Using DNA markers to trace pedigrees and population substructure and identify associations between major histocompatibility regions and disease resistance in rainbow trout (*Oncorhynchus mykiss*)

Nathan Allen Johnson

Thesis submitted to the Faculty of Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Fisheries and Wildlife Sciences

Dr. Eric M. Hallerman, Chair and Department Head

Dr. Yniv Palti, Committee Member

Dr. Klaus D. Elgert, Committee Member

June 21, 2007

Blacksburg, VA

Keywords: Multiplex; Parentage assignment; Pedigree; *Oncorhynchus mykiss;* Strain; Subpopulations; Bacterial coldwater disease; MHC; Disease association; Linkage disequilibrium; Rainbow trout fry syndrome; Microsatellites; *Flavobacterium psychrophilum*

Using DNA markers to trace pedigrees and population substructure and identify associations between major histocompatibility regions and disease resistance in rainbow trout (*Oncorhynchus mykiss*)

Nathan A. Johnson

Eric M. Hallerman, Chair

Fisheries and Wildlife Sciences

(Abstract)

Examination of variation at polymorphic microsatellite loci is a widely accepted method for determining parentage and examining genetic diversity within rainbow trout (*Oncorhynchus mykiss*) breeding programs. Genotyping costs are considerable; therefore, I developed a single-step method of co-amplifying twelve microsatellite loci in two hexaplex reactions. The protocol is designed to promote reproducible results. I applied the protocol to samples previously analyzed at the National Center for Cool and Coldwater Aquaculture (NCCCWA) with previously reported marker sets for a comparison of results. Each marker within the multiplex system was evaluated for duplication, null alleles, physical linkage, and probability of genotyping errors. Data from four of the 12 markers were excluded from parental analysis based on these criteria. Parental assignments were compared to those of a previous study that used five independently amplified microsatellites. Percentages of progeny assigned to parents were higher using the subset of eight markers from the multiplex system than with five markers used in the previous study (98% vs. 92%). Through multiplexing, use of additional markers improved parental allocation while also improving efficiency by reducing the number of PCR reactions and genotyping runs required. I evaluated the methods further through estimation of *F*-statistics, pairwise genetic distances, and cluster analysis among brood-years at the NCCCWA facility. These estimates were compared to those from nine independently amplified microsatellites used in a previous study. F_{st} metrics calculated between brood-years showed similar values of genetic differentiation using both marker sets. Estimates of individual pairwise genetic distances were

used for constructing neighbor-joining trees. Both marker-sets yielded trees that showed similar subpopulation structuring and agreed with results from a model-based cluster analysis and available pedigree information. These approaches for detecting population substructure and admixture portions within individuals are particularly useful for new breeding programs where the founders' relatedness is unknown. The 2005 NCCCWA brood-year (75 full-sib families) was challenged with *Flavobacterium psychrophilum*, the causative agent of bacterial coldwater disease (BCWD). The overall mortality rate was 70%, with large variation among families. Resistance to the disease was assessed by monitoring post-challenge days-to-death. Phenotypic variation and additive genetic variation were estimated using mixed models of survival analysis. The microsatellite markers used were previously isolated from BAC clones that harbor genes of interest and mapped onto the rainbow trout genetic linkage map. A general relationship between *UBA* gene sequence types and *MH-IA*-linked microsatellite alleles indicated that microsatellites mapped near or within specific major histocompatibility (MH) loci reliably mark sequence variation at MH genes. The parents and grandparents of the 2005 brood-year families were genotyped with markers linked to the four MH genomic regions (*MH-IA*, *MH-IB*, *TAP1*, and *MH-II*) to assess linkage disequilibrium (LD) between those genomic regions and resistance to BCWD. Family analysis suggested that *MH-IB* and *MH-II* markers are linked to BCWD survivability. Tests for disease association at the population level substantiated the involvement of *MH-IB* with disease resistance. The impact of MH sequence variation on selective breeding for disease resistance is discussed in the context of aquaculture production.

