACCTGCAGAAGCTGAACCCAGAG-3'	65
LDH-B_R	5'-TGAGATCAGCCACACTCAGG-3'	65
LDH-C_F	5'-ATGGCCTCAATTCTGCAGAAGCTC-3'	65
LDH-C_R	5'-TTACACGTCTTTCAGGTCCTTCTGGATA-3'	65

Figure 3.32 The tissue expression of LDH-A in rainbow smelt. Nine rainbow smelt tissues were brain, eye, gill, muscle, heart, liver, head kidney, spleen and gonad. + control: genomic DNA; - control: dH_2O. (a) β-actin amplification (b) LDH-A amplification (c) LDH-B amplification (d) LDH-C amplification.

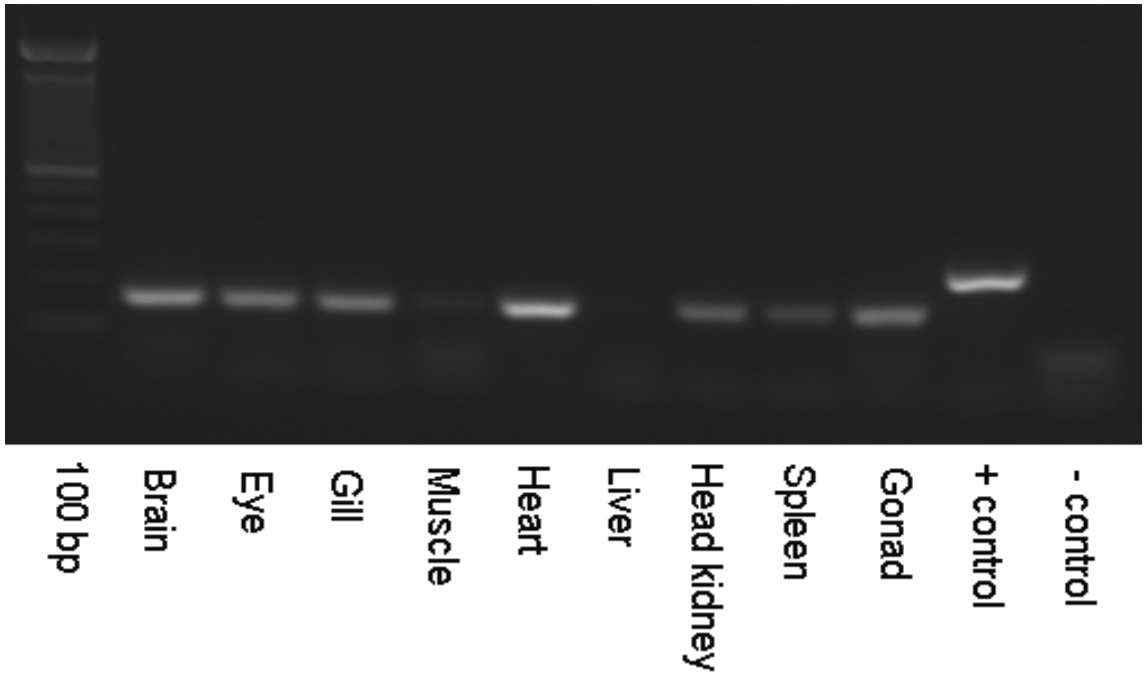
(a)



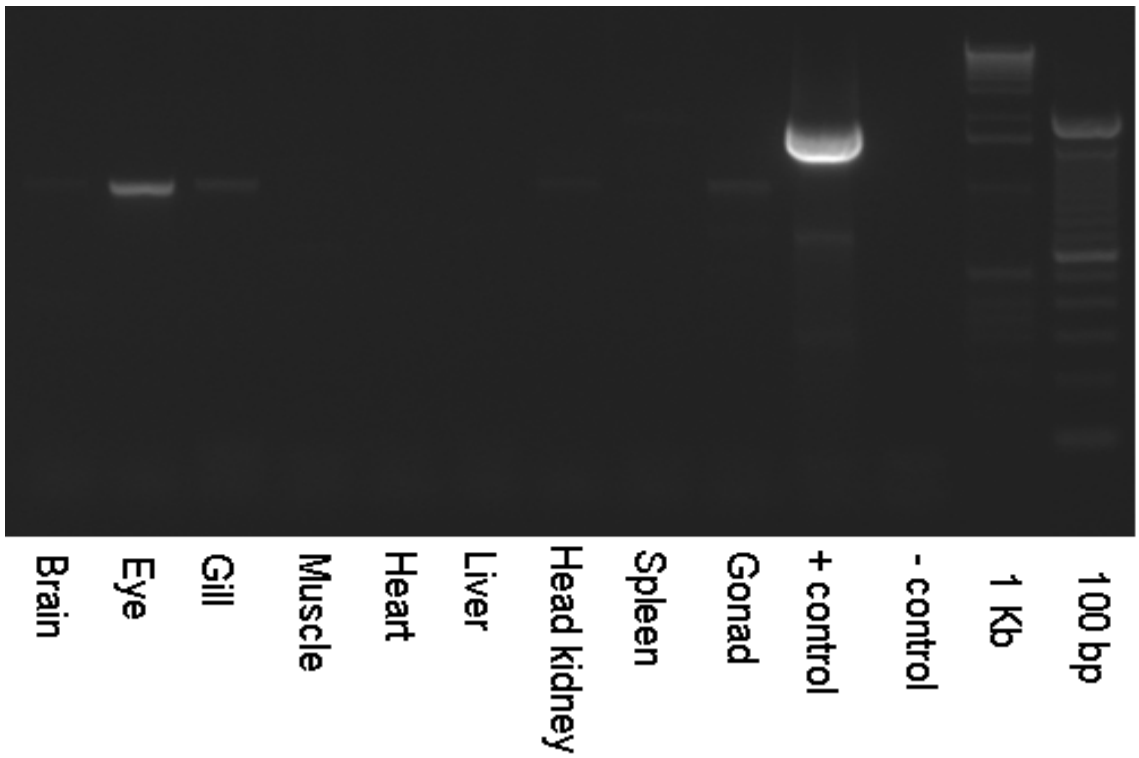
(b)



