Mapping of sex-determining candidate genes in Atlantic salmon (Salmo salar)

Karlegging av kjønnsbestemmende kandidatgener i Atlantisk laks (Salmo salar)

Anders Mühlbradt

Norwegian University of Life Sciences
Department of animal and aquacultural sciences
Master Thesis 60 credits 2011

Preface

This work is a result of a final thesis project for a master's degree in molecular genomics at the Norwegian University of Life Sciences. The work was carried out at the Center for Integrative Genetics in Ås.

I would like to thank my advisor, Sigbjørn Lien, for all help and guidance throughout this thesis.

Thanks to Arne Roseth for the help in the laboratory and Kristil Kindem Sundsaasen for help with the TyperAnalyzer software. I would also like to thank Simon Taylor for all the help with the SNP detection and SAM tools software.

Finally, I want to thank my parents and my friends who have been of great support throughout the work on this thesis.

Ås, May 13., 2011

Anders Mühlbradt

Abstract

Salmo salar (Atlantic salmon) is a species of great economical importance and is also of scientific interest because of its tetraploid state.

Chromosomes 2 and 3 are proved to be the sex-determining chromosomes in Atlantic salmon and the purpose of this study was to map genes associated with sex-determination in other species to these chromosomes.

A total of 65 candidate genes were collected from articles regarding sex-determination and BLASTed against the salmon UniGene Build database and salmon Arachne 1.96x assembly. A SNP detection was performed followed by genotying of the SNPs.

29 of the genes were successfully mapped either by two-point linkage mapping or by BLAST search performed against the Salmon Celera 3.1x assembly, which contains scaffolds and markers that the genes can be aligned to. Out of these, the genes AM-DSX and RBP1 were mapped to sex-determining chromosome 3, while BRD3 was mapped to chromosome 2. The gene FOXL2 was mapped to chromosome 14, a chromosome homologous to chromosome 3. These four genes can be considered factors involved in sex-determination in Atlantic salmon based on their function and location, while the genes mapped to other chromosomes are ruled out.

Sammendrag

Salmo salar (atlantisk laks) er en viktig økonomisk art og er også av interesse innen forskning på grunn av at den er tetraploid.

Kromosom 2 og 3 er bevist å være de kjønnsbestemmende kromosomene I atlantisk laks. Målet med denne oppgaven var å kartlegge gener assossoiert med kjønns-bestemmelse til disse kromosomene.

Totalt 65 kandidatgener ble samlet fra artikler som omhandler kjønnsbestemmelse og søkt opp mot Salmon UniGene Build database og Salmon Arachne 1.96x assembliet ved hjelp av BLAST. En SNP deteksjon ble utført, fulgt opp av en genotyping av SNPene.

29 av genene ble kartlagt enten ved hjelp av two-point linkage kartlegging eller ved kartlegging gjort med hjelp av BLAST mot Salmon Celerea 3.1x assembliet som inneholder scaffold med markører som genene blir kartlagt til. Av disse genene, ble AM-DSX og RBP1 kartlagt til kromosom 3, mens BRD3 ble kartlagt til kromosom 2. FOXL2-genet ble kartlagt til kromosom 14, som er en paralog til kromosom 3. Disse fire genene kan bli sett på som faktorer involvert i kjønnsbestemmelse i atlantisk laks basert på deres funksjon og plassering, mens resten av genene kartlagt i denne oppgaven kan bli avskrevet som gener involvert i kjønnsbestemmelse i atlantisk laks.

Table of Contents

1. Introduction

1.1. Atlantic salmon

The Atlantic salmon (*salmo salar*) is a valued species in sport fishing and an important contributor to aquaculture production all over the world. In 2010, Norway alone exported aquacultured salmon to 96 different countries with a total value of 31.4 billion NOK (Eksportutvalgetforfisk 2011). The species is a member of the family Salmonidae, which consists of the three sub-families Coregoninae, Thymallinae and Salmoninae. The sub-family Salmoninae includes charr, trout and salmon, whereof Atlantic salmon is one of the species. Figure 1 presents an overview of the Salmonidae with the sub-families and the most important genera.

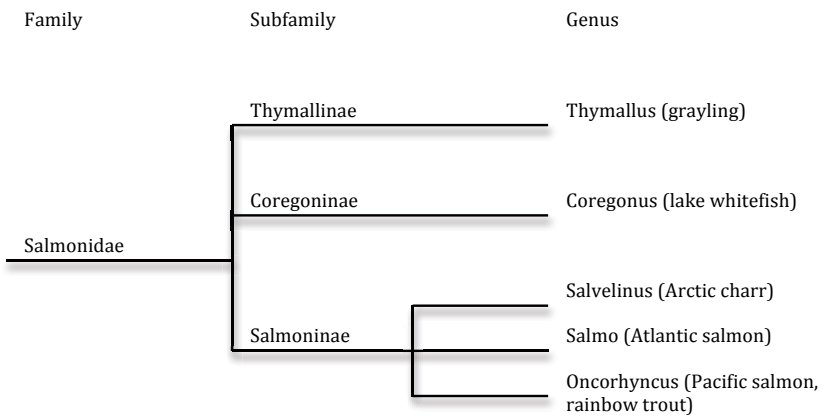

Figure 1: Overview of the family Salmonidae, with the 3 subfamilies and some of the most common genera within each of the subfamilies (Davidson et al. 2009).

Most of the fish found today belongs to the teleosts. Teleosts are very diverse and reproduce in several ways; they can carry live brood, lay eggs or release gametes into plankton. They have evolved in a relatively short period of time and dates back to about 200 million years (Desjardins & Fernald 2009).

The family Salmonidae is suspected to share an whole genome duplication in the early evolution of the family 25 – 100 million years ago (Allendorf & Thorgaard 1984). A proof

of this is that Salmonids have around double the amount of DNA as their closest relatives, which are Esociformes and Osmeriformes. Salmonidae average genome size is approximately almost 3000Mb, while the average genome size for Esociformes and Osmeriformes are approximately 1200Mb and 1900Mb (Gregory 2011). Since the genome duplication event, the genomes of salmonids have been reverting towards a diploid state by differentiating duplicate chromosome sets into distinct pairs of homeologs (Ohno et al. 1968).

A challenge concerning genome duplication is how to deal with sex determination. One way salmonids may have dealt with sex-determination is by deleting one copy of the sex determining locus, or by recruiting a duplicated transcription factor to become a novel sex determining gene(Force et al. 1999). It is likely to expect that different lineages implement different genes from what they originally had, something that will start speciation, but because the molecular basis for sex-determination is unsolved, it is not known which strategies for sex determination different salmonids have adopted.

As salmonids are at an early stage in sex chromosome differentiation they provide unique opportunities to reveal fundamental knowledge regarding evolution of sex-determination. Moreover, there is an interest in breeding all-female stocks of Atlantic salmon in aquaculture (Davidson et al. 2009) and information about sex-determining loci can help to separate males from females before they become fully developed. All-female stocks will cause consistent growth patterns and prevent genetic pollution within the wild salmon environment.

1.2. Chromosomes

A chromosome is a physical unit, which contains a DNA sequence with many genes. Each gene is located at a specific location on the chromosome, called a locus. A metacentric chromosome is made up of two chromatids with a long- and a short arm. The end of each chromosome arm is called telomere. These chromatids are connected in the centromere. All human chromosomes are metacentric. However, Atlantic salmon also has a different type of chromosome, called an acrocentric chromosome. An acrocentric chromosome has only one arm, with the centromere located at one end. Figure 2 shows

schematically how a typical metacentric chromosome in the metaphase is arranged into pairs, while figure 3 illustrate how an acrocentric chromosome will look like.

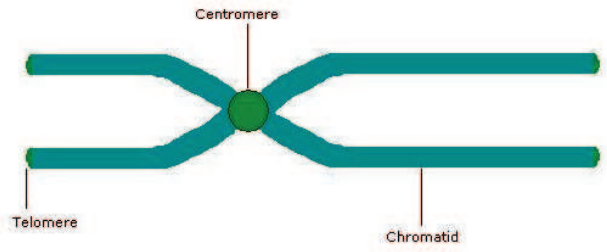

Figure 2: Paired metaphase chromosome. The two chromosomes are joined at the centromere. The arms of the chromosome are called chromatids, while the end of each chromatid is called telomeres (Brown 2006).

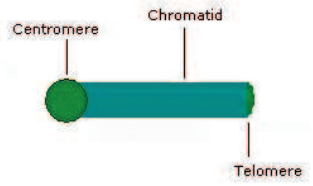

Figure 3: An acrocentric chromosome, with the centromere at the end of the chromosome.

Out of the 29 chromosomes in Atlantic salmon of European origin, only the eight first chromosomes are metacentric, while the rest are acrocentric. An overview of the chromosomes of Atlantic salmon is given in figure 4.

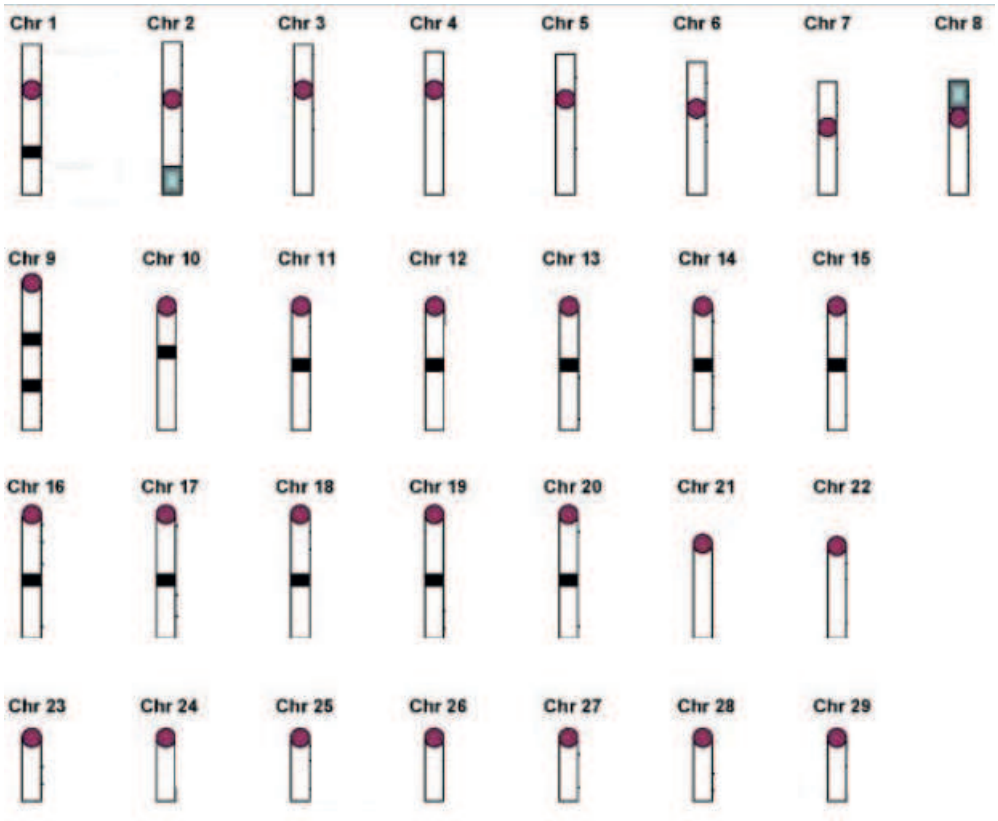

Figure 4. An overview of the chromosomes in the European Atlantic salmon. Out of 29 chromosomes, the 8 first chromosomes are metacentric, while the remaining are acrocentric. The figure is taken from (Phillips et al. 2009).

1.3. Genome duplication

Genome duplication occurs when an error during meiosis leads to production of gametes that are diploid instead of haploid. When two diploid gametes fuse, they will form a type of autopolyploid, which in the case of salmonids, will give a tetraploid cell. Autopolyploidy is not uncommon, and is especially frequent among plants. Autopolyploidy is a way for speciation to occur, because an autopolyploid can reproduce successfully since each chromosome has a homologous partner with which it can form a bivalent during meiosis (Brown 2006). Figure 5 shows how polyploidzation occurs.

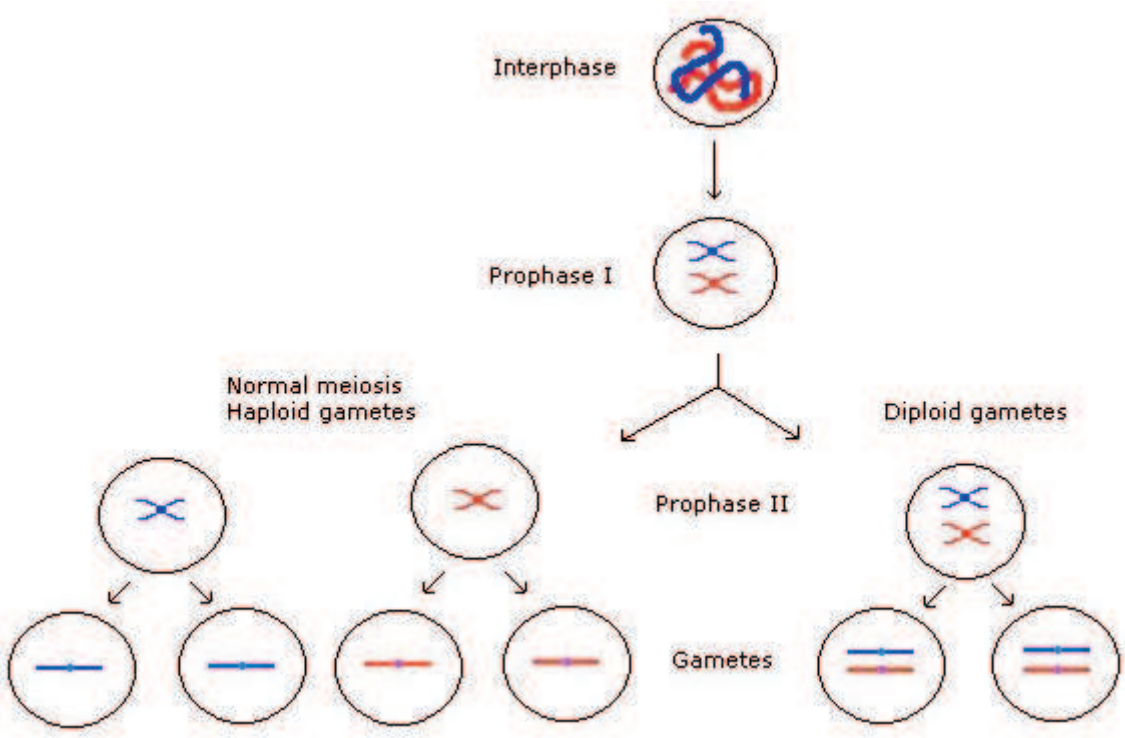

Figure 5: A haploid meiosis (left) versus a polyploid meiosis (right). With autopolyploidzation, an irregular event occurs between prophase I and prophase II that causes the pairs of homologous chromosomes to not be separated into different nuclei. The result of this is that the gametes will be diploid (Brown 2006).

Because of this tetraploid state, many genes in Atlantic salmon have paralogs. A paralog is a homologous sequence that has been separated by a gene duplication event, like the genome duplication that the common ancestor of salmonids went through. Paralog sequences do not have identical nucleotide sequences, but they are usually highly similarity, since they come from the same starting sequence. Two paralogues will be approximately 85% - 90% identical. The function of the two paralogs will most likely be the same, or at least similar (Brown 2006).

1.4. Sex determination and development

Sex determination is the process that establishes the sex of an organism. Sex determination can be strictly genetic, as in mammals and birds, but also strictly environmental, like in various reptiles, or genetic but reversible by environmental factors, like in many fish and amphibians (Stelkens & Wedekind 2010). Fish displays a

wide array of mechanisms determining sex, something that makes them attractive when it comes to studying sex determination and the evolution of sex chromosomes (Schultheis et al. 2009).

To understand the process of sex-determination it is important to examine the way cells and organs are involved in development of the gonad. The gonad is an organ that produces sex cells: either an ovary in females, or testis in males (Purves et al. 2004). As in most other vertebrates, fertilization and egg activation in fish is followed by repeated mitotic cell division. A blastula is produced, which is made up of cells with a broad development potential (Nilsson & Cloud 1992).The next stage is the gastrula stage, where the germ layers and specialized cells and tissue are developed through interactions and differentiation of the cells. The gonad region forms as a longitudinal thickening of mesoderm, which is one of the primary germ cell layers in the early development of an embryo, that goes in to the coelomic cavity ventral to the developing kidney and lateral to the dorsal mesentery (Devlin & Nagahama 2002).

The gonad is similar in fish as in other vertebrates, with germ cells, which are reproductive cells or gametes (Purves W K 2004),and somatic cells. Germ cells have the potential to divide mitotically and enter meiosis, while somatic cells differentiate into structural and endocrine cells. In fish, like in many other vertebrates, the developmental origin of germ cells and somatic cells differs. Germ cells are usually formed by cells developing around specialized granular cytoplasm, which is often found in vegetal regions of an unfertilized eggs (Wei & Mahowald 1994). This develops primordial germ cells, which later develops into regular germ cells. Somatic cells, on the other hand, derive from the genital ridge epithelium, or tissue. Before differentiation, somatic cells seem to be similar in what is expected to develop into males and females.

For development of an ovary, somatic cells and primordial germ cells differentiates to form follicles, which comprise of oocytes surrounded by an inner granulose and an outer thecal layer (Nagahama et al. 1982). Ovarian development can first be detected by profilation of some somatic cells and early oocyte differentiation before the ovarian cavity forms. Development of testis usually happens later than ovarian differentiation (Nakamura et al. 1998).

There are certain genes that control ovarian development, while others control testicular development. The sex of an individual is decided by the strength of the genetic factors it receives from its parents. However, the final sexual stage of a fish does not always reflect the initial gonadal pathway taken (Devlin & Nagahama 2002).