Acknowledgements

The following thesis, while an individual work, benefited from the insights and direction of several people. First, my Committee Chair, Dr. Eric Hallerman, exemplifies the high quality scholarship to which I aspire and admire. Throughout the past eight years, he has been instrumental in my passion for aquatic genetic research through four-and-a-half courses, unprecedented mentorship, and partnership on a diversity of inspirational research. His friendship and professional relationship with Dr. Yniv Palti created a unique situation for me to conduct this research in a USDA-ARS facility off Virginia Tech's campus. Dr. Yniv Palti has been a dedicated research advisor and is responsible for making this project both possible and successful. He not only provided me with the provisions necessary (state-of-the-art laboratory, funding, technical assistance, etc.) but also demonstrated patience and determination to help me develop necessary skills for the advancement of my career as a scientist. Dr. Klaus Elgert formally introduced me to the field of immunology in my first graduate course. He stimulated my immunological interests by encouraging me to present papers and a final project on fish diseases in a class dominated by premedical students. His immunological expertise and instructive comments at every stage of the thesis process were instrumental to the quality of this document. All three of these gentlemen allied as a committee and provided insights that guided and challenged my thinking, substantially improving the finished product. A cohort of other colleagues and friends significantly contributed as well. The statistical expertise of Dr. Roger Vallejo and Dr. Jeffery Silverstein's fish breeding ingenuity allowed for the best available scientific analyses and publication quality documents. The encouragement, mentorship, and assistance of Dr. Caird Rexroad III began during my first trip to Leetown and still continue today. Dr. Scott Gahr, M. Renee Fincham, Lindsey Pierce, Kristy Shewbridge, Roseanna Long, Dr. Issa Coulibaly, Dr. Fernanda Rodriguez, and Jill Corbett all contributed technical support in the laboratory and I greatly appreciate their assistance and contributions. In fact, Dr. William Hershberger and his group at the NCCCWA have been helpful in every way possible and I thank all of them. I also would like to thank Dr. Tim King and recognize his laboratory group at the USGS Leetown Science Center for all their assistance and friendship over the past two years. Finally, I wish to thank my entire family for their encouragement while continuing my education. My girlfriend, Nicole Meyer, and her family provided on-going support throughout the thesis

process, as well as emotional stability critical for completing the project. My parents instilled in me, from an early age, the confidence and ambition to make my dreams realities and they continue to provide loving support along my way. I cannot end without mentioning those who provided inspiration during the times we shared.. I hope to continue enlightening others through my research and inspiring them to respect our miraculous planet and the marvels of nature it holds. It is for this cause that I dedicate this work.

Table of Contents

Chapter 1. Development and evaluation of a new microsatellite multiplex system for parental allocation and management of rainbow trout (*Oncorhynchus mykiss*) broodstocks

List of Tables

Chapter 1. Development and evaluation of a new microsatellite multiplex system for parental allocation and management of rainbow trout (*Oncorhynchus mykiss*) broodstocks

List of Figures

Chapter 2. Association of major histocompatibility genetic markers with resistance to bacterial coldwater disease in rainbow trout (*Oncorhynchus mykiss*)

List of Tables

Chapter 2. Association of major histocompatibility genetic markers with resistance to bacterial coldwater disease in rainbow trout (*Oncorhynchus mykiss*)

List of Figures

Overview

The rainbow trout (*Oncorhynchus mykiss*) is one of the most commercially important aquaculture species worldwide. It is a member of the Family Salmonidae and is native to the North Pacific Rim of North America and Asia. The McCloud stock was the first ever cultured in 1874 in California (Jenkins and Burkhead 1994). Now this species is widely distributed throughout the world due to its popularity as both a food fish and a sportfish. The long history of intensive hatchery selection has led to a considerable amount of research on this species, and its known culture and biological qualities make it an exceptional model organism for many types of studies.

The Agricultural Research Service (ARS) division of the United States Department of Agriculture (USDA) developed the National Center for Cool and Coldwater Aquaculture (NCCCWA) facility in Kearneysville, WV to focus on rainbow trout aquaculture. The mission of this facility is to support and enhance the nation's cool and cold water aquaculture production through research and technology transfer. My thesis research was conducted at this facility, and its focus was to investigate the applicability of specific DNA markers in genomic studies to improve broodstock management and identify associations of markers with resistance to bacterial diseases in rainbow trout. The facility manages even- and odd-year broodstocks and selectively breeds each for a different set of traits. The even-year selection program mainly focuses on growth traits and diet supplementation, whereas disease resistance is the main breeding priority in the odd-year broodstock.