(c)


100 bp
Brain
Eye
Gill
Muscle
Heart
Liver
Head kidney
Spleen
Gonad
+ control
- control

(d)


Brain
Eye
Gill
Muscle
Heart
Liver
Head kidney
Spleen
Gonad
+ control
- control
1 Kb
100 bp

3.6.2 Search for promoters of the LDH-B genes

Another method of examining the subfunctionalization of LDH-B1 and LDH-B2 in Atlantic salmon using rainbow smelt as a reference is promoter analysis. This assumes that despite my results, LDH-B is in fact expressed in liver and heart and that it represents the ancestral state from which the Atlantic salmon LDH-B genes have diverged through subfunctionalization and differential loss or modification of liver and heart promoter regions. Keith Boroevich helped me to do this analysis. A 1000 bp upstream region of six LDH-Bs (two LDH-Bs from Atlantic salmon, one LDH-B from rainbow smelt, stickleback, medaka and tetraodon) were retrieved and submitted into JASPAR: The high-quality transcription factor binding profile database (http://jaspar.genereg.net/) (Wasserman and Sandelin 2004). There were 81 putative transcription factor binding sites (TFBS) identified in all six sequences, and 11 identified only in five sequences. TAL1-TCF3 was identified three to four times in all six sequences except Atlantic salmon LDH-B1, and HLF was found one to two times in all sequences except Atlantic salmon LDH-B2. These results suggested that differential loss of TAL1-TCF3 in salmonid LDH-B1 and HLF in salmonid LDH-B2 could explain the subfunctionalization that is observed.

In order to test if these two TFBS are specific to LDH genes, the 1000 bp upstream region of all annotated genes (20256) in the medaka genome were scanned for TAL1-TCF3 and HLF. The result showed that 86% of the total genes (17480) in medaka contain HLF sites, and 79% (16093) contain TAL1-TCF3 sites. Due to the absence of a fish regulatory region and TFBS database and a lack of

a reference genome sequence for salmonids, I am unable to make any conclusions concerning this intriguing result at this time.

3.7 Comparison of the rainbow smelt and Atlantic salmon LDH genomic regions

The rediploidization process that returns a duplicated tetraploid genome to a stable diploid state probably requires considerable genome rearrangements (Wolfe 2001). These genome rearrangements are thought to be particularly important in the case of autotetraploidy as the two pairs of identical homeologous chromosomes change such that the homeologues no longer interact with one another. Therefore, I searched for conservation of synteny and changes such as inversions and deletions in the genomic regions containing LDH genes in rainbow smelt and the corresponding regions of Atlantic salmon.

3.7.1 Search for conservation of synteny in regions of the genome surrounding the LDH genes in rainbow smelt and Atlantic salmon

The syntenic comparisons of each LDH BAC between rainbow smelt and Atlantic salmon were based on the rainbow smelt LDH BAC annotations shown in Figures 3.33, 3.34 and 3.35 and what has been done for the corresponding LDH containing BACs from Atlantic salmon (Lubieiecki et al. in preparation). The syntenic blocks showed the order of neighboring genes around each LDH gene.

The LDH-A group contained syntenic blocks from Atlantic salmon, LDH-A1 and LDH-A2, and rainbow smelt LDH-A; the LDH-B group contained syntenic blocks from Atlantic salmon, LDH-B1 and LDH-B2, and rainbow smelt LDH-B; and the LDH-C group contained syntenic blocks from Atlantic salmon LDH-C and

rainbow smelt LDH-C. In each group, the same genes were highlighted as same colour. The results are shown in Figures 3.33 – 3.35.

In the LDH-A group, the BACs containing the LDH-A1 (S0052D13) and LDH-A2 (S0069I14) have apparently no genes in common on either side of the LDH-A gene (Figure 3.33). One gene, TSG101, is located adjacent to both the rainbow smelt LDH-A gene and the Atlantic salmon LDH-A2, but the orientation of transcription of TSG101 relative to that of LDH-A is different in each case. These results suggest that considerable rearrangement has occurred in this region of the Atlantic salmon genome. However, it is not possible to predict what the ancestral state was.

The LDH-B group shows considerable conservation of synteny within the Atlantic salmon genome and between the Atlantic salmon and rainbow smelt genomes (Figure 3.34). In the overlap of the BACs containing the Atlantic salmon LDH-B1 (S0225J21) and LDH-B2 (S0276I15) genes, there are nine genes whose orders of transcription have been conserved. Seven of these common genes were also annotated in the rainbow smelt LDH-B containing BAC, and they are in the same order and transcriptional orientation. A closer examination of the rainbow smelt BAC sequence is warranted to search for C120RF39 between GOT1B and GYS2 and MPCP between LDHB and TMPO. In addition, the rainbow smelt BAC contains STRAP above TMPO as does the Atlantic salmon BAC that contains LDH-B2. These comparisons show that unlike the LDH-A genes in Atlantic salmon, there has been little or no genomic reorganization around the LDH-B genes. Perhaps there has been a selective pressure to retain

this configuration after the subfunctionalization of the LDH-B genes in this species.

A comparison of the LDH-C genomic regions of rainbow smelt and Atlantic salmon shows two genes in common located upstream of LDH-C, MPCP and TMPO, with SLC25A3 in between LDH-C and MPCP in Atlantic salmon (Figure 3.35). As above, it is worth examining this region in more detail in the rainbow smelt BAC to see if the annotation pipeline has missed SLC25A3. GOT1B is downstream of LDH-C in both species, but thereafter there are no genes in common (SLC35B4 and CHCHD3 in Atlantic salmon, and NET1, ASB9 and RAP140 in rainbow smelt). There are several genes at the LDH-B loci that also occur at the LDH-C loci. The gene order and transcriptional orientation of TMPO – (SLC25A3) – MPCP – LDH-B/C is conserved at these loci. GOT1B is found downstream of the LDH-B and LDH-C loci, but the transcriptional orientation is reversed, suggesting that a local inversion has occurred. The finding of conservation of synteny between the LDH-B and LDH-C loci supports the statement that LDH-C is derived by the gene duplication from LDH-B in teleosts (Whitt et al. 1975).

