

Sex-associated DNA markers in the Atlantic halibut

(Hippoglossus hippoglossus)

Munevver ORAL*

*Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland,
UK*

**Supervisors: ¹Andrew Davie, ¹Christos Palaiokostas, ¹John Taggart,
¹Michaël Bekaert, ¹Dave Penman, ¹Herve Migaud, ¹Brendan McAndrew**

¹Institute of Aquaculture, Pathfoot Building, University of Stirling,
Stirling FK9 4LA, Scotland, UK

*Author for correspondence

Munevver ORAL

Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, UK

Tel: +44 [0] 7507468460; Email: munevverorall@gmail.com

Abstract

Atlantic halibut (*Hippoglossus hippoglossus*) is one of the candidates for cold water marine aquaculture. Females grow faster and mature later than males which makes it desirable to have monosex female production in commercial scale. This study focuses on the development of sex-associated markers that can be used to assist development of monosex female production by decreasing the laborious process of progeny testing. Sexed individuals from former studies which had balanced sex ratios (50% and 47.5% females respectively) were used to generate Restriction Associated DNA RAD tags from next generation sequencing technology. Restriction Associated DNA (RAD) tags were used to generate five putative Y contigs and eleven allele specific SNP assays from 67,466 identified markers. Sex association was investigated by analysing seven SNP loci using allele-specific PCR. Two of these loci (*Hhi58665* and *Hhi41238*) predicted the phenotypic sex of unrelated offspring with 93% accuracy within the entire dataset ($p < 10^{-15}$) whereas putative Y markers failed. These results represent an example of the potential of next generation sequencing systems with high throughput to analyse sex in aquaculture species. Understanding sex determination mechanism and developing sex associated markers will give a boost to halibut industry, shortening the time for monosex female production to come to market and establish stable commercial production, decreasing the costs of farming.

Key words: Allele specific PCR, SNPs, sex determination, RAD tags, *Hippoglossus*

I. Introduction

Interest in diversification following the success of salmonid farming led to development of farming of new aquatic species (Forster, 1999). One of the proposed candidates for producers was Atlantic halibut, *Hippoglossus hippoglossus*, a flatfish species that lives in cold water in the North Atlantic. Countries which had salmon farms started investments in farming of Atlantic halibut, taking into account the price of halibut (2005-2010 market value for halibut is £6.2/kg while salmon is £3.1/kg, FAO Aquaculture Production Statistics). Moreover, Atlantic halibut exists within a niche market, mainly being bought by high-end buyers including big hotels and luxury restaurants which make species desirable for producers (Forster, 1999). The leader of the industry is Norway which started attempts on research in 1974 when adult males and females were caught from wild and stripped for external fertilisation. Since then several commercial production trials were established. Attempts in Norway pushed the development of big scale production in the country (from 14 hatcheries to 3 remaining) and spread around the world including UK in 1983 at Sea Fish Industry Authority's Marine Farming Unit, Ardtoe. Mainly UK's halibut production is supplied from Scotland currently from just one farm left (from 4 hatcheries in the past) which is the largest halibut hatchery, Otter Ferry Seafish, in UK. In Iceland, the first farming experiments began in 1985 and significant progress has been done in following years with investments of Fiskey's hatchery which collapsed recently in 2011. Table-1 shows the world production of the Atlantic halibut and the major players from the beginning. Other entrepreneur countries are Canada (3 hatcheries to 1 remaining), USA (one research hatchery) more recently Chile (Bromage *et al.*, 2000, Mangor-Jensen *et al.*, 1998, Alvial and Manriquez, 1999).

Even though the attempts have been made for more than three decades; still, there are a range of difficulties about commercial halibut production. The first difficulty associated with production is lack of knowledge about husbandry particularly during the larval development stage and weaning. The species has a long embryonic yolk sac stage, approximately 265 degree-days, which is very long time compared to other cultured marine species (Opstad *et al.*, 1998). High mortality (up to 50% with deformities) during developmental stage and fragility of prolonged yolk sac stage of halibut represent major bottlenecks in the establishment of reliable production system (Mangor-Jensen *et al.*, 1998). Apart from complex biology of fish, long grow out time represents another challenge in halibut intensive culture. In the wild, males mature at around 4.5 years old

with the average weighing of 1.7 kg and 55 cm length whereas females mature around 7 years old weighing 18 kg and more than 110 cm length (Jakupsstovu and Haug, 1988). In the culture conditions males mature at around 2 years old while females start maturation around 6 years old. Both these examples depict the fact that Atlantic halibut has sexual dimorphism in terms of growing performance. Faster development of females means that females reach market size (3-5 kg for supermarkets; >5 kg desired by high-end buyers) before maturation which allows fish to make use of all energy for just somatic growth (Hendry *et al.*, 2002, Tveldt *et al.*, 2006). On the other hand, males become mature at a younger age (around 2 years old) with smaller size and also require more time for market size compared to females, meaning males share energy for growth and maturation. The biological process of male sexual maturation causes big economic losses in whole production cycle. Thus growth performances of females (>25%) are much greater than males. In other words, the positive development of all female stocks in all countries would improve the economics of on-growing and resulted in 25% less juveniles being required to grow the same tonnage. Not only this but also reduction in variance at size of harvest could also be considered an advantage of mono-sex female culture for Atlantic halibut (Beardmore *et al.*, 2001). Therefore from the industrial point of view having all-female population by manipulating gender is desirable in order to increase profitability. Even though there are numerous artificial ways to manipulate sex ratio which will be discussed later, the foremost respect should have good understanding of sex determination system in fish with the focus of halibut.

Fish show very varied mechanisms of sex determination compared to other vertebrates including gonochorism or hermaphroditism, XX/XY or WZ/ZZ, multiple sex chromosomes, mixed Genotypic Sex Determination (GSD) and Temperature Sex Determination (TSD) or unisexual species. All these different kind of mechanisms require at least a basic understanding of sex determination/differentiation if manipulation of sex ratio is desired (Beardmore *et al.*, 2001; Piferrer *et al.*, 2008). Many aquatic species show sex-related differences, including salmonids, tilapias, sea basses and Atlantic halibut is one of those. The karyotype of Atlantic halibut has been described by Brown *et al.* (1997) and reported $2n=48$ telocentric chromosomes which is typical for flatfish species. The whole genome of Atlantic halibut has not been sequenced yet. A genetic linkage map has been constructed on the basis of 258 microsatellites and 346 AFLPs (Reid *et al.*, 2007). The fish has relatively small genome (0.73 pg of DNA/haploid cell) and there is evidence

suggesting that the Atlantic halibut has an XX/XY sex determination system (Reid *et al.*, 2007). Despite the fact that using a linkage mapping approach, Quantitative Trait Loci (QTLs) can be detected and sex determining genes can be placed into one or more linkage groups (Penman and Piferrer, 2008), this was not done in halibut due to the progeny in the mapping family being too small to sex (Reid *et al.*, 2007).

After understanding of sex determination system, sex ratio can be manipulated to the desired one. Currently, there are three main techniques for producing mono-sex culture by taking into account success of higher average growth rate; gynogenesis, temperature manipulations and endocrine manipulations (hormonal sex reversal). Gynogenesis refers to the process of only maternal DNA inheritance by inactivating paternal genetic information and inducing duplication of the maternal chromosome set (Devlin and Nagahama, 2002). Even though the technique is used to elucidate sex determining mechanism, it is not feasible for commercial scale mono-sex production due to low survival rates and poor hatch, as has been reported in Atlantic halibut (Piferrer *et al.*, 2004, Tveldt *et al.*, 2006). The second technique to achieve mono-sex culture is temperature manipulation. Some scientists argue that experimental and hatchery temperatures under which skewed sex ratios have been observed are often far away from natural conditions in nature, so the observed effects on sex ratio do not constitute “real” TSD (i.e. an adaptive response). Moreover, temperature acting on sex differentiation can induce sex reversal in species regardless of homomorphic or heteromorphic sex chromosomes. It has not been reported any significant effect of temperature manipulations in Atlantic halibut including recent study by Hughes *et al.*, 2007 halibut juveniles were reared at 7 °C, 12 °C and 15 °C, sex ratio were not significantly different from 1:1 in any group hence concluded that sex in Atlantic halibut is not influenced by temperature. Patterns of sex ratio responses to temperature in finfish species have been discussed in detail by Ospina-Alvarez *et al.*, 2008.

The methodology of choice for halibut production is endocrine manipulations which can be used directly (on fish for market) or indirectly (on broodstock, to produce neomales or neofemales for crossing) to manipulate the sex ratio of populations either to produce females (feminisation) or males (masculinisation) with the help of injection, in-feed or bath treatment. The first method for endocrine manipulations is direct hormone treatment which can be applied to any species regardless of sex determination system or chromosome structure; homogametic or heterogametic sex. Undifferentiated fish are treated by hormone (oestrogen/androgen or their synthetic analogies) during the labile period and sex ratio is

manipulated as desired. Despite the (apparent) simplicity of method, those fish are not suitable for breeding programmes due to obscurity of genotype and phenotype of fish and public concern about direct exposure of hormone for human consumption (Piferrer, 2001).

Indirect hormone treatment requires an understanding of sex determination system. The method can easily be applied for species in which females are homogametic sex (XX/XY; Female/Male) for feminisation and males are homogametic sex (WZ/ZZ; Female/Male) for masculinisation with the exception of YY viable males and Genetically Male Tilapia (GMT) have been developed in Nile tilapia. For example, indirect feminisation process starts with androgen treatment of undifferentiated mixed fish. After first treatment normal males (XY) and neomales (XX: genotypically female but phenotypically male) individuals are produced. After that by crossing neomales (XX) and normal females (XX), all-female commercial production is achieved. But this process brings together some problems too like doubling the time, effort, space and cost of all production which would be excess to 5-6 years to produce neomales and test progeny. Although all these disadvantageous still indirect hormone treatment is the methodology of choice for monosex halibut commercial production. Therefore, there is an emergent need to develop quicker techniques. The biggest advantageous of using indirect hormone treatment is untreated fish of the desired sex can be grown out and harvested with high success rates. Moreover there is not any marketing issue due to the fact that there is no direct exposure to steroids of fish for human consumption. In other words, the key fact about indirect hormone treatment is sex determination system is manipulated in broodstock to result in progeny which are all genetically the same sex. Indirect hormone treatment has already been applied for commercial scale production of rainbow trout, Atlantic halibut and many other finfish species (Beardmore et al., 2001; Nagahama, 2005; Piferrer, 2001).

The problem of progeny testing can potentially be reduced or eliminated by using genotyping. Sex-specific DNA probes allow the fast identification of genotypic sex and have been successfully developed (*OtY-1*, *GH-P*) in Chinook salmon (*Oncorhynchus tshawytscha*) (Devlin et al., 1994). The same markers have been applied to related species such as chum (*O.keta*) and pink salmon (*O.gorbuscha*), however these markers cannot identify sex in the same family members including Atlantic salmon and rainbow trout (Devlin et al., 1994). More recently in Pacific Halibut (*Hippoglossus stenolepis* is a species from the same family "Pleuronectidae" as Atlantic halibut) three out of sixteen microsatellite loci that were correlated to female gender. These three loci were screened in

550 individuals and results showed that they successfully identified sex in 92%. Same markers have not been tried in Atlantic halibut yet (Galindo *et al.*, 2011). Thus, genomic technologies capable to answer our questions in terms of defining associations with sex, requires investment in terms of both money and time. Once it has been achieved, the technology can give a boost to halibut industry through shortening of challenging production time needed for market size fish.

Advances in DNA technologies and methodologies can help to improve our understanding about fish sex determination/differentiation system. However, there are still gaps in practice and theoretical data due to the size of the genome and the absence of large sex-specific chromosomal regions in most aquaculture species (Penman and Piferrer, 2008). Even though using a linkage mapping approach may help us to detect many sex-linked markers in species with simple or complex genetic (or GSD + TSD) systems, but the ability to detect sex-specific markers will still depend on the size of the sex-specific region in the genome (in simple sex determination systems only). Even if this is limited just to a single gene, it can still be (very) difficult to detect truly sex-specific markers. The main difficulty is the unknown location of sex determining gene(s) in order to minimise the probability of recombination between marker and gene(s). So as to define the sex determining gene(s) huge numbers of markers (hundreds of thousands) are required which does not seem possible. Despite Devlin *et al.*, 1994 Chinook salmon success, considerable effort has been done in order to detect more sex linked markers in variety of fish species but only small numbers of cases of success have been reported. Techniques have been used to screen for sex-linked markers are Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Restriction Fragment Length Polymorphism (RFLP), microsatellites and Single Nucleotide Polymorphism (SNPs) respectively. There is an increasing need to be able to develop appropriate markers quickly and cheaply in order to decrease time consuming progeny testing.

There is a huge interest to use SNPs as informative markers due to their abundance in eukaryotic genomes (occurring approximately once every 1000 base pairs) (Kwok and Chen, 2003). There are different ways to detect SNPs including DNA sequencing, primer extension assay, designing of allele-specific oligos or gene-chip technology (Kwok & Chen, 2003). Due to the requirement for expensive DNA sequencing, isolating and typing large numbers of SNPs was until recently very expensive and time consuming. After the discovery of a new sequencing method, Restriction-site Associated DNA, RAD tag

sequencing, (using Illumina Genome Analyzer sequencer), very comprehensive datasets can be developed and utilised much more efficiently (Baird et al., 2008). Major benefits of the RAD sequencing technology are a significant increase in the number of markers and an appreciable decrease in cost compared to older sequencing technologies. Furthermore depending on the number of restriction enzyme sites, the number of markers can be increased (Baird et al., 2008). RAD has also high coverage for every SNP which makes advantageous to avoid genotypic errors. Next generation sequencing (2nd Generation Sequencing) generates up to 100 gb per run (100 million reads and each read 100bp). These new technologies are now available and produce a very large amount of data rapidly and relatively cheaply (Liu, 2011). The high throughput characteristic of RAD Sequencing allows the technique to be used to isolate SNPs, create linkage maps and perform QTL analysis (including disease resistance, spawning time, sex association) in a single run. Many SNPs can be identified and associations can be detected with the help of this technology, which is also much cheaper than previous sequencing methods (Liu, 2011). However the output of this technology needs to be verified and the associations formally analysed.

Allele-specific PCR assays for SNPs, as cost-effective as standard PCR, can be developed and carried out using standard instrument and machines found in most laboratories and designed from high-throughput RAD data. This cost-effective protocol has been adopted for a range of studies including population genetics, pharmacologic studies (Wangkumhang *et al.*, 2007). The system relies on the discrimination power of a novel form of competitive allele specific PCR. SNP chips can also be developed to assay numerous SNPs simultaneously but these are more expensive (Liu, 2011).

In addition to SNP markers associated with a trait of interest (i.e. gender) RAD technology could also generate whole sequence contigs specific to the trait of interest. These ultimately would be far easier to assay than the SNP markers as they could be amplified by routine PCR methodologies. However care should be taken in assuming that such putative Y specific contigs genuine and not the result of technical problems or mutation sites.

The overall aim of the study is to verify the existence of sex-associated (sex-specific and/or sex-linked) markers by using allele-specific SNP assays and PCR for putative Y contigs. The study focuses on markers identified in an existing RAD sequencing dataset and tries to develop assays for sex-associated markers which can be used to assist in development of

mono-sex female production in Atlantic halibut. Even though this fish has complex biology and it has been very challenging to establish a stable production cycle, with the help of such improvements it is hoped that the volume of production will be increased due to its high market price.

II. Materials and Methods

All laboratory work was performed between 23th of April and 19th of July 2012 in the Molecular Biology Laboratory of the Institute of Aquaculture at the University of Stirling, UK.

1. DNA Samples

DNA for the initial RAD sequencing project came from two mixed sex families produced during Mairi's PhD study from University of Stirling (2 family produced: Fam B has 2 parents and 28 offspring, Fam C has 2 parents and 62 individuals) (Cowan, 2010). The parents and two offspring from each cross (one of each gender) were used as a test panel for the initial optimisation and verification of each assays performance. Thereafter, for the allele specific PCR studies (see below) an unrelated historic DNA dataset was used to verify the robustness of the chosen markers. All DNA samples from the historic DNA dataset were collected in 1999 by Anu Frank-Laware who was a PhD student in University of Stirling, aimed at genetic management of Atlantic halibut (*Hippoglossus hippoglossus*) using 7 microsatellite loci from Otter Ferry Seafish Limited, which is located on the West Coast of Scotland on the East Coast of Loch Fyne (Frank-lawale, 2005). DNA was extracted from 350 individuals' blood samples including parents and offsprings from each family modifying the method from Taggart *et al.*, (1992) and stored at +4°C.

2. Specifying of the quality and the quantity of DNA

The amount of DNA for every sample was quantified using NanoDrop® ND-1000 (NanoDrop Technologies, Inc, Wilmington, Delaware USA) Spectrophotometry. The quantities obtained were highly variable between samples and ranged from 29 -1908 ng/µl. Due to the fact that high concentration of DNA can inhibit PCR, samples with high DNA

concentrations were diluted to ~50-100ng/µl using ddH₂O to be used in PCR amplifications.

3. Putative Y contigs

Five Putative Y-specific contigs generated by RAD data (i.e. Sequence information present in DNA from male parents in mapping families but not in female) were analysed. Primers for those loci were generated by NCBI primer design tool [<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>] (Table 2). For the first trial, PCRs for these five loci were run at an annealing temperature of 59 °C. In this trial, the RAD family test panel as explained above was run.

Each PCR reaction consisted of 2.25 mM MgCl₂, 150 mM of each dNTP mix, 0.8 µM of each forward and reverse primer, 10X PCR buffer (detergent and MgCl₂ free-K Biosciences Beverly, US), 0.02 units of *Taq* DNA Polymerase (K Biosciences Beverly, US) and ~50ng genomic DNA template. The total volume of every reaction was 20 µl. Reactions were conducted in a Biometra Gradient PCR machine under the following thermal cycle protocol: Initial denaturation step of 15 minutes at 95 °C followed by 40 cycles of; 30 seconds at 95 °C (denaturation), 30 seconds at Ta °C (annealing) and 60 seconds at 72 °C (extension), followed a final extension step at 72 °C for 7 minutes to promote 3' adenylation and samples were held at +20 °C. Table 2 shows nucleotide sequence, product size and T_m of Putative Y primers used in the study. PCR products were loaded into 1% agarose gel (1% agarose gels in 1X (Trisima Base-Acetic Acid-EDTA) TAE buffer were run at 90 Volts for 30 minutes) to visualise products and thus reveal sex association.

4. Allele Specific PCR (AS-PCR)

Eleven SNP loci (*Hhi52750*, *Hhi12577*, *Hhi18571*, *Hhi47769*, *Hhi42798*, *Hhi10170*, *Hhi16608*, *Hhi58665*, *Hhi5064*, *Hhi45159*, *Hhi41238*) generated by RAD data were analysed for sex-association. A web based allele-specific primer design application [<http://bioinfo.biotec.or.th/WASP/>] was used so as to detect SNPs by designing the three required primers (Wangkumhang et al., 2007). Sequence data from the RAD dataset was uploaded with the SNP annotated along with desired primer parameters including primer length (18-26 bp), GC% concentrations (40-65%), maximum Ta difference (<6 °C), maximum self complementarity (4bp), maximum 3' self complementarity (3bp), maximum

PolyX in primer (3 bp) and a product size between 100-300 bp. For each locus the primer set which was closest to these criteria was ordered (Table 3).

For the first trial of each locus, a commercially prepared master mix, MyTaq™ HS Mix (Bioline Reagents Ltd) which contains MgCl₂, PCR buffer and specific *Taq* enzyme, was used in order to reduce non-specific amplifications and any pipetting errors. PCR mastermix was chosen for the first step as while it is comparatively expensive, it uses proprietary chemistries which makes it highly specific and thus ideal for these high specificity assays. Each reaction was set up following commercial instructions of MyTaq™, consisting of half of total volume per reaction MyTaq™ HS Mix, 0.4 mM each forward and reverse primer and 200ng genomic DNA template. The total volume of every reaction was 10µl and each locus was optimised using the same test panel as described above. Reactions were conducted in a Biometra Gradient PCR machine under the following thermal cycle protocol which recommended by company: Initial denaturation step of 1 minute at 95 °C followed by 35 cycles of, 15 seconds at 94 °C (denaturation), 15 seconds at 55 °C (annealing) and then 15 seconds at 72 °C (extension). No final extension step was used for commercial MyTaq™ HS Mix PCR. Amplicons were loaded into 1% agarose gel in order to visualise products and thus reveal sex association.

Once each locus was verified using MyTaq™, the assay was converted and then optimised to Klear *Taq* (K Biosciences Beverly, US) due to the fact that Klear *Taq* is much cheaper per sample for further genotyping. Each reaction for Klear *Taq* consisted of 2.25mM MgCl₂, 150mM of each dNTP_{mix}, 0.8µM each forward and reverse primer, 10X PCR buffer (detergent and MgCl₂ free, K Biosciences Beverly US), 0.02 units of *Taq* DNA Polymerase (K Biosciences Beverly US), and ~50ng genomic DNA template. The total volume of every reaction was 10 µl. Reactions were conducted in a Biometra Gradient PCR machine under the following thermal cycle protocol: Initial denaturation step of 15 minutes at 95 °C followed by varied cycles between 30-40 of, 15 seconds at 95 °C (denaturation), 10 seconds at Ta °C (annealing), 20 seconds at 72 °C (extension) and a final extension step was used at 72 °C for 7 minutes to promote 3' adenylation and samples were held at +20 °C. PCR products were loaded into 1-1.5% agarose gel (depended of size of amplicon, products are less than 200 bp visualised in 1.5% concentration while bigger products [>200bp] loaded into 1% agarose gel in order to visualise products).

Two loci (*Hhi52750* and *Hhi12577*) were located within discontinuous contigs with the SNP itself being close to the unknown sequence region. The initial amplicons for both loci were larger than expected (*Hhi52570* >1kbp; *Hhi12577* ~600bp). In order to design assays with smaller products both contigs were re-sequenced using a Beckman 8800 autosequencer and the primer sets then redesigned as above with a known product size.

Hhi5064 and *Hhi41238* required additional optimisation due to extra bands around the expected product. Therefore TouchDown-PCR was used to increase specific binding and avoid non-specific bands (Table 3).

Once each allele-specific locus were optimised with the same panel (RAD parents and 2 families) unrelated halibut families' parents were run for each marker. Five families (Sire: 17A and 37A; Dam: 10A-11A-08A, 64A and 3A) which had enough individuals (n=18, 20, 20, 18, 9 respectively) were identified. Table 4 shows the pedigree structure of unrelated halibut families, number of offsprings and the code of chosen individuals. From each family 10 males and 10 females were chosen to test robustness of allele specific loci but due to limited number of individuals in some families maximum possible number of individuals were chosen. Totally 85 offsprings from unrelated 5 halibut families were tested with parents. In the gel electrophoresis imaging of the assays results a double loading system was used to screen individuals: once AS-PCR was carried out the common allele of each locus was loaded onto the gel and an electric current passed over the gel, then after 10 minutes the electricity was stopped and the male-associated allele was loaded (same lane per individual). The electricity was put back on and the gel left for a further approx. 20 mins before imaging.

Statistical analysis

Statistical analyses were conducted using *R* v.2.15.1 statistic programme. Association for every SNP was checked for each family separately and across families using a Bernoulli GLM (Generalized Linear Model) so as to assess whether there were significant differences in genotypic data between males and females. After p-values were calculated for the entire data set, Tmap v.1.1 was used to calculate genetic distance between the best performing markers.

III. Results

1. Putative Y Contigs

Five Putative Y contigs (*HalY59970*, *HalY62243*, *HalY62390*, *HalY65579*, *HalY63667*) generated by RAD data were amplified in both genders regardless of sex. From out of five putative Y contigs only one of those, *HalY-62390*, required extra optimisation step apart from temperature gradient and optimised at 59 °C with 2mM, 2.25mM and 2.50 mM MgCl₂ concentrations, Figure 1 depicts the result of optimisation in *HalY-62390*. Figure 2 shows a typical example of putative Y results from contig-*HalY59970*.

2. Allele Specific PCR

Out of eleven SNP loci for which AS-PCRs were developed (*Hhi52750*, *Hhi12577*, *Hhi18571*, *Hhi47769*, *Hhi42798*, *Hhi10170*, *Hhi16608*, *Hhi58665*, *Hhi5064*, *Hhi45159*, *Hhi41238*) generated by RAD data only 7 of them (*Hhi18571*, *Hhi47769*, *Hhi16608*, *Hhi58665*, *Hhi5064*, *Hhi45159*, *Hhi41238*) showed sex association in that RAD family test panel (Table 3). Figure 3 shows a typical example (*Hhi58665*) of an optimised AS-PCR with clear sex association in the RAD test panel and Figure 4 shows the same loci tested in unrelated families' offsprings used in the second part of the study.

Out of eleven loci just seven of them showed sex association across RAD family test panel. The other four loci (*Hhi52750*, *Hhi12577*, *Hhi42798*, *Hhi10170*) showed different patterns. Two loci (*Hhi52750* and *Hhi12577*) were discontinuous constructs therefore further re-sequencing was used to allow the redesign of the assays and amplify the SNP located area specifically. Other two loci (*Hhi42798*, *Hhi10170*) were amplified in both genders regardless of sex. Finally these four loci were not used to screen genotype of unrelated offspring.

Eighty-five offsprings from five different families were genotyped for the seven SNPs above. Out of seven SNPs three of those (*Hhi18571*, *Hhi5064*, *Hhi47769*) showed significant segregation distortion ($p < 0.01$) and were therefore removed from any further analysis. Table 5 shows the genotypic proportion of allele specific markers through unrelated halibut families and association rate of each locus through five families.

The p-values from the Bernoulli GLM for every family separately showed significant association for the remaining 4 SNPs (*Hhi58665*, *Hhi45159*, *Hhi16608* and *Hhi41238*)

through Families 2-5. In Family 1 only *Hhi41238* showed significant association ($p < 0.05$). In family 2, apart from *Hhi16608* locus (p -value cannot be calculated due to unexpected pattern of genotyping results-See Table 5 for distribution of genotypes) all other loci (*Hhi58665* $p = 1.90 \times 10^{-6}$, *Hhi45159* $p = 4.02 \times 10^{-3}$ and *Hhi41238* $p = 1.90 \times 10^{-6}$) showed significant associations. The same pattern in family 2 was observed in family 3 too, p -values calculated *Hhi58665* $p = 2.94 \times 10^{-7}$, *Hhi45159* $p = 4.06 \times 10^{-3}$, *Hhi16608* $p = 4.56 \times 10^{-1}$ *Hhi41238* $p = 9.62 \times 10^{-6}$ respectively in family 3. In family 4 only *Hhi45159* locus did not show high association ($P > 0.05$) Apart from this loci all others showed high associations (*Hhi58665* $p = 6.96 \times 10^{-4}$, *Hhi47769* $p = 1.52 \times 10^{-5}$, *Hhi16608* $p = 5.48 \times 10^{-3}$, *Hhi41238* $p = 4.52 \times 10^{-6}$). In family 5 apart from *Hhi45159* locus ($p > 0.05$) other loci showed the same p value = 4.37×10^{-4} *Hhi58665*, *Hhi16608* and *Hhi41238*.

The results from the GLM model across all the families showed significant association for all 4 SNPs. The highest association was observed in *Hhi58665* and *Hhi41238* with sex in the entire dataset ($p < 10^{-15}$). These two best performing allele specific loci (*Hhi58665* and *Hhi41238*) predict the sex of 93% of male and female individuals through four families. The genetic distance between these two markers above was 5 cM (Figure 5).

VI. Discussion and Conclusion

This study is the first demonstration of sex-associated SNP markers for Atlantic halibut. Using next generation sequencing technology and its high throughput a 5cM area in Atlantic halibut genome which has strong association (93% $p < 0.05$) with sex has been localised. These results will help further studies to identify the sex determining regions/genes in Atlantic halibut and commercially these results can help to shorten the time taken to get monosex female production to the market.

No putative Y contigs showed sex association and primer pairs amplified specific products regardless of sex. This clearly suggests that these sequences are not sex specific and therefore it remains to be determined why these sequence outputs were associated with gender following the RAD sequencing analysis. One possible reason could be that for these contigs, the primers used in the initial RAD library construction could be located on sex associated mutation site.

Allele specific PCR trials contained a large range of optimisation steps for each locus themselves as it is necessary to ensure the specificity of the assays. Except for the

polymorphic base of the SNP, the only difference in the AS-PCR primers is an altered penultimate base which can be used to enhance reaction specificity deliberately (Wangkumhang *et al.*, 2007). It is preferred to introduce a weak base (like A/T instead using a strong base like G/C). Since SNPs are located on the 3' end of allele specific primers they bind to template DNA when perfect complementation occurs (Hirotzu *et al.*, 2010). One of the two primers per locus calls "common type" which would expect to see in all individuals while the other one ("gender specific" for this study) anneals only the existence of SNP. Thus, if a single base polymorphism occurs, the genotyping results can be observed by simply comparing the length of PCR products. Although optimisation process of allele specific PCR reaction seems to be achievable by optimising standard PCR conditions, due to the factors has been discussed above optimisation process per each locus required more time than standard PCRs'. For example, in all cases SNPs were located on the allele specific forward primers (apart from one exception, *Hhi45159* locus, SNPs was located on reverse primers) and majority of cases each allele assay was optimised in different conditions despite the fact that each forward primer has only one base pair difference (Table-5).

Initial optimisation process of majority of allele specific loci non specific bigger bands were seen than expected product size in females. Temperature gradient was used ($\pm 5^{\circ}\text{C}$) for the initial optimisation steps. In order to make reactions more specific annealing time was decrease from 15 seconds to 10 seconds and extension time was decreased from 30 seconds to 20 seconds too. Secondly PCR cycles were adjusted bearing into mind that every cycle the amount of DNA twice what it was before (30-40 cycles). If nonspecific bands existence remained then MgCl_2 concentration was decreased stepwise taking into consideration of MgCl_2 is the cofactor of the enzyme, *Taq* polymerase, and the high concentrations of this compound might cause non-specific binding (Markoulatos *et al.*, 2002). Nonspecific binding problem with two loci (*Hhi5064* and *Hhi41238*) emerged after initial optimisation of both loci with test panel particularly in unrelated historic DNA samples from offsprings. In order to define problem clearly 2 individuals from test panel and unrelated historic DNA samples from offspring were run together in the same conditions. Results showed that some individuals from unrelated offsprings had proximate two bands, smaller and bigger, than the anticipated product. In order to increase specificity Touchdown PCR was used. The method allows primer to avoid nonspecific binding by

decreasing annealing temperature gradually after very specific base pairing between the primer and the template DNA (Don *et al.*, 1991).

Following optimisation all 7 loci which showed sex association in the RAD panel were then used to screen an unrelated DNA dataset. Following preliminary analysis of these results 3 SNPs (Hhi18571, Hhi5064, Hhi47769) were removed from further analysis as they represented significant segregation distortion ($p < 0.01$). This phenomenon refers to the observed genotypes significantly deviating from expected Mendelian frequency. Even though many explanations can be offered due to different populations having different Mendelian ratio, still the most sensible explanation is due viability selection on linked marker loci (Zhan and Xu, 2011).

For the remaining four loci, genotyping results of eighty-five individuals showed varied results in different families. Results of family 2 to 5 represented high sex-association whereas family 1 did not show the similar pattern. This can be due to small number of individuals in some families or can be explained different phase between marker allele and gene(s) in different families (Mackay, 2001). It should be noted that the fifth family's female genotype was unknown prior to this analysis. However, by interference by comparing sire and offspring genotype and even though their existed a limited number of individuals ($\Sigma 9$, 5F and 4M) female called 64A appeared to be homozygous for all SNP loci screened (Table 5).

This work has focused on the design and validation of individual allele specific PCR assays for each locus, which while effective is very time consuming and not capable of large throughput efficiently. Advancements in DNA technology particularly in next generation sequence systems have made available high throughput and cost effective systems to assay SNPs such as the Applied Biosystems Open Array System and the Fluidigm Dynamic Array Systems with smaller number of validated SNPs in rapid genotyping compared to TaqMan based platforms (Liu, 2011). More recently Mass Array Systems, SNPstreams and Illumina's Golden Gate systems also offer highest throughput per plate (Liu, 2011). One promising assay type for SNP genotyping is KASP (K Bioscience, Hertz, UK) which has high assay conversion rate compared to TaqMan probes (90% compared to ~60% TaqMan). KASP offers cost effective run with high number of individuals and high reproducibility (Tobin *et al.*, 2008). One of the most important features of this assay type is offers more accurate output compared to Allele Specific-PCR

and much more effective due to reduced laboratory effort. Therefore it is recommended that any further SNP studies should use one of those assay types discussed above to save time and provide greater accuracy even though allele specific PCR can be run with basic instruments found in most laboratories.

Even though single QTL analyses from the RAD dataset shows accumulation and a single peak within linkage group LG13. It is clear from the current study that some loci (*Hhi58665* and *Hhi41238*) showed better performance than others. This must be due to their locations and the linkage of these markers being closer to the sex determining region. The best performing allele specific loci (*Hhi58665* and *Hhi41238*) predict phenotypic sex with 93% accuracy across families 2 to 5. The remaining 7% of individuals whose sex was opposite and cannot be predicted, can be attributed to errors in phenotyping or recombination (Lee, *et al.*, 2003). Even though it can be due to additional genetic factors such as parental conditions or environmental factors like temperature alternatively but most likely reason is due to recombination. In order to define sex more clearly by markers, an increased number of genotyped individuals would be necessary. Alternatively, another improvement would be an increased number of markers to screen, which could be generated with help of new technology including RAD. If a RAD library was constructed using an enzyme with a smaller recognition site like *EcoRI* (GAATTC) as opposed to the standard such as *SbfI* (CCTGCAGG) this would provide a greater density of markers (Baird *et al.*, 2008). Since *EcoRI* has 6 base pair recognition sites the possibility to get more marker increase. As further step of increased markers located on genomic map in close distance major sex determining gene can be identified and can help by narrowing down the area of interest. Notwithstanding, the current results did identify a region with strong association to sex in the Halibut genome either between or close to loci *Hhi58665* and *Hhi41238*. The gap between these loci was calculated 5cM (Figure 5). This depicts the way of measuring genetic distance in genome. 1cM is equivalent to 1% recombination between markers. It should be stressed that the equivalent of centi Morgan to base pairs varies according to species and also across the genome. A centi Morgan can correspond to a span of between ten thousand to a million nucleotide base pairs which typically translated into ten or hundred genes (Lynch and Walsh, 1998).

Even though this study proves the accuracy of RAD technology high throughput more study is required so as to verify and analyse association formally. This study is significant due to the first demonstration of sex-associated SNP markers in Atlantic halibut. Seven

markers identified in this study are enough as a starting point so as to predict sex. But still there is a need to define more markers in order to increase accuracy. Once a range of sex associated markers are robustly verified the technology can give a boost to halibut industry through shortening the time to verification of neomale broodstock following sex reversal treatment. Genotyping will reduce the need to sacrifice fish for sex determination and will remove the need to hold stocks of fish for histologic sex determination ultimately the greatest saving will be in time where neomales can be identified within 1 year of treatment by genotyping as opposed to *circa* 5 years post treatment by progeny testing. It is hoped that this will result in a greater uptake of neomale production in halibut hatcheries worldwide which will result in an increase in all female production and hereby stable commercial production of Atlantic halibut can be established with increased volumes.

Acknowledgements

I would like to thank first and foremost the Turkish Government, Ministry of Education for giving me the opportunity to have a Master degree abroad and supporting me financially. I am grateful to all my supervisors from the Institute of Aquaculture, University of Stirling including Dr. David Penman who has helped with his discussions and broaden my horizons, Dr. Andrew Davie who has helped with his laboratory skills and getting me stress every time in order to utilise my potential perfectly. Many thanks to PhD. Christos Palaikostas with his expertise in statistical analysis and motivation continuously. I would also like to thank Dr. Michaël Bekaert for his loadstar chats on the way of RAD technology and Dr. John Taggart for sharing his huge experiences and stuff in laboratory. Finally I am grateful to all staff of Otter Ferry Seafish LTD including Mr Alastair Barge who is Managing Director of the company; Mr Dave Patterson is Technical Director of the company and Antonios Em. Chalaris is Assistant Project Manager for sharing their experiences and hospitality.

Along the year I have had great time with my classmates from all around the world I thank them very much for their valuable company.

Special thanks must go to my family for their priceless support, firm belief about me and encouragements during all my life.

REFERENCES

- Allen, A. M., Barker, G. L. a, Berry, S. T., Coghill, J. a, Gwilliam, R., Kirby, S., Robinson, P., et al. (2011). Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant biotechnology journal*, 9(9), 1086–99.
- Alvial, A., & Manriquez, J. (1999). Diversification of flatfish culture in Chile, *Aquaculture*, 176 65–73.
- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188(4), 799-808.
- Baird, N. a, Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. a, Selker, E. U., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3(10).
- Baroiller, J. F., D’Cotta, H., Bezault, E., Wessels, S., & Hoerstgen-Schwark, G. (2009). Tilapia sex determination: Where temperature and genetics meet. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, 153(1), 30-8.
- Baroiller, J.-F., Guiguen, Y., & Fostier, a. (1999). Endocrine and environmental aspects of sex differentiation in fish. *Cellular and Molecular Life Sciences*, 55(7), 910.
- Beardmore, J. ., Mair, G., & Lewis, R. . (2001). Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture*, 197(1-4), 283-301.
- Bjirrnsson, B. (1995). The growth pattern and sexual maturation of Atlantic halibut (*Hippoglossus hippoglossus* L .) reared in large tanks for 3 years, *Aquaculture*, 138(281-290).
- Bromage N, Mazorra C, Bruce M, Brown N, and Shields R. 2000. Halibut culture. In: Encyclopedia of Aquaculture (RR Stickney, ed.) pp.425-432, John Wiley and Sons, N.Y
- Cowan, M., Migaud, H., Davie, A. (2010) Research and development of stock management strategies to optimise growth potential in on growing of marine fish. SARF 027 Report pp 100.
- Davey, J. (2010). RAD Sequencing : A method for Population Genomics RAD Sequencing : A method for Population Genomics *Briefings in functional genomics*, 9(5-6), 416-23.
- Davey, J. W., Davey, J. L., Blaxter, M. L., & Blaxter, M. W. (2010). RADSeq: next-generation population genetics. *Briefings in functional genomics*, 9(5-6), 416-23.
- Devlin, R. H., McNeil, B. K., Solar, I. I., & Donaldson, E. M. (1994). A rapid PCR-based test for Y-chromosomal DNA allows simple production of all-female strains of Chinook salmon. *Aquaculture*, 128(3-4), 211-220.

- Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences, *Aquaculture*, 208(191–364).
- Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., & Mattick, J. S. (1991). “Touchdown” PCR to circumvent spurious priming during gene amplification. *Nucleic acids research*, 19(14), 4008.
- FAO (2010) Aquaculture production 2010, Fisheries statistics, Retrieved July 2, 2011, from <http://www.fao.org/fishery/statistics/global-aquaculture-production/query/en>
- Forster, J., (1999). Halibut farming its development and likely impact on the market for wild Alaska halibut, A report for Alaska department of Commerce and Economic Development Division of Trade and Development, (907) 465-2017.
- Frank-Lawale, A. S. (2005). *Genetic Management of the Atlantic halibut (Hippoglossus hippoglossus)*, PhD thesis, University of Stirling Institute of Aquaculture, Scotland, UK.
- Galindo, H. M., Loher, T., & Hauser, L. (2011). Genetic sex identification and the potential evolution of sex determination in Pacific halibut (*Hippoglossus stenolepis*). *Marine biotechnology (New York, N.Y.)*, 13(5), 1027-37.
- Hendry, C. I., Martin-Robichaud, D. J., & Benfey, T. J. (2003). Hormonal sex reversal of Atlantic halibut (*Hippoglossus hippoglossus*). *Aquaculture*, 219(1-4), 769-781.
- Hirotsu, N., Murakami, N., Kashiwagi, T., Ujiie, K., & Ishimaru, K. (2010). Protocol: a simple gel-free method for SNP genotyping using allele-specific primers in rice and other plant species. *Plant methods*, 6, 12.
- Hughes, V., Benfey, T. J., & Martin-Robichaud, D. J. (2007). Effect of rearing temperature on sex ratio in juvenile Atlantic halibut, *Hippoglossus hippoglossus*. *Environmental Biology of Fishes*, 81(4), 415–419.
- Kwok, P.-Yan, & Chen, X. (2003). Detection of Single Nucleotide Polymorphisms, *Current Issues in Molecular Biology*, 5(43-60).
- Lui, Z. (2011), *Next generation sequencing & whole genome selection in Aquaculture*, Iowa, USA: Wiley-Blackwell Publishing, 2011. Print.
- Lynch, M., Walsh, B. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associated, Inc.pp 413-424.
- Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35:303–39.
- Mangor-Jensen, A., Harboe, T., Shields R J., Gara B., Naas K E., (1998) Atlantic halibut, *Hippoglossus hippoglossus* L., larvae cultivation literature, including a bibliography. *Aquaculture Research*, (29), 857-886.

- Markoulatos, P., Siafakas, N., & Moncany, M. (2002). Multiplex Polymerase Chain Reaction : A Practical Approach, *51*(October 2001), 47–51.
- Nakamura, M., Kobayashi, T., & Chang, X.-tian. (1998). Gonadal Sex Differentiation in Teleost Fish. *The Journal of Experimental Zoology* 281:362–372.
- Obstad, I., Bergh O., Skiftesfik, A B. (1998). Large-scale rearing of Atlantic halibut, *Hippoglossus hippoglossus* L., yolk sac larvae: effects of flow rate on growth, survival and accumulation of bacteria. *Aquaculture Research*, (29), 893-898
- Ospina-Alvarez, N., & Piferrer, F. (2008). Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS one*, 3(7).
- Pandian, T. J., & Sheela, S. G. (1995). Hormonal induction of sex reversal in fish. *Aquaculture*, 138(1-4), 1-22.
- Penman, D. J., & Piferrer, F. (2008). Fish Gonadogenesis. Part I: Genetic and Environmental Mechanisms of Sex Determination. *Reviews in Fisheries Science*, 16(sup1), 16-34.
- Piferrer, F. (2001). Endocrine sex control strategies for the feminization of teleost fish. *Aquaculture*, 197(1-4), 229-281.
- Piferrer, F., & Guiguen, Y. (2008). Fish Gonadogenesis. Part II: Molecular Biology and Genomics of Sex Differentiation. *Reviews in Fisheries Science*, 16(sup1), 35-55.
- Reid, D. P., Smith, C.-A., Rommens, M., Blanchard, B., Martin-Robichaud, D., & Reith, M. (2007). A Genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics*, 177(2), 1193-205.
- Rowe, H. C., Renaut, S., & Guggisberg, a. (2011). RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, 20(17), 3499-502.
- Strüssmann, C. A., & Nakamura, M. (2002). Morphology, endocrinology, and environmental modulation of gonadal sex differentiation in teleost fishes, *Fish Physiology and Biochemistry* 26: 13–29.
- Stuart, E. J., Power, J. E., Benfey, T. J., Wolf, G., & Blanchard, B. (2010). Guidelines for Cage Culture of Atlantic Halibut in Canadian Maritime Waters Canadian Technical Report of Fisheries and Aquatic Sciences 2860.
- Taggart, J.B., Hynes, R.A., Prodoehl, P.A and Ferguson, A. (1992) A simple protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology* 40, 963-965.
- Tobin, M. D., Timpson, N. J., Wain, L. V., Ring, S., Jones, L. R., Emmett, P. M., Palmer, T. M., et al. (2008). Common variation in the WNK1 gene and blood pressure in childhood: the Avon Longitudinal Study of Parents and Children. *Hypertension*, 52(5), 974–9.

- Tvedt, H. B., Benfey, T. J., Martin-Robichaud, D. J., McGowan, C., & Reith, M. (2006). Gynogenesis and sex determination in Atlantic Halibut (*Hippoglossus hippoglossus*). *Aquaculture*, 252(2-4), 573-583.
- Valenzuela, N., Adams, D. C., Janzen, F. J., & Url, S. (2003). Pattern Does Not Equal Process: Exactly When Is Sex Environmentally Determined? *The American Naturalist* 161(4), 676-683.
- Wang, L. H., & Tsai, C. L. (2000). Effects of temperature on the deformity and sex differentiation of tilapia, *Oreochromis mossambicus*. *The Journal of experimental zoology*, 286(5), 534-7.
- Wangkumhang, P., Chaichoompu, K., Ngamphiw, C., Ruangrit, U., Chanprasert, J., Assawamakin, A., & Tongsim, S. (2007). WASP: a Web-based Allele-Specific PCR assay designing tool for detecting SNPs and mutations. *BMC genomics*, 8, 275.
- Zhan, H., & Xu, S. (2011). Generalized linear mixed model for segregation distortion analysis. *BMC Genetics*, 12(1), 97.

Table-1: The World production of the Atlantic halibut, FAO (tonnes)

Countries	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Iceland	2	8	13	34	93	120	95	0	0	110	31	39	49	72
Norway	0	0	0	0	0	424	427	631	1 197	1 185	2 308	1 587	1 568	1 610
UK	0	0	4	1	80	187	187 <i>F</i>	187	272	233	147	206	189	139
Grand Total	2	8	17	35	173	731	709	818	1 469	1 528	2 486	1 832	1 806	1 821

Table 2: Nucleotide sequence, product size and Ta of Putative Y primers used in the study.

Locus	Primer Sequences (5'→3')	Product Size (bp)	Ta (°C)
<i>HalY-59970</i>	F: GGCTAACGCTCCGCTCAA R: ACATTCTTGTTGGGATGGCCTGG	~ 100	59
<i>HalY-62243</i>	F: TTCTTCTTGCCGCTGCCCT R: TGCACGTGGCCTTTCTGCCT	~ 100-200	61
<i>HalY-62390</i>	F: TGAGCACCAGGCTCCTTGTTGG R: CGTGCGTGTGTGCACGTAGGTT	~ 100	59
<i>HalY-63667</i>	F: AGATGATGCCGCCATTTTGACATGA R: TGCACATGGTCTCTCTCACACACA	~ 100	61
<i>HalY-65579</i>	F: TGCAGCTTGGCTCCCTCTGTT R: TACCGTTTGCACAGCCCCAC	~ 200	61

Table 3: Nucleotide sequence, product size, Ta and assay type [either MyTaq™ or Clear Taq (Qiagen)] and optimised PCR conditions of all Allele Specific primers used in the study.

Locus	Primer Sequences (5'→3')	Product Size (bp)	PCR Conditions
<i>Hhi52750</i> ^a	F..[C]: ATCTCACAGTGGCCCTGGATGTC F..[T]: ATCTCACAGTGGCCCTGGATGTT R: GTCTTTTCTTGCAGGTTGTTTACG	~ 100	Ta: 55°C 35 cycles
<i>Hhi12577</i> ^a	F..[C]: AGGACACCATCACAGCTCTGGTC F..[T]: AGGACACCATCACAGCTCTGGTT R: CCATTGCAAATTTCCCCAAAACGCA *R_100bp:ACAGCAGCAGGAAGGAACGACC *R_323bp:TGCAATGCTACAGCAGCAAAGTGG	~ 100	Ta: 55°C 35 cycles
¹ <i>Hhi18571</i> ^b	F..[C]: AGCAGCTGGTGGATAGAGGC F..[A]: AGCAGCTGGTGGATAGAGGA R: ACAGAGTGAGGCTCAGGACG	~ 100	Ta: 59°C 30 cycles
¹ <i>Hhi47769</i> ^b	F..[G]: TCATCTGTCACCTTGTGCATTG F..[A]: TCATCTGTCACCTTGTGCATTA R: CGGAGGAGGAATAAAGAGC	100-200	Ta: 57°C 30 cycles
<i>Hhi42798</i> ^b	F..[G]: ATCATCAATATCATCTTCAGCAAAG F..[C]: ATCATCAATATCATCTTCAGCAAAC R: CTCATCATCCAAATCAGCGA	~ 100- 200	Ta: 55°C 30 cycles
<i>Hhi10170</i> ^b	F..[G]: CCTTAACTCCTCAAACCTAACGG F..[A]: CCTTAACTCCTCAAACCTAACGA R: TCAACAGGATTCACACAACACA	~100- 200	Ta: 56°C 35 cycles
¹ <i>Hhi16608</i> ^b	F..[C]: TCGACAAATCTGAAGCTGGAGTC F..[A]: TCGACAAATCTGAAGCTGGAGTA R: CTCCTCTGCGTCTGTTCAGG	~ 100- 200	F..[A] Ta:58°C, 30cyc F..[C] Ta: 57°C, 35cyc
¹ <i>Hhi58665</i> ^b	F..[C]: TACTGCTCTTGACTGCCTCGC F..[G]: TACTGCTCTTGACTGCCTCGG R: GCCACTTTCTGATGGTTTCGTC	~ 300	Ta: 59°C 30 cycles
¹ <i>Hhi5064</i> ^b	F..[G]: GACTACAACAATGCAAATGGAGAG F..[A]: GACTACAACAATGCAAATGGAGAA R: AGCGTCTGAGGGAGTCAGAG	~ 400- 500	F..[G] Ta:57°C, 30cyc F..[A] TouchDownPCR Ta: 59°C 5cyc Ta: 58°C 5cyc Ta: 57°C 20cyc
¹ <i>Hhi45159</i> ^b	R..[C]: ACTTAATGTTCCCTCCAGATAGATG R..[T]: ACTTAATGTTCCCTCCAGATAGATA F: GGTTGAACCTTGTATGTACATTGTC	~ 100- 200	Ta: 57°C 35 cycles
¹ <i>Hhi41238</i> ^b	F..[G]: CAGGCTTCAGGCCTTTTCG F..[T]: CAGGCTTCAGGCCTTTTCT R: AAGCTTCCCTGTGCTCAAGT	~ 500	F..[G] Ta:55°C, 35 cyc F..[T] TouchDownPCR Ta: 56°C 5cyc Ta: 55°C 5cyc Ta: 54°C 30cyc

*: After sequencing new Reverse primers were designed. ^a: Assay type optimised using MyTaq™

¹: loci were used to scan offsprings.

^b: Assay type optimised using Klear Taq

Table 4: Pedigree structure of unrelated halibut families, number of offsprings and the code of chosen individuals.

♀ \ ♂	17A	37A	Chosen offsprings	Family No	Codes of chosen individuals
08A	57(26F:31M)		18(10F:8M)	Fam-1	F: NO121-NO146-NO149-NO152-NO158-18-26-28-40-61 M: NO21-NO62-NO107-NO144 NO147-NO148-NO156-NO163 F: NO120-NO130-NO132-NO137-NO139-NO174-NO175-NO177-NO178-NO185 M: NO103-NO114-NO116-NO122-NO126-NO128-NO138-NO140-NO141-NO155 F: NO191-NO209-NO210-NO219-NO277-NO280-NO290-NO301-NO304-NO305 M: NO201-NO202-NO205-NO208-NO229-NO261-NO274-NO281-NO60-NO78 F: NO153-NO252-NO289-NO299-NO317-NO326-NO58-NM18-NM114 M: NO14-NO212-NO217-NO240-NO296-NO322-NM126-NM40-NM66 F: NO166-NO169-NO171-NO189-NO29 M: NO10-NO341-NO157-NM36
10A	210(114F:96M)		20(10F:10M)	Fam-2	
11A	73(41F:32M)		20(10F:10M)	Fam-3	
03A		18(9F:9M)	18(9F:9M)	Fam-4	
*64A	10(6F:4M)		9(5F:4M)	Fam-5	

*: Missing parent

Table 5: Genotypic proportion of Allele specific markers through unrelated halibut families and association rate of each locus through five families.

Markers	Family 1				Family 2				Family 3				Family 4				Family 5				^e Association Rate			
	^a G	^b F	^c M	p-value	^a G	^b F	^c M	p-value	^a G	^b F	^c M	p-value	^a G	^b F	^c M	p-value	^a G ^d	^b F	^c M	p-value				
Hhi18571	AA	6	3	6.49 10 ⁻²	AA	9	-	7.62 10 ⁻⁶ ***	AA	4	-	6.02 10 ⁻⁴ ***	AA	7	-	1.88 10 ⁻³ **	AA	1	-	2.52 10 ⁻²	76			
	AB	1	4		AB	-	8		AB	6	3		AB	2	4		AB	4	1			BB	-	3
					BB	-	6		BB	-	2		BB	-	2		BB	-	3					
Hhi58665	AA	5	1	7.11 10 ^{-1*}	AA	10	-	1.90 10 ⁻⁶ ***	AA	10	-	2.94 10 ⁻⁷ ***	AA	9	2	6.96 10 ⁻⁴ ***	AA	5	-	4.37 10 ⁻⁴ ***	93			
	AB	2	4		AB	-	9		AB	-	9		AB	-	5		AB	-	4					
		2	2																					
Hhi47769	AA	4	1	1.80 10 ⁻¹	AA	8	-	2.12 10 ⁻⁴ ***	AA	7	1	2.12 10 ⁻⁴ ***	AA	8	-	1.52 10 ⁻⁵ ***	AA	-	-	2.28 10 ⁻²	79			
	AB	5	7		AB	2	1		AB	3	9		AB	-	6		AB	5	1					
		1	-																					
Hhi45159	AA	4	1	3.59 10 ⁻¹	AA	3	-	4.02 10 ⁻³ **	AA	6	-	4.06 10 ⁻³ **	AA	2	-	9.24 10 ⁻²	AA	-	-	1.82 10 ⁻¹	61			
	AB	5	5		AB	7	4		AB	4	8		AB	6	9		AB	5	3					
		1	2																					
Hhi5064	AA	2	2	1	AA	-	-	0.5	AA	4	-	5.63 10 ⁻²	AA	5	1	1.79 10 ⁻³ **	AA	1	-	2.96 10 ⁻²	65			
	AB	6	6		AB	9	9		AB	5	6		AB	-	5		AB	4	4					
		-	-																					
Hhi16608	AA	3	-	1.0 10 ⁻¹	AA	-	-	-	AA	-	-	4.56 10 ⁻¹	AA	2	-	5.48 10 ⁻³ **	AA	5	-	4.37 10 ⁻⁴ ***	56			
	AB	4	3		AB	10	9		AB	6	5		AB	5	2		AB	-	4					
		3	5																					
Hhi41238	AA	7	1	1.84 10 ⁻²	AA	10	-	1.90 10 ⁻⁶ ***	AA	10	1	9.62 10 ⁻⁶ ***	AA	6	-	4.52 10 ⁻⁶ ***	AA	5	-	4.37 10 ⁻⁴ ***	93			
	AB	3	7		AB	-	9		AB	-	8		AB	-	6		AB	-	4					

^a: Genotype of individuals

^b: Female

^c: Male

^d:Missing parent (Genotype of Female parent is not known.)

^e: Percentage of association between genotype and marker (%)

*p < 0.05.

**p < 0.01.

***p < 0.001.

Figures

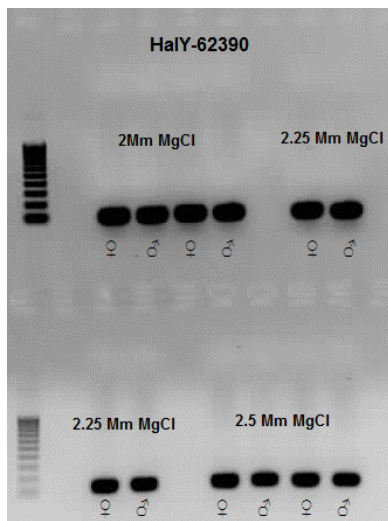


Figure 1

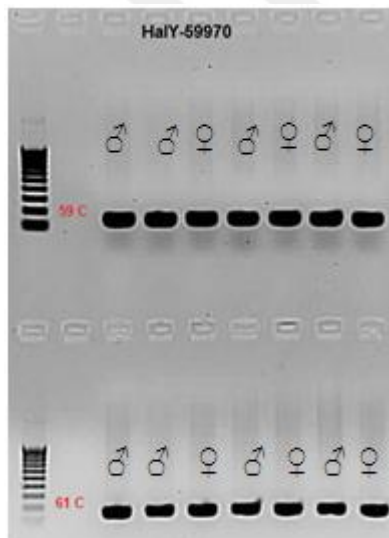


Figure 2

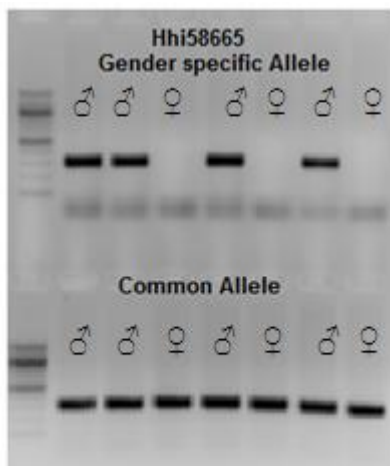


Figure 3

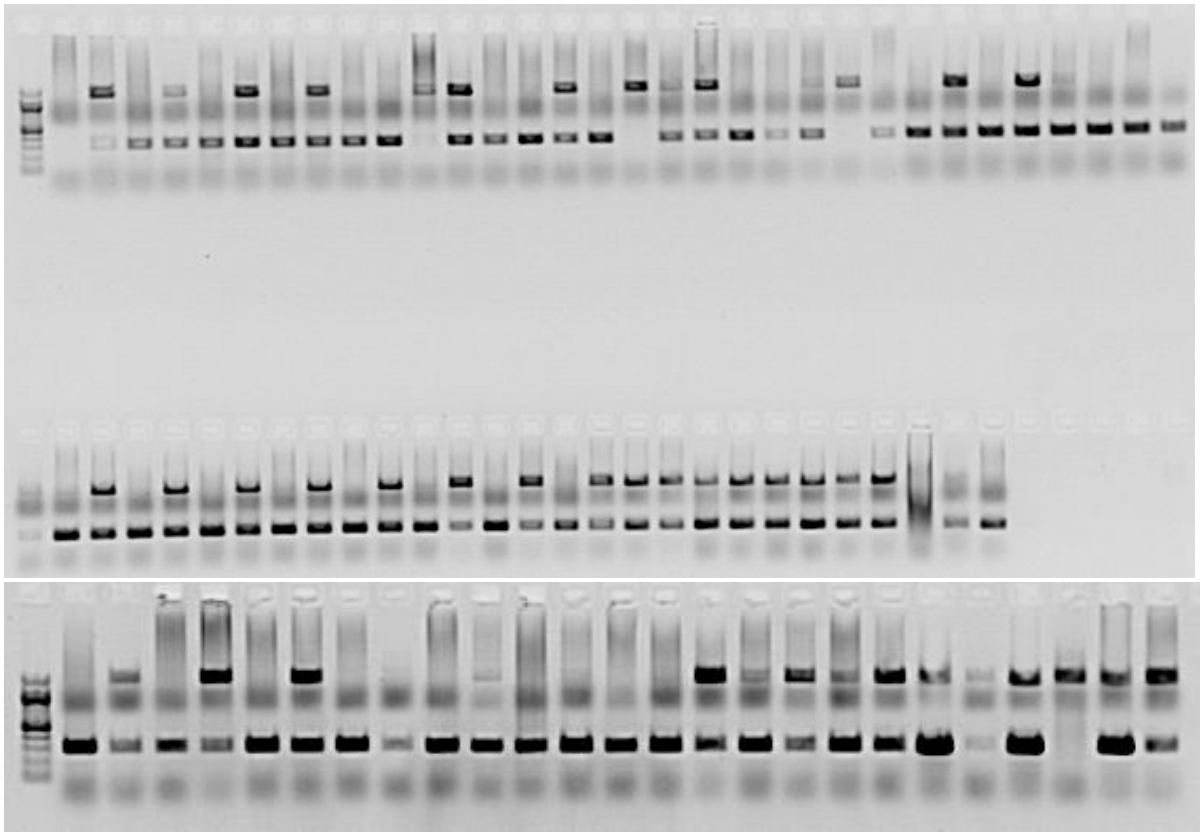


Figure 4

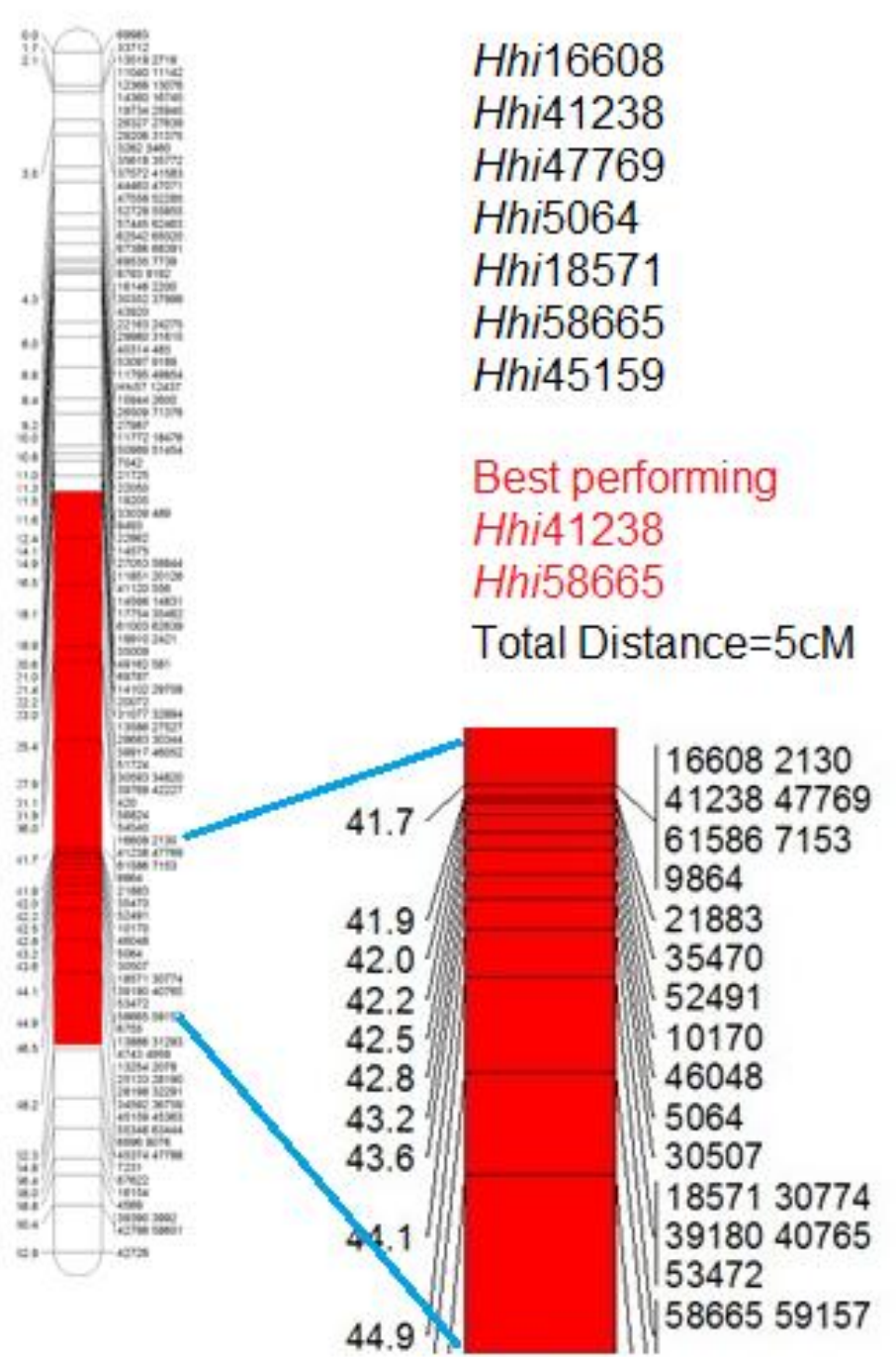


Figure 5

Captions for Tables

Table 1

The World production of the Atlantic halibut, FAO (tonnes)

Table 2

Nucleotide sequence, product size and Ta of Putative Y primers used in the study.

Table 3

Nucleotide sequence, product size, Ta and assay type [either MyTaq™ or Clear Taq (Qiagen)] and optimised PCR conditions of all Allele Specific primers used in the study.

Table 4

Pedigree structure of unrelated halibut families, number of offsprings and the code of chosen individuals.

Table 5

Genotypic proportion of Allele specific markers through unrelated halibut families and association rate of each locus through five families.

Captions for Figures

Figure 1

The result of optimisation in *HalY-62390* Putative Y contigs.

Figure 2

HalY-59970, one of the examples of Putative Y locus through test panel used in study.

Figure 3

Hhi58665, one of the examples of allele specific locus through test panel used in study.

Figure 4

Hhi58665, one of the examples of allele specific locus through unrelated families' offsprings used in the study.

Figure 5

Genomic map of Linkage group 13 (LG13) which contains all RAD markers that was generated previously with the focus of SNPs that used in this study.