Salmonids are a gonochoristic species, which means that individuals belonging to salmonids can only have one of at least two distinct sexes. As soon as sex has been decided, the gonadal differentiation proceeds down a developmental pathway that develops either fully differentiated testes or ovary (Hawkins 1994). Different from gonochoristic species sex in several other fish species may be affected by environmental factors, such as temperature could affect sex. Pollution and hormones could also have an affect on sex in several fish species (Baroiller et al. 2009).

Even though the mechanism of sex determination in salmon is strictly genetic, salmon is able to be sex reversed before fully developed. Hormones like methyltestosterone and estradiol-17ß are factors that could change the sex phenotype in early development (Devlin & Nagahama 2002). This happens in the way that testosterone given after sexual differentiation can stop gametogenesis, which stops oogenesis in females and gives regression of the testis in males(Billard et al. 1982). The result of this is that a sex reversed female that mate with a normal female, would only have female offspring. On the other hand, if a sex reversed male mate with a normal male, only ¾ of the offspring will be male (Johnstone et al. 1979).The reason for this is that the sex chromosomes will not change even though the phenotypic sex changes. Figure 6 shows why a female mating with a sex-reversed female will always produce female offspring and why a male mating with a sex-reversed male will still produce a number of female offspring.

a.	X	X
X	XX	XX
X	XX	XX

b.	X	Y
X	XX	XY
Y	XY	YY

Figure 6: a. Only female XX can be produced from salmon with XX chromosomes. b. ¼ of the offspring will be XX-females when two XY salmons reproduce. The rest will be XY-males.

1.5. Sex genes in other species

In many mammals, the gene SRY, a dominant gene on the Y-chromosome, controls the male development switch (Gubbay et al. 1990). *SRY* is a High Mobility Group Box bearing gene, also known as a SOX gene, which has an important regulatory role in transcription. It is likely that the SRY gene has derived from another SOX gene, SOX3. SRY and SOX3 interact with SOX9 for testis differentiation in mammals (Graves 1998). In fish, however, different sex determining genes seems to be chosen in different species.

Even though the sex-determining region has been identified in several teleost species (Artieri et al. 2006) it is only in *Oryzias latipes* (medaka), that the sex-determining gene has been found and characterized. In 2002, Matsuda et al. identified DMY, in some studies also designated DMRT1, as the sex-determining gene (Matsuda et al. 2002). The DMY gene encodes a transcription factor with an intertwined Zn-finger DNA binding domain and is highly conserved through evolution (Schartl 2004). The approach used to find these genes, linkage mapping with BAC positional cloning of physical genomic libraries, has been used a number of times to try identify the sex determining gene of other fish species (Davidson et al. 2009).

1.6. Chromosomes 2 and 3 are sex chromosomes

Sex determining genes can be spread across the genome, located on certain chromosomes, or even restricted to a single locus (Devlin & Nagahama 2002). Several salmonid species have been shown to possess a genetic mechanism for sex

determination characterized by the XY chromosomal system, even though environmental factors such as temperature or hormones could affect the sex phenotype.

Chromosome 2 is proved to be the sex-determining chromosome (Artieri et al. 2006). By using probes from sex-linked microsatellite markers, Artieri et al. screened a bacterial artificial chromosome (BAC) library. The BACs containing sex-linked microsatellites and their contigs were identified and representative BACs were placed on the salmon chromosomes by using fluorescent in situ hybridization (FISH). This identified the a single locus mapping near a block of repetitive DNA at the en of the q- arm on chromosome 2 as a sex determining loci (Artieri et al. 2006).

However, recent studies of data from a large number of families from a Norwegian aquaculture strain suggests that also chromosome 3 is involved in sex determination, as shown in figure 7 (Lien et al. Subm).

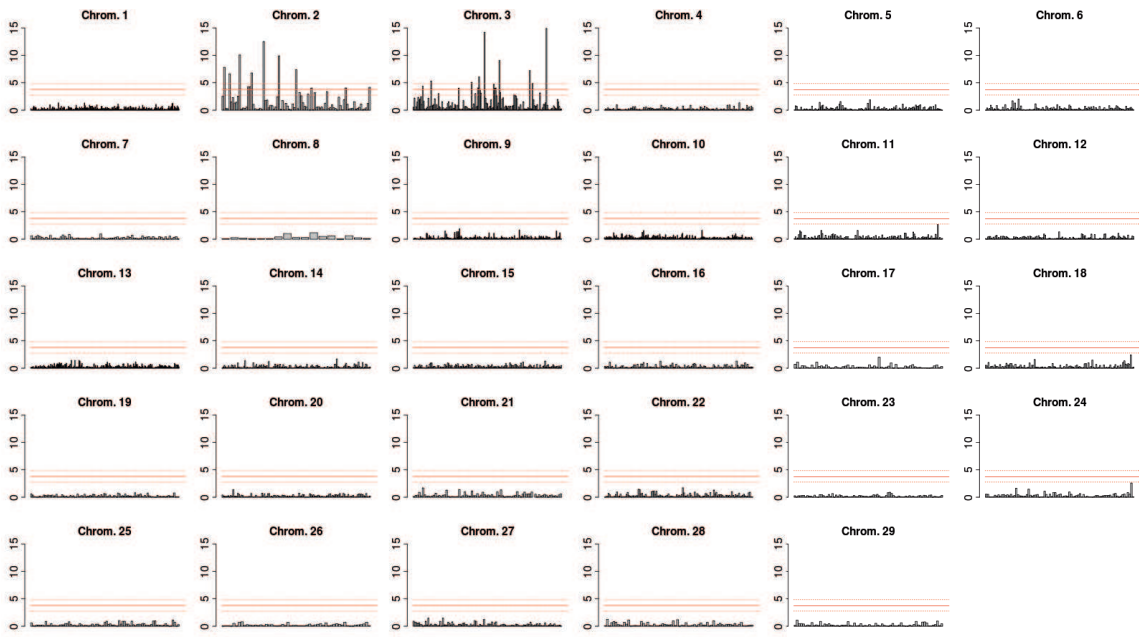

Figure 7: The figure shows the probability of genes involved in sex determination located on Atlantic salmon chromosomes. The x-axis display SNPs, y-axis shows the probability and the red line parallel to the x-axis is the significance level. Significant results were only found for chromosomes 2 and 3. The figure is taken from (Lien et al. Subm).

Genes that have earlier been ruled out as sex-determining genes because they could not be located on chromosome 2 now needs to be evaluated again, because there is a chance they can be located on chromosome 3.

1.7. Linkage mapping and physical mapping

The creation of a linkage map is of great help when studying the salmon genome. This has been done both by Gilbey at al. and Moen et al. (Gilbey et al. 2004; Moen et al. 2008). Genetic linkage is the tendency of certain loci or alleles to be inherited together (Wikipedia 2011). Genetic loci that are physically close to one another on the same chromosome tend to stay together during meiosis, and are thus genetically linked. A linkage map is a genetical map of a species or a population that show the position of known genes and/or markers relative to each other when considering recombination frequency, instead of specific physical distance at the chromosome. Markers used to construct linkage maps are usually either single nucleotide polymorphism (SNP).

Bacterial artifical chromosomes, also known as BACs, are based on the fertility plasmid, F-plasmid. BAC is the most used vectors for cloning DNA and can clone fragments of 300kb and longer (Brown 2006).They are used to sequence the genome of organisms by amplifying a piece of an organisms DNA as an insertion in a BAC and then sequenced. The sequenced parts are then rearranged and show the genomic sequence of the entire organism (Wikipedia 2011).

Fluorescent in situ Hybridization (FISH) has been of great value in the characterization of the salmonid chromosomes (Davidson et al. 2009). FISH is used to find and locate the presence or absence of certain DNA sequences on chromosomes. The technique is based on mapping marker locations by hybridizing a probe containing the marker to intact chromosomes.

1.8. Homology with *Gasterosteus aculeatus*

The genome of three-spined stickleback, *Gasterosteus aculeatus,* has been sequenced and can be used in comparative studies to extrapolate information less studied fish species, like Atlantic salmon. Recently, a syntetic relationships between Atlantic salmon and the reference genome sequence of three-spined stickleback was used to identify

homeologous chromosomal regions in the salmon genome (figure 8). Information from this study can also be used to predict location of candidate genes for sex-determination in salmon.

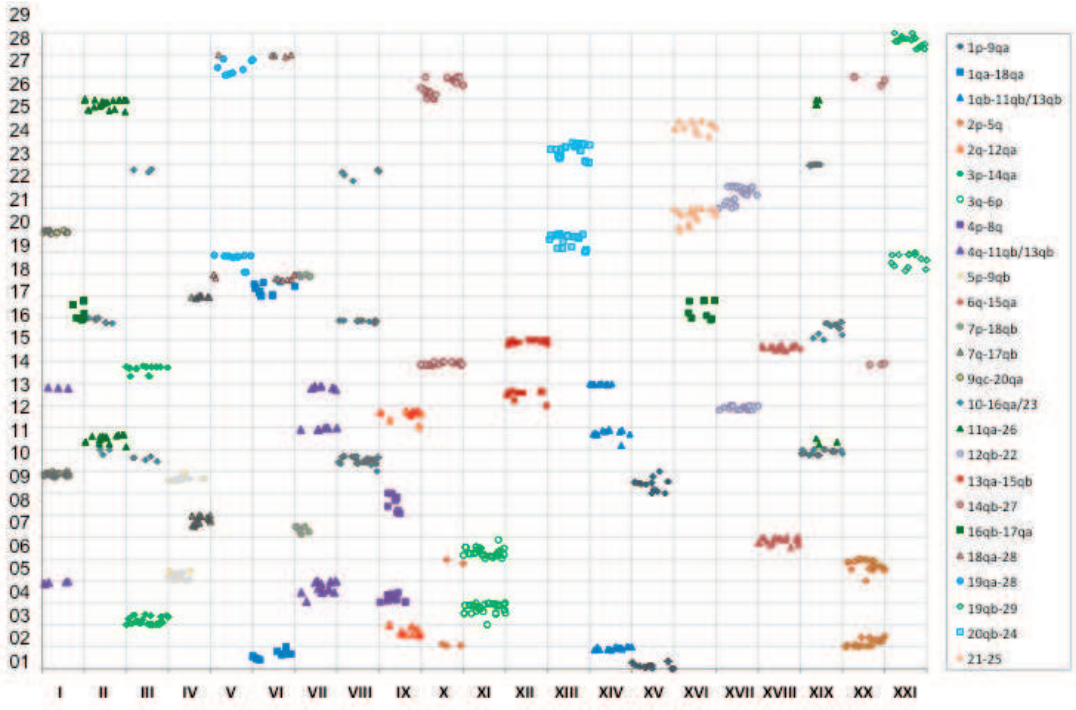

Figure 8: Comparison of syntenic relationships between stickleback (x-axis) and salmon (y-axis). The dots in the diagram are chromosome areas. The overview on the right shows which chromosome areas are homeologous to each other. (The figure is taken from (Lien et al. Subm).

From figure 8 we can see that chromosomes III, IX and XX in stickleback correspond to sex-determining chromosomes 2 and 3 in salmon. As a consequence the most potent candidate genes for sex determination in salmon may be found on these stickleback chromosomes.

1.9. Genome assemblies

The genome is the entire genetic complement of a living organism, in other words, all of the organisms DNA. Deoxyribonucleic acid, DNA, contains the genetic instructions used in development and functionality of any living organism. The genome includes the sequence of all the chromosomes and any DNA in organelles (Lewin 2006).

To construct a genome assembly, the entire organisms DNA have to be sequenced. It is not possible to sequence an entire genome in a single experiment, therefore the DNA is broken into fragments and the sequence is found for each fragment. The fragments are

joined together by finding overlaps and are then linked together in order to form a complete sequence of the genome. For large genomes, however, this shotgun method cannot be used all by itself. Usually genome maps are needed to help the assembly of the final sequence. The genome map shows positions of genes and other markers, like microsatellites and SNPs, and is used as a guide when sequencing genomes. The method combining the shotgun method and genome maps is called the whole-genome shotgun method (Brown 2006).

Another common technique used for sequencing genomes is the clone contig method. A contig is a contiguous genomic sequence where the order of the bases is well known. This method also takes advantage of the shotgun method. The difference is that the clone contig method use pieces of DNA with a known position on the genome map. The piece of DNA is then divided into fragments suitable for shotgun sequencing and put back together. The now sequenced piece of DNA is then placed back on its position on the genome map.

A third method of making an assembly of a genome is to construct a scaffold assembly. A scaffold is a part of a genome sequence consisting of contigs and gaps. When two sequenced ends of a fragment overlap with other reads in two different contigs, a gap occurs. The length of the gap can be estimated since the approximate lengths of the fragments are known (JGI_Genome_Portal 2011). Scaffolds are the result of a complete sequence assembly. The various markers on the scaffold are used to place the scaffold on the genome map.

There is one database and two assemblies used in this thesis. The Salmon UniGene Build database was available when this work started. The Salmon UniGene database is made up of mRNAs and Expressed sequence tags (ESTs) available at (http://www.ncbi.nlm.nih.gov/UniGene), as shown in table 1. An EST is cDNA that is sequenced in order to get quick and easy access to the genes in a genome. cDNA is made by converting mRNA into double-stranded DNA and because the mRNA in a cell is derived from protein-coding genes, the cDNA and the ESTs will represent the genes expressed in cell where the mRNA was prepared (Brown 2006).

Table 1. Number sequences included in the *salmo salar* UniGene assembly. The table is taken from http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=8030

Salmo salar: UniGene Build #28

ESTs are from dbEST through 04 Oct 2010

16,608	mRNAs
0	Models
0	HTC
155,145	EST, 3'reads
219,990	EST, 5'reads
76,129	EST, other/unknown
467,872	total sequences in clusters

Salmon Arachne 1.96x is a genome assembly built from 1.96x coverage of the salmon genome using Sanger reads from the salmon genome-sequencing project (Davidson et al. 2010). As indicated by its name the assembly was built by the Arachne assembler(Batzoglou et al. 2002). The assembly consists of 500827 contigs, with an average sequence length of 3135 bases. N_{50} contig size for this assembly is 4134. N_{50} means that of 50% of the contigs in this assembly will be longer than 4134 bases.

Salmon Celera 3.1x is a genome assembly built from 3.1x coverage of the salmon genome using Celera assembler (http://sourceforge.net/projects/wgs-assembler/) and data from the salmon genome-sequencing project. The assembly was produced in the beginning of April 2011. This assembly contains 204817 scaffolds with an average length of 12561 bases and also 433625 contigs with an average length of 3160 bases. N_{50} contigs and N_{50} scaffold sizes were 5742 and 80145, respectively. Only scaffolds from this assembly were used in this thesis.

1.10. Aim of the thesis

The goal of this thesis is to map genes that are involved in sex-determination and see if they exist on the sex determining chromosomes in Atlantic salmon. This will be done by searching for genes involved in sex determination in other species, find out if they exist in salmon, and then map them genetically and by BLAST against the Salmon Celera 3.1x assembly, which aligns gene sequences with markers already existing in the scaffolds.

2. Materials and Methods

The materials and methods of this thesis can be divided into three parts. First part describes the search for candidate genes in literature and sees if the genes exist in both stickleback and the Salmon UniGene Build and Salmon Arachne 1.96x assembly for Atlantic salmon. Second part covers the masking of sequences from the Salmon Arachne 1.96x assembly, as well as SNP detection. The third, and last part covers the genotyping and mapping of the candidate genes, both by genetic mapping and scaffold mapping.

2.1. Find candidate genes

The first step in this thesis was to search for genes that are involved in sex-determination in the literature databases ISI Web of Knowledge and PubMed. ISI Web of knowledge is a search database that gives access to articles related to sciences, as well as social science and (ISI_Web_Of_Knowledge 2010). PubMed is a similar database containing citations from life science journals and books and is developed and maintained by the National Center for Biotechnology (NCBI) in USA (PubMed_Central 2010).At this stage there was no limitation in what specific part the candidate gene had in sex determination, nor what organism in which it was found.

2.2. Positioning in the stickleback genome

The candidates found in the literature search were searched for in the UCSC genome browser (UCSC_GenomeBrowser 2006) or NCBI gene browser(NCBI 2011) to see if they were annotated or existed in the stickleback- or the Atlantic salmon genome. Hits found in these genomes, were recorded along with accession number and position in the stickleback genome.

The UCSC Genome Bioinformatics site contains reference sequences to a large collection of genomes. The site is developed and maintained by the Genome Bioinformatics Group at the University of California Santa Cruz.

By using the Genome Browser tool it is possible to locate specific genes and sequences on chromosomes in a given species, like the stickleback. The stickleback genome used at the UCSC-site is sequenced to 6X coverage and about 87% of the sequence has been anchored to chromosomes (UCSCGenomeBrowser 2006). In the UCSC genome browser

it is also possible to find the accession number and FASTA sequences for genes. FASTA format is the standard arrangement for nucleotide sequences when uploaded into tools such as BLASTn.

Basic Local Alignment Search Tool , or BLAST is an algorithm that finds regions of similarity between gene sequences. In this thesis, the nucleotide search function has been used (BLASTn), but it is also possible to search in protein sequences. It is possible to compare a reference sequence to a database, and also compare two or more different sequences.

If a gene was found in Atlantic salmon, genes found in closest relative species was recorded, preferably other salmonids and other fishes, but in certain cases also *mus musculus* (mouse) and *apis mellifera* (honeybee) was used.

2.3. BLAST against salmon UniGene Build #28 database

The BLAST searches were performed on a local implementation of an online software resource, Galaxy (http://main.g2.bx.psu.edu/) within CIGENE. Galaxy is a web-based tool that allows scientists to quickly analyze and multiple alignments and compare genomic annotations. The software was developed at Penn State University, USA (Zhang et al. 2000).

FASTA-sequences for the candidate genes were uploaded into Galaxy in two documents: one containing sequences found in salmon and another containing the remaining sequences. This because the sequences from salmon were expected to give better hits when using megaBLAST against the salmon UniGene Build database than the others, and for that reason different search settings were applied.

The salmon sequences was BLASTed towards the salmon UniGene Build database with similarity set to 80% and word size 28, while the rest used similarity set to 70% and word size 16. For BLAST search within the salmon sequences similarity was set as low as 80% in order to identify paralogues within the salmon genome for which we expect 85 – 95% similarity (Davidson et al. 2010).

The results of the megaBLAST were joined and compared to the Salmon UniGene Build database by using Galaxy. The best hit in the Salmon UniGene Build database for each of

the candidates was recorded, along with possible paralogs. The selection of sequences was based upon identity (%) between the FASTA-file and the UniGene Build database sequence, and BLAST-score. Comparing bit scores is a normal way to compare alignment scores from different BLAST searches.

Some candidates did not produce any hits in UniGene Build database at all. In these cases, the sequences that were originally uploaded into Galaxy (sequences from the literature search) were also used in the next step.

Some genes, even with different accession numbers and original sequences were aligned with the same sequence in the assembly. The reason for this could be repeats, or that the genes are in the same gene family and therefore are already very similar.

2.4. BLAST against the salmon genome salmon Arachne 1.96x assembly

All the sequences selected in the UniGene Build database, along with the sequences with no hits in this assembly, were uploaded into Galaxy again and used in BLAST search against the Salmon Arachne 1.96x assembly in the same way as with the UniGene Build database.

Similar sequences in the sox-genes caused similar hits in the Salmon Arachne 1.96x assembly. Due to these problems with the SOX-gene family, the entire procedure above was double-checked and repeated with these genes, including double-checking accession numbers and positions in the stickleback genome.

2.5. Masking

The salmon genome is known to contain long and frequent repeats, together with the size of the salmon genome make it a challenging species to study (Davidson et al. 2010).

RepeatMasker (RepeatMasker 2003) is a program that finds repeats in given sequences. In this case the program was given the FASTA-sequences from the salmon Arachne 1.96x assembly and found repeats in those sequences. The output of the program is an annotation of repeats that are present in every query sequence, as well as a new version of the sequence where all annotated repeats have been replaced by N (instead of the

nucleotide bases A, C, G, T).

Prior to BLAST searches sequences were repeat masked using a repeat database generated in CIGENE using RepeatModeler (RepeatModeler 2003) and data from the Salmon Arachne 1.96x assembly. A comparison of BLAST hits for unmasked and masked sequences showed that results from masked sequences were much better and easier to interpret. A BLAST search for the FASTA-sequences picked from the Salmon UniGene Build database against the Salmon Arachne 1.96 assembly gave approximately 3000 hits, whereas masked sequences produced approximately 800 hits.

2.6. SNP detection

A single nucleotide polymorphism, SNP, is a point mutation within the DNA (Brown 2006). The genetic code consists of the letters A (adenine), C (cytosine), G (guanine) and T (thymine). A SNP variation occurs when a single nucleotide, for example A, replaces on of the other three letters; C, G or T. Example of a SNP: AAGT***A***CTC in one individual and AAGT***G***CTC in another. A SNP has to be represented in at least 1% of a population for it to be considered a proper SNP. SNPs are very common markers used in the mapping of genomes.

In order to perform the SNP detection 45 individuals of Atlantic salmon was resequenced by creating Illumina reads with a length of 100bp each. The genome coverage on each of these resequenced individuals was 2x and the reference-sequence used in the resequencing was the Salmon Arachne 1.96x assembly (http://www.illumina.com).

The masked sequences from section 2.5.2 were made into FASTA-files and a SNP search was performed. The SNP-search was performed using the Burrows-Wheeler Alignment BWA tool (http://bio-bwa.sourceforge.net/) and the Utilities for the Sequence Alignment/Map (SAM) format (http://samtools.sourceforge.net/).

The Burrows-Wheeler Alignment Tool is a tool that aligns short sequences to database sequences. It is based on two Burrows-Wheeler Transform algorithms, one for queries up to 200bp with low error rate, and one for long reads with more errors.

It is, however, not the BWA tool that finds the SNP. BWA produces an alignment in SAM (sequence alignment/map) format, which can be used by samtools to detect SNPs. Samtools sorts, merges and indexes the alignments.

Several SNPs from the SNP detection was picked for every gene. For a SNP to good enough to be designed onto a primer, the SNP need a length of 100bp on each side.

2.7. Genotyping

All the SNP genotyping done in section 2.7 was performed as described below and in accordance with the protocol "iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassArray® System Through Single Base Primer Extension with Mass-Modified Terminators." (Oeth et al. 2005) (http://www.sequenom.com). A complete overview of the genotyping is given in Appendix I.

2.8. TyperAnalyzer

The four chips were uploaded in Mass ARRAY TyperAnalyzer v3.3 and all the SNPs were quality checked by seeing if they had an even distribution of homozygotes and heterozygotes.

A "Best call probability report" was made for the four chips and the Sequenom reports were uploaded into Galaxy and converted into an excel-file.

2.9. Two-point analysis

A linkage map has been constructed based on the genotyping of 3297 individuals from 143 different families of Atlantic salmon consisting of 5650 SNPs (Lien et al. Subm). The SNPs genotyped in this thesis were genotyped in a subset of the families mentioned above.

A two-point analysis estimates the recombination between a marker (SNP) and a locus/scaffold and finds a unique location for the marker. It identifies linkage between marker pairs so that linked markers can be placed on individual chromosomes or marker groups.

The result of the two-point analysis will map the genotypes to chromosome positions, based on recombination frequency, LOD-score and meioses.

Recombination frequency is the number of recombinants divided by the total number offspring. LOD-score (likelihood of odds) compares likelihood of obtaining data if two loci are indeed linked, with the likelihood of observing the same data by chance. A high LOD score means there is linkage. The formula for estimating LOD-score is as follows:

LOD = log10 likelihood if loci are linked / likelihood if the loci are not linked

The LOD-score needs to be above 3.0 to be significant. It is the combination of a high LOD-score and a low recombination frequency that is sought after.

To perform the two-point analysis, the program CRI-MAP was used(Green P 1990). CI-MAP is a program that is made to allow rapid, largely automated construction of multilocus linkage maps (Green P 1990).

2.10. BLAST against the Salmon Celera 3.1x assembly

Another way of mapping genes to chromosomes is by performing a BLAST search against an assembly containing scaffolds. The query sequences will be aligned against markers already existing on the scaffolds and will thus provide an accurate position on a chromosome.

FASTA files from the masked salmon Arachne 1.96x assembly hits were used in BLAST searches against Salmon Celera 3.1x assembly, using Galaxy.

The best hits were picked from the scaffold BLAST based on identity and BLAST-score. These hits were then joined and compared with salmon Celera 3.1x assembly. This way the query sequences got assigned to different scaffolds.

Lastly, the results from the two-point analysis were compared to the results from the BLAST search against the Salmon Celera 3.1x assembly to see if the two methods mapped the genes to the same chromosomes. The results here were also compared to the chromosome positions found in stickleback (section 2.2).

3. Results and Discussion

The results and discussion section can be divided into three main parts. The first part covers the selection of candidate genes from the literature, positioning of the genes in stickleback as well as detection of corresponding DNA sequences in Atlantic salmon. The second part covers the SNP detection and the genotyping in salmon. The third and last part shows the mapping of genes to chromosomes. This was achieved both by linkage analysis and by BLAST searches against scaffolds with a known chromosome position. This last part also evaluates the genes mapped to chromosomes 2 and 3.

3.1. Literature search and detection of corresponding sequences in Atlantic salmon

Appendix II give an overview of genes involved in sex-determination found in the initial literature search. After the candidates were found, their accession number and position in the stickleback genome was obtained by using the UCSC genome browser (http://genome.ucsc.edu/). The genes accession numbers and stickleback position are shown in Appendix I as well. For genes with unknown gene sequences (mRNA or ESTs) in Atlantic salmon or stickleback, sequences from other organisms were used, preferably other salmonids and teleosts, but in some cases also *c.elegans* and *mus musculus*.

The search for candidate genes in articles is a rather time consuming task. For this thesis we ended up with 65 genes after several weeks of work. We feel that the genes covered by this thesis represents a very relevant set of possible sex determining, but as new articles on the topic are released frequently and it is hard to keep track with development within the field. The candidate search was very broad and as along as a gene was involved in sex determination or development it was included in the study. With respect to selection of candidate genes that may represent a master switch of sex determination one could argue that the literature search could have been stricter. However, as much is unknown about sex-determination in salmon, we decided to test a large number of candidate genes.

3.2. BLAST against salmon UniGene Build #28 database

When performing a BLAST search, short pieces of sequence of a given length are used to find an initial match between the query sequence and the match in the database. The length of this short piece of sequence is the word size. A shorter word size will give a more sensitive search, while a longer word size will give a longer, faster search. Similarity describe how similar the query sequence and hit sequence are to each other, shown in percent (%).

The default word size when performing a nucleotide BLAST search is 28, which in this thesis was used for the BLAST searches using salmon sequences. The reason for this was that these sequences were expected to find almost identical hits in the assembly. For the other sequences, word size 16 was used, something which will give the search a higher sensitivity. The higher the word size is, the more identical the hit will be.

Most genes produced reliable hits, some also with paralogues, when aligning with the salmon UniGene Build database using BLAST. However, not all sequences were found in this assembly. Instead of already excluding these genes, their original sequences were kept. Even though a certain sequence could not be found in the salmon UniGene Build database there was still a chance that it could be found in the salmon Arachne 1.96x assembly of the salmon genome. Appendix III shows an overview of genes found in salmon according to the salmon UniGene Build database, and genes that was not found. Some of the genes also had potential paralogues that are listed in Appendix II.

3.2.1. Sox genes

In several articles, SOX genes were mentioned as genes involved in sex-determination. Some SOX genes were mentioned more frequently than others, such as SOX03 and SOX09. However, since the SOX-gene family seems to contain many genes that could be involved in sex determination in one way or another, we decided to include as much as 19 sox genes in this thesis. One major problem including 19 genes in BLAST searches was that a larger number of them ended up with identical hits, even though the FASTA sequences uploaded in the program are different. All the SOX-genes were then double-checked to see that their sequences were different, and all SOX-genes were run through BLAST towards the Salmon Arachne 1.96x assembly once more. Notably, also SRY derives from a SOX gene (Bowles et al. 2000), and that is the probably the reason for

why it produced the best same hit in the BLAST search got the same results as the SOX03 gene.

3.3. BLAST search against the salmon Arachne 1.96x assembly

A BLAST search in salmon Arachne 1.96x assembly identified reliable hits for 48 out of the 65 genes. Out of these, potential paralogues were identified for 16 genes. Selected hits in the salmon genome salmon Arachne 1.96x assembly are shown in Appendix IV. Among 3139 putative SNPs in these genes, 88 SNPs were picked for further use and are displayed in Appendix V.

3.4. Genotyping

MassARRAY genotyping assays were designed for 81 out of 89 putative SNPs in four multiplexes; two 23-plexes, one 20-plex and one 15-plex. The remaining 8 SNPs could not be designed. Table 5 lists these SNPs and includes the error report on why they could not be designed.

Table 5: The SNPs that could not be placed in wells. The error report tells why these SNPs were dismissed.

SNP_ID	ErrorReport
Aldh3a1_5773	No primers found on 3' side of SNP: too much repeated subsequence
Aldh3a1_5990	No primers found on 5' side of SNP: too much repeated subsequence
Dmrt2b_852	No primers found on 5' side of SNP: too many unknown bases
Gata4_191	Not enough bases of 5' side of SNP to design primers
Shbgb_1585	Invalid SNP syntax for SNP: [/]
Sox04_1231	No primers found on 5' side of SNP: too much repeated subsequence
Tbx1_para_3279	Failed to design stop mix for reverse extend primer: cannot design primer/analyte within mass window. Failed to design stop mix for forward extend primer: cannot design primer/analyte within mass window.
Wt1_para_7278	No primers found on 5' side of SNP: too much repeated subsequence

However, as it was expected that some of the putative SNPs might turn out to be false positives and not real SNPs, more than one putative SNP was picked for every gene (see section 3.3). Appendix VI lists a table where the distribution of the assays designed are shown, including the multiplex they were placed in.

A family material consisting of 380 samples in 21 families were genotyped for these four multiplexes and analyzed using the MassARRAY SpectroTYPER RT v3.4 software. After an automized genotype calling, data was manually inspected using the MassARRAY TyperAnalyzer v3.3 software. Markers were classified into 5 different classes: SNPs, MSV-3, PSVs, MONO and FAIL. Genotyping results representing these classes are shown in figures 11, 12, 13, 14 and 15.

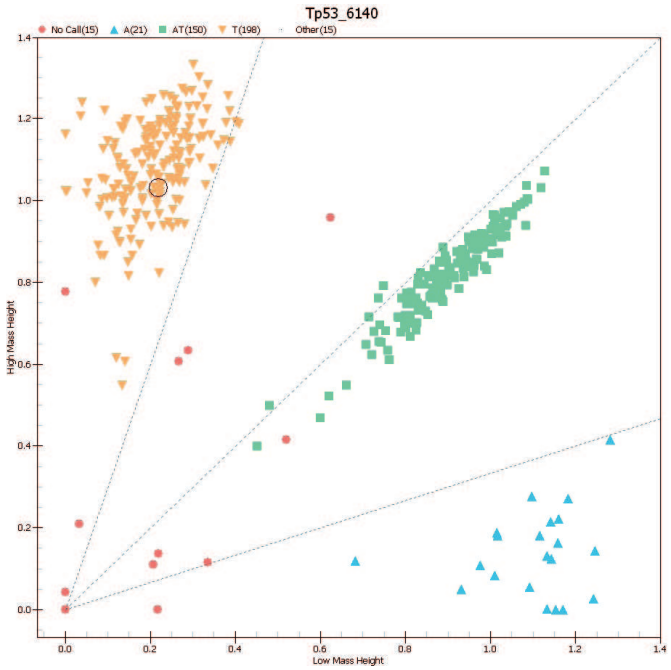

Figure 11. A typical example of how a SNP will be displayed in TyperAnalyzer v3.3. Three clusters display homozygote T (orange), heterozygote AT (green) and homozygote A (blue). The red dots displayed are either blanks (no DNA in well) or wells with weak/unclear signal.

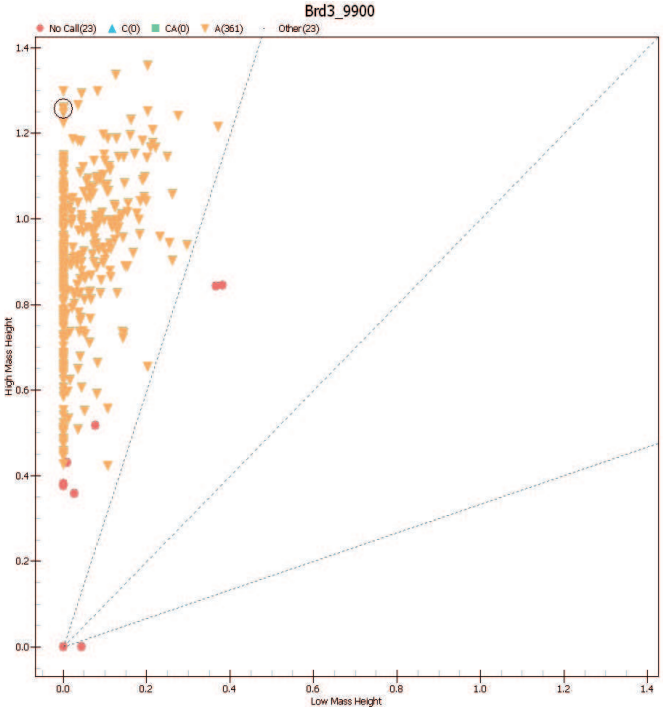

Figure 12. MONO. A typical example of how a monomorphic marker will be displayed in TyperAnayzer v3.3 (in this case heterozygote A).

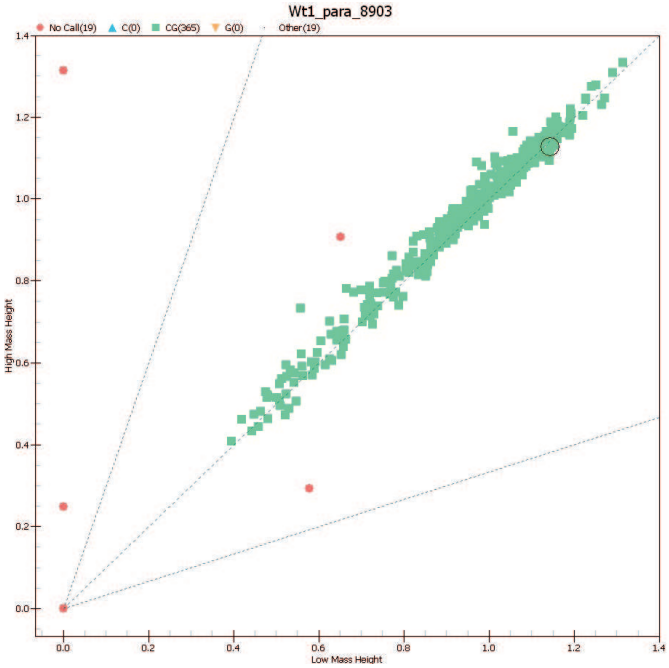

Figure 13. Example of how a paralogous sequence variant (PSV) will appear in TyperAnalyzer v3.3. Only heterozygotes appears in the plot.

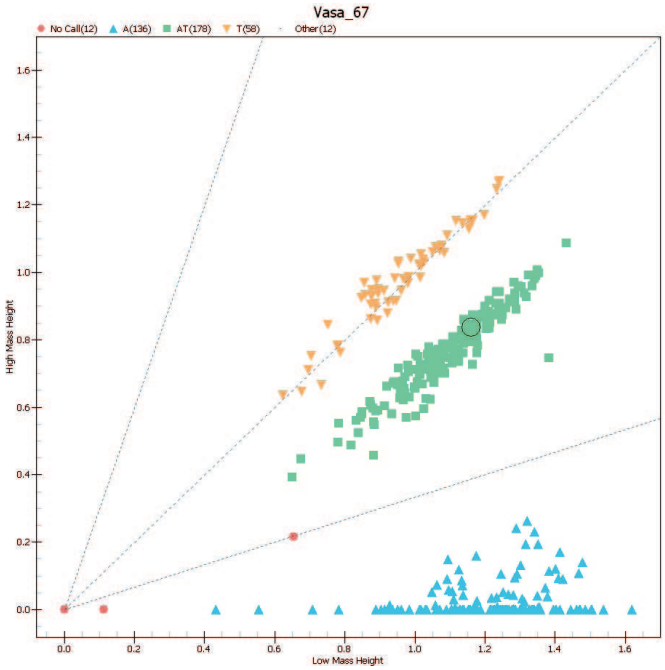

Figure 14. MSV-3. This shows how a typical MSV-3 may look like in TyperAnalyzer v3.3. Instead of being clustered as a normal SNP, plots are skewed to one side.

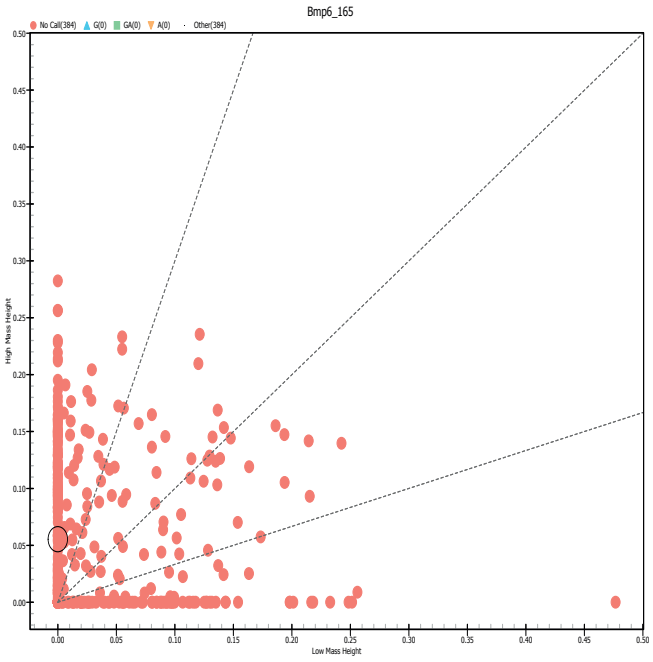

Figure 15. FAIL. This figure shows how an array that could not be designed will look like.

Table 6 presents all the markers uploaded in the Typer Analyzer v3.3 software and their classification.

Table 6: Markers uploaded into MassARRAY Typer software. Markers were sorted into 5 different classes; SNP, MSV-3, PSV, FAIL and MONO. For the markers that turned out to be monomorphic (MONO), the allele is also given.

Marker ID	Class	Marker ID	Class	Marker ID	Class
Aldh3a1_3040	SNP	Gsdf_8703	MONO A	Sox11_258	MONO G
Amh_1137	MONO A	Larp1_2862	SNP	Sox11_para_2898	PSV
Amh_3718	SNP	Larp1_para_1965	MONO A	Sox12_1208	FAIL
Bmp6_1384	MONO T	Lhx9_1928	MONO T	Sox12_469	SNP
Bmp6_165	FAIL	Lhx9_5633	MONO A	Sox13_856	MONO T
Bmp6_para_167	FAIL	Lim1_10920	MONO C	Sox19_2748	SNP
Bmp6_para_176	FAIL	Lim1_8568	MONO A	Sparc_2077	SNP
Brd3_1015	SNP	Mep1_1532	MONO C	Sparc_2411	MONO G
Brd3_9613	MONO A	Mep1_4877	MONO T	Sry_2416	MONO C
Brd3_9900	MONO A	Mep1_728	MONO A	Sry_7513	MONO T
Brd3_para_1540	MONO G	Mis12_3207	SNP	Ssb-4_1172	MONO T
Brd3_para_1709	SNP	Ppt1_2246	FAIL	Sxl_3541	MONO C
Cyp19a_1599	PSV	Sf1_para_691	SNP	Sxl_para_1106	MONO T
Cyp19a_1620	PSV	Shbgb_5575	MONO T	Tbx1_1234	MONO G
Dmrt1_2031	MONO T	Shbgb_para_1384	MONO T	Tbx1_para_1971	PSV
Dmrt1_3081	MONO A	Sox01_1406	PSV	Tp53_4247	MONO T
Dmrt2b_4439	MONO C	Sox01_6368	SNP	Tp53_6140	SNP
Dmrt3_1688	MONO G	Sox02_4652	MONO C	Tp53_6198	MONO T
Dmrt3_2279	PSV	Sox02_6616	SNP	Tra2b_4050	PSV
Foxl2_12516	MONO C	Sox04_para_6340	PSV	Tra2b_6890	SNP
Foxl2_5632	SNP	Sox05_3620	MONO C	Tra2b_para_9011	MONO C
Foxl2_8960	FAIL	Sox06_3565	SNP	Vasa_67	MSV-3
Foxl2_para_10210	MONO C	Sox06_6105	MONO C	Vcp_2698	SNP
Foxl2_para_10283	MONO G	Sox06_7845	MONO T	Vnn1_542	MONO C
Fru_3622	MONO T	Sox10_2096	MONO C	Wt1_132	FAIL
Gata4_221	MONO A	Sox10_3309	PSV	Wt1_para_8903	PSV
Gsdf_3248	MONO A	Sox10_3349	FAIL	Wt1_para_9590	PSV

Table 7 (below) presents the number of markers that fell into each class.

Table 7: The total number of SNPs, MSV-3s, PSVs, MONOs and FAIL.

SNPs:	17
MSV-3s:	1
PSVs:	11
MONO:	44
FAIL:	8
Total:	81

It was a bit disappointing that our genotyping revealed that only 18 out of 81 assays turned out to be 'real' SNPs or MSV-3s. A large fraction of putative SNPs turned out to be monomorphic markers (MONO) or paralogous sequemce variants (PSVs). The most likely explanation this is that errors in the resequencing data have been misinterpreted as SNPs. Evidently, one weakness in the SNP detection was low sequence coverage (approximately 2x coverage) per individual sequenced. Thus an alternative way of analyzing the data would be to mix the sequences of the 45 individuals into a gene pool. This pool would also generate 90x coverage, in contrast to the 2x coverage for each individual. Also, the conditions in the Burrows Wheeler Alignment tool could have been more stringent to avoid mixed reads from paralogous regions.

A larger fraction of markers within the duplicate regions in the Atlantic salmon genome is expected to be MSVs (Danzmann et al. 2008). As shown in the figure 14, MSV genotype clusters looks like a normal SNP, but everything is skewed to one side. The reason for this is that their assay signal reflects a mixture of four alleles. Based on manual inspection of data using the MassARRAY TyperAnalyzer v3.3 software, we judge that approximately 5,5% of the markers in this work was MSVs, which is lower than the 9,5% estimated previously from microsatellite data (Danzmann et al. 2008). The reason for the low frequency of duplicated markers is most likely caused by the low amount of markers mapped in general in this thesis. Since only 18 markers in this thesis were

'real,' the portion of MSVs could easily be random and is not representative for Atlantic salmon as a whole.

3.5. Linkage mapping

18 SNPs were genotyped in a set of 380 individuals from 21 different families. These families are from a subset of a larger material used to construct a high-density SNP-map for Atlantic salmon (Lien et al. Subm).

The results from the linkage analysis combining the two datasets using the two-point option in CRIMAP (Green P 1990) are presented in table 8.

Table 8: The SNPs mapped by two-point analysis against markers in a genetic map constructed for Atlantic salmon (Lien et al. Subm). The best LOD-score is between the candidate gene SNPs mapped and previously mapped markers, along with the recombination rate. The chromosome each SNP was mapped to is given as well an approximate distance in cM downward on the chromosome for both male and female. The meioses column presents informative meioses for the SNPs.

SNP	*Best LOD-score*	*Recombination*	*Chromosome*	*cM female*	*cM male*	*Meioses*
Aldh3a1	19,87	0	9	60 - 89	3,4 - 5,7	159
Amh	12,42	0,05	10	37,5 - 48,9	1,6 - 1,7	54
Brd3	15,21	0	5	70 - 90	43,8 - 44,1	114
Brd3 para	31,91	0	27	19 - 36	0,6 - 1,2	245
Foxl2	12,04	0	14	40 - 65	0,9 - 2,3	43
Larp1	25,92	0	10	54 - 86	2,1 - 8,9	172
Mis12	69,84	0	13	57 - 77	1,9 - 5,1	356
Sf1 para	31,44	0	18	38 - 66	1,7 - 27,2	223
Sox01	38,73	0	17	44,4 - 68,2	0,8 - 6	275
Sox02	15,4	0	10	34,2 - 53,9	1,6 - 2,1	71
Sox06	36,22	0	11	42,2 - 83,3	1,1 - 2,8	263
Sox12	12,94	0	25	34,8 - 54,7	0,2 - 2	57
Sox19	12,06	0	7	20,8 - 95,3	57,3 - 58,1	71
Sparc	46,99	0	9	59,2 - 96,1	2,3 - 5,7	261
Tp53	34,46	0,02	4	39,6 - 77,3	2,5 - 3,4	247
Tra2b	30,71	0	16	53,4 - 62,3	2 - 9,6	155
Vasa	36,72	0	27	35,8 - 53,4	1,2 - 3,5	308
Vcp	47,56	0	1	115,3 - 134,3	49,9 - 52,4	311

Most likely chromosomal positions for the 18 genes that were determined based on recombination rate (as low as possible) and LOD scores (as high as possible). The LOD score is a result of recombination frequency and informative meioses. However, in this study where only 380 individuals were used in the SNP detection and genotyping it's difficult to place the SNP to an accurate position on the chromosome since the number of meioses is relatively low. If more individuals were included in a study, it should have

been possible to get more accurate the positioning of the genes within a chromosomal region.

In Atlantic salmon there is a great difference between males and females in terms of recombination rates (Danzmann et al. 2008; Lien et al. Subm; Moen et al. 2008). Figure 16 presents an overview of salmon chromosome 2 with male and female marker maps taken from Lien et al. (Lien et al. Subm). The figure illustrates a general recombination pattern in male salmon characterized by high recombination in telomeric regions, but low on the other chromosomal regions. This low recombination rate in males affects our ability to precisely position genes within certain regions.

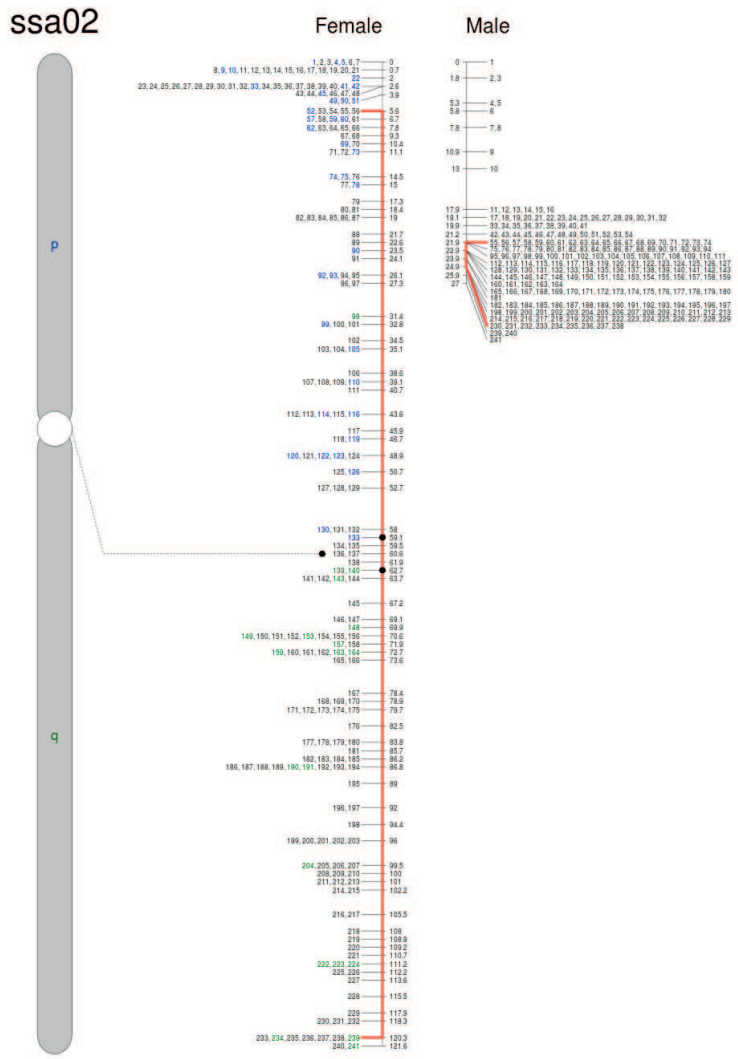

Figure 16. Atlantic salmon chromosome 2 (Lien et al. Subm). Red areas in female and male correspond with each other, something that shows that the recombination rate is much lower in females than males.

3.6. BLAST against scaffolds in the Salmon Celera 3.1x assembly

Recently an assembly of salmon genome (Salmon Celera 3.1x) was generated in CIGENE. Mapping of genes to scaffolds with SNP positions in the dense linkage constructed by Lien et al. (Lien et al. Subm) allow for mapping of additional candidate genes to chromosomes.

Table 9 shows genes mapped by the Salmon Celera 3.1x assembly. Appendix VII also presents the genes that were found in the BLAST search between the masked sequences

and the Salmon Celera 3.1x assembly. Seventy sequences were selected from the BLAST search against the Salmon Celera 3.1x assembly. However, only 22 of these scaffolds had been mapped to a chromosome by a scaffold already positioned by the dense SNP-map generated by Lien et al., (Subm.). This leaves out 37 sequences positioned to scaffolds not yet assigned to any specific chromosomes.

Table 9. The table shows genes positioned by BLAST between the query sequences and the Salmon Celera 3.1x assembly. The table presents what scaffold the gene is mapped to, scaffold length, and chromosome number. The SNP number tells what chromosome is closest to the gene, and cM female the distance downward on the chromosome. The identity, e-value and BLAST-score show the quality of the hit between the query sequence and Salmon Celera 3.1x assembly.

Gene	*Scaffold*	*Ssa*	*SNP number*	*cM female*	*Identity %*	*E-value*	*BLAST-score*
Aldh3a1	scf15186897	ssa09qb	156	69,5	100	0	1350
Aldh3a1|para	scf15186897	ssa09qb	156	69,5	94,19	0	848
Am-dsx	scf15179227	ssa03q	280	112	100	0	801
Amh	scf15097739	ssa10qb	66	45,1	100	0	872
bmp6	scf15190741	ssa22q	48	37,4	84,42	6,00E-22	111
Brd3	scf15193845	ssa05q	165	81	93,88	5,00E-53	214
Brd3	scf15113107	ssa02p	98	31,4	100	0	646
Dmrt2b	scf15192969	ssa23q	24	13,2	100	0	918
fancl	scf15193879	ssa18qa	52	42,3	100	0	971
Ppt1	scf15192456	ssa27q	27	19,8	100	2,00E-31	143
Rbp1	scf15188577	ssa03p	46	4,9	89,19	4,00E-37	161
Sf1	scf15194451	ssa18qa	101	2,5	92,98	7,00E-52	210
Sox06	scf15193093	ssa23q			100	0	1298
Sox09/Sox17	scf15158750	ssa29q	24	25,1	99,71	0	680
Sox16	scf15173171	ssa28q	13	8,3	82,64	1,00E-55	222
Tbx1	scf15188114	ssa24q	40	37,3	86,67	1,00E-133	482
tp53	scf15118478	ssa04q	77	52,1	100	3,00E-113	416
vasa	scf15204390	ssa27q	117	51,8	89,73	1,00E-75	289
vasa	scf15166306	ssa04q	75	48	100	2,00E-46	192
Vcp	scf15192588	ssa01qb	370	134	100	6,00E-64	250
Vnn1	scf15131889	ssa18qa	45	38,4	100	3,00E-54	218

Even though most genes presented in table 9 produced good hits between the query sequence and the Salmon Celera 3.1x assembly, several genes (BMP6, RBP1, SOX16, TBX1 and one of the VASA versions) had a low identity between the query sequence and the best hit in the assembly. The gene BMP6 also had a low BLAST score. The reason for this is that after the scaffold BLAST these sequences were picked out as potential paralogues.

The salmon Celera 3.1x assembly only assigned a chromosome to the gene SOX06. There were no markers on the same scaffold to give a more precise position. This could be an error.

SOX09 and SOX17 produced identical hits in the salmon Celera 3.1x assembly. It is hard to know which one of these two that is actually mapped and therefore it is annotated as SOX09/SOX17 at this point, but problem this will be addressed in section 3.8.

3.7. Comparison of linkage mapping results and BLAST hits against Salmon Celera 3.1x scaffolds positioned to chromosomes

Table 10 show the linkage mapping results compared to the BLAST search against the Salmon Celera 3.1x scaffolds positioned on chromosomes by Lien et al. (Subm)

Table 10: Linkage mapping vs Salmon Celera 3.1x assembly.

SNP	*Two-point linkage analysis*	*Salmon Celera 3.1x*
Aldh3a1	09	09qb
Amh	10	10qb
Brd3	05	02p / 05q
Brd3 para	27	
Foxl2	14	
Larp1	10	
Mis12	13	
Sf1 para	18	
Sox01	17	
Sox02	10	
Sox06	11	
Sox12	25	
Sox19	07	
Sparc	09	
Tp53	04	04q
Tra2b	16	
Vasa	27	27q/04q
Vcp	01	01qb

From the table 10 we see that genes mapped both by the two-point linkage analysis and BLAST against the Salmon Celera 3.1x assembly gave similar chromosomal locations.

In table 8, BRD3 para, which is a sequence selected in the Salmon Arachne 1.96x assembly as a potential paralogue for the BRD3-gene, is mapped to chromosome 27. This is a chromosome not homologous to the chromosomes 2 and 5, where the BRD3-gene was mapped. The reason for this could be that rather than finding the paralogue for this gene, a different gene in the same gene family was found. If BRD3 and BRD3 para were paralogous they would be located on chromosomes that are homologous to each other.

3.8. Homology with stickleback

Table 11 presents the chromosome position of the genes mapped in this thesis along with their homologues. Their syntenic chromosomal regions in stickleback are also listed, as well as the predicted location of the candidate genes in the stickleback genome (from Appendix II).

Table 11. The table presents the location of the genes mapped to salmon chromosomes (SsaChr), as well as their homologous areas (Ssa Homologues). The two columns on the left shows the chromosomes in stickleback that are syntenic to the mapped salmon chromosomes (StckChr Homeo.), as well as the expected syntenic chromosome in stickleback based on the gene's location in stickleback from the UCSC Genome Browser (UCSC StckChr).

SNP	SsaChr	Ssa Homologues	StckChr Homeo.	UCSC StckChr
Aldh3a1	ssa09qb	ssa05p	IV	I
Aldh3a1 para	ssa09qb	ssa05p	IV	I
Am-dsx	ssa03q	ssa06p	XI	
Amh	ssa10	ssa16qa/ssa23	VIII/XIX	VIII
bmp6	ssa22q	ssa12qb	XVII	XXI
Brd3	ssa05q	ssa02p	XX	XX
Brd3	ssa02p	ssa05q	XX	XX
Brd3 para	ssa27	ssa14qb	X	XX
Dmrt2b	ssa23q	ssa10/ssa16qa	VIII/XIX	VIII
Fancl	ssa18qa	ssa01qa/ssa28	VI	
Foxl2	ssa14	ssa03p/ssa27	III/X	I
Larp1	ssa10	ssa16qa/ssa23	VIII/XIX	XV
Mis12	ssa13	ssa04q/ssa11qb/ssa15qb	I/VII/XII	II
Ppt1	ssa27q	ssa14qb	X	X
Rbp1	ssa03p	ssa14qa	III	III
Sf1	ssa18qa	ssa01qa/ssa28	VI	
Sf1 para	ssa18	ssa01qa/ssa28	VI	
Sox01	ssa17	ssa07q/ssa16qb	IV/XVI	XVI
Sox02	ssa10	ssa16qa/ssa23	VIII/XIX	VIII
Sox06	ssa11	ssa26/ssa13qb/ssa04q/ssa01qb	II/VII/XIV/XIX	II
Sox06	ssa23	ssa10/ssa16qa	III/VIII/XIX	II
Sox09/Sox17	ssa29q	ssa19qb	XXI	XXI/XVIII
Sox12	ssa25	ssa21	XVI	XVI
Sox16	ssa28q	ssa18qa/ssa19qa	VI	I
Sox19	ssa07	ssa17qb/ssa18qb	IV/VII	VII
Sparc	ssa09	ssa20qa/ssa05p/ssa01p	I/IV/XV	IV
Tbx1	ssa24q	ssa20qb	XIII	XIII
Tp53	ssa04q	ssa13qb/ssa11qb	VII	VII
Tra2b	ssa16	ssa10/ssa17qa/ssa23	I/II/VIII/XIX	XV
Vasa	ssa27q	ssa14qb	X	X
Vasa	ssa04q	ssa13qb/ssa11qb	VII	X
Vcp	ssa01qb	ssa18/ssa09/ssa11/ssa13	XIV	I
Vnn1	ssa18qa	ssa01/ssa28	VI	VI

Out of 33 gene sequences, 17 were mapped in accordance with expected syntenic chromosomal regions between stickleback and Atlantic salmon. Seven genes were mapped to conflicting regions and four of the genes in table 11 could not be located in

stickleback by the UCSC Genome Browser. The VASA gene was mapped to chromosome 27 by the two-point analysis and to chromosome positions 4q and 27q by the scaffold mapping. Table 9 shows that the identity of the VASA-gene mapped to salmon chromosome position 27q is 89,73%, while the identity of the VASA gene mapped to chromosome position 04q is 100%. In table 11, only the version of the VASA-gene that was mapped to salmon chromosome position 27q matches its expected position in stickleback. Because of the homeology with the chromosome position 27q is most likely the correct location for this gene, although the vasa version mapped to chromosome position 04 could be in the same gene family, but not a paralogue because of the lack of homology between salmon chromosomes 4 and 27.

The mapping of the gene sequence corresponding to SOX09 or SOX17 show a match up with the location of the SOX09 gene in stickleback, and therefore we can conclude that the gene mapped is most likely SOX09 and not SOX17.

A reason for why some of the genes mapped in this thesis did not match the expected position in stickleback could be that the sequences producing best hits in the Salmon Celera 3.1x assembly, or even the Salmon Arachne 1.96x assembly or Salmon UniGene Build database, not were the sequences representing the correct gene. In all sequences there were often more than one good hit for each gene, as well as potential paralogues. If another sequence for a gene was selected, it could perhaps been mapped to a different chromosome. Genes in the same gene family are usually very similar, and that difference could cause the gene to be mapped to a different chromosome expected.

Another reason for why the expected homeology between the salmon chromosomes and stickleback chromosomes does not always match be that the syntenic map created by Lien et al. (Subm.) does not necessarily give a complete picture of the syntenic relationship between these two species.

3.1. Assemblies

It is still possible that several of the genes that could not be mapped still exist in the Atlantic salmon genome. The reason for why they were not mapped is because the current assemblies for Atlantic salmon are incomplete. If work covered by this thesis

was tried again in a few years when there is a complete salmon genome reference sequence available, results would probably have changed dramatically.

3.2. Genes mapped to chromosomes 2 or 3

There were four genes mapped to chromosomes 2 and 3, including one gene that was mapped to a chromosome homologous to chromosome 3. The gene BRD3 was mapped to chromosome 5 in the two-point analysis and both chromosomes 2 and 5 in the salmon Celera 3.1x assembly. Atlantic salmon chromosomes 2q and 5p are homologous chromosome arms. In stickleback, the BRD3-gene is located on chromosome XX, containing a syntenis region to chromosomes 2p and 5q in Atlantic salmon. This gene was mapped to the p-arm of chromosome 2 and not the q arm, where Artieri et al. (Artieri et al. 2006) located the sex-determining loci to be located. However, because of the low recombination present in the male chromosome, it is still possible that the p-arm also is involved in sex-determination. BRD3 is a gene that is expressed in pre-Sertoli cells on the genital ridge of mouse at the time of sex-determination (Boyer et al. 2004). Pre-Sertoli cells are the first cell types to show sex specific differentiation (Cory et al. 2007).

AM-DSX was mapped to chromosome 3 in Atlantic salmon but could not be found in stickleback. The AM-DSX gene is a version of the DSX, doublesex, gene found in honeybee. It encodes for a highly conserved transcription factor in metazoans and is involved in the regulation of the expression of genes involved in sexual phenotype formation(Hodgkin 2002). It has functions in both sexes; it can encode either for a male- or female-specific isoform (Baker & Wolfner 1988).

RBP1 was placed on chromosome III in stickleback, as well as in Atlantic salmon. This RBP1 sequence was picked as a potential paralogue in the Salmon Celera 3.1x assembly, but because the version of the gene that showed 100% identity with the Salmon Celera 3.1x could not be mapped, it is hard to know whether the sequence being mapped is an RBP1 paralogue or another gene with high similarity of RBP1. However, the fact that the gene's mapped location of the gene in Atlantic salmon matches the predicted syntenic region in stickleback indicates that it is the Rbp1 gene or a paralogue that is mapped. RBP1 is a gene involved in alternative splicing of the sex-determining gene in *Drosophila*

melanogaster, DSX, which is highly similar to the AM-DSX gene mapped to chromosome 3 in this thesis (Cristino et al. 2006).

FOXL2 was mapped to chromosome 14qa in Atlantic salmon, which is homologous to chromosome 3p. Therefore, FOXL2 could still be a potent candidate as a sex-determining gene in Atlantic salmon. FOXL2 is a gene that is involved in follicular development and ovary regulation in vertebrates and individuals missing this gene will appear male (Loffler et al. 2003).

The rest of the genes mapped in this thesis can be ruled out as sex-determining factors in Atlantic salmon because they are not located at the sex-determining chromosomes or their homologues.

4. Conclusion

Out of the 65 genes found in literature, 29 different genes were successfully mapped to chromosomes in Atlantic salmon. The genes BRD3, AM-DSX and RBP1 may be candidate genes for sex-determination in Atlantic salmon because they map to chromosomes 2 or 3. Also FOXL2 is a strong candidate, as it was mapped to chromosome 14qa, a chromosome arm homologous to chromosome 3p. Genes not mapped to the two sex-determining chromosomes, or their homologs, are to be ruled out as candidates for sex-determination in Atlantic salmon, while further investigations should be made for BRD3, AM-DSX, RBP1 and FOXL2.

In order to map more genes that potentially are involved in sex-determination in Atlantic salmon, a reference genome with high coverage is of high importance(Davidson et al. 2010). It is therefore important not to rule out genes that could not be mapped by this thesis completely, because these could still exist in areas of the salmon genome with low coverage.

5. Abbreviations

bp – base pair

cM – centiMorgan

Mb – Megabases

mg - milligram

mM - millimolar

ng - nanogram

nM - nanomolar

rpm – rounds per minute

µL – Microlitre

µM - micrometer

6. References

Alfaqih, M. A., Brunelli, J. P., Drew, R. E. & Thorgaard, G. H. (2009). Mapping of five candidate sex-determining loci in rainbow trout (Oncorhynchus mykiss). *BMC Genetics*, 10.

Allendorf & Thorgaard. (1984). Tetraploidy and the evolution of salmonid fishes. *Evolutionary Genetics of Fishes*: 1 - 46.

Artieri, C. G., Mitchell, L. A., Ng, S. H. S., Parisotto, S. E., Danzmann, R. G., Hoyheim, B., Phillips, R. B., Morasch, M., Koop, B. F. & Davidson, W. S. (2006). Identification of the sex-determining locus of Atlantic salmon (Salmo salar) on chromosome 2. *Cytogenetic and Genome Research*, 112 (1-2): 152-159.

Baker, B. S. & Wolfner, M. F. (1988). A MOLECULAR ANALYSIS OF DOUBLESEX, A BIFUNCTIONAL GENE THAT CONTROLS BOTH MALE AND FEMALE SEXUAL-DIFFERENTIATION IN DROSOPHILA-MELANOGASTER. *Genes & Development*, 2 (4): 477-489.

Baroiller, J. F., D'Cotta, H. & Saillant, E. (2009). Environmental Effects on Fish Sex Determination and Differentiation. *Sexual Development*, 3 (2-3): 118-135.

Batzoglou, S., Jaffe, D. B., Stanley, K., Butler, J., Gnerre, S., Mauceli, E., Berger, B., Mesirov, J. P. & Lander, E. S. (2002). ARACHNE: A whole-genome shotgun assembler. *Genome Research*, 12 (1): 177-189.

Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E. & Omholt, S. W. (2003). The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell*, 114 (4): 419-429.

Billard, R., Richard, M. & Rombauts, R. (1982). INHIBITION OF SPERMATOGENESIS AND VITELLOGENESIS IN RAINBOW-TROUT BY HORMONAL ADDITIVES IN THE DIET. *Progressive Fish-Culturist*, 44 (1): 15-18.

Bobe, J., Mahe, S., Nguyen, T., Rime, H., Vizziano, D., Fostier, A. & Guiguen, Y. (2008). Identification of a novel Sex Hormone Binding Globulin (SHBG) in the rainbow trout (Oncorhynchys mykiss) ovary. *Cybium*, 32 (2): 151-152.

Bobe, J., Guiguen, Y. & Fostier, A. (2010). Diversity and biological significance of sex hormone-binding globulin in fish, an evolutionary perspective. *Molecular and Cellular Endocrinology*, 316 (1): 66-78.

Bowles, J., Schepers, G. & Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Developmental Biology*, 227 (2): 239-255.

Boyer, A., Lussier, J. G., Sinclair, A. H., McClive, P. J. & Silversides, D. W. (2004). Pre-Sertoli specific gene expression profiling reveals differential expression of Ppt1 and Brd3 genes within the mouse genital ridge at the time of sex determination. *Biology of Reproduction*, 71 (3): 820-827.

Brown, T. A. (2006). *Genomes - Third Editiom*: Garland Science.

Cory, A. T., Boyer, A., Pilon, N., Lussier, J. G. & Silversides, D. W. (2007). Presumptive pre-Sertoli cells express genes involved in cell proliferation and cell signalling during a critical window in early testis differentiation. *Molecular Reproduction and Development*, 74 (12): 1491-1504.

Cristino, A. S., do Nascimento, A. M., Costa, L. D. & Simoes, Z. L. P. (2006). A comparative analysis of highly conserved sex-determining genes between Apis mellifera and Drosophila melanogaster. *Genetics and Molecular Research*, 5 (1): 154-168.

Danzmann, R. G., Davidson, E. A., Ferguson, M. M., Gharbi, K., Koop, B. F., Hoyheim, B., Lien, S., Lubieniecki, K. P., Moghadam, H. K., Parks, J., et al. (2008). Distribution of

ancestral proto-Actinopterygian chromosome arms within the genome of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). *BMC Genomics*, 9.

Davidson, W. S., Huang, T. K., Fujiki, K., von Schalburg, K. R. & Koop, B. F. (2009). The Sex Determining Loci and Sex Chromosomes in the Family Salmonidae. *Sexual Development*, 3 (2-3): 78-87.

Davidson, W. S., Koop, B. F., Jones, S. J. M., Iturra, P., Vidal, R., Maass, A., Jonassen, I., Lien, S. & Omholt, S. W. (2010). Sequencing the genome of the Atlantic salmon (Salmo salar). *Genome Biology*, 11 (9).

Desjardins, J. K. & Fernald, R. D. (2009). Fish sex: why so diverse? *Current Opinion in Neurobiology*, 19 (6): 648-653.

Devlin, R. H. & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*, 208 (3-4): 191-364.

Eksportutvalgetforfisk. (2011). http://www.seafood.no/binary?id=138749.

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

Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M. & Noce, T. (1994). ISOLATION OF A DEAD-FAMILY PROTEIN GENE THAT ENCODES A MURINE HOMOLOG OF DROSOPHILA-VASA AND ITS SPECIFIC EXPRESSION IN GERM-CELL LINEAGE. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (25): 12258-12262.

Fukada, S., Tanaka, M., Iwaya, M., Nakajima, M. & Nagahama, Y. (1995). THE SOX GENE FAMILY AND ITS EXPRESSION DURING EMBRYOGENESIS IN THE TELEOST FISH, MEDAKA (ORYZIAS-LATIPES). *Development Growth & Differentiation*, 37 (4): 379-385.

Gilbey, J., Verspoor, E., McLay, A. & Houlihan, D. (2004). A microsatellite linkage map for Atlantic salmon (Salmo salar). *Animal Genetics*, 35 (2): 98-105.

Graves, J. A. M. (1998). Interactions between SRY and SOX genes in mammalian sex determination. *Bioessays*, 20 (3): 264-269.

Green P, F. K., Crooks S. (1990). Documentation for CRI-MAP, version 2.4. *Washington University Scool of Medicine, St. Louis, Missouri*.

Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M. B., Bonaldo, M. D., Smith, L., Tymowska-Lalanne, Z., Wells, C., et al. (2000). Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development. *Human Molecular Genetics*, 9 (10): 1553-1560.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. & Lovellbadge, R. (1990). A GENE-MAPPING TO THE SEX-DETERMINING REGION OF THE MOUSE Y-CHROMOSOME IS A MEMBER OF A NOVEL FAMILY OF EMBRYONICALLY EXPRESSED GENES. *Nature*, 346 (6281): 245-250.

Guiguen, Y., Fostier, A., Piferrer, F. & Chang, C. F. (2010). Ovarian aromatase and estrogens: A pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology*, 165 (3): 352-366.

Hale, M. C., Jackson, J. R. & DeWoody, J. A. (2010). Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (Acipenser fulvescens). *Genetica*, 138 (7): 745-756.

Hasselmann, M., Gempe, T., Schiott, M., Nunes-Silva, C. G., Otte, M. & Beye, M. (2008). Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature*, 454 (7203): 519-U7.

Hawkins, J. R. (1994). SEX DETERMINATION. *Human Molecular Genetics*, 3: 1463-1467.

Hodgkin, J. (2002). The remarkable ubiquity of DM domain factors as regulators of sexual phenotype: ancestry or aptitude? *Genes & Development*, 16 (18): 2322-2326.

Hoshino, A., Koide, M., Ono, T. & Yasugi, S. (2005). Sex-specific and left-right asymmetric expression pattern of Bmp7 in the gonad of normal and sex-reversed chicken embryos. *Development Growth & Differentiation*, 47 (2): 65-74.

ISI_Web_Of_Knowledge. (2010). *http://apps.isiknowledge.com*.

JGI_Genome_Portal. (2011). http://genome.jgi-psf.org/help/scaffolds.html.

Johnstone, R., Simpson, T. H., Youngson, A. F. & Whitehead, C. (1979). SEX REVERSAL IN SALMONID CULTURE .2. PROGENY OF SEX-REVERSED RAINBOW-TROUT. *Aquaculture*, 18 (1): 13-19.

Kasturi, P., Zanetti, S., Passannante, M., Saudan, Z., Muller, F. & Puoti, A. (2010). The C. elegans sex determination protein MOG-3 functions in meiosis and binds to the CSL co-repressor CIR-1. *Developmental Biology*, 344 (2): 593-602.

Lewin, B. (2006). *Essential Genes*: Pearson Education, Inc.

Lien, S., Gidskehaug, L., Hayes, B., Moen, T., Davidson, W. S., Omholt, S. & Kent, M. (Subm). A dense SNP-based linkage map for Atlantic salmon (salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *Genome Biology*.

Loffler, K. A., Zarkower, D. & Koopman, P. (2003). Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: FOXL2 is a conserved, early-acting gene in vertebrate ovarian development. *Endocrinology*, 144 (7): 3237-3243.

Manolakou, P., Lavranos, G. & Angelopoulou, R. (2006). Molecular patterns of sex determination in the animal kingdom: a comparative study of the biology of reproduction. *Reproductive Biology and Endocrinology*, 4.

Marin, I. & Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science*, 281 (5385): 1990-1994.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417 (6888): 559-563.

Miyamoto, Y., Taniguchi, H., Hamel, F., Silversides, D. W. & Viger, R. S. (2008). A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *Bmc Molecular Biology*, 9.

Moen, T., Hayes, B., Baranski, M., Berg, P. R., Kjoglum, S., Koop, B. F., Davidson, W. S., Omholt, S. W. & Lien, S. (2008). A linkage map of the Atlantic salmon (Salmo salar) based on EST-derived SNP markers. *Bmc Genomics*, 9.

Nagahama, Y., Kagawa, H. & Young, G. (1982). CELLULAR SOURCES OF SEX STEROIDS IN TELEOST GONADS. *Canadian Journal of Fisheries and Aquatic Sciences*, 39 (1): 56-64.

Nakamura, M., Kobayashi, T., Chang, X. T. & Nagahama, Y. (1998). Gonadal sex differentiation in teleost fish. *Journal of Experimental Zoology*, 281 (5): 362-372.

Nakamura, M. (2010). The Mechanism of Sex Determination in Vertebrates-Are Sex Steroids the Key-Factor? *Journal of Experimental Zoology Part a-Ecological Genetics and Physiology*, 313A (7): 381-398.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z. H., Haaf, T., Shimizu, N., Shima, A., et al. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (18): 11778-11783.

NCBI. (2011). http://www.ncbi.nlm.nih.gov/.

Nilsson, E. E. & Cloud, J. G. (1992). RAINBOW-TROUT CHIMERAS PRODUCED BY INJECTION OF BLASTOMERES INTO RECIPIENT BLASTULAS. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (20): 9425-9428.

Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van den Boom, D. & Jurinke, C. (2005). iPlex Assay; Increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with mass-modified terminators. *Sequenom*.

Ohno, S., Wolf, U. & Atkin, N. B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas*, 59: 169-187.

Oshima, Y., Noguchi, K. & Nakamura, M. (2007). Expression of Lhx9 Isoforms in the developing gonads of Rana rugosa. *Zoological Science*, 24 (8): 798-802.

Phillips, R. B., Keatley, K. A., Morasch, M. R., Ventura, A. B., Lubieniecki, K. P., Koop, B. F., Danzmann, R. G. & Davidson, W. S. (2009). Assignment of Atlantic salmon (Salmo salar) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (Oncorhynchus mykiss). *BMC Genetics*, 10.

PubMed_Central. (2010). http://www.ncbi.nlm.nih.gov/pmc/.

Purves W K, S. D., Orians G H, Heller H C. (2004). *Life: The science of biology*, vol. 1: Sinauer Associates.

Purves, W. K., Sadava, D., Heller, H. C. & Orians, G. H. (2004). *Life: The science of biology*, vol. 1: Sinauer Associates.

RepeatMasker. (2003). http://www.repeatmasker.org/.

Rodriguez-Mari, A., Canestro, C., BreMiller, R. A., Nguyen-Johnson, A., Asakawa, K., Kawakami, K. & Postlethwait, J. H. (2010). Sex Reversal in Zebrafish fancl Mutants Is Caused by Tp53-Mediated Germ Cell Apoptosis. *Plos Genetics*, 6 (7).

Sandra, G. E. & Norma, M. M. (2010). Sexual determination and differentiation in teleost fish. *Reviews in Fish Biology and Fisheries*, 20 (1): 101-121.

Schartl, M. (2004). A comparative view on sex determination in medaka. *Mechanisms of Development*, 121 (7-8): 639-645.

Schultheis, C., Bohne, A., Schartl, M., Volff, J. N. & Galiana-Arnoux, D. (2009). Sex Determination Diversity and Sex Chromosome Evolution in Poeciliid Fish. *Sexual Development*, 3 (2-3): 68-77.

Shibata, Y., Paul-Prasanth, B., Suzuki, A., Usami, T., Nakamoto, M., Matsuda, M. & Nagahama, Y. (2010). Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka. *Gene Expression Patterns*, 10 (6): 283-289.

Shirak, A., Seroussi, E., Cnaani, A., Howe, A. E., Domokhovsky, R., Zilberman, N., Kocher, T. D., Hulata, G. & Ron, M. (2006). Amh and Dmrta2 genes map to tilapia (Oreochromis spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics*, 174 (3): 1573-1581.

Smith, L., Van Hateren, N., Willan, J., Romero, R., Blanco, G., Siggers, P., Walsh, J., Banerjee, R., Denny, P., Ponting, C., et al. (2003). Candidate testis-determining gene, Maestro (Mro), encodes a novel HEAT repeat protein. *Developmental Dynamics*, 227 (4): 600-607.

Stelkens, R. B. & Wedekind, C. (2010). Environmental sex reversal, Trojan sex genes, and sex ratio adjustment: conditions and population consequences. *Molecular Ecology*, 19 (4): 627-646.

Szatkowska, M., Zych, S. & Kulig, H. (2003). Sex determination in mammals. *Medycyna Weterynaryjna*, 59 (10): 847-852.

UCSC_GenomeBrowser. (2006). http://genome.ucsc.edu/.

UCSCGenomeBrowser. (2006). http://genome.ucsc.edu/.

Verhulst, E. C., van de Zande, L. & Beukeboom, L. W. (2010). Insect sex determination: it all evolves around transformer. *Current Opinion in Genetics & Development*, 20 (4): 376-383.

von Schalburg, K. R., Yasuike, M., Yazawa, R., de Boer, J. G., Reid, L., So, S., Robb, A., Rondeau, E. B., Phillips, R. B., Davidson, W. S., et al. (2011). Regulation and expression of sexual differentiation factors in embryonic and extragonadal tissues of Atlantic salmon. *Bmc Genomics*, 12.

Wang, Z. L., Zha, X. F., He, N. J., Xiang, Z. H. & Xia, Q. Y. (2010). Molecular cloning and expression analysis of Bmrbp1, the Bombyx mori homologue of the Drosophila gene rbp1. *Molecular Biology Reports*, 37 (5): 2525-2531.

Wedekind, S. (2009). Tackling the diversity of sex determination. *Biol. Lett.*, 6: 7-9.

Wei, G. & Mahowald, A. P. (1994). THE GERMLINE - FAMILIAR AND NEWLY UNCOVERED PROPERTIES. *Annual Review of Genetics*, 28: 309-324.

Wikipedia. (2011). *http://en.wikipedia.org/wiki/Bacterial_Artificial_Chromosome*.

Yano, A., Nicol, B., Guerin, A. & Guiguen, Y. (2011). The Duplicated Rainbow Trout (Oncorhynchus mykiss) T-Box Transcription Factors 1, tbx1a and tbx1b, are Up-Regulated During Testicular Development. *Molecular Reproduction and Development*, 78 (3): 172-180.

Zanin, E., Pacquelet, A., Scheckel, C., Ciosk, R. & Gotta, M. (2010). LARP-1 promotes oogenesis by repressing fem-3 in the C. elegans germline. *Journal of Cell Science*, 123 (16): 2717-2724.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7 (1-2): 203-214.

Zhou, X., Li, Q., Lu, H., Chen, H., Guo, Y. Q., Cheng, H. H. & Zhou, R. J. (2008). Fish specific duplication of Dmrt2: Characterization of zebrafish Dmrt2b. *Biochimie*, 90 (6): 878-887.

7. Appendix

7.1. Appendix I

7.1.1. Assay design

MassArray AssayDesigner software was used to make the iPlex assays. When the assays were designed, primers for these assays were ordered. A primer is a short oligonucleotide that is attached to a single stranded DNA molecule. This will provide a starting point for strand synthesis, for example PCR.

7.1.2. DNA isolation

DNA from 380 Atlantic salmon individuals from different families was previously isolated and used in this thesis.

7.1.3. Polymerase Chain Reaction (PCR)

PCR is a method that gives an exponential amplification of a selected region of the DNA.

The primers that contain sequences matching the target region and a DNA polymerase enable selective and repeated amplification. The DNA generated will again be used a template for replication, something that will give an exponential amplification.

To perform the PCR reaction, the items listed in table 2 were combined.

Table 2: PCR mix.

Reagent	Conc in 5µL	Volume (1 reaction)	Volume (384 reactions)
Nanopure H_20	NA	1.850µL	888µL
PCR Buffer with $MgCl_2$ (10x)	1.25x	0.625µL	300µL
$MgCl_2$ (25mM)	1.625mM	0.325µL	156µL
dNTP mix (25mM)	500µM	0.100µL	48µL
Primer mix (500nm each)	100nM	1.000µL	480µL
Genomic DNA (5 – 10ng/µL)	5 – 10ng/reaction	1.000µL	480µL
Hotstar Taq®(5U/µL)	0.5U/reaction	0.100µL	48µL
Total		5.000µL	2400µL

The samples were then mixed and the PCR reaction was cycled in a thermoscycler in the following way:

-94C^0 for 15 minutes

-94C^0 for 20 seconds

-56C^0 for 30 seconds

-72C^0 for 1 minute

} 45 cycles

-72C^0 for 3 minutes

-4C^0 forever

7.1.4. SAP Treatment

Often, excess dNTP that remains after PCR interfere with enzymatic reactions involving DNA synthesis. Shrimp Alkaline Phosphatase (SAP) is used to dephosphorylate unincorporated dNTPs.

SAP treatment was performed by combining the items shown in table 3.

Table 3. SAP mix

Reagent	Volume (1 reaction)	Volume (384 reactions)
Nanopure H_2O	1.530µL	734.4µL
10x SAP buffer	0.170µL	81.6µL
SAP enzyme (1U/µL)	0.300µL	144.0µL
Total	2.000µL	960.0µL

2µl of the SAP mix was added to 5µl of PCR reaction for all samples, which were then gently vortexed. The SAP treated PCR reaction was then incubated as follows:

-37C^0 for 20 minutes

-85C^0 for 5 minutes

-4C^0 forever

7.1.5. iPlex reaction

The items showed in table 4 were combined to perform the iPlex primer extension.

Table4. iPlex mix

Reagent	Conc. In 9µL	Volume (1 reaction)	Volume(384reactions)
Nanopure H_2O	NA	0.755µL	362.40µL
iPlex buffer (10x)	0.222x	0.200µL	96.00µL
iPlex termination mix	1x	0.200µL	96.00 µL
Primer mix (7µM: 14µM)	0.625µM: 1.25µM	0.804µL	385.92µL
iPlex enzyme	1x	0.041µL	19.68µL
Total		2.000µL	960.00µL

The iPlex reaction was gently vortexed and cycled by using a two-step 200 short cycles program in a thermocycler as follows:

-94C^0 for 30 seconds

-94C^0 for 5 seconds

-52C^0 for 5 seconds — 5 cycles — 40 cycles

-80C^0 for 5 seconds

-72C^0 for 3 minutes

-4C^0 forever

7.1.6. Clear resin

The iPlex reaction needs to be desalted to optimize mass spectrometric analysis. The samples were diluted with 25µl of water and 6mg of resin and centrifuged at 4000rpm to get the resin packed into the bottom of the well.

7.1.7. MALDI-TOF MS analysis

A nanodispenser was used to dispense the reaction products onto a SpectroCHIP® bioarray containing 384 elements. The iPlex SpectroCHIPS were analyzed by MALDI-TOF MS.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometer, MALDI-TOF MS, is a type of mass spectronomy used in proteomics. Peptides are ionized by a pulse of energy from a laser and accelerated down the column in the mass spectrometer to the reflector and then onto the detector. The time-of-flight is measured for the peptide, and is dependant upon the peptide's mass-to-charge ratio.

A MassArray Workstation was used to process and analyze the iPlex SpectroCHIP bioarrays.

7.2. Appendix II

Overview of candidate genes found in literature search, along with accession number and chromosome position in the stickleback genome.

Gene	Full name	Function	Reference	Accession number	Stickleback position
Aldh3a1	Aldehyde dehydrogenase 3 family, member A1	Gene over-expressed in developing testis compared with ovary	(Boyer et al. 2004)	BT045844	chrI:18367241-18372639
Am-dsx	Dsx doublesex	Conserved sex-specific transcription factor, honeybee	(Hasselmann et al. 2008; Wedekind 2009)	XM_002056562.1	
Amh	Anti-Mullerian hormone	Important gene in sex-determining pathway in mammals	(Alfaqih et al. 2009; Shirak et al. 2006)	FJ773241	chrVIII:15987194-15987552
bmp6	Bone morphogenetic protein 6		(Hoshino et al. 2005)	BC165212	chrXXI:3831600-3862459
Brd3	Bromodomain containing 3	Expressed at the genital ridge at time of sex determination in mouse	(Boyer et al. 2004)	DQ139863	chrXX:15425295-15428547
Csd	Complementary sex determiner	Complementary sex-determining gene in honeybee	(Beye et al. 2003; Wedekind 2009)	AY569721.1	
Cyp19a	Cytochrome P450, family 19, subfamily a	Gonadal aromatase	(Guiguen et al. 2010)	FJ773242	chrII:18039953-18041963
Dax1	Orphan nuclear receptor dax1	Gene of mammalian SD pathway. Tested for linkage to Y chromosome in rainbow trout.	(Alfaqih et al. 2009; von Schalburg et al. 2011)	AB253513	chrXVI:6930062-6931516
Dax2	Orphan nuclear receptor dax2	Associated to the sex determining process in vertebrates	(Davidson et al. 2009; von Schalburg et al. 2011)	DQ269441	chrXVI:6930780-6931204
Dmrt1	doublesex and mab-3 related transcription factor 1	Expresses in the embryonic gonad in vertebrates	(Matsuda et al. 2002; Nanda et al. 2002)	AF209095	chrXIII:19260865-19268641
Dmrt2b	Doublesex and mab-3 related transcription factor 2	Involved in regulating sexual development in vertebrates	(Zhou et al. 2008)	DQ307066	chrVIII:7292074-7294419
Dmrt3	doublesex and mab-3 related transcription factor 3	Expressed in the embryonic gonad in vertebrates	(Zhou et al. 2008)	AY621083	chrXIII:19242635-19245213
DMY	DMY protein	Sex determining gene in the medaka	(Matsuda et al. 2002)	AB111358	chrXIII:19251514-19268686
Fancl	Fanconi anemia, complementation group L	Expressed in developing germ cells in bipotential gonads at the critical time of sexual fate determination	(Rodriguez-Mari et al. 2010)	BC055509	
Fem	Feminizer	Feminizer gene in honeybee	(Wedekind 2009)(Hasselmann, 2008)	EU101388.1	

Fgf9	Fibroblast growth factor 9	Glia-activating factor	(Nakamura 2010)	BC099872	chrIV:11966280-11970315
Foxl2	Forkhead box L2	Follicular developm. and ovary funct.	(Baroiller, 2009; (Loffler et al. 2003)	AB252055	chrI:1811639-1812556
Fru	Fruitless	Determiner of male sexual behavior in insects	(Verhulst et al. 2010)	BT047704	
Ftz-F1	Fushi tarazu factor 1	Regulator of sex steroid production	(Manolakou et al. 2006)(Hofsten 2005; Krovel, 2004)	AY879313	chrXIII:732331-745073
Gata4	GATA binding protein 4	Regualte transcription of genes required of mammalian sex determination and development	(Davidson et al. 2009; Miyamoto et al. 2008)	HM475152	chrXVIII:10178943-10179497
Gsdf	Gonadal somatic cell derived factor	Involved in early testicular differentiation in medaka	(Shibata et al. 2010)	DQ489287.1	chrXIV:3575485-3577470
Igf1	Insulin-like growth factor 1	Insulin-like-growth-factor, tilapia	(Baroiller, 2009; Berishvili, 2006)	EF432852	chrIV:32098250-32108229
Larp1	La ribonucleoprotein domain family, member 1	Regulator of sex determination in C. elegans	(Zanin et al. 2010)	EZ789304	chrXV:719505-721303
Lhx9	LIM homeobox 9	Expressed in developing gonads	(Nakamura 2010; Oshima et al. 2007)	AY534647	chrVIII:8960269-8965720
LIM1	LIM-homeodomain protein	Genetic element involved in sex in mammals	(Szatkowska et al. 2003)	L37802	chrI:19051879-19058785
Mep1	Meprin 1	Necessary for oocyte production and somatic differentiation	(Kasturi et al. 2010)	BC085666	chrXVIII:918005-921652
Mis12	MIND kinetochore complex component	Anti-Müllerian hormone	(Sandra & Norma 2010)	BT056345	chrII:13580807-13582114
Mro	Maestro	Candidate testis-determining gene in mouse	(Smith et al. 2003)	NM_027741.2	
Ppt1	Palmitoyl-protein thioesterase 1	Expressed at the genital ridge at time of sex determination in mouse	(Boyer et al. 2004)	BT026911	chrX:2614046-2618466
Rbp1	RNA-binding protein 1	Involved in alternative splicing of the pre-mRNA of Drosophila sex-determining gene dsx.	(Wang et al. 2010)	BT027529	chrIII:8071205-8073800
Rspo1	R-spondin homolog	Gene involved in the vertebrate sex-determination cascade	(Hale et al. 2010)	BC075907	chrXX:17223708-17228139
Sf1	Splicing factor 1	Steroidogenic factor. Plays a role in transcriptional regulation of aromatase	(Davidson et al. 2009; von Schalburg et al. 2011)	BC163938	
Shbgb	Sex hormone-binding globulin beta form	Strongly expressed in ovary in rainbow trout	(Bobe et al. 2008; Bobe et al. 2010)	EF577269.1	chrII:7712632-7713485
Sox01	SRY-box containing gene 1	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	AB242327	chrXVI:7107947-7109323

Sox02	SRY-box containing gene 2	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	AB242329	chrVIII:6157588-6159515
Sox03	SRY-box containing gene 3	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BC092845	chrIV:5786326-5787968
Sox04	SRY-box containing gene 4	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT059154	chrXX:7263735-7267139
Sox05	SRY-box containing gene 5	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	EF577484	chrXIX:7849957-7895787
Sox06	SRY-box containing gene 6	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	EU532205	chrII:9549106-9622576
Sox07	SRY-box containing gene 7	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT071875	chrI:22919282-23146252
Sox08	SRY-box containing gene 8	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	DQ222970	chrXI:9561734-9563288
Sox09	SRY-box containing gene 9	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT027942	chrXXI:9690607-9692667
Sox10	SRY-box containing gene10	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BC163868	chrIX:840851-1139576
Sox11	SRY-box containing gene 11	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT059525	chrXV:6609369-6611563
Sox12	SRY-box containing gene 12	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT028068	chrXVI:7107383-7109312
Sox13	SRY-box containing gene 13	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	FJ895596	chrXVII:11232062-11243854
Sox14	SRY-box containing gene 14	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BC108033	chrVIII:3002182-3003756
Sox15	SRY-box containing gene 15	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BC119067	chrVII:10357222-10357479
Sox16	SRY-box containing gene 16	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	L29084	chrI:22995741-22995899
Sox17	SRY-box containing gene 17	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BT026959	chrXVIII:10826850-10829918
Sox18	SRY-box containing gene 18	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	FJ895599	chrXII:13373253-13376656
Sox19	SRY-box containing gene 19	Expressed during embryogenesis in teleosts	(Fukada et al. 1995)	BC078221	chrVII:10357234-10359541
Sparc	Secreted acidic cysteine rich glycoprotein	Gene over-expressed in developing testis compared with ovary	(Boyer et al. 2004)	DT969625	chrIV:4486109-4491419
Sry	Sex determining region Y	Primary male-determining factor in most mammals. Se introduction	(Artieri et al. 2006; Nanda et al. 2002)	AJ245396	chrIV:5786543-5788177
SSB-4	splA/ryanodine receptor domain and SOCS box containing 4	Plays a role in the early development of germ cells in zebrafish	(Nanda et al. 2002; Sandra & Norma 2010)	BC134936	chrIII:8179290-8202036
Sxl	Sex lethal	Top of SD cascade in drosphilia	(Marin & Baker 1998; Schartl 2004)	BT028153	chrVIII:17255045-17260661

Tbx1	t-box transcription factor	Involved in embryogenesis and organogenesis	(Yano et al. 2011)	AY277631	chrXIII:15665783-15672536
tp53	Tumor protein p53	Induces sex reversal after the mutation of a DNA-repair pathway gene in zebrafish	(Rodriguez-Mari et al. 2010)	BT026593	chrVII:4324272-4329205
tra(tra2b)	Transformer	Central player in the evolution of sex determination in insects.	(Verhulst et al. 2010)	BT045504	chrXV:8495570-8497467
Vasa	Vasa	Gene overexpressed in developing testis compared with ovary	(Fujiwara et al. 1994; Manolakou et al. 2006)	EZ797383	chrX:9086205-9086829
Vcp	Valosin-containing protein	Sex related in zebrafish, only in gonads	(Boyer et al. 2004)	CD505418	chrI:26988970-26993741
Vnn1	Vanin1	Gene overexpressed in developing testis compared with ovary	(Boyer et al. 2004; Grimmond et al. 2000)	BT026963	chrVI:610365-621924
Wnt4	Wingless-type MMTV integration site family, member 4	Genes involved in the vertebrate sex-determination cascade	(Hale et al. 2010)	U25141	chrXVII:11160291-11163025
Wt1	Wilms tumor 1 homolog	Involved in the vertebrate sex determination cascade	(Hale et al. 2010)	BT059480	chrXIX:3225572-3240225
Xol	XO lethal	Top of SD cascade in C.elegans	(Schartl 2004)	NM_001029248.1	

7.3. Appendix III

The table below shows what genes were found in Atlantic salmon after the unigene BLAST performed in Galaxy, along with the sequence bit-score and identity in percent. It also shows possible paralogues for several of the genes, marked with an "x" in the "Paralog"-column, along with its bit-score and identity. The genes that have another species than Salmo salar in the "Species"-column are the genes that did not have any good hits in the unigene-assembly and therefore kept their original sequences.

Gene	Species	BLAST-score	Identity %	Paralog	BLAST-score	Identity %
Aldh3a1	Salmo salar	3915	100	x	914	94,19
Am-dsx	Salmo salar	50	100			
Amh	Salmo salar	50	91,89			
bmp6	Salmo salar	172	83,21			
Brd3	Salmo salar	1118	97,32			
csd	Apis mellifera					
Cyp19a	Salmo salar	137	82,89			
Dax1	Oryzias latipes					
Dax2	Oreochromis niloticus					
Dmrt1	Oncorhyncus mykiss					
Dmrt2b	Salmo salar	139	84,7			
Dmrt3	Danio rerio					
DMY	Oryzias latipes					
fancl	Salmo salar	141	80,84			
fem	Apis mellifera					
Fgf9	Mus musculus					
Foxl2	Salmo salar	60	94,74			
fru	Salmo salar	2238	100	x	1084	93,72
Ftz-F1	Salmo salar	331	92			
Gata4	Salmo salar	745	100			
Gsdf	Salmo salar	938	94,31			
Igf1	Salmo salar	1822	100			
Larp1	Salmo salar	559	98,34			
Lhx9	Oryzias latipes					
LIM1	Salmo salar	123	89,29			
Mep1	Salmo salar	77,8	94,12			

Mis12	Salmo salar	2377	100	x	626	91,97
Mro	Mus musculus					
Ppt1	Salmo salar	319	80,95			
Rbp1	Salmo salar	357	87,4			
Rspo1	Danio rerio					
Sf1	Salmo salar	75,8	82,81			
Shbgb	Salmo salar	1314	96,72			
Sox01	Salmo salar	333	82,85			
Sox02	Salmo salar	805	85,68			
Sox03	Salmo salar	706	89,04			
Sox04	Salmo salar	2034	100			
Sox05	Oryzias latipes					
Sox06	Danio rerio					
Sox07	Salmo salar	1788	100			
Sox08	Salmo salar	1211	100			
Sox09	Salmo salar	389	84,27			
Sox10	Danio rerio					
Sox11	Salmo salar	4175	100			
Sox12	Stickleback					
Sox13	Oryzias latipes					
Sox14	Salmo salar	125	82,57			
Sox15	Mus musculus					
Sox16	Mus musculus					
Sox17	Stickleback					
Sox18	Oryzias latipes					
Sox19	Danio rerio					
Sparc	Salmo salar	519	84,2			
Sry	Salmo salar	722	89,43			
SSB-4	Salmo salar	61,9	94,87			
Sxl	Salmo salar	434	81,43			
Tbx1	Salmo salar	97,6	83,03			
tp53	Salmo salar	113	86,26			
tra2b	Salmo salar	3471	100	x	446	87,3
vasa	Salmo salar	377	88,99			

Vcp	Salmo salar	101	80,07			
Vnn1	Salmo salar	119	79,94			
Wnt4	Danio rerio					
Wt1	Salmo salar	4849	100	x	311	86,07
xol	Caenorhabditis elegans					

7.4. Appendix IV

Sequences selected from arachne assembly after removing repeats. 1 in the left column means it was the prime selection, while 2 means it was selected as a probable paralog. The table shows the gene name, what contig the BLAST sequence hit, identity in % of the hit and e-value and bit-score. The table also shows the start- (start seq) and end (end seq) positions in the query sequence along with length of the sequence hit (length seq), as well as start- and end positions in the hit, along with length of the hit.

	Gene	Contig	Identity %	Start seq	End seq	Start hit	End hit	Length seq	Length hit	E-value	BLAST-score
1	Aldh3a1	3086	**100**	1295	2002	9103	8396	707	707	0	**1350**
2	Aldh3a1	328899	**89,22**	591	756	591	756	165	-165	2,00E-44	**186**
1	Am-dsx	47433	**100**	194	597	3205	3608	403	-403	0	**801**
2	Am-dsx	306119	**95,33**	105	597	1349	860	492	489	0	**791**
1	Amh	1158	**100**	63	502	2706	3145	439	-439	0	**872**
1	bmp6	292939	**100**	55	170	1220	1105	115	115	5,00E-58	**230**
2	bmp6	58891	**84,42**	168	319	4483	4634	151	-151	3,00E-22	**111**
1	Brd3	19734	**97,55**	142	467	7027	6702	325	325	3,00E-164	**583**
2	Brd3	285660	**90,2**	198	441	1452	1209	243	243	4,00E-77	**293**
1	Cyp19a	131600	**98,78**	110	354	1459	1215	244	244	4,00E-128	**462**
1	Dax1	38585	**85,65**	373	587	295	509	214	-214	7,00E-43	**180**
1	Dmrt1	153561	**96,64**	866	1014	1625	1773	148	-148	1,00E-65	**256**
1	Dmrt2b	153822	**88,41**	5	463	4133	4586	458	-453	4,00E-132	**476**
1	Dmrt3	237168	**87,42**	53	352	2364	2065	299	299	8,00E-77	**293**
1	fancl	186346	**91,14**	27	105	1892	1970	78	-78	3,00E-19	**101**
1	Foxl2	24908	**100**	5	849	11144	10300	844	844	0	**1675**
2	Foxl2	32199	**93,56**	5	846	10131	10980	841	-849	0	**1241**
1	fru	205014	**99,5**	532	1129	946	347	597	599	0	**1164**
1	fru|para	308298	**100**	6	252	1106	1352	246	-246	3,00E-136	**490**
1	Gata4	329555	**100**	170	338	2397	2229	168	168	8,00E-90	**335**
1	Gsdf	67113	**84,66**	306	466	3162	3322	160	-160	3,00E-25	**121**

1	Igf1	122954	**100**	430	693	1477	1214	263	263	3,00E-146	**523**
2	Igf1	97742	**94,26**	691	890	1204	1411	199	-207	8,00E-82	**309**
1	Larp1	296205	**99,56**	256	707	2706	3158	451	-452	0	**880**
2	Larp1	132785	**91,98**	256	492	1040	1276	236	-236	6,00E-85	**319**
1	Lhx9	94990	**89,36**	124	311	306	119	187	187	3,00E-53	**214**
1	LIM1	36118	**90,99**	2	322	9060	9381	320	-321	1,00E-82	**311**
1	Mep1	26282	**100**	1	219	1696	1914	218	-218	1,00E-119	**434**
1	Mis12	25087	**99,59**	48	1266	2342	3558	1218	-1216	0	**2278**
1	Ppt1	35342	**95,28**	262	387	1	127	125	-126	1,00E-49	**202**
1	Sf1	16866	**100**	5	70	319	254	65	65	3,00E-28	**131**
2	Sf1	453844	**92,98**	69	239	1122	1292	170	-170	4,00E-52	**210**
1	Shbgb	156204	**100**	278	469	4361	4170	191	191	2,00E-103	**381**
2	Shbgb	373630	**89,02**	1	163	1490	1330	162	160	2,00E-42	**178**
1	sox01	128661	**99,65**	1	571	3157	2587	570	570	0	**1116**
1	sox02	4110	**100**	2010	3982	4001	2029	1972	1972	0	**3869**
1	sox03	36013	**100**	1	836	8628	9463	835	-835	0	**1616**
1	sox04	65696	**99,91**	1373	2482	5931	4822	1109	1109	0	**2026**
2	sox04	115355	**84,72**	626	836	6292	6502	210	-210	3,00E-35	**157**
1	sox05	180964	**93,16**	1625	1740	3736	3851	115	-115	2,00E-38	**167**
1	sox06	118503	**88,53**	1817	2032	3285	3070	215	215	2,00E-57	**230**
1	sox07	388189	**100**	1111	1249	1825	1963	138	-138	2,00E-71	**276**
1	sox09	68819	**93,11**	55	350	2165	1861	295	304	1,00E-117	**428**
1	sox10	22318	**84,84**	1534	1775	1177	1418	241	-241	2,00E-44	**186**
1	sox11	17756	**100**	1	2231	3749	1519	2230	2230	0	**4173**
2	sox11	92274	**89,77**	121	1278	4862	6027	1157	-1165	0	**1352**
1	sox12	217913	**88,12**	236	655	504	85	419	419	1,00E-119	**436**

1	sox13	332043	**91,3**	1585	1767	1031	849	182	182	2,00E-59	**236**
1	sox19	68688	**80,8**	356	847	2155	2646	491	-491	8,00E-53	**214**
1	Sparc	81860	**100**	947	1371	3308	3732	424	-424	0	**842**
1	Sry	36013	**100**	1	836	8628	9463	835	-835	0	**1616**
1	SSB-4	109444	**99,49**	159	743	156	741	584	-585	0	**1035**
1	Sxl	79386	**99,91**	770	1862	1874	2967	1092	-1093	0	**2034**
2	Sxl	359836	**92,1**	770	1848	752	1855	1078	-1103	0	**1356**
1	Tbx1	456355	**99,74**	5	770	110	875	765	-765	0	**1455**
2	Tbx1	45283	**86,67**	5	544	5502	6040	539	-538	9,00E-134	**482**
1	tp53	52504	**99,89**	1949	2843	438	1332	894	-894	0	**1766**
1	tra2b	50273	**99,81**	1015	1549	2332	1798	534	534	0	**1053**
1	tra2b|para	94974	**94,37**	295	364	1077	1009	69	68	2,00E-20	**105**
1	vasa	159994	**99,45**	283	464	240	59	181	181	5,00E-95	**353**
1	Vcp	41651	**95,57**	125	282	2541	2695	157	-154	2,00E-65	**254**
1	Vnn1	95796	**99,75**	453	846	672	279	393	393	0	**773**
1	Wt1	176038	**100**	41	748	487	1194	707	-707	0	**1320**
1	Wt1|para	46633	**97,95**	474	863	9717	9327	389	390	0	**710**

7.5. Appendix V

The table displays the putative SNPs selected for each gene after the SNP detection in section 2.6. SNPs were selected based on frequency.

SNP		Distribution			Frequency			
Aldh3a1_3040	A	|:AG-2:AA-21:GG-7|30|63|A-0.73G-0.27	30	63	**A**	0,73	**G**	0,27
Aldh3a1_5773	G	|:TG-6:TT-4:GG-11|21|35|T-0.33G-0.67	21	35	**T**	0,33	**G**	0,67
Aldh3a1_5990	T	|:TG-1:TT-12:GG-5|18|32|T-0.69G-0.31	18	32	**T**	0,69	**G**	0,31
Amh_1137	A	|:CC-2:AA-21:CA-4|27|67|A-0.85C-0.15	27	67	**A**	0,85	**C**	0,15
Amh_3718	T	|CC-3TT-10CT-0|13|35|T-0.77C-0.23	13	35	**T**	0,77	**C**	0,23
Bmp6_165	A	|AA-10GG-1AG-0|11|20|A-0.91G-0.09	11	20	**A**	0,91	**G**	0,09
Bmp6|para_167	A	|AA-17GG-1AG-0|18|36|A-0.94G-0.06	18	36	**A**	0,94	**G**	0,06
Bmp6|para_176	G	|:AG-1:AA-2:GG-15|18|42|A-0.14G-0.86	18	42	**A**	0,14	**G**	0,86
Brd3_1015	C	|:CC-14:AA-9:CA-1|24|78|A-0.40C-0.60	24	78	**A**	0,4	**C**	0,6
Brd3_9613	A	|:AG-2:AA-27:GG-3|32|59|A-0.88G-0.12	32	59	**A**	0,88	**G**	0,12
Brd3|para_1540	G	|:TG-2:TT-1:GG-25|28|66|T-0.07G-0.93	28	66	**T**	0,07	**G**	0,93
Brd3|para_1709	T	|CC-2TT-10CT-0|12|27|T-0.83C-0.17	12	27	**T**	0,83	**C**	0,17
Cyp19a_1599	A	|:AG-1:AA-17:GG-1|19|69|A-0.92G-0.08	19	69	**A**	0,92	**G**	0,08
Cyp19a_1620	A	|:AG-2:AA-21:GG-8|31|84|A-0.71G-0.29	31	84	**A**	0,71	**G**	0,29
Dax1	T	|TG-1TT-17GG-0|18|29|T-0.97G-0.03	18	29	**T**	0,97	**G**	0,03
Dmrt1_2031	T	|TT-4CT-1CC-0|5|14|T-0.90C-0.10	5	14	**T**	0,9	**C**	0,1
Dmrt1_3081	A	|:AG-7:AA-15:GG-4|26|90|A-0.71G-0.29	26	90	**A**	0,71	**G**	0,29
Dmrt2b_4439	C	|CC-18AA-1CA-0|19|28|A-0.05C-0.95	19	28	**A**	0,05	**C**	0,95
Dmrt2b_852	T	|:CC-5:TT-10:CT-1|16|28|T-0.66C-0.34	16	28	**T**	0,66	**C**	0,34
Dmrt3_1688	G	|TT-1GG-18TG-0|19|34|T-0.05G-0.95	19	34	**T**	0,05	**G**	0,95
Dmrt3_2279	T	|CC-1TT-19CT-0|20|39|T-0.95C-0.05	20	39	**T**	0,95	**C**	0,05
Fancl	T	|TT-23GG-1TG-0|24|44|T-0.96G-	24	44	**T**	0,96	**G**	0,04

		0.04						
Foxl2_12516	A	|CC-2CA-15AA-0|17|20|A-0.44C-0.56	17	20	**A**	0,44	**C**	0,56
Foxl2_5632	G	|AA-4GG-13AG-0|17|28|A-0.24G-0.76	17	28	**A**	0,24	**G**	0,76
Foxl2_8960	T	|TT-12GG-13TG-0|25|66|T-0.48G-0.52	25	66	**T**	0,48	**G**	0,52
Foxl2|para_10210	G	|CC-8GG-12CG-0|20|40|C-0.40G-0.60	20	40	**C**	0,4	**G**	0,6
Foxl2|para_10283	T	|TT-11GG-4TG-0|15|20|T-0.73G-0.27	15	20	**T**	0,73	**G**	0,27
Fru|para	G	|AA-1GG-25AG-0|26|49|A-0.04G-0.96	26	49	**A**	0,04	**G**	0,96
Gsdf_3248	A	|:AG-1:AA-21:GG-1|23|39|A-0.93G-0.07	23	39	**A**	0,93	**G**	0,07
Gsdf_8703	A	|AG-2AA-16GG-0|18|29|A-0.94G-0.06	18	29	**A**	0,94	**G**	0,06
Larp1_2862	A	|:AT-11:AA-4:TT-4|19|26|A-0.50T-0.50	19	26	**A**	0,5	**T**	0,5
Larp1|para_1965	A	|CC-1AA-16CA-0|17|24|A-0.94C-0.06	17	24	**A**	0,94	**C**	0,06
Lhx9_1928	T	|:TG-1:TT-22:GG-1|24|53|T-0.94G-0.06	24	53	**T**	0,94	**G**	0,06
Lhx9_5633	G	|AA-9GG-13AG-0|22|31|A-0.41G-0.59	22	31	**A**	0,41	**G**	0,59
Lim1_10920	C	|:CC-12:TT-1:CT-7|20|57|T-0.23C-0.78	20	57	**T**	0,23	**C**	0,78
Lim1_8568	A	|:CC-1:AA-15:CA-2|18|35|A-0.89C-0.11	18	35	**A**	0,89	**C**	0,11
Mep1_4877	T	|CC-1TT-20CT-0|21|39|T-0.95C-0.05	21	39	**T**	0,95	**C**	0,05
Mep1_728	A	|CC-1AA-16CA-0|17|28|A-0.94C-0.06	17	28	**A**	0,94	**C**	0,06
Mis12_3207	T	|CC-2TT-17CT-0|19|45|T-0.89C-0.11	19	45	**T**	0,89	**C**	0,11
Ppt1_2246	T	|:CC-9:TT-15:CT-4|28|57|T-0.61C-0.39	28	57	**T**	0,61	**C**	0,39
Sf1	T	|TT-27GG-1TG-0|28|56|T-0.96G-0.04	28	56	**T**	0,96	**G**	0,04
Sf1|para_691	G	|:CC-8:CG-1:GG-22|31|99|C-0.27G-0.73	31	99	**C**	0,27	**G**	0,73
Shbgb|para_1384	T	|TT-21GG-1TG-0|22|43|T-0.95G-0.05	22	43	**T**	0,95	**G**	0,05
Sox01_1406	A	|AG-8AA-16GG-0|24|60|A-0.83G-0.17	24	60	**A**	0,83	**G**	0,17
Sox01_6368	A	|:CC-3:AA-14:CA-1|18|46|A-0.81C-0.19	18	46	**A**	0,81	**C**	0,19

Sox02_4652	C	|CC-18TT-1CT-0|19|34|T-0.05C-0.95	19	34	**T**	0,05	**C**	0,95
Sox02_6616	G	|:CC-1:CG-1:GG-20|22|47|C-0.07G-0.93	22	47	**C**	0,07	**G**	0,93
Sox04|para_6340	G	|:AG-1:AA-1:GG-24|26|47|A-0.06G-0.94	26	47	**A**	0,06	**G**	0,94
Sox05_3620	C	|CC-20AA-1CA-0|21|26|A-0.05C-0.95	21	26	**A**	0,05	**C**	0,95
Sox06_3565	A	|:AG-10:AA-6:GG-10|26|56|A-0.42G-0.58	26	56	**A**	0,42	**G**	0,58
Sox06_6105	C	|:CC-18:CG-2:GG-1|21|42|C-0.90G-0.10	21	42	**C**	0,9	**G**	0,1
Sox06_7845	T	|:TG-1:TT-18:GG-2|21|47|T-0.88G-0.12	21	47	**T**	0,88	**G**	0,12
Sox10_2096	C	|:CC-17:GG-2:CG-10|29|68|C-0.76G-0.24	29	68	**C**	0,76	**G**	0,24
Sox10_3309	T	|:CC-6:TT-17:CT-2|25|59|T-0.72C-0.28	25	59	**T**	0,72	**C**	0,28
Sox10_3349	C	|CC-16GG-3CG-0|19|52|C-0.84G-0.16	19	52	**C**	0,84	**G**	0,16
Sox11_258	T	|TT-13GG-2TG-0|15|37|T-0.87G-0.13	15	37	**T**	0,87	**G**	0,13
Sox11|para_2898	T	|CC-2TT-28CT-0|30|52|T-0.93C-0.07	30	52	**T**	0,93	**C**	0,07
Sox12_1208	A	|:AG-1:AA-13:GG-5|19|33|A-0.71G-0.29	19	33	**A**	0,71	**G**	0,29
Sox12_469	G	|CC-9GG-12CG-0|21|37|C-0.43G-0.57	21	37	**C**	0,43	**G**	0,57
Sox13_856	T	|TT-19GG-1TG-0|20|33|T-0.95G-0.05	20	33	**T**	0,95	**G**	0,05
Sox19_2748	G	|:CC-1:CG-2:GG-25|28|58|C-0.07G-0.93	28	58	**C**	0,07	**G**	0,93
Sparc_2077	A	|:CC-7:AA-9:CA-12|28|62|A-0.54C-0.46	28	62	**A**	0,54	**C**	0,46
Sparc_2411	G	|CC-3GG-18CG-0|21|38|C-0.14G-0.86	21	38	**C**	0,14	**G**	0,86
Sry_2416	C	|CC-16AA-1CA-0|17|28|A-0.06C-0.94	17	28	**A**	0,06	**C**	0,94
Sry_7513	T	|AA-2TT-28AT-0|30|54|A-0.07T-0.93	30	54	**A**	0,07	**T**	0,93
Ssb-4	G	|AA-1GG-22AG-0|23|36|A-0.04G-0.96	23	36	**A**	0,04	**G**	0,96
Ssb-4	T	|AA-1TT-26AT-0|27|51|A-0.04T-0.96	27	51	**A**	0,04	**T**	0,96
Ssb-4_1172	T	|:TG-2:TT-32:GG-1|35|71|T-0.94G-0.06	35	71	**T**	0,94	**G**	0,06
Sxl_3541	C	|AA-1CC-20AC-0|21|37|A-0.05C-0.95	21	37	**A**	0,05	**C**	0,95

Sxl|para_1106	T	|:CC-1:TT-19:CT-1|21|35|T-0.93C-0.07	21	35	**T**	0,93	**C**	0,07
Tbx1_1234	G	|AA-1GG-18AG-0|19|36|A-0.05G-0.95	19	36	**A**	0,05	**G**	0,95
Tbx1|para_1971	A	|CC-2AA-21CA-0|23|46|A-0.91C-0.09	23	46	**A**	0,91	**C**	0,09
Tbx1|para_3279	G	|AA-1GG-11AG-0|12|25|A-0.08G-0.92	12	25	**A**	0,08	**G**	0,92
Tp53_4247	T	|:TG-1:TT-18:GG-1|20|47|T-0.93G-0.07	20	47	**T**	0,93	**G**	0,07
Tp53_6140	A	|AA-9TT-1AT-0|10|40|A-0.90T-0.10	10	40	**A**	0,9	**T**	0,1
Tp53_6198	T	|TT-14GG-1TG-0|15|44|T-0.93G-0.07	15	44	**T**	0,93	**G**	0,07
Tra2b_4050	T	|:CC-1:TT-26:CT-1|28|58|T-0.95C-0.05	28	58	**T**	0,95	**C**	0,05
Tra2b_6890	G	|CC-2GG-15CG-0|17|32|C-0.12G-0.88	17	32	**C**	0,12	**G**	0,88
Tra2b|para_9011	C	|:CC-24:AA-1:CA-1|26|59|A-0.06C-0.94	26	59	**A**	0,06	**C**	0,94
Vasa_67	T	|:AT-2:AA-14:TT-16|32|77|A-0.47T-0.53	32	77	**A**	0,47	**T**	0,53
Vcp_2698	A	|CC-1AA-12CA-0|13|28|A-0.92C-0.08	13	28	**A**	0,92	**C**	0,08
Vnn1_542	C	|CC-28AA-3CA-0|31|62|A-0.10C-0.90	31	62	**A**	0,1	**C**	0,9
Wt1_132	T	|CC-1TT-20CT-0|21|36|T-0.95C-0.05	21	36	**T**	0,95	**C**	0,05
Wt1|para_7278	T	|AA-1TT-15AT-0|16|20|A-0.06T-0.94	16	20	**A**	0,06	**T**	0,94
Wt1|para_8903	G	|:CC-4:CG-3:GG-24|31|69|C-0.18G-0.82	31	69	**C**	0,18	**G**	0,82
Wt1|para_9590	T	|:CC-1:TT-13:CT-1|15|29|T-0.90C-0.10	15	29	**T**	0,9	**C**	0,1

7.6. Appendix VI

The table below displays the MassARRAY multiplexes and what SNPs went into what multiplex. There are four multiplexes with a total of 81 SNPs.

WELL	TERM	SNP_ID
W1	iPLEX	Sry_7513
W1	iPLEX	Shbgb_para_1384
W1	iPLEX	Amh_1137
W1	iPLEX	Dmrt3_1688
W1	iPLEX	Sox01_6368
W1	iPLEX	Dmrt2b_4439
W1	iPLEX	Sox02_6616
W1	iPLEX	Brd3_9613
W1	iPLEX	Bmp6_165
W1	iPLEX	Wt1_para_9590
W1	iPLEX	Mep1_4877
W1	iPLEX	Bmp6_1384
W1	iPLEX	Gsdf_8703
W1	iPLEX	Gata4_221
W1	iPLEX	Tbx1_para_1971
W1	iPLEX	Sxl_3541
W1	iPLEX	Tbx1_1234
W1	iPLEX	Tra2b_para_9011
W1	iPLEX	Wt1_132
W1	iPLEX	Ssb-4_1172
W1	iPLEX	Sox06_6105
W1	iPLEX	Sox12_1208
W1	iPLEX	Vcp_2698
W2	iPLEX	Lhx9_5633
W2	iPLEX	Amh_3718
W2	iPLEX	Larp1_para_1965
W2	iPLEX	Sox04_para_6340
W2	iPLEX	Bmp6_para_176
W2	iPLEX	Sox05_3620
W2	iPLEX	Lim1_8568
W2	iPLEX	Sry_2416
W2	iPLEX	Sox02_4652
W2	iPLEX	Gsdf_3248
W2	iPLEX	Sox12_469
W2	iPLEX	Sox01_1406
W2	iPLEX	Sox06_3565
W2	iPLEX	Fru_3622
W2	iPLEX	Shbgb_5575
W2	iPLEX	Sox13_856
W2	iPLEX	Foxl2_para_10283
W2	iPLEX	Sox06_7845
W2	iPLEX	Tp53_4247
W2	iPLEX	Tp53_6140
W2	iPLEX	Ppt1_2246
W2	iPLEX	Foxl2_5632
W2	iPLEX	Cyp19a_1599
W3	iPLEX	Sxl_para_1106
W3	iPLEX	Sparc_2411
W3	iPLEX	Brd3_1015
W3	iPLEX	Dmrt1_2031
W3	iPLEX	Aldh3a1_3040
W3	iPLEX	Brd3_para_1709
W3	iPLEX	Tra2b_6890
W3	iPLEX	Tra2b_4050
W3	iPLEX	Vnn1_542
W3	iPLEX	Sox11_para_2898
W3	iPLEX	Sf1_para_691
W3	iPLEX	Sox10_3349
W3	iPLEX	Brd3_9900
W3	iPLEX	Foxl2_12516
W3	iPLEX	Dmrt1_3081
W3	iPLEX	Sox19_2748
W3	iPLEX	Wt1_para_8903
W3	iPLEX	Mis12_3207
W3	iPLEX	Sox11_258
W3	iPLEX	Mep1_1532
W4	iPLEX	Cyp19a_1620
W4	iPLEX	Vasa_67
W4	iPLEX	Bmp6_para_167
W4	iPLEX	Tp53_6198
W4	iPLEX	Foxl2_8960
W4	iPLEX	Lim1_10920
W4	iPLEX	Sox10_2096

W4	iPLEX	Sox10_3309
W4	iPLEX	Brd3_para_1540
W4	iPLEX	Mep1_728
W4	iPLEX	Dmrt3_2279
W4	iPLEX	Foxl2_para_10210
W4	iPLEX	Sparc_2077
W4	iPLEX	Larp1_2862
W4	iPLEX	Lhx9_1928

7.7. Appendix VII

The table below shows the result of the megaBLAST against the scaffold assembly and the scaffold mapping, including those genes that could not be mapped by scaffolds or fingerprints. The columns show the scaffold the gene sequence is located in, the chromosome the scaffold is located on, along with the hit's identity, e-value and bit-score. The fingerprint is also shown. The chrosmosmes shaded grey or yellow are for genes mapped by scaffolds. "x" in the SsaChr column means that the gene could not be mapped to a chromosome. In the fingerprint column, the "x" means that the gene couldn't be assigned to any fingerprint.

SNP	Scaffold	Scaffold Length	SsaChr	Identity %	E-value	BLAST score	Fingerprint
Aldh3a1	scf15186897	131 892	ssa09	100	0	1350	fps4250
Aldh3a1|para	scf15186897	131 892	ssa09	94,19	0	848	fps4250
Am-dsx	scf15179227	9 767	ssa03	100	0	801	fps367
Amh	scf15097739	184 336	ssa10	100	0	872	fps1652
bmp6	scf15190741	224 571	ssa22	84,42	6,00E-22	111	fps1387
bmp6	scf15173904	9 871	x	99,5	0	727	fps568
Brd3	scf15193845	113 545	ssa05	93,88	5,00E-53	214	fps5544
Brd3	scf15113107	110 669	ssa02	100	0	646	fps349
Cyp19a	scf15137108	7 450	x	98,78	7,00E-128	462	x
Dmrt1	scf15186772	8 787	x	96,64	2,00E-65	256	x
Dmrt2b	scf15098855	46 957	x	88,41	6,00E-132	476	x
Dmrt2b	scf15192969	216 011	ssa23	100	0	918	fps4585
Dmrt3|daniorerio	scf15152654	12 418	x	87,42	1,00E-76	293	fps4033
fancl	scf15193879	148 801	ssa18	100	0	971	fps3628
Foxl2	scf15102427	8 107	x	86,71	2,00E-132	478	x
Foxl2	scf15188563	83 241	x	100	0	1675	fps1743
fru	scf15155360	118 555	x	99,5	0	1164	fps2075
fru|para	scf15124131	172 917	x	100	5,00E-136	490	x
Gata4	scf15105265	23 654	x	87,22	2,00E-26	125	x

Gata4	scf15184911	175 318	ssa06	100	1,00E-89	335	fps1461
Gsdf	scf15153204	83 814	x	84,66	4,00E-25	121	fps286
Gsdf	scf15163005	10 098	x	100	1,00E-173	614	x
Larp1	scf15130158	91 331	x	99,56	0	880	fps486
Lhx9|medaka	scf15191338	127 091	x	84,24	2,00E-64	252	fps1618
LIM1	scf15160112	10 422	x	93,8	4,00E-124	450	x
Mep1	scf15123091	46 787	x	100	2,00E-119	434	x
Mis12	scf15190809	8 260	ssa1/13	86,88	0	795	fps328
Mis12	scf15189589	273 462	x	99,67	0	2286	fps7331
Ppt1	scf15192456	293 928	ssa27	100	2,00E-31	143	fps1317
Rbp1	scf15188577	271 788	ssa03	89,19	4,00E-37	161	fps877
Rbp1	scf15144724	117 270	x	100	1,00E-102	379	x
Sf1	scf15194451	770 215	ssa18	92,98	7,00E-52	210	
Sf1	scf15246873	1 573	x	100	1,00E-74	285	x
Shbgb	scf15194529	401 512	ssa09/28	88,71	2,00E-49	202	fps2333
Shbgb	scf15100754	25 687	x	100	4,00E-103	381	x
Sox01/Sox12	scf15125643	15 093	x	88,07	6,00E-166	589	x
Sox01/Sox12	scf15171912	12 426	ssa12	100	2,00E-172	611	fps2932
Sox02	scf15170786	168 559	ssa19	82,8	1,00E-32	147	fps1709
Sox02	scf15133941	9 473	ssa19	100	0	825	fps1831
Sox03	scf15149721	12 565	x	89,77	0	1352	fps2628
Sox03	scf15186386	123 970	ssa06	100	0	4173	fps496
Sox04/Sox14	scf15182287	18 215	x	100	0	1790	x
Sox06	scf15129926	276 217	ssa10	95,46	0	1064	fps352
Sox06	scf15193093	362 220	ssa23	100	0	1298	fps5539
Sox07	scf15173706	6 678	x	100	4,00E-71	276	x

Sox09/Sox17	scf15158750	2 108	ssa29	99,71	0	680	fps80
Sox10	scf15134628	97 789	ssa19	93,11	1,00E-117	428	fps253
Sox11/Sox13	scf15105042	8 294	ssa25	85,52	5E-44	184	fps39
Sox11/Sox13	scf15097978	114 827	ssa05	100	0	1616	fps762
Sox16|mouse	scf15173171	195 335	ssa28	82,64	1,00E-55	222	fps2166
Sox16|mouse	scf15152123	4 699	x	98,98	0	932	x
Sox18	scf15170786	168 559	ssa19	82,8	1E-32	147	fps1709
Sox18	scf15133941	9 473	ssa19	100	0	825	fps1831
Sox19	scf15152123	4 699	x	100	0	3406	x
Sparc	scf15108219	14 999	x	86,94	6,00E-60	238	x
Sparc	scf15152409	5 659	x	100	0	842	x
Sry	scf15171912	12 426	ssa12	86,88	3,00E-51	208	fps2932
Sry	scf15097978	114 827	ssa05	100	0	1616	fps762
SSB-4	scf15140303	155 180	ssa14	99,49	0	1035	fps764
Sxl	scf15116667	84 743	x	92,1	0	1356	x
Sxl	scf15181786	13 163	x	99,91	0	2034	x
Tbx1	scf15188114	415 547	ssa24	86,67	1,00E-133	482	fps1337
Tbx1	scf15099451	61 690	x	99,74	0	1455	fps2953
tp53	scf15118478	235 050	ssa04	100	3,00E-113	416	fps1538
tra2b	scf15157312	110 565	x	99,81	0	1053	fps3865
vasa	scf15204390	2 142	ssa27	89,73	1,00E-75	289	
vasa	scf15166306	202 943	ssa04	100	2,00E-46	192	
Vcp	scf15192588	63 106	ssa01	100	6,00E-64	250	x
Vnn1	scf15131889	172 144	ssa18	100	3,00E-54	218	fps2152
Wt1	scf15194398	674 408	ssa16	100	7,00E-18	99,6	fps88