Figure 3.33 Syntenic comparison of LDH-A between rainbow smelt and Atlantic salmon.

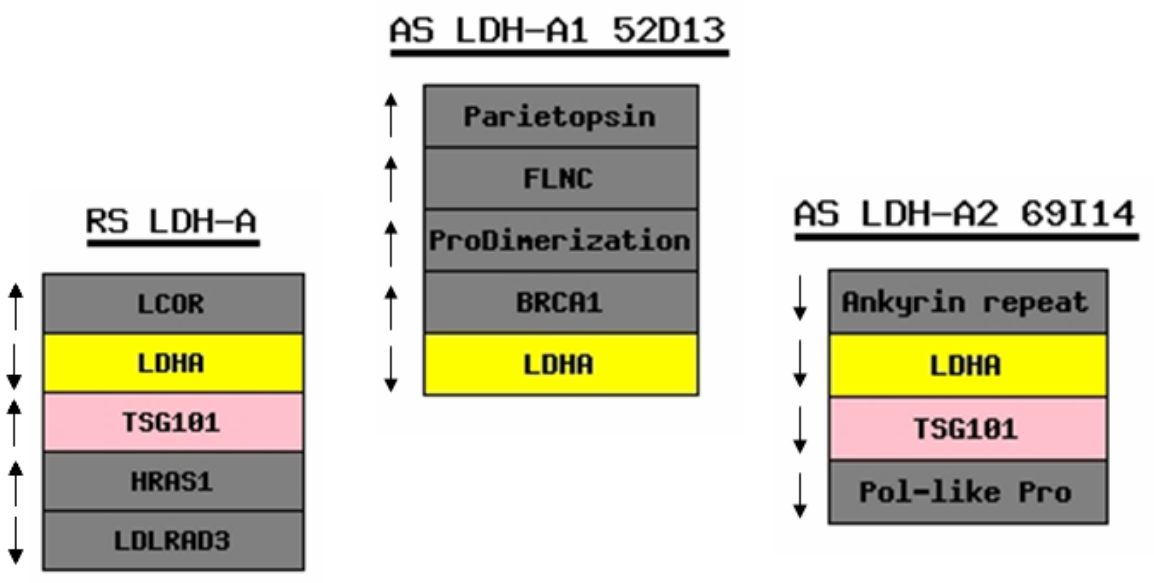


Figure 3.34 Syntenic comparison of LDH-B between rainbow smelt and Atlantic salmon.

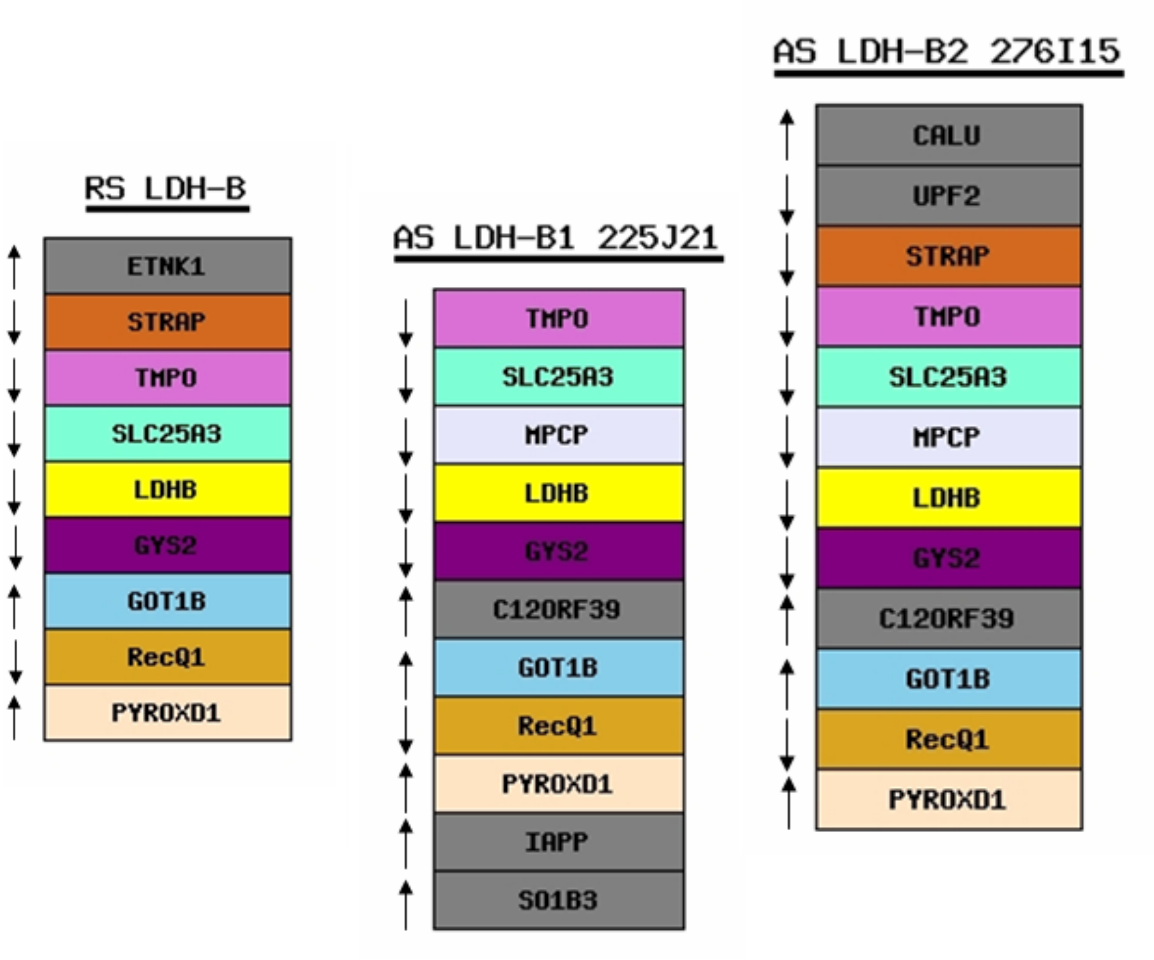
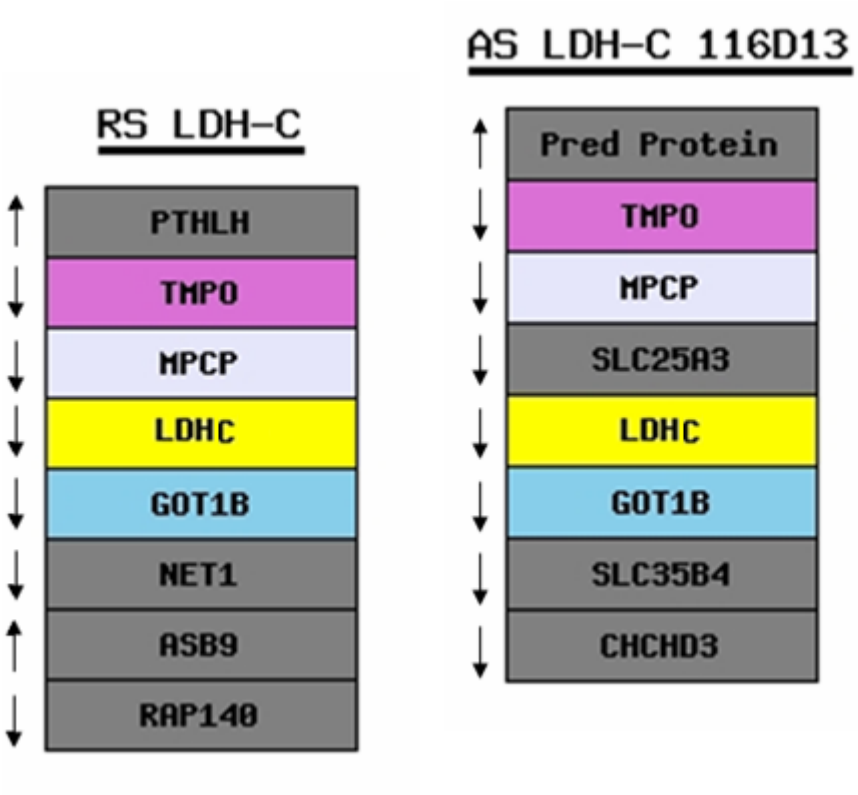


Figure 3.35 Syntenic comparison of LDH-C between rainbow smelt and Atlantic salmon.



CHAPTER 4: DISCUSSION

The objective of this project involved the characterization of LDH genes from rainbow smelt so that they could act as a diploid out-group for comparisons with the LDH gene family in salmonids. This then allowed me to better understand how duplicated genes evolve by carrying out a phylogenetic analysis and examining factors such as tissue expression and selective pressures. Moreover, the characterization of rainbow smelt genes will aid in the future annotation of the Atlantic salmon genome as it is sequenced.

4.1 Evolution of genome duplication in fish compared to frog system

Gene duplication plays a fundamental role in biological evolution. Three fates of gene duplication were proposed in the DDC model (Force et al. 1999). Several studies have focused on the divergence of duplicated genes arising from a genome duplication in vertebrates. For example, the African clawed frog allotetraploid-derived *Xenopus laevis* and the diploid *Xenopus tropicalis* were used as a model to study the signatures of selection and to measure the evolutionary divergence between triplets, in which a single gene in *X. tropicalis* corresponds to two paralogous genes within *X. laevis* arising from the whole genome duplication (Hellsten et al. 2007). Their results based on the EST sequences of the triplets supported the hypothesis that duplicated genes are retained under the process of subfunctionalization and relaxation of constraint on

both copies of an ancestral gene (Hellsten et al. 2007). The pairwise d_N/d_S ratio within triplets indicated that purifying selection had occurred, but the ratio between paralogues in *X. laevis* was higher relative to their *X. tropicalis* orthologues (Morin et al. 2006). They suggested that there were asymmetric evolutionary rates within the triplets of the *Xenopus* family. One of the paralogues evolved faster than the other in *X. laevis* and the single gene in *X. tropicalis*, and this is consistent with the neofunctionalization (Chain and Evans 2006).

In contrast to the *Xenopus* genome duplication study, I used the well-characterized LDH genes instead of EST sequences in my study. I have characterized three LDH genes and sequenced BACs containing each of the three LDH genes in rainbow smelt, and used them as representative diploid reference genes to study the fate of paralogous genes formed by the salmonid genome duplication.

The whole genome duplication study in the *Xenopus* family did not use an ancestral out-group of vertebrate to measure the distances and rates of change along the lineages leading to the diploid and tetraploid *Xenopus* species. In my study, I chose zebrafish, tetraodon, takifugu, medaka, killifish, stickleback and dogfish as a teleost out-group to study the amino acid substitutions and evolutionary rates in the lineages of the salmonid and rainbow smelt system. The results show that the rate of change differs with respect to LDH type: LDH-C accepting mutations faster than LDH-B which is evolving faster than LDH-A. In addition, in each case, the rate along the salmonid lineage is greater than in the rainbow smelt lineage. Moreover, the *Xenopus* study did not have a sister

tetraploid species to compare the paralogues with *X. laevis*. In my project, I chose rainbow trout as a sister group to Atlantic salmon to identify amino acid changes after the salmonid genome duplication but before the speciation of Atlantic salmon and rainbow trout. By taking this approach I was able to show that the patterns of amino acid substitutions varies between LDH-A and LDH-B. In particular, there is a lineage specific increase in the rate of amino acid substitutions in the LDH-A1 lineage in Atlantic salmon. However, in LDH-B there is a burst of amino acid substitutions after the salmonid genome duplication but before the salmonid speciation occurs.

Synonymous changes (d_S) can be used as an indicator to measure of evolutionary distances between paralogues and orthologues in pairwise comparisons of Atlantic salmon, rainbow trout and rainbow smelt. The results I obtained are consistent with the predicted evolutionary relationships of the species and gene duplicates. Hellsen et al. (2006) used a different method to measure the evolutionary divergence between *X. laevis* and *X. tropicalis* orthologues and paralogues. They used the transversion rate at four-fold degenerate synonymous codon positions (4DTv) as the indicator. They list several advantages of using transversions rather than total nucleotide substitutions: (1) transversions have a slower rate of occurrence than transitions; (2) transversions provide a simpler situation which was not required to deal with multi-substitution corrections; and (3) transversions are insensitive to protein function such as GC content and methylation. It would be worthwhile to apply the 4DTv methods to the LDH gene model to determine if it reduced possible

substitution errors and improved the accuracy of d_S measurement for testing the divergence between pairwise sequences.

The sequence of the BAC DNA of each LDH gene in Atlantic salmon and rainbow smelt allowed me to look at the regions surrounding the LDH loci in different genomes so that I could see if genome rearrangements had occurred after the whole genome duplication between diploid and tetraploid species. I searched for conservation of synteny and changes such as inversions and deletions in the genomic regions containing LDH genes in rainbow smelt and the corresponding regions of Atlantic salmon. The results suggest that considerable rearrangement occurred in the LDH-A regions of the Atlantic salmon genome. The finding of conservation of synteny between the LDH-B and LDH-C loci supports the statement that LDH-C is derived by the gene duplication from LDH-B in teleosts (Whitt et al. 1975). For the further study of genome reorganization in the LDH regions, I would like to take a more detailed investigation based on inter-genic sequences surrounding the LDH genes in rainbow smelt and use this information to compare with the corresponding regions in the Atlantic salmon genome. This would enable me to determine if there are repetitive elements in these regions and if they are conserved or may have played a role in genome rearrangements and reorganizations.

4.2 Future work

From my study on the characterization and evolution of the LDH genes in rainbow smelt and salmonids, several points arise which lead to suggestions for further work on this project.

All of the LDH genes appear to have evolved under negative selection with few nonsynonymous changes between salmonids and rainbow smelt. However, some non-conservative amino acid changes were observed. Although these changes did not occur at positions that have been identified as part of the active site in LDH, they may cause subtle changes in kinetic parameters. It would be interesting to be able to produce these proteins and carry out enzyme measurements to determine if this is the case. Also, it would be possible to put the amino acid sequences into the 3D model of LDH (Eventoff et al. 1977) to see if this predicted changes in the LDH function.

The tissue expression study suggested that the rainbow smelt LDH-B gene is not expressed in liver. This result is not consistent with the subfunctionalization model for the LDH-B gene, which notes that the Atlantic salmon LDH-B1 is expressed in liver and LDH-B2 predominates in heart, and predicts that the rainbow smelt LDH-B should be expressed in both of these tissues. A qPCR analysis may resolve this issue. By measuring accurately the relative amounts of LDH transcripts in each salmonid and rainbow smelt tissue it may be possible to explore the putative subfunctionalization in LDH-B. The bioinformatic analysis of TFBS in the LDH-B regulatory region among Atlantic salmon, rainbow smelt and other teleosts failed to identify specific TFBS in 1000 bp upstream of the initiation codon in the LDH-B genes of Atlantic salmon and rainbow smelt. This is probably due to the small database of TFBS and the lack of information for fish TFBS. A goal of a future study should be to understand how promoters of LDH genes operate and to identify complementary mutations

that regulate different tissue expressions in salmonid paralogues. Having the genome sequences of Atlantic salmon and rainbow trout will make this type of analysis possible.

REFERENCE LIST

Allendorf, F.W. & Thorgaard, G.H., (1984). Tetraploidy and the evolution of salmonid fishes. In: Turner BJ ed. *Evolutionary Genetics of Fishes (New York: Plenum Press),* 1-53.

Allendorf, F. W., & Danzmann, R. G. (1997). Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics, 145*(4), 1083-1092.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C. & Fritz, A. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science (New York, N.Y.), 282*(5394), 1711-1714.

Amores, A., Suzuki, T., Yan, Y. L., Pomeroy, J., Singer, A. & Amemiya, C. (2004). Developmental roles of pufferfish hox clusters and genome evolution in ray-fin fish. *Genome Research, 14*(1), 1-10.

Appella, E. & Markert, C. L. (1961). Dissociation of Lactate Dehydrogenase into Subunits with Guanidine Hydrochloride. *Biochem. Biophys. Res. Commun. 6,* 161-176.

Bailey, G. S. & Wilson, A. C. (1968). Homologies between isoenzymes of fishes and those of higher vertebrates. evidence for multiple H-4 lactate dehydrogenases in trout. *The Journal of Biological Chemistry, 243*(22), 5843-5853.

Blanco, A. & Zinkham, W. H. (1963). Lactate dehydrogenases in human testes. *Science (New York, N.Y.), 139*(3555), 601-602.

Boyer, S. H., Fainer, D. C. & Watson-Williams, E. J. (1963). Lactate dehydrogenase variant from human blood: Evidence for molecular subunits. *Science (New York, N.Y.), 141*, 642-643.

Brosius, J. (1991). Retroposons-seeds of evolution. *Science, 251*(4995), 753.

Cahn, R. D., Zwilling, E., Kaplan, N. O. & Levine, L. (1962). Nature and development of lactic dehydrogenases: The two major types of this enzyme form molecular hybrids which change in makeup during development. *Science (New York, N.Y.), 136*(3520), 962-969.

Chain, F. J. & Evans, B. J. (2006). Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *xenopus laevis*. *PLoS Genetics, 2*(4), e56.

Cristescu, M. E., Innes, D. J., Stillman, J. H. & Crease, T. J. (2008). D- and L-lactate dehydrogenases during invertebrate evolution. *BMC Evolutionary Biology, 8*, 268.

Darwin, Charles (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st ed.), London.

Dingerkus, G. & Howell, W. M. (1976). Karyotypic analysis and evidence of tetraploidy in the north american paddlefish, polyodon spathula. *Science (New York, N.Y.), 194*(4267), 842-844.

Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H. J., Meyer, H. & Keil, W. (1977). Structural adaptations of lactate dehydrogenase isozymes. *Proceedings of the National Academy of Sciences of the United States of America, 74*(7), 2677-2681.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. error probabilities. *Genome Research, 8*(3), 186-194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. accuracy assessment. *Genome Research, 8*(3), 175-185.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics, 151*(4), 1531-1545.

Friedman, R. & Hughes, A. L. (2001). Pattern and timing of gene duplication in animal genomes. *Genome Research, 11*(11), 1842-1847.

Gordon, D., Abajian, C. & Green, P. (1998). Consed: A graphical tool for sequence finishing. *Genome Research, 8*(3), 195-202.

Graur, D. & Li, W. (2000). *Fundamentals of molecular evolution* (2nd ed.). Sunderland, Ma.: Sinauer.

Gregory. T.R. (2005). *The evolution of the genome*. (1st ed.). Elsevier Academic Press, MA, USA.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews.Genetics, 10*(8), 551-564.

Hellsten, U., Khokha, M. K., Grammer, T. C., Harland, R. M., Richardson, P. & Rokhsar, D. S. (2007). Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *xenopus laevis*. *BMC Biology, 5*, 31.

Hoegg, S., Brinkmann, H., Taylor, J. S. & Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *Journal of Molecular Evolution, 59*(2), 190-203.

Hoegg, S. & Meyer, A. (2005). Hox clusters as models for vertebrate genome evolution. *Trends in Genetics : TIG, 21*(8), 421-424.

Holland, P. W. & Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. *Developmental Biology, 173*(2), 382-395.

Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development (Cambridge, England).Supplement,* 125-133.

Holland, P. W. H. (1997). Vertebrate evolution: Something fishy about hox genes. *Current Biology, 7*(9), R570-R572.

Holmes, R. S. & Markert, C. L. (1969). Immunochemical homologies among subunits of trout lactate dehydrogenase isozymes. *Proceedings of the National Academy of Sciences of the United States of America, 64*(1), 205-210.

Holmes, R. S. (1972). Evolution of lactate dehydrogenase genes. *FEBS Letters, 28* (1), 51-55.

Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K. & Bragin, E. (2009). Ensembl 2009. *Nucleic Acids Research, 37*, D690-7.

Hunter, R. L. & Markert, C. L. (1957). Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science, 125*(3261), 1294-1295.

Huss, S.S. (1995) Quality and quality changes in fresh fish. Food and Agriculture Organization of the United Nations. Rome, Italy.

Jacobs, G. H., Neitz, M., Deegan, J. F., & Neitz, J. (1996). Trichromatic colour vision in new world monkeys. *Nature, 382*(6587), 156-158.

Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N. & Mauceli, E. (2004). Genome duplication in the teleost fish tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature, 431*(7011), 946-957.

Kaessmann, H., Vinckenbosch, N. & Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. *Nature Reviews.Genetics, 10*(1), 19-31.

Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. (2008). Splign: Algorithms for computing spliced alignments with identification of paralogs. *Biology Direct, 3*, 20.

Kochhar, S., Hunziker, P. E., Leong-Morgenthaler, P. & Hottinger, H. (1992). Evolutionary relationship of NAD(+)-dependent D-lactate dehydrogenase: Comparison of primary structure of 2-hydroxy acid dehydrogenases. *Biochemical and Biophysical Research Communications, 184*(1), 60-66.

Koop, B. F. & Davidson, W. S. (2008) Genomics and the Genome Duplication in Salmonids. *Fisheries for Global Welfare and Environment, 5th Wold Fisheries Congress*, 77-86.

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

Larhammar, D., Lundin, L. G. & Hallbook, F. (2002). The human hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Research, 12*(12), 1910-1920.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A. & McWilliam, H. (2007). Clustal W and clustal X version 2.0. *Bioinformatics (Oxford, England), 23*(21), 2947-2948.

Lemons, D. & McGinnis, W. (2006). Genomic evolution of hox gene clusters. *Science (New York, N.Y.), 313*(5795), 1918-1922.

Li, S. S.-L., Fitch, W. M., Pan, Y.-C. E. & Sharief, F. S. (1983). Evolutionary relationships of vertebrate lactate dehydrogenase isozymes A4 (muscle), B4 (heart) and C4 (testis).*J. Biol. Chem. 258,* 7029-7032.

Li, Y. J., Tsoi, S. C., Mannen, H. & Shoei-lung Li, S. (2002). Phylogenetic analysis of vertebrate lactate dehydrogenase (LDH) multigene families. *Journal of Molecular Evolution, 54*(5), 614-624.

Lim, S. T. & Bailey, G. S. (1977). Gene duplication in salmonid fishes: Evidence for duplicated but catalytically equivalent A4 lactate dehydrogenases. *Biochemical Genetics, 15*(7-8), 707-721.

Lynch, M. & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.), 290*(5494), 1151-1155.

Málaga-Trillo E. & Meyer A. (2001). Genome Duplications and Accelerated Evolution of Hox Genes and Cluster Architecture in Teleost Fishes. *Am Zool*. *41,* 676-686.

Mannen, H., Tsoi, S. C., Krushkal, J. S., Li, W. H. & Li, S. S. (1997). The cDNA cloning and molecular evolution of reptile and pigeon lactate dehydrogenase isozymes. *Molecular Biology and Evolution, 14*(11), 1081-1087.

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

Markert, C. L. (1963). Lactate dehydrogenase isozymes: Dissociation and recombination of subunits. *Science (New York, N.Y.), 140*(3573), 1329-1330.

Markert, C. L. (1968). The molecular basis for isozymes. *Annals of the New York Academy of Sciences, 151*(1), 14-40.

Markert, C. L. & Møller, F. (1959). Multiple forms of enzymes: Tissue, ontogenetic, and species specific patterns. *Proceedings of the National Academy of Sciences of the United States of America, 45*(5), 753-763.

Markert, C. L., Shaklee, J. B. & Whitt, G. S. (1975). Evolution of a gene. multiple genes for LDH isozymes provide a model of the evolution of gene structure, function and regulation. *Science (New York, N.Y.), 189*(4197), 102-114.

McKay, S. J., Trautner, J., Smith, M. J., Koop, B. F. & Devlin, R. H. (2004). Evolution of duplicated growth hormone genes in autotetraploid salmonid fishes. *Genome / National Research Council Canada, 47*(4), 714-723.

Moghadam, H. K., Ferguson, M. M. & Danzmann, R. G. (2005a). Evidence for hox gene duplication in rainbow trout (O*ncorhynchus mykiss*): A tetraploid model species. *Journal of Molecular Evolution, 61*(6), 804-818.

Moghadam, H. K., Ferguson, M. M. & Danzmann, R. G. (2005b). Evolution of hox clusters in salmonidae: A comparative analysis between atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). *Journal of Molecular Evolution, 61*(5), 636-649.

Morin, R. D., Chang, E., Petrescu, A., Liao, N., Griffith, M. & Chow, W. (2006). Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome remodeling. *Genome Research, 16*(6), 796-803.

Nakae, Y. & Stoward, P. J. (1994). The diverse michaelis constants and maximum velocities of lactate dehydrogenase in situ in various types of cell. *The Histochemical Journal, 26*(4), 292-297.

Nance, W. E., Claflin, A. & Smithies, O. (1963). Lactic dehydrogenase: Genetic control in man. *Science (New York, N.Y.), 142*, 1075-1077.

Nelson, J. S. (1994). *Fishes of the world* (3rd ed.). Wiley and Sons, New York.

Nelson, J. S. (2006). *Fishes of the world* (4th ed.). Hoboken, N.J.: J. Wiley.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution, 3*(5), 418-426.

Ohno, S., Muramoto, J., Stenius, C., Christian, L., Kittrell, W. A. & Atkin, N. B. (1969). Microchromosomes in holocephalian, chondrostean and holostean fishes. *Chromosoma, 26*(1), 35-40.

Ohno, S. (1970a). *Evolution by gene duplication. -*. Berlin; New York: Springer-Verlag.

Ohno S. (1970b). The Enormous Diversity in Genome Sizes of Fish as a Reflection of Nature's Extensive Experiments with Gene Duplication. *Transactions of the American Fisheries Society*. *99,*120-130.

Phyllips, R.B., & Rab, P. (2001). Chromosome evolution in the Salmonidae (Pisces): An update. *Biol Rev, 76,* 1-25.

Pond, S. L. & Frost, S. D. (2005a). Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics (Oxford, England), 21*(10), 2531-2533.

Pond, S. L. & Frost, S. D. (2005b). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution, 22*(5), 1208-1222.

Postlethwait, J. H., Woods, I. G., Ngo-Hazelett, P., Yan, Y. L., Kelly, P. D. & Chu, F. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Research, 10*(12), 1890-1902.

Prohaska, S. J. & Stadler, P. F. (2004). The duplication of the hox gene clusters in teleost fishes. *Theory in Biosciences, 123*(1), 89-110.

Rehse P.H., & Davidson W.S. (1986). Evolutionary Relationship of a Fish C Type Lactate Dehydrogenase to Other Vertebrate Lactate Dehydrogenase Isozymes. *Can.J.Fish.Aquat. Sci*. *43,*1045-1051.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K. & Nakazato, T. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science (New York, N.Y.), 301*(5637), 1211-1216.

Rise, M. L., von Schalburg, K. R., Brown, G. D., Mawer, M. A., Devlin, R. H. & Kuipers, N. (2004). Development and application of a salmonid EST database and cDNA microarray: Data mining and interspecific hybridization characteristics. *Genome Research, 14*(3), 478-490.

Schalburg, K., Leong, J., Cooper, G., Robb, A., Beetz-Sargent, M., Lieph, R., Holt, R., Moore, R., Ewart, K., Driedzic, W., ten Hallers, B., Zhu, B., Jong, P., Davidson, W. & Koop, B. (2008). Rainbow Smelt (*Osmerus mordax*) Genomic Library and EST Resources. *Marine Biotechnology, 10* (5), 487-491.

Sensabaugh, G. F.,Jr. & Kaplan, N. O. (1972). A lactate dehydeogenase specific to the liver of gadoid fish. *The Journal of Biological Chemistry, 247*(2), 585-593.

Shaw, C. R. & Barto, E. (1963). Genetic evidence for the subunit structure of lactate dehydrogenase isozymes. *Proceedings of the National Academy of Sciences of the United States of America, 50*, 211-214.

Shiina, T., Dijkstra, J. M., Shimizu, S., Watanabe, A., Yanagiya, K. & Kiryu, I. (2005). Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (oncorhynchus mykiss), a species with a presumably recent tetraploid ancestry. *Immunogenetics, 56*(12), 878-893.

