Applications of low and high density SNP arrays to Atlantic salmon breeding: Parameters affecting parentage assignment accuracy and mapping quantitative trait loci for infectious salmon anemia resistance

by

Forest Michel Dussault

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
© Forest Michel Dussault, September, 2016
Applications of single nucleotide polymorphisms (SNPs) to the Atlantic salmon (Salmo salar) aquaculture industry can reduce genotyping costs while increasing sample throughput, thereby benefiting commercial producers. Two applications were investigated: 1) The effect of total number of SNPs and minor allele frequency (MAF) on parentage assignment accuracy and 2) Scanning for infectious salmon anemia (ISA) resistance quantitative trait loci (QTL) in a commercial strain of Atlantic salmon using a 50K SNP assay. 

1) Results confirmed minimum criteria necessary for a parentage assignment rate of >98% when considering MAF and SNP number; when average MAF = 0.30, a minimum of 60 SNPs are necessary, when average MAF = 0.10, a minimum of 150 SNPs are necessary. 2) Following a scan using a female linkage map created for this experiment, a total of six QTL were detected across six chromosomes in the Saint John River strain of Atlantic salmon for ISA resistance.
ACKNOWLEDGEMENTS

I’d like to thank my advisor Dr. Elizabeth Boulding as well as the members of my advisory committee, Dr. Roselynn Stevenson and Dr. Brian Glebe. Their expertise and support was much appreciated in this project.

This project was made possible through funding provided by the Genomic Applications Partnership Program (GAPP) available through Genome Canada and Genome Atlantic. This research, as well as my internship at Cooke Aquaculture, were partially funded through the MITACS Accelerate Program. I’d like to thank all members of the GAPP project, including Dr. Keng Pee Ang, Dr. J.A.K. Elliott, Frank Powell, Erica Harvey, Melissa Holborn, Dr. Jane Tosh, Dr. Brian Glebe, and Dr. Larry Schaeffer.

Thank you to all parties involved in the 2011 ISA disease challenge: Dr. Anthony J. Manning (RPC) for his role in the experimental design, protocol development, data management, and project management in collaboration with DFO SABS, Novartis, and Cooke Aquaculture Inc. Steven Leadbeater (DFO SABS) for his role in animal husbandry, disease challenge oversight, data collection, fin clip collection, and biosecurity and animal care compliance responsibilities. Dr. Brian Glebe for activity oversight and management, project authorization, protocol development, and communication with the Canadian Food Inspection Agency (CFIA) in meeting biosecurity requirements. Finally, Novartis Animal Health staff for their role in management and funding of the project that facilitated the ISA disease challenges.

This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.

Finally, I’d like to thank my friends, colleagues, and family for their support through this project.
TABLE OF CONTENTS

ABSTRACT ..................................................................................................................... ii
ACKNOWLEDGEMENTS ................................................................................................... iii
LIST OF FIGURES ........................................................................................................ vi
LIST OF TABLES ........................................................................................................... vii
LIST OF APPENDICES .................................................................................................. viii

CHAPTER I: GENERAL INTRODUCTION ...................................................................... 1
  1.1: Aquaculture of Atlantic salmon ............................................................................. 1
  1.2: Atlantic salmon genetics and genomic applications ............................................ 3
  1.3: General questions and goals .............................................................................. 4
  1.4: References ........................................................................................................... 6

CHAPTER II: THE EFFECT OF MINOR ALLELE FREQUENCY ON PARENTAGE
ASSIGNMENT ACCURACY IN FARMED ATLANTIC SALMON (Salmo salar) ............... 9
  2.1: Introduction ......................................................................................................... 9
  2.2. Materials and methods ....................................................................................... 11
      2.2.1. Animals and genotyping ............................................................................ 11
      2.2.2. Filtering and selecting SNPs .................................................................. 12
      2.2.3. Pedigree verification ............................................................................. 13
      2.2.4. Simulations .......................................................................................... 14
      2.2.5. Parentage assignment with empirical data ............................................. 14
  2.3. Results ................................................................................................................ 15
      2.3.1. Properties of SNP markers .................................................................. 15
      2.3.2. Simulations and empirical results ......................................................... 16
  2.4. Discussion ......................................................................................................... 17
  2.5. Figures and Tables ............................................................................................ 19
  2.6. References ......................................................................................................... 27

CHAPTER III: MAPPING QUANTITATIVE TRAIT LOCI FOR INFECTIOUS SALMON
ANEMIA (ISA) RESISTANCE IN A NORTH AMERICAN STRAIN OF ATLANTIC SALMON 30
  3.1. Introduction ........................................................................................................ 30
      3.1.1. Background .......................................................................................... 30
      3.1.2. Infectious salmon anemia virus (ISAV) .................................................. 31
      3.1.3. Selection on Atlantic salmon ................................................................ 32
      3.1.4. Mapping ISA-resistance to the Atlantic salmon genome ..................... 34
      3.1.5. Questions and goals ............................................................................. 35
  3.2. Materials and Methods ...................................................................................... 36
CHAPTER IV: CONCLUSIONS

4.1. The utility of low cost SNP assays

4.2. The potential benefits of SNPs in aquaculture

4.3. References

APPENDIX A

APPENDIX B
LIST OF FIGURES

Figure 2.1: Simulations showing relationship between the number of SNPs and the number of simulated ‘strict assignments’ assigned by CERVUS across five levels of average minor allele frequency in three families of Atlantic salmon ................................................................. 19

Figure 2.2: Empirical results showing relationship between the number of SNPs and the proportion of correctly assigned Atlantic salmon offspring by CERVUS to a validated pedigree across five levels of average minor allele frequency using an empirical data set across three populations ............................................................................................................ 20

Figure 2.3: Empirical results showing the proportion of Atlantic salmon parents incorrectly assigned by each year class, with number of SNPs and minor allele frequency group both indicated .......................................................................................................................... 21

Figure 2.4: Relationship between q, the minor allele frequency (MAF), and the empirical value of expected heterozygosity (2pq) in three populations of Atlantic salmon (2006, 2007, 2008) .................................................. 23

Figure 3.1: Survivorship curve for the infectious salmon anemia (ISA) cohabitation disease challenge conducted on Atlantic salmon used for QTL mapping experiment ........................................... 45

Figure 3.2: Response to infectious salmon anemia (ISA) cohabitation disease challenge by Saint John River (SJR) strain Atlantic salmon individuals from 10 different full-sibling families.......................... 46

Figure 3.3: Sex-specific linkage maps for a) female (composite for families F02 and F03) and b) male (family F02) Atlantic salmon generated using OneMap software and merged with LPmerge .......................................................................................................................... 47

Figure 3.4: Significant ISA QTL graphs generated in GridQTL detected at the chromosome-wide level in Atlantic salmon .................................................................................................................................. 48
LIST OF TABLES

Table 2.1: Numbers of Atlantic salmon available in empirical dataset with parentage validated with microsatellites used to evaluate the effect of minor allele frequency (MAF) and number of SNPs on parentage assignment accuracy. ................................................................. 24

Table 2.2: Estimated average MAF for each SNP group across the 2006YC, 2007YC, and 2008YC populations of Atlantic salmon. ........................................................................................................ 25

Table 2.3: Expected heterozygosity for each SNP subset for each Atlantic salmon population calculated in CERVUS ............................................................................................................................................. 26

Table 3.1: ISA Disease trial conducted in 2011 at DFO SABS using 2010 year class Atlantic salmon sourced from Cooke Aquaculture’s Saint John River (SJR) aquacultural strain. .......... 49

Table 3.2: All GridQTL results for ISA-resistance in the Saint John River strain of Atlantic salmon QTL mapping experiment. ........................................................................................................ 50
LIST OF APPENDICES

APPENDIX A: Python script for reformatting tabular genotype data as indicated in the script to a format more amenable to input into CERVUS ........................................................................................................ 61

APPENDIX B: Python written for pulling a specific list of SNP markers specified in a supplementary '.txt' file from a '.csv' file with genotypes listed in tabular format ....................... 63
CHAPTER I: GENERAL INTRODUCTION

1.1: AQUACULTURE OF ATLANTIC SALMON

Atlantic salmon (*Salmo salar*) are native to the subarctic and temperate regions of the North Atlantic Ocean, and have a complex life history that spans across both fresh and marine water environments (Thorstad *et al*., 2012). The aquaculture of Atlantic salmon on a commercial scale is a relatively recent development in the business of animal husbandry, beginning in the 1970s in Norway (Anderson, 1997; Gjøen & Bentsen, 1997). Since its inception, the Atlantic salmon aquaculture industry has experienced massive growth, expanding from Norway to Canada, Scotland, Chile, the Faroe Islands, and others. Farmed Atlantic salmon now contributes to over 50% of the global salmon market, with over 2.3 million tonnes produced in 2014 (FAO, 2014).

The natural life cycle of Atlantic salmon begins in a freshwater environment between September and February, where females lay their eggs on the gravel bottoms of river beds in preparation for external fertilization by males (Thorstad *et al*., 2012). Upon hatching in the spring, juvenile fish known as parr begin to hatch. The juvenile fish remain in the freshwater river system for one to eight years, after which they undergo smoltification in preparation for their journey to the ocean (Klemetsen *et al*., 2003). Salmon that have undergone smoltification are known as smolts, a life stage in which the fish has become physiologically prepared for a marine system. As smolts, the salmon will travel to the ocean where they take between one and five years to mature into adults, eventually returning to spawn in those same river beds in which they were hatched (Aas *et al*., 2010). Unlike some other salmonids, the Atlantic salmon is iteroparous, meaning they have the potential to spawn more than once (Thorstad *et al*., 2012).

Accommodating the life history described above in an aquaculture setting is a complex task, requiring many specialized facilities to rear the fish from egg to adult. The most common approach to the aquaculture of Atlantic salmon is to use ‘sea cages’, which are large enclosures located in the ocean where adults are left to develop for up to two years until harvest. Subsequent generations of aquaculture stock are obtained through the breeding of fish referred to as broodstock, which are sexually mature
adults carefully selected for desirable production traits. Once a set of broodstock has been selected, the females and males are stripped of their roe and milt respectively, and many families (often hundreds) are produced through fertilization under tightly controlled conditions. These fertilized eggs are then disinfected and transferred to hatching trays. Eventually, the eggs hatch, and alevin stage Atlantic salmon emerge with their yolk sacs still attached. The yolk sacs are eventually absorbed, and at this point the salmon have developed to the fry stage. They are then transferred to tanks with conditions designed to emulate those found in the rivers from which they would emerge in a natural setting. From this point, they are left to grow, and eventually smoltify to prepare for their transfer to sea cages, where they will do the bulk of their growing.

With the daunting growth of the Atlantic salmon aquaculture industry, producers seek improvement to their breeding and production programs through many techniques and technologies (Gjedrem, 2012; Mackay et al., 2009). Traditional selective breeding remains the de facto method for increasing simple desirable traits like body weight (Gjøen & Bentsen, 1997). To achieve this, individuals with superior phenotypes are chosen for broodstock and subsequent generations are bred using these fish.

Genetic improvement in livestock is achieved through selective breeding or cross breeding. Artificial insemination and embryo transfer have allowed selective and cross breeding to be administered on a larger scale, with more rapid genetic gains as a result. Many techniques have been developed to evaluate animals and to select the optimal set of animals to achieve gains in a desired trait. Estimated breeding values (EBVs), genomic selection, marker-assisted selection, among other techniques, allow rapid improvement of desired traits through selective breeding (Goddard & Hayes, 2009). Despite the advent of genetic improvement in traditionally farmed animals, uptake of genomic technologies has been slow in aquaculture, with estimates suggesting only 10% of aquaculture production worldwide is based on genetically improved animals (Gjedrem et al, 2012).
1.2: ATLANTIC SALMON GENETICS AND GENOMIC APPLICATIONS

Single-nucleotide polymorphisms (SNPs) are abundant, biallelic molecular markers that have been used extensively in a large number of diverse biological fields. SNPs have proven to be a valuable asset to Atlantic salmon aquaculture operations with numerous applications. Parentage assignment is an important aspect in determining pedigree structure in Atlantic salmon. Parents can now be easily assigned to their offspring using a relatively small number of SNP markers, with many algorithms and software packages now available (Jones et al., 2010). SNPs also have value for the detection of quantitative trait loci (QTL). A QTL is a region of the genome that is associated with a phenotype (i.e. weight, carcass quality, disease resistance) and explains some proportion of the observed phenotypic variation in a population. A helpful and extensive review on the genetics of quantitative traits is available from Mackay et al. (2009). QTL have been mapped for many traits in Atlantic salmon: resistance to pancreas disease (Gonen et al., 2015); body-weight (Reid et al., 2005; Gutierrez et al., 2012); resistance to infectious pancreatic necrosis (Moen et al., 2009); body shape (Boulding et al., 2008); maturation traits (Gutierrez et al., 2014; Pedersen et al., 2013) and many more. Through the mapping of these QTL associated with traits of interest, producers can eventually implement marker-assisted selection programs where animals with favourable genotypes are selected for broodstock (Gjedrem, 2012). Marker-assisted selection has a particular advantage in improving traits that are difficult to traditionally select for, such as flesh colour, meat quality, or disease resistance.

The genome of the common ancestor of the salmonid family is believed to have undergone a whole-genome duplication event 25-100 million years ago, leading to a tetraploid state (Allendorf & Thorgaard, 1984). Since this event, there has been selective pressure leading to the reversion of much of the genome to a diploid state, currently leaving roughly a third of the modern salmonid genome as tetraploid (Gidskehaug et al., 2011). Despite the gradual reversion, this duplication event adds a layer of complexity to QTL detection not typically seen in diploid organisms. This is due to the presence of multisite variants (MSVs), which indicate two copies of a locus with the same set of segregating alleles.
instead of the single locus with two alleles seen in a typical SNP. To compensate for MSVs, SNPs belonging to only the diploid portion of the genome are often considered for QTL detection (Gidskehaug et al., 2011). Paralogous sequence variants (PSVs) also obfuscate the search for SNPs (Fredman et al., 2004). PSVs occur when the sequences of two paralogs have a single nucleotide difference, yet the base substitution does not segregate within either paralog. To compensate for this, putative SNPs that contain no homozygotes in any samples are deemed PSVs, and are often removed from the analysis (Gidskehaug et al., 2011). Atlantic salmon have also undergone significant divergence between North American and European strains as a result of geographical isolation between the populations (Davidson et al., 1989; Bourret et al., 2013). This separation of North American and European strains of Atlantic salmon is an important distinction to make when comparing genomic study results between the two groups, especially in the case of QTL mapping experiments.

1.3: GENERAL QUESTIONS AND GOALS

Designing a SNP assay intended for accurate parentage assignment that is both cost-effective and accurate can be a valuable asset in the aquaculture industry. By analyzing various properties of SNPs on the population level, inferences can be made to evaluate their potential to verify or to establish accurate pedigrees. Chapter II describes a study of the effect of minor allele frequency (MAF) on parentage assignment accuracy. There have been many studies detailing the effect of numbers of SNPs on parentage assignment accuracy, though I wanted to expand on this past work by generating empirical results on the effect of MAF on parentage assignment accuracy, with the final aim of contributing to the design of low-cost SNP arrays in the future. I used three separate generations of Atlantic salmon and their corresponding genotypes derived from a 6000 (6K) SNP array to create a dataset that allowed for my analysis.

Chapter III describes a disease challenge on full-sib families of Atlantic salmon, and the subsequent mapping of infectious salmon anemia-resistance QTL using a dense linkage map composed of SNP markers. Infectious salmon anemia (ISA) is a disease of considerable economic importance to the
Atlantic salmon aquaculture industry, resulting in massive economic losses for producers when an outbreak occurs. As little has been done to identify genetic associations with ISA-resistance, a scan of the Atlantic salmon genome using a dense SNP linkage map could provide valuable evidence for ISA-resistance QTL. Any detected QTL could later be leveraged in breeding programs to select for more resistant Atlantic salmon.
1.4: REFERENCES


CHAPTER II: THE EFFECT OF MINOR ALLELE FREQUENCY ON PARENTAGE ASSIGNMENT ACCURACY IN FARMED ATLANTIC SALMON (SALMO SALAR)

2.1: INTRODUCTION

Single-nucleotide polymorphisms (SNPs) are likely to supersede microsatellites for parentage assignment in the future due to their increasingly low cost per-genotype and amenability to machine scoring (Anderson & Garza, 2006). SNPs are highly abundant in the genome, are amenable to high throughput processing, have lower mutation rates than microsatellites, and allow for easy standardization across laboratories (Anderson & Garza, 2006). SNPs are also typically biallelic in populations, and therefore tend to contain less polymorphic information than microsatellite markers used in parentage assays. This limit on the amount of polymorphic information available is countered by the shear abundance of SNPs within the Atlantic salmon genome, and the low cost per SNP genotype.

Physically tagging animals for later identification in aquaculture operations is often a difficult, or impossible task given the small size of the animals at the time of hatching (Yue & Xia, 2014). A common approach to maintain the veracity of pedigrees in farmed aquatic animals is to rear them in family designated tanks, where animals can eventually be tagged once they reach adequate size (Vandeputte & Haffray, 2014). This method is costly as it requires a large amount of operational floor space, as well as specially designated tanks to isolate families from each other, with large scale breeding programs often supporting hundreds of families at a time. In addition to the cost, the separate rearing of families can lead to environmental effects that could be confounded with genetic effects in later assessments (Falconer & Mackay, 1996). Due to the high cost of this family-based tank setup, and the potential for environmental effects, communal rearing followed by determining parentage using molecular markers is preferable as the cost of genotyping per individual continues to decline.

Parentage assignment using molecular markers is typically conducted using exclusion-based methods or likelihood-based methods. A review by Jones et al. (2010) details parentage methods with
molecular markers extensively. The authors of the review argue that exclusion-based methods are straightforward; given the rules of Mendelian inheritance in diploid organisms, the parent and offspring must share in common at least one allele per locus. If this rule is violated at a particular locus between an offspring and putative parent, that parent can then be excluded. These sequential pairwise comparisons continue, and the parent with the least errors can be established. Jones et al. (2010) concluded that strict exclusion-based methods are not ideal given that they do not allow for genotyping errors. In addition, complete exclusion may be difficult to achieve without a large number of markers, given a large pool of offspring and potential parents. Many methods have since been established that provide a more flexible approach to parentage assignment. The parentage program CERVUS (Kalinowski et al., 2007), the primary program used in this study, uses the categorical allocation method instead of strict exclusion, where scoring errors, mutations and confidence of parentage assignments can be assessed to determine the most likely parent pair for offspring.

Estimates in the literature suggest varying conversion rates for resolving power between SNPs and microsatellites. Fisher et al. (2009) found a 40-SNP panel was a comparable diagnostic tool compared to a 14-microsatellite panel in dairy cattle, and Hauser et al. (2011) found that an 80-SNP panel was more powerful than an 11-microsatellite panel in sockeye salmon. The resolving power of a SNP is contingent on having a high level of polymorphism in the population, which is often reported as the minor allele frequency (MAF, or q). MAF reflects the allele that occurs least frequently at a particular SNP location when evaluating SNP genotypes within a population. Estimates vary for the minimum number of SNPs required to assign parentage. Simulations conducted by Anderson & Garza (2006) determined that 60-100 SNPs were likely to be sufficient for accurate parentage assignment. Liu et al. (2015) developed a SNP panel for the rainbow trout with a high average MAF (q = 0.37), and they found that 48 SNPs were able to provide accurate parental assignment 98.4% of the time.

MAF is an important component to consider when designing assays intended for parentage assignment through SNP-based testing, and can be helpful in determining the number of SNPs required
for accurate assignment when designing low-density assays. High MAF values (ideally approaching the maximum value of 0.5) are desirable when determining parentage, given that these high MAF SNPs provide more polymorphic information for downstream analysis. Through simulations, Anderson & Garza (2006) found that SNPs are most powerful for parentage assignment when the MAF (q) = 0.5, though the gain in power from frequencies from 0.2 - 0.3 is greater than the gain from 0.4 - 0.5, and false-positive rates decrease as q approaches 0.5. Strucken et al. (2016) found that panels with higher MAFs were able to resolve parentage more accurately with fewer SNPs using several breeds of cattle. Baruch & Weller (2008) found through simulations that non-exclusion probability decreased with increasing minimum MAF cut-off levels at the same number of SNPs on a panel.

Little has been done to assess the effect of MAF on parentage assignment beyond simulations (Anderson & Garza, 2006; Baruch et al., 2008). This paper aims to provide additional insight into the effect of MAF on parentage assignment accuracy using both simulations and an empirical dataset with several populations of farmed Atlantic salmon. The main question posed by this study is the following: How does manipulating the number of SNPs available in tandem with different levels of average MAF affect parentage assignment accuracy?

### 2.2. MATERIALS AND METHODS

#### 2.2.1. Animals and genotyping

Genetic material from three populations of the Saint John River (SJR) aquacultural strain of Atlantic salmon provided by Cooke Aquaculture Inc. were used in this study. The first population of fish was made up of 65 dams, 39 sires, and 146 offspring. The second population was made up of 58 dams, 45 sires and 165 offspring. The third population was made up of 47 dams, 39 sires, and 105 offspring. These populations will be hereafter referred to as 2006YC, 2007YC and 2008YC, respectively where YC refers to Year Class. The familial relationships of these fish were previously established through the use of microsatellite markers by the New Brunswick Provincial Research and Productivity Council (RPC) on
behalf of Cooke Aquaculture Inc., though I performed my own parentage analysis to verify this pedigree. These initial populations were then subject to a filtering process for genotype quality and availability, which reduced the numbers of fish available across each group.

All fin clip material from the three populations of SJR Atlantic salmon considered in this study were genotyped using a custom Atlantic salmon iSelect™ 6K bead array created by Illumina (San Diego, USA) and 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. Genotypes were available for both the offspring and all potential parents across the three populations of fish used in this study. The 6K array was designed for the European Atlantic salmon subspecies, and as a result there are SNPs on the array that were found to be multi-site variants (MSVs), did not amplify, or were fixed in the SJR population and thus had to be removed from the analysis.

2.2.2. Filtering and selecting SNPs

In addition to removing SNPs with low call rates, all SNP data were visually inspected in the Illumina software GenomeStudio©. This process allowed identification of SNPs that would not be useful for downstream analysis. Through examination of the polar co-ordinate graphs produced by GenomeStudio© for each individual SNP within the population, those SNPs that appeared to be monomorphic or MSVs in the SJR population of Atlantic salmon were flagged, and excluded when exporting genotypic data for each individual. This left only bona fide binary SNPs for subsequent analysis. The genotypic data for each year class of fish was then exported from GenomeStudio© and converted to a format compatible with the CERVUS (v3.0.7) parentage program using an ad-hoc script I wrote in Python 3.5 and can be found in Appendix A. After the filtering process, a pool of 3980 SNPs remained to create in silico assay designs using varying numbers of SNPs at varying average MAFs. The goal of this technique was to test the effect of varying average MAF values on parentage assignment across varying numbers of SNP markers.
The MAF for each SNP was calculated individually for each generation, and a subset was selected that approximated a target average value \( q = 0.40, 0.30, 0.20, 0.15, 0.10 \). From these average MAF groups, subsets of SNPs were selected that varied in number \( n = 150, 125, 100, 75, 60, 50, 25 \), each to be tested for accuracy in pedigree reconstruction. These SNPs were selected from a the larger 6K genotype dataset using an *ad hoc* Python script I wrote that can be found in Appendix B.

### 2.2.3 Pedigree verification

The previously confirmed microsatellite-derived pedigrees were embedded in the 6K bead array dataset available in GenomeStudio©. Using a large subset of SNP markers \( n > 200 \) with high minor allele frequencies \( q > 0.35 \), the genotypes of which did not differ significantly from Hardy-Weinberg Equilibrium, parentage assignment was conducted through CERVUS on each individual in order to verify the results of the pedigree previously established by microsatellite genotyping. The individuals considered for this experiment were then selected if there was a complete consensus between the SNP parental assignment through CERVUS and the previously established microsatellite parental assignment. When the result between the CERVUS parental assignment and the microsatellite pedigree differed for a particular fish through a dam mismatch, sire mismatch, or both, that offspring was removed from the dataset, as its true parentage could not be confirmed. Parents that were not assigned to any of the offspring as dam or sire were also removed from the dataset, as these data were not useful. In total, 21 offspring, 3 dams, and 3 sires were removed from the 2006YC population; 41 offspring, 10 dams, and 3 sires were removed from the 2007YC population, and 15 offspring, 7 dams, and 4 sires were removed from the 2008YC population.

Once a verified subset of offspring and parents with full consensus between CERVUS and the microsatellite pedigree was identified for each of the three populations, the experiment proceeded. After filtering the pedigree, 2006YC contained 125 offspring, 62 dams, and 36 sires; 2007YC contained 124 offspring, 48 dams, and 39 sires; 2008YC contained 90 offspring, 40 dams, and 35 sires (Table 2.1).
2.2.4. Simulations

CERVUS requires generation of simulated data before proceeding with parental assignment, which is accomplished using the Simulation module provided within the software. These simulations are used to determine how large the logarithm-of-odds (LOD) of the most likely candidate parent must be in order for parentage to be assigned to a particular individual. The simulation uses empirical allele frequencies determined from the CERVUS allele frequency module, which uses observed genotypes of the parents and offspring. The simulation then takes the following input parameters into account (the bracketed term contains the input values that were used): number of offspring (n = 10,000), number of candidate males (n = number of candidate males in respective population), number of candidate females (n = number of candidate females in respective population), proportion of males sampled (1.0), proportion of females sampled (1.0), proportion of loci typed (0.98), and proportion of loci mistyped (0.02). While genotyping error was set to 2%, genotyping errors for SNP markers are reported to be generally under 1% (Saunders et al., 2007; Wiggans et al., 2009). Any other parameters within CERVUS not specified were left as defaults. A total of 105 parental assignments were conducted (35 per year class), consisting of 7 sets of SNPs (n = 150, 125, 100, 75, 60, 50, 25) tested against 5 target average MAF groups (q = 0.40, 0.30, 0.20, 0.15, 0.10) for each of the three year classes analyzed. Once simulations were complete for all 105 groups considered in this study (Table 2.2), the number of strict assignments of the 10,000 simulated offspring considered were extracted from each output file.

2.2.5. Parentage assignment with empirical data

Parentage assignment was conducted using the CERVUS Parent Pair (Sexes Known) module. This module requires a list of dams, sires, offspring, simulation results, and a file containing all genotypes. The output file produced contains LOD scores, which are evaluated according to the simulation results. The estimated dam and sire is listed for each offspring, and the significance of each,
along with the likelihood of the trio (offspring, sire, dam) is reported according to a 95% confidence interval. Each parental assignment was obtained from the CERVUS output file and compared to the verified pedigree to determine the proportion of correct assignments, which is referred to as assignment accuracy.

When using the CERVUS Parent Pair (Sexes Known) module, a list of all potential sires and dams were provided to the program for each respective population. The required allele frequency file was previously generated in the simulation stage of each parentage assignment test, as well as the required simulation file. The “most-likely parent” for each offspring was selected from the CERVUS dropdown menu when selecting report criteria for each offspring.

In my preliminary analyses, comparisons were made between the software packages, COLONY (Jones & Wang, 2010) and SNPPIT (Anderson, 2010) as alternatives to CERVUS. The time required to produce a single run in COLONY took many hours longer than a single run in CERVUS or SNPPIT, making it unfeasible for conducting the 105 runs necessary for my experimental design. SNPPIT was able to compute parentage at a very rapid rate, completing a run in only a few minutes, while CERVUS could take up to 30 minutes to complete a single run. Despite the speed differences, CERVUS was selected for ease of use and accuracy. Liu et al. (2015) found that SNPPIT was slightly less accurate for parentage assignment than CERVUS, and my preliminary work suggested the same trend.

2.3. RESULTS

2.3.1. Properties of SNP markers

The average MAF of each year class of fish for each SNP level was very close to the target average MAF in the majority of cases (Table 2.2). For the 0.40 average MAF group, the 2006YC and 2007YC populations deviated the most from the intended target, with the lowest average MAF estimated at 0.37. The average values of the expected heterozygosity across all loci for each group were also determined in CERVUS (Table 2.3).
2.3.2. Simulations and empirical results

CERVUS was able to generate lists of the most likely pairs of parents for each offspring tested for each subset of SNP parameter values for SNP number and MAF. Parentage assignments were then compared to the previously validated pedigree to determine if one or both of the parents were assigned correctly. At 150 SNPs, the proportion of accurately assigned offspring across all populations ranged from 98.6% (q = 0.10) to 100% (q = 0.40), increasing along with MAF (Figure 2.1). At 125 SNPs this trend continued, with accuracies ranging from 97.6% (q = 0.10) to 100% (q = 0.40). As the number of SNPs decreased, the spread between the accuracies reported also increased according to MAF (Figure 2.1). Accuracies span the greatest range from 7.8% (q = 0.10) in 2008YC to 84.0% (q = 0.40) in 2006YC at the lowest level of 25 SNPs.

The results from the simulations of 10,000 individuals conducted in CERVUS show a similar trend to the empirical data (Figure 2.2). The proportions of accurate assignments in the empirical dataset are typically higher than those in the simulations at the lower SNP levels (n = 25, 50). This discrepancy is likely because a ‘correct assignment’ was determined in the empirical data if there was consensus with the pedigree, regardless of the statistical confidence of the assignment. This contrasts with the simulation results, which were only given ‘strict’ status if an assignment fell within the 95% confidence level. The 2008YC simulation results differed substantially from the 2008YC empirical analysis when q = 0.10, as the empirical assignment accuracies were smaller than the simulations. Despite 95% confidence in the majority of trio assignments, many assignments conflicted with the verified pedigree. These empirical results also show variation in the number of parents incorrectly assigned across each year class, with decreasing numbers of SNPs and decreasing average MAF both showing a trend of increasing numbers of misassignments (Figure 2.3). Far more instances in which both parents were misassigned were observed at the lower values of average MAF, especially at the 25 SNP level across all year classes.

As average heterozygosity is sometimes reported instead of MAF as an attribute of SNPs used in parentage studies (Weinman et al., 2015), I also included expected heterozygosity in my results in
addition to the average MAFs for each group tested (Table 2.3). As expected, the calculated heterozygosity values are strongly correlated with the calculated average MAF (Figure 2.4).

2.4. DISCUSSION

Low-density SNP assays can be used to provide information on numerous traits, and can also provide a means to establish accurate parentage assignments in aquaculture populations. By reducing the number of SNPs required to genotype each individual in a population, the cost per genotyped individual can be reduced accordingly. By establishing minimum values for SNP parameters, such as number of SNP loci genotyped per individual, and average MAF, assays developed by aquaculture operations can be optimized to maximize accuracy, while also minimizing the cost associated with genotyping. My results fill a gap in the literature by detailing the relationship between the number of SNPs needed to achieve a given level of parentage assignment as a function of MAF.

Decreasing accuracy in parental assignment is an expected observation when testing decreasing numbers of available SNP genotypes and decreasing average MAF in populations, the effects of which are both detailed in simulation work by Anderson & Garza (2006). These simulations first demonstrated that 60-100 SNPs may provide accurate pedigree reconstruction, this number being contingent on family structure and on MAF, with higher MAF values being desirable for increased assignment power. The results generated by Liu et al. (2015) with an average MAF of 0.34 for their SNP subset show a similar trend to the 0.3 and 0.4 average MAF subsets in my dataset, with accuracy decreasing substantially below the 50 SNP mark. This result regarding numbers of SNPs has since been empirically analyzed by many studies, spanning many species (Tokarska et al., 2009; Harlizius et al., 2011; Hauser et al., 2011; Steele et al., 2013; Liu et al., 2015; Weinman et al., 2015) typically using sets of high average MAF SNPs (q > 0.35).

Family-based structure allows for accurate estimates of the heritability of traits of interest. Through accurate estimates of trait heritability, molecular markers can be used to exploit genetic variation in a
population, utilizing selective breeding for gains in important quantitative traits (Sonesson & Meuwissen, 2009). These genetic gains and their accuracy are contingent on the availability of accurate pedigrees. As low-density SNP assays can be a valuable tool in aquaculture breeding operations, this work can help inform their design in the future when desirable SNPs of very high MAF (approaching q = 0.5) are not immediately available. By considering both the MAF and the number of SNPs available for a panel, one can estimate with reasonable certainty how effective that panel will be at assigning parentage, given complete availability of parental genotypes. By creating groups comprised of varying numbers of SNPs across varying target average MAF values derived as subsets from the initial 6K genotype array results, I was able to generate results detailing the precise effect of these parameters on parentage assignment accuracy (Figure 2.1).

Future studies regarding the accuracy of parental assignments may benefit from the consideration of some factors not included in this study, such as parental relatedness, which can have a substantial effect on assignment success (Marshall et al., 1998). That said, the results presented here can be used to make informed decisions on the potential accuracy of low-density SNP assays across varying numbers of available SNP genotypes and average MAF values. Expected heterozygosity can be used instead of average MAF, as illustrated by the very strong correlation between the two measures (Figure 2.4). My results show that the number of SNPs needed for a desired level of parentage assignment accuracy is indeed strongly influenced by average MAF. My results also confirmed minimum criteria necessary for high accuracy parentage assignment (>98%) when considering MAF and SNP number for a number of scenarios (Figure 2.1 and Figure 2.2). Valuable statements like the following can be made with the aid of my results: when average MAF = 0.30, a minimum of 60 SNPs are necessary for 98% accuracy; when average MAF = 0.10, a minimum of 150 SNPs are necessary for 98% accuracy. These results will provide a useful reference point for aquaculture producers, especially when evaluating the potential for accurate assignment for low-density SNP assays that are in the early development stages.
2.5. FIGURES AND TABLES

Figure 2.1: Simulations showing relationship between the number of SNPs and the number of simulated ‘strict assignments’ assigned by CERVUS across five levels of average minor allele frequency in three families of Atlantic salmon. Strict assignments are offspring to parent assignments conducted by CERVUS that have been deemed to be significant at a 95% confidence level. The x-axis represents the total target number of SNPs per group. The ‘AVG.MAF’ label above each panel represents the target average minor allele frequency for each group.
Figure 2.2: Empirical results showing relationship between the number of SNPs and the proportion of correctly assigned Atlantic salmon offspring by CERVUS to a validated pedigree across five levels of average minor allele frequency using an empirical data set across three populations. The x-axis represents the total target number of SNPs per group. The ‘AVG.MAF’ label above each panel represents the target average minor allele frequency for each group.
Figure 2.3: Empirical results showing the number of Atlantic salmon parents incorrectly assigned by each year class, with the number of SNPs and minor allele frequency group both indicated. The x-axis represents the total target number of SNPs per group. The ‘AVG.MAF’ label above each panel represents the target average minor allele frequency for each group.
Figure 2.4: Relationship between $q$, the minor allele frequency (MAF), and the empirical value of expected heterozygosity ($2pq$) in three populations of Atlantic salmon (2006, 2007, and 2008). Points are colour coded by each average MAF dataset that was constructed from the empirical data.
Table 2.1: Numbers of Atlantic salmon available in empirical dataset with parentage validated with microsatellites used to evaluate the effect of minor allele frequency (MAF) and number of SNPs on parentage assignment accuracy.

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of offspring</th>
<th>No. of dams</th>
<th>No. of sires</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006YC</td>
<td>125</td>
<td>62</td>
<td>36</td>
</tr>
<tr>
<td>2007YC</td>
<td>124</td>
<td>48</td>
<td>39</td>
</tr>
<tr>
<td>2008YC</td>
<td>90</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>339</td>
<td>150</td>
<td>110</td>
</tr>
</tbody>
</table>
Table 2.2: Estimated average MAF for each SNP group across the 2006YC, 2007YC, and 2008YC populations of Atlantic salmon. The ‘Target MAF’ column represents the desired average minor allele frequency for a group of selected SNPs. Each group is presented, including the 3 populations of fish referred to respectively as 2006YC, 2007YC, and 2008YC. Each group is further split across each SNP group (25, 50, 60, 75, 100, 125, and 150 SNPs). Actual average MAF value for each group were obtained from the 6K genotype dataset.

<table>
<thead>
<tr>
<th>Target MAF</th>
<th>25 SNPs</th>
<th>50 SNPs</th>
<th>60 SNPs</th>
<th>75 SNPs</th>
<th>100 SNPs</th>
<th>125 SNPs</th>
<th>150 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.122</td>
<td>0.118</td>
<td>0.117</td>
<td>0.117</td>
<td>0.119</td>
<td>0.120</td>
<td>0.120</td>
</tr>
<tr>
<td>0.15</td>
<td>0.154</td>
<td>0.153</td>
<td>0.160</td>
<td>0.159</td>
<td>0.164</td>
<td>0.163</td>
<td>0.166</td>
</tr>
<tr>
<td>0.20</td>
<td>0.210</td>
<td>0.212</td>
<td>0.202</td>
<td>0.209</td>
<td>0.209</td>
<td>0.211</td>
<td>0.213</td>
</tr>
<tr>
<td>0.30</td>
<td>0.297</td>
<td>0.288</td>
<td>0.291</td>
<td>0.290</td>
<td>0.296</td>
<td>0.297</td>
<td>0.297</td>
</tr>
<tr>
<td>0.40</td>
<td>0.396</td>
<td>0.374</td>
<td>0.374</td>
<td>0.372</td>
<td>0.374</td>
<td>0.373</td>
<td>0.377</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target MAF</th>
<th>2007YC</th>
<th>25 SNPs</th>
<th>50 SNPs</th>
<th>60 SNPs</th>
<th>75 SNPs</th>
<th>100 SNPs</th>
<th>125 SNPs</th>
<th>150 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.095</td>
<td>0.099</td>
<td>0.100</td>
<td>0.098</td>
<td>0.095</td>
<td>0.094</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.147</td>
<td>0.153</td>
<td>0.152</td>
<td>0.150</td>
<td>0.144</td>
<td>0.147</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.204</td>
<td>0.201</td>
<td>0.212</td>
<td>0.203</td>
<td>0.202</td>
<td>0.199</td>
<td>0.197</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.302</td>
<td>0.315</td>
<td>0.310</td>
<td>0.312</td>
<td>0.306</td>
<td>0.305</td>
<td>0.304</td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.373</td>
<td>0.386</td>
<td>0.383</td>
<td>0.380</td>
<td>0.379</td>
<td>0.380</td>
<td>0.374</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target MAF</th>
<th>2008YC</th>
<th>25 SNPs</th>
<th>50 SNPs</th>
<th>60 SNPs</th>
<th>75 SNPs</th>
<th>100 SNPs</th>
<th>125 SNPs</th>
<th>150 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.101</td>
<td>0.100</td>
<td>0.100</td>
<td>0.101</td>
<td>0.101</td>
<td>0.10</td>
<td>0.101</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.151</td>
<td>0.150</td>
<td>0.150</td>
<td>0.151</td>
<td>0.151</td>
<td>0.152</td>
<td>0.152</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.205</td>
<td>0.203</td>
<td>0.203</td>
<td>0.207</td>
<td>0.206</td>
<td>0.205</td>
<td>0.205</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.304</td>
<td>0.305</td>
<td>0.307</td>
<td>0.308</td>
<td>0.309</td>
<td>0.307</td>
<td>0.305</td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.401</td>
<td>0.399</td>
<td>0.401</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.400</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3: Expected heterozygosity for each SNP subset for each Atlantic salmon population calculated in CERVUS. Values are presented according to target average minor allele frequency (“Target MAF”) and target number of SNPs. The strong correlation between MAF, or \( q \), and heterozygosity is expected given that heterozygosity can be represented by \( 2pq \).

<table>
<thead>
<tr>
<th>Target MAF</th>
<th>25 SNPs</th>
<th>50 SNPs</th>
<th>60 SNPs</th>
<th>75 SNPs</th>
<th>100 SNPs</th>
<th>125 SNPs</th>
<th>150 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.212</td>
<td>0.206</td>
<td>0.203</td>
<td>0.204</td>
<td>0.206</td>
<td>0.209</td>
<td>0.208</td>
</tr>
<tr>
<td>0.15</td>
<td>0.258</td>
<td>0.256</td>
<td>0.263</td>
<td>0.264</td>
<td>0.270</td>
<td>0.269</td>
<td>0.273</td>
</tr>
<tr>
<td>0.20</td>
<td>0.326</td>
<td>0.328</td>
<td>0.315</td>
<td>0.324</td>
<td>0.325</td>
<td>0.328</td>
<td>0.331</td>
</tr>
<tr>
<td>0.30</td>
<td>0.409</td>
<td>0.403</td>
<td>0.406</td>
<td>0.405</td>
<td>0.409</td>
<td>0.410</td>
<td>0.410</td>
</tr>
<tr>
<td>0.40</td>
<td>0.471</td>
<td>0.460</td>
<td>0.455</td>
<td>0.458</td>
<td>0.459</td>
<td>0.460</td>
<td>0.464</td>
</tr>
<tr>
<td>2007YC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.170</td>
<td>0.175</td>
<td>0.177</td>
<td>0.174</td>
<td>0.169</td>
<td>0.167</td>
<td>0.169</td>
</tr>
<tr>
<td>0.15</td>
<td>0.248</td>
<td>0.255</td>
<td>0.252</td>
<td>0.250</td>
<td>0.243</td>
<td>0.246</td>
<td>0.246</td>
</tr>
<tr>
<td>0.20</td>
<td>0.318</td>
<td>0.313</td>
<td>0.325</td>
<td>0.315</td>
<td>0.316</td>
<td>0.312</td>
<td>0.310</td>
</tr>
<tr>
<td>0.30</td>
<td>0.412</td>
<td>0.424</td>
<td>0.421</td>
<td>0.422</td>
<td>0.417</td>
<td>0.416</td>
<td>0.415</td>
</tr>
<tr>
<td>0.40</td>
<td>0.455</td>
<td>0.464</td>
<td>0.463</td>
<td>0.461</td>
<td>0.464</td>
<td>0.464</td>
<td>0.459</td>
</tr>
<tr>
<td>2008YC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.152</td>
<td>0.174</td>
<td>0.174</td>
<td>0.173</td>
<td>0.192</td>
<td>0.201</td>
<td>0.204</td>
</tr>
<tr>
<td>0.15</td>
<td>0.251</td>
<td>0.246</td>
<td>0.227</td>
<td>0.246</td>
<td>0.248</td>
<td>0.248</td>
<td>0.248</td>
</tr>
<tr>
<td>0.20</td>
<td>0.321</td>
<td>0.314</td>
<td>0.314</td>
<td>0.319</td>
<td>0.317</td>
<td>0.317</td>
<td>0.317</td>
</tr>
<tr>
<td>0.30</td>
<td>0.421</td>
<td>0.421</td>
<td>0.423</td>
<td>0.424</td>
<td>0.425</td>
<td>0.422</td>
<td>0.421</td>
</tr>
<tr>
<td>0.40</td>
<td>0.482</td>
<td>0.480</td>
<td>0.480</td>
<td>0.480</td>
<td>0.479</td>
<td>0.479</td>
<td>0.479</td>
</tr>
</tbody>
</table>
2.6. REFERENCES


CHAPTER III: MAPPING QUANTITATIVE TRAIT LOCI FOR INFECTIOUS SALMON ANEMIA (ISA) RESISTANCE IN A NORTH AMERICAN STRAIN OF ATLANTIC SALMON

3.1. INTRODUCTION

3.1.1. Background

Infectious salmon anemia (ISA) is a waterborne viral disease that affects farmed Atlantic salmon (*Salmo salar*) on a global scale, leading to severe economic losses due to both the high virulence of the virus and the high mortality rates in Atlantic salmon caused by the disease (Murray & Peeler, 2005). The prevalence of ISA is of great importance to the aquaculture industry, with outbreaks across Chile, Norway, the state of Maine, Maritime provinces of Canada, the Faroe Islands, and Scotland reported over the last three decades (Rimstad et al., 2011). Selection on aquacultural stock of Atlantic salmon for resistance to ISA has been an objective for some breeding programs since the mid-1990s (Moen et al., 2007). Despite ongoing research and preventative measures taken, the threat of ISA outbreaks remains an issue for Atlantic salmon aquaculture producers.

Outbreaks of ISA were first detected in Norway in 1984 (Jarp & Karlsen, 1997) and have since persisted as a global issue within the aquaculture industry. There has been a significant effort to contain and minimize the damage caused by ISA, with solutions ranging from aquacultural management changes (Murray and Peeler, 2005), vaccine development (Mikalsen et al., 2005) and selective breeding for resistance (Kjøglum et al., 2008). The economic impacts of ISA outbreaks are severe, with the cost to Canadian producers alone estimated to be $14 million CAD annually (Hastings et al., 1999). The impact of ISA on wild populations of Atlantic salmon is not currently known (Falk et al., 1997). However, the high density of fish in aquaculture pens likely exacerbates losses caused by ISA in farmed populations and may increase the threat to wild salmon through virus shedding into the environment.
3.1.2. Infectious salmon anemia virus (ISAV)

ISAV is a negative sense single-stranded RNA (ssRNA) virus in the Orthomyxoviridae family, the sole member of the genus, ISAVirus (Rimstad & Mjaaland, 2002), and a close relative of the influenza viruses. Signs of the disease include sluggishness, anemia, and haemorrhage of eyes and skin, all of which relate to circulatory disturbances. The progressive anemia that regularly occurs will result in ‘watery blood’ indicated by a low haematocrit score (Rimstad et al., 2011). ISAV has a tropism for the endothelial cells of the vasculature (Aamelfot et al., 2012), which is unsurprising given the extensive damage that is observed in the circulatory system of an infected fish. Studies on differential gene expression as a result of ISA have shown that many genes related to oxygen transportation are under-expressed post-injection with a virulent strain of ISAV (LeBlanc et al., 2010). Symptoms of the disease may vary depending on the virulence of the particular strain of ISA, viral dose, environmental conditions, management practices and genetic makeup of the infected fish (Rimstad et al., 2011). The majority of ISA outbreaks have occurred in the seawater stage of farmed Atlantic salmon, with very few outbreaks documented in the freshwater stage of growth (Lyngstad et al., 2008). Due to the high mortality associated with ISA and the highly communicable nature of ISAV, many aquaculture operations now cull entire net pen populations upon detection of ISA to minimize the potential for further spread. Atlantic salmon producers are now required by law to notify the Canadian Food Inspection Agency (CFIA) of any suspected ISA infections. Depending on the situation, the movements of the fish may be restricted and the humane destruction of the fish may be required.

Transmission of the virus seems to be mostly horizontal (Lyngstad et al., 2008), though recent evidence suggests that ISAV may also be transmitted vertically (from parent to offspring) to the gametes of infected broodstock (Marshall et al., 2014). Horizontal transmission of ISAV can occur through vector organisms, non-biological routes, or salmon effluent (blood, mucous, faeces, urine and skin). Studies on Scottish aquaculture operations have linked shipping containers to the transmission of ISAV among populations (Murray et al., 2002). This is an example for the potential of non-biological transmission.
Several members of the family Salmonidae, such as the rainbow trout (*Oncorhynchus mykiss*), are known to act as carriers for ISAV, though the virus does not replicate or generate a symptomatic response in these species (Nylund *et al.*, 1997). In addition, the salmon louse (*Lepeophtheirus salmonis*) is believed to be a vector leading to outbreaks of ISA while simultaneously leading to a host of other health issues in Atlantic salmon (Nylund *et al.*, 1994; Grimnes & Jakobsen, 1996). ISAV is also capable of surviving in water without a host for up to 20 hours, leading to the possibility of passive transmission in addition to the vertical and horizontal pathways already detailed (Cottet *et al.*, 2011).

### 3.1.3. Selection on Atlantic salmon

Quantitative trait loci (QTL) are regions of the genome that at least partially determine a quantitative (or continuous) phenotypic trait. The QTL itself comprises one or more variable loci that are associated with variation in the trait (Mackay *et al.*, 2009). Through statistical associations between molecular markers and observations on traits of interest, we are able to identify QTL location in the genome and in some cases the magnitude of phenotypic variance attributable to them. A common molecular marker used for QTL mapping studies is the single nucleotide polymorphism (SNP), which is a single base variant at a given locus within a population (Mackay *et al.*, 2009). If a SNP is in linkage disequilibrium with a QTL, they are associated to some extent, and may be transferred together during recombination events. Molecular markers with close linkage to a QTL of interest are desirable, as this increases the likelihood that the markers will be useful for tracking QTL throughout subsequent generations.

Traits of economic importance have been the primary focus of QTL mapping studies. For instance, in Atlantic salmon, SNPs have been used to determine which loci affect fillet texture and fat content (Sodeland *et al.*, 2013), growth rate (Gutierrez *et al.*, 2012), adult maturation (Pedersen *et al.*, 2013) and other traits such as body shape (Boulding *et al.*, 2008). Knowledge of QTL locations and their
-associated molecular markers can allow the design of breeding schemes using marker-assisted selection to select for traits of economic importance.

An estimated breeding value (EBV) is a value widely used in animal breeding, and represents the additive component of variance in a trait that an animal will pass to its offspring. Typically, phenotypic selection for growth rate is the priority for most breeding programmes in Atlantic salmon. However, with the use of statistical models that predict breeding values, breeders can simultaneously select for other traits of economic interest such as carcass quality and growth rate (Quinton et al., 2005). With the advent of lower cost SNP-genotyping arrays for Atlantic salmon, the potential for genomic data to be included in selection algorithms to produce genomic-estimated breeding values (GEBVs) is now possible. Genomic selection is a novel iteration of marker-assisted selection where many marker effects are estimated simultaneously. This method of genomic selection has the potential to increase selection accuracy, increase genetic gain, and to reduce inbreeding levels for aquacultural populations (Sonesson & Meuwissen, 2009). Nielsen et al. (2009) found that an aquaculture program could benefit from genomic selection, though at that time the cost of applying the technology on a large scale was prohibitive.

In order for selection on a trait to be successful within a commercial population, the trait must be heritable. Heritability (more specifically, narrow-sense heritability, or \( h^2 \)) is a measure of the proportion of observed phenotypic variance for a trait attributable to additive genetic variation. Successful selection on a specific trait requires that it has a non-zero value of heritability (Goddard & Hayes, 2009). ISA-resistance has been shown to be a heritable trait in some Norwegian populations of farmed salmon: Kjøglum et al. (2008) estimated the narrow-sense heritability of the trait to be 0.37, while Ødegård et al. (2007) found a heritability of 0.32. These high heritability values suggest that Atlantic salmon will respond to selection for resistance to ISA, allowing aquaculture producers to eventually employ genomic or marker-assisted selection to breed more disease-resistant fish.
3.1.4. Mapping ISA-resistance to the Atlantic salmon genome

Using microsatellite markers, a QTL for ISA-resistance in a Norwegian strain of *Salmo salar* was mapped by Moen *et al.* (2007) to Linkage Group 8, which corresponds to Chromosome 15 (Phillips *et al.*, 2009). This resistance QTL accounted for 6% of the phenotypic variation observed in their test population, further suggesting that marker-assisted selection targeting QTL for ISA resistance is possible. Given that sparsely interspersed microsatellite markers were used for this study, Moen *et al.* (2007) suggested that further fine mapping would allow identification of molecular markers that are in population-wide association with ISA-resistance QTL. Li *et al.* (2011) suggested that the QTL identified by Moen *et al.* (2007) could be associated with the gene HIV-EP2/MBP-2. Eventually identifying a tightly linked marker for this particular gene (or any others associated with the QTL) could prove to be valuable as a reliable set of SNP markers indicative of ISA-resistance has not yet been developed for the Saint John River aquacultural strain. With a suite of closely linked SNP markers indicating QTL for ISA-resistance, a low density SNP-assay could be produced, allowing for selection of resistant broodstock in a commercial setting at low cost.

Moen *et al.* (2011) identified a major resistance QTL for the economically important disease infectious pancreatic necrosis (IPN), another disease of Atlantic salmon that leads to major losses in the Atlantic salmon aquaculture industry. This QTL explained 29% of the phenotypic variation observed in their test population, suggesting that there is a major gene (or several) represented by this QTL. Selection for broodstock with SNP markers tightly linked to this QTL could result in a marked change in disease occurrence in aquacultural populations, thereby reducing losses. While these results are promising for IPN resistance, it is worth noting that finding a single QTL that explains the bulk of variation in a trait is a rare occurrence. Typically, traits are polygenic, meaning they are under the influence of many genes, each exerting an additive effect (Mackay *et al.*, 2009). This polygenic tendency is representative of the Fisher’s infinitesimal model (1918), which assumes that genetic variation is attributable to many genes exerting a small quantitative effect on any given trait (Barton & Keightley, 2002). Given the large scale application
of the infinitesimal model in quantitative genetics, and with previous studies (Moen et al. 2007) only managing to identify a single QTL with a relatively small effect on ISA-resistance, it is reasonable to predict that no single major gene controls the bulk of the resistance mechanism for ISA in Atlantic salmon.

A new unpublished custom Affymetrix Axiom 50K 384-peg SNP-array for a North American subspecies of Atlantic salmon was used for my study. This will enable finer resolution and greater genome SNP coverage than achieved in previous studies, which will aid in the search for additional QTL or more reliable markers.

3.1.5. Questions and goals

My main research question was as follows: are there any detectable ISA-resistance QTL in the Saint John River (SJR) strain of Atlantic salmon? To answer this question, ISA disease challenges had been previously conducted on full-sib families of Atlantic salmon, where fish that survived were deemed ‘resistant’. These full-sib families and their parents were then genotyped to search for associations between genotype and the ISA-resistant phenotype. I generated both male and female linkage maps specific to the SJR population. To locate the map position of any ISA-resistance QTL segregating within any of my six full sibling families I then uploaded: 1) phenotypes from a disease challenge on 6 families, 2) the SNP genotypes derived from a new custom 50K array, and 3) my previously constructed linkage maps to the GridQTL software server and ran the sib pair analysis module. Identification and location of these QTL could enable implementation of a commercial marker-assisted selection program for ISA-resistance for the SJR strain of Atlantic salmon, eventually fostering a more resistant population.
3.2. MATERIALS AND METHODS

3.2.1. ISAV culture

A highly virulent North American strain of ISAV was provided by Novartis to perform the disease challenge. The virus was cultured and administered by Novartis staff. Further details on the culturing of this virus are not available due to the proprietary nature of this information in respect to Novartis’s intellectual property.

3.2.2. ISAV disease challenge

The disease challenge was conducted in the now former quarantine laboratory of the St. Andrews Biological Station (SABS) of the Department of Fisheries and Oceans (DFO), New Brunswick in 2011. This facility had the capacity to use recirculated natural sea water (LeBlanc et al., 2010; 2012; Laflamme et al., 2013). The disease challenges were conducted as part of a larger project led by Novartis Animal Health in collaboration with Cooke Aquaculture Inc., the New Brunswick Research Productivity Council (RPC), and DFO SABS. All fish were provided by Cooke Aquaculture Inc. and constituted a saline-injected control group obtained from a larger experiment. Fish were maintained (daily feeding, temperature monitoring, density monitoring) by SABS staff. All fish had reached the smolt stage of their lifecycle (approximately 1-year old fish) and were housed in seawater. A total of 10 full-sib families were challenged using a cohabitation method of infection (n = 586), where an additional group of fish comprising a proportion of the total population per tank were intra-peritoneally (ip) injected with a virulent strain of ISAV. The ip-injected fish were evenly allocated into two separate 3000L tanks, with the remaining fish distributed across the tanks to cohabitate with the ip-injected fish, allowing for passive infection of these individuals. Every fish was tagged with a passive-induction transponder (PIT) tag for later identification and fin clipped. Fish mortalities were recorded for passively-infected fish, and these fish were removed from the tanks upon death. In addition, non-ISAV specific mortalities were monitored. The disease trial continued for approximately 2 months, and was terminated once mortality
rates per day had levelled off (Figure 3.1). Fish that survived the trial and remained once mortalities had leveled off were deemed ‘resistant’. Canadian Food Inspection Agency (CFIA) approval of the operation of the SABS facility was obtained for the 2011 challenge, and approval was given for the facility to operate as a Level 2 in vivo aquatic containment facility.

3.2.3. DNA extraction, genotyping, parentage

Six of the 2010-year class families from the 2011 ISA disease challenge were selected for genotyping, totalling 321 individuals (Table 3.1). Families were selected based off of their performance in the disease trial, with preference for families that showed a high degree of phenotypic variation (Figure 3.2). Parentage assignments of the fish were initially determined by RPC using microsatellite marker genotypes for Cooke Aquaculture Inc. Adipose fin clips from the challenged fish were used for the high molecular weight DNA extraction using a standard extraction protocol with the QIAGEN DNeasy-96 blood and tissue extraction kit at the University of Guelph. This extracted DNA was sent to the Centre of Integrative Genetics (CIGENE), Norwegian University of Life Sciences, and genotyped on the NA Ssa 50K custom Axiom 50K SNP array developed specifically for the SJR strain of Atlantic salmon. This custom 50K SNP array was designed in collaboration between the University of Guelph, Cooke Aquaculture Inc. and CIGENE as part of Genome Canada’s Genomic Applications Partnership Program (GAPP). The new custom array was primarily based on a previous 220K array designed by CIGENE, intended for use on a European strain of Atlantic salmon, but also included SNPs useful for North American populations from a 6K SNP array (Lien et al. 2011). The parents of the challenged fish had already been genotyped on this higher density 220K-SNP array. The vast majority of SNPs available on the 50K chip are also present on the 220K chip, allowing use of these parental genotyping results for the downstream QTL mapping experiment.
3.2.4 Linkage maps

Genotypes from individuals and their parents were used to construct both female- and male-specific linkage maps necessary for the downstream QTL analysis, because the existing linkage map for North American Atlantic salmon has fewer than 4K SNPs on it (Brenna-Hansen et al., 2012). The female map was constructed from families F02 and F03, which were merged using the R package LPmerge (Endelman & Plomion, 2014). The male map was constructed only from family F02. Maps were generated using the R package OneMap (Margarido et al., 2007), a program initially developed for constructing linkage maps of outcrossing plant species within full-sib families using a maximum-likelihood approach for linkage estimation. A substantial amount of marker filtering was necessary to produce the linkage maps. For the female specific map for each family, only markers for which the dam was heterozygous and the sire homozygous were used, whereas the opposite was used for the male specific map. Markers that had low call rates, high error rates, or were significantly different from expected Mendelian segregation ratios in the family were also removed. After filtering, a total of 11,056 SNP markers remained to be positioned on the female map, and a total of 5,622 on the male map. Two-point recombination fractions were estimated with the rf.2pts() function within OneMap, with LOD = 6 specified as a parameter, and the remaining available parameters left as defaults. Markers were assigned to linkage groups using the make.seq() function. Finally, markers were ordered using the order.seq() function, which attempts to determine the best order of markers. I observed the expected large recombination differences between markers within males and females, and consequently only used the higher resolution composite female map in my subsequent QTL mapping analysis. Finally, the linkage maps were visualized using the Rqtl R library (Broman et al., 2003). All OneMap R scripts were executed on the servers of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca), a consortium of universities and colleges, including the University of Guelph.
3.2.5. **QTL mapping**

QTL mapping was conducted using the GridQTL web application (Seaton *et al.*, 2006). The sib-pair model was used, which takes advantage of the full-sib structure of the pedigree and genotypes of the parents. GridQTL uses a linear regression-based interval mapping approach for QTL detection (Knott *et al.*, 1996). The method works using a multiple-marker based approach, where the probability of inheriting an allele at a particular marker location is determined, and then used to produce information content values for each marker. Phenotypes are then regressed on the probabilities of inheriting particular alleles. In this study, the phenotypes were the binary outcome of the disease challenge (0 = survived, 1 = mortality). F-ratios calculated at each defined interval along the chromosome were then used to represent the strength of evidence of a QTL at a particular location. Significance thresholds for a QTL can be established at either the chromosome- or experiment-wide level using permutation tests. A total of 1,000 permutations were used for both the chromosome- and experiment-wide tests, producing F-ratio thresholds for significance for a P-value = 0.05 and for a P-value=0.01. I conducted 1,000 permutation tests on all 27 chromosomes when using the sib-pair module to generate results. All results generated by GridQTL were downloaded from the web module.

3.3. **RESULTS**

3.3.1. **Linkage map**

Separate male and female maps containing 27 linkage groups were produced using SNP data from each of two families, F02 and F03. The female map was a composite of F02 and F03 linkage maps, while the male map was produced from only F02. In total, genotypes from 125 individuals were used to construct the linkage maps. A total of 5,622 SNP markers were placed on the male map, and a total of 11,056 SNP markers were placed on the composite female map. There were indications of heterozygosity in at least one of the four parents in: a) the fusion of chromosomes Ssa26 and Ssa28, b) Ssa08 and Ssa29, and c) translocation of the p-arm of Ssa01 to Ssa23 in the results generated by OneMap. Nevertheless,
only the 27 most frequent linkage groups were used in the final maps to simplify the QTL mapping analysis. A graphical display of the linkage maps was generated in OneMap and visualized using Rqtl (Figure 3.3).

3.3.2. Sib-pair analysis results

Evidence for putative chromosome-wide QTL affecting ISA-resistance were detected on six chromosomes of the 27 scanned (Table 3.2). Using the sib-pair model, QTL with chromosome-wide significance at the P < 0.05 level were detected on six chromosomes. The names of these chromosomes cannot yet be provided due to an intellectual property agreement with Cooke Aquaculture Inc. and as such will be referred to alphabetically in no particular order. None of these QTL were significant at the P < 0.01 level, and no QTL were significant at the experiment-wide level.

One of the most significant QTL detected appeared at 132 cM on Chromosome B with an F-statistic of 26.62 (Table 3.2). This QTL position is flanked by the marker AX-87139480 at 112.74 cM and AX-87300089 at 148.36 cM (Figure 3.4b). A second QTL was detected on Chromosome A at 72 cM with an F-value of 20.1 (thresholds of 18.19 and 26.94 for P < 0.05 and P < 0.01, respectively) which is flanked by markers AX-87542706 at 70.77 cM, and AX-87206939 at 72.29 cM (Figure 3.4a). A third QTL was detected on Chromosome C at 18 cM with an F-statistic of 23.95 (thresholds of 19.80 and 29.15 for P < 0.05 and P < 0.01 respectively), and was flanked by markers AX-87441663 at 16.26 cM and AX-87531376 at 18.01 cM (Figure 3.4c). The fourth QTL detected on Chromosome D is located at 18 cM with an F-statistic of 17.87 (thresholds are 17.31 and 28.17 for P < 0.05 and P < 0.01 respectively); it is flanked by markers AX-87290930 at 12.63 cM and AX-87620330 at 18.99 cM (Figure 3.4d). The fifth QTL detected on Chromosome E was located at the 54 cM position with an F-statistic of 21.24 (thresholds of 19.948298 and 27.57819 for P < 0.05 and P < 0.01 respectively); it is flanked by markers AX-87165194 at 53.71 cM and AX-87078554 at 56.79 cM (Figure 3.4e). Finally, the sixth QTL was detected on Chromosome F at the 204 cM position with an F-statistic of 20.38 (thresholds of 18.518509
and 30.876396 for $P < 0.05$ and $P < 0.01$ respectively); its nearest marker is AX-87076036 at the 170.98 cM position (Figure 3.4f).

### 3.4. DISCUSSION

The goal of this research was to determine potential QTL linked with ISA resistance in the SJR strain of Atlantic salmon. Through mapping of QTL linked to ISA resistance, fish could eventually be selected for breeding based on their genotyping results, with a preference for ISA resistant individuals. Through continual selection of more resistant individuals, incidences of ISA outbreaks could be reduced, thereby mitigating potential economic or ecological impacts caused by ISA. This study has identified six suggestive QTL for ISA-resistance, all of which are significant at the chromosome-wide level and act as a reference point for future studies seeking to validate these QTL. The identification of these QTL mark the initial steps taken in the long-term implementation of a marker-assisted selection breeding program that targets broodstock for ISA resistance, thereby reducing incidences of ISA outbreaks in aquaculture operations.

This experiment used phenotypes and tissue samples from an ISA-resistance disease trial conducted in 2011 in collaboration with RPC, DFO-SABS, Novartis Animal Health, and Cooke Aquaculture. Data was generated and analysed for six full-sib families of Atlantic salmon. Genotyping was conducted on a custom 50K SNP array specifically developed for the Saint John River (SJR) strain of Atlantic salmon. A female specific linkage map was generated through the OneMap R library, which was used for QTL mapping using the GridQTL software using their sib-pair module to account for the full-sib family structure.

QTL mapping has been used to map many traits in Atlantic salmon, though ISA-resistance has been rarely addressed. A previous experiment by Moen et al. (2007) identified a QTL for ISA-resistance on Ssa15 that accounted for 6% of the phenotypic variation observed in their trial population composed of European Atlantic salmon, though they only used a sparse set of 8 microsatellite markers. A subsequent
study by Li et al. (2011) investigated the QTL on Ssa15 further and suggested that the HIV-EP2/MBP-2 gene could be related, as it may influence several other genes that have been implicated in the immune response to infection by ISAV. I expected to detect ISA resistance QTL in the SJR population of Atlantic salmon, given the past work by Moen et al. (2007), as well as moderate estimates of the heritability of the ISA resistance phenotype (Gjoen et al., 1997). I did not detect any ISA-resistance QTL on Ssa15 as in Moen et al. (2007), but instead on six other chromosomes (Figure 3.4); these chromosome names cannot yet be provided due to an intellectual property agreement with Cooke Aquaculture Inc. Due to limitations of the sib-pair model used by GridQTL, I could not quantify the amount of phenotypic variance attributable to any of the QTL detected. This could be done with my dataset in the future by using the half-sibling module in GridQTL and data from a more extensive disease challenge. The QTL that I detected were all significant at the P < 0.05 level evaluated at the chromosome-wide level, rather than the experiment-wide level. These results mean that these QTL are ‘suggestive’ (Everett & Seeb, 2014) and need to be confirmed in future studies.

Many studies have examined gene expression changes in Atlantic salmon tissues upon exposure to ISA (Jorgensen et al., 2007; McBeath et al., 2007; Jorgensen et al., 2008; Schiotz et al., 2008; Workenhe et al., 2009). There have also been associations made between resistance against ISA and major histocompatibility complex (MHC) gene polymorphism in Atlantic salmon (Grimholt et al., 2003), which provided a reference point for potential sources of resistance. LeBlanc et al. (2012) characterized differential genes expressed in head-kidney tissue between primary and secondary exposures to ISAV in Atlantic salmon, providing valuable insight into the immune response of the fish upon primary and secondary infections. Aamelfot et al. (2012) identified a receptor on Atlantic salmon endothelial cells that correlated. These expression studies and receptor studies provide potential starting points for investigating regions located near the QTL detected in my study that may be involved in the genetic basis of ISA-resistance.
Previous studies have reported large differences in sex-specific recombination rates in Atlantic salmon, with a heavy skew towards many additional recombination events for females compared to males (Danzmann et al., 2008; Moen et al., 2008). Lien et al. (2011) identified sex-specific recombination patterns in a European strain of Atlantic salmon using a 6.5K SNP assay; their female map spanned 2402.3 cM, and the male map was 1746.2 cM, though their ratio of female to male recombination events was smaller than reported in previous studies due to increased coverage of telomeric regions, where most male recombination takes place. Lien et al. (2011) also predicted that as marker coverage increased, the map distances between female and male maps would converge. In contrast to their cM distances, the female linkage map that I produced spanned 3622.6 cM (Figure 3.3a), while the male linkage map was 1295.5 cM (Figure 3.3b), showing that despite significantly increased coverage, there was no convergence between the maps; additional coverage in the telomeric regions may be necessary.

The most common number of linkage groups found in my linkage maps was the 27 shown in my composite female linkage map. However, a few parents were heterozygous for particular linkage groups and therefore showed 28 linkage groups. Brenna-Hansen et al. (2012) also found an average of 27 linkage groups in the linkage maps that they constructed for the North American salmon. They note that one fluorescently-labelled individual was polymorphic for the linkage group Ssa26/28 (Fig. 6.2 in Brenna-Hansen et al., 2012). This chromosome Ssa26/28 is interesting as it is one of three linkage groups that are fused in North American Atlantic salmon (Brenna-Hansen et al., 2012) but not in European Atlantic salmon (Lien et al., 2011).

These results identifying six suggestive QTL for ISA resistance in the SJR strain of Atlantic salmon are a starting point for further investigation and study. Additional studies involving larger numbers of families, many half-siblings, and additional individuals per-family will allow for more powerful detection of QTL linked to ISA resistance. Through the eventual detection of QTL for ISA resistance detected at the experiment-wide level, aquaculture producers of the SJR strain of Atlantic salmon can consider validating these QTL through further disease challenges, and eventually implement
them into marker-assisted selection programs to meet their breeding objectives. In addition to the suggestive QTL detected in this study, high-density, sex-specific linkage maps were produced, which will aid in the mapping of other QTL for various phenotypic traits of interest in the future. These maps also provide a starting point for further refinement through the addition of further families, allowing for even higher density maps to be produced.
3.5. FIGURES AND TABLES

Figure 3.1: Survivorship curve for the infectious salmon anemia (ISA) cohabitation disease challenge conducted on Atlantic salmon used for QTL mapping experiment. Fish that survived the trial were recorded as ‘resistant’ to ISA.
Figure 3.2: Response to infectious salmon anemia (ISA) cohabitation disease challenge by Saint John River (SJR) strain Atlantic salmon individuals from 10 different full-sibling families. Fish were provided by Cooke Aquaculture from their 2010 year-class as part of a larger animal health experiment; this group represents a saline-injected control group that was used for this experiment. Fish were approximately 1-year old and had undergone smoltification. Challenged fish were housed in the quarantine lab at DFO SABS in seawater split across two 3,000L tanks.
Figure 3.3: Sex-specific linkage maps for a) female (composite for families F02 and F03) and b) male (family F02) Atlantic salmon generated using OneMap software and merged with LPmerge. Markers have visualized in Rqtl been 'jittered' for clarity using the jittermap function, which visually spreads overlapping markers represented by ticks on each chromosome apart by a marginal amount.
Figure 3.4: Significant ISA QTL graphs generated in GridQTL detected at the chromosome-wide level in Atlantic salmon. F-statistic thresholds at the $P < 0.05$ level were determined through 1,000 permutations. Chromosome and cM position for significant QTL are shown in the figure as follows a) Chromosome A at 72 cM, b) Chromosome B at 135 cM, c) Chromosome C at 18 cM, d) Chromosome D at 18 cM, e) Chromosome E at 54 cM, f) Chromosome F at 204 cM.
Table 3.1: ISA Disease trial conducted in 2011 at DFO SABS using 2010-year class Atlantic salmon sourced from Cooke Aquaculture’s Saint John River (SJR) aquacultural strain\(^1\).

<table>
<thead>
<tr>
<th>Family</th>
<th>No. offspring</th>
<th>Dam</th>
<th>Sire(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F02</td>
<td>25</td>
<td>2006OB0339</td>
<td>2006OB0636</td>
</tr>
<tr>
<td>F03</td>
<td>69</td>
<td>2006OB0510</td>
<td>2006OB1872</td>
</tr>
<tr>
<td>F04</td>
<td>58</td>
<td>2006OB1297</td>
<td>2006OB1214</td>
</tr>
<tr>
<td>F06</td>
<td>59</td>
<td>2006OB1489</td>
<td>2006OB1890</td>
</tr>
<tr>
<td>F07</td>
<td>40</td>
<td>2006OB1785</td>
<td>2006OB1406</td>
</tr>
<tr>
<td>F09</td>
<td>70</td>
<td>2006OB2112</td>
<td>2006OB1872</td>
</tr>
</tbody>
</table>

\(^1\)A total of 321 offspring across six families are included.

\(^2\)F03 and F09 are paternal half sibling families.
Table 3.2: All GridQTL results for ISA-resistance in the Saint John River strain of Atlantic salmon QTL mapping experiment. LOD scores (logarithm of the odds to the base 10) are presented as calculated by GridQTL. Values presented in “P < 0.05” and “P < 0.01” columns represent F-statistic thresholds necessary for significance at those levels, respectively.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Location (cM)</th>
<th>t-statistic</th>
<th>F-statistic</th>
<th>LOD</th>
<th>P &lt; 0.05</th>
<th>P &lt; 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>34</td>
<td>-3.68</td>
<td>13.56</td>
<td>2.944</td>
<td>21.904926</td>
<td>33.444843</td>
</tr>
<tr>
<td>I</td>
<td>124</td>
<td>-3.24</td>
<td>10.49</td>
<td>2.278</td>
<td>18.977842</td>
<td>32.44552</td>
</tr>
<tr>
<td>H</td>
<td>130</td>
<td>-3.66</td>
<td>13.38</td>
<td>2.906</td>
<td>21.917036</td>
<td>32.99629</td>
</tr>
<tr>
<td>K</td>
<td>105</td>
<td>-4.5</td>
<td>20.23</td>
<td>4.393</td>
<td>23.652159</td>
<td>32.67288</td>
</tr>
<tr>
<td>A</td>
<td>72</td>
<td>-4.48</td>
<td>20.1</td>
<td>4.365</td>
<td>18.194754</td>
<td>26.93503</td>
</tr>
<tr>
<td>L</td>
<td>70</td>
<td>-2.91</td>
<td>8.47</td>
<td>1.839</td>
<td>19.898813</td>
<td>30.795229</td>
</tr>
<tr>
<td>M</td>
<td>92</td>
<td>-3.06</td>
<td>9.34</td>
<td>2.029</td>
<td>20.368284</td>
<td>27.953245</td>
</tr>
<tr>
<td>N</td>
<td>59</td>
<td>-2.89</td>
<td>8.35</td>
<td>1.813</td>
<td>23.531357</td>
<td>34.487705</td>
</tr>
<tr>
<td>O</td>
<td>82</td>
<td>-3.32</td>
<td>11</td>
<td>2.388</td>
<td>18.979088</td>
<td>31.480556</td>
</tr>
<tr>
<td>B</td>
<td>135</td>
<td>-5.16</td>
<td>26.62</td>
<td>5.781</td>
<td>22.220491</td>
<td>34.024696</td>
</tr>
<tr>
<td>P</td>
<td>84</td>
<td>-2.59</td>
<td>6.71</td>
<td>1.456</td>
<td>17.241814</td>
<td>25.411997</td>
</tr>
<tr>
<td>Q</td>
<td>2</td>
<td>-4.13</td>
<td>17.06</td>
<td>3.704</td>
<td>21.07636</td>
<td>30.51892</td>
</tr>
<tr>
<td>C</td>
<td>18</td>
<td>-4.89</td>
<td>23.95</td>
<td>5.201</td>
<td>19.801964</td>
<td>29.15263</td>
</tr>
<tr>
<td>R</td>
<td>125</td>
<td>-3.16</td>
<td>9.99</td>
<td>2.17</td>
<td>22.229097</td>
<td>30.34534</td>
</tr>
<tr>
<td>S</td>
<td>38</td>
<td>-2.63</td>
<td>6.94</td>
<td>1.507</td>
<td>19.592089</td>
<td>31.472034</td>
</tr>
<tr>
<td>T</td>
<td>58</td>
<td>-3.63</td>
<td>13.15</td>
<td>2.855</td>
<td>21.10171</td>
<td>29.586939</td>
</tr>
<tr>
<td>I</td>
<td>48</td>
<td>-2.91</td>
<td>8.44</td>
<td>1.833</td>
<td>21.795795</td>
<td>33.979935</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
<td>-4.23</td>
<td>17.87</td>
<td>3.88</td>
<td>17.30691</td>
<td>28.166103</td>
</tr>
<tr>
<td>V</td>
<td>66</td>
<td>-5.13</td>
<td>26.27</td>
<td>5.704</td>
<td>26.496155</td>
<td>48.70676</td>
</tr>
<tr>
<td>W</td>
<td>23</td>
<td>-1.82</td>
<td>3.32</td>
<td>0.721</td>
<td>19.14795</td>
<td>26.975224</td>
</tr>
<tr>
<td>E</td>
<td>54</td>
<td>-4.61</td>
<td>21.24</td>
<td>4.613</td>
<td>19.948298</td>
<td>27.579819</td>
</tr>
<tr>
<td>F</td>
<td>204</td>
<td>-4.51</td>
<td>20.38</td>
<td>4.426</td>
<td>18.518509</td>
<td>30.876396</td>
</tr>
<tr>
<td>Y</td>
<td>30</td>
<td>-2.38</td>
<td>5.66</td>
<td>1.229</td>
<td>18.61597</td>
<td>26.289604</td>
</tr>
<tr>
<td>A1</td>
<td>101</td>
<td>-2.87</td>
<td>8.24</td>
<td>1.789</td>
<td>19.940166</td>
<td>29.496151</td>
</tr>
</tbody>
</table>

1Locations along each chromosome are represented in centiMorgan (cM) distance. LOD scores indicated are approximate.
3.6. REFERENCES


CHAPTER IV: CONCLUSIONS

4.1. THE UTILITY OF LOW COST SNP ASSAYS

My thesis has contributed to the goal of lowering genotyping costs and increasing sample throughput by determining the minimum number of SNPs needed to accurately estimate parentage for a typical Atlantic salmon breeding program. I have also shown that a 50K SNP chip was a sufficiently high density to enable the detection of QTL for ISA-resistance.

Through simulations and the empirical analysis of three separate year classes of fish, I was able to provide an in depth look at the effect of MAF on parentage assignment accuracy. My empirical findings obtained by comparing simulations of allele frequencies from an actual Atlantic salmon breeding program with actual parentage results from the same population were predicted by more general simulations initially conducted by Anderson & Garza (2006, Figure 5). My work is valuable to the aquaculture industry in general, as smaller budgets often limit the size of assays and panels available. My findings provide some valuable insight into what the smallest number of SNPs of a particular MAF are needed for a given accuracy of parentage assignment. By considering these results, a more cost effective low-density SNP parentage assay can be designed. This work was made possible through extensive previous genotyping conducted on a 6K assay, which allowed for large groups of SNPs with similar MAF values to be pulled and applied to parentage analysis.

The QTL mapping aspect of this thesis also involves the improvement of aquaculture through lower cost genotyping with SNP markers. The QTL mapping required me to construct a genetic linkage map that was entirely dependent on accurate pedigrees. The QTL mapping itself was done within full-sibling families. In future, estimating conventional and genomic breeding values is entirely dependent on accurate parentage assays.

More work must be done to detect QTL with the potential to eventually be implemented into breeding programs, as no experiment-wide QTL for ISA resistance were detected. New experiments would benefit from larger family sizes, additional families, a further refined linkage map, or a genome-
wide association study (GWAS) designed to locate SNPs associated with ISA resistance. Work conducted by Moen et al. (2007) paved the way for ISA-resistance QTL mapping, though this work was conducted using a European strain of Atlantic salmon and sparsely positioned microsatellite markers. By using a dense 50K chip to establish new linkage maps specific to the SJR population, I was able to search for QTL specific to this group. The QTL detected in my analysis had not been previously identified and were located on six separate chromosomes. This work provides a starting point for breeding Atlantic salmon that are resistant to ISA and the eventual implementation of marker-assisted selection if this proves to be an economically viable option.

4.2. THE POTENTIAL BENEFITS OF SNPS IN AQUACULTURE

SNPs have now become the molecular marker of choice for many tasks rooted in genomics in the Atlantic salmon aquaculture industry. In addition to parentage assignment and QTL mapping, GWAS and genomic estimated breeding values (GEBVs) are being applied with great success in the field of animal breeding with the use of SNP markers. The relationship between a particular genotype and phenotype can be determined through GWAS, much in the same way QTL mapping can relate particular loci to a phenotype. Both of these techniques are now primarily driven by SNPs and can allow for marker-assisted selection if valuable QTL or associated SNPs are discovered.

GEBVs are another emerging and valuable tool that leverages SNPs. GEBVs act as an extension of traditional estimated breeding values (EBVs) which are used to estimate the overall genetic value of an animal for a particular trait and allow more informed breeding decisions to be made. GEBVs allow the incorporation of potentially massive genotypic datasets into the calculations of genetic value, allowing much earlier but still accurate estimates of the potential value of a particular animal to a breeding company (Schaeffer, 2006). By using very dense marker sets, the effects of all QTL present in the genome can potentially be captured through linear models that assign an additive value to every single SNP marker based on its contribution to a particular phenotype (Hayes et al., 2009). The methodology of
estimating GEBV for aquaculture species is still being developed, but should be especially useful for ranking the genetic merit of candidate broodstock within full sibling families for “difficult” traits. Difficult traits are defined as those that must be measured on close relatives of future broodstock, such as most disease resistance or carcass traits (Meuwissen et al., 2015). Genomic selection is expected to increase the gain per generation in difficult traits once the cost of genotyping is sufficiently reduced and SNP chip genotyping is currently the only economical way to apply this to aquaculture.
4.3. REFERENCES


APPENDIX A

```python
import pandas as pd
import sys
import csv
import os

print('Python version ' + sys.version)
print('Pandas version: ' + pd.__version__)

#Enter the name of the file you want to convert from alphabetical to numeric (2 column per marker) format.
#Do not include the extension.
filename = "2008fish_filtered"

#Set 6K file name
data_6k = filename + "_.csv"

#Read desired SNP list to be selected
snp_list = open("2008fish_snplist.txt", "r")
good_snp_list = snp_list.read().split("n")

#Read data to pandas dataframe
df_6k = pd.read_csv(data_6k, sep=';', usecols=good_snp_list, dtype=object)

df_6k = df_6k.sort_index(axis=1)

cols_6k = df_6k.columns.tolist()

cols_6k = cols_6k[1:-1] + cols_6k[-1:len(cols_6k)]

#Select columns
sorted_df_6k = df_6k[cols_6k]

#Throw a delimiter between everything, replace alpha genotypes
#Homozygotes
sorted_df_6k = sorted_df_6k.replace("AA", "1$1")
sorted_df_6k = sorted_df_6k.replace("CC", "2$2")
sorted_df_6k = sorted_df_6k.replace("GG", "3$3")
sorted_df_6k = sorted_df_6k.replace("TT", "4$4")

#Missing genotypes
sorted_df_6k = sorted_df_6k.replace("--", "0$0")

#Heterozygotes
sorted_df_6k = sorted_df_6k.replace("AT", "1$4")
sorted_df_6k = sorted_df_6k.replace("TA", "4$1")
sorted_df_6k = sorted_df_6k.replace("AG", "1$3")
sorted_df_6k = sorted_df_6k.replace("GA", "3$1")
sorted_df_6k = sorted_df_6k.replace("AC", "1$2")
sorted_df_6k = sorted_df_6k.replace("CA", "2$1")
sorted_df_6k = sorted_df_6k.replace("CG", "3$2")
sorted_df_6k = sorted_df_6k.replace("GC", "2$3")
sorted_df_6k = sorted_df_6k.replace("CT", "2$4")
sorted_df_6k = sorted_df_6k.replace("TC", "4$2")
sorted_df_6k = sorted_df_6k.replace("GT", "4$3")
sorted_df_6k = sorted_df_6k.replace("TG", "4$3")
```
#Output genotypes to .csv files
59. ```python
sorted_df_6k.to_csv(filename + '_temp.csv', sep="","", header=False, index=False)
```  
60. ```python
temp6k = pd.read_csv(filename + '_temp.csv', index_col=False, header=None, delimiter=r'[,\s]', engine='python')
```  
61. ```python
temp6k.to_csv(filename + '_output_processed.csv', sep="","", header=False, index=False)
```  
62. ```
63. #Cleanup
64. os.remove(filename + '_temp.csv')
65. print("SCRIPT COMPLETE")
66. ```
APPENDIX B

1. #This script will take a genotype file and marker list file and select SNP columns according to the list.
2. 3. #file: GENOTYPE.TXT
4. #SAMPLE ID  MARKER_1  MARKER_2  MARKER_3  ...
5. #SAMPLE001  AA  AG  TT  ...
6. 7. #file: MARKERLIST.TXT
8. #Sample Name
9. #MARKER_1
10. #MARKER_3
11. #...etc
12. 13. import pandas as pd
14. import sys
15. 16. print('Python version: ' + sys.version)
17. print('Pandas version: ' + pd.__version__)
18. 19. #Set genotype file name
20. genotypes = "AllGenotypes_2012"
21. 22. #Set SNP list name
23. snplist_file = "MAF10_150SNP_2012"
24. 25. #Define function
26. def pick_snps(genotypes, snplist_file):
27.     snp_list_data = open(snplist_file + "_txt", "r") #list of markers you want (must be .txt)
28.     snp_list = snp_list_data.read().split("n") #this will store the list in snp_list
29.     df_genotypes = pd.read_csv(genotypes + ".csv", sep=',', usecols=snp_list, dtype=object) #use pandas to choose columns
30.     output_filename = snplist_file + ".Converted.csv"
31.     df_genotypes.to_csv(output_filename, sep=',', header=True, index=False)
32. 33. #run pick_snps function
34. pick_snps(genotypes, snplist_file)