Genome wide association study for salmon lice (*Lepeophtheirus salmonis*) resistance in Atlantic salmon (*Salmo salar*)

by

Christina Marie Rochus

A Thesis
presented to
the University of Guelph

In partial fulfilment of requirements
for the degree of
Master of Science
in
Integrative Biology

Guelph, Ontario, Canada

ABSTRACT

GENOME WIDE ASSOCIATION STUDY FOR SALMON LICE (*Lepeophtheirus salmonis*) RESISTANCE IN ATLANTIC SALMON (*Salmo salar*)

Christina Marie Rochus
University of Guelph, 2013

Advisor:
Dr. Elizabeth Grace Boulding

My objective was to detect single nucleotide polymorphism (SNP) associations with salmon lice resistance in an aquacultural population of Saint John River Atlantic salmon. In 2011 and 2012 I challenged recent smolts with copepodids from 42 and 47 families respectively. Fish were euthanized once the lice reached the chalimus stages and lice count, sex, tank and weight were recorded. I used a multiple trait model to estimate breeding values for parents of challenged fish using fresh water weights collected on the parent generation and the salt water weights and lice counts collected on the challenged fish. Using 299 individuals that had deregressed estimated breeding values and had been genotyped for 3638 SNPs, I detected 70 SNP associations using a forward regression. I was able to detect SNP associations with salmon lice resistance which, with further research could lead to genomic selection for this economically important trait.

ACKNOWLEDGMENTS

Finally, I would like to thank my friends and family for all of their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

CHAPTER 1: GENERAL INTRODUCTION

1.1 Wild and cultured Atlantic salmon

Atlantic salmon (*Salmo salar*) are a very diverse species with a complex life cycle. Native to the subarctic and temperate regions of the North Atlantic Ocean, many Atlantic salmon are anadromous, able to live in both fresh and marine water environments (Thorstad et al., 2012). For these salmon, mature adult males and females migrate from the ocean up rivers to spawn where females lay their eggs in the gravel bottom of the river bed, which are then fertilized by the males between September and February (Thorstad et al., 2011). After hatching in the spring and emerging from the gravel, juvenile salmon, parr, remain in the river for one to eight years until they are ready to smolt and ocean conditions are favourable (Hvidsten et al., 1998, Klemetsen et al., 2003). Smolts are physiologically and morphologically prepared to live in a marine environment and travel to the ocean where they can take one to five years to become mature adults, ready to return to the river they came from and spawn (Evans and Claiborne, 2006). The Atlantic salmon is iteroparous having the potential to spawn more than once in their life (Thorstad et al., 2011).

Atlantic salmon genetics show the diversity of this species. Previous work including protein electrophoretic studies and mitochondrial DNA restriction fragment length polymorphisms have been able to demonstrate that there is evidence for divergence between European and North American Atlantic salmon (Davidson et al., 1988). On the chromosome level, a large difference can be seen with European Atlantic salmon generally having 29 pairs of chromosomes and 74 chromosome arms (Hartley, 1987) and North American Atlantic salmon generally having 27 pairs of chromosomes and 72 chromosome arms (Roberts, 1970). Brenna-Hansen et al. (2012) compared the linkage maps of European and North American Atlantic salmon and found that a

large proportion of single nucleotide polymorphisms (SNPs), DNA markers, that had been found in European Atlantic salmon were also informative in the North American population and marker order was highly conserved. There were some large differences between the two subspecies: chromosomes identified as 26 and 28 in European Atlantic salmon were fused in the North American population, chromosomes 8 and 29 were also fused in the North American population and chromosome 23 and the p arm of chromosome 1 were fused together (Brenna-Hansen et al., 2012).

The first salmon farm in Canada was located in New Brunswick starting in 1979 and since that time salmon production has become a major Canadian aquaculture product (Surprenant, 2010). In 2009, Canadian aquaculture production was valued at approximately $800 million with $653 million attributed to salmon culture (Aquaculture Canada: Facts and Figures, 2011). By province, British Columbia produced the most salmon, valued at $349 million, followed by New Brunswick with $159 million (Aquaculture Canada: Facts and Figures, 2011). Salmon production in New Brunswick is limited to the Saint John River strain, an aquacultural strain derived from wild Atlantic salmon from the Saint John River and mariculture of this salmon is also popular in Nova Scotia and Maine (Aquaculture Canada: Facts and Figures, 2011).

Atlantic salmon production differs depending on the location and company. Salmon kept for breeding purposes, broodstock, in New Brunswick can remain in fresh water facilities their entire lives and when adults are mature and ready to spawn in the late fall, milt and eggs are stripped from males and females respectively (Quinton, 2005). Broodstock are generally only used once for breeding purposes and because sperm and eggs are collected, families of fish can be made quite easily for a breeding program (Quinton, 2005). Fish that are destined for food production are hatched and raised in fresh water until they smolt and at this time they are transferred to net

pens in the ocean and grown out until they are ready for harvesting (Quinton, 2005).

1.2 The Salmon louse (*Lepeophtheirus salmonis*)

The salmon louse (*Lepeophtheirus salmonis*) is a marine ectoparasite that feeds on both wild and farmed salmon (Kolstad et al., 2005). Salmon lice begin as eggs attached to strings on their mother which hatch into a free swimming stage that does not need to eat, called nauplii (Connors et al., 2011). Nauplii become copepodids and must seek out a host in order to survive tethering themselves to skin or fins of the fish with their frontal filament and remaining attached throughout the chalimus stages (Hayward et al., 2011). Chalimus stages are followed by two mobile stages; pre adult and adult stages and lice sexually reproduce when they are mature adults (Connors et al., 2011). Salmon lice growth is dependent on water temperature and the complete life cycle takes 40 days for males and 52 days for females at $10^{o}C$ (Johnson and Albright, 1991). From the copepodid through to the adult stages, lice diet consists mostly of fish skin, although blood and mucous are also often ingested (Pike and Wadsworth, 1999). Effects of this parasite on Atlantic salmon include dermal ulceration, osmotic imbalance, physiological stress, suppressed immune function and secondary infection and all of these can result in reduced fish growth, reduced feed conversion ratio and market downgrading (Mustafa et al., 2000).

As reported by the Food and Agriculture Organization (FAO), the cost of lice treatment control in Atlantic Canada was 0.10 €/kg salmon produced in 2008 (Costello, 2008). There are many different species of lice that affect salmonids and aquaculture production but the salmon louse is particularly problematic for production in Eastern North America and Europe (Pike and Wadsworth, 1999). Besides treatments already mentioned, there are costs associated with public perception, secondary infections for fish and preventative measures and research for new solutions (Costello, 2008). In October 2012, 84 whole Atlantic salmon were pulled from grocery

stores in Atlantic Canada after an anti-aquaculture activist posted pictures of purchased fish with lice on them on Facebook. This recall was despite Health Canada's statement that lice are not harmful for humans and that lice are also found on wild salmon (CBC News, 2012). Sea lice are thought to be a good vector for disease. Researchers have demonstrated in both laboratory tanks and commercial net pen situations that lice transfer hosts very easily (Ritchie, 1997). Ritchie (1997) reported that 61% of male and 69% of female lice transferred to new hosts in 3 days in the laboratory setting while 63% of male and 52% of female lice transferred to new hosts in 4 days in a commercial net pen. With their ability to travel between fish and infectious salmon anemia virus being isolated from salmon lice (Nylund et al., 1994), they could be participating in the spread of disease among salmonids.

Another problem with salmon lice is the concern that they may become resistant to chemical and non-chemical treatments used for controlling their numbers (Jones et al., 2012). Emamectin benzoate (SLICE™) is a chemical treatment which is delivered via fish feed and because of its efficacy and its ability to affect all stages of salmon lice it has been used for the majority of treatments for years in New Brunswick (Jones et al., 2012). There is now some evidence that the efficacy is being reduced in the Bay of Fundy (Jones et al., 2012).

1.3 Breeding for salmon lice resistance

Breeding Atlantic salmon that are more resistant to salmon lice is an opportunity for producers to manage sea lice effectively and to reduce the use of chemical treatments. Kolstad et al. (2005) estimated heritability for salmon lice resistance in European Atlantic salmon using a natural salmon lice infection occurring in a commercial setting. These researchers used lice count as a measure for resistance and found it had a low heritability of 0.14 with heritabilities for motile and sessile lice estimated to be 0.02 and 0.12 respectively and estimated that genetic

correlation between motile and sessile lice counts was high (r=0.98) (Kolstad et al., 2005). Kolstad et al. (2005) also estimated resistance in a laboratory setting where fish were infected in tanks (h^2=0.26) and that the genetic correlation between lice number in a controlled challenge and in a natural infection was high (r=0.88). Through their findings, these researchers concluded that future work could be done using controlled challenges by counting sessile lice, both of which make testing of resistance to salmon lice much easier (Kolstad et al., 2005). Gjerde et al. (2011) also estimated salmon lice resistance in European Atlantic salmon using a controlled challenge and counting sessile lice (h^2=0.33) but also estimated heritability for lice density (h^2=0.26) where lice density= lice count/ weight$^{2/3}$.

In 2010 a Norwegian company, SalmoBreed, which sells European Atlantic salmon eggs to producers began advertising eggs for fish that are more resistant to salmon lice (SalmoBreed AS, nd.). The website says that "SalmoBreed Exclusive Lice" Atlantic salmon are also being selected for gains in harvest weight and resistance to infectious pancreatic necrosis, infectious salmon anemia and furunculosis with a calculated progress for salmon lice resistance for the 2012/2013 season at 15-20% compared to the average of the parent generation. The sires of these fish are selected from the 10% of families with the lowest salmon lice counts of 300 families that are challenged and the dams are from the SalmoBreed broodstock lines (SalmoBreed AS, nd.).

One study of European Atlantic salmon has also looked for quantitative trait loci (QTL) for salmon lice resistance on two chromosomes. A QTL is a location in the genome that affects a quantitative trait (Snustad and Simmons, 2006). Gharbi et al. (2009) looked for QTL on two chromosomes known to contain major histocompatibility complex genes. The researchers found one QTL on chromosome 12 accounting for 12.9% of within-family variance for lice count

(Gharbi et al., 2009).

The genetic component of salmon lice resistance in Atlantic salmon could also be investigated using a genome wide association study (GWAS). GWAS uses markers across the genome to estimate all the genetic effects contributing to a trait. Many economic traits have been shown to be under the control of many genes and unlike QTL detection studies, GWAS can also have the ability to detect the smaller genetic effects (Goddard and Hayes, 2007). To help increase the ability of GWAS to detect markers that are truly associated with a trait, pedigree information and location of markers can be used and multiple tests can be accounted for. In my study population, there are a relatively small number of individuals with few records making forward regression a good method for the GWAS as it is not as stringent and therefore some of the smaller substitution effects can be estimated (Gu et al., 2010). Allele substitution effects, the difference between the average allele effects (Falconer, 1981), are important because they can be used for selection of future broodstock using genomics.

To date there are no published studies for salmon lice resistance in North American Atlantic salmon. Furthermore there are no published studies with North American or European Atlantic salmon that look at salmon lice resistance and use genetic markers that are on every chromosome.

1.4 Objectives, hypotheses and predictions

The main objective of this research was to detect single nucleotide polymorphism (SNP), associations with salmon lice resistance in an aquacultural population of Saint John River Atlantic salmon belonging to the North American subspecies using deregressed estimated breeding values (EBVs) and a 6000 array chip. My hypotheses and predictions were as follows:

H1: A genome wide association study (GWAS) using forward regression will result in the

detection of significant associations between salmon lice count EBVs and SNPs because salmon lice count is a heritable trait.

P1: I predicted that I would find SNPs that were associated with salmon lice count because it has been shown to be a heritable trait in the European Atlantic salmon (Kolstad et al., 2005, Gjerde et al., 2011) and a quantitative trait loci (QTL) has been previously detected on chromosome 12 (Gharbi et al., 2009).

H2: Comparison of associated SNPs between two year classes will result in a significant difference in allele substitution effects because the GWAS will also be detecting family effect.

P2: I predicted that the associated SNPs found would have different allele substitution effects for the two year classes I used. Estimated allele substitution effects are going to be different because dams and sires are not used for more than one year class.

CHAPTER 2: GENOME WIDE ASSOCIATION STUDIES FOR SALMON LOUSE (*Lepeophtheirus salmonis*) RESISTANCE IN ATLANTIC SALMON (*Salmo salar*)

2.1 Introduction

The salmon louse (*Lepeophtheirus salmonis*) is a marine ectoparasite found on salmonids globally. There are many different species of lice that negatively impact aquaculture production but the salmon louse is particularly problematic in eastern North America, and Europe (Costello, 2008). The salmon louse eats skin, mucus and blood of its host (Pike and Wadsworth, 1999) and an infestation can lead to open sores on skin, osmotic imbalance, physiological stress, suppressed immune function and secondary infection (Mustafa et al., 2000). Maricultured Atlantic salmon being can have reduced fish growth, reduced feed conversion ratio and market downgrading (Mustafa et al., 2000). There is also the possibility that salmon lice are a good vector of disease as researchers have demonstrated in both laboratory tanks and commercial net pen situations that lice transfer hosts very easily (Ritchie, 1997) and that it may be possible for lice to be vectors of disease like infectious salmon anemia virus (Nylund et al., 1994). Salmon lice can be treated using chemicals like emamectin benzoate, which is a popular chemical treatment for salmon lice delivered via fish feed (Jones et al., 2012). Because of the efficacy of emamectin benzoate and its ability to affect all stages of the salmon louse, it has been used for the majority of treatments for years and there is now some evidence its efficacy is being reduced in the Bay of Fundy (Jones et al., 2012).

Adult female salmon lice have two eggs strings attached that their eggs hatch from (Connors et al., 2011). When salmon lice first hatch, they are in the nauplii (two stages) and are free swimming and then become copepodids (two stages) and must seek out a host in order to survive (Pike and Wadsworth, 1999). Once the host is found, lice have four chalimus stages where the louse is tethered by their frontal filament to skin or fins of the fish (Hayward et al., 2011).

Chalimus stages are followed by mobile stages; pre adult and adult stages and sexually reproduce when mature (Connors et al., 2011). Salmon lice are highly affected by temperature and can complete their life cycle in 40 days for males and 52 days for females at 10°C (Johnson and Albright, 1991).

Researchers have been investigating the viability of breeding for salmon lice resistance. Kolstad et al. (2005) and Gjerde et al. (2011) looked at European Atlantic salmon susceptibility to salmon lice in a controlled challenge test estimating heritability to be 0.26 and 0.33. Kolstad et al. (2005) also estimated heritability for a natural infection occurring in a commercial setting and found total lice count had a low heritability of 0.14. These researchers estimated genetic correlation between salmon lice count in a controlled challenge and in a natural infection (r=0.88) and because of the high correlation, challenges done in a more controlled setting are acceptable (Kolstad et al., 2005). Gjerde et al. (2011) estimated heritability for lice density (h^2=0.26) where lice density= lice count/ weight$^{2/3}$.

One study of European Atlantic salmon has also looked at quantitative trait loci (QTL) for salmon lice resistance on two chromosomes. Gharbi et al. (2009) looked for QTL on two chromosomes known to contain major histocompatibility complex genes. The researchers found one QTL on chromosome 12 accounting for 12.9% of within-family variance for lice abundance.

Molecular genetics could be applied to breeding programs and gains could be larger using these tools when traditional selection is difficult (Goddard and Hayes, 2007). With marker assisted selection, animals are genotyped and effects of markers linked to QTL are estimated for a trait, genotypes and phenotypes are collected on another group of individuals to validate the effects (Goddard and Hayes, 2007). For broodstock candidates, they would not be challenged with lice, rather, they would be genotyped for markers for the QTLs and then all the effects

would be summed to obtain the genomic estimated breeding values (Goddard and Hayes, 2007). Genomic selection is a form of marker assisted selection where instead of a small number of markers, DNA markers covering the entire genome are used so that all QTL associated with a trait of interest are in linkage disequilibrium with at least one marker and can therefore be used for selection (Goddard and Hayes, 2007). In the case of genomic selection, marker allele substitution effects are estimated and they are summed to obtain the genomic estimated breeding values (Goddard and Hayes, 2007). Selection for a difficult to select for trait like salmon lice resistance could really benefit from genomic selection.

To date there are no published studies for salmon lice resistance in North American Atlantic salmon. Furthermore, there are no published studies with North American or European Atlantic salmon that look at salmon lice resistance and use genetic markers that are on all chromosomes. The objective of this research was to detect single nucleotide polymorphism (SNP) associations with salmon lice resistance in an aquacultural population of Saint John River Atlantic salmon using deregressed estimated breeding values and a 6000 array chip.

2.2 Material and Methods

Three generations of North American Atlantic salmon from a Saint John River aquacultural population were used in this study. Familial relationships were previously estimated using microsatellite data for the second generation and were known for the third generation because full siblings were kept in separate tanks until they were large enough (8 grams) to be uniquely identified using passive inductive transponder (PIT) tags. The second generation was made of 214 full sibling families (112 sires and 213 dams) with an average of 17 fish in each. The third generation was made of 90 full sibling families (68 sires and 90 dams) with an average of 20 fish in each. 240 of 436 fish in the first generation and 299 of 36333 in the second generation were

genotyped while fish from the third generation (N=1803) were challenged with salmon lice.

Salmon lice counts and salt water weights from the third generation along with the fresh water

weights from the second generation were used to estimate breeding values (EBVs) for salmon

lice count. A forward regression was used to look for significant associations between DNA

markers, single nucleotide polymorphisms (SNPs), and the deregressed EBVs for salmon lice

count using the 299 fish from the second generation that had been genotyped. Figure 1 depicts

an overview of the methods for this study.

2.2.1 Genotyping

The first two generations of Atlantic salmon from our Saint John River aquacultural

population were genotyped using a custom Atlantic salmon iSelect™ 6 K bead array by Illumina

(San Diego, USA) described by Brenna-Hansen et al. (2012). All genotyping was carried out at

the Centre of Integrative Genetics (CIGENE), Norwegian University of Life Sciences, Ås,

Norway. This bead array was designed for the European Atlantic salmon subspecies and

includes both single nucleotide polymorphisms (SNPs) and multisite variants (MSVs).

Salmonids have experienced a whole genome duplication event relatively recently which made

them tetraploids (Danzmann et al., 2008). They are currently in the process of returning to

diploidy so at the moment only a third of the genome remains tetraploid (Gidskehaug et al.,

2011). MSVs are informative but are located in tetraploid regions of the genome and more

complicated to work with than SNPs.

Once all fish were genotyped, each locus was inspected using GenomeStudio© (Illumina).

This was to determine whether it was polymorphic or not in our population. Monomorphic loci

have alleles that are fixed and therefore not informative. I also inspected polymorphic loci to

determine whether they were SNPs, MSVs or paralogous sequence variants (PSVs). MSVs can

be placed in one of three categories: MSV5 where five different genotype combinations are present in the population (AAAA, AAAB, AABB, ABBB and BBBB), MSV4 where only 4 of the 5 genotypes are present in the population, or MSV3 where one of the paralogues is fixed and only 3 of the 5 genotypes are present in the population (Gidskehaug et al., 2011). A PSV, which can easily be mistaken for a SNP, occurs when there are two paralogs that have a base pair difference but there is no segregation between either and they are not informative (Gidskehaug et al., 2011). GenomeStudio cannot automatically identify PSVs or MSVs therefore I identified them through visual inspection of GenomeStudio's polar co-ordinate graphs with signal intensity, norm R, on the y axis and allele frequency, norm theta, on the x axis. Only loci identified as SNPs in our population were used for analysis.

Genotypes from the first generation of the study population were used to estimate familial relationships (full sibling and half sibling) amongst the fish using the pedigree estimating software, COLONY (Wang, 2012). I used 493 SNPs from the 6 K array with a call rate of 1 and a minor allele frequency from 0.12 to 0.50 for 240 fish in the first generation in this program. Allelic dropout rate and the rate of other genotyping errors of the markers were assumed to be zero. Sibship size prior, average paternal and maternal sibship sizes, were both set to 1 and females were assumed to be monogamous and males polygamous. I determined the described program settings by using the SNP genotypes from the second generation who had their pedigree previously estimated by microsatellite genotyping.

2.2.2 Salmon lice challenges

Saint John River Atlantic salmon used in the salmon lice challenges, hatched in early 2010, for the September 2011 challenge, and early 2011, for the November 2012 challenges, in a New Brunswick hatchery. Fish were raised in fresh water with families kept in separate tanks until

they were large enough to identify individuals with a passive inductive transponder (PIT) tag. Salmon lice challenges were carried out in the quarantine laboratory at the St. Andrew's Biological Station, Fisheries and Oceans Canada, St. Andrews, New Brunswick.

In 2011, salmon lice were donated by harvest boats belonging to Cooke Aquaculture Inc.. Egg strings from adult, mature female salmon lice were hatched at the facility at the Huntsman Marine Sciences Centre. Once lice hatched and reached the copepodid stage their numbers were estimated by sampling five 1 ml aliquots and counting the number of lice in each. The average number of lice in a 1 ml aliquot was then used to obtain an approximate number of lice needed for the challenge. On September 8, 2011, 830 recent smolts, from 42 full sibling families from Cooke Aquaculture's breeding program, were challenged in two tanks by adding 30 copepodids per fish to tanks and water flow was stopped for an hour (while maintaining oxygen levels between 80% and 110% saturation). Four days later two fish were removed and all lice were counted. Because of the low numbers of salmon lice on the fish, I decided to infect with an additional 60 lice/fish which were added using the same methods as before and lice were permitted to grow for four more days. All fish were euthanized by an overdose of the anesthetic, tricane methanesulphonate (TMS). Fish were bagged individually and frozen. Frozen fish were thawed and then lice were counted using dissection microscopes. Besides total lice counts, I also collected: length, weight, sex, and tank. Lice that were counted were in the chalimus I, II, III and IV stages. Counting salmon lice on fish that had been frozen and then thawed was difficult so I decided to try different methods in 2012.

In 2012 salmon lice were again donated by harvest boats belonging to Cooke Aquaculture Inc.. Eggs hatched from the mature adult female salmon lice with egg strings were purchased from the Huntsman Marine Sciences Centre, St. Andrews, New Brunswick. Salmon lice were

hatched and counted at the Huntsman Marine Sciences Centre. Copepodids were then taken to the St Andrews Biological Station quarantine laboratory where recent smolts were kept. Fish from 47 full sibling families from Cooke Aquaculture's breeding program were infected with 100 lice per fish and at the time of infection, water flow was stopped with oxygen levels between 80-110% saturation. We again waited for lice to reach the chalimus stages before counting. In 2012, fish were euthanized and had their salmon lice counts and weights taken immediately afterwards and therefore we set up a schedule of trials so that all lice counting could be done on chalimus stages. Table 1 is the schedule of staggered salmon lice infections in November and December 2012 and the dates of salmon euthanasia and lice counting. Tanks 5, 6, 7 and 8 had 100 fish each except for the infection on November 28 (tank 5 and 6 had 50 fish each). Tank 1 and 2 were stocked with 200 fish each. On the day that lice were counted, fish were euthanized by an overdose of TMS. Salmon lice counted were in the chalimus I, II, III, and IV stages. Salmon lice were counted by professional lice counters from the Huntsman Marine Science Centre while illuminating the freshly killed fish with bright light, and unlike the challenge in the previous year, no microscope was required.

   I used SAS (SAS Institute, Inc., Cary, NC) to examine three traits; fresh water weights at three years of age from the second generation of fish and salt water weights and salmon lice counts at one and a half years of age from the third generation of fish. I first looked at summary statistics to see if I needed to transform any of the traits. Data should be transformed to make the distribution normal and to make variance independent of the mean (Falconer, 1981). For those reasons, I transformed salt water weights and salmon lice counts and used the natural logarithm of salt water weight and square root of salmon lice counts instead. The natural logarithm transformation is suggested to make variance independent of mean and the square root

14

transformation is suggested to make count data normally distributed (Sokal and Rohlf, 1995). I

used the SAS procedure, PROC GLM, to identify effects that were significant for fresh water

weights, salt water weights and salmon lice counts using general linear models. The linear

model that I used to explain variation in fresh water weight was:

$$y_{1ijk} = F_{1i} + CG_{1j} * S_{1k} + e_{1ijk} \tag{5}$$

where:

y_{1ijk} is fresh water weight (kg) at three years of age from the second generation,

F_{1i} is the family effect,

CG_{1j} is the contemporary group which includes year and tank,

S_{1k} is sex (male or female) and

e_{1ijk} is the random residual effect.

 The general linear model that I used to explain variation in the natural logarithmic

transformed salt water weights was:

$$y_{2ijk} = F_{2i} + CG_{2j} * S_{2k} + e_{2ijk} \tag{6}$$

where:

y_{2ijk} is natural logarithm of salt water weights at one and a half years of age that were measured

in grams from the third generation,

F_{2i} is the family effect,

CG_{2j} is the contemporary group which includes year, tank and time of measurement,

S_{2k} is sex (male or female) and

e_{2ijk} is the random residual effect.

 The general linear model that I used to explain variation in square root transformed salmon

lice counts was:

15

$$y_{3ijk} = F_{3i} + CG_{3j} * S_{3k} + e_{3ijk} \tag{7}$$

where:

y_{3ijk} is the square root of lice counts from the third generation,

F_{3i} is the family effect,

CG_{3j} is the contemporary group which includes year, tank and time of measurement,

S_{3k} is sex (male or female) and

e_{3ijk} is the random residual effect.

2.2.3 Estimation of breeding values

I used ASReml (Gilmour et al., 2006) to estimate breeding values (EBVs) with a multiple trait model using fresh water weights from the second generation, salt water weights and salmon lice counts from the third generation and a pedigree. The multiple trait model that I used for fresh water weights from the second generation of fish, salt water weights and salmon lice counts for the third generation of fish was:

$$y_{nijk} = CG_{ni} * S_{nj} + a_{nk} + e_{nijk} \tag{8}$$

where:

y_{nijk} is the phenotypic records for trait n; fresh water weight; salt water weight or salmon lice count,

CG_{ni} is the contemporary group (fixed effect) which includes year and tank,

S_{nj} is sex (male or female) (fixed effect),

a_{nk} is the animal effect (random effect) and

e_{nijk} is the residual effect (random effect).

The same multiple trait model was also run with salt water weights and salmon lice counts not transformed and a multiple trait model with just two traits was run (transformed salt water

weights and salmon lice counts).

A multiple trait model was used to take advantage of simultaneous analysis of a low heritable trait, salmon lice count and two higher heritability traits, salt and fresh water weights, which can increase accuracy for the trait with a low heritability (Schaeffer, 1984). The inclusion of both fresh water weights and salt water weights increased the number of records on individuals which could increase accuracy of EBVs (Muir, 2007). Also, the inclusion of fresh water weights from the second generation allowed fish that had been genotyped, but that did not have any progeny challenged with salmon lice, to have a lice count EBV distinguishable from their siblings. ASReml estimated breeding values (EBVs) for all three traits for all fish included in the pedigree, and genetic parameters including heritability, variances and genetic and phenotypic correlations. The pedigree for the three generations included the relationships estimated among the first generation with SNP genotypes. For those individuals in the first generation that were not genotyped, phantom parents were assigned by hatch year and sex of parent.

I used the methods described by Garrick et al. (2011) to deregress EBVs. This was done because EBVs, calculated using best linear unbiased predictor (BLUP), have less variance than phenotypes and this leads to the best performing animals being under evaluated and the worst performing animals being over evaluated (Garrick et al., 2011). EBVs are shrunk towards the mean based on reliability which is affected by the amount of data for each individual (Garrick et al., 2011). The researchers also proposed a weighting calculation based on reliability so that EBVs with a higher reliability have more weighting in further analyses and heterogeneous variances of deregressed EBVs are accounted for (Garrick et al., 2011).

For the forward regression I used EBVs for salmon lice counts rather than records on individuals because traits with low heritability because with enough records from relatives,

EBVs can be more accurate. For example, using the heritability of lice resistance from Kolstad et al. (2005), if an individual has a record for itself accuracy is the square root of heritability (Mrode and Thompson, 2005) or $\sqrt{0.26} = 0.51$ in this case. The predicted accuracy of the EBV for single records on n full-sib progeny from Cameron (1997) is:

$$accuracy = \frac{1}{2}\sqrt{\frac{nh^2}{1+(n-1)\frac{1}{2}h^2}} \qquad (1)$$

Using a heritability of 0.26 (Kolstad et al., 2005) and considering that the average family size was 20, the expected accuracy for a fish in the parent generation would be 0.61. The observed EBV accuracy was calculated in ASReml by:

$$accuracy = \sqrt{1 - (SE^2/\sigma_a^2)} \qquad (2)$$

where SE is the standard error for the EBV, and σ_a^2 is the additive genetic variance (Gilmour et al., 2006). Another reason for not getting phenotypic records directly from fish that are genotyped is that in this case they were broodstock and in this aquacultural population broodstock are kept in fresh water for their entire lives. Since salmon lice can only survive in salt water, the broodstock cannot be challenged and therefore relatives must be challenged instead.

2.2.4 Forward regression

For the genome wide association study, I used a forward regression using PROC GLMSELECT in SAS (SAS Institute, Inc., Cary, NC). The procedure finds the single nucleotide polymorphism (SNP) that has the most significant F-statistic for the deregressed estimated breeding values (EBVs). That SNP is added to the model and then the procedure finds the next SNP with the most significant F-statistic for the deregressed EBVs after taking into account the effect of the first SNP already found. Then the procedure finds the next SNP with the most significant F-statistic for the deregressed EBVs after taking into account the effects of the SNPs

18

already found. The procedure continues to add the SNP with the next most significant F-statistic

until adding a SNP would produce a higher Akaike information criterion (AICC) then when the

previous SNP was added. AICC is calculated in SAS by:

$$1 + ln\left(\frac{SSE}{n}\right) + \frac{2(p+1)}{n-p-2} \qquad (3)$$

where SSE is the residual sums of squares, n is the number of observations and p is number of

parameters including the intercept (Hurvich and Tsai 1989).

   For SNPs that were found to be associated with salmon lice count, I looked up their

chromosome location on the published North American linkage map (Brenna-Hansen et al.,

2012). SNPs that were not on the North American map had the chromosome for the European

subspecies of Atlantic salmon reported instead (Lien et al., 2011).

   The published European map (Lien et al., 2011) includes supplementary materials containing

short DNA sequences for each SNP locus with the position of the SNP or SNPs indicated. I

blasted the translated protein and DNA sequences of the first 20 SNPs entered into the forward

regression using NCBI blastx and NCBI nucleotide blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi)

respectively. I recorded any similar annotated protein or DNA sequences from NCBI that had E-

values $< 10^{-3}$.

   I used the SNPs that were significantly associated with the EBVs in a weighted multiple

linear regression using PROC GLM in SAS. For this weighted multiple linear regression I used

all of the SNPs that were in the results of the forward regression as fixed effects (in the same

order). This procedure estimated the difference between the average allele effects which is also

known as the allele substitution effect (Falconer, 1981).

   I also calculated the partial coefficient of determination (partial R^2) for each SNP:

$$partial\ R^2 = \frac{SSE_{reduced} - SSE_{full}}{SSE_{reduced}} \qquad (4)$$

where SSE_{full} is the residual sums of squares from the multiple linear regression with all SNPs

included and $SSE_{reduced}$ is the residual sums of squares from the multiple linear regression with

all but the SNP of interest included (Neter et al., 1990). The partial R^2 is a measure of the effect

of each separate SNP on salmon lice count EBVs after accounting for the effects of all other

SNPs that were found to be associated (Neter et al., 1990). SNPs that are added sooner in the

forward regression don't necessarily have a higher partial R^2 or allele substitution effect and that

can be explained as SNPs were entered into the model based on the significance of their F-

statistic.

I identified allele substitution effects different from zero and then used them in SAS to run

two more weighted multiple linear regressions, one for each separate year class of the second

generation, 2006 and 2007, to estimate the allele substitution effects for the two years. I used a

paired t-test to determine if the allele substitution effects were different between the two year

classes.

2.3 Results

2.3.1 Genotyping

A total of 4467 single nucleotide polymorphisms (SNPs) were identified on the 6K chip as

suitable for the studied Saint John River aquacultural strain of North American Atlantic salmon

(Table 2). Only 502 loci were monomorphic and therefore not informative (Table 2). There

were 573 multisite variants (MSVs), 12 unscorable loci and 11 loci that were unable to

hybridize, all of which were not used in this study (Table 2).

SNP genotypes were used to identify close relatives among the first generation of fish. In the

first generation 240 of the 436 fish were genotyped. COLONY estimated that these 240 fish

belonged to 116 full sibling families made from 116 dams and 65 sires. The estimated full and half sibling family relationships were then recorded in the pedigree. Of the remaining 196 fish, 18 had family information previously known. Those 18 fish belonged in 13 full sibling families made from 13 dams and 13 sires. The remaining 178 fish were assigned phantom parents by year of hatch and sex of parent which resulted in 6 full sibling families (six phantom dams and six phantom sires). These relationships were included in the pedigree used for the multiple trait model and can be found in Appendix B.

2.3.2 Salmon lice challenges

Summary statistics for fresh water weights, salt water weights and salmon lice counts are presented in Table 3. Salt water weights were transformed because mean and variance of weight were correlated for contemporary group-sex subclasses. This positive correlation is shown in Figure 2a. Transforming salt water weights by taking the natural logarithm solved this problem and the new means and variances are plotted in Figure 2b. Lice counts were also transformed because of a positive linear relationship between lice count means and lice count variance for contemporary group-sex subclasses (Figure 3a). Figure 3b shows the transformed lice count means and variance for contemporary by sex subclasses. The linear model that I used to explain variation in fresh water weight had an R^2 value of 0.78. Further results can be found in Appendix A i. The general linear model that I used to explain variation in the natural logarithmic transformed salt water weight had an R^2 value of 0.79. Further results can be found in Appendix A ii. The general linear model that I used to explain variation in square root transformed salmon lice counts had an R^2 value of 0.54. Further results can be found in Appendix A iii.

2.3.3 Estimation of breeding values

Estimated phenotypic and genetic correlations are reported in Tables 4a for transformed and in 4b for non-transformed data. Heritabilities for fresh water weights, salt water weights and salmon lice counts were reported in Table 4a and 4b using transformed and not transformed data. Salmon lice count heritability for data not transformed and transformed was 0.17 and 0.29 respectively.

Table 5 has summary statistics for accuracies for two and three trait model estimated breeding values (EBVs) for the 299 fish that were used for the forward regression (in the second generation and genotyped). There was a difference between the two models with the three trait model with higher accuracies by 0.0702 (paired t-test, p<0.0001). Of the 299 fish there were four main groups; 155 were parents of salmon lice challenged fish; 130 were aunts or uncles; seven were half aunts or uncles and seven were not parents, aunts/uncles or half aunts/uncles but were more distantly related. Figure 4 reports the distribution of the EBV accuracies for the 299 fish used for the genome wide association study (GWAS) where 188 of 299 fish had accuracies above the accuracy of a single observation on an individual (estimated to be 0.51). Figure 4 shows the accuracies broken down by the four groups previously mentioned.

Deregressed EBVs for the 299 fish that had genotypes can be found in Appendix C of the additional materials.

2.3.4 Forward Regression

70 single nucleotide polymorphisms (SNPs) were found to be associated with salmon lice count. These SNPs were located across the genome on all chromosomes except two, *Ssa* 25 and *Ssa* 26/28. These 70 SNPs are in Appendix D along with order of entry into the model, corrected Akaike information criterion (AICC), F-statistic, P-value, partial coefficient of determination (partial R^2) and allele substitution effect and standard error. Figure 5 is a graph depicting the

chromosome each associated SNP is located on and the P-value for the F-statistic to enter from the forward regression. NCBI blastx and nucleotide blast results for the first 20 SNPs entered in the forward regression can be found in Table 6. Gene or protein identifications were found for 15 of the 20 SNPs. Figure 6 shows the distribution of back transformed allele substitution effects and Figure 7 shows the distribution of partial R^2 for all of the 70 SNPs found to be associated with salmon lice count. 68 of the 70 SNPs had an allele substitution effect different from zero. The difference between the estimated allele substitution effects for the two separate year classes was not significant (paired t-test, P=0.058) as seen in Table 7.

2.4 Discussion

The heritabilities estimated for salmon lice count (0.17) and transformed salmon lice count (0.29) in my study were in the same range as previous studies for the European subspecies of Atlantic salmon. Both of the previous estimates had heritabilities that were moderate: Kolstad et al. (2005) estimated heritability to be 0.26 and Gjerde et al. (2011) estimated heritability to be 0.33. Fitness traits, such as resistance to parasites, usually have low heritabilities and high environmental variances because of strong directional selection (Falconer, 1981) and they are often influenced by many other traits (Price and Schluter, 1991). Therefore, moderate heritabilities that were estimated in previous literature (Kolstad et al., 2005, Gjerde et al., 2011) and in my study are expected because they were estimated from controlled challenges conducted in laboratories where environmental variance is reduced. This reduced environmental variance would be expected to increase my statistical power to detect SNPs associated with increased salmon lice counts in my subsequent analyses.

The results of my genome wide association study (GWAS) showed that salmon lice resistance is a quantitative trait. This was evident from the distribution of back transformed allele

23

substitution effect and partial R^2 for the 70 single nucleotide polymorphisms (SNPs) found to be associated in Figures 6 and 7: there were many small effects and few large effects (Snustad and Simmons, 2006). The distributions of SNP effects was very similar to the hypothetical distribution of effects for frequency of bristles in *Drosophila melanogaster* (Shrimpton and Robertson, 1988) and the distribution of QTL effects in pigs and dairy cattle using a meta-analysis of published estimates (Hayes and Goddard, 2001). Many traits in humans and traits of economic importance in livestock are now known to be affected by many genes with most gene effects being quite small (Goddard and Hayes, 2007). The gamma distribution of allele substitution effects that I found is similar to that predicted by theory and suggests that the SNPs I found associated with genetic resistance to salmon lice resistance in Atlantic salmon are real.

I used a forward regression for the GWAS because of its simplicity as it only required deregressed estimated breeding values (EBVs) and SNPs. Other GWAS based on separate SNP analysis have reduced power to detect weak genetic effects and can be too stringent when sample size is small for a complex disease (Gu et al., 2010) because they account for multiple tests, marker location and pedigree. When correcting for multiple tests you can estimate Type I error using permutations and increase the power over that resulting from a Bonferonni correction (Churchill and Doerge, 1994). Linkage is also important because if a SNP is in or closest to a QTL then the SNPs in the immediate area will also be associated with the trait of interest leading allele substitution effects to be overestimated (Gu et al., 2009). Although forward regression does not use marker location information, it does somewhat take care of the problem as a SNP is unlikely to be added to the forward regression if a SNP that was close to it had a significant F-statistic and was already included in the results. Forward regression is not seen in breeding literature often but there have been a few other GWAS using F-statistics for entering SNPs into

the model including in studies for maize leaf architecture (Tian et al., 2011) and kernel composition (Cook et al., 2012) as well as for meat and carcass traits in Australian cattle (Bolormaa et al., 2011). SAS does not recommend the use of the F-statistic to stop the addition of new SNPs to the forward regression model because they don't follow an F-distribution and p-values can't reliably be used as probabilities resulting in over or underfitting of a model (SAS Institute, Inc., Cary, NC). This explains why the final model from the forward regression (using the AICC value to stop) included 14 SNPs with a p-value greater than 0.05 in my results.

The results of the forward regression provided SNPs that were associated with salmon lice count because SNPs were directly involved in the genetic control of the trait or because they were tightly linked to the functional region (Lynch and Walsh, 1998).

To determine whether SNP associations were because of direct involvement in the gene, I blasted the first 20 SNPs in the forward regression. The second SNP entered into the forward regression is involved in a gene for immunoglobulin D (IgD), which is found in the membrane of B cells (Kindt et al., 2007). Although IgD's function is unknown, B cells are important in the antibody mediated, adaptive immune system (Kindt et al., 2007). Three other interesting findings from blasting SNPs were; the sixth SNP added, osteoclast stimulating factor 1; the twelfth SNP added, apoptosis inhibitor 5; and the twentieth SNP added, cytochrome c-like. Osteoclasts are macrophage-like cells that free calcium from bones (Kindt et al., 2007) and salmon lice disturb and damage scales, which are made of calcium, to reach the skin of Atlantic salmon (Pike and Wadsworth, 1999). Apoptosis is a type of programmed cell death and is considered part of the immune system (Kindt et al., 2007) and cytochrome c triggers one of the pathways which leads to apoptosis (Tizard, 2009).

I also looked for SNPs that are related to other traits of interest by looking at the literature for

QTL detection. I then compared chromosome number found in Table 8. Gharbi et al. (2008) have already detected QTL for salmon lice resistance. Their study only looked at linkage group 15 and linkage group 6, the locations of *UBA* a locus for major histocompatibility complex I and *DAA* a locus for major histocompatibility complex II. The researchers used three families that were shown to have significant associations between lice counts and marker variation at *UBA*, *DAA* or both (Glover et al., 2007). They found one QTL on chromosome 12 with 12.9% of within-family variance explained. In my study, three SNPs were found on chromosome 12 with the first appearing in step 9 of my forward regression.

Gyrodactylus salaris, a freshwater ectoparasite of salmonids first seen in Norway has also had QTL detected (Gilbey et al., 2006). The researchers used backcrosses of Baltic and Scottish Atlantic salmon and 39 microsatellite markers. 10 QTL accounted for 27.3% of total variation in parasite control found on chromosomes 10, 13, 24, 3, 23, 4, 15, 17, 16, and 12 (Gilbey et al., 2006). The researchers also found that fish size (length, weight, condition) did not have an effect. The large number of QTL fits with my finding of many SNP associations if these are both quantitative traits controlled relatively similarly.

QTL have recently been detected for two health traits of interest to the salmon industry, infectious pancreatic necrosis (IPN) and infectious salmon anemia (ISA). IPN has a large QTL located on chromosome 26 which accounts for 70.3% of genetic variance (Moen et al., 2009). Moen et al. (2007) detected a QTL on chromosome 15 which accounted for 6% of phenotypic variance. Both studies used European Atlantic salmon. I did not find any SNPs on chromosome 26 but did find four on chromosome 15.

Finally, body weight and condition are very important production traits and Reid et al. (2005) detected QTL for both traits from Saint John River Atlantic salmon from a broodstock program

in New Brunswick. They found QTL for body weight on 15 and 3 and QTL for body condition on 21, 3, 13, and 10. Even though my study did not find a genetic correlation between body weight and salmon louse count, future studies should also estimate correlations between these two traits because a larger fish with larger surface area might have more lice.

Genomic selection might be useful for selecting for salmon lice resistance because it is a difficult trait to select for with a low heritability. Traits with those two features benefit the most from genomic selection (Goddard and Hayes, 2007). Sonesson and Meuwissen (2009) used simulation to examine strategies for genomic selection in aquaculture breeding programs and found that there were long term increases in accuracy when the trait of interest was measured every generation. Their simulation used 3000 selection candidates, 3000 tested siblings, 100 full sibling families and a heritability of 0.4 for their trait of interest (Sonesson and Meuwissen, 2009). Despite this benefit in accuracy, estimated breeding values (EBVs) may be currently more accurate for this population. My study used a small number of individuals and markers to estimate allele substitution effects so I would not expect to see a similar increase in accuracy in my population. I tested only 1733 progeny and genotyped only 299 individuals whereas they simulated 3000 tested and 3000 genotyped siblings. The heritability they used in their simulation of 0.4 was also much higher than lice count the heritabilities that I observed of 0.17-0.33. Their simulation also used a smaller diploid genome with 10 chromosomes that were each 100 cM in length and more markers than I had (10 000 SNP markers). It is difficult to properly estimate allele substitution effects with the small amount of data I had and some SNP effects may not have been detected (Hayes et al., 2009). In contrast, EBVs include all the QTL for a trait of interest while genomic selection only uses markers estimated to have an effect (Goddard, 2009).

Before genomic selection can happen, allele substitution effects should be estimated accurately. In this study I used two year classes and because sires and dams are only used once for breeding (with very few exceptions) no fish in one year class is a sibling or half sibling with fish in another year class, although there may be cousins. Fish from both year classes are descended from a common wild population from the Saint John River in New Brunswick so they should have the same very close linkage disequilibrium between QTL and markers (Dekkers, 2004). If this LD was captured then allele substitution effects would likely be the same however the low density of markers (less than 4000) that I used likely did not capture it. Allele substitution effects are unlikely to be the same when comparing them from these two groups because they represented different families, each possibly with different linkage phases. However, there was not quite a significant difference ($p=0.058$) when I compared the allele substitution effects from the two year classes (Table 7). It seems like there may be a difference in the allele substitution effects between the two year classes but there was insufficient data to detect it. It is important to remember allele substitution effects may be different for each year class and that I may not have had sufficient LD to detect all the QTL effects, which is what is needed for genomic selection (Dekkers, 2004). Therefore it may be premature to apply my results to a breeding program.

5. Conclusions

Salmon lice resistance is a quantitative trait in the Saint John River aquacultural Atlantic salmon population that I studied. Using a forward regression, I detected 70 SNPs associated with salmon lice resistance. SNPs were located on every chromosome except *Ssa* 25 and *Ssa* 26/29. There was no evidence of a difference in the allele substitution effects for the associated SNPs between the two year classes used in the study. In future, salmon lice resistance, a

heritable trait, might possibly be changed through selection using markers but more research is

needed.

CHAPTER 3: GENERAL CONCLUSIONS

The objective of this research was to detect single nucleotide (SNP), associations with salmon lice resistance in an aquacultural population of Saint John River Atlantic salmon using deregressed estimated breeding values (EBVs) and a 6000 array chip. I detected 70 SNPs that were significantly associated with deregressed EBVs for salmon louse count using two different year classes of fish. Fish are only used once for breeding and so is important to remember that allele substitution effects estimated for one year class may not be the same for another. There was a trend towards different allele substitution effects for the 70 associated SNPs between the two year classes but this difference was not statistically significant.

The identification of markers significantly associated with salmon lice count shows promise for genomic selection, the use of markers across the genome which can include all of the quantitative trait loci (QTL) that contribute to a trait (Hayes et al., 2009). The advantage of genomic selection would be that you could select the more resistant broodstock, fish never exposed to salmon lice, for breeding purposes just by using DNA marker information (Hayes et al., 2009). Schaeffer (2006) demonstrated how to replace or modify a dairy breeding program using genomic selection, showing that there could be a doubling of genetic change at 92% of the current cost of the program. Hayes et al. (2009) showed that for dairy breeding programs in Australia, the Netherlands, New Zealand and the United States of America, reliabilities calculated from genomic estimated breeding values (GEBVs) were much greater than breeding values from parental averages, where the GEBVs were calculated from a combination of breeding values from parental averages, pedigree information and genomic index using selection index theory.   Hayes et al. (2009) said that accuracy for genomic selection is under control of two things that can be changed: sufficient linkage disequilibrium between markers and QTL and

the number of animals with phenotypes and genotypes in the reference population. Dairy cattle have been genotyped for many SNPs (tens of thousands in Hayes et al., 2009) in comparison to the approximately 4000 SNPs that worked for the Atlantic salmon population in this study. Also, Hayes et al. (2009) reviewed national breeding programs where thousands of bulls were genotyped while our study only had 299 fish genotyped. Dairy cattle breeding programs which use genomic selection also receive gains through reduction in generation interval (Goddard and Hayes, 2007) and unfortunately, an Atlantic salmon breeding program would not receive the same benefit as an individual will have its own record and records of siblings before it even reaches maturity. Simulation studies done by Nielsen et al. (2009) showed that an aquaculture program can benefit from genomic selection as the genomic breeding values had good accuracies but the authors admitted that cost of genotyping is definitely a limiting factor for genomic breeding to be put into practice.

There are many benefits of genomic selection but for now more phenotypic and genotypic records should be collected before this is implemented in the Canadian aquaculture industry.

REFERENCES

Aquaculture Canada: facts and figures. 2011. Fisheries and Oceans Canada retrieved 31 January 2013 from http://www.dfo-mpo.gc.ca/aquaculture/ref/stats/aqua-ff-fc-2009-eng.htm#ch21.

Bolormaa, S., Porto Neto, L.R., Zhang, Y.D., Bunch, R.J., Harrison, B.E., Goddard, M.E., Barendse, W., 2011. A genome-wide association study of meat and carcass traits in Australian cattle. Journal of Animal Science 89, 2297-2309.

Brenna-Hansen, S., Li, J., Kent, M.P., Boulding, E.G., Dominik, S., Davidson, W.S., Lien, S., 2012. Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence *in situ* hybridization analysis. BMC Genomics 13, 432.

Cameron, N.D., 1997. Selection indices and prediction of genetic merit in animal breeding. CAB International: New York.

Churchill, G.A., Doerge, R.W., 1994. Empirical threshold values for quantitative trait mapping. Genetics 138, 963-971.

Connors, B.M., Lagasse, C., Dill, L.M., 2011. What's love got to do with it? Ontogenetic changes in drivers of dispersal in a marine ectoparasite. Behavioral Ecology 22, 588-593.

Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.S., Flint-Garcia, S.A., 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiology 158, 824-834.

Costello, M.J., 2008. The global economic cost of sea lice to the salmonid farming industry. Journal of Fish Diseases 32, 115-118.

Danzmann, R.G., Davidson, E.A., Ferguson, M.M., Gharbi, K., Koop, B.F., Hoyheim, B., Lien,

S., Lubieniecki, K.P., Moghadam, H.K., Park, J., Phillips, R.B., Davidson, W.S., 2008. Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). BMC Genomics 9, 557.

Davidson, W.S., Birt, T.P., Green, J.M., 1988. A review of genetic variation in Atlantic salmon, *Salmo salar L.*, and its importance for stock identification, enhancement programmes and aquaculture. Journal of Fish Biology 34, 547-560.

Dekkers, J.C.M., 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. Journal of Animal Science 82, 313-328.

Evans, D.H., Claiborne, J.B., 2006. The Physiology of Fishes. CRC Press Taylor and Francis Group: Boca Raton, Florida, USA.

Falconer, D.S., 1981. Introduction to quantitative genetics, second edition. Longman House: Burnt Mill, Harlow, Essex, UK.

Garrick, D.J., Taylor, J.F., Fernando, R.L., 2011. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution 41, 55.

Gharbi, K., Glover, K.A., Stone, L.C., MacDonald, E.S., Matthews, L., Grimholt, U., Stear, M.J., 2009. Genetic dissection of MHC-associated susceptibility to *Lepeophtheirus salmonis* in Atlantic salmon. BMC Genetics 10, 20.

Gidskehaug, L., Kent, M., Hayes, B.J., and Lien, S., 2011. Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. Bioinformatics 27, 303-310.

Gilbey, J., Verspoor, E., Mo, T.A., Sterud, E., Olstad, K., Hytterød, S., Jones, C., Noble, L.,

2006. Identification of genetic markers associated with *Gyrodactylus salaris* resistance in Atlantic salmon *Salmo salar*. Diseases of Aquatic Organisms 71, 119-129.

Gilmour, A.R., Gogel, B.J., Cullins, B.R., Thompson, R., 2006. ASReml user guide release 2.0. VSN International Limited: Hemel Hempstead, UK.

Gjerde, B., Ødegård, J., Thorland, I., 2011. Estimates of genetic variation in the susceptibility of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. Aquaculture 314, 66-72.

Glover, K.A., Grimholt, U., Bakke, H.G., Nilsen, F., Storset, A., Skaala, O., 2007. Major histocompatibility complex (MHC) variation and susceptibility to the sea louse *Lepeophtheirus salmonis* in Atlantic salmon *Salmo salar*. Diseases of Aquatic Organisms, 76, 57-65.

Goddard, M., 2009. Genomic selection: prediction of accuracy and maximization of long term response. Genetica 136, 245-257.

Goddard, M. E., Hayes, B.J., 2007. Genomic selection. Journal of Animal Breeding and Genetics 124, 323-330.

Gu, X., Frankowski, R.F., Rosner, G.L., Relling, M., Peng, B., Amos, C., 2010. A modified forward multiple regression in high-density genome-wide association studies for complex traits. Genetic Epidemiology 33, 518-525.

Hartley, S.E., 1987. The chromosomes of salmonid fishes. Biological Reviews 62, 197-214.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92, 433-443.

Hayes, B., Goddard, M.E., 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genetic Selection Evolution 33, 209-229.

Hayward C.J., Andrews M., Nowak B.F., 2011. Introduction: *Lepeoptheirus salmonis* – a

    remarkable success story. In Jones, S., Beamish, R., (Eds.). Salmon lice: An integrated

    approach to understanding parasite abundance and distribution. John Wiley & Sons Ltd.:

    Chichester, UK.

Hurvich, C.M., Tsai, C., 1989. Regression and time series model selection in small samples.

    Biometrika 76, 297-307.

Johnson, S.C., Albright, L.J., 1991. Development, growth, and survival of *Lepeophtheirus*

    salmonis (Copepoda: Caligidae) under laboratory conditions. Journal of the Marine

    Biological Association of the United Kingdom 71, 425-436.

Jones, P.G., Hammell, K.L., Dohoo, I.R., Revie, C.W., 2012. Effectiveness of emanmectin

    benzoate for treatment of *Lepeophtheirus salmonis* on farmed Atlantic salmon *Salmo*

    salar in the Bay of Fundy, Canada. Diseases of Aquatic Organisms 102, 53-64.

Kindt, T.J., Goldsby, R.A., Osborne, B.A., 2007. Kuby Immunology: Sixth edition. W.H.

    Freeman and Company: New York.

Hvidsten, N.A., Heggberget, T.G., Jensen, A.J., 1998. Sea water temperature at Atlantic salmon

    smolt enterance. Nordic Journal of Freshwater Research 74, 79–86.

Klemetsen, A., Amundsen, P.A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F.,

    Mortensen, E., 2003. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and

    Arctic charr *Salvelinus alpinus* (L.): A review of aspects of their life histories. Ecology

    of Freshwater Fish 12, 1–59.

Kolstad, K., Heuch, P.A., Gjerde, B., Gjedrem, T., Salte, R., 2005. Genetic variation in

    resistance of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*.

    Aquaculture 247, 145-151.

Lande, R., Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124, 743-756.

Lien, S., Gidskehaug, L., Moen, T., Hayes, B.J., Berg, P.R., Davidson, W.S., Omholt, S.W., Kent, M.P., 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics 12, 615.

Lynch, M., Walsh B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Incorporated: Sunderland, MA, USA.

Moen, T., Baranski, M., Sonesson, A.K., Kjøglum, S., 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. BMC Genomics 10, 368.

Moen, T., Sonesson, A.K., Hayes, B., Lien, S., Munck, H., Meuwissen, T.H.E., 2007. Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. BMC Genetics 8, 53.

Mrode, R. A., Thompson, R., 2005. Linear models for the prediction of animal breeding values. CABI Publishers: Wallingford, UK.

Muir, W.M., 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. Journal of Animal Breeding and Genetics 124, 342-355.

Mustafa, A., Rankaduwa, W., Campbell, P., 2000. Estimating the cost of sea lice to salmon aquaculture in Eastern Canada. Canadian Veterinary Journal 41, 54-56.

Neter, J., W. Wasserman, and M. H. Kutner. 1990. Applied linear statistical models: regression,

analysis of variance and experimental design, third edition. Irwin: Homewood, IL, USA.

Nielsen, H.M., Sonesson, A.K., Yazdi, H., Meuwissen, T.H.E., 2009. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. Aquaculture 289, 259-264.

Nylund, A., Hovland, T., Hodneland, K., Nilsen, F., Løvik, P., 1994. Mechanisms for transmission of infectious salmon anaemia (ISA). Diseases of Aquatic Organisms 19, 95-100.

Pike, A.W., Wadsworth, S.L., 1999. Sealice on salmonids: Their biology and control. Advances in Parasitology 44, 233-337.

Price, T., Schluter, D., 1991. On the low heritability of life-history traits. Evolution 45, 853-861.

Quinton, C.D., 2005. Genetic improvement of Atlantic salmon for New Brunswick aquaculture. Ph.D. Thesis. University of Guelph: Guelph, ON, Canada.

Reid, D.P., Szanto, A., Glebe, B., Danzmann, R.G., Ferguson, M.M., 2004. QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). Heredity 94, 166-172.

Ritchie, G., 1997. The host transfer ability of *Lepeophtheirus salmonis* (Copepoda: Caligidae) from farmed Atlantic salmon, *Salmo salar L.* Journal of Fish Diseases 20, 153-157.

Roberts, F.L., 1970. Atlantic salmon (*Salmo salar*) chromosomes and speciation. Transactions of the American Fisheries Society. 99, 105-111.

SalmoBreed AS, nd. Fish with increased resistance against salmon lice infestation. SalmoBreed retrieved 11 March 2013 from http://www.salmobreed.no/en/products/salmobreed-exclusive/sb-exclusive-lice.

Schaeffer, L.R., 1984. Sire and cow evaluation under multiple trait models. Journal of Dairy Science 67, 1567-1580.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123, 218-223.

Shrimpton, A.E., Robertson, A., 1988. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics 118, 445-459.

Snustad, D.P., Simmons, M.J., 2006. Principles of Genetics: Fourth Edition. John Wiley and Sons, Incorporated: Hoboken, NJ, USA.

CBC News, 2012. Sobeys confirms some recalled salmon had sea lice. Canadian Broadcasting Corporation retrieved 9 March 2013 from http://www.cbc.ca/news/canada/nova-scotia/story/2012/10/22/ns-fish-farming-debate.html.

Sokal, R.R., Rohlf, F.J., 1995. Biometry third edition. W.H. Freeman and Company: New York.

Sonesson, A.K., Meuwissen, T.H.E., 2009. Testing strategies for genomic selection in aquaculture for breeding programs. Genetic Selection Evolution 41, 37.

Surprenant, D., 2010. Aquaculture in Eastern Canada. Parliament of Canada retrieved 31 January 2013 from http://www.parl.gc.ca/Content/LOP/ResearchPublications/2010-13-e.htm#a8.

Taylor, J.F., 2013. Implementation and accuracy of genomic selection, Aquaculture, http://dx.doi.org/10.1016/j.aquaculture.2013.02.017.

Thorstad, E.B., Whoriskey, F., Rikardsen, A.H., Aarestrup, K., 2011. Aquatic Nomads: The Life and Migrations of the Atlantic Salmon. In Aas, Ø.,Einum, S., Klemetsen, A., Skurdal, J. (Eds.). Atlantic Salmon Ecology. Wiley-Blackwell: West Sussex, UK.

Thorstad, E.B., Whoriskey, F., Uglem, I., Moore, A., Rikardsen, A.H., Finstad, B., 2012. A

    critical life stage of the Atlantic salmon *Salmo salar*: behavior and survival during the

    smolt and initial post-smolt migration. Journal of Fish Biology 81, 500-542.

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R.,

    McMullen, M.D., Holland, J.B., Buckler, E.S., 2011. Genome-wide association study of

    lead architecture in the maize nested association mapping population. Nature Genetics

    43, 159-164.

Tizard, I.R., 2009. Veterinary immunology, eighth edition. Elsevier Incorporated: St. Louis,

    Missouri, USA.

Wang, J., 2012. Computationally efficient sibship and parentage assignment from multilocus

    marker data. Genetics 191, 183-194.

Table 1: The schedule of staggered salmon lice infections in November and December 2012 and the dates of salmon euthanasia and lice counting

Date infected	Date counted	Tank
5 November 2012	13 November 2012	8
5 November 2012	20 November 2012	8
6 November 2012	21 November 2012	7
7 November 2012	22 November 2012	6
8 November 2012	23 November 2012	5
16 November 2012	29 November 2012 and 30 November 2012	1
19 November 2012	3 December 2012 and 4 December 2012	2
22 November 2012	6 December 2012	7
23 November 2012	7 December 2012	8
26 November 2012	10 December 2012	7
28 November 2012	12 December 2012	5 and 6

The schedule for Atlantic salmon challenged with salmon lice in 2012. Tank, date when fish were challenged as well as the date fish were euthanized and lice counts were recorded are included.

Table 2: Results for an Illumina iSelect™ array that was custom designed for European Atlantic salmon tested in a North American aquacultural population of Saint John River Atlantic salmon.

Classification		Number
Fail		11
Monomorphic		502
Multisite variant (MSV)		573
	MSV3 452	
	MSV4 59	
	MSV5 62	
Single nucleotide polymorphism (SNP)		4467
Unknown		12
Total		5565

539 fish from a North American Atlantic salmon population were genotyped using an array panel designed for European Atlantic salmon. SNPs and MSVs demonstrate polymorphism and are useful. The monomorphic classification was for alleles that are no polymorphic in our population and the fail and unknown classifications were alleles that could not be classified and therefore not useful.

Table 3: Summary statistics for fresh water weights for the second generation and salt water weights, transformed salt water weights, salmon lice counts and transformed salmon lice counts for the third generation

	Fresh water weight (kg)	Salt water weight (g)	ln(Salt water weight)	Salmon lice count	√Salmon lice count
N	3588	1803	1803	1733	1733
Mean	3.41	375.67	5.74	32.29	5.22
Median	3.06	340.90	5.83	26	5.10
Standard deviation	1.34	221.05	0.64	26.82	2.25
Skewness	1.25	0.79	-0.33	2.00	0.43
Kurtosis	1.14	-0.09	-0.40	7.29	0.64
Range	1.06-9.00	29.1-1177.5	3.37-7.07	0-269	0.00-16.40

Table 4a: Phenotypic correlations, above diagonal, and genetic correlations, below diagonal, between fresh water weight, transformed salt water weight and transformed salmon lice count and estimated heritabilies, on diagonal, for each trait (associated standard errors in brackets)

	Fresh water weight (kg)	ln(Salt water weight)	√Salmon lice count
Fresh water weight (kg)	0.60 (0.06)	-0.04 (0.08)	0.03 (0.08)
ln(Salt water weight)	-0.08 (0.18)	0.34 (0.06)	0.08 (0.03)
√Salmon lice count	0.08 (0.19)	0.00 (0.17)	0.29 (0.05)

Table 4b: Phenotypic correlations, above diagonal, and genetic correlations, below diagonal, between fresh water weight, salt water weight and salmon lice count and estimated heritabilies, on diagonal, for each trait (associated standard errors in brackets)

	Fresh water weight (kg)	Salt water weight (g)	Salmon lice count
Fresh water weight (kg)	0.60 (0.06)	0.08 (0.10)	0.00 (0.07)
Salt water weight (g)	0.14 (0.17)	0.54 (0.07)	-0.03 (0.04)
Salmon lice count	0.01 (0.22)	-0.14 (0.18)	0.17 (0.04)

Table 5: Accuracy of estimated breeding values (EBVs) for salmon lice count for the multiple trait model with transformed salt water weight and transformed salmon lice count and for the multiple trait model for fresh water weight, transformed salt water weight and transformed salmon lice count for the second generation of fish that were genotyped (N=299)

	Multiple trait model with 2 traits	Multiple trait model with 3 traits
Mean	0.50	0.57
Standard deviation	0.14	0.15
Minimum value	0.07	0.10
Maximum value	0.76	0.82

Table 6: Single nucleotide polymorphism (SNP) name, chromosome and partial R^2 of first 20 SNPs entered in the forward regression along with results from blasting (gene and E value)

Step entered	Name	Gene	E	Chromosome	Partial R^2
1	ESTNV_24376_271	*Salmo salar* clone HM4_0644 member RAS oncogene family (rab5a1) mRNA, complete cds	2E-114	2	0.00
2	GCR_cBin8670_Ctg1_184	*Oncorhynchus mykiss* immunoglobulin heavy chain constant region (ighd) gene, partial cds	3.00E-77	7	0.12
3	ESTNV_36936_1258	No DNA sequence available yet		15	0.29
4	GCR_cBin27009_Ctg1_308	*Salmo salar* clone 272P16 chaperonin gene, complete cds; myosin 1 gene, partial cds; and TCR-gamma locus, partial sequence	2E-11	9	0.11
5	ESTNV_26999_214	PREDICTED: hypothetical protein LOC100149633 [*Danio rerio*]	4E-29	22	0.08
6	ESTV_16677_620	*Salmo salar* osteoclast stimulating factor 1 (ostf1), mRNA	0	24	0.38
7	GCR_cBin19665_Ctg1_346_V2			14	0.00
8	GCR_cBin10989_Ctg1_237			1/23	0.03
9	BASS113_B6A_E01_397	*Salmo salar* clone BHMS7-057 microsatellite sequence	3E-19	12	0.08
10	GCR_cBin11486_Ctg1_236	*Salmo salar* neurogranin, TIP41-like protein (TIP41), MHC class II antigen beta chain (Sasa-DBB), MHC class II antigen alpha chain (Sasa-DBA), leucine rich repeat containing 35-like protein, and alpha-tectorin-like	2E-21	21	0.08

		protein genes, complete cds			
11	GCR_cBin37279_Ctg1_86			9	0.16
12	ESTNV_27066_1033	Apoptosis inhibitor 5 [*Salmo salar*]	6.00E-34	10	0.03
13	ESTNV_20525_311	*Salmo salar* clone ssal-eve-540-327 FAM49B putative mRNA, complete cds	3E-164	8/29	0.13
14	GCR_cBin36852_Ctg1_144	*Salmo salar* retinoic acid receptor gamma b (Rargb) gene, partial cds; and homeobox proteins HoxC13ba, HoxC12ba, HoxC11ba, HoxC10ba, HoxC9ba, HoxC8ba, HoxC6ba, HoxC5ba, and HoxC4ba genes, complete cds	3E-17	18	0.19
15	GCR_cBin17883_Ctg1_169	Mouse DNA sequence from clone RP23-156E17 on chromosome 2, complete sequence	2E-04	24	0.01
16	ESTV_20585_205	PREDICTED: (Japanese medaka) *Oryzias latipes* N-acetylgalactosamine-6-sulfatase-like, transcript variant 2 (LOC101171730), mRNA	1E-36	8/29	0.16
17	ESTNV_31115_754			15	0.06
18	ESTV_17011_96	40S ribosomal protein S21 [*Salmo salar*]	2.00E-63	15	0.14
19	ESTNV_34665_1373	Cbp/p300-interacting transactivator 2 [*Salmo salar*]	1.00E-124	14	0.12
20	ESTNV_36823_515	cytochrome c-like [*Salmo salar*]	5.00E-67	21	0.17

The first 20 SNPs found to be associated with salmon lice count in the forward regression were blasted to find the genes they are found in.

Table 7: Average allele substitution effect estimated for parents and relatives of the 2011 and 2012 year classes that were genotyped and paired t-test result for the allele substitution effect difference between the two year classes

	Number of parents genotyped	Number of other relatives genotyped	Average allele substitution effect
2011	75	98	0.04
2012	80	46	-0.11
Difference			0.15 (P=0.058)

Table 8: Previous QTL detection work and locations of QTL mapped to chromosomes for health and production traits for Atlantic salmon

Trait	Chromosome		
Salmon lice count	12	12.9% of within family variation	(Gharbi et al., 2008)
Gyrodactylus salaris **count**	10, 13, 24, 3, 23, 4, 15, 17, 16, and 12	27.3% of phenotypic variation	(Gilbey et al., 2006)
Infectious pancreatic necrosis (IPN)	26	70% of genetic variation	(Moen et al., 2009)
	6	0.9% of genetic variation	(Moen et al., 2009)
Infectious salmon anemia virus (ISA)	15	6% of phenotypic variation	(Moen et al., 2007)
Body weight	15	20.1% of phenotypic variation	(Reid et al., 2005)
	3	12.4% or 11.7% of phenotypic variation	(Reid et al., 2005)
Body condition score	21	24.9% of phenotypic variation	(Reid et al., 2005)
	3	16.9% of phenotypic variation	(Reid et al., 2005)
	13	17.6% of phenotypic variation	(Reid et al., 2005)
	10	17.6% of phenotypic variation	(Reid et al., 2005)

Figure 1: Methodology
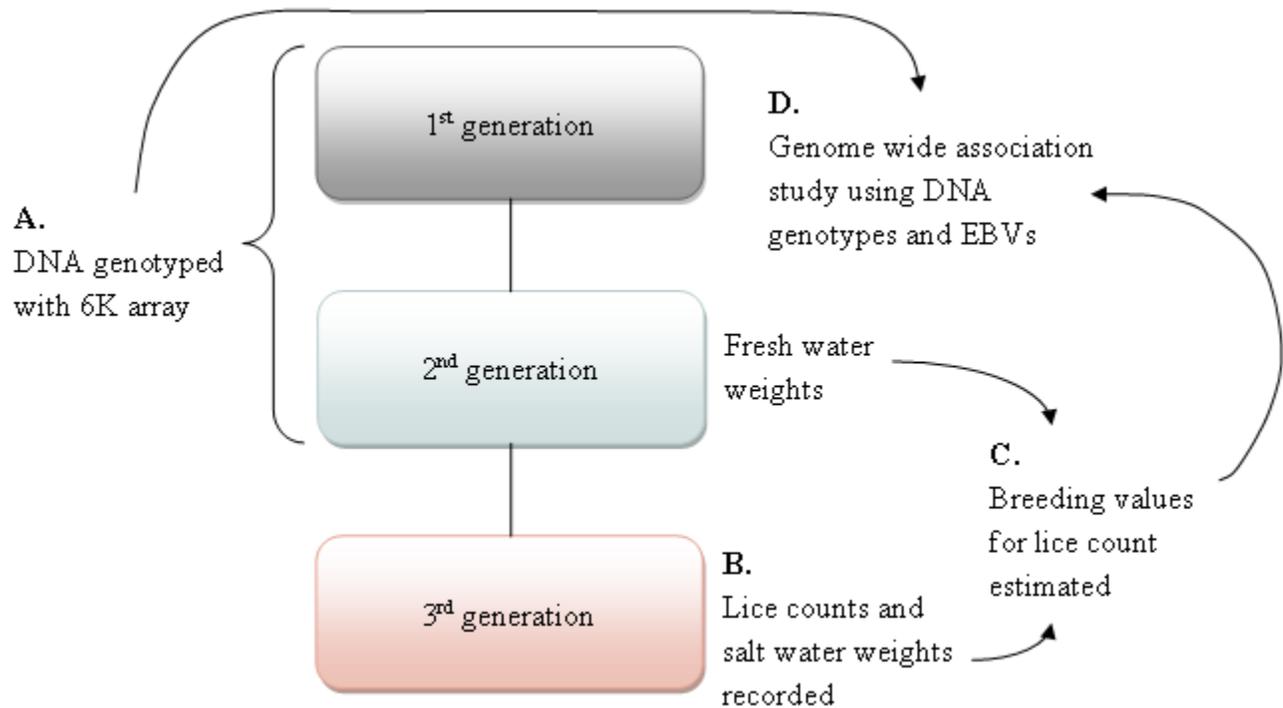
Using three generations from an aquaucultural population of Atlantic salmon, with known pedigree:

A. 240 of 436 fish in the first generation and 299 of 5588 fish in the second generation were genotyped.

B. All fish in the third generation (N=1803) were challenged with salmon lice and had salmon lice counts and salt water weight recorded.

C. Breeding values for lice count were estimated using a multiple trait model that used fresh water weight records from 3588 of the fish in the second generation and lice counts and salt water weights from the third generation.

D. Using the 299 fish in the second generation that were genotyped and their estimated breeding values for salmon lice count, associations were detected.

Figure 2a: Salt water weight mean and variance for contemporary group by sex subclasses
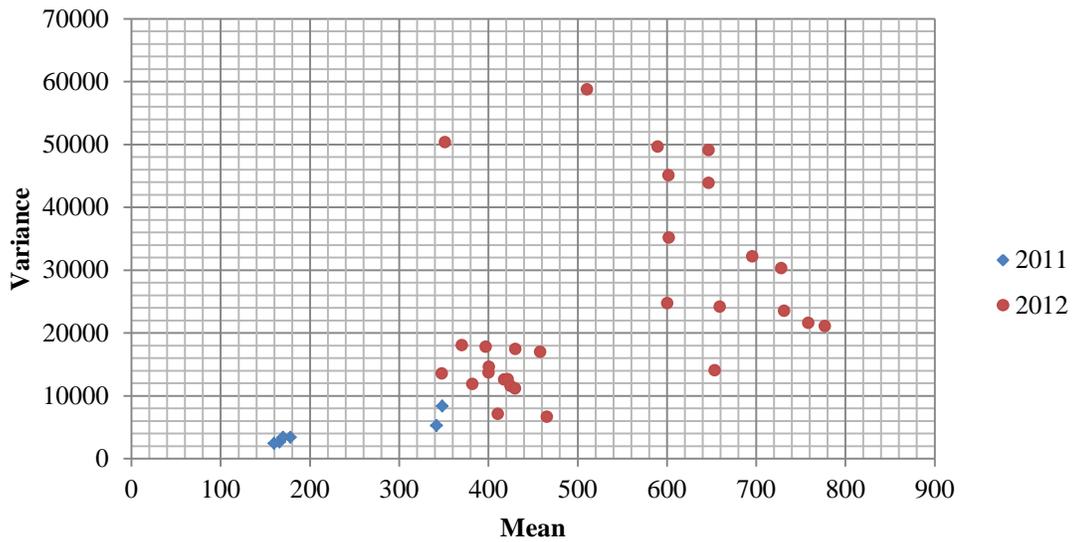
Figure 2b: Natural logarithm transformed salt water weight mean and variance for contemporary group by sex subclasses
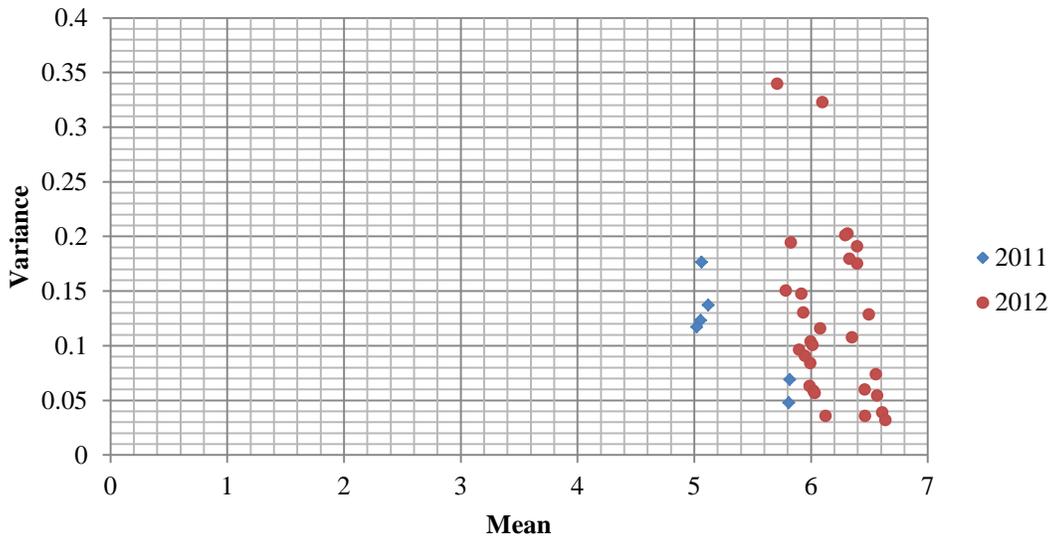
Figure 2 shows the relationship between contemporary group by sex subclasses salt water weight means and variances for the study population of Atlantic salmon challenged in 2011 and 2012. Figure 2a shows a positive relationship while Figure 2b salt water weights have been transformed and no longer have a positive relationship.

Figure 3a: Salmon lice counts mean and variance for contemporary group by sex subclasses
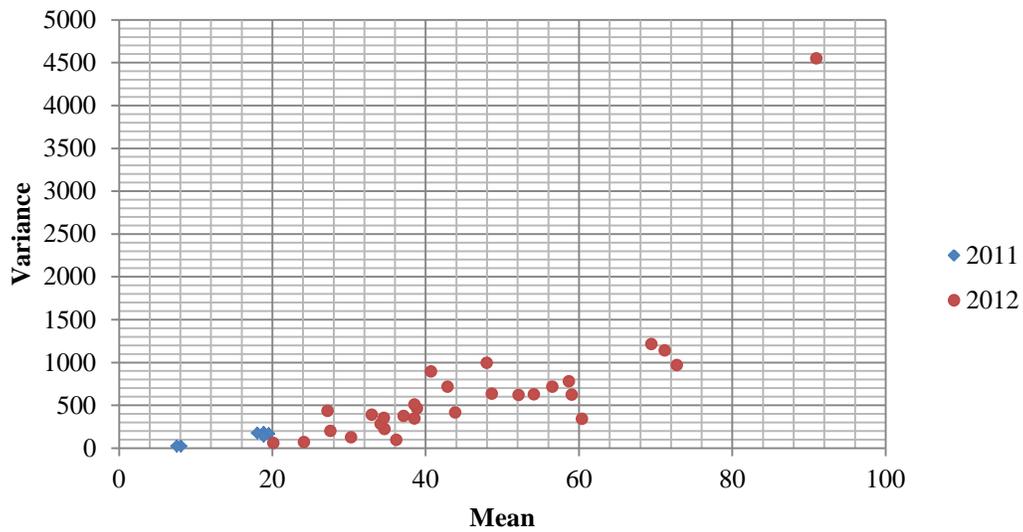
Figure 3b: Square root transformed salmon lice counts mean and variance for contemporary group by sex subclasses
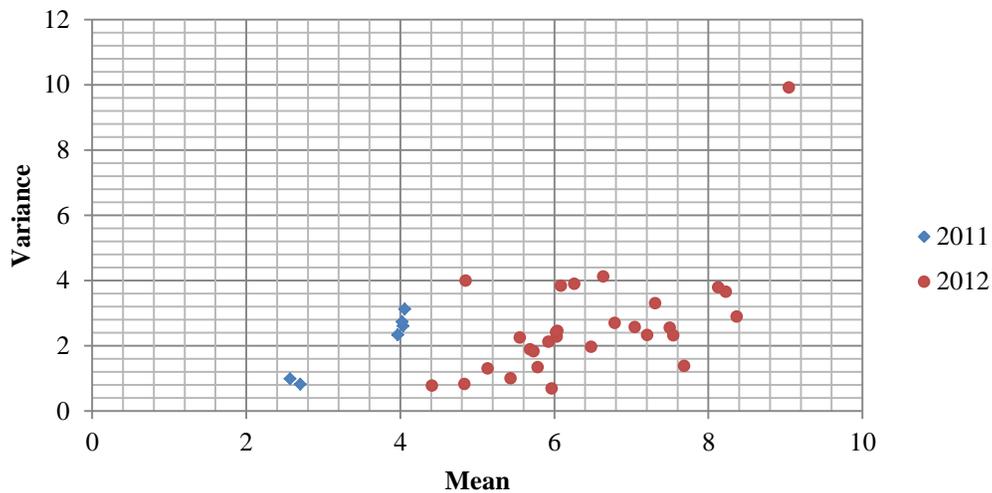
Figure 3 shows the relationship between contemporary group by sex subclasses lice count means and variances for the study population of Atlantic salmon challenged in 2011 and 2012. Figure 2a shows a positive relationship while Figure 2b lice counts have been transformed and no longer have a positive relationship.

Figure 4: Distribution of accuracies of estimated breeding values for transformed salmon lice count of the 299 fish from the second generation of fish that were genotyped

- no offspring, nieces/nephews or half nieces/nephews challenged
- half nieces/nephews challenged
- nieces/nephews challenged
- offspring challenged
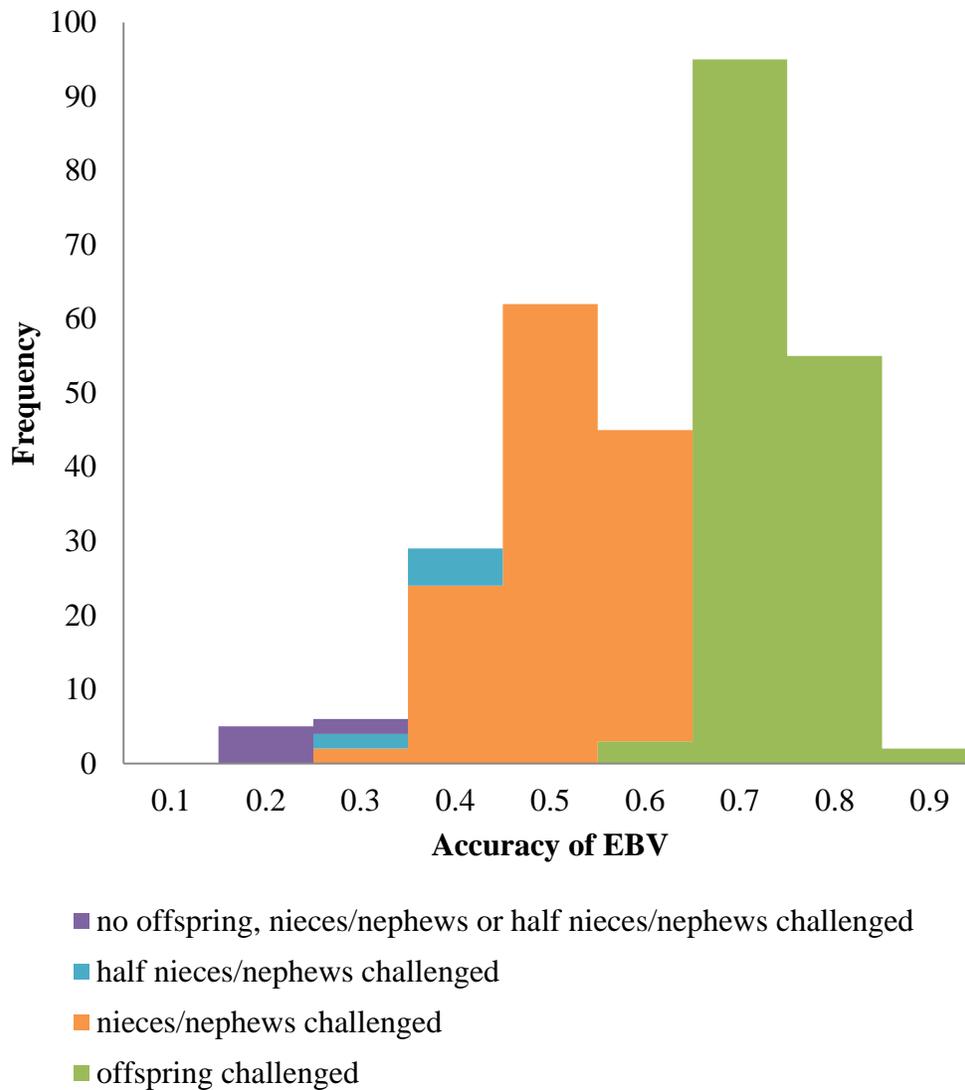
Estimated breeding value (EBV) accuracies for the 299 fish in the second generation that were genotyped separated into four groups which indicate the closest relationship they had to challenged fish.

Figure 5: Manhattan plot of the 70 SNPs associated with salmon lice count
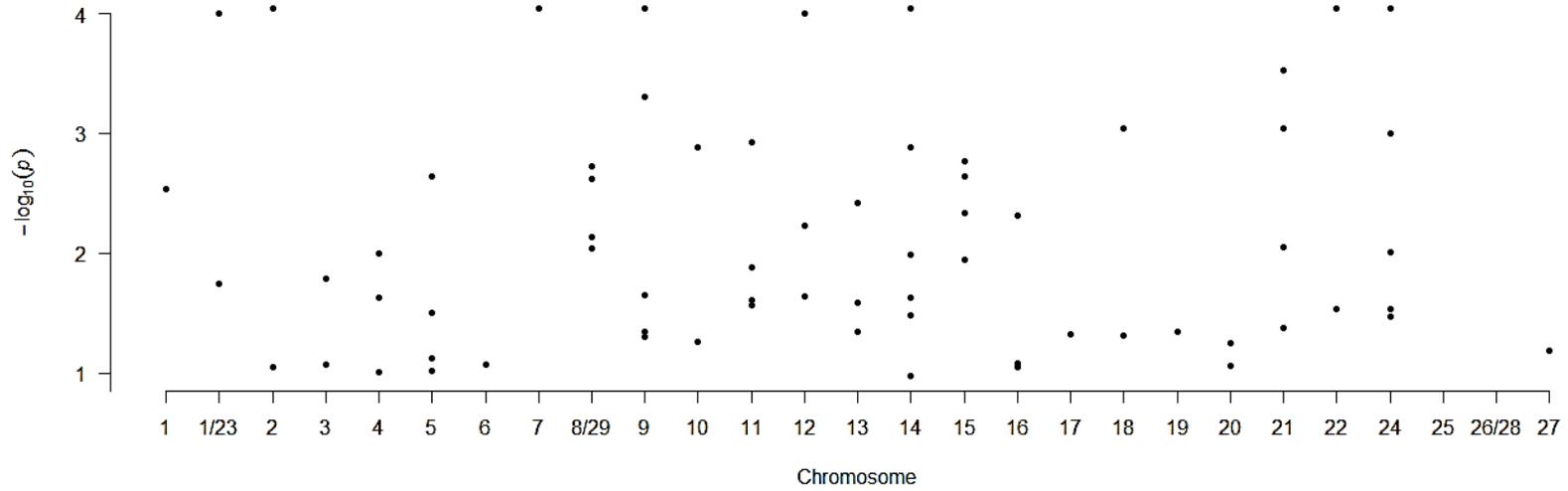
70 SNPs associated with salmon lice count by chromosome number (x axis) and by P-value from the forward regression for SNP to enter the model.

Figure 6: Back transformed allele substitution effects of the 70 SNPs associated with salmon lice count
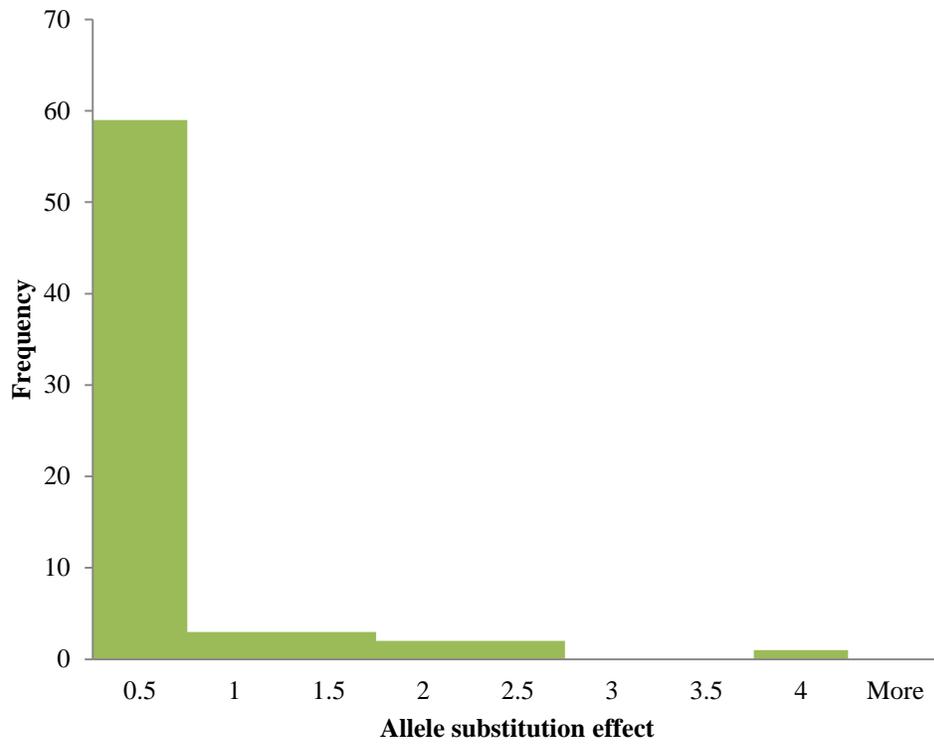
Frequencies of back transformed allele substitution effects estimated using a multiple linear regression in SAS.

Figure 7: Partial coefficient of determination (partial R^2) for the 70 SNPs associated with salmon lice count
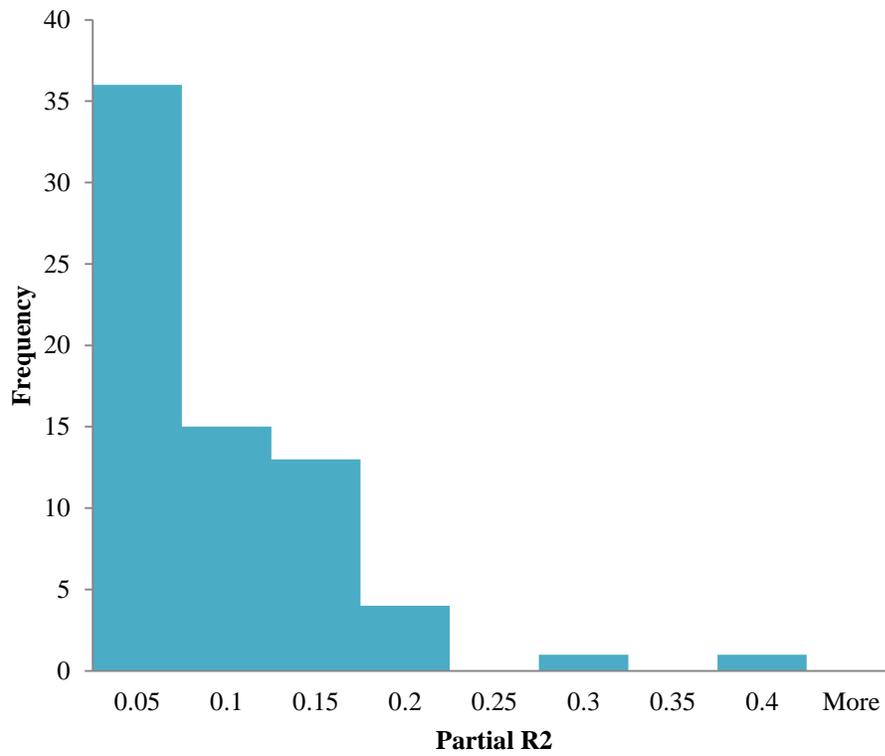
Frequencies of partial R^2 for the SNPs associated with salmon lice count estimated using SAS.

Appendix A i: PROC GLM results for fresh water weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	219	5036.632	22.99832	54.12	<.0001
Error	3368	1431.117	0.424916		
Corrected Total	3587	6467.749			

R-Square	Coeff Var	Root MSE	fwwt Mean
0.77873	19.1164	0.651856	3.40993

Source	DF	Type I SS	Mean Square	F Value	Pr>F
fam	213	4886.581	22.94169	53.99	<.0001
CG*sex	6	150.0517	25.00862	58.86	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr>F
fam	212	518.6006	2.446229	5.76	<.0001
CG*sex	6	150.0517	25.00862	58.86	<.0001

Appendix A ii: PROC GLM results for salt water weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	122	69120021	566557.6	50.28	<.0001
Error	1680	18930557	11268.19		
Corrected Total	1802	88050579			

R-Square	Coeff Var	Root MSE	fwwt Mean
0.785004	28.25639	106.1517	375.6733

Source	DF	Type I SS	Mean Square	F Value	Pr>F
dam	88	46779078	531580.4	47.18	<.0001
CG*sex	34	22340943	657086.6	58.31	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr>F
dam	87	4575765	52595	4.67	<.0001
CG*sex	34	22340943	657086.6	58.31	<.0001

Appendix A iii: PROC GLM results for salmon lice count

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	122	675704.6	5538.563	15.65	<.0001
Error	1610	569759.1	353.888		
Corrected Total	1732	1245464			

R-Square	Coeff Var	Root MSE	fwwt Mean
0.542533	58.26189	18.8119	32.28852

Source	DF	Type I SS	Mean Square	F Value	Pr>F
dam	88	471715.6	5360.405	15.15	<.0001
CG*sex	34	203989	5999.677	16.95	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr>F
dam	87	78164.33	898.4406	2.54	<.0001
CG*sex	34	203989	5999.677	16.95	<.0001