Spring, J. (2003). Major transitions in evolution by genome fusions: From prokaryotes to eukaryotes, metazoans, bilaterians and vertebrates. *Journal of Structural and Functional Genomics, 3*(1-4), 19-25.

Stock, D. W. & Whitt, G. S. (1992). Evolutionary implications of the cDNA sequence of the single lactate dehydrogenase of a lamprey. *Proceedings of the National Academy of Sciences of the United States of America, 89*(5), 1799-1803.

Suyama, M., Torrents, D. & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research, 34*(Web Server issue), W609-12.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution, 24*(8), 1596-1599.

Taylor, S.S., Oxley, S.S., Allison, W.S. & Naplan, N.O. (1973). Aminoacid sequence of Dogfish M_4 Lactate Dehydrogenase. *Proc. Nat. Acad. Sci. USA*. *70* (6), 1790-1794.

Taylor, J. S., Van de Peer, Y., Braasch, I., & Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 356*(1414), 1661-1679.

Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. & Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Research, 13*(3), 382-390.

Taylor, J. S. & Raes, J. (2004). Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics, 38*(1), 615.

Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L. & Ristow, S. S. (2002). Status and opportunities for genomics research with rainbow trout. *Comparative Biochemistry and Physiology.Part B, Biochemistry & Molecular Biology, 133*(4), 609-646.

Trask, B. J. (2002). Human cytogenetics: 46 chromosomes, 46 years and counting. *Nature Reviews.Genetics, 3*(10), 769-778.

Tsoi, S. C. & Li, S. S. (1994). The nucleotide and deduced amino-acid sequences of a cDNA encoding lactate dehydrogenase from caenorhabditis elegans: The evolutionary relationships of lactate dehydrogenases from mammals, birds, amphibian, fish, nematode, plants, bacteria, mycoplasma, and plasmodium. *Biochemical and Biophysical Research Communications, 205*(1), 558-564.

Tsuji, S., Qureshi, M. A., Hou, E. W., Fitch, W. M. & Li, S. S. (1994). Evolutionary relationships of lactate dehydrogenases (LDHs) from mammals, birds, an amphibian, fish, barley, and bacteria: LDH cDNA sequences from xenopus, pig, and rat. *Proceedings of the National Academy of Sciences of the United States of America, 91*(20), 9392-9396.

Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A. & Van de Peer, Y. (2004). Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proceedings of the National Academy of Sciences of the United States of America, 101*(6), 1638-1643.

Vinals, C., Depiereux, E. & Feytmans, E. (1993). Prediction of structurally conserved regions of D-specific hydroxy acid dehydrogenases by multiple alignment with formate dehydrogenase. *Biochemical and Biophysical Research Communications, 192*(1), 182-188.

Volff, J. N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity, 94*(3), 280-294.

Wasserman, W. W. & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nature Reviews.Genetics, 5*(4), 276-287.

Whitt, G. S. (1969). Homology of lactate dehydrogenase genes: E gene function in the teleost nervous system. *Science (New York, N.Y.), 166*(909), 1156-1158.

Whitt, G. S. (1970). Development genetics of the lactate dehydrogenase isozymes of fish. *The Journal of Experimental Zoology, 175*(1), 1-35.

Whitt, G. S. & Booth, G. M. (1970). Localization of lactate dehydrogenase activity in the cells of the fish (xiphorphorus helleri) eye. *The Journal of Experimental Zoology, 174*(2), 215-224.

Whitt, G.S., Shaklee, J.B. & Markert, C.L. (1975). *Isozymes IV Genetics and Evolution.* Markert C.L., Ed.: 381-414. Academic Press, New York.

Wistow, G. J., Mulders, J. W., & de Jong, W. W. (1987). The enzyme lactate dehydrogenase as a structural protein in avian and crocodilian lenses. *Nature, 326*(6113), 622-624.

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

Yokoyama, S. & Yokoyama, R. (1989). Molecular evolution of human visual pigment genes. *Molecular Biology and Evolution, 6*(2), 186-197.

Zhang, J., Zhang, Y. P. & Rosenberg, H. F. (2002). Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics, 30*(4), 411-415.

Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology & Evolution, 18*(6), 292-298.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W., & Wilson, .A.C. (1980). Rapid duplication and loss of genes coding for the alpha chains of haemoglobin. *Proc Natl Acad Sci USA, 77*(4), 2158-62.

Zinkham, W. H. (1968). Lactate dehydrogenase isozymes of testis and sperm: Biological and biochemical properties and genetic control. *Annals of the New York Academy of Sciences, 151*(1), 598-610.

Zinkham, W. H., Isensee, H. & Renwick, J. H. (1969). Linkage of lactate dehydrogenase B and C loci in pigeons. *Science (New York, N.Y.), 164*(876), 185-187.

APPENDIX 1: LIST OF WEBSITES

Website Name	Website	Goal
consortium for Genomics Research on All Salmon Project	http://web.uvic.ca/grasp/	a major player internationally in salmonid genomics.
Animal Genome Size Database	http://www.genomesize.com/	a comprehensive catalogue of animal genome size data. Haploid DNA contents (C-values, in picograms).
Genomics Research on All Salmon Project	http://grasp.mbb.sfu.ca/	all aspects of genomics research on Atlantic salmon
Splign: Spliced Alignments	http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi?textpage=overview&level=form	a utility for computing cDNA-to-Genomic, or spliced sequence alignments.
ExPASy Translation Tool	http://www.expasy.ch/tools/dna.html	a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.
Adaptive Evolution Server: DATAMONKEY	http://www.datamonkey.org/dataupload.php	a webserver to test the signature of positive or negative selection.
PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments	http://www.bork.embl.de/pal2nal/	a program that converts a multiple sequence alignment of proteins and the corresponding DNA (or mRNA) sequences into a codon alignment and test dN and dS value.
JASPAR: The high-quality transcription factor binding profile database	http://jaspar.genereg.net/	a curated, non-redundant transcription factor binding sites for multicellular eukaryotes.