Monitoring genetic diversity in the broodstocks is performed on a yearly basis by microsatellite variability analysis using fluorescently-labeled polymerase chain reaction (PCR) primers and automated DNA sequencers. DNA microsatellites are repeating motifs of 1-8 nucleotides that are widely dispersed throughout the nuclear genome, usually characterized as di- tri- or tetra-nucleotide repeats. These markers exhibit high co-dominant, Mendelian inheritance and tend to be highly polymorphic due to their high mutational rate. These properties make them a popular choice for a variety of aquaculture applications, such as parentage assignment in mixed family groups and genetic diversity estimation to monitor levels of inbreeding in selective breeding programs (Silverstein et al., 2004). Numerous polymorphic markers are generally required to assign parentage and to characterize intra- or inter-population variation, but the

number needed depends in large part on the properties of the loci used (Bernatchez and Duchesne 2000). The costs associated with these analyses are considerable for large-scale commercial breeding programs. Therefore, I developed a multiplex PCR system capable of high throughput and high-resolution single-step co-amplification of multiple microsatellite markers. This technique has shown great success in a variety of aquatic species (O'Reilly et al., 1998; Fishback et al., 1999; Pruett et al., 2005; Lerceteau-Kohler and Weiss 2006; Porta et al., 2006), but protocols currently available for rainbow trout were not suitable for multiple applications due to genotyping difficulties, duplicated loci, and null alleles. With the recent availability of new microsatellite libraries for this species, I created a new system using microsatellite tracts isolated at the NCCCWA.

It is important to evaluate multiplex performance once a protocol has been established to ensure its applicability and reliability. Evaluation begins by ensuring that alleles are consistently amplified and accurately scored. This is tested by amplifying the same DNA template several times and determining whether there is consistency in size and signal intensity of fragments. Error rates for parentage assignment can be tested against a previous study by using genotypic data from multiplex analysis as input for a software package designed for parentage assignment, such as CERVUS (Slate et al., 2000) or PAPA (Duchesne et al., 2002). The next approach is to demonstrate the feasibility of using the multiplex sets for broodstock management by assessing variation in hatchery stocks. The stocks of interest are the even- and odd- brood-years at NCCCWA. Genetic variation within these populations was monitored each year by determining F_{st}, F_{is}, and F_{it} estimates from nine individually amplified DNA microsatellite loci. I re-analyzed the brood-years from 2002-2005 and compared parameter estimates to determine whether the multiplex marker sets were adequate for monitoring genetic variation. The methods I developed set out how the twelve microsatellite loci are grouped, co-amplified, and evaluated for reliability as tools for several aquaculture applications.

The second major focus of my thesis involved only the odd-year broodstock lines, which are selected for disease resistance. The gram-negative bacterium *Flavobacterium psychrophilum* (*Fp*) poses major fish health concerns in trout and salmon hatcheries around the world and has been documented to cause mortalities >90% (Cipriano and Holt 2005). The disease caused by this bacterium is commonly called bacterial coldwater disease (BCWD) and particularly affects juvenile fish before immune system development, which limits the use of vaccination as a

control or treatment. Genes encoded by the major histocompatibility complex (MHC) are obvious candidates for investigation following bacterial challenges since they are known to play a direct role in immune function and adaptive immunity.

Challenge experiments are the most direct tool for measuring the general competency of the immune system. In fishes, this is a successful approach to determine associations and linkages between MHC alleles and disease resistance and susceptibility to pathogens (Langefors et al., 2001; Palti et al., 2001; Grimholt et al., 2003). In this study, I tested the hypothesis that variation of microsatellite markers linked to rainbow trout MH genes is associated with variation in resistance to BCWD caused by *Flavobacterium psychrophilum*. Variability of the *UBA* gene sequence was compared to genotypic variation at *MH-IA* microsatellites to test the relationship between gene sequence and microsatellite alleles. With the knowledge of families with high or low survival rates after a bacterial challenge, I screened for correspondence between MH gene markers and resistance to BCWD, using microsatellite markers linked to four particular MH gene regions (*MH-IA*, *MH-IB*, *TAP1*, and *MH-II*).

References

Bernatchez, L. and Duchesne, P. (2000). Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Canadian Journal of Fisheries and Aquatic Sciences 57: 1-12.

Cipriano, R. C. and Holt, R. A. (2005). *Flavobacterium psychrophilum*, cause of bacterial cold-water disease and rainbow trout fry syndrome. Fish Disease Leaflet No. 86. United States Dept. of the Interior. U.S. Geological Service, National Fish Health Research Laboratory, Kearneysville, WV.

Duchesne, P., et al. (2002). PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. Molecular Ecology Notes 2: 191-193.

Fishback, A. G., et al. (1999). Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout (*Oncorhynchus mykiss*). Aquaculture 172: 247-254.

Grimholt, U., et al. (2003). MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-219.

