CHARACTERIZATION OF THE ATLANTIC SALMON SEX-DETERMINING CHROMOSOME

by

Carlo Artieri
BSc (Hons), Dalhousie University, 2003

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

MASTER OF SCIENCE

In the
Department of Molecular Biology and Biochemistry

© Carlo Artieri 2005

SIMON FRASER UNIVERSITY

Spring 2005

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author.

APPROVAL

Name: **Carlo Giovanni Artieri**

Degree: **Master of Science**

Title of Thesis: **Characterization of the Atlantic salmon sex-determining chromosome**

Examining Committee:

Chair: **Dr. Margo Moore**
Associate Professor, Department of Biological Sciences

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

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

---

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

---

Dr. Robert H. Devlin
Public Examiner
Research Scientist, Department of Fisheries and Oceans Canada

Date Defended/Approved: April 18, 2005

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

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

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection.

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

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

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

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

W. A. C. Bennett Library
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

I have integrated data from linkage mapping, physical mapping and karyotyping in order to characterize the sex chromosomes in Atlantic salmon (*Salmo salar*). The primary genetic sex-determing signal, *SEX,* has been mapped to Atlantic salmon microsatellite linkage group 1 (ASL1). I have used probes designed from the flanking regions of these sex-linked microsatellite markers to screen a bacterial artificial chromosome (BAC) library, representing an 11.7X coverage of the genome, which has been Hind III fingerprinted and assembled into contigs. BACs containing sex-linked microsatellites and their related contigs have been identified and representative BACs have been placed on Atlantic salmon chromosomes by fluorescent *in situ* hybridization (FISH). This identified chromosome 2, as the sex-chromosome and allowed me to orient ASL1 with respect to chromosome 2. The region containing *SEX* appears to lie on the long arm between marker Ssa202DU and a region of heterochromatin identified by DAPI staining.

DEDICATION

To my high-school principal, who told me that I was not "university material". Thank you for giving me a goal to pursue and thus the drive to succeed.

> *"Education is the ability to listen to almost anything without losing your temper or your self-confidence."*
>
> - Robert Frost (1874 – 1963)

ACKNOWLEDGEMENTS

I would sincerely like to thank my supervisor, Dr. William Davidson, for giving me the opportunity to work on this project. I would also like to thank all of my lab mates for creating an atmosphere in which my knowledge and interest in science could grow and develop. Special thanks are reserved for the members of Dr. Ruth Phillips' lab at Washington State University for all of the help with cytogenetics, Dr. Roy Danzmann at the University of Guelph for assistance with linkage mapping as well as the members of the British Columbia Genome Sciences Centre for assistance in physical mapping and bioinformatics tools. Finally, I would like to thank my family and friends for their support and last but not least, my girlfriend Julie for sticking with me during the "trying" periods of this project.

TABLE OF CONTENTS

Approval ii

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Figures ix

List of Tables x

Chapter 1 Introduction 1

1.1 Sexual Reproduction 1
1.2 Sex-Determining Mechanisms 4
1.2.1 Genetic mechanisms 4
1.2.2 Environmental Mechanisms 5
1.3 Sex-Determining Pathways 6
1.3.1 Description of sex-determining pathways 6
1.3.2 Sex-determination in humans 7
1.3.3 Sex-determination in *D. melanogaster* 9
1.3.4 Sex-determination in *C. elegans* 12
1.3.5 Evolutionary conservation in sex-determining pathways 14
1.4 Sex-Determination in Fishes 15
1.4.1 Teleost fishes 15
1.4.2 Characterization of teleost sex-determining chromosomes 16
1.5 Salmonid Fishes 16
1.5.1 Family Salmonidae 16
1.5.2 Sex-determination in Salmonids 18
1.6 Atlantic salmon (*Salmo salar*) 23
1.6.1 The Atlantic salmon karyotype 23
1.6.2 Genetic linkage mapping in Atlantic salmon 24
1.6.3 Sex-determination in Atlantic salmon 25
1.6.4 Genomics Research on Atlantic Salmon Project (GRASP) resources 25
1.7 Approach to Identify the *SEX* in Atlantic salmon 27
1.7.1 Previous approaches to characterization of *SEX* in other species 27
1.8 Aim of the Thesis 29

Chapter 2 Materials and Methods 30

2.1 Integration of Atlantic Salmon Genomic and Physical Maps 30
2.1.1 BAC Library Screening 30
2.1.2 Positive Conformation by PCR 31
2.2 Physical Map *in silico* Analysis and Extension 32
2.3 BAC-end Sequencing and Novel Gene Identification 32
2.4 Chromosome "Walking" 34
2.4.1 Probe design 34
2.4.2 Confirmation of primer viability by PCR 35
2.4.3 BAC library screening 35
2.4.4 Analysis of clone S0503M23 36
2.5 Fluorescent *in situ* Hybridization 37
2.6 Sequencing and Analysis of Clone S0493K22 38
2.6.1 Shotgun library construction 38
2.6.2 Assay for the presence of *E. Coli* genomic contamination 40
2.6.3 Sequencing of shotgun library 40
2.6.4 Gene identification 41
2.6.5 Repetitive element analysis 42

Chapter 3 Results 43

3.1 Integration of Atlantic Salmon Genetic and Physical Maps 43
3.1.1 Atlantic salmon *SEX*-Associated Genetic Linkage Map 43
3.1.2 Incorporation of *SEX*-linked BACs into the Physical Map 50
3.2 Extension of Coverage of *SEX*-linked Physical Map 54
3.2.1 *In silico* contig extension using FPC 54
3.2.2 BAC-end sequencing 57
3.2.3 Chromosome walking 62
3.3 Fluorescent *in situ* Hybridization of Sex-Linked Clones 81
3.3.1 Identification of the Atlantic salmon sex-determining chromosome 81
3.3.2 Orientation of linkage map relative to the sex-determining chromosome 84
3.3.3 Chromosome mapping of duplicate microsatellite loci 86
3.4 Analysis of BAC clone S0493K22 89
3.4.1 Generation and sequencing of a shotgun library of clone S0493K22 89
3.4.2 Sequence annotation 92
3.4.3 Gene synteny analysis 96
3.4.4 Repetitive element content analysis 98

Chapter 4 Discussion 99

4.1 Position of *SEX* on the Y-chromosome of Atlantic Salmon Based on the Genetic Linkage Map 99
4.2 Integration of ASL1 with the Hind III Digested Physical Map 100
4.2.1 BAC library screening 100
4.2.2 Incorporation of *SEX*-linked BACs into the physical map 101
4.3 Extension of Coverage of the Sex Chromosome 102
4.3.1 *In silico* contig extension via FPC 102

4.3.2 BAC-end sequencing of contig terminal clones 104
4.3.3 Chromosome-walking from sex-linked contigs 106
4.4 Fluorescent in situ Hybridization of Sex-Linked Clones 114
4.4.1 Identification of the Atlantic salmon sex chromosome 114
4.4.2 Orientation of the linkage map relative to the sex chromosome 114
4.4.3 Fluorescent *in situ* hybridisation (FISH) of duplicate loci 118
4.5 Analysis of BAC clone S0493K22 121
4.5.1 Sequence location 121
4.5.2 Sequence annotation 122
4.5.3 Analysis of repetitive DNA content 123

References 125

LIST OF FIGURES

Figure 1.1: The "double-repressor" hypothesis in humans. 8

Figure 1.2: The *Drosophila melanogaster* sex-determination pathway. 11

Figure 1.3: The *Caenorhabditis elegans* sex-determining pathway. 13

Figure 1.4: Phylogenetic tree of Salmonid fishes. 17

Figure 3.1: Atlantic Salmon Linkage Group 1 (ASL1). 44

Figure 3.2: CHORI-214 Atlantic salmon BAC library membrane hybridized with the OMM1278 marker probe. 49

Figure 3.3: Contig 2708 as viewed with iCE version 3.4. 53

Figure 3.4: Integration of ASL1 with the Hind III fingerprinted physical map. 56

Figure 3.5: Generalized flow-scheme of chromosome walking procedure. .. 67

Figure 3.6: CHORI-214 Atlantic salmon genomic BAC library high-density spotted membrane, probed with a repetitive element-containing probe. 72

Figure 3.7: Representation of contig-contig joining via BAC-end sequence designed probes. 77

Figure 3.8: Contig 2133 of the Hind III fingerprinted physical map as viewed in iCE 3.4. 79

Figure 3.9: PCR amplification of clones in contig 2133 using S0503M23 end-primers..... 80

Figure 3.10: Results of Fluorescent *in situ* hybridization. 83

Figure 3.11: Orientation of ASL1 with respect to the sex chromosome. 85

Figure 3.12: Fluorescent *in situ* hybridization of duplicate microsatellite loci. 89

Figure 3.13: Build of clone S0493K22 as viewed using Assembly View in Consed. 91

Figure 3.14: Annotated Genscan output of contig 19-13. 93

Figure 3.15: Annotated Genscan output of contig 16-17. 94

LIST OF TABLES

Table 3.1: Microsatellite markers used in the integration of the genetic linkage map with the genomic BAC library. 46

Table 3.2: Hind III fingerprinted physical map history. 51

Table 3.3: Tabulated results of BLASTn search against the GenBank non-redundant database using BAC-end sequences. 59

Table 3.4: Genes identified from BAC-end sequences. 61

Table 3.5: Clones selected for chromosome walking. 63

Table 3.6: Results of the 34 probes used in the screening. 69

Table 3.7: Syntenic relationship of genes identified on clone S0493K22. 97

CHAPTER 1 INTRODUCTION

1.1 Sexual Reproduction

One of the most fundamental biological processes is the determination of phenotypic sex followed by the subsequent differentiation of an organism into a particular reproductive fate. This differentiation includes not only the development of reproductive organs, but also affects almost all aspects of an organism including social behaviour, physiology and morphology (Beye et al., 2003).

A basic question in the study of evolutionary biology asks; 'what advantages do sexual organisms (having a life cycle with alternating syngamy and meiosis) confer over asexual organisms (producing an offspring from single mitotically derived cells)?' Despite almost a century of debate, the major forces driving the evolution and maintenance of sexual reproduction from an originally asexual population remain poorly understood (Kondrashov, 1993). From a purely energetic standpoint, sexual reproduction is at a disadvantage, as it requires more energy to retain two of each haploid genome within the cell, as well as the energy expended in mating with another organism (Lively and Loyd, 1990).

Two principle advantages conferred by sexual reproducing organisms are that they are less prone to the accumulation of deleterious mutations and that they can adapt more rapidly to changing ecological environments (Kondrashov, 1993). These observations have led to the two most widely accepted hypotheses as to why sexual reproduction evolved, and why it has been maintained.

Kondrashov's 'Mutational Deterministic' (MD) hypothesis (Kondrashov, 1993) proposes that a high genomic rate of deleterious mutation may have been the principle drive towards the evolution of sexual reproduction from asexual reproduction. Simply put, the cost of perfecting one's genomic repair machinery increases as one nears perfection. Therefore, it may be energetically favourable to allow some mistakes to accumulate if, through the process of recombination, one can purge them later. Under this model, sexual reproduction seems to be favoured over asexual reproduction when the number of deleterious mutations per diploid genome per generation (U) exceeds 1 event per generation when it is assumed that newly arising asexual populations are mutation free (Charlesworth, 1990). Below such a rate for U, an asexual population will overwhelm a sexually reproducing one, because of its greater productivity.

This hypothesis has not been without criticism. Keightley and Eyre-Walker (2000, 2001) compared the genomes of related organisms in order to test whether U was sufficiently near or above the one event per generation required in order for the MD hypothesis to be the sole factor maintaining sexual reproduction among these organisms. While Keightley and Eyre-Walker found values for U above 2 for species with long generation times (such as long-lived mammals), species with short generation times had U values as low as 0.05. Based on these results, the authors suggested that the ability of recombination to purge deleterious mutations alone was not sufficient to maintain sexual reproduction among species with short generation times, and therefore sexual reproduction must confer additional advantages.

Another long-standing theory as to the basis of sexual reproduction is known as the 'Red Queen' (RQ) hypothesis, coined by VanValen (1973), so named after Lewis Carroll's classic novel, *Through the Looking Glass* (Caroll, 1872). The hypothesis is based on the idea that in tightly co-evolved interactions between species (such as a predator and its prey or a parasite and its host), rapid evolutionary change in one species, may lead to extinction in the other. Therefore, each species must continue to evolve in order to remain extant; or as the Red Queen tells Alice, "[I]t takes all the running you can do to keep in the same place." (Caroll, 1872). The hypothesis was further expanded by Bell (1982) through the idea that co-evolution of such co-dependent species could lead to sustained oscillations of genotype. Host organisms capable of asexual reproduction may select for sexual reproduction in a situation of co-evolution with parasites, as a mechanism to reduce the chance of infection in its offspring.

Recent mathematical modelling of species-species interactions under conditions of co-evolution has shown that the RQ hypothesis only holds under conditions of strong selection at very few fitness-associated loci (Otto and Nuismer, 2004). Under conditions of weak selection, or selection acting upon multiple fitness-associated loci, recombination conferred by sexual reproduction is more likely to break apart fit gene combinations and produce less-fit recombinant genotypes. Therefore, although the RQ might explain the origin of a small amount of sex and recombination, much like the MD hypothesis, it is unlikely that it alone can explain the maintenance of sexual reproduction and the high rates of recombination observed in extant species.

While neither the MD nor the RQ hypotheses alone seem sufficient to explain why sexual reproduction has been favoured and maintained among species throughout evolutionary history, they may play a role in the evolution and maintenance of sex by acting together along with other factors such as mutation or random genetic drift. Synergistic effects have been found in models incorporating both the MD and RQ hypotheses (Howard and Lively, 1998) suggesting that only by incorporating elements from various models can we produce a rigorous hypothesis as to the origins of sex.

Our understanding of the evolution of sexual reproduction remains limited because of our inadequate understanding of the molecular mechanisms underlying sex-determination in extant species as well as their ancestors. Expanding our knowledge of these mechanisms is a crucial first step in the understanding of reproduction as a whole.

1.2 Sex-Determining Mechanisms

The diverse mechanisms, by which sex is inherited in organisms with separates sexes, can be divided into two broad categories: genetic mechanisms and environmental mechanisms (Bull, 1983).

1.2.1 Genetic mechanisms

A genetic sex-determination mechanism involves the inheritance of specific elements at the genotypic level, which cause a zygote to develop into a particular sexual fate. Various examples of genetic sex determining mechanisms include:

Male Heterogamety- All males are XY, females XX and therefore all zygotes are either XY or XX. The X and the Y are the inherited basis of sex-determination and sex is determined at conception.

Female Heterogamety- The reverse of male heterogamety where females are ZW, males are ZZ and all zygotes are either ZW or ZZ. Depending on the zygotic mechanism, sex may be determined before fertilization but no later than at the time of the formation of the zygotic nucleus.

Polyfactorial Sex-Determination- Sex is determined in the zygote by multiple factors of small additive effect.

Arrhenotoky (Haplo-diploidy)- Males arise from unfertilized eggs and females from fertilized eggs.

1.2.2 Environmental Mechanisms

Organisms employing environmental sex-determining (ESD) mechanisms use external stimuli in order to determine phenotypic sex. In certain species, sex may be determined during embryogenesis, in response to stimuli such as temperature or salinity. Males develop below a given threshold and females develop at or above the same threshold. Species using ESD may also be able to change their sex after they have reached full sexual maturity. Examples of this case include certain species where individuals may be sensitive to particular population dynamics and males will spontaneously become females when females are scarce.

In organisms employing ESD, the individual's genotype has little influence on whether it becomes male or female. However, it is not assumed that sex-determination occurs without the action of genes, only that the products of such genes are sensitive to the environment.

1.3 Sex-Determining Pathways

1.3.1 Description of sex-determining pathways

A sex-determining pathway can be viewed as a complex series of interacting biochemical processes that ultimately lead to sex-cell determination and differentiation (Devlin and Nagahama, 2002). Despite many years of study, the mechanisms comprising the sex-determining pathways of most species remain elusive. Though the primary sex-determining signal has been identified in most mammals, the nematode, *Caenorhabditis elegans*, the fruit fly, *Drosophila melanogaster*, the honeybee, *Apis mellifera*, and the Japanese Medakafish, *Oryzias latipes*, the downstream effectors of their sex pathways have not all been characterized (Berta et al., 1990; Beye et al., 2003; for review see Stothard and Pilgrim, 2003; Schütt and Nothiger, 2000; Swain, 2002). With only a small representative group of which any substantive amount of sex pathway characterization has been performed, our understanding of how these pathways have evolved (from a common ancestor or independently) remains limited.

Following is a brief discussion of what is known about the sex-determining pathways of humans, *D. melanogaster* and *C. elegans*.

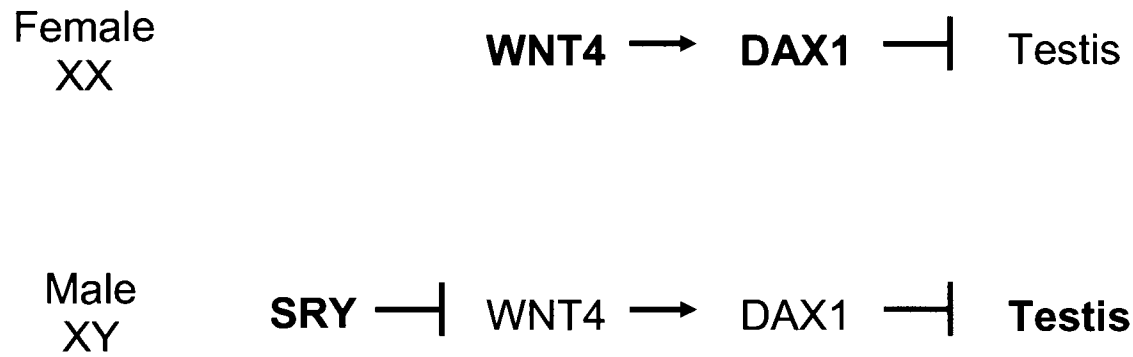
1.3.2 Sex-determination in humans

Humans employ a genetic sex-determination mechanism in which males are the heterogametic sex (XY) and females, homogametic (XX). It is firmly established that the gene *SRY*, which lies on the Y chromosome is the primary genetic signal and that its presence induces male differentiation (Scherer, 2002). Nevertheless, the direct gene target(s) of *SRY* remain unknown. As *SRY* is a member of the HMG domain family of transcription factors, its mode of action is most likely to act as a transcriptional regulator, activating or inhibiting the production of downstream elements.

SRY is thought to act through a "double-repressor" system, in which it inhibits a gene or genes that themselves inhibit masculinization (McElreavey et al., 1993; Vilain, 2002; Scherer, 2002). This was first postulated upon observation that not all XX male individuals expressed *SRY* (acquired via a translocation event), suggesting that they were defective in expression of a downstream male-repressor, the supposed target of *SRY*. Expression studies have led to the model that *SRY* inhibits the action of DAX1, a known testis inhibitor, perhaps through the action of WNT4 (Fig. 1.1) (for review see Vilain, 2002).

Several sexual development pathologies, including full sex-reversal, have been observed in individuals defective in the expression of several other sex-related genes such as *SOX9*, *DMRT1* and *AMH* among others. It is most likely that there are still several downstream elements and cofactors whose positions within the overall sex pathway have yet to be determined.

Figure 1.1 The "double-repressor" hypothesis in humans. Females express WNT4, which is believed to be a signal for the activation of DAX1, a known testis inhibitor. In males, SRY is believed to inhibit the expression of WNT4, thus preventing it from it from activating the expression of DAX1. Proteins whose names are in bold are being expressed, whereas those not in bold are being inhibited. Testis written in bold represents activation of the male developmental pathway. Testis not written in bold represents suppression of the male pathway and activation of the female pathway.



1.3.3 Sex-determination in *D. melanogaster*

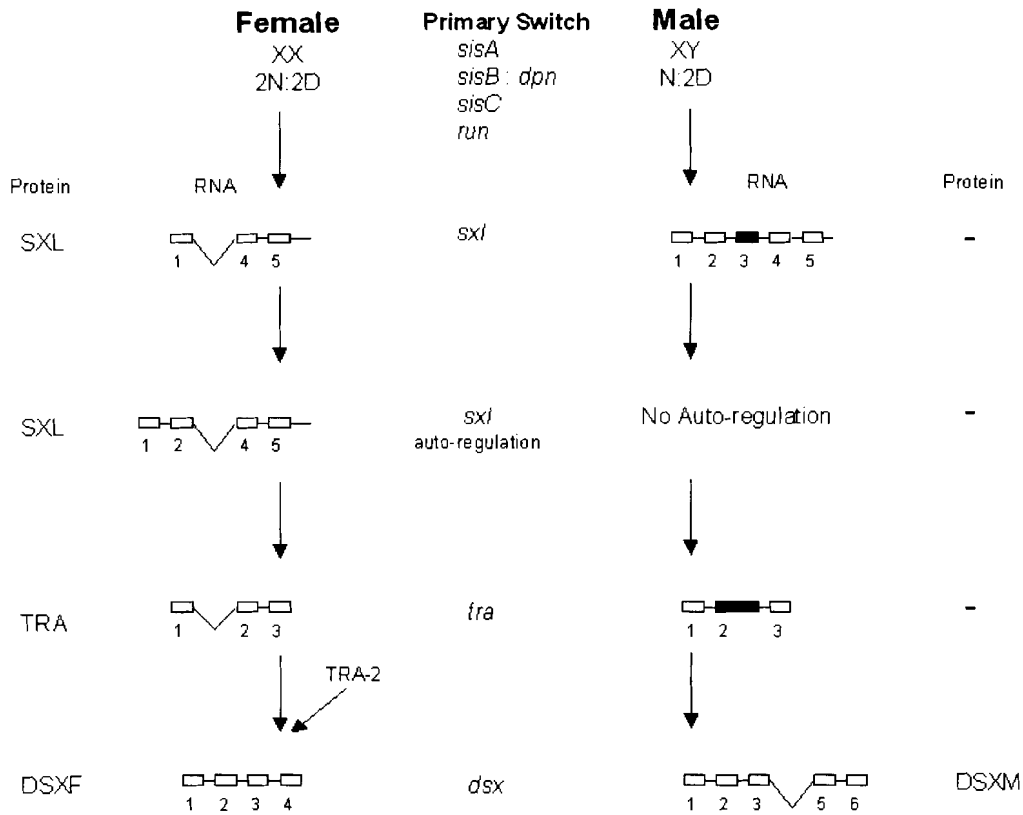
Like in humans and other mammals, sex is determined through a genetic mechanism in the fruit fly, *Drosophila melanogaster* (Fig. 1.2) (Pomiankowski et al., 2004; for review see Schütt and Nöthiger, 2000). The primary genetic signal is provided by the ratio of X-linked numerator genes (N) [the three *sisterless* genes (*sisA,B,C*) and *runt* (*run*)] to the autosomal denominator gene (D), *deadpan* (*dpn*). The ratio in females, who inherit two X chromosomes, is 2:2, whereas in males (bearing only one X chromosome) it is 1:2.

The products of the N and D genes are transcription factors that regulate the expression of the gene *sex-lethal* (*sxl*). *Sex-lethal* has a premature stop codon in exon 3, which if not removed during RNA processing, will lead to the production of non-functional SXL protein. In females, the double dose of X linked numerator genes activates an early promoter of *sxl* (*Pe*) which produces RNA transcripts in which exons 2 and 3 are spliced out. This results in the production of active SXL protein. SXL also acts as a splice enhancer, binding its own pre-mRNA, and enforcing the removal of exon 3. Thus the production of active SXL in females activates an auto-regulatory loop, ensuring the continued production of functional protein. In males, insufficient dosage of the numerator genes leads to the production of prematurely truncated *sxl* and thus the functional SXL protein is not produced, and the auto-regulatory loop is never established.

Active SXL is an RNA binding protein that regulates the production of the next gene in the sex pathway, *transformer* (*tra*). Like *sxl*, *tra* contains a series of premature stop codons at the beginning of exon 2. In females, the action of SXL forces the use of a cryptic splice site downstream of the stop codons in exon 2, preventing premature termination and resulting in the production of active TRA protein. In males, the lack of active SXL prevents the removal of the stop codons and no functional TRA is produced.

TRA is another RNA binding protein that causes alternative splicing of the gene *doublesex* (*dsx*), the next element downstream in the pathway. In females, TRA, in conjunction with the cofactor TRA-2, initiates an alternative splicing pattern in *dsx* including exons 1-4, whose protein product is known as DSXF. In males, absence of TRA causes *dsx* to splice into its default form, removing exon 4. Thus, the male mRNA transcript contains exons 1-3 and 5-6, resulting in the production of the male specific isoform, DSXM. Both DSX isoforms act as transcription factors that enhance and repress the activity of downstream developmental genes in a sex-specific manner, and implement either of the two different routes of sexual differentiation.

Figure 1.2 The *Drosophila melanogaster* sex-determining pathway. Gene names are shown in italics whereas the names of proteins are shown in capital letters. mRNAs are represented by the blue and red boxes. Blue boxes are translated exons whereas red boxes represent exons with pre-mature stop codons. Naming abbreviations are as follows: sis, sisterless; run, runt; dpn, deadpan; sxl, sex-lethal; tra, transformer; dsx, doublesex.

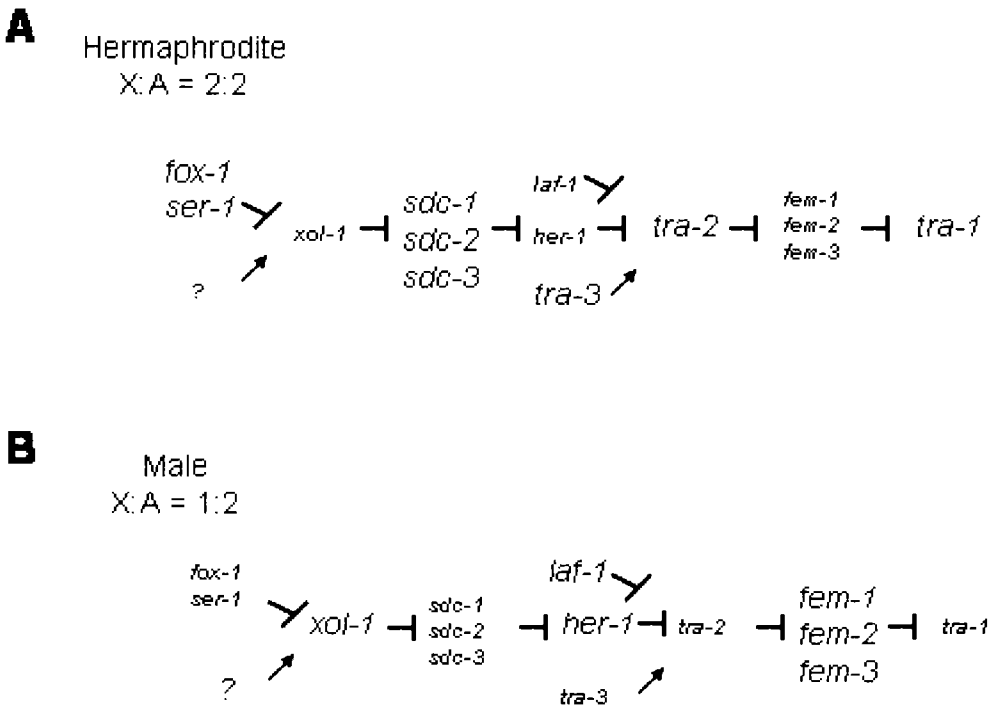


1.3.4 Sex-determination in *C. elegans*

Like the sex pathway in *Drosophila*, the primary genetic signal in the nematode, *Caenorhabditis elegans*, is the ratio of X chromosomes to autosomes (Fig. 1.3) (for review see Stothard and Pilgrim, 2003). However, unlike in *Drosophila*, *C. elegans* exists as either a male (inheriting one X chromosome) or a hermaphrodite (inheriting two X chromosomes). In hermaphrodites (Fig. 1.3A), proper dosage of the X linked genes, *fox-1* and *sex-1,* acts to reduce expression of *xol-1*. This in turn prevents the xol-1's inhibitory effect on the hermaphrodite promoting *sdc* genes (*sdc-1,2* and *3*). The *sdc* genes again act as transcriptional repressors of the *her-1* gene, which itself is a repressor of *tra-2* expression. So consequently in the case of hermaphrodites, tra-2 is able to inhibit the *fem* genes (*fem-1,2* and *3*), preventing them from inhibiting *tra-1,* which acts as a global regulator of sex, promoting hermaphrodite development. In males, the insufficient initial dosage of *fox-1* and *ser-1* lead to opposite repression effects, culminating in the inactivation of *tra-1* (Fig. 1.3B). This initiates the default global pathway of male sexual development.

Although the molecular mechanisms of the primary genes involved in sex determination in *C. elegans* have been well characterized, many trans-acting cofactors feed into the pathway at various stages. Mutations in these cofactors lead to improper sexual development and further characterization of these genes is required in order to understand their effects on the overall pathway.

Figure 1.3 The *Caenorhabditis elegans* sex-determining pathway. The state of sex-determining pathway gene expression is shown for hermaphrodites (A) and males (B) Genes shown in large font are active whereas the expression of those in small font is inhibited. X = copies of X chromosomes and A = copies of autosomes.



1.3.5 Evolutionary conservation in sex-determining pathways

Recent work performed in an attempt to understand the molecular basis of the origin of sex-determination has identified related sex-determining genes throughout a wide variety of taxa (Raymond et al., 1998; Raymond et al., 2000; Ottolenghi et al., 2000, 2002; Ottolenghi and McElreavey, 2000). The "DM" genes, so named because they contain a conserved DNA binding motif known as a DM domain, are known to be involved at some level of sex-determination in a wide variety of species. The global sex regulator, *dsx,* in *Drosophila*, *DMRT1* in mammals and *mab-3*, a downstream masculanizing factor in *C. elegans,* are all DM domain-containing genes. The primary genetic sex-determining signal in the medakafish, DMY is also DM gene (Matsuda et al., 2002; Matsuda et al., 2003), and *DMRT1* homologs have been identified on the sex chromosomes of avians (for review see Marshall Graves and Shetty, 2001; Smith and Sinclair, 2004), as well as in the genomes of environmentally sex-determining organisms such as reptiles and amphibians (Torres Maldonado et al., 2002).

The observation that there is evolutionary conservation of downstream elements within sex-determining pathways lends credence to the theory that these pathways may evolve through the process of "retrograde evolution" that is, the successive addition of upstream control elements to an ancient conserved downstream module (Wilkins, 1995; Pomiankowski et al., 2004).

Identification of the primary genetic sex-determining switch of additional organisms is the first step in the functional characterization of their sex-determining pathways. Such characterization remains critical to increasing our comprehension of the evolution of sex-determining mechanisms.

1.4 Sex-Determination in Fishes

1.4.1 Teleost fishes

Teleost fish (class *Teleostei*) are an attractive group of organisms for the study of sex-determination because of the wide variety of sex-determining mechanisms they employ (Potts and Wooton, 1984). With the sole exception of DMY in the medakafish, *O. latipes*, no primary sex-determining genetic signal has been identified within the teleosts (Kondo et al., 2003). Even the primary genetic mechanisms behind sexual differentiation have yet to be elucidated for zebrafish (*Danio rerio*) and pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*) whose genomes have been sequenced (Aparicio et al., 2002; Jaillon et al., 2004).

Teleost fishes are also a particularly suitable group for the study of the evolution of sex chromosomes, as many species of fish that employ genetic sex-determining mechanisms have sex chromosomes that have not yet undergone visible differentiation at the cytogenetic level (Thorgaard, 1983; Iturra et al., 2001; Woram et al., 2003).

1.4.2 Characterization of teleost sex-determining chromosomes

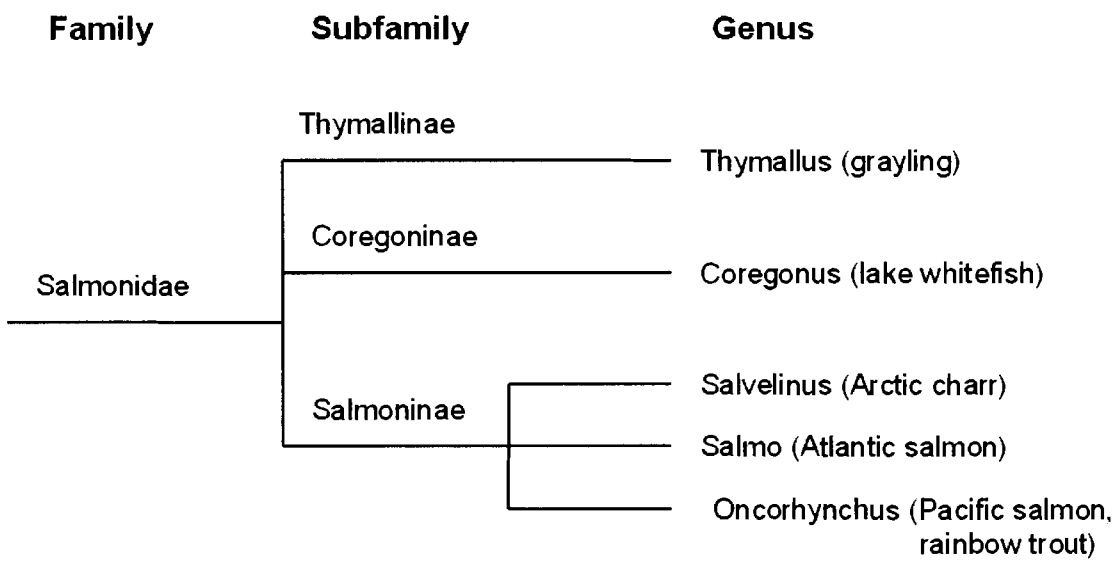
More than 1700 species of fish have been cytogenetically characterized (Arkhipchuk, 1995) of which 176 species (10.4%) have been found to have morphologically distinct sex chromosomes (reviewed in Devlin and Nagahama, 2002).

1.5 Salmonid Fishes

1.5.1 Family Salmonidae

The family Salmonidae is comprised of three sub-families: Coregoninae (whitefishes and ciscos), Thymallinae (graylings), and Salmoninae (trouts, salmons and charrs) (Fig. 1.4) (Nelson, 1984). The sub-family Salmoninae has been particularly well studied as they have long been of interest to commercial and sport fishing, as well as a large aquaculture industry surrounding many of its members, including Atlantic salmon (*Salmo salar*), Arctic charr (*Salvelinus alpinus*) and rainbow trout (*Oncorhynchus mykiss*). Rainbow trout has also been used traditionally as an important fish physiological model organism. Salmonidae are native to the Northern hemisphere; however, they have been introduced to many different areas of the world, as a source of game fish and farming stock. Salmonids pursue anadramous and freshwater life-cycles; nevertheless, spawning always takes place in freshwater. Some species (sockeye salmon, *Oncorhynchus nerka* and chinook salmon, *Oncorhynchus tshawytscha*) die after spawning, whereas others such as (Atlantic salmon, brown trout (*Salmo trutta*), and Arctic charr) are repeat spawners (Nelson, 1984).

Figure 1.4 Phylogenetic tree of Salmonid fishes. The phylogenetic relationship between the three subfamilies of the Salmonidae family is shown with species examples provided for each genus.



It is believed that a genome duplication event occurred at the Teleost divergence, contributing to their successful radiation (Ohno, 1970; Allendorf and Thorgaard, 1984; Hoegg et al., 2004; Christoffels et al., 2004). This is evidenced by the fact that ray-finned fishes (*Actinopterygii* in which the Teleosts are included) have two paralagous copies of many genes, whereas tetrapods have only one (Hoegg et al., 2004).

The common ancestor to the Salmonids is believed to have undergone a separate tetraploidization event between 25 to 120 million years ago (mya). Four major lines of evidence support this hypothesis: (1) Salmonid fishes have approximately twice the amount of DNA per cell as their nearest relatives. (2) Salmonids typically have about 100 chromosome arms, which is twice the amount typically seen in closely related species. (3) Multivalents have commonly been observed in meiotic preparations from several males in salmonid species. (4) Salmonids show a high incidence of duplicated enzyme loci (for review see Allendorf and Thorgaard, 1984; Ohno et al., 1968). Recent work by Mitchell (2004) has placed the Salmonid specific genome duplication event towards the more ancient side of the predicted time, between 80 and 120 mya.

1.5.2 Sex-determination in Salmonids

Salmonids possess a genetic mechanism of sex-determination; however, like most fish, they are very labile during early development and can be directed by exogenous factors such as hormones (Devlin and Nagahama, 2002). Once a particular developmental profile has been selected, the state of gonadal

differentiation may then be stably perpetuated throughout subsequent development.

1.5.2.1 Evidence of male heterogamety in salmonids

Genetic sex-determination employing male heterogamety has long been accepted as the general rule in salmonids, although the sex chromosomes of most species have not yet been identified (reviewed in Phillips and Ráb, 2001). Primary evidence for the presence of an XY sex-determination mechanism is derived from the sex ratios of the progeny of hormonally sex-reversed individuals.

Addition of methyltestosterone (inducing male development) or estradiol-17β (inducing female development) to breeding tanks during early development will cause spontaneous sex reversal (reviewed in Devlin and Nagahama, 2002). It has been observed that the offspring of sex-reversed female salmonids (being phenotypically male while genotypically female) crossed with normal females are all female. Therefore, females must be homogametic, XX, and males heterogametic, XY (Johnstone et al., 1979; Hunter et al., 1982, 1983; Johnstone and Youngson, 1984).

The identification of sex-linked and sex-specific markers has lent support to models of male heterogamety in several species such as Arctic charr (*S. alpinus*), lake trout (*Salvelinus namaycush*), Masu salmon (*Oncorhynchus masou)*, pink salmon (*Oncorhynchus gorbuscha*), chum salmon (*Oncorhynchus keta*), brown trout (*S. trutta*), rainbow trout (*O. mykiss*), coho salmon

(*Oncorhynchus kisutch*) and chinook salmon (*O. Tshawytscha*) (May et al., 1989; Du et al., 1993; Forbes et al., 1994; Prodöhl et al., 1994; Young et al., 1998; Nakayama et al., 1999; Sakamoto et al., 2000; Devlin et al., 1991, 1998, 2001; Zhang et al., 2001; Stein et al., 2002). In each species, marker PCR product length polymorphisms were observed such that males were heterozygous at the allele and females were homozygous. Genetic linkage mapping of such markers placed them within the sex linkage group, suggesting that the polymorphisms observed are differences between alleles located on the Y and X chromosomes, respectively (Young et al., 1998; Sakamoto et al., 2000; Devlin et al., 2001).

1.5.2.2 Salmonid sex chromosomes

Within Salmonid species, morphologically distinct sex chromosomes are not generally observable. Nevertheless, heteromorphic sex chromosomes have been identified in Least cisco (*Coregonus sardinella*), rainbow trout (*O. mykiss*), sockeye salmon (*O. nerka*), and lake trout (*S. namaycush*) (reviewed in Phillips and Ráb, 2001).

The formation of heteromorphic sex chromosomes often involves heterochromatin addition to the X chromosome (Phillips and Ráb, 2001). This was observed in lake trout and brook trout (*Salvelinus fontinalis*) as the largest pair of submetacentrics have inherited an X-specific heterochromatin block at the end of their short arms, identifying them as the sex chromosome pair (Phillips and Ihssen, 1985; Phillips et al., 2002). Size-differences have been observed between the sex chromosomes among certain hatchery populations of rainbow trout (*O. mykiss*) suggesting that these populations represent an early stage in

visible sex chromosome differentiation (Thorgaard, 1977). Interestingly, rainbow trout populations lacking any heteromorphisms have also been observed, indicating that chromosomal rearrangements differentiating the sex chromosomes of this species are still in the process of fixation (Thorgaard, 1983).

Fluorescent *in situ* hybridization (FISH) probes designed from sex-specific markers identified in several salmonid species have greatly facilitated the identification of the sex chromosome pair within each of their karyotypes (Reed et al., 1995; Moran et al., 1996; Iturra et al., 1998, 2001; Phillips, 2001; Phillips et al., 2001, 2002; Stein et al., 2001.) Interestingly, when these markers were used as hybridization probes across species, they hybridized to one or more pairs of autosomes, leading to the conclusion that sex-chromosomes differ among species within the salmonids (Phillips et al., 2001)

Overall, current cytogenetic data support the assumption that salmonids, like many other fishes, represent early stages in sex chromosome differentiation (Phillips et al., 2001). Consistent with this assumption is the observation that YY males are viable and fertile (Hunter et al., 1982; Chevassus, 1988; Onozato, 1989; Devlin et al., 2001), suggesting that the Y and X chromosomes still retain similar gene compliments.

1.5.2.3 Sex-linkage in Salmonids

Microsatellite linkage maps have been constructed for a number of economically important salmonid species including rainbow trout (*O. mykiss*) (Sakamoto et al., 2000), brown trout (*S. trutta*) (Gharbi, 2001), Atlantic salmon

(*S. salar*) (Gilbey et al., 2004; Moen et al. 2004; R. Danzmann, unpublished results; B. Hoyheim, unpublished results) and Arctic charr (*S. alpinus*) (Woram et al., 2004). Linkage data indicate that there is a lack of conservation among the phenotypic sex-determining loci (thereafter denoted as *SEX*) among salmonid species (Woram et al., 2003).

This was first evidenced by a study conducted within the genus *Salvelinus* (May et al., 1989), which found that sex-linked allozyme markers in Arctic charr (*S. alpinus*) were not linked to *SEX* in either brook trout (*S. fontinalis*) or lake trout (*S. namaycush*). It was also shown in the genus *Oncorhynchus* that a growth hormone marker, known to be linked to *SEX* in coho salmon (*O. kisutch*), chinook salmon (*O. tshawytscha*) and Masu salmon (*O. masou*) was not sex-linked in amago salmon (*Oncorhynchus rhodurus*) or rainbow trout (*O. mykiss*) (Forbes et al., 1994; Nakayama et al., 1999; Zhang et al., 2001). In the genus *Salmo*, a minisatellite locus shown to be in tight association with *SEX* in brown trout (*S. trutta*) mapped to an autosomal pair in Atlantic salmon (*S. salar*) (Taggart et al., 1995).

A recent study by Woram et al. (2003) compared the sex linkage groups of Arctic charr (*S. alpinus*), brown trout (*S. trutta*), Atlantic salmon (*S. salar*) and rainbow trout (*O. mykiss*). It was observed that the position of the sex-determining locus was not conserved with respect to synteny among sex-linkage groups between species. In fact, sex-linked microsatellite markers able to amplify across species, were found to be in autosomal, homologous linkage groups in other species. The genetic linkage maps indicate that *SEX* is located at the end

of the sex linkage groups of Atlantic salmon, brown trout and Arctic charr. Although it is known that the distal ends of linkage groups are not necessarily coincidental with the telomeric ends of chromosomes, large recombination distances between *SEX* and its associated markers is suggestive of terminal placement. Terminal placement is also supported by gene centromere mapping by gynogenesis in chinook salmon (Devlin et al., 2001) as well as the observation that molecular markers tightly associated with SEX in this species, hybridize to a near-terminal position (Stein et al., 2001). In the case of rainbow trout, *SEX* is not found on the distal end of the linkage group; rather, it has been mapped close to putative centromeric markers, suggesting that the sex chromosome in rainbow trout was created through translocation of an ancestral sex chromosomal segment to an autosome (Thorgaard, 1977; Woram et al., 2003). The assumed telomeric placement of *SEX* in Arctic charr, brown trout and Atlantic salmon is consistent with an earlier hypothesis that the sex-determining locus has been transposing within the genomes of some salmonids without relocating adjacent markers, thus causing disruption of sex-linkage among species (Phillips et al., 2001).

1.6 Atlantic salmon (*Salmo salar*)

1.6.1 The Atlantic salmon karyotype

Like other salmonids, Atlantic salmon shows extensive interindividual chromosome polymorphism resulting from Robertsonian translocations. As well, some intraindividual polymorphism is observed around a constant number of chromosome arms (NF) (Hartley and Horne, 1984). In a study of Atlantic salmon

caught in Scotland, Hartley and Horne (1984) found three different diploid number of chromosomes in *S. salar* individuals: 56, 57 and 58 around a constant NF of 74. The 2n=58 karyotype was the most common. Also, 5 out of 30 individuals' karyotypes showed intraindividual chromosome polymorphism around NF=74. Comparison of C-banding patterns (which stains heterochromatin) in Atlantic salmon, brown trout and rainbow trout, showed that Atlantic salmon has more heterochromatin than either of the other two species. It was also observed that Atlantic salmon possesses less chromosome arms (NF=74) than either brown trout (NF=102) or rainbow trout (NF=104) and that it has more acrocentric chromosomes than metatcentric chromosomes. This suggests that Atlantic salmon chromosomes have undergone a larger number of inversions relative to other salmonid species, from these three genera, during the process of re-diploidization after the ancestral salmonid genome duplication.

1.6.2 Genetic linkage mapping in Atlantic salmon

As previously indicated, microsatellite based (R. Danzmann, unpublished results; B. Hoyheim, unpublished results; Gilbey et al., 2004) and amplified fragment length polymorphism (AFLP) (Moen et al., 2004) genetic linkage maps have been generated for Atlantic salmon. As in the case of previous linkage maps generated for other salmonids, a large difference in recombination rates between males and females was observed. Gilbey et al. (2004) found that female recombination rates were in the order of 3.92 fold higher than that of males: the largest sex-specific recombination rate difference reported for a vertebrate species. It has been previously suggested that the large sex-specific difference in

recombination rates is due to suppression of recombination in males due to the structural constraints imposed in crossing over during multivalent pairing (for review see Allendorf and Thorgaard, 1984; Sakamoto et al., 2000; Nichols et al., 2003).

1.6.3 Sex-determination in Atlantic salmon

Nothing is known about the molecular basis of sex-determination in Atlantic salmon, and the sex chromosome pair has not been identified. As previously indicated, *SEX* appears to be located on the telomere of the chromosome equivalent to linkage group 1 (ASL1) (R. Danzmann unpublished results; B. Hoyheim unpublished results). In addition, no sex-specific markers have been identified in Atlantic salmon. Despite an extensive bulked segregant analysis performed by McGowan and Davidson (1998) of 1152 random amplified polymorphic DNAs (RAPDs), none were found to be tightly associated with *SEX*. These results further support the theory that the homologous chromosomes on which the sex-determining factor is found, are not extensively differentiated in Atlantic salmon.

1.6.4 Genomics Research on Atlantic Salmon Project (GRASP) resources

The Genomics Research on Atlantic salmon Project (GRASP) is a major initiative bringing together various universities and research institutions on an international scale, in order to better understand the genome of Atlantic salmon. As part of the overall GRASP project a number of resources have been developed, based on existing genome technologies.

1.6.4.1 Construction of a large-insert BAC genomic library

The CHORI-214 Atlantic salmon BAC library (Thorsen et al., submitted; http://bacpac.chori.org/salmon214.htm) was constructed in three segments using DNA from a single male Atlantic salmon from a Norwegian aquaculture strain. DNA was extracted from sperm, partially digested with *Eco*RI, size fractionated, and then ligated into pTARBAC2.1 (Zeng et al., 2001; Osoegawa et al., 2004). Approximately 298,820 clones were produced and arrayed into 816 384 well micro-titer plates and also gridded onto seventeen 22x22 cm nylon high-density filters for screening by probe hybridization. Each hybridization membrane represents over 18,000 distinct salmon BAC clones, stamped in duplicate. The library's three segments comprise plates 1 – 288 (segment 1), plates 289 – 576 (segment 2), and finally plates 577 – 816 (segment 3). Thorsen et al. (submitted) estimated the average insert size of 658 random clones (249 from segment 1, 218 from segment 2, and 191 from segment 3) by digestion with *Not*I followed by pulse-field gel electrophoresis (Osoegawa et al., 1998). The average insert sizes of the library's three segments were 189 kb, 190 kb, and 186 kb for segments 1, 2, and 3, respectively. The percentages of empty wells for each of segments 1, 2, and 3 are 1.79, 2.56 and 3.23 %, respectively. The percentages of non-insert clones in each of the library's three segments are 1.2, 2.8, and 2.5 %, respectively. The haploid C-value for Atlantic salmon has recently been estimated as 3.27 pg (Hardie et al., 2002; Hardie and Hebert, 2003) which agrees well with the 3.10 pg that is listed in the Genome Size Database (http://www.genomesize.com/fish.htm). This translates into a genome size of

approximately 3 x 10^9 bp and therefore, this BAC library represents an 18.8 fold coverage of the Atlantic salmon genome.

1.6.4.2 Generation of a Hind III fingerprinted physical map

Hind III fingerprinting was performed on the first 200,640 clones of the CHORI-214 library in order to generate a physical map of the Atlantic salmon genome (Ng et al., submitted). 185,938 clones gave informative fingerprints after Hind III digestion and were used in the assembly of the library using FPC version 7.0 (Soderland et al., 1997). The average insert size calculated from the addition of fingerprint bands is 186,000 base pairs, therefore providing an 11.5 fold coverage of the Atlantic salmon genome. The physical BAC fingerprint map for Atlantic salmon has been made publicly available using the internet Contig Explorer (iCE) version 3.4 (Fjell et al., 2003; available at http://ice.bcgsc.ca/) which is a web-based application that allows viewing of all the physical map data.

1.7 Approach to Identify the *SEX* in Atlantic salmon

1.7.1 Previous approaches to characterization of *SEX* in other species

Previous approaches used to identify *SEX* in other species were successful due to the integration of genetic linkage maps and large-insert genomic libraries (Beye et al., 2003; Matsuda et al., 2002). In the honeybee (*A. mellifera*), *SEX* was mapped between a RAPD based genetic marker and one obtained by bulk segregation analysis using multilocus fingerprinting. Probes were designed to the flanking regions of these markers and used to probe a 5 fold genomic coverage BAC library and a 20 fold genomic coverage cosmid

library. Probes were designed to the end sequences of these clones and used to perform chromosome walking towards the sex-determining factor. This technique, coupled with positional cloning of *SEX* allowed Beye et al. (2003) to narrow down the region containing *SEX* to a 24 kb segment. This segment was shotgun sequenced to a 12 fold coverage, and it was discovered that the gene *csd* is the primary genetic signal of the sex-determining pathway in the honeybee.

A similar approach was taken in the discovery of *DMY*, the primary genetic signal in the sex-determining pathway of medakafish (Matsuda et al., 2002), in which chromosome walking within a large-insert BAC library was also used to identify a BAC clone containing the *SEX* locus and identify it.

Recently, Peichel et al. (2004) integrated a BAC genomic library with a microsatellite genetic linkage map of the threespine stickleback (*Gasterosteus aculeatus*), a fish which, as in the case of Atlantic salmon, does not demonstrate morphologically distinct sex chromosomes. *G. aculeatus* like Atlantic salmon possesses an XY system of sex-determination. A previous study had noted a sex-specific polymorphism in the isocitrate dehydrogenase gene (*IDH*) (Withler et al., 1985). When this gene was incorporated onto the *G. aculeatus* linkage map, it was found to map to the *SEX*-containing linkage group. Clones positive for *IDH* were identified within the BAC genomic library and were segregated as being from the X or Y chromosome based on the sex-specific polymorphisms. Whole shotgun sequencing of four BACs covering homologous segments of the X and Y chromosomes revealed that, though the chromosomes appear homomorphic at

the cytogenetic level, at the sequence level there is a significant amount of degeneration. This is evidenced by the large accumulation of repetitive sequence accumulation on the Y chromosome, a feature commonly associated with the initial formation of distinct, heteromorphic sex chromosomes.

1.8 Aim of the Thesis

The purpose of this thesis is to integrate the *SEX*-containing microsatellite linkage group ASL1 with the Hind III fingerprinted physical map and the Atlantic salmon karyotype in order to characterize the sex-determining chromosome and identify the sex-determining region of the genome.

CHAPTER 2 MATERIALS AND METHODS

2.1 Integration of Atlantic Salmon Genomic and Physical Maps

2.1.1 BAC Library Screening

Sixteen 40mer oligonucleotide probes, as well as 20mer reverse complement PCR primers were designed to the flanking regions of sex-linked microsatellite markers using cprimer software (http://iubio.bio.indiana.edu/soft/molbio/ mac/cprimer.hqx). The 40mers and a *C. briggsae* 40-mer overgo reference probe (5'- GTTGCCAAATTCCGAGATCTTGG CGACGAAGCCACATGAT-3') were each labeled at the 5'end with ^{32}P at 37°C in 10 µl volume consisting of 1 µl (0.37MBq) $\gamma^{32}P$-ATP (Perkin Elmer), 10 U T4 oligonucleotide kinase (Invitrogen), 2 µl of Invitrogen 5X Forward Reaction buffer (350 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 500 mM KCl and 5 mM 2-mercaptoethanol) and 5ml of dH_2O. Each high-density BAC filter (http://bacpac.chori.org/salmon214.htm) was pre-hybridized at 65°C for two hours in 50 ml of 5X SSC, 20 mM Na_2HPO_4 (pH 7.2), 7% SDS and 1X Denhardt's solution. To each pre-hybridized filter, 1.6 µl of radioactively labeled probe and 1.6 µl of radioactively labeled *C. briggsae* overgo probe were added and allowed to hybridize at 65°C overnight (~16 hours). Filters were then washed twice at 50°C for one hour in 100 ml of 3X SSC, 10X Denhardt's solution, 5% SDS and 25mM NaH_2PO_4 (pH 7.5) followed by a single one hour wash in 100 ml

of 1X SSC and 1% SDS. Filters were then wrapped in plastic film and allowed to expose on storage phosphor-screens (Molecular Dynamics) overnight (~16 hours). Hybridization signals were then detected using a STORM 820 Phosphorimager (Amersham Biosciences).

2.1.2 Positive Confirmation by PCR

BAC clones positive by hybridization were picked from the CHORI-214 Atlantic salmon genomic library and grown in 5 ml of 2X YT enriched growth medium (Sambrook et al., 1989) containing 20 μg/ml chloramphenicol for 16 hours at 37°C. DNA was isolated from 3 ml of each culture using the Perfectprep© 96-well vacuum manifold BAC prepping kit (Brinkmann Instruments). Each BAC that was positive by hybridization for a particular microsatellite's flanking region probe was tested for the presence of that microsatellite by PCR using appropriate primers and annealing temperatures (Table 3.1). PCR amplifications were performed in a T3 Thermocycler (Biometra) in 25 μl reaction volumes containing between 10 and 20 ng of BAC DNA, 1X PCR buffer (1.5mM $MgCl_2$, Amersham), 50 nM dNTP (Amersham), 0.05 U of Taq DNA polymerase (Invitrogen) and 0.4 μM forward and reverse PCR primers. The PCR temperature profile consisted of an initial denaturation step of 95°C for 5 min; 34 cycles of 95°C for 30 sec, 1 min at specific annealing temperature, and 72°C for 1 min; and a final extension step of 72°C for 2 min. PCR products were separated on a 2.5% agarose gel containing 1X TBE and 0.5 ug/ml ethidium

bromide. The DNA fragments were visualized using a UV trans-illuminator (Ultra-Violet Products).

2.2 Physical Map *in silico* Analysis and Extension

BAC clones containing sex-linked microsatellite markers, confirmed by both hybridization and PCR were then associated with contigs in the Atlantic Salmon genomic physical map (Ng et al., submitted), using FPC ver. 7.0 (Soderlund et al., 1997, http://www.genome.arizona.edu/fpc/) and iCE ver. 3.4 (Fjell et al., 2003, http://ice.bcgsc.ca) software. End-terminal clones from sex-linked microsatellite marker positive BAC contigs and positive singletons were end-to-end joined against the entire physical map using FPC ver. 7.0 with the following parameters: Tolerance = 7, Cutoff = 1e-09, Use CpM off, Log off and stdout on.

2.3 BAC-end Sequencing and Novel Gene Identification

Bacterial colonies were picked manually from the library plates and re-arrayed into a 96-well, deep well culture block (Beckman) containing 1.2 ml of 2xYT medium containing 12.5 ug/ml chloramphenicol. The plates were sealed with AirPore™ tape (Qiagen) and incubated overnight in a shaking incubator at 37°C for 16h, following which the bacterial cells were collected by centrifugation. DNA was isolated as described in Schein et al., (2004) and suspended in 60 µl of 1X TE. DNA Sequencing reactions were assembled in 384-well clear optical reaction plates (Applied Biosystems) using a Tango workstation (Robbins Scientific). In each 8 µl reaction (total volume) the following were added: 5 µl of

purified BAC DNA (~45 ng/μl), 0.7 μl of sequencing primer (5 pmol/μl, Invitrogen), 0.3 μl of Ultrapure water (Gibco), and 2.0 μl of BigDye v.3.1 ready reaction mix (Applied Biosystems). Sequence data were obtained using the T7 primer (5'-TAA TACGACTCACTATAGGG-3') and the pTARBAC13 P2 primer (5'-TCCCGAATTG ACTAGTGGGTA-3'). Thermal cycling was performed on PTC-225 thermal cyclers (MJ Research) with parameters of 85 cycles at 96° C for 10 seconds, 48° C for 5 seconds using T7 primer or 54° C for 5 seconds using pTARBAC13P2, 60° C for 4 minutes, followed by incubation at 4° C. Reaction products were precipitated by adding 40 μl of 75% isopropanol per well followed by centrifugation at 2750 x g for 30 minutes in an Eppendorf 5810R centrifuge. The isopropanol was immediately decanted and reaction products washed with 70% ethanol. The 384-well cycle plates were allowed to dry inverted for 15 minutes. Samples were suspended in 10 μl of Ultrapure water and analyzed using a 3700 DNA analyzer (Applied Biosystems). Sequence data were evaluated using PHRED software (Ewing et al., 1998) and BLASTn aligned (Altschul et al., 1990) against the non-redundant GenBank database at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/entrez/). BLAST likelihood values of E-0.05 or lower were considered significant. BLAST hits to genes were only considered significant if they met the above criteria and also aligned to known or predicted exons within a gene sequence. All BAC-end sequences were submitted to GenBank and can be found under accession numbers CW883115 to CW883293.

2.4 Chromosome "Walking"

2.4.1 Probe design

Hybridization oligonucleotide probes and primers were designed from BAC end sequences using the Primer premier version 5.0 software (http://www.premierbiosoft.com/primerdesign/index.html) using the software's automatic search function across the entire vector-trimmed BAC-end sequence, generating complementary primers of 20 +/- 2 nucleotides in length and producing a PCR fragment of between 150 and 500 nucleotides in length. Under the "Automatic Search Parameters" menu, all options were selected with the exception of "false priming". The best software designed primer combination was considered suitable if it had no reported possible false priming sites. Whenever suitable primers were designed, one of the two primers (forward or reverse) was manually extended to approximately 40 nucleotides and its annealing temperature (T_m) was evaluated using the software. In cases where one of the two primers could be extend such that it generated a hybridization probe with a Tm above 70°C, without generating potential false priming sites, the two primers were considered suitable and used for screening of the CHORI-214 Atlantic salmon genomic BAC library. Probe/primer combinations were designed from 45 contig and super-contig-terminal representative BAC-ends in this fashion. Primers were ordered in 96-well plate format from Illumina Inc. (http://www.illumina.com).

2.4.2 Confirmation of primer viability by PCR

Five µl of 25 % glycerized freezer stocks of the BAC clones from which each primer/probe set was diluted with 195 µl of distilled H_20 and used as template DNA for PCR reactions using the primer/probe sets. PCR amplifications were performed in a TGradient temperature gradient thermocycler (Biometra) in 25 µl reaction volumes containing 10 µl of BAC DNA template, 1X PCR buffer (1.5 mM $MgCl_2$, Amersham), 50 nM dNTP (Amersham), 0.05 U of Taq DNA polymerase (Invitrogen) and 0.4 µM forward and reverse PCR primers. The PCR temperature profile consisted of an initial denaturation step of 95°C for 5 min; 34 cycles of 95°C for 30 sec, 1 min at a gradient of annealing temperatures, and 72°C for 1 min; and a final extension step of 72°C for 2 min. Six separate PCR reactions were performed for each primer set such that each reaction used a separate annealing temperature (42.5°C, 44.8°C, 50.5°C, 56.5°C, 62.4°C and 66.9°C). PCR products were separated on a 1.0% agarose gel containing 1X TBE and 0.5 ug/ml ethidium bromide. The DNA fragments were visualized using a UV trans-illuminator (Ultra-Violet Products). The 11 primer sets that did not amplify fragments at any of the annealing temperatures were rejected and not used in subsequent analysis. Optimal annealing temperature for each primer set was determined empirically from the gradient of annealing temperatures.

2.4.3 BAC library screening

The approximately 40 nucleotide hybridization probe of each of the probe/primer sets was used to screen the CHORI-214 Atlantic salmon genomic

BAC library. The probes and a *C. briggsae* 40-mer overgo reference probe (5'-GTTGCCAAATTCCGAGATCTTGGCGACGAAGCCACATGAT-3') were each labeled at the 5' end with ^{32}P at 37°C in 10 µl volume consisting of 1 µl (0.37MBq) $\gamma^{32}P$-ATP (Perkin Elmer), 10 U T4 oligonucleotide kinase (Invitrogen), 2 µl of Invitrogen 5X Forward Reaction buffer (350 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 500 mM KCl and 5 mM 2-mercaptoethanol) and 5 ml of dH_2O. High-density BAC filters (http://bacpac.chori.org/salmon214.htm) were pre-hybridized in groups of three at 65°C for one hour in 100 ml of 5X SSC, 0.5% SDS and 10X Denhardt's solution. To each pre-hybridized filter, 2.5 µl of radioactively labeled probe and 5.0 µl of radioactively labeled *C. briggsae* overgo probe were added and allowed to hybridize at 65°C overnight (~16 hours). Filters were then washed three times at 50°C for one hour in 150 ml of 1X SSC and 0.5% SDS. Filters were then wrapped in plastic film and allowed to expose on storage phosphor-screens (Molecular Dynamics) overnight (~16 hours). Hybridization signals were then detected using a STORM 820 Phosphorimager (Amersham Biosciences).

2.4.4 Analysis of clone S0503M23

BAC clones S0069C07, S0070C22, S0100N19, S0113N21, S0191I09 and S0503M23 were picked from the CHORI-214 Atlantic salmon genomic library and cultured overnight (~16 hours) in LB broth containing 12.5 µg/ml chloramphenicol. Twenty-five % glycerized freezer stocks were made from each overnight culture. Five µl of glycerized freezer stock was diluted in 195 µl of H_20 for use as a PCR template. PCR was performed on each of the clones using both

the S0503M23T7 and S0503M23SP6 primers in 25 μl reaction volumes containing 10 μl of BAC DNA template, 1X PCR buffer (1.5 mM $MgCl_2$, Amersham), 50 nM dNTP (Amersham), 0.05 U of Taq DNA polymerase (Invitrogen) and 0.4 μM forward and reverse PCR primers. The PCR temperature profile consisted of an initial denaturation step of 95°C for 5 min; 34 cycles of 95°C for 30 sec, 45 sec at annealing temperatures (56.5°C for the S0503M23T7 primers, and 50.5°C for the S0503M23SP6 primer set), 72°C for 1 min; and a final extension step of 72°C for 5 min. PCR products were separated on a 1.0% agarose gel containing 1X TBE and 0.5 ug/ml ethidium bromide. The DNA fragments were visualized using a UV trans-illuminator (Ultra-Violet Products).

2.5 Fluorescent *in situ* Hybridization

Blood was cultured from the Norwegian strain of Atlantic salmon using standard methods (Reed and Phillips, 1995). DNA was isolated from five BAC clones using the Qiagen Midi-Preparation kit. BAC clones were labeled with Spectrum Orange using a nick translation kit (Vysis, Inc.). Human placental DNA (2 μg) and Cot-1 DNA (1 μg, prepared from Atlantic salmon) were added to the probe mixture for blocking. Hybridizations were carried out at 37^0C overnight and post-hybridization washes were as recommended by the manufacturer (Vysis, Inc.) with minor modifications (Phillips and Reed, 2000). Antibodies to Spectrum Orange (Molecular Probes) were used to amplify the signal in some cases. Slides were counter-stained with 4,6-diamidino-2-phenylindole (DAPI) at a concentration of 125 ng DAPI in 1 ml antifade solution. Images were captured

with a Sensys camera and analyzed with Cytovision Genus (Applied Imaging, Inc.) software. Chromosomes were arranged according to size within the meta-centric/submeta-centric (pairs 1-8) and acro-centric (pairs 9-29) groups. Chromosome 2 can be distinguished on the basis of size and presence of a large DAPI bright block of repetitive DNA at the telomere of the long arm.

2.6 Sequencing and Analysis of Clone S0493K22

2.6.1 Shotgun library construction

Clone S0493K22 was prepped using Quiagen's Large-Construct Kit. Approximately 15 µg of isolated BAC DNA was sheared by sonication for each sample. Sonication efficiency was verified by agarose gel electrophoresis. Sonicated DNA was end repaired using the Epicentre End-It DNA End-Repair Kit. All end-repair reactions were carried out at room temperature in a 50 µL final volume containing 5 µL each of dNTP mix, ATP, and 10X buffer from the kit, as well as 1 µL of End-Repair Enzyme Mix and 34 µL of DNA, and were incubated for 45 minutes. The reactions were subsequently terminated by heating the samples at 70°C for 10 minutes.

Sonicated, end-repaired BAC DNA was size fractionated by gel electrophoresis, using a 1 % agarose gel made with 1X TAE. Upon completion of electrophoresis, the lane containing 1 Kb standard ladder (Invitrogen) was cut off from the rest of the gel and allowed to stain in a 0.5 µg/ml ethidium bromide solution for 45 minutes. The ladder was then visualized under ultraviolet light, and notches were made in the gel at the 2 and 5 Kb standard sizes. The

remainder of the original gel was then compared to the notched ladder portion, and DNA between 2 and 5 kb was cut out from the gel and subsequently extracted using the Qiagen Gel Extraction Kit. DNA concentration was then measured by spectrophotometry and the samples were diluted to 25 ng/µL.

The 2 to 5 Kb size fractionated portion of DNA was then ligated into pUC19 vector. Positive (standard insert DNA) as well as negative (no-insert) controls were also performed. The ligation reactions were carried out using T4 DNA Ligase from Invitrogen in 20 µL reaction volumes containing 10 U of ligase, 100 ng of insert DNA, 20 ng of *Sma*I digested, phosphatase-treated pUC19 vector, and 4 µL of 5X reaction buffer. The samples were incubated overnight (approximately 18 hours) at 14°C, and then stored at -20°C until they were used for transformation reactions.

The ligation reactions were used to transform 100 µL of XL-1 Blue Super-Competent Cells (Stratagene) using a heat shock method. The cells were first allowed to thaw while sitting on ice. 4 µL of beta-mercaptoethanol was added to each tube, and incubated for 10 minutes on ice. 2.5 µL of the ligation reaction was mixed with 100 µL of super-competent cells and then incubated on ice for 30 minutes. The cells were then heat shocked for 45 seconds at 42°C (with vigorous swirling for the first ten seconds) and then incubated on ice for 2 minutes. After the addition of 900 µL of SOC medium, an outgrowth step was performed by shaking at 250 rpm for 1 hour at 37°C. Finally, 250 µL of each sample was plated onto an LB agar plate containing 200 µg/mL ampicillin. 375 µL of X-gal (20mg/ml) and 225 µL of isopropyl-beta-D-thiogalactopyranoside (IPTG)

(200mg/ml) were also transferred onto the plates, and the resulting mixture was spread evenly on the plate surface. The plates were then incubated for overnight (~20 hours at 37°C).

2.6.2 Assay for the presence of *E. coli* genomic contamination

Twenty-four random colonies were picked from those plated from the transformation and cultured overnight (~20 hours) in 5 ml of LB broth containing 200 μg/ml ampicillin. Three ml of each tube were then prepped using a mini-prep kit (Quiagen). The isolated plasmid DNA was then used for insert sequencing using the dideoxy terminator method. 300 ng of plasmid DNA was mixed with 5pmol of M13 reverse primer, and 4 μL of sequencing premix (Amersham) in a final volume of 10 μL. Sequencing reactions were carried out with the following thermal profile: 30 cycles of 94°C for 20 sec, 50°C for 20sec and 60°C for 1min. Following amplification, unincorporated dideoxynucleotides were removed from each reaction mixture by ethanol precipitation, and the reactions were resuspended in 2 μL of diformamide loading dye. Sequence analysis was carried out on an ABI 377 DNA Sequencer (Applied Biosystems). The resulting sequences were then subjected to BLASTn searches of the non-redundant nucleotide database at Genbank in order to determine the amount of *E. coli* genomic contamination.

2.6.3 Sequencing of shotgun library

A total of 1544 sequence reads were obtained from random clones picked from the plated library. Sequencing was performed at the Genome British

Columbia sequencing platform at the University of Victoria. Sequence reads were analyzed using the Phred/Phrap/Consed suite of programs. Crossmatch was run to remove sequence reads containing *E.coli* genomic DNA as well as the vector sequence (pTARBAC2.1).

2.6.4 Gene identification

Two super-contigs were predicted by Phrap (contig 19-13 containing 5 smaller contigs and contig 16-17 containing 3 smaller contigs). An arbitrary number (100) unknown nucleotides (N) were placed between each of the contig-contig joins within each super-contig predicted by Phrap, as the actual nucleotide distances between contigs is unknown. The two super-contigs were searched for predicted genes using Genscan (Burge and Karlin, 1997; http://genes.mit.edu/GENSCAN.html) as well as tBLASTx aligned against the non-redundant GenBank database and the Atlantic salmon and rainbow trout Gene Indices databases at TIGR (http://www.tigr.org/tdb/tgi/). The Genscan gene predictions (Five in contig 19-13 and three in contig 16-17) were then also tBLASTn aligned against both the non-redundant and TIGR Gene Indices databases. The results of both rounds of BLAST database alignments were compared. The tBLASTn alignments using clone sequence did not find any additional genes not predicted by Genscan and thus only the Genscan gene predictions were used for further analysis.

One Genscan gene prediction was identified as a Tc-1 like Transposase and thus removed from further analysis. Of the remaining seven gene predictions only those with an arbitrary e-15 or lower BLAST expect value in both the

Genbank non-redundant database and the TIGR Gene Indices were counted as valid predictions and annotated. The best BLAST alignments (lowest expect value) were used to search the Ensembl Genome Browser (www.ensembl.org) in order to determine potential gene function and assign gene chromosome positions between "best hits" across organisms.

2.6.5 Repetitive element analysis

Both super-contigs (contigs 19-13 and 16-17) were run through Repeat Masker software version 3.0 (www.repeatmasker.org; Smit et al., 1996-2004) using the Salmonid specific repeat database developed by Ng et al. (unpublished results). The masked sequences were then run again through Repeat Masker using the *Takifugu rubripes*, *Danio rerio* and Other Fish (non-pufferfish, non-zebrafish) specific databases in a stepwise fashion, excluding previously masked sequences from subsequent searches. The total amount of masked bases were added and used to determine the total percentage of repetitive content of clone S0493K22.

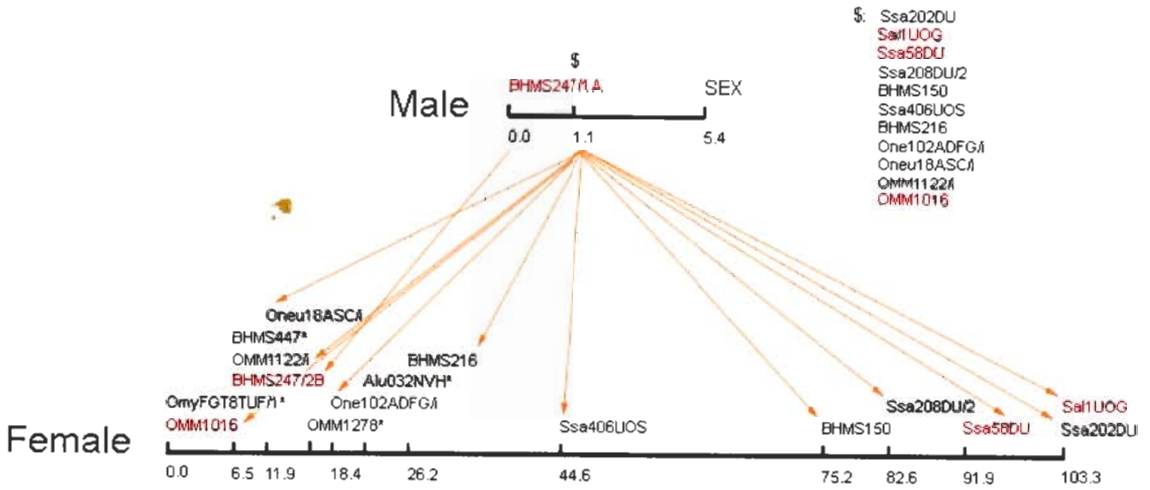
CHAPTER 3 RESULTS

3.1 Integration of Atlantic Salmon Genetic and Physical Maps

3.1.1 Atlantic salmon *SEX*-Associated Genetic Linkage Map

I took as my starting point microsatellite genetic linkage maps that had been generated in part by the SALMAP consortium (R. Danzmann and B. Hoyheim, unpublished results; Woram et al., 2003). Separate male and female linkage maps have been generated as female salmonids possess a much greater recombination rate than males (Sakamoto et al., 2000; Woram et al., 2004). The linkage maps contain an average of 290 microsatellite markers covering 31 - 33 linkage groups (linkage maps are available at http://grasp.mbb.sfu.ca). Atlantic salmon linkage group 1 (ASL1) has been designated as the linkage group containing *SEX*, based on its association with the male phenotype. Composite maps of both male and female ASL1 were generated by the integration of all sex-linked markers for which sequences could be obtained (Fig. 3.1). Placement of markers upon the composite map was based on comparison of recombination distances of shared microsatellites between integrated maps. Distances shown are drawn from the map with the most markers.

Figure 3.1 Atlantic Salmon Linkage Group 1 (ASL1). Relative positions of microsatellite markers incorporated from the SALMAP linkage maps (R. Danzmann and B. Hoyheim, unpublished results) are shown for both male and female versions of ASL1. Markers are shown with associated recombination distances in centiMorgans based on the map containing the most microsatellite markers. Markers without recombination distances have been incorporated from additional maps (as indicated by *). Markers shown in black have been integrated with the CHORI-214 Atlantic salmon BAC genomic library. Probes designed for markers shown in red were not represented in the BAC library. Arrows indicate positions of microsatellite markers on the female linkage map, relative to the male linkage map.



3.1.1.1 Designing of Microsatellite-Specific Hybridization Probes/PCR Primers

Oligonucleotide primers were designed to the flanking regions of sixteen microsatellite markers such that each pair consisted of a primer of approximately 40 nucleotides, which could be used as a probe for Southern hybridization, and a complimentary primer for use in PCR. Microsatellite primer sequences and hybridization/annealing conditions are shown in Table 3.1.

Table 3.1 Microsatellite markers used in the integration of the genetic linkage map with the genomic BAC library. (F-) Forward primer, (R-) Reverse primer. In cases where there is no probe listed, the approximately 40 nucleotide forward or reverse primer was used as the probe.

Marker	Name	Primers	Annealing Temp (°C)	GenBank accession	Positive Contig(s) or Clones	Duplicate Locus
1	Ssa202DU	Probe-TAGAATATGGCTTAGACACAAATCAGGGAACTTGACC F-CTTGGAATATCTAGAATATGGC R-TTCATGTGTTAATGTTGCGTG	51.4	U43695	S0581B12, S0605H01 S0612P18	No
2	Alu032NVH	F-GCCCAGTGCGGGAAAATTTGGAGACTACCTCTGCGATGCG R-ATCCTCAATGTTCTTCAATGGAA	56.5	AF271435	ctg526	No
3	BHMS150	Probe-CATCCCCACAAGGACTGTAAAACCAAAACCCACACAC F-GTTTTGCTTCTCGCTCTCTT R-TTACTTTACCTATTTTTGGCAGG	56.5	AF256699	ctg4421	No
4	Oneu18ASC	F-GTTCATCGTCCAGACTCCCAG R-AGCACTGTCATTAGGGTTATTTTTAGAGACTCCCCAGATA	51.4	U56718	S0017N03, S0119E21	Yes
5	Ssa208DU	Probe-GACTTTACCTTGACCTTCTCTGATACCTTGTTGTAG F-AAACCTGTGAGCTGGAACA R-CCACAACAGAGTATCAGCTG	51.4	AF019161	ctg2133, ctg5898	LG 6
6	Ssa406UOS	F-TATTGTATGGTTCTTTTTATTGTTGTTACTGATTGTCCTGTTGAC R-AGTTGGTTGTTGTGCTGCCG	51.4	AJ402723	S0068C06, S0116A05	No
7	BHMS216	Probe-AAGAATGGAAAGCAGGGAAAAGGTATGAAGAGAGAAC F-TATGAAGAGAGAACCCTTGAACA R-AGTTACACATTACTTGGCAGGAC	50.5	AF256728	ctg1812	No
8	BHMS447	F-CTGAGCAGGTCGCAAGTCTT R-TGAACTCTGCTTCTTCTCAACCAGGAATGTAAACAGGGCT	55.9	AF256856	ctg6488	LG 12
9	BHMS247	F-TGGCAACAGGGTATCCCACCGAGGCGCCCAGCGAGC R-TTAGCTGATTGTATTGTGAGTATGAG	55.9	AF256684	No Positives	No
10	One102ADFG	Probe-ATTTGGGATTATTCTTACTTTGGCTGTTTCAGGTAT F-TTTCAGACATAAATCGTTTCTCA R-TCACTGCCCTACAACAGAAGTCTC	56.5	AF274518	ctg5234, ctg3284	Yes
11	Ssa58DU	Probe-TCCAATCCCTCTCCTGGCTTACTTACTCAGTTGAATAGGT F-CCCTCTCCTGGCTTACTTAC	51.5	AF019180	No Positives	No

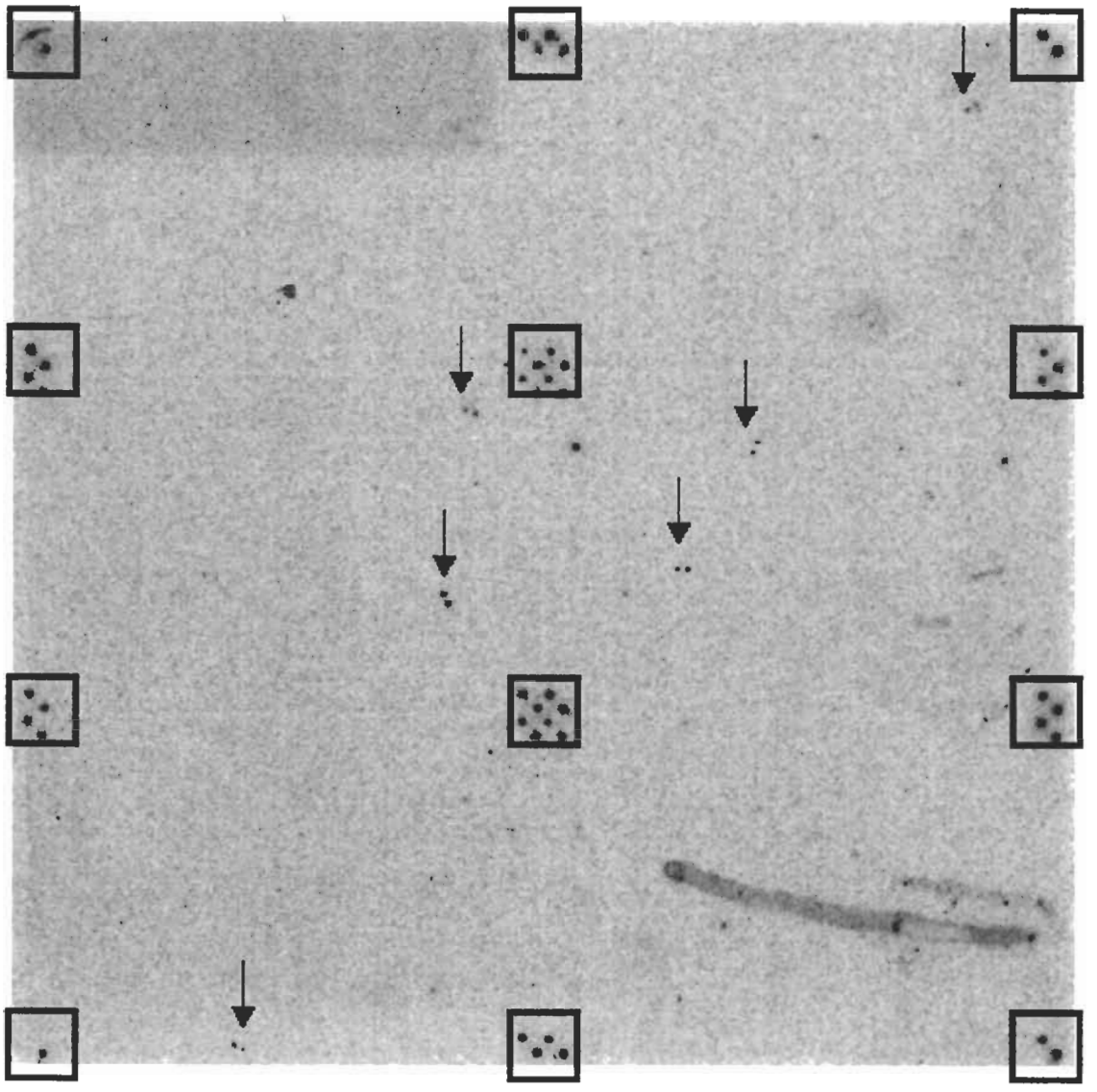
		R-GAGGGACTAATGACCTTGAG				
12	OmyFGT8TUF	F-CCAGAAATTAGATACAGGTTTGTGAGAGTTGATATTAAAGTGTTG R-TCCCTCCTCAGAATACCAAT	62.0	T. Sakamoto unpublished	ctg749	LG 12
13	Sal1UOG	F-GAGACACTGAGCCTCTTCGC R-CTTAACAGCAGCACAGCATTTCAGCGCCTCAACTCTCCAA	59.4	R. Danzmann unpublished	No Positives	No
14	OMM1278	F-AGTTGTTGATGAGAGCACACTCCCCACCCTTATGGAC R-CCATCAAGGGTAAACTGCCC	62.4	AF470042	ctg2708	No
15	OMM1122	F-CAAGGTAACTGTAGGTAATCATTTTGAAGGAATCTCCAGGTA R-TCTTGTGAGGTTTTCTGTCCT	62.4	AF352773	ctg351, ctg476	Yes
16	OMM1016	F-AGACCACACTGTTAAAGCAA R-CATCCACATGTCACGGCTGAGATACAAACCAGAGC	44.8	AF346676	No Positives	No

3.1.1.2 Large-insert BAC Library Probing

Each of the sixteen oligonucleotide hybridization probes was hybridized to the first twelve high-density BAC spotted membranes of the CHORI-214 Atlantic salmon BAC library. The first twelve high-density membranes represent a 13.8 fold coverage of the Atlantic salmon genome; the first 11.7 fold of which is the Hind III fingerprinted physical map (Ng et al., submitted). In cases where no positive hybridization signals were observed (markers Ssa58DU, Sal1UoG, OMM1016 and BHMS247), the oligonucleotides were then used to probe the remaining five high-density membranes (representing an 18.8 fold total genomic coverage).

Eleven of the sixteen microsatellite marker probes gave positive hybridization signals to BAC clones within the Hind III fingerprinted portion of the BAC library (an example of which is shown in Fig. 3.2). The Ssa202DU probe hybridized to three clones on filter 12, which were outside of the fingerprinted portion of the library. Markers Ssa58DU, Sal1UoG, BHMS247 and OMM1016 did not hybridize to any positive clones within the 18.8 fold genomic coverage library. Although PCR primers designed to the flanking regions of markers Ssa58DU, Sal1UoG, BHMS247 and OMM1016 PCR amplify observable products from Atlantic salmon genomic DNA, the region of the genome that they occupy (presumably along the sex chromosome) may not be represented in the BAC library.

Figure 3.2 CHORI-214 Atlantic salmon BAC library membrane hybridized with the OMM1278 marker probe. Arrows indicate positive clones which are identified by their characteristic duplicate spots in predetermined orientations. Boxes indicate the *Caenorhabditis briggsae* overgo orientation spots, which allow proper membrane orientation during analysis.



3.1.2 Incorporation of *SEX*-linked BACs into the Physical Map

BACs that were positive by hybridization for microsatellite flanking sequences were examined for the presence of the corresponding microsatellite by PCR using the oligonucleotide primers described in Table 3.1. The PCR confirmed BACs were then placed into contigs of the Atlantic salmon physical map (Ng et al., submitted) (Fig. 3.3) [analysis performed with FPC ver. 7.0 (Soderlund et al., 1997) and iCE ver. 3.0 (Fjell et al., 2003) publicly available at http://ice.bcgsc.ca].

3.1.2.1 Hind III fingerprinted physical map history

The Atlantic salmon Hind III fingerprinted physical map has gone through multiple builds at various FPC physical map likelihood cut-off stringencies, as well as contamination analysis methods. The build history of the physical map including dates, FPC likelihood cut-off values, number of contigs and number of singletons can be found in Table 3.2. All analyses described in this thesis use the most recent build (build 031125, November 25th 2003); however, studies I performed prior to this build may have employed data from previous builds. If such information is relevant to the study, it shall be noted in the text.

Table 3.2 **Hind III fingerprinted physical map history.** The Hind III physical map was originally built at an FPC likelihood cut-off value of e-16 (build 030326). It was then DQed (build 030326DQed), a process which involves removing all clones of questionable length or band intensity (suggesting that they are contaminated). The map was then examined using the software Mapmop (build 031015) (BC Genome Science Center, Physical Mapping Group [www.bcgsc.ca/lab/mapping/]) which uses an alternative method to FPC's own DQ function to remove clones suspected of contamination. Contigs within the map were then *in silico* end-to-end joined using FPC at a cut-off likelihood value of e-10 to produce build 031125. Finally, an alternative build at likelihood cut-off e-12 DQed was produced (build e-12DQed).

Build #	Date (DD/MM/YY)	FPC Likelihood Cut-off	Contigs	Singletons	Information
030326	06/05/03	e-16	7544	38014	
030326DQed	19/06/03	e-16	7695	38014	Q-clones Removed
031015	15/10/03	e-16	7666	38031	Mapmoped
e-12Dqed	25/11/03	e-16	5825	37888	End-End joined at e-10
031125	31/10/03	e-12	4354	27906	Q-clones Removed

3.1.2.2 Physical map integration

Microsatellite markers Oneu18ASC and Ssa406UOS each mapped to a pair of singletons, and thus were not incorporated into contigs. As indicated above, the Ssa202DU marker hybridized to clones outside those Hind III fingerprinted in the physical map, and therefore also could not be incorporated into contigs. The probes designed for the Ssa208DU, One102ADFG and OMM1122 markers (which are known to be duplicated within the Atlantic salmon genome), identified two separate non-collapsible positive contigs each (contig 5898 and contig 2133 for Ssa208DU, contig 3284 and contig 5234 for One102ADFG and contig 476 and contig 351 for OMM1122). Two other *SEX*-linked loci identified as duplicated by linkage analysis, BHMS447 and OmyFGT8TUF, each identified only a single positive contig (contig 6488 and contig 759, respectively). The hybridization probes designed for the flanking regions of the remaining microsatellite markers not known to be duplicated, each hybridized to a single contig: BHMS150 (contig 4481), BHMS216 (contig 1812), BHMS447 (contig 6488), Alu032NVH (contig 526) and OMM1278 (contig 2708)

Figure 3.3 Contig 2708 as viewed with iCE version 3.4. OMM1278 positive clones are highlighted in yellow.

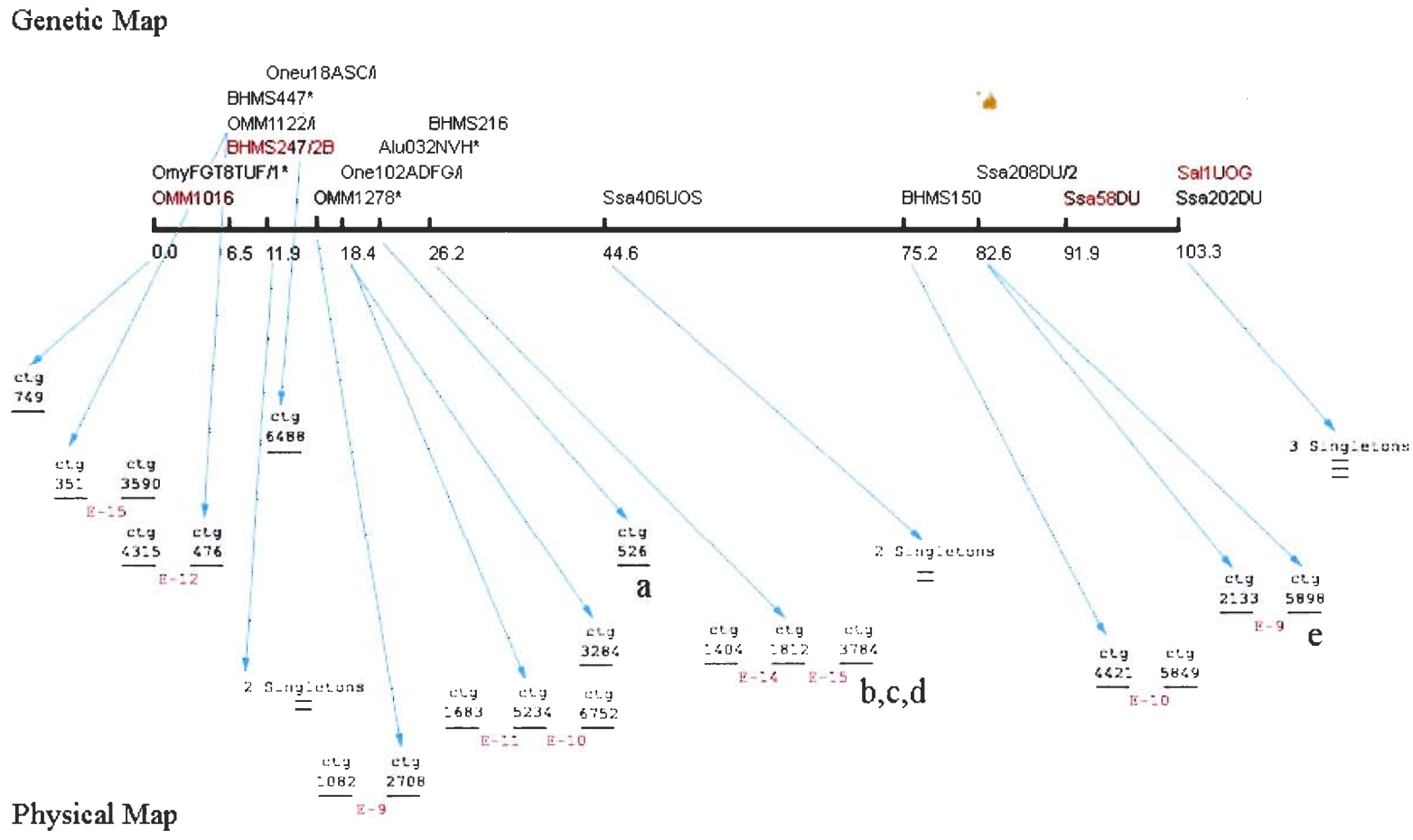
3.2 Extension of Coverage of *SEX*-linked Physical Map

3.2.1 *In silico* contig extension using FPC

End-terminal clones from sex-linked microsatellite marker positive BAC contigs and super-contigs as well as positive singletons were end-to-end joined using FPC ver. 7.0 (Soderlund et al., 1997). Query clones that scored hits to more than a single target BAC within another contig at a likelihood value of E-9 or better were considered significant. Putative super-contig joins as well as their likelihood E-value of the best joins are shown in Fig. 3.4. I was unable to collapse singleton BACs containing microsatellite markers (S0017N03 and S0119E21 positive for Oneu18ASC, and S0068C06 and S0116A05 positive for Ssa406UoS) into contigs (against themselves or the Atlantic salmon physical map). Clones S0119E21 and S0068C06 produced too few bands when digested with Hind III to be used by FPC. This may suggest that they contain repetitive elements preventing Hind III fingerprinting (Ng et al., submitted). Approximately 60 percent (7 of 12) of microsatellite marker containing contigs could be extended from at least one end into larger super-contigs. The two contigs positive for the Ssa208DU locus (see above) were joined at a threshold acceptable likelihood value of E-9. This could indicate that the duplicated regions of the genome containing the Ssa208DU marker retain significant sequence similarity, and this may be a complicating factor for the analysis of the Atlantic salmon genome. Alternatively, the genomic region containing the Ssa208DU marker could have

undergone a tandem duplication event, and therefore the two positive contigs are adjacent to one another on the sex chromosome.

Figure 3.4 Integration of ASL1 with the HindIII fingerprinted physical map. Arrows show contigs associated with sex-linked microsatellites; the female linkage map is shown as a positional reference. The best FPC likelihood scores are shown between contig-to-contig joins. Markers shown in black have been integrated with the CHORI-214 Atlantic salmon BAC genomic library. Letters show contig positions of five identified genes: a, Mitogen Activated Protein Kinase Kinase Kinase; b, Ubiquitin Conjugating Enzyme; c, Lamin; d, Predicted Protein RKHD1; and e, Homeodomain Box MSX.



3.2.2 BAC-end sequencing

The sequences of the ends of ninety representative contig end-terminal clones (according to physical map build 031015 of October 15th 2003 [Table 3.2]) were determined from *SEX*-linked microsatellite marker containing contigs as well as end-to-end joined contigs (BAC-end sequences have been submitted to GenBank under accession numbers CW883115 to CW883293). Clones were sequenced from both ends of the pTARBAC2.1 vector (Zeng et al., 2001) using the vector's own universal T7 and SP6 primers as provided from CHORI (http://bacpac.chori.org/cyclesere.htm). Sequences obtained using the T7 primer were generally approximately 500 nucleotides in length. Sequences obtained using the SP6 primer were significantly shorter, approximately 250 nucleotides in length.

Clones were selected for sequencing in order to design hybridization probes with the intent of probing the BAC library in order to confirm putative *in silico* contig joins, as well as extend the physical coverage of ASL1 from the ends of super-contigs. In some cases, two clones were selected from the same end of a single contig in order to provide multiple potential sequences from which to design primers in the event of poor sequencing results. Sequences from the BAC-ends were BLASTn-aligned against the entire non-redundant GenBank database (Altschul et al., 1990), and results of the BLAST analysis are shown in Table 3.2. Most BAC-end sequences did not produce significant BLAST hits. Approximately 10 percent (19 of 180 BAC-end sequences) produced reliable hits

(E-5 or lower) to known salmonid transposable and repetitive elements, of which the TC-1 like transposable element was represented in thirteen clones. Nineteen percent of the BAC-end sequences produced highly reliable BLAST hits to the flanking regions of known genes and microsatellite markers, some of which are mapped in the sex-chromosome. This suggests that these regions contain as yet uncharacterized repetitive elements within the Atlantic salmon genome.

The sequence obtained from the T7 primer read of S0017N03, a singleton positive for Oneu18ASC, produced a BLAST hit with a strong likelihood value (3E-27) to the Oneu18ASC sequence (GenBank accession number U56718), thus validating the efficacy of our approach.

Table 3.3 Tabulated results of BLASTn search against the GenBank non-redundant database using BAC-end sequences.

BLASTn Result	Number	Percentage (%)
No Signifcant Hits	120	66.7
Known Transposable Elements	13	7.2
Known Repetitive Elements	6	3.3
Microsatellite Marker Sequence	1	0.55
Flanking Region of Genes and Microsatellite Markers	34	18.9
Genes	5	2.8
No Sequence From Clone	1	0.55
	180	

3.2.2.1 Identification of novel genes

Five genes, not previously mapped in Atlantic salmon, were identified from reliable BLASTn hits (Table 3.3; Locations of genes along the physical map are shown in Fig. 3.4). Analysis of the gene positions using LocusLink (NCBI) did not reveal any conservation of synteny in the genomes of two fish (*Danio rerio* and *Tetraodon nigroviridis*), nor was any synteny observed in mouse. Shared synteny was observed for human and Atlantic salmon for mitogen activated protein kinase kinase kinase and lamin as both are found on human chromosome 1. However, the genes are found on opposite arms of the largest human chromosome, and their placement on chromosome 1 is more likely the product of genome rearrangement rather than preservation of an ancestral state.

Table 3.4 **Genes identified from BAC-end sequences.** Syntenic relationships for each of the five genes identified from BAC-end sequencing was ascertained for Danio, Tetraodon, mouse and human using LocusLink at the National Center for Biotechnology Information at http://www.ncbi.nlm.nih.gov/.

					Homologous Locus Position			
Name	**Genbank Accession Number**	**BLASTn e-value**	**Clone End-Sequence**	**Contig**	Danio	Tetraodon	Mouse	Human
Ubiquitin Conjugating Enzyme	NM_201489	e-25	S0232D13 T7 Read	3784	LG 4	Chr 1	Chr 6	Chr 7q32
Mitogen Activated Protein Kinase Kinase Kinase	NM_016693	e-05	S0017M11 T7 Read	526	N/A	Not Mapped	Chr 4	Chr 1p36.11
Lamin	BC033088	e-17	S0061N21 T7 Read	3784	LG 16	Chr 12	Chr 3	Chr 1q21.2-3
Predicted Protein RKHD1	XM_524036	e-17	S0213H09 T7 Read	3784	N/A	N/A	Chr 10	Chr 19p13.3
Homeodomain Box MSX	NM_031059	e-09	S0220F20 T7 Read	5898	LG 14	N/A	Chr 5	Chr 4p16

3.2.3 Chromosome walking

3.2.3.1 Probe design

An approximately 40 nucleotide hybridization probe and corresponding 20 nucleotide complementary PCR primer were designed from the BAC-end sequences using the Primer Premier version 5 software (http://www.premierbiosoft.com/primerdesign/index.html). Sequence reads were considered suitable for primer design if the PCR primers could produce a fragment of 100 nucleotides or greater length, and the primers had appropriate hybridization temperatures (greater than 75°C for the hybridization probe and greater than 50°C for the complementary PCR primer). In addition, only one sequenced clone from either end of a contig was selected for primer design. Forty-five of the BAC-end sequences were considered suitable for primer design under the above criteria, representing 25 of the 90 individual clones selected for sequencing. A list of end-sequenced clones selected for chromosome walking as well as their corresponding primers and PCR conditions are shown in Table 3.4. Whenever possible, primers were designed from both forward and reverse end-reads in order to orient end-clones with respect to their contigs.

Table 3.5 Clones selected for chromosome walking. Clone names are shown along with the contigs in which they are found as well as the microsatellite (MS) markers to which they are associated. Primer sequences are shown for the approximately 40 nucleotide hybridization probe and its complimentary PCR primer. All hybridization primers have been designed such that they will hybridize at an annealing temperature (T_m) of 65°C. PCR empirically determined T_ms are shown for each primer pair.

Clone	Associated MS Marker	Contig	Read	Primers / Probe	PCR T_m (°C)
S0612P18	Ssa202DU	Singleton	T7	F- GACTGAAACGCCATTATCT	56.5
				R- GAAAATACACCCCTACCATCATCCCTACCTCACAACAC	
S0581B12	Ssa202DU	Singleton	SP6	F- TACACCTCATTCCACTTTCT	62.5
				R- GGTTAAAAAACATATTTCGATTGAATCTGGTTACTACTCCT	
S0295I12	Alu032NVH	526	T7	F- GAGCCCTTGGGAAGTAATA	50.5
				R- GCTACCAATTGCCTAGCCAGTTGTAGCTCAATCTTGGGTG	
			SP6	F- GTACCATAGCTTTCTTCAGCA	63.4
				R- GGGGTGGAGGGTGAGTAGACAGAGGAAGGGAACAGGA	
S0230N16	Alu032NVH	526	T7	F- GGACTGGAGGGTAGCATAG	Fail
				R- TGAATAGGAGAAAAAACAAGCAACAAACAAAGAAGATAGG	
			SP6	F- AATCCCATCTGTCTTTGCT	56.4
				R- AAGCCTTGCTTTGCCTTTTTGACGGTTTGCACTGATAC	
S0360O08	Alu032NVH	526	T7	F- ACATCAGTACCATAGAGTGCG	62.4
				R- AGTTAGATGGACATGACACAGGTGAAACAGACAGAGGGG	
			SP6	F- ATGCCTCTGCTGGTCTACTTAGCTTTATCGTCAGATATCC	50.5
				R- TGGACGGAAATTCTTGATA	
S0179B16	Alu032NVH	526	T7	F- ACAGTGTCACCAGCAAAGC	62.4
				R- AGGACACCAATAAGAAGAAGAGACTTGCTTGGG	
			SP6	F- GCAGCCCTTATTCGGGAGAGTATTTTCTGCCTTGACCAGA	56.6
				R- GTTTGCATTTTGCGCTTTG	
S0139D08	Alu032NVH	526	T7	F- ACCTTCTACCATACCCTGT	56.5
				R- TTTTCTTAACCTGACAAATAGGTTGTTCCATCAATCTAAT	
			SP6	F- ACCAGTTACTTGAATCTTTGAAAAAGCGTATTCTCCAATGG	56.5
				R- AGCGTGCCTTATTACTACA	

S0336N24	BHMS150	4421	T7	F- CGGGATCTGGATGGGAGACCAGATGCTGCTGGAAGTGG	62.5
				R- TTATGGCAATGACCTGGGG	
			SP6	F- CTCACCTAATCCCCACATA	56.5
				R- GGATTAATATTTCTACCAGATAAATGCTTCTGAAAGTATTGAACT	
S0397C07	BHMS150	4421	T7	F- GGGCGAGAAGAAGTCACAG	Fail
				R- CAAATCAAAGAGTGTGAATAGGAAATTATCAGGCGGAG	
S0017N03	Oneu18ASC	Singleton	T7	F- CTCACAGAAGTCCTCCACC	62.5
				R- GGGGAGGGGGGGGTTTTACACATTTGTCTTGAACGCAC	
S0042N21	Ssa208DU	5898	T7	F- GTATGGTTTGTCCCTGTGC	62.5
				R- AGGAACTTAAAACGTTCCACCTTCTCCGCTACTGTCTG	
			SP6	F- GACCACCATTCTCCTTATTTAGCAAAGCGCATGGATCAA	62.5
				R- TGCTCACAGTTATTTCATCTAC	
S0220F20	Ssa208DU	5898	T7	F- TCTACGGACATTCTAACGCCTACCCCAGGCCCACT	56.5
				R- CACCATCTTTCAATCAACAA	
			SP6	F- CAAGTCATGCCAAGAAAATCCACGCTTGATCTAATTTTCTG	Fail
				R- TATTTCTTGAACTGCATTG	
S0503M23	Ssa208DU	2133	T7	F- CTGGAGTACAGCACCTTTGATTTACCTCCG	56.5
				R- TCGCTACACCTCTTCATCT	
			SP6	F- CAAGTGAAAGGTGAAGAAA	50.5
				R- TCTTCGCAGCAATATCTGCAGAAATAGGATGGTTGT	
S0068C06	Ssa406UOS	No-digest	T7	F- CCCCAGACCTCACAACCATAAAGGGCAATGGGTTCT	62.5
				R- CGCCAGGACCACAAATACA	
S0116A05	Ssa406UOS	Singleton	T7	F- CAACACCACAGTCCTCTACCTGAAAGAGGTCACCAAG	62.5
				R- ATAAACCAATAACGCCCAG	
			SP6	F- CCGCTTCTCACAGACATTCGGTACAGCATGGG	62.5
				R- CATCGTATTCCGCTCCATC	
S0142L05	BHMS216	1812	T7	F- AATGGTTTCATACTTCCTTG	Fail
				R- CGTTTCTGCCATGCTTAGTGGTGTATTCTTCTCGGT	
			SP6	F- TATGCTGCTCAGGAAAAGAGAAAATAAAGAAGC	Fail
				R- TGCTGGTAACGGAGGAATC	
S0213H09	BHMS216	1812	T7	F- TAGACAAGCAGGGATAAATG	56.5
				R- GCTAAGATTGTTACACCTACACGGTAATACGAAGCAAT	
			SP6	F- ACCAACGCCACATGATAAATGCATGGCCAGCCCC	Fail

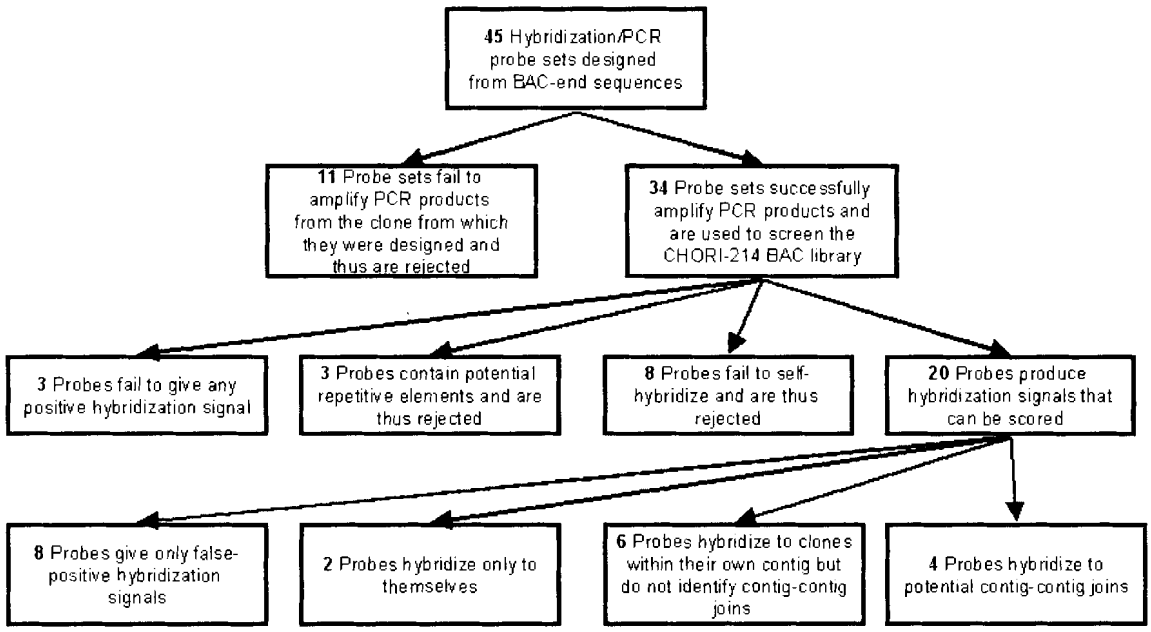
				R- CTGTTACTGCCGTCACTTTC	
S0205P08	BHMS216	1812	T7	F- CTTTAGAACTTAAGCCAGAGGAGAGGCCTCAGAC	62.5
				R- CAGAGGGAAGGCGAGTATG	
			SP6	F- TATCTGGACTGATGCTGGCCACTGATTGGTGCTAAAACAC	56.5
				R- ACCCATACCCATTGCTTAG	
S0232D13	BHMS216	1812	T7	F- ATCTTTGAGTCGTTCCTGCCTCAGCTGCTGGCATACC	56.5
				R- TGTTAATCAAGGCGTAGGG	
			SP6	F- AATGCTTGGTTGCTTTTAT	50.5
				R- GCTGTGGGCTTTAAATTTGCCTTTTGACATCCAGATA	
S0343D17	BHMS247	6488	T7	F- TGCCGACCGCATTATCTAC	Fail
				R- GACTGGAATTGGAATGTTGGGAACCGACTGTGA	
			SP6	F- ACAGCCAGAGGTAAGGACAGAACTGACATGTTAACC	56.5
				R- CAGGAGGTGGTAATGAGGT	
S0042I14	BHMS447	6488	T7	F- CTCATTTACCATTGCGACC	44.8
				R- CATAGCCCATCTATAATTTAGCCCAAACAACTACCTCT	
			SP6	F- TTTCCCAGGCTGAGATAGT	56.5
				R- GTTTAATGGTCTCTTTTTCAAGCTTAAAAGGATAAACATT	
S0372C06	BHMS447	6488	T7	F- GTAGGGATGGTGCCAGGTT	44.8
				R- AACCAAGCCATGAGGTTGAAGGAATTGTCGG	
			SP6	F- TCTGTGCTCAGTGGGATGG	Fail
				R- CCACCTGTAGTTCTCTGACTCTAATGGGAGTGGATC	
S0382A05	One102ADFG	5234	T7	F- CGTGAGTAGTTTGCATCGT	50.5
				R- TTAGTTTAAAATCAGCTGTGTTTGCTAGGGATAAGG	
			SP6	F- TGGTCCTGTCACCTCTGTTCATTTACTGAAAAACTAGC	Fail
				R- CCTCTTATCTCCTCCTTCCC	
S0370K19	One102ADFG	5234	T7	F- ATCCTCTGCCTCAATCTCC	62.5
				R- GGGTTATGTGACTTGGTGTCTATCGTGCCATCGTGTT	
			SP6	F- CAGGGCTTCAGAGGCTATTCCGATGGACCGTAACTAAC	Fail
				R- GTCAACCTTTGTTTTGGCA	
S0459F08	One102ADFG	3284	T7	F- GTATGGTTTGTCCCTGTGC	Fail
				R- AACTTAAAACGTTCCACCTTCTCCGCTACTGTCTG	
			SP6	F- GACCACCATTCTCCTTATTTAGCAAAGCGCATGGATC	44.8
				R- TGCTCACAGTTATTTCATCTAC	

3.2.3.2 BAC library probing

A generalized scheme of the CHORI-214 genomic BAC library probing using the BAC-end sequence designed primers can be found in Fig. 3.5. PCR was performed on each individual clone using its own primer set(s) in order to confirm that the oligonucleotide primers were functional. Eleven out of the 45 primer sets failed to produce visible PCR fragments. These primer sets were considered unsuitable for BAC library screening as their failure to produce PCR products suggests that the primers do not bind specifically to the sites for which they were designed. This could possibly be due to the fact that they were designed from poor BAC-sequence reads, or that they contain as yet uncharacterized repetitive elements within the Atlantic salmon genome which may have been present in multiple copies along the BAC clone to which the were designed.

Hereafter, the following convention is used for the naming of primers designed from BAC-end sequences: Primers are given the name of the clone they were designed from followed by the sequencing primer used in order to obtain the sequence. For example, the hybridization primer designed from the sequence obtained from clone S0612P18 using the T7 primer is named S0612P18T7.

Figure 3.5 Generalized flow-scheme of chromosome walking procedure.



The hybridization probes S0612P18T7 and S0581B12SP6, both positive for the presence of the Ssa202DU microsatellite marker, were used to screen all seventeen high-density spotted membranes of the BAC library. As the Ssa202DU marker is considered to be the marker closest to *SEX* in most linkage maps (R. Danzmann unpublished results; B. Hoyheim unpublished results), the first twelve filters (representing the entire Hind III fingerprinted portion of the library) were screened in order to identify any possible end-joins within the library and collapse the Ssa202DU singletons into a contig. In addition, all three clones identified as positive for the Ssa202DU marker fell outside of the Hind III fingerprinted portion of the BAC library; therefore, membranes thirteen through seventeen were also screened in order to observe the presence of self-hybridization. The hybridization probes of the remaining 32 primer sets were used to screen only the first six high-density membranes of the CHORI-214 BAC-library, representing a 6.8 fold coverage of the Atlantic salmon genome. The results of the BAC library screening are found in Table 3.5.

Table 3.6 **Results of the 34 probes used in the screening.** Clone names are shown along with the contigs in which they are found as well as the microsatellite (MS) markers to which they are associated.

Probe	Associated MS Marker	Contig	Self-Hybridized?	Probing Result
S0612P18T7	Ssa202DU	Singleton	Yes	Clones S0493K22, S0534O16, S0552D23
S0581B12SP6	Ssa202DU	Singleton	Yes	No positives
S0295I12T7	Alu032NVH	526	Not Probed	Clones in contig 526
S0295I12SP6			Not Probed	Repetitive
S0230N16SP6	Alu032NVH	526	Yes	Clones in contig 526
S0360O08T7	Alu032NVH	526	Not Probed	False positives
S0360O08SP6			Not Probed	Clones in contig 526
S0179B16T7	Alu032NVH	526	No	Repetitive
S0179B16SP6			Yes	Clones in contig 526
S0139D08T7	Alu032NVH	526	No	False positives
S0139D08SP6			No	No positives
S0336N24T7	BHMS150	4421	Not Probed	Repetitive
S0336N24SP6			Not Probed	No positives
S0017N03T7	Oneu18ASC	Singleton	No	False positives
S0042N21T7	Ssa208DU	5898	Yes	Clones S0039P19 and S0251P06 of ctg2133
S0042N21SP6			Yes	False positives

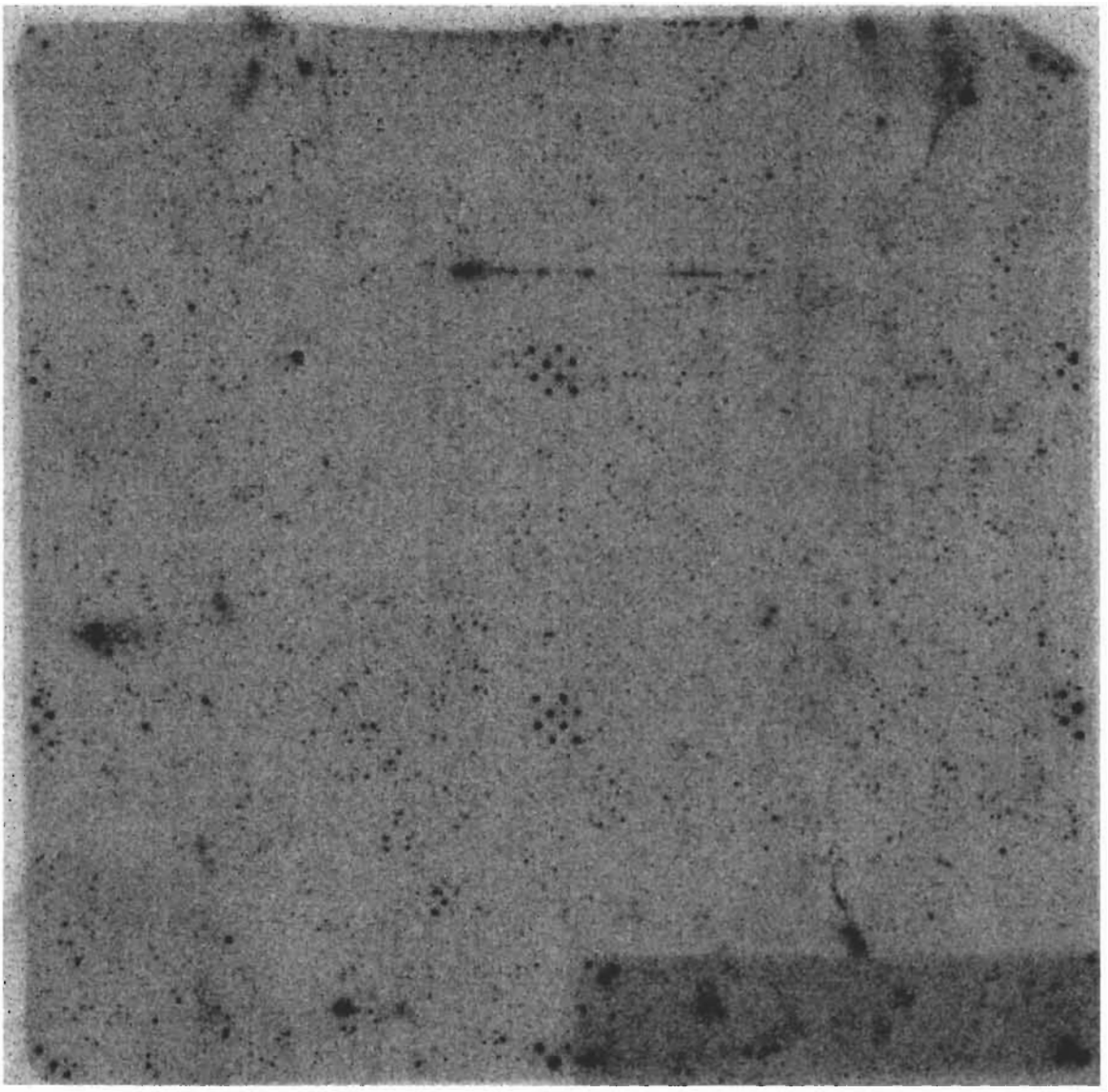
S0220F20T7	Ssa208DU	5898	Yes	False positives
S0503M23T7	Ssa208DU	2133	Not Probed	False positives
S0503M23SP6			Not Probed	False positives
S0068C06T7	Ssa406UOS	No-digest	No	Repetitive
S0116A05T7	Ssa406UOS	Singleton	Yes	Clones S0143M20, S0221C23 and S0275J14 of contig 5324 and S0255J14, a singleton
S0116A05SP6			Yes	No Positives
S0213H09T7	BHMS216	1812	Yes	False positives
S0205P08T7	BHMS216	1812	Yes	False positives
S0205P08SP6			Yes	Clones in contig 1812
S0232D13T7	BHMS216	1812	Yes	False positives
S0232D13SP6			No	False positives
S0343D17SP6	BHMS447	6488	Not Probed	No positives
S0042I14T7	BHMS447	6488	No	Repetitive
S0042I14SP6			No	False positives
S0372C06T7	BHMS447	6488	Not Probed	Repetitive
S0382A05T7	One102ADFG	5234	Not Probed	Clones in contig 5234
S0370K19T7	One102ADFG	5234	Not Probed	Clones from contig 5234 and S0054G19, S0114M21 and S0198B08 from contig 2133
S0459F08SP6	One102ADFG	3284	Not Probed	No positives

3.2.3.2.1 Rejected hybridization probes

Primers used to screen the BAC library that were designed from clones present on the first six high-density spotted membranes were assessed for the presence of self-hybridization. Eight out of the 34 probes used in the screening (S0179B16T7, S0139D08T7, S0139D08SP6, S0017N03T7, S0068C06, 232D13SP6, S0042I14T7 and S0042I14SP6) failed to self-hybridize, and therefore were not included in further analyses.

Three of the remaining 26 hybridization probes used to screen the BAC library produced extremely noisy hybridization signals (S0295I12SP6, S0336N24T7 and S0372C06T7) as they hybridized to large numbers of clones in a non-contig specific manner, an example of which is shown in Fig. 3.6. These clones were examined for the presence of identified Atlantic salmon repetitive elements using the Repeat Masker software version 3.0 (www.repeatmasker.org; Smit et al., 1996-2004) employing a database of repetitive elements identified from numerous Atlantic salmon sequences including expressed sequence tags (ESTs), whole shotgun sequenced BAC clones and BAC-end sequences (S. Ng, unpublished data). None of the clones contained sequences that were masked by the Repeat Masker software, which most likely indicates that they contain uncharacterized repetitive elements within the Atlantic salmon genome.

Figure 3.6 CHORI-214 Atlantic salmon genomic BAC library high-density spotted membrane, probed with a repetitive element-containing probe. The probe designed for the S0295I12SP6 sequence was identified as containing potentially uncharacterized repetitive elements within the Atlantic salmon genome. Large numbers of positive clones appear in a non-contig specific manner.



Three of the remaining 23 probes used to screen the BAC library showed no positive hybridization signals whatsoever: S0336N24SP6, S0343D17SP6 and S0459F08SP6. As stated above, the SP6 primer produced significantly shorter sequence results. The SP6 primer site of the pTARBAC2.1 vector is known to demonstrate some degree of polymorphism (J. Stott, personal communication) which can lead to poor sequencing results. It is possible that the primers designed for these three clones demonstrated a significant degree of sequence dissimilarity with the clone target sequence, such that they were able to amplify PCR products on BAC clone DNA, but were unable to bind during hybridization, under the stringency of the membrane probing wash conditions. It should also be noted that the clones from which these primers were designed were not located on the first six high-density spotted membranes, and therefore the presence of self-hybridization was not tested.

3.2.3.2.2 Scored probes

Twenty out of the 34 probes used in the screening produced non-repetitive hybridization signals on at least one of the first six membranes of the BAC library. Positive hybridization spots were first scored and their respective contigs were analyzed in FPC version 7.0 and iCE version 3.4. In the event that the probe hybridized to two or more clones within a large contig, the clones were subjected to PCR amplification using the original hybridization oligonucleotide used to identify it as well as its corresponding reverse primer. If the probe hybridized to a single clone without hybridizing to corresponding overlapping clones within a contig, the hybridization was rejected as a false positive. Such false positives

may have been the result of background radioactivity spots mimicking the pattern of orientation used to identify the BAC clones spotted on the high-density membranes (Fig. 3.2). It is also possible that such false positives were able to bind weakly to the oligonucleotide hybridization probes, such that visible hybridization signals were observed in some of these clones, whereas such signals in other overlapping clones were not visible above background levels.

Eight of the scored probes (S0360O08T7, S0042N21SP6, S0220F20T7, S0503M23T7, S0503M23SP6, S0213H09T7, S0205P08T7 and S0232D13T7) produced only false positive hybridization signals. Unfortunately, clones S0360O08 and S0503M23 were not found on the six hybridization filters that were screened. Therefore, I was unable to verify the presence of self-hybridization for probes designed from these clones. Probes S0581B12SP6 and S0116A05SP6 hybridized only to themselves, and demonstrated no additional positives. It is unexpected that both the S0503M23T7 and S0503M23SP6 probes would both bind only false positives given that clone S0503M23 is a terminal clone of contig 2133. At least one of the clone end-probes should have identified other overlapping clones within contig 2133.

Hybridization probes S0295I12T7, S0230N16SP6, S0360O08SP6, S0179B16SP6, S0205P08SP6 and S0382A05T7 hybridized to clones in the contigs to which they belong; however, they failed to identify any new contig-contig joins.

Hybridization probe S0612P18T7, hybridized to PCR confirmed clones S0493K22, S0534O16 and S0552D23, collapsing them into a contig. Clones

S0534O16 and S0552D23 are outside of the Hind III fingerprinted portion of the BAC library. I performed *in silico* end-to-end joining of singleton S0493K22 in FPC version 7.0 according to the same conditions used in the section 3.2.1. FPC identified a potential end-to-end join between clone S0493K22 and another singleton, S0279G12, at a likelihood score of e-13.

Hybridization probe S0042N21T7 identified clones S0039P19 and S0251P06 of contig 2133. As shown in the integration of the ASL1 linkage group and physical map (Fig. 3.4), the contig in which clone S0042N21 is found, 5898, is potentially joined *in silico,* at an FPC likelihood value of E-9, to contig 2133. Both contigs 5898 and 2133 contain clones identified as positive for the microsatellite marker Ssa208DU. That a hybridization probe designed from a clone from contig 5898 hybridizes to clones within contig 2133 suggests two possibilities: A) Microsatellite marker Ssa208DU is tandemly duplicated along the sex-chromosome such that it is found in two separate contigs that join to one another; or B) As microsatellite marker Ssa208DU is known to be duplicated within the Atlantic salmon genome through linkage mapping (B. Hoyheim, unpublised results), contigs 5898 and 2133 represent the two separate, non-syntenic Ssa208DU loci, which retain enough sequence similarity for probes designed from the end sequence of clones from one contig, to hybridize to the conserved region in the other contig.

Hybridization probe S0116A05T7 identified clones S0143M20, S0221C23, and S0275J14 of contig 5324 as well as the singleton S0255F17. These results collapse the singletons S0116A05 and S0255F17 into contig 5324. *In silico*

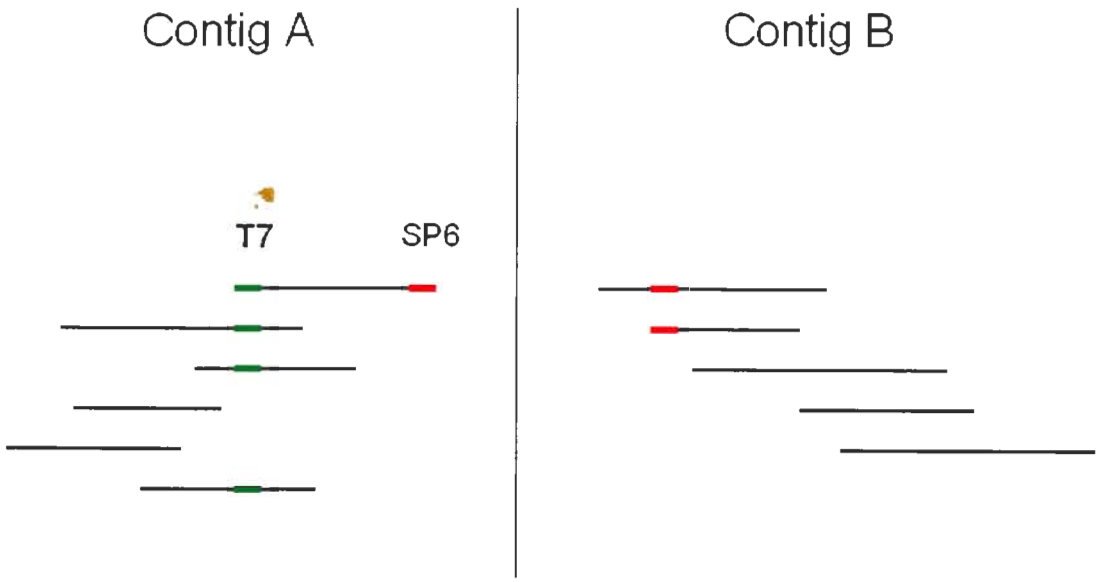
contig extension was attempted in FPC version 7 under the conditions employed previously; however, I did not observe any significant end-to-end joins.

Hybridization probe S0370K19T7 identified clones within the contig in which it is found (contig 5234) as well as clones S0054G19, S0114M21 and S0198B08 in contig 3284. Both contigs contain clones containing the One102ADFG microsatellite marker. This marker is known to be duplicated within the Atlantic salmon genome, and contigs 3284 and 5238 most likely represent each of the duplicate One102ADFG loci. It is likely that the sequence surrounding the two separate One102ADFG loci are significantly conserved, such that hybridization probes designed from clones in one contig are hybridizing to the same sequence in clones located in the other contig.

3.2.3.3 Analysis of clone S0503M23 T7 and SP6 primers within contig 2133

As previously observed, both end-sequence designed probes from clone S0503M23 gave only false hybridization signals. Neither of them identified other overlapping clones within the contig in which they are found (contig 2133). Fig. 3.7 demonstrates my expectations concerning the positive hybridization signals I would observe when both T7 and SP6 end sequence designed probes from a single contig terminal clone were used to probe the BAC library.

Figure 3.7 Representation of contig-contig joining via BAC-end sequence designed probes. The hybridization probe designed from the T7 sequence of the terminal clone of contig A (represented in green) hybridizes to clones within contig A, indicating that the SP6 sequence designed primer (represented in red) is the terminus of contig A. In this hypothetical situation, contig B would not have enough overlap between its terminal clone(s) and the terminal clone(s) of contig A for FPC to collapse both contigs into a super-contig. However, the small amount of overlap between the two contigs does include the SP6 sequence from contig A and thus the probe is able to identify terminal clones from contig B as being joined to contig A.



I would have expected either the S0503M23T7 or S0503M23SP6 probes to have behaved like the "green" probe of Fig. 3.7, and identified positive overlapping clones from contig 2133. It is entirely possible that the other probe identify no potential contig-contig joins, if no such joins exist.

In order to test the reliability of the end-sequence designed S0503M23 probe/primers, I selected five additional clones in addition to clone S0503M23 from contig 2133 and attempted to identify whether the S0503M23T7 and/or S0503M23SP6 primers would amplify PCR products from their sequence. The relative positions of the selected clones (S0069C07, S0070C22, S0100N19, S0113N21 and S0191I09) as well as clone S0503M23 within contig 2133 are shown in Fig. 3.8. Previous work has shown that graphical representation of the physical map as shown in software such as iCE or FPC gives only a rough representation of contig size and shape (J. Schein, personal communication). Overlap between clones within a given contig is not drawn to scale.

What is immediately apparent from Fig. 3.8 is that clone S0503M23 is not a contig terminal clone as suggested by the most recent build of the physical map (Table 3.2). As previously indicated, terminal clones were selected based on an earlier build of the physical map (build 031015), in which clone S0503M23 was observed to be contig terminal.

PCR was performed on each of the five clones and clone S0503M23 using both the S0503M23T7 and S0503M23SP6 PCR primers sets. The PCR products were subsequently analyzed using agarose gel electrophoresis, the results of which are shown in Fig. 3.9.

Figure 3.8 Contig 2133 of the Hind III fingerprinted physical map as viewed in iCE 3.4. The clones selected for PCR analysis using S0503M23T7 and S0503M23SP6 primers are shown in yellow.

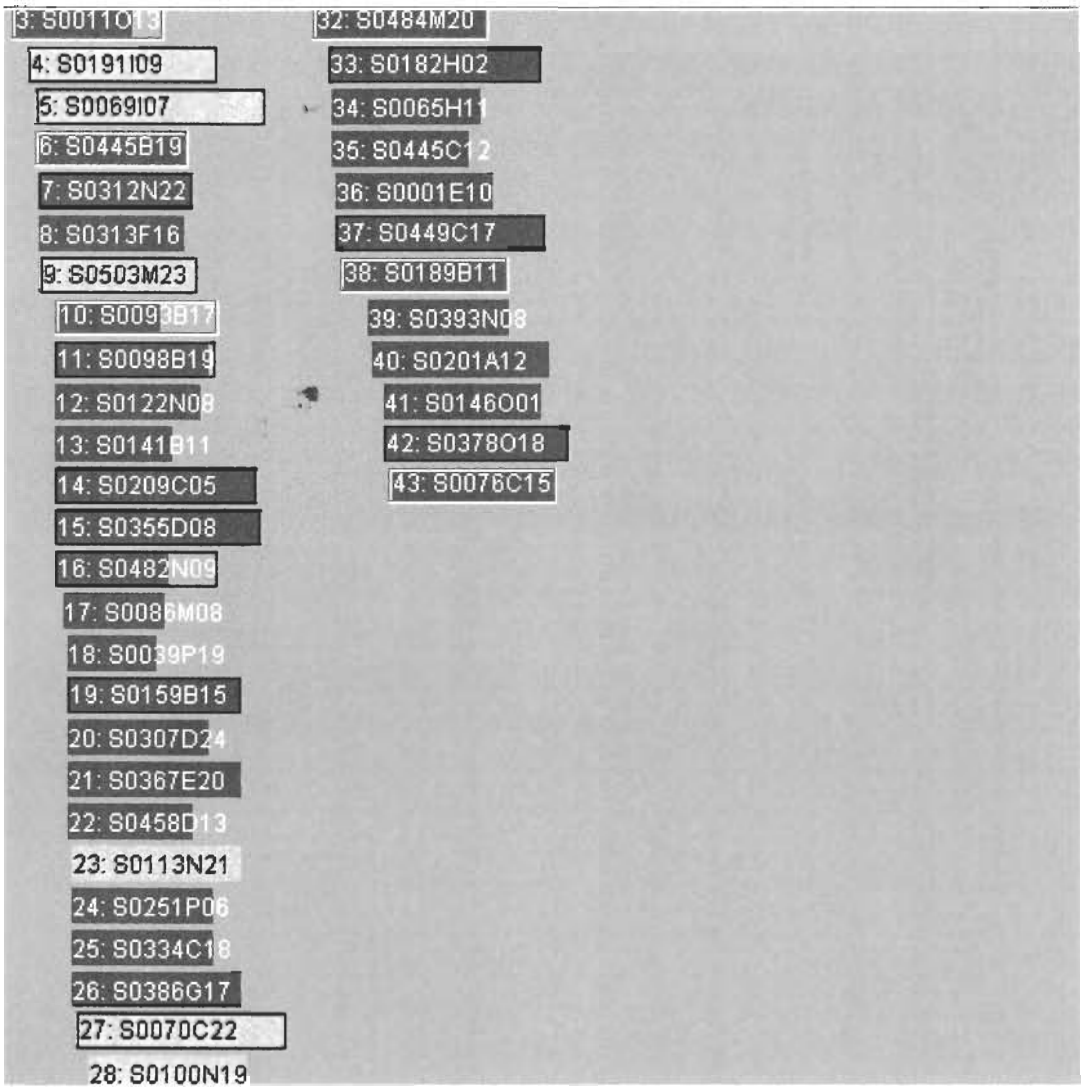
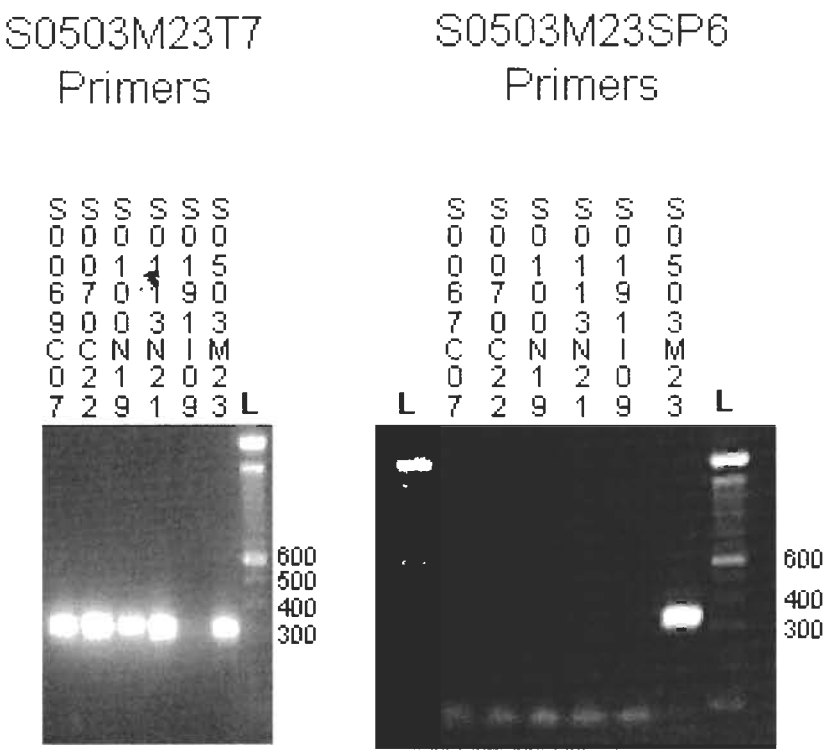


Figure 3.9 PCR amplification of clones in contig 2133 using S0503M23 end-primers. PCR products were run on a 1% Agarose gel in 1X TBE containing 50 µg/ml ethidium bromide. Clone names are indicated above each gel lane. Lanes marked "**L**" contain 100 base pair standard ladder (Invitrogen) in which the bright band is an approximately 600 base pair fragment and each subsequent band below it decreases in size by approximately 100 base pairs.



The results of the PCR amplification using the S0503M23T7 primer indicate that the T7 terminus of clone S0503M23 lies buried within overlapping clones of contig 2133. However, if the graphical representation of contig 2133 was correct with respect to the degree of overlap between clones, clone S0191I09 should have produced an amplification product. Similarly, the S0503M23SP6 primers amplify a product only from clone S0503M23 suggesting that this BAC is, in fact, contig terminal. As previously indicated the graphical representation of contigs within the iCE software is a rough approximation (B. Koop, personal communication). These data confirm that clone S0503M23 is contig terminal.

3.3 Fluorescent *in situ* Hybridization of Sex-Linked Clones

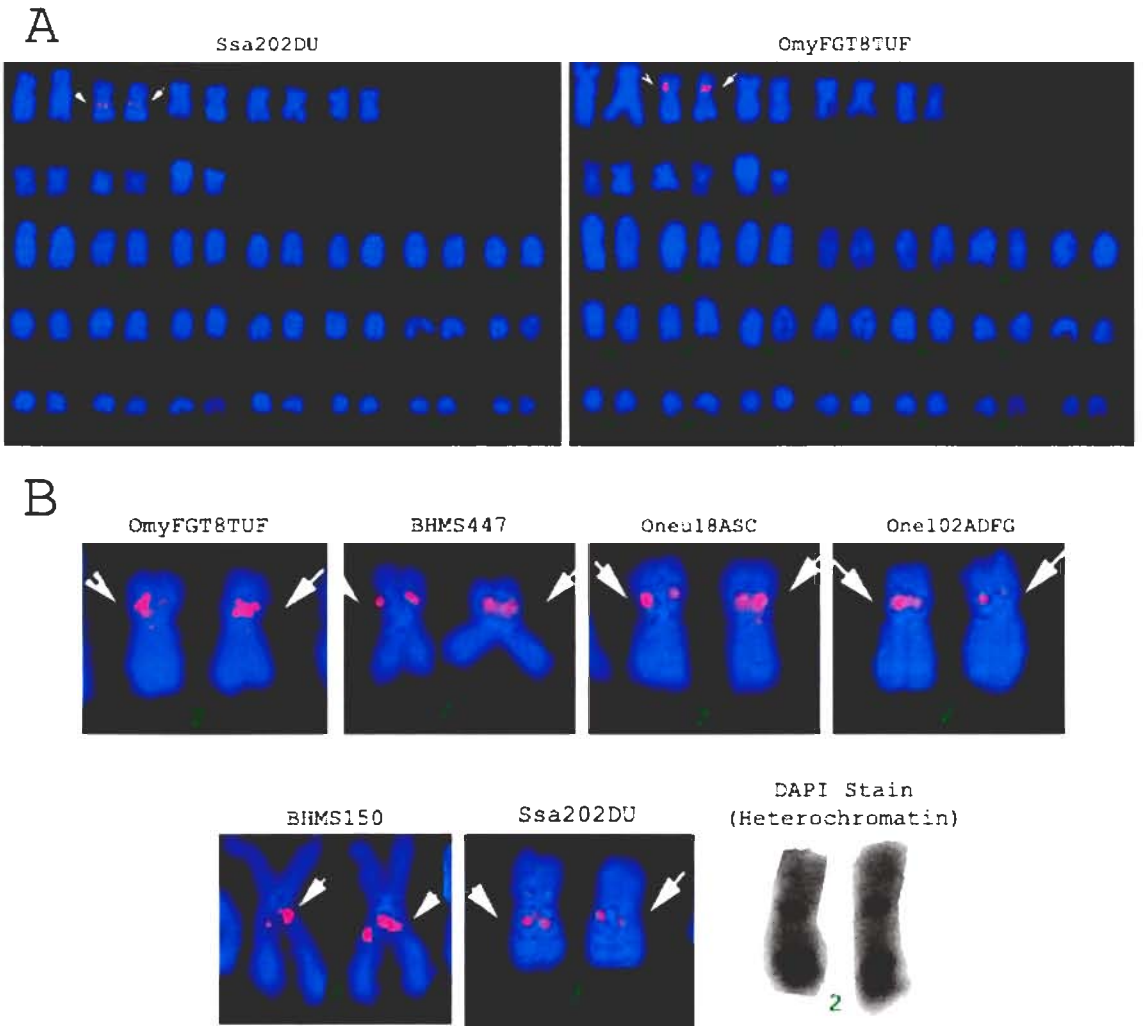
3.3.1 Identification of the Atlantic salmon sex-determining chromosome

I selected representative BAC clones positive by hybridization and PCR for markers Ssa202DU (clone S0605H01), BHMS150 (clone S0336N24), One102ADFG (clone S0018G23), BHMS447 (clone S0051B22), Oneu18ASC (clone S0017N03) and OmyFGT8TUF (clone S0227A12), and sent them to Dr. Ruth Phillips' lab at Washington State University, Vancouver WA, where they were used to probe chromosomes prepared from a male Atlantic salmon, via fluorescent *in situ* hybridization (FISH).

Results produced by FISH using a BAC containing either the Ssa202DU or OmyFGT8TUF markers are shown in Fig. 3.10A. All clones used in the FISH analysis hybridized to chromosome 2, the second largest metacentric chromosome, revealing it as bearing the sex-determining region. It should be

noted that both of the chromosome 2 homologs gave positive hybridization results with the BACs in all cases (Fig. 3.10B).

Figure 3.10 Results of Fluorescent in situ hybridization. (A) BAC clones positive by hybridization for microsatellite markers Ssa202DU and OmyFGT8TUF hybridize by FISH to chromosome 2. (B) Relative positions of microsatellite markers along the sex-chromosome as shown through FISH as well as the position of heterochromatin on the sex-chromosome (black areas) as shown by 4,6-diamidino-2-phenylindole (DAPI) staining. The clone used as a hybridization probe for each of the microsatellite markers is as follows:

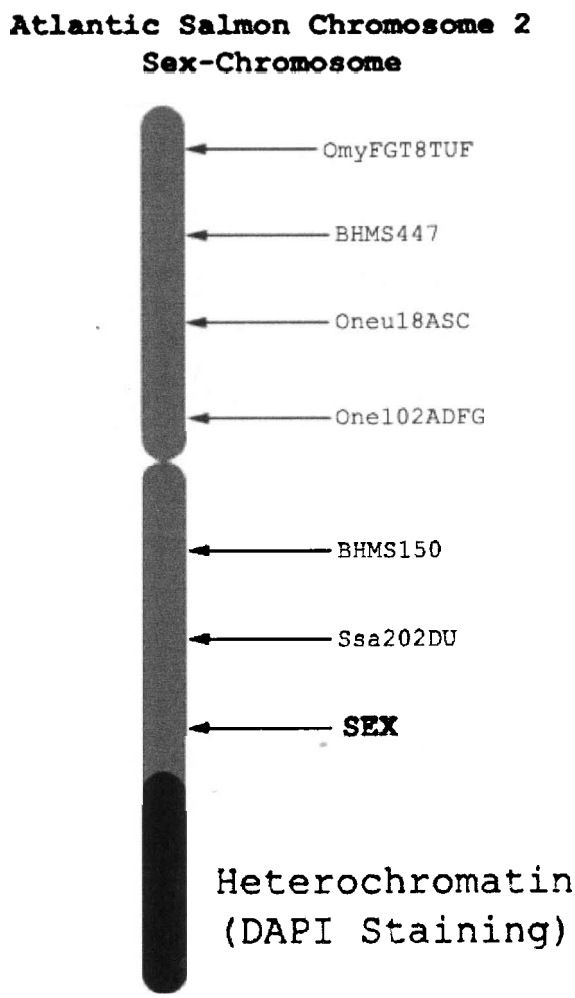


3.3.2 Orientation of linkage map relative to the sex-determining chromosome

The relative order of FISH hybridized BACs along chromosome 2 correlates well with the order of the corresponding microsatellite markers obtained from linkage analysis (Fig. 3.11). As is further illustrated in Fig. 3.11, mapping of multiple BACs along chromosome 2 allowed us to orient this chromosome with respect to the linkage map. 4,6-diamidino-2-phenylindole (DAPI) staining of Atlantic salmon chromosomes revealed that there is a large region of heterochromatin on the telomeric end of the long arm of chromosome 2 (Fig. 3.10B). The sex-linkage data combined with the information obtained from cytogenetic analysis suggest that *SEX* lies past the Ssa202DU marker on the long arm of Atlantic salmon chromosome 2.

Heterochromatin
(DAPI Staining)

Figure 3.11 Orientation of ASL1 with respect to the sex chromosome. Integration of FISH data showing that the relative positions of microsatellite markers as shown by FISH correlate with the genetic linkage maps. SEX is predicted to be located between the Ssa202DU marker and a large region of heterochromatin on the long arm of chromosome 2



3.3.3 Chromosome mapping of duplicate microsatellite loci

Both of the non-overlapping singletons positive for microsatellite markers Oneu18ASC (S0017N03 and S0119E21) and Ssa406UOS (S0068C06 and S0116A05) were separately hybridized on to chromosomes using FISH in order to ascertain whether they represented duplicate marker loci within the Atlantic salmon genome (Fig. 3.12).

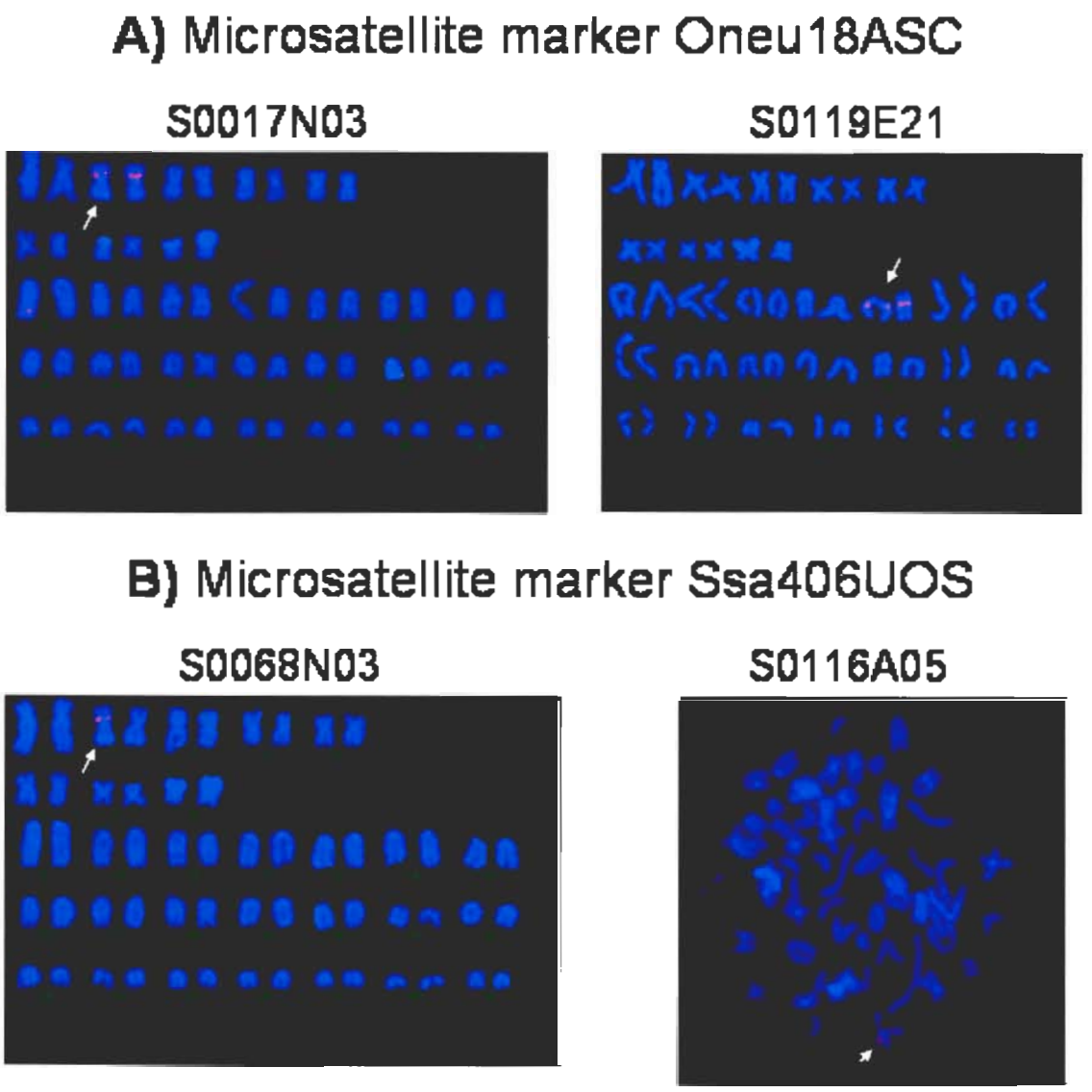
In the case of the Oneu18ASC microsatellite marker (Fig. 3.12A), clone S0017N03 hybridized to the short arm of the sex chromosome pair (chromosome 2) whereas clone S0119E21 hybridized to a medium-sized autosomal metacentric chromosome. Salmonid chromosomes do not demonstrate GC-rich and GC-poor regions, and thus a technique such as G-banding, which is very useful in differentiating mammalian chromosomes, cannot be used to differentiate chromosomes of similar shape and size in Atlantic salmon (Ocalewicz et al., 2003). Although Fig. 3.12 indicates that clone S0119E21 hybridizes to chromosome thirteen, I cannot conclusively identify whether this is actually chromosome eleven, twelve or thirteen as they are all metacentrics of similar size.

In the case of the Ssa406UOS microsatellite marker (Fig. 3.12B), clone S0068C06 hybridized to the short arm of the sex chromosome whereas clone S0116A05 hybridized to a medium-sized metacentric chromosome. In the case of both Ssa406UOS positive clones hybridization was observed on only one chromosome of the chromosome pair. This is most likely due to weak hybridization of the clone (Ruth Philips, personal communication). For example, if

clone S0116A05 contained repetitive elements, it could hybridize more strongly to one chromosome because it possesses more copies of the repeat than its homolog (perhaps as a result of somatic cell unequal crossing over).

Although these data demonstrate that non-collapsible contigs positive for the same microsatellite marker represent residual duplicated loci from the ancestral salmonid tetraploidization, I am currently unable to determine whether both duplicated loci lie on the same or separate autosomal chromosome(s). This could be resolved via dual-hybridization of both non-sex linked clones simultaneously, testing for the presence of hybridization to the same metacentric autosomal chromosome pair.

Figure 3.12 Fluorescent *in situ* hybridization of duplicate microsatellite loci. Non-collapsible singleton clones identified as positive for the presence of microsatellite markers Oneu18ASC (A) and Ssa406UOS (B) were placed onto Atlantic salmon chromosomes using fluorescent *in situ* hybridization. White arrows indicate location of fluorescent signal. In both cases, one of the two clones hybridized to the sex-determining chromosome, while the other hybridized to a medium-sized autosomal metacentric chromosome, indicating that the singletons represent conserved duplicate loci within the Atlantic salmon genome.



3.4 Analysis of BAC clone S0493K22

In order to gain a better understanding of the genomic composition of the sex chromosome, I generated a shotgun library of clone S0493K22 and sequenced it to approximately 8.5 fold coverage. S0493K22 was found to be physically overlapping a clone containing the Ssa202DU marker (S0612P18) through chromosome "walking" (Table 3.6). As well, this was the only clone in the region of the Ssa202DU marker that had been Hind III fingerprinted, and thus it was possible to estimate the clone's size, which is crucial to determining sequence coverage. In addition, prior to sequencing, S0493K22 was placed on Atlantic salmon chromosomes by FISH and was localized to the same region as the Ssa202DU marker (Fig. 3.10).

3.4.1 Generation and sequencing of a shotgun library of clone S0493K22

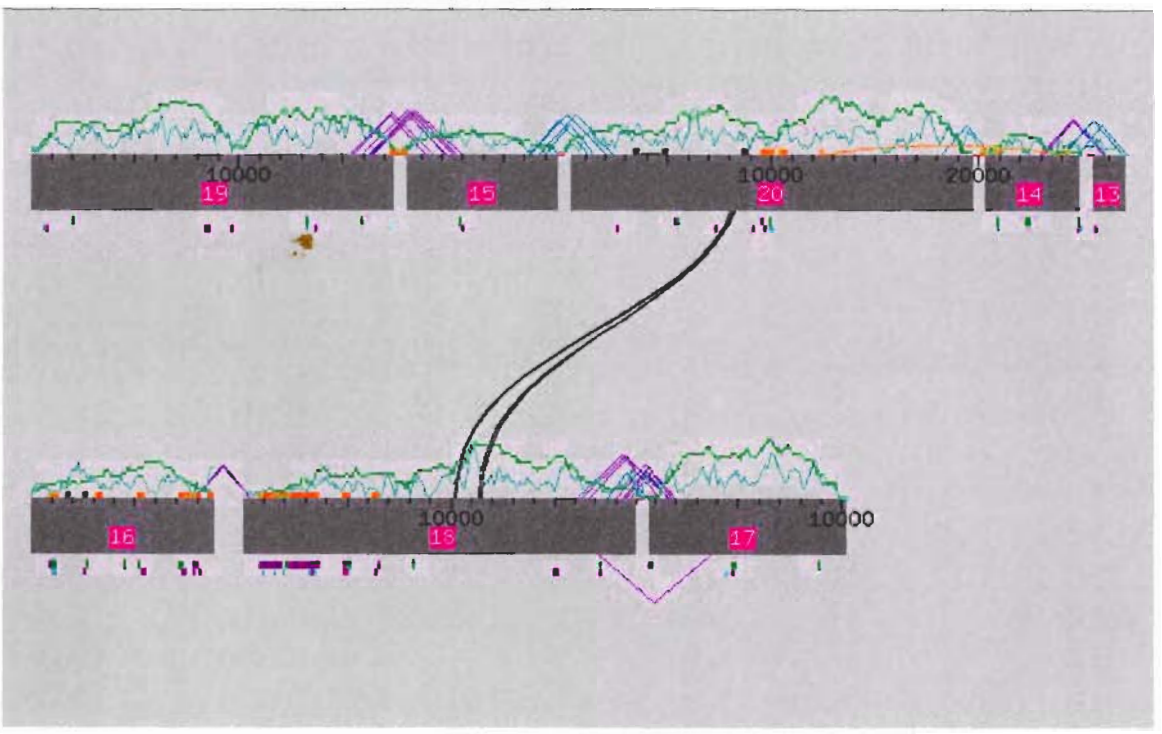
A single colony of clone S0493K22 was isolated, grown, confirmed by PCR and prepared using Quiagen's Large-Construct DNA kit, which isolates large-insert clone DNA while minimizing *E. coli* genomic contamination. BAC clone DNA was subsequently sheared by sonication, end-repaired and a 2 to 5 Kb fragment size range was isolated using agarose gel electrophoresis. The fragment mix was then ligated into Sma I digested, phosphatase treated pUC19, and transformed via heat-shock into XL1-blue super-competent *E. coli* cells. The transformed cells were then plated onto ampicillin containing agar plates, using X-gal and IPTG in order to be able to carry out blue-white selection of insert containing hybrid plasmids.

Twenty-four transformed clones were then selected for sequencing in order to obtain an estimate of possible *E. coli* genomic contamination. No *E. coli* genomic contamination was detected within the 24 sequenced clones. 1544 library clones were successfully sequenced at the sequencing centre at the University of Victoria. Clone S0493K22 is predicted to be approximately 91 Kb in length, according to the Hind III digest in FPC, and 1544 total sequences of 500 bp average read length represents an 8.5 fold depth of coverage.

3.4.1.1 Contig assembly using Phred/Phrap and Consed software

The Phred/Phrap/Consed suite of programs was used to assemble the sequence data into contigs (Ewing et al, 1998; Gordon et al., 1998). The initial build using Phrap yielded eight contigs that were collapsed into two super-contigs based on forward and reverse read consistency. The total sequence length of the two super contigs is 90,943 bp, which is quite near the total clone insert length of (91,818 bp) as determined by addition of Hind III fragment sizes in iCE version 3.4. The resulting two super-contig build as displayed using Consed is shown in Fig. 3.13. Hereafter the larger super-contig will be referred to as contig 19-13 (based on the Phrap assembly) and the smaller super-contig will be referred to as contig 16-17.

Figure 3.13 Build of clone S0493K22 as viewed using Assembly View in Consed. Contigs are represented by the thick dark-grey rectangles with numbers indicating sequence length in base pairs. Purple and red lines represent forward and reverse read pairs that are consistent in distance from one another when on top of the contigs, and inconsistent when below. Orange lines represent tandem repeats while black lines represent inverted repeats. Finally sequence depth is represented by the bright green line above contigs while the dark green line represents forward and reverse read pair consistency.



3.4.2 Sequence annotation

Potential genes were identified by BLASTn sequence alignment of the two super-contigs against both the GenBank non-redundant (NCBI) and the Rainbow trout and Atlantic salmon EST (TIGR, http://www.tigr.org/tdb/tgi/) databases in conjunction with Genscan (Burge and Karlin, 1997; http://genes.mit.edu/GENSCAN.html) gene predictions. The sequence contained a total of seven transcripts as predicted by Genscan (four in contig 19-13 and three in contig 16-17).

Each Genscan predicted peptide sequence was tBLASTn aligned against the GenBank non-redundant database as well as the TIGR Rainbow trout and Atlantic salmon Gene Indices in order to look for evidence to support the gene predictions. I annotated the Genscan predictions that produced at lease one alignment from the non-redundant database, combined with supporting EST data, at a minimal threshold expect value of e-15 or lower (in each database). The annotated Genscan outputs for contigs 19-13 and 16-17 can be found in Figs. 3.14 and 3.15, respectively.

Figure 3.14 Annotated Genscan output of contig 19-13. Predicted peptides are represented as blue arrows which are divided into exons. The tail-end of arrows represent the first exon of the gene, while arrowheads represent the gene's terminal exon. Genes above the ruled line are located on the sequence's plus strand, while genes below the line are on the minus strand. Annotated gene names appear above the predicted genes.

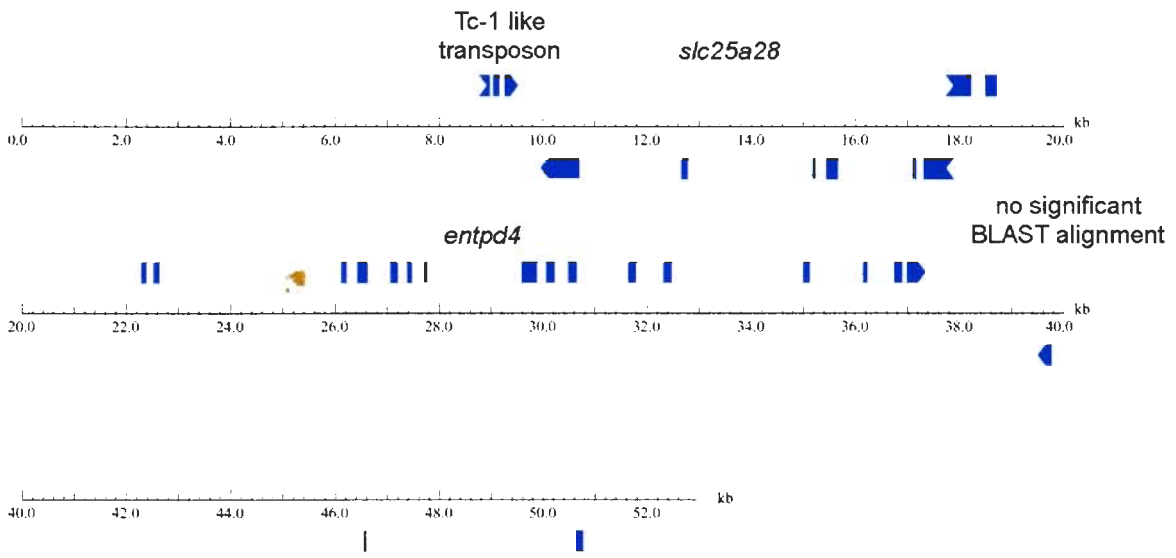
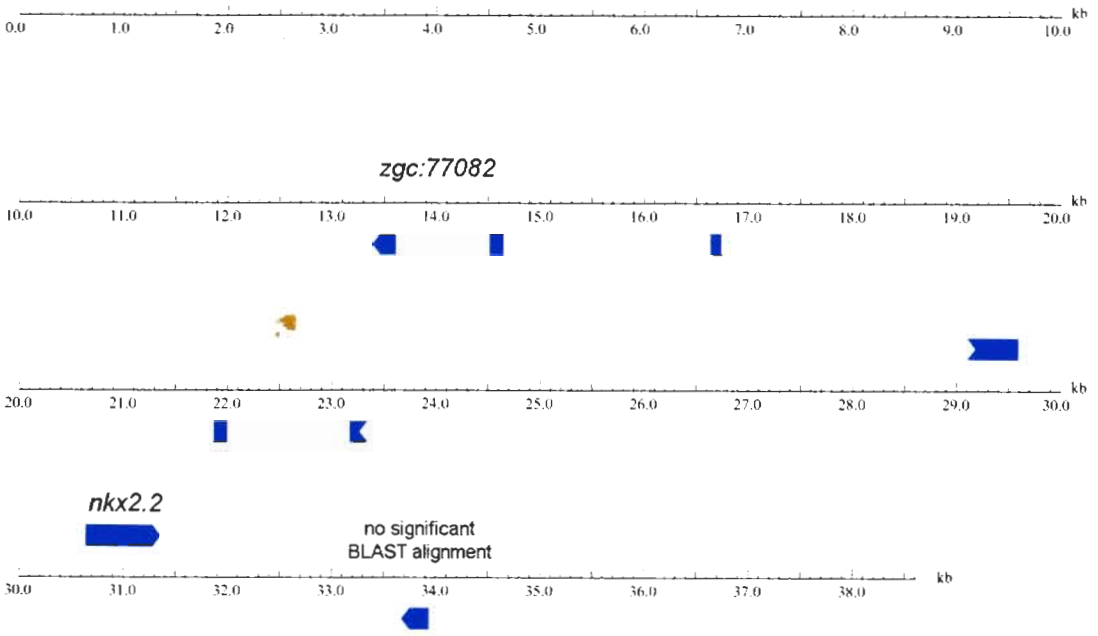


Figure 3.15 Annotated Genscan output of contig 16-17. Predicted peptides are represented as blue arrows which are divided into exons. The tail-end of arrows represent the first exon of the gene, while arrowheads represent the gene's terminal exon. Genes above the ruled line are located on the sequence's plus strand, while genes below the line are on the minus strand. Annotated gene names appear above the predicted genes.



3.4.2.1 Gene descriptions

The following is a description of each of the genes identified in clone S0493K22.

3.4.2.1.1 Slc25a28

Solute carrier family 25, member 28, *slc25a28* (GenBank Accession # NM_145156; Ensembl Gene ID # ENSMUSG00000040414), is an integral membrane protein located in the inner membrane of the mitochondria. In mice, it acts as a mitochondrial RNA splicing protein and suppresses a mitochondrial splice defect in the first intron of the *cob* gene. It may act as a carrier, exerting its suppressor activity via modulation of solute concentrations in the mitochondrion (possibly of cations).

3.4.2.1.2 Entpd4

Ectonucleoside triphosphate diphosphohydrolase 4, *entpd4* (GenBank Accession # NM_004901; Ensemble Gene ID # ENSG00000197217), is an integral golgi membrane protein that preferentially hydrolyzes nucleoside 5'-diphosphates; nucleoside 5'-triphosphates are hydrolyzed only to a minor extent. In addition Uracil 5'-diphosphate is preferentially hydrolyzed over Guanine, Cytosine or Thymidine 5'-diphosphate. Neither Adenosine 5'-di, nor triphosphate can act as substrates for the protein.

3.4.2.1.3 Zgc:77082

Zgc:77082 (GenBank Accession # BC066708; Ensembl Gene ID # ENSDARG00000018944) is a predicted protein in zebrafish, *Danio rerio*. This

prediction is based on either GeneWise or GenScan gene predictions of zebrafish genomic DNA and is supported by full-length cDNA obtained from a zebrafish EST library. Its function is unknown.

3.4.2.1.4 Nkx2.2

Homeobox protein Nkx-2.2, *nkx2.2* (GenBank Accession # AAC83132; Ensembl Gene ID # ENSG00000125820), is a transcription factor that promotes oligo-dendrocyte formation in conjunction with another transcription factor, *olig2*. *Nkx2.2* is has been implicated in spinal cord and central nervous system development in humans.

3.4.3 Gene synteny analysis

I analyzed genes orthologous to the four discovered on clone S0493K22 for evidence of conservation of synteny among ancestral relationships, in order to determine the correct orientation of the two super-contigs to one another. The results of this analysis are shown in Table 3.7. Orthologs and orthologous gene positions were determined using Ensembl's Genome Browser (www.ensembl.org).

Table 3.7 Syntenic relationship of genes identified on clone S0493K22. Genes are noted with their respected tBLASNn expect values (e-value) for the non-redundant (nr) GenBank database and the TIGR Gene Indices (EST). The nr e-value is that of the highest scoring hit to the gene itself, whereas the EST e-value is that of the highest scoring EST. The chromosome on which the homologous locus is found in *Danio rerio*, *Tetraodon nigroviridis*, mouse, human and *Drosophila melanogaster* is indicated.

				Homologous Locus Position				
Gene	**Genbank Accession Number**	**tBLASTn e-value (nr)**	**tBLASTn e-value (EST)**	Danio	Tetraodon	Mouse	Human	D. melanogaster
slc25a28	NM_145156	1e-110	1.3e-89	Chr 18	Chr 17	Chr 19	Chr 10	Chr 3R_24
Entpd4	NM_004901	3e-73	9e-92	Chr 17	Chr 12	Chr 14	Chr 8	N/A
zgc:77082	BC066708	2e-53	7.6e-64	Chr 11	N/A	Chr 19	Chr 10	N/A
nkx2.2	AAC83132	4e-89	2e-50	Chr 17	Chr 10	Chr 2	Chr 20	Chr X

There appears to be conservation of synteny between genes *slc25a28* and *zgc:77082* between Atlantic salmon and both mouse and human. In addition, synteny appears to be conserved between genes *entpd4* and *nkx2.2* between Atlantic salmon and *Danio*. Unfortunately, this information does not allow me to collapse super contigs 19-13 and 16-17 as suggesting that *entpd4* and *nxk2.2* are adjacent to one another due to shared synteny in *Danio* contradicts the suggestion the *slc25a28* is adjacent to *zgc:77082* based on shared synteny in human and mouse. Traditional sequence finishing would be required to accurately collapse the two super-contigs.

3.4.4 Repetitive element content analysis

In order to estimate the repetitive element content of clone S0493K22, the two super-contig sequences were run through the Atlantic salmon Repeat Masker database (S. Ng, unpublished results). In addition, the Atlantic salmon specific repeat masked sequence was run through the *Danio*, *Fugu* and other fish specific repeat databases (www.repeatmasker.org).

A total of 27,471 base pairs were masked due to Atlantic salmon specific repeats. In addition, 2,379 base pairs were masked from the additional fish databases. Therefore, clone S0493K22 appears to be 32.8 percent repetitive (29,850 repetitive bases / 90,943 total bases * 100). This number is comparable to the percent repetitive numbers seen in other fully sequenced Atlantic salmon genome insert containing BACs (Mitchell, 2004).

CHAPTER 4 DISCUSSION

The objective of this study involved the integration of Atlantic salmon microsatellite genetic linkage maps associated with the sex-determining factor with a Hind III fingerprinted genomic physical map as well as Atlantic salmon chromosomes. This allowed me to identify the sex chromosome pair in Atlantic salmon and characterize the sex chromosome itself as well as refine the sex-determining region.

4.1 Position of *SEX* on the Y-chromosome of Atlantic Salmon Based on the Genetic Linkage Map

All genetic linkage maps integrated in the study placed the *SEX* locus on the distal end of the *SEX* linkage group, ASL1 (Fig. 3.1) (R. Danzmann, unpublished results; B. Hoyheim, unpublished results). It is known that distal ends of linkage groups do not necessarily coincide with the telomeric ends of chromosomes (Woram et al., 2003); however, there is evidence supporting a more telomeric placement of *SEX*. According to Wright Jr. et al.'s (1983) model of chromosome pairing, only the telomeric segments of homologously paired multivalents may recombine during male meiosis, due to the physical constraints imposed by such multivalent pairing. Additionally, increased male recombination distances towards the ends of several linkage groups in the linkage maps of rainbow trout (Sakamoto et al., 2000) and brown trout (Gharbi, 2001) indicate that higher recombination rates near the telomeres of chromosomes may

represent a common trend in male meiosis among all salmonids (Gharbi, 2001). The male ASL1 map may demonstrate poor marker resolution due to the recombination constraints imposed by multivalent pairing. There is however, recombination between Ssa202DU (the microsatellite marker closest to *SEX*) and SEX itself (4.3 centiMorgans as shown in Fig. 3.1), which suggests that *SEX* is near the telomere, away from the constraints of multivalent pairing.

4.2 Integration of ASL1 with the Hind III Digested Physical Map

4.2.1 BAC library screening

All sixteen microsatellite markers on ASL1 amplified PCR products using Atlantic salmon genomic DNA as the template. However, hybridization probes designed to the flanking regions of markers Sal1UoG, Ssa208DU, BHMS247 and OMM1016 failed to identify any positive BAC clones within the 18.8 fold genomic coverage BAC library. It is likely that these microsatellite markers are located in genomic regions not represented within the BAC library. Such unrepresented regions may contain large numbers of repetitive elements, leading to a lack of EcoR I cut sites, preventing them from being incorporated into the library. Current work by Ng et al. (submitted) has identified BAC clones not incorporated into the physical map, as they are refractive to Hind III digestion and therefore, fingerprinting. Characterization of such "Hind III refractive" BAC clones has identified tandemly repeating sequences spanning these BACs, in which there are little or no Hind III cut sites. Markers Sal1UoG and Ssa58DU are adjacent to

one another on the linkage map. Similarly, markers BHMS247 and OMM1016 are also adjacent to one another. This suggests that these regions of the sex chromosome may contain repetitive elements, within which there are little or no EcoR I cut sites. Generation of an additional BAC library based on partial digestion with a different restriction enzyme may allow the cloning of such unrepresented areas in the CHORI-214 large-insert BAC genomic library. A Hind III fingerprinted physical map of such a library could result in the collapse of contigs in the current physical map separated by such non-represented regions.

4.2.2 Incorporation of *SEX*-linked BACs into the physical map

The common ancestor to all salmonids underwent a tetraploidization event, and extant species of salmonids are still undergoing the "rediploidization" process of restoring disomic inheritance (Allendorf and Thorgaard, 1984). Six of the microsatellite loci used to screen the BAC library were known to be duplicated from linkage analysis and the amplification of products from the loci located on both homeologous chromosome pairs (Table 1) (R. Danzmann, unpublished results; B. Hoyheim, unpublished results). Probing of the BAC library however, revealed evidence of duplication at three additional loci, Ssa208DU, One102ADFG and OMM1122, as their probes identified clones in separate, non-collapsible contigs. The Oneu18ASC marker identified two positive singletons; however, these singletons could not be collapsed into a contig, suggesting that they were each from a different genomic region. Markers BHMS447 and OmyFGT8TUF hybridized only to clones within a single contig each. Nevertheless, this does not necessarily contradict the observation of duplication

during the linkage analysis. Three possible explanations for the lack of identification of separate non-collapsible BAC clones are: A) The sequences surrounding each of the duplicate loci are in the process of diverging such that the region in which the original PCR primers designed in the linkage analysis is conserved, whereas the homeologous sequence has diverged significantly; B) That the regions of sequence surrounding the duplicate loci are conserved, such that they retain similar Hind III cut sites, and thus are being built into the same contig by FPC; or C) It is possible that the genomic segment surrounding one of the two loci is not represented in the Hind III fingerprinted portion of the BAC library.

4.3 Extension of Coverage of the Sex Chromosome

4.3.1 *In silico* contig extension via FPC

Performing end-to-end joins from terminal contig clones allowed me to collapse seven out of twelve (60%) of the sex-linked contigs into larger super-contigs. These contigs were not collapsed in the original FPC build of the Atlantic salmon physical map as the cut-off stringency of this build was set quite high (FPC likelihood value of E-16 or better) in order to ensure that the contigs of the physical map were reliably assembled (Ng et al., submitted). However, multiple clones from one contig producing end-to-end joins to another contig at a likelihood value of E-9 or lower are most likely significant (J. Schein, personal communication)

Although the FPC version 7.0 software is able to determine the nucleotide length of individual clones within a contig, it is unable to accurately determine the relative amount of overlap between individual clones within a contig, or between clones from contigs that have been end-to-end joined. Therefore, it is not possible to get an accurate estimate of how many nucleotides of sequence along the sex chromosome have been identified through genetic and physical map integration.

Markers Oneu18ASC and Ssa406UOS identified only singleton BACs, which could not be collapsed into contigs via end-to-end joining. In the case of Oneu18ASC, one of the two clones for which it was positive, S0119E21, demonstrated a Hind III fingerprint of only eight bands, which is likely too few to correctly incorporate the clone's fingerprint into the FPC build of the physical map. The lack of bands produced by Hind III digestion suggests that this clone has a high content of repetitive elements.

In the case of Ssa406UOS, singleton S0068C06 was flagged as "non-digestible" by Bandleader (Fuhrmann et al., 2003). Bandleader is a software program used to screen Hind III fingerprint patterns for the presence of clones that demonstrate unsuitable banding patterns either due to lack of Hind III cut-sites within the clone, or due to clone well contamination during fingerprinting. These data, taken with the fact that the clones that did produce usable banding patterns were not built into contigs, are suggestive that both of these microsatellite markers lie within regions of the genome that are resistant to incorporation into the BAC library. These regions may also simultaneously be

refractive to Hind III digestion. As discussed previously, such regions may arise due to the presence of repeating units within the sequence of these areas of the genome.

4.3.2 BAC-end sequencing of contig terminal clones

4.3.2.1 BAC-end sequencing

Sequences obtained using the pTARBAC2.1 vector's T7 sequencing primer were considerably longer and of higher quality than those obtained using the SP6 primer. Previous BAC-end sequencing efforts using genomic large-insert BAC libraries, ligated into the pTARBAC2.1 vector, have also met with similar reduced quality sequences using the SP6 primer (J. Stott, personal communication). Recently, we designed a new "SP6" primer upstream of the universal primer on the pTARBAC2.1 vector and employed it in a large BAC-end sequencing project (C. Artieri and S. Ng, unpublished results). The new primer sequence is as follows, in 5' to 3' orientation: ACTGTGGCTTGTTTTACAATTT. Sequences obtained from this new "SP6" primer were of comparable quality to those obtained from the T7 primer, indicating that the poor-quality sequences were due to poor design of the original SP6 primer.

4.3.2.2 Identification of novel genes

All five genes identified from sex-linked clones (Table 3.3) were identified from sequences obtained using the T7 sequencing primer. It is not surprising that no genes were identified from SP6 derived sequences, considering their poor quality.

Interestingly, three of the genes were obtained from end sequences of clones within contig 3784, which is associated with microsatellite marker Alu032NVH. This may suggest that this contig represents a gene-rich area of the sex chromosome; however, my sample size of sequence is small and not uniformly distributed along the length of the chromosome. Therefore, such a hypothesis is purely speculative. Nevertheless, microsatellite markers are often identified from ESTs, and therefore it is likely that genes would be found in or near microsatellite marker containing clones.

The lack of observable shared synteny among the five identified genes within the organisms used in the comparison (Atlantic salmon, zebrafish, pufferfish, mouse and human) is not unexpected as these genes are found along the length of the Atlantic salmon sex-chromosome. Many chromosomal rearrangements have occurred between Atlantic salmon and its sister species in the Salmonidae (Hartley and Horne, 1984); therefore, given the divergence times between Atlantic salmon and human (approximately 450 mya) or Atlantic salmon and *Danio* (approximately 200 mya) (Kumar and Hedges, 1998; Nei and Glazko, 2002; Van de Peer, 2004), it is not unexpected that such chromosomal rearrangement events have broken the synteny between these genes.

Work performed by Mitchell (2004) has demonstrated a surprising amount of conservation of synteny between Atlantic salmon and other fishes such as *Danio* and even mammals with respect to a group of genes all located within a single Atlantic salmon BAC clone. Therefore, I may have expected to have observed shared synteny between the three genes found in contig 3784

(Ubiquitin Conjugating Enzyme, Lamin and the predicted protein RKHD1); however, it appears that chromosomal rearrangement events have shuffled even these relatively nearby genes.

4.3.3 Chromosome-walking from sex-linked contigs

4.3.3.1 Probe design

The hybridization probes designed to extend the physical map coverage of the sex chromosome suffered from a high-degree of failure. Only 20 hybridization probes out of the 45 designed produced scorable hybridization patterns when used to probe the high-density BAC clone spotted membranes. The reason behind the failure of many of these probes may lie in the method by which the probes were designed.

Primer sets were designed from BAC-end sequence based on which primer sequences would result in optimal PCR primers using Primer Premier version 5.0. Subsequently one of the two approximately 20 nucleotide PCR primers was extended to approximately 40 nucleotides in order to create a hybridization probe of suitable annealing temperature. I may have achieved a greater success rate during hybridization if I had employed a strategy involving generating optimal hybridization primers independently of separately designed PCR primers. Additionally, due to the large volume of primers being designed, I did not inspect each of the BAC-end sequence chromatograms. It is possible that errors in PHRED software automated base-calling occurred (Ewing et al., 1998), resulting in lack of sequence identity between hybridization primers and the actual genomic sequence.

4.3.3.2 BAC library probing results

4.3.3.2.1 Primer sets that failed to amplify PCR products

Eleven out of the 45 primer sets (31 percent) designed from BAC-end sequences failed to amplify PCR products from the BACs from which they were designed. Five out of the failed primer sets were designed from sequences obtained using the pTARBAC2.1 T7 sequencing primer while the remaining six were designed from sequences obtained using the SP6 primer. I may have expected to see more SP6 sequence designed primer sets fail, rather than T7, as the SP6 sequences were generally of poorer quality. As PCR amplification failure occurred equally in sequences obtained using both primers, it is likely that the cause of primer failure lies in the primer design itself. A more stringent selection of acceptable BAC-end sequences suitable for primer design may have resulted in less failed primer pairs.

4.3.3.2.2 Primer sets that failed to produce appropriate hybridization signals

Three out of the 34 hybridization probes used to probe the BAC library did not give any hybridization signal. Additionally, eight hybridization probes used to probe the BAC library failed to hybridize to the BAC clones from which they were designed. These results are particularly suggestive of incongruence of sequence between the hybridization primers and the actual sequence of the clone from which they were designed. It is likely that I would have observed additional instances of non self-hybridization had I probed all high-density BAC clone spotted membranes with hybridization probes designed from BAC-end sequence of clones not located within the first six membranes.

The three hybridization probes that gave repetitive hybridization signals reveal that there are still many uncharacterized repetitive elements within the Atlantic salmon genome. The Atlantic salmon-specific repeat database (S. Ng, unpublished results) failed to detect the presence of repetitive elements within the sequences of any of these probes. It is crucial that more potential repetitive elements be identified if better probes are to be designed for use in future hybridization studies.

4.3.3.2.3 Primer sets that successfully hybridized

Twenty out of the original 45 primer sets (44.4 percent) designed to chromosome walk from the ends of sex-linked contigs produced scorable hybridization signals. I chose to employ stringent criteria for what could be considered a suitable potential positive hybridization signal before selecting such clones for PCR confirmation. Clones were only considered potentially positive, and thus suitable for PCR confirmation if: A) The clone was the only clone positive by hybridization within a contig and that clone had significant regions free of possible overlapping or buried clones in FPC; and B) Multiple, overlapping clones within a contig were positive by hybridization at which point each of them was selected for PCR confirmation. Large numbers of false positive clones were observed during the original screening of the BAC library that identified the microsatellite positive clones and contigs (Table 3.1). Therefore, in the interest of managing the large number of potential positive clones, such selection criteria were adopted.

Hybridization probes were designed for both ends (T7 and SP6) of contig terminal BAC clones whenever possible. In the case of such terminal BACs, I would have expected to observe one of the two probes hybridizing to clones within the contig, while the other probe hybridizes to the clone itself, as well as any potential contig-contig joins (Fig. 3.10).

Chromosome "walking" is only possible in the event that two contigs contain clones with some degree of overlapping sequence, and that A) this overlap is insufficient in length to produce enough common Hind III fingerprint bands, or that B) the sequence of this overlapping region is deficient in Hind III digestion sites. In both cases, there would be an insufficient number of Hind III fingerprinting bands produced for the FPC software to collapse the two contigs. It would be unlikely to observe a situation in which both T7 and SP6 sequence designed probes hybridize to the same clone in a potential contig-contig join as in this case, the entirety of one clone's sequence would be buried within the other clone and I would expect FPC to have collapsed these clones *in silico*.

Eight of the scored probes that neither hybridize to clones within their own contigs, nor identified any contig-contig joins (S0360O08T7, S0042N21SP6, S0220F20T7, S0213H09T7, S0205P08T7, S0232D13T7, S0581B12SP6 and S0116A05SP6), are likely probes designed from contig terminal sequence for which no potential contig-contig joins are represented within the BAC genomic library. Given the conditions necessary for chromosome "walking", it is likely that I would have observed contig terminal end-sequence probes hybridizing only to themselves without revealing any new contig-contig joins.

Hybridization probes S0295I12T7, S0230N16SP6, S0360O08SP6, S0179B16SP6, S0205P08SP6 and S0382A05T7 are examples of the "green" probe in Fig. 3.7 as they hybridize to themselves as well as additional clones within the contig in which they lie. Such hybridization signals allow the orientation of these clones relative to their own contigs such that I now know that the opposite BAC-ends of these clones are representative of contig-terminal sequence, and thus are the BAC-ends from which potential chromosome "walking" will occur.

Hybridization probe S0612P18T7 (positive for the presence of the Ssa202DU microsatellite marker) collapsed clones S0493K22, S0534O16 and S0552D23 into a contig. Like clone S0612P18 itself, clones S0534O16 and S0552D23 are outside of the Hind III fingerprinted portion of the BAC genomic library. Clone S0493K22; however, is a singleton within the physical map with a Hind III fingerprint containing 17 bands. Two lines of evidence suggest that the Ssa202DU marker lies in a region of the sex chromosome that is rich in the presence of repetitive elements. First, probing of the entire 18.8 fold genomic coverage BAC library has identified only three Ssa202DU positive clones (S0581B12, S0605H01 and 612P12) which would indicate that the region containing the Ssa202DU marker is underrepresented in the BAC library and second, clone S0493K22, the only Hind III fingerprinted, Ssa202DU associated clone, has a Hind III fingerprint of only 17 bands, which is far below the library's average number of bands per clone of 40 (Ng et al., submitted). Both these data taken together suggest that there is a scarcity of both EcoR I and Hind III

digestions sites in this region of the genome, which previous work (Ng et al., submitted) has shown to be indicative of regions containing repetitive segments.

Hybridization probe S0042N21T7, which is found in contig 5898 and is positive for the presence of the Ssa208DU marker, hybridized to clones within contig 2133, also positive for the presence of the Ssa208DU marker. Microsatellite marker Ssa208DU is known to be duplicated within the Atlantic salmon genome as it is found in ASL1 and ASL6 (R. Danzmann, unpublished results; B. Hoyheim, unpublished results). Therefore, it was not unexpected to find that the hybridization probe designed for the flanking region of the Ssa208DU marker hybridizes to two separate non-collapsible contigs, suggesting that these contigs each represent one of the two homeologous Ssa208DU loci. It was unexpected, however, to observe that contigs 5898 and 2133 join to each other via *in silico* end-to-end joining in FPC, albeit at the minimum threshold I considered significant in this study (FPC likelihood value of e-9). This *in silico*, potential contig-contig join was confirmed via the results of the BAC-end sequencing; however, despite the fact that these results suggest that the Ssa208DU marker is tandemly duplicated along the sex chromosome, they are perhaps equally suggestive of a phenomenon that will have a far greater impact on future studies involving Atlantic salmon genomics. It is possible that contigs 5898 and 2133 do in fact represent the separate homeologous loci of the Ssa208DU marker, but that genomic regions surrounding these two loci have not diverged sufficiently in sequence to be discriminated. As previously discussed, salmonid fishes are still in a state of pseudo-tetraploidy, and are thus attractive

organisms in which to study the consequences of whole genome duplication. However, to perform such studies, it will be imperative that we be able to differentiate between the duplicated loci. A future attempt to sequence the entire Atlantic salmon genome via the whole-shotgun approach could run into difficulties at the assembly stage if the software is unable to differentiate between the two loci, due to high similarity of sequence. It must not be ignored; however, that at the stringency at which the physical map was built (FPC likelihood value of e-16) these two Ssa208DU loci were placed into two separate, non-collapsible contigs. This is indicative of sufficient Hind III restriction pattern divergence such that future studies will be manageable.

The case of hybridization probe S0370K19T7 is another potential example of conservation of sequence similarity surrounding duplicate loci of a microsatellite marker. Probe S0370K19T7 identified clones within contig 5234 (in which it is found) as well as contig 3284. Both contigs have been identified as positive for the duplicated One102ADFG marker, and are believed to represent each of the duplicate loci. *In silico* BAC end-to-end joining analysis does not suggest a potential contig-contig join of the two One102ADFG contigs, which may be indicative that the homeologous One102ADFG loci possess a greater level of sequence divergence than those of the Ssa208DU loci. However, these results suggest that BAC-end sequence designed probes are still able to hybridize across loci, and therefore there is a significant level of sequence conservation between homeologous sections of the Atlantic salmon genome.

The collapse of singleton clone S0116A05, positive for the Ssa406UOS microsatellite marker, into contig 5324 is demonstrative of the utility of the chromosome "walking" technique. The initial physical coverage of this Ssa406UOS locus was limited to a single clone; however, through chromosome "walking" I have extended this coverage into a large contig. Maximizing the amount of physical coverage around microsatellite loci will be crucial in facilitating any future marker-assisted analyses performed on the Atlantic salmon genome.

4.3.3.2.3.1 Analysis of clone S0503M23 in contig 2133

As previously indicated, neither S0503M23T7 nor S0503M23SP6 hybridization probes identified any positive clones when used to screen the first six high-density spotted membranes of the genomic BAC library (Table 3.6). However, both sets of primers did successfully amplify PCR products from clone S0503M23 (Table 3.5), suggesting that the end-sequences from which they were designed were identical to that of clone S0503M23. The results of the amplification of PCR products from overlapping clones using the S0503M23T7 primers (Fig. 3.10) indicates that I should have observed hybridization positives for these same probes during BAC library screening with the S0503M23T7 probe. It is important to note that the hybridization T_m is more stringent than that of the PCR (65°C in the case of hybridization and 56.5°C during PCR). It is possible that at the hybridization T_m, the S0503M23T7 probe was unable to anneal to the clones from which PCR products were produced. Repeating the hybridization using a less stringent T_m could validate this hypothesis.

4.4 Fluorescent in situ Hybridization of Sex-Linked Clones

4.4.1 Identification of the Atlantic salmon sex chromosome

I report here the first successful identification of the Atlantic salmon sex chromosome. Representative sex-linked microsatellite marker containing BACs all hybridized via FISH to chromosome two, the second largest meta-centric.

In all cases, both chromosome two homologues gave positive FISH signals, which support the hypothesis that salmonid sex chromosomes are representative of an early stage in sex chromosome evolution, such that they still retain similar sequence, and similar gene compliments. Peichel et al.'s (2004) study of the sex chromosomes of the threespine stickleback found that despite being homomorphic, the sex-chromosomes have begun differentiating in sequence, such that they retain only approximately 63.7 percent sequence identity. Given that Atlantic salmon also has homomorphic sex chromosomes and that microsatellite markers found on the Y chromosome are also found on the X (as is the case in threespine stickleback), I would expect to see similar evidence that Atlantic salmon is in the initial stages of sex-chromosome differentiation. However, validation of such a hypothesis will require first, the identification of sex-specific markers and second, sequencing of male and female chromosome-specific BACs representing a homologous region.

4.4.2 Orientation of the linkage map relative to the sex chromosome

Comparative analysis of the position of hybridized sex-linked BAC clones along the sex chromosome with their relative positions along the integrated sex-linkage map (R. Danzmann, unpublished results; B. Hoyheim, unpublished

results), has allowed me to confirm the reliability of the linkage map (Fig. 3.8). Such validation of the positional relationship of microsatellite markers is vital for future studies performed using these genetic linkage maps. In addition, understanding the positional relationship of microsatellite markers along the sex chromosome also allows us to localize the coverage of the chromosome within the Hind III fingerprinted physical map.

This analysis has also allowed me to orient the linkage map relative to the sex chromosome. According to the integrated linkage map, the Ssa202DU microsatellite marker is closest to *SEX* (Fig. 3.1). FISH localizes the Ssa202DU containing BAC on the long arm of chromosome two, quite close to the centromere (Fig. 3.7B).

4.4.2.1 Possible locations of *SEX* on the Atlantic salmon sex chromosome

Interpretation of these data suggests three possibilities for the placement of *SEX* on the sex chromosome of Atlantic salmon, each requiring a different bevy of strategies and techniques necessary for its elucidation:

1) *SEX lies between the Ssa202DU marker and the region of heterochromatin*

As shown in Fig. 3.7B, DAPI staining has revealed a large area of heterochromatin, spanning approximately half of the long arm of the sex-chromosome. Tightly condensed heterochromatic DNA is known to be relatively gene poor (Wallrath, 1998). In addition, previous studies have shown that genes found within DAPI stained heterochromatin may lack active transcription (Schulz

and Tyler, 2005), which may suggest that the primary sex-determining factor is located within the euchromatic region of the sex chromosome. The region of heterochromatin appears to span the entire distance from the middle of the long arm of the sex chromosome to the telomere. Therefore, *SEX* may lie near the region of heterochromatin, between it and the Ssa202DU marker. In this case, the chromosome "walking" technique may be employed in order to walk along the euchromatic segment of the sex chromosome, assuming that this entire region is represented within the CHORI-214 Atlantic salmon genomic BAC library. Unfortunately, the correlation between recombination distance (both male and female) and nucleotide distance in Atlantic salmon is unknown and therefore, although it is known that *SEX* lies approximately 4.4 centiMorgans away from the Ssa202DU marker according to the male ASL1 map (Fig. 3.1), we are currently unable to estimate the corresponding distance that would be necessary to "walk" along the sex chromosome. This distance could be impractical to breach using the chromosome "walking" technique, and some estimate of the actual distance would most likely be required before attempting such a study.

2) *SEX lies near the telomere, on the other side of the region of heterochromatin*

As previously discussed, various lines of evidence suggest that the sex-determining factor in Atlantic salmon, brown trout and Arctic charr is located on a Y chromosome specific telomeric segment of the sex chromosome in these species (Woram et al., 2003). If this is the case in Atlantic salmon, the Ssa202DU marker is most likely too far away from the telomere of the long-arm of

chromosome two for the chromosome "walking" technique to be employed as a method for identifying the sex-determining locus. Additionally, regions of tightly condensed heterochromatic DNA are often associated with tandem satellite repeats and are likely underrepresented in the genomic BAC library (for review see Dimitri et al., 2004), thus making chromosome walking impossible. In such a scenario, identification of the primary sex-determining locus would almost certainly require the identification of markers more tightly associated with *SEX*, preferably even sex specific markers such as the growth hormone pseudogene in rainbow trout (Du et al., 1993). Localization of such a marker to the telomere of the sex chromosome would provide a starting point from which chromosome "walking" could be attempted, without the caveat of having to "walk" across regions of heterochromatin.

3) *SEX lies within the large region of heterochromatin*

As nothing is known about the primary genetic mechanism of sex-determination in salmonid species, it is entirely possible that the primary sex-determining locus lies within the heterochromatic segment of the sex chromosome. It also remains entirely possible that *SEX* is near the telomere, such that the telomeric model of salmonid *SEX* loci is valid. Under this scenario, it is quite likely that the primary genetic sex-determining factor is not represented within the genomic BAC library, and that techniques such as chromosome "walking" will be ineffective. In such a case, alternative methodologies would have to be employed, such as screening EST libraries for candidate sex-

determining genes based on known motifs (e.g., the DM domain of the sex-determining gene in medaka; Matsuda et al., 2002). A large number of salmonid cDNA libraries have been developed as part of the GRASP project, using mRNA isolated from various tissues and developmental stages (Rise et al., 2004). Despite this, it is possible that such libraries do not contain any ESTs representative of the primary-genetic sex-determining factor in Atlantic salmon as such genes in other non-fish vertebrate species are known to have narrow temporal and tissue-specific expression windows, during male embryogenesis (Jeske et al., 1995; Payen et al., 1996; Meyers-Wallen, 2003). Recently, however, it has been found that the primary sex-determining factor of the Japanese medakafish is expressed in various tissues throughout the life of male medaka (Ohmuro-Matsuyama et al., 2003). Since DMY is the only primary sex-determining signal identified in any teleost fish to date, this may suggest that fish as a group may demonstrate such a non-temporal and tissue specific pattern of expression, and that the currently available cDNA/EST libraries do in fact contain the sex-determining factor, waiting only to be identified.

4.4.3 Fluorescent *in situ* hybridisation (FISH) of duplicate loci

FISH analysis of non-collapsible clones each positive for the same microsatellite marker (Ssa406UOS or Oneu18ASC) has revealed that these clones represent duplicate loci within the Atlantic salmon genome. These loci are likely conserved remnants of the original ancestral salmonid tetraploidization event (Allendorf and Thorgaard, 1984).

Unfortunately, as previously discussed, I am unable to differentiate between Atlantic salmon chromosomes of similar shape and size. The chromosomes of fish in general do not display G-banding patterns (Ocalewicz et al., 2003). In addition, C-banding patterns, such as those seen through DAPI staining (which stains heterochromatin), have provided little help in differentiating between many of the medium-sized metacentric chromosomes found in the Atlantic salmon genome. Therefore I am unable to determine whether the non-sex linked duplicated loci of Ssa406UOS (clone S0116A05) and Oneu18ASC (clone S0119E21) are both located on the same chromosome. Elucidation of this situation could be achieved through dual-hybridization; that is, hybridizing both duplicate loci on the same chromosome spread, each in a different colour. Using such a technique, it would be evident whether or not they both hybridized to the same chromosome.

Such information would be useful in determining whether the duplicated homeologous portions of the sex chromosome remain in large segments, or if chromosomal rearrangement events have interspersed duplicate loci throughout the genome. Linkage data (R. Danzmann, unpublished results; B. Hoyheim, unpublished results) indicate that the non-sex linked duplicate markers, BHMS447 and OmyFGT8TUF, are both in linkage group twelve, whereas the non-sex linked duplicate marker, Ssa208DU, is in linkage group 6 (Table 3.1). The FISH results performed in this study (Fig. 3.7) indicate that the OmyFGT8TUF and BHMS447 markers are located on the short arm of the sex chromosome. The Ssa208DU marker is located between the BHMS150 and

Ssa202DU markers (Fig. 3.1) both of which are located on the long-arm of the sex chromosome (Fig. 3.7). Taken together, these data suggest that while there have been portions of the ancestral genomic segments that make up the current sex chromosome conserved in large homeologous segments, there has been at least some chromosomal rearrangement, perhaps with respect to the sex chromosome's arms. However, it is impossible to tell at this point whether the current, complete sex chromosome, or the separate chromosomal segments represent the ancestral state of these portions of the genome. Comparative analysis of the co-linearity of conserved sex-chromosomal markers between Atlantic salmon and a closely related, non-salmonid ancestor (such as the Northern pike, *Esox lucius*) could allow us to determine what was the ancestral state of the Atlantic salmon sex chromosome, and perhaps provide information as to how the genome has reordered itself after the tetraploidization event.

The observation that Ssa406UOS hybridized to only one chromosome of both pairs for which it is positive (Fig. 3.9) may suggest that the marker is only located on one pair of the two homologous chromosomes. Furthermore, clone S0068C06 hybridized to only one of two different sex chromosomes, this may suggest that the Ssa406UOS marker clone contains a sex-specific element. It is possible that clone S0068C06 hybridized more strongly to one chromosome rather than the other (perhaps because it possesses more copies of a repetitive element than its homolog, as a result of somatic cell unequal crossing over); however, it is unlikely that the region surrounding Ssa406UOS contains any sex-specific elements. Linkage analysis (Fig. 3.1) has shown that the Ssa406UOS is

far removed from the male sex-determining factor; therefore, it is unlikely that such an element would not have recombined on to both the X and Y chromosomes, such that sex specificity could no longer be observed.

4.5 Analysis of BAC clone S0493K22

4.5.1 Sequence location

Clone S0493K22 was identified through chromosome "walking" off the end of clone S0612P22. However, I am unable to determine whether this clone lies closer to or further from *SEX*. Regardless, even in the event that clone S0493K22 lies closer to SEX than clone S0612P22, it is highly unlikely that the *SEX* locus itself would be found within it. According to the male ASL1, *SEX* is 4.4 centiMorgans away from the Ssa202DU marker. Although we are currently unable to correlate genetic distance to physical distance in base pairs within Salmonids, it has been estimated that 1 centiMorgan in humans, corresponds to roughly 1,000,000 base pairs (http://www.genome.iastate.edu/edu/doe/prim2.html). Therefore, it would not be unreasonable to suggest a similar correlation in Salmonids, given that both species have similar genomic sizes (3.3 giga base pairs in humans and 3.0 giga base pairs in Atlantic salmon). Even assuming that this estimate is wrong by an order of magnitude, 4.4 centiMorgans still corresponds to well over 92,000 base pairs, which is the approximate size of clone S0493K22.

4.5.2 Sequence annotation

A previous study by Mitchell (2004) demonstrated near perfect conservation of synteny within a wide variety of taxa among 10 genes located within a single BAC clone of approximately 200 Kb in length. The four genes identified in clone S0493K22 however, are not well conserved in synteny. It is unlikely that an increased number of chromosomal rearrangement events, leading to broken synteny from an ancestral state, are influenced by the presence of these genes on the sex chromosome. It is known that chromosomes homologous to Atlantic salmon chromosome two are not the sex chromosome in other salmonids such as brown trout, rainbow trout or Arctic charr (Woram et al., 2003). The telomeric model of sex chromosome evolution within the Salmonids would suggest that Atlantic salmon chromosome two is not necessarily the ancestral Salmonid sex chromosome, and thus the recombination event leading to broken synteny among these genes could have occurred on an ancestral autosome.

The three genes whose functions have been identified through sequence similarity in other organisms (*slc25a28*, *entpd4* and *nkx2.2*) are not known to have a role in sex-determination. Though the function of predicted protein *zgc:77082* is unknown, it does not bear sequence similarity to any known sex-determining genes. This, in addition to the results discussed above, suggests that it is an unlikely candidate as the primary sex-determining factor.

4.5.3 Analysis of repetitive DNA content

Accumulation of transposable elements as well as repetitive DNA is predicted to occur on the Y chromosome (Charlesworth, 1991). Unlike the study of threespine stickleback performed by Peichel et al. (2004) which demonstrated that the Y chromosome had accumulated significantly more repetitive elements than the X, the lack of sex-specific elements within the sequence of S0493K22 has made it impossible for me to determine whether the clone comes from the X or Y chromosome. However, the percentage of repetitive DNA within the clone's sequence, 32.8 percent, is not higher than that seen in previous analyses of fully sequenced BAC clones (Mitchell, 2004). This suggests two possibilities: A) That clone S0493K22 is still too far away from *SEX* to have begun showing signs of sex-specific chromosome degeneration, or B) that clone S0493K22 represents sequence on the X chromosome, and thus has not undergone significant repetitive DNA accumulation from that seen elsewhere within the genome.

Finally, heterochromatic regions are known to mostly consist of repetitive DNA sequences which may be essential to their function (Wallrath, 1998). As previously discussed, the long arm of Atlantic salmon chromosome two (which bears the *SEX* locus) contains a large segment of heterochromatin, as revealed by DAPI staining (Fig. 3.10). As clone S0493K22 does not contain an above average percentage of repetitive DNA, it is unlikely that the Ssa202DU marker is significantly near the heterochromatic region. This is also supported by the fluorescent *in situ* hybridization data (Fig. 3.10). This information therefore does not rule out the possibility that the *SEX* locus lies between the Ssa202DU marker

and the large region of heterochromatin as there appears to be a significant area of DNA between the Ssa202DU region and the beginning of the heterochromatic region.

REFERENCES

Allendorf FW and GH Thorgaard. 1984. *Tetraploidy and the evolution of salmonid fishes*, pp. 1-53 in Evolutionary Genetics of Fishes, edited by BJ Turner. Plenum Press, New York.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science.* **297**:1301-1310.

Arkhipchuk VV. 1995. Role of chromosomal and mutations in the evolution of bony fishes. *Hydrobiol. J.* **31**:55-65.

Bell G. 1982. The Masterpiece of Nature: The Evolution and Genetics of Sexuality. University of California Press, Berkeley. 635 p.

Berta P, Hawkins JR, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. 1990. Genetic evidence equating SRY and the testis-determining factor. *Nature.* **348**:448-450.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW. 2003. The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell.* **114**:419-429.

Brunner B, Hornung U, Shan Z, Nanda I, Kondo M, Zend-Ajusch E, Haaf T, Ropers HH, Shima A, Schmid M, Kalscheuer VM, Schartl M. 2001. Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for DMRT1. *Genomics.* **77**:8-17.

Bull JJ. 1983. Evolution of sex determining mechanisms. Benjamin/Cummings publishing, Don Mills. 316 p.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J. of Mol. Biol.* **268**: 78-94.

Carroll L. 1872. Through the looking glass and what Alice found there. Macmillan, London. 240 p.

Charlesworth B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55:**199-221.

Charlesworth B. 1991. The evolution of sex chromosomes. *Science.* **251:**1030-1033.

Chevassus B. 1998. "Modification du phenotype sexuel et du mode de reproduction chez les poisons salmonidés: Inversion sexuelle, gynogénèse, hybridation interspécifique et polypoïdisation." Ph D. dissertation, Université de Paris XI, Orsay, France.

Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **21:**1146-1151.

Devlin RH, MacNeil BK, Groves TDD, Donaldson EM. 1991. Isolation of a DNA Y-chromosomal probe capable of determining genetic sex in Chinook salmon (*Oncorhynchus tshawytscha*). *Can. J. Fish Aquat. Sci.* **48:**1606-1612.

Devlin RH, Stone GW, Smailus DE. 1998. Extensive direct-tandem organization of a long repeat DNA sequence on the Y chromosome of Chinook salmon (Oncorhynchus Tshawytscha). *J. Mol. Evol.* **46:**277-287.

Devlin RH, Biagi CA, Smailus DE. 2001. Genetic mapping of Y-chromosomal DNA markers in Pacific salmon. *Genetica.* **111:**43-58.

Devlin RH, Nagahama Y. 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological and environmental influences. *Aquaculture.* **208:**191-364.

Dimitri P, Corradini N, Rossi F, Verni F. 2004. The paradox of functional heterochromatin. *Bioessays.* **27:**29-41.

Du SJ, Devlin RH, Hew CL. 1993. Genomic structure of growth hormone genes in chinook salmon (Oncorhynchus tshawytscha): presence of two functional genes, GH-I and GH-II, and a male-specific pseudogene, GH-psi. *DNA Cell. Biol.* **12:**739-51

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

Fjell CD, Bosdet I, Schein JE, Jones SJ, Marra MA. 2003. Internet Contig Explorer (iCE)--a tool for visualizing clone fingerprint maps. *Genome Res.* **13:**1244-1249.

Forbes SH, Knudsen KL, North TW, Allendorf FW. 1994. One of two growth hormone genes in coho salmon is sex-linked. *Proc. Natl. Acad. Sci. U. S. A.* **91**:1628-1631.

Fuhrmann DR, Krzywinski MI, Chiu R, Saeedi P, Schein JE, Bosdet IE, Chinwalla A, Hillier LW, Waterston RH, McPherson JD, Jones SJ, Marra MA. 2003. Software for automated analysis of DNA fingerprinting gels. *Genome Res.* **13**:940-953.

Gharbi K. 2001. "Construction d'une carte génétique partielle du génome tétraploide de la truite commune (*Salmo trutta*)". Ph.D. dissertation, Institut National Agronomique Paris-Grignon, Paris, France.

Gilbey J, Verspoor E, McLay A, Houlihan D. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Anim. Genet.* **35**:98-105.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195-202.

Hardie DC, Gregory TR, Hebert PD. 2002. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J. Histochem. Cytochem.* **50**:735-749.

Hardie DC, Hebert PD. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome.* **46**:683-706.

Hartley SE, Horne MT. 1984. Chromosome relationships in the genus *Salmo. Chromosoma.* **90**:229-237.

Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **59**:190-203.

Howard RS, Lively CM. 1998. The maintenance of sex by parasitism and mutation accumulation under epistatic fitness functions. *Evolution.* **52**:604-610.

Hunter GA, Donaldson EM, Stoss J, Baker II. 1983. Production of monosex female groups of chinook salmon (*Oncorhynchus tshawytscha*) by the fertilization of normal ova with sperm from sex-reversed females. *Aquaculture.* **33**:355-364.

Hunter GA, Donaldson EM, Goetz FW, Edgell PR. 1982. Production of all female and sterile Coho salmon, and experimental evidence of male heterogamety. *Trans. Am. Fish Soc.* **111**:367-372.

Iturra P. Medrano JF, Bagley M, Lam N, Vergara N, Marin JC. 1998. Identification of sex chromosome molecular markers using RAPDs and fluorescent in situ hybridization in rainbow trout. *Genetica.* **101**:209-213.

Iturra P, Lam N, de la Fuente M, Vergara N, Medrano JF. 2001. Characterization of sex chromosomes in rainbow trout and coho salmon using fluorescence in situ hybridization (FISH). *Genetica.* **111**:125-131.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. 2004. *Nature.* **431:**946-957.

Jeske YW, Bowles J, Greenfield A, Koopman P. 1995. Expression of a linear Sry transcript in the mouse genital ridge. *Nat. Genet.* **10:**480-482.

Johnstone R, Simpsons TH, Youngson AF, Whitehead C. 1979. Sex reversal in salmonid culture. Part II. The progeny of sex reversed rainbow trout. *Aquaculture.* **18:**13-19.

Johnstone R, Youngson AF. 1984. The progeny of sex-inverted female Atlantic salmon (*Salmo salar* L.). Aquaculture. **37:**179-182.

Keightley PD, Eyre-Walker A. 2000. Deleterious mutations and the evolution of sex. *Science.* **290:**331-333.

Keightley PD, Eyre-Walker A. 2001. Response to Kondrashov. *Trends. Genet.* **17:**77-78.

Kondo M, Nanda I, Hornung U, Asakawa S, Shimizu N, Mitani H, Schmid M, Shima A, Schartl M. 2003. Absence of the candidate male sex-determining gene dmrt1b(Y) of medaka from other fish species. *Curr. Biol.* **13:**416-420.

Kondrashov AS. 1993. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84:**372-387.

Kumar S, Hedges B. 1998. A molecular timescale for vertebrate evolution. *Nature.* **392:**917-920.

Lively CM, Loyd DG. 1990. The cost of biparental sex under individual selection. *Am. Nat.* **135:**489-500.

Marshall Graves JA, Shetty S. 2001. Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *J. Exp. Zool.* **290:**449-462.

Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, Hori H, Hamaguchi S, Sakaizumi M. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417:**559-563.

Matsuda M, Sato T, Toyazaki Y, Nagahama Y, Hamaguchi S, Sakaizumi M. 2003. Oryzias curvinotus has DMY, a gene that is required for male development in the medaka, O. latipes. *Zoolog. Sci.* **20:**159-161.

May B, Johnson KR, Wright JE Jr. 1989. Sex linkage in salmonids: evidence from a hybridized genome of brook trout and Arctic charr. *Biochem. Genet.* **27**:291-301.

McElreavey K, Vilain E, Abbas N, Herskowitz I, Fellous M. 1993. A regulatory cascade hypothesis for mammalian sex determination: SRY represses a negative regulator of male development. *Proc. Natl. Acad. Sci. U.S.A.* **90**:3368-3372.

McGowan C, Davidson WS. 1998. The RAPD technique fails to detect a male-specific marker in Atlantic salmon. *J. Fish Biol.* **53**:1134-1136.

Meyers-Wallen VN. 2003. Sry and Sox9 Expression During Canine Gonadal Sex Determination Assayed by Quantitative Reverse Transcription-Polymerase Chain Reaction. *Mol. Reprod. Dev.* **65**:373-381.

Mitchell L. 2004. "The Fate of Duplicated Regions of the Atlantic Salmon (*Salmo salar*) Genome". M.Sc. dissertation. Simon Fraser University, Burnaby, Canada.

Moen T, Hoyheim B, Munck H, Gomez-Raya L. 2004. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim. Genet.* **35**:81-92.

Moran P, Martinez JL, Garcia-Vazquez E, Pendas AM. 1996. Sex chromosome linkage of 5S rDNA in rainbow trout (*Oncorhynchus mykiss*). *Cytogenet Cell Genet.* **75**:145-150.

Nagai K. 2001. Molecular evolution of Sry and Sox gene. *Gene.* **270**:161-169.

Nakayama I, Biagi CA, Koide N, Devlin RH.1999. Identification of a sex-linked GH pseudogene in one of two species of Japanese salmon (Oncorhynchus masou and O. rhodurus). *Aquaculture.* **173**:65-72.

Nei M, Glazko GV. 2002. The Wilhelmine E. Key 2001 Invitational Lecture. Estimation of divergence times for a few mammalian and several primate species. *J. Hered.* **93**:157-164.

Nelson, JS. 1994. *Fishes of the World, 3rd Edition.* Wiley and Sons, New York.

Nichols KM, Young WP, Danzmann RG, Robison BD, Rexroad C, Noakes M, Phillips RB, Bentzen P, Spies I, Knudsen K, Allendorf FW, Cunningham BM, Brunelli J, Zhang H, Ristow S, Drew R, Brown KH, Wheeler PA, Thorgaard GH. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet.* **34**:102-115.

Ocalewicz K, Jankun M, Luczynski M. 2003. Homologous chromosomes characteristics by sequential banding procedures in rainbow trout *Oncorhynchus mykiss*. *Folia. Biol.* (Krakow). **51**:41-46.

Ohmuro-Matsuyama Y, Matsuda M, Kobayashi T, Ikeuchi T, Nagahama Y. 2003. Expression of DMY and DMRT1 in various tissues of the medaka (Oryzias latipes). *Zoolog. Sci.* **20**:1395-1398. Erratum in *Zoolog. Sci.* **20**:1491.

Ohno S, U Wolf, and NB Atkin. 1968. Evolution from fish to mammals by gene duplication. *Hereditas.* **59:**169-187.

Ohno S. 1970. 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.

Onozato H. 1989. Androgenesis in masu salmon. *Physiol. Ecol. Japan. Spec.* **1:**543.

Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, Catanese JJ, de Jong PJ. 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics.* **52:**1-8.

Osoegawa K, Zhu B, Shu C-L, Ren T, Cao Q, Vessere GM, Lutz MM, Jensen-Seaman MI, Zhao S, de Jong P. 2004. BAC resources for the rat genome project. *Gen. Res.* **14:**780-785.

Otto SP, Nuismer SL. 2004. Species interactions and the evolution of sex. *Science.* **304:**1018-1020.

Ottolenghi C, Veitia R, Barbieri M, Fellous M, McElreavey K. 2000. The human doublesex-related gene, DMRT2, is homologous to a gene involved in somitogenesis and encodes a potential bicistronic transcript. *Genomics* **64**:179–186.

Ottolenghi C, McElreavey K. 2000. Deletions of 9p and the quest for a conserved mechanism of sex determination. *Mol. Genet. Metab.* **71**:397–404.

Ottolenghi C, Fellous M, Barbieri M, McElreavey K. 2002. Novel paralogy relations among human chromosomes support a link between the phylogeny of doublesex-related genes and the evolution of sex determination. *Genomics.* **79:**333-343.

Payen E, Pailhoux E, Abou Merhi R, Gianquinto L, Kirszenbaum M, Locatelli A, Cotinot C. 1996. Characterization of ovine SRY transcript and developmental expression of genes involved in sexual differentiation. Int. J. Dev. Biol. **40:**567-575.

Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D, Kingsley DM. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* **14:**1416-1424.

Phillips R, Ihssen PE. 1985. Chromosome banding in salmonid fish: nucleolar organizer regions in Salmo and Salvelinus. *Can. J. Genet. Cytol.* **27:**433-440.

Phillips RB, Reed KM. 2000. Localization of repetitive DNAs to zebrafish chromosomes using multi-color fluorescence *in situ* hybridization. *Chromosome Res.* **8:**27-35.

Phillips RB. 2001. Application of fluorescence *in situ* hybridization (FISH) to fish genetics and genome mapping. *Marine Biotechnol.* **3:**5145-5152.

Phillips R, Ráb P. 2001. Chromosome evolution in the Salmonidae (Pisces): an update. *Biol. Rev.* **76:**1-25.

Phillips RB, Konkol NR, Reed KM, Stein JD. 2001. Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus*, *Salmo* and *Salvelinus* (Salmonidae). *Genetica.* **111:**119-123.

Phillips RB, Matsuoka MP, Reed KM. 2002. Characterization of charr chromosomes using fluorescence in situ hybridization. *Env. Biol. Fishes.* **64:**223-228.

Pomiankowski A, Nothiger R, Wilkins A. 2004. The evolution of the Drosophila sex-determination pathway. *Genetics.* **166:**1761-1773.

Potts, GW and Wooton, RJ. 1984. *Preface* pp. vii-viii in Fish Reproduction: Strategies and Tactics, edited by GW Potts and RJ Wooton. Academic Press, Toronto. 410 p.

Prodöhl PA, Taggart JB, Ferguson A. 1994. Single locus inheritance and joint segregation analysis of minisatellite (VNTR) DNA loci in brown trout (*Salmo trutta L.*) *Heredity.* **73:**556-566.

Raymond CS, Shamu CE, Shen MM, Seifert KJ, Hirsch B, Hodgkin J, Zarkower D. 1998. Evidence for evolutionary conservation of sex-determining genes. *Nature.* **391:**691-695.

Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. 2000. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14:**2587-2595.

Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SR, Smailus DE, Jones SJ, Schein JE, Marra MA, Butterfield YS, Stott JM, Ng SH, Davidson WS, Koop BF. 2004. Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res.* **14:**478-90.

Reed KM, Phillips RB. 1995. Molecular cytogenetic analysis of the double CMA3 Chromosome in lake trout, *Salvelinus namaycush*. *Cytogenet. Cell. Genet.* **70:**104-107.

Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm LE, Guyomard R, Hoyheim B. 2000. A microsatellite linkage map of rainbow trout (Oncorhynchus mykiss) characterized by large sex-specific differences in recombination rates. *Genetics.* **155:**1331-1345.

Sambrook, J, Fritsch EF, Maniatis T. 1999. Molecular Cloning: A Laboratory Manual, 2nd. Ed. Cold Spring Harbor Laboratory Press, New York.

Schein JE, Kucaba TA, Sekhon M, Smailus D, Waterston RH, Marra MA. 2004. High-throughput BAC fingerprinting. In *Methods in molecular biology*, Vol

255. *Bacterial artificial chromosomes: Library construction, physical mapping and sequencing* (eds. S. Zhao and M. Stodolsky), pp. 143-156. Humana Press Inc., Totawa, NJ.

Scherer G. 2002. *The molecular genetic jigsaw puzzle of vertebrate sex determination and its missing pieces*, pp. 225-239 in The Genetics and Biology of Sex Determination, edited by D Chadwick and J Goode. John Wiley & Sons Ltd, West Sussex. 266 p.

Schulz L, Tyler J. 2005. Heterochromatin focuses on senescence. *Mol. Cell.* **17:**168-170.

Schütt C, Nöthiger R. 2000. Structure, function and evolution of sex-determining systems in Dipteran insects. *Development.* **127:**667-677.

Smit, AFA, Hubley, R, Green, P. 1996-2004. *RepeatMasker Open-3.0.* <http://www.repeatmasker.org>.

Smith CA, Sinclair AH. 2004. Sex determination: insights from the chicken. *Bioessays.* **26:**120-132.

Soderlund C, Longden I, Mott R. 1997. FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13:**523-35.

Stein J, Phillips RB, Devlin RH. 2001. Identification of the Y chromosome in chinook salmon (*Oncorhynchus tshawytscha*). *Cytogenet Cell Genet.* **92:**108-110.

Stein JD, Reed KM, Phillips RB. 2002. Isolation and characterization of a sex-linked microsatellite locus from the lake trout sex chromosome. *Cytogenet. Cell Genet.* **92:**108-110.

Stothard P, Pilgrim D. 2003. Sex-determination gene and pathway evolution in nematodes. *BioEssays.* **25:**221-231.

Swain A. 2002. Vertebrate Sex Determination: A New Player in the Field. *Curr. Biol.* **12:**602-603.

Taggart JB, Prodöhl PA, Ferguson A. 1995. Genetic markers for Atlantic salmon (*Salmo salar L.*): Single locus inheritance and joint segregation analyses of minisatellite (VNTR) DNA loci. *Anim. Genet.* **26:**13-20.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nuc. Acids Res.* **24:**4876-4882.

Thorgaard GH. 1977. Hetermorphic Sex Chromosomes in Male Rainbow Trout. *Science.* **196:**900-902.

Thorgaard GH. 1983. Chromosomal differences among rainbow trout populations. *Copeia.* **3:**650-662.

Torres Maldonado LC, Landa Piedra A, Moreno Mendoza N, Marmolejo Valencia A, Meza Martínez A, Merchant Larios H. 2002. Expression profiles of *Dax1*, *Dmrt1*, and *Sox9* during temperature sex determination in the

gonads of the sea turtle *Lepidochelys olivacea*. *Gen. Comp. Endocrin.* **129**:20-26.

Van de Peer Y. 2004. Computational approaches to unveiling ancient genome duplications. *Nature Reviews, Genetics*. **5**:752-763.

Van Valen L. 1973. A new evolutionary law. *Evol. Theory* **1**:1-30.

Vilain E. 2002. *Anomalies of human sexual development: clinical aspects and genetic analysis*, pp. 43-53 in The Genetics and Biology of Sex Determination, edited by D Chadwick and J Goode. John Wiley & Sons Ltd, West Sussex. 266 p.

Wallrath LL. 1998. Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**:147-153.

Whitler RE, McPhail JD, Devlin RH. 1985. Genetic variability in freshwater and anadromous sticklebacks (Gasterosteus aculeatus) of Southern British Columbia. *Can. J. Zool.* **63**:528-533.

Wilkins AS. 1995. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays.* **17**:71-77.

Woram RA, Gharbi K, Sakamoto T, Hoyheim B, Holm LE, Naish K, McGowan C, Ferguson MM, Phillips RB, Stein J, Guyomard R, Cairney M, Taggart JB, Powell R, Davidson W, Danzmann RG. 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* **13**:272-80.

Woram RA, McGowan C, Stout JA, Gharbi K, Ferguson MM, Hoyheim B, Davidson EA, Davidson WS, Rexroad C, Danzmann RG. 2004. A genetic linkage map for Arctic char (Salvelinus alpinus): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome.* **47**:304-15.

Wright Jr. JE, Johnson K, Hollister A, May B. 1983. Meiotic models to explain classical linkage, pseudolinkage and chromosome pairing in tetraploid derivative salmonid genomes. *Isozymes: Curr. Top. Biol. Med. Res.* **10**:239-260.

Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH. 1998. A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics.* **148**:839-850.

Zeng C, Kouprina N, Zhu B, Cairo A, Hoek M, Cross G, Osoegawa K, Larionov V, de Jong P. 2001. Large-insert BAC/YAC libraries for selective re-isolation of genomic regions by homologous recombination in yeast. *Genomics.* **77**:27-34.

Zhang Q, Nakayama I, Fujiwara A, Kobayashi T, Oohara, Masaoka T, Kitamura S, Devlin RH. 2001. Sex identification by male-specific growth hormone pseudogene (GH-psi) in Oncorhynchus masou complex and a related hybrid. *Genetica.* **111**:111-118