Jenkins, R. E. and Burkhead, N. M. (1994). Freshwater Fishes of Virginia. Bethesda, MD, American Fisheries Society.

Langefors, A., et al. (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proceedings of the Royal Society of London. 268: 479-485.

Lerceteau-Kohler, E. and Weiss, S. (2006). Development of a multiplex PCR microsatellite assay in brown trout *Salmo trutta*, and its potential application for the genus. Aquaculture 258: 641-645.

O'Reilly, P. T., et al. (1998). Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. Animal Genetics 29: 363-370.

Palti, Y., et al. (2001). Association between DNA polymorphisms tightly linked to MHC class II genes and IHN virus resistance in backcrosses of rainbow and cutthroat trout. Aquaculture 194: 283-289.

Porta, J., et al. (2006). Development of a microsatellite multiplex PCR for Senegalese sole (*Solea senegalensis*) and its application to broodstock management. Aquaculture 256: 159-166.

Pruett, C. L., et al. (2005). Microsatellite DNA markers for population genetic studies and parentage assignment in cobia, *Rachycentron canadum*. Molecular Ecology Notes 5: 84-86.

Silverstein, J. T., et al. (2004). Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). Aquaculture Research 35: 40-48.

Slate, J., et al. (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. Molecular Ecology 9: 801-808.

Chapter 1

Development and evaluation of a new microsatellite multiplex system for parental allocation and management of rainbow trout (*Oncorhynchus mykiss*) broodstocks

Abstract

Examination of microsatellite variation at several polymorphic loci is a widely accepted method for determining parentage and examining genetic diversity within rainbow trout (*Oncorhynchus mykiss*) breeding programs. Genotyping costs are considerable; therefore, I developed a single-step method of co-amplifying twelve microsatellites in two hexaplex reactions to reduce costs and improve efficiency. The protocol is designed to promote reproducible results. I applied the protocol to samples analyzed at the National Center for Cool and Coldwater Aquaculture (NCCCWA) with previously reported marker sets for a comparison of results. Each marker within the multiplex system was evaluated for duplication, null alleles, physical linkage, and probability of genotyping errors. Data from four of the 12 markers were excluded from parental analysis based on these criteria. Parental assignments were compared to those of a previous study that used five independently amplified microsatellites. Percentages of progeny assigned to parents were higher using the subset of eight markers from the multiplex system than with five markers used in the previous study (98% vs. 92%). Through multiplexing, additional markers were used to improve parental allocation while also improving efficiency by reducing the number of PCR reactions and genotyping runs required. I evaluated my methods further through estimation of *F*- statistics, pairwise genetic distances, and cluster analysis among brood-years at the NCCCWA facility. These measures were compared to estimates from nine independently amplified microsatellites used in a previous study. The F_{st} results calculated between brood-years show similar values of genetic differentiation using both marker sets. Estimates of individual pairwise genetic distances were used for constructing neighbor-joining trees. Both marker-sets yielded trees that show similar subpopulation structuring and agreed with results from a model-based cluster analysis and available pedigree information. These approaches for detecting population substructure and admixture portions within individuals are particularly useful for new breeding programs where the founders' relatedness is unknown.

1. Introduction

DNA microsatellites are powerful genetic markers frequently applied to studies in the research disciplines of aquaculture genetics, evolution, conservation, and fisheries management (Chistiakov et al., 2006). Microsatellites can be utilized as genetic tags to determine parentage in mixed-family studies (Palti et al., 2006), to assess genetic processes underlying reproduction of hatchery broodstocks through estimations of effective population sizes, and to reconstruct pedigrees (Porta et al., 2006). In other studies, microsatellites were used to identify strains of origin and to estimate genetic diversity in hatchery broodstocks (Overturf et al., 2003; Silverstein et al., 2004). They also remain the most common DNA marker used in the mapping of quantitative trait loci (QTL) affecting complex quantitative traits in fish (Danzmann et al., 1999; Ozaki et al., 2001).

The costs associated with these marker applications are considerable. A method commonly used to reduce costs and maximize efficiency for amplifying microsatellites is called multiplexing. This method utilizes a single-step of co-amplifying multiple microsatellite markers in a single polymerase chain reaction (PCR). Multiplex PCR systems have been developed for a variety of aquaculture species, including rainbow trout (Fishback et al., 1999), Atlantic salmon (O'Reilly et al., 1996), cobia (Pruett et al., 2005), brown trout (Lerceteau-Kohler and Weiss 2006), and Senegalese sole (Porta et al., 2006). However, a major limitation in developing multiplex systems for salmonids that prove useful in a variety of applications for a variety of strains has been the high frequency of duplicated marker loci and 'null' alleles (Fishback et al., 1999; Rexroad and Palti 2003). The dramatic increase over recent years in the number of microsatellites available for rainbow trout presents an opportunity for rigorous marker screening, selection, and evaluation for multiplexing in this species. Once a multiplex panel is selected and optimized, it is important to validate the assay to determine whether the properties of the loci used in populations under investigation are reliable for desired applications.

In this paper, I report the development of an optimized protocol for co-amplifying twelve DNA microsatellite loci in two hexaplex reactions for rainbow trout. The protocol was applied to collect genotypic data from samples previously analyzed to evaluate the efficiency and accuracy of the new multiplexes as tools for determining parentage and managing hatchery broodstocks. Parental assignments were calculated and compared to results of a previous study (Palti et al.,

2006) that used five independently amplified microsatellite loci to assign parentage in a mixed-family stock. Secondly, measures of genetic variation were estimated within four brood-years of rainbow trout (2002-2005) maintained at the National Center for Cool and Cold Water Aquaculture (NCCCWA) in Kearneysville, WV using the multiplex marker system. Both the even and odd brood-years (BY) at the NCCCWA were founded by intermixing commercial strains of rainbow trout. The BY2002 was founded by individuals from Troutlodge Inc. (Sumner, WA) and the University of Washington (UW, Seattle, WA). The founding strains for the BY2003 were Troutlodge, UW, Mt. Shasta (from Ennis National Fish Hatchery, Ennis, MT) and House Creek (College of Southern Idaho, Twin Falls, ID). Levels of genetic diversity and inbreeding were monitored previously in the founding populations (BY2002 and BY2003) and within the parents of the first generation (BY2004 and BY2005) using a set of nine independently amplified microsatellite loci (Silverstein et al., 2004). I compared the effectiveness of the newly-developed multiplex marker system and the nine marker set for estimating genetic differentiation between populations and for detecting statistical evidence of subpopulation structure within each brood-year.

2. Materials and methods

2.1. Genomic DNA isolation

Two-hundred and twenty DNA samples (30 parents and 190 progeny) from Troutlodge, Inc. were previously extracted for parental assignments (Palti et al., 2006), and 592 samples from the NCCCWA brood-years 2002-2005 were extracted for measuring genetic variation as previously described (Silverstein et al., 2004). DNA samples were stored in low TE buffer (10 mM Tris, 0.1mM EDTA, pH 8.0) and quantified on a spectrophotometer to determine DNA concentration. Samples with A260/280 ratios less than 1.75 were re-extracted. Working stocks of DNA were diluted to 12.5 ng/µl for PCR.

2.2. Marker selection and screening

Twelve microsatellite loci were selected from the literature and organized in two hexa-multiplexes. The criteria used for marker selection were non-duplication, no known null alleles, at least 50% heterozygosity, and a minimum of five alleles based on initial characterization.

Four loci (*OMM5132, OMM5233, OMM5177, OMM5264*) were previously characterized by Coulibaly et al. (2005), two (*OMM1303, OMM1325*) by Palti et al. (2002), two (*OMM5007, OMM5047*) by Rexroad et al. (2005), two (*OMM1008, OMM1051*) by Rexroad et al. (2002a) and two (*OMM1097, OMM1088*) by Rexroad et al. (2002b). Linkage between microsatellite loci at the time of selection was determined from the NCCCWA genetic linkage map, and the occurrence of null alleles was determined from NCCCWA mapping reference families (Rexroad et al., 2006). Loci names, GenBank accession numbers, primer sequences, dye labels, primer concentrations, annealing temperatures (T_A), and $MgCl_2$ (mM) concentrations are grouped by hexaplex and reported in Table 1. Primer pairs were synthesized by Alpha DNA (Montreal, Quebec, Canada) and Applied Biosystems (PE Biosystems, Foster City, CA, USA), and forward primers were labeled with one of three fluorescent dyes: 6Fam, Hex, or Ned. The specific dye chosen for each marker was based on allele size distribution and optimized polymerase chain reaction (PCR) conditions. Primers were resuspended and diluted to a working concentration of 40 µM. During initial screening, a panel of eight DNA samples from the NCCCWA reference families was amplified in individual reactions for all markers to determine allele sizes and signal intensity. Pairs of markers then were combined for co-amplification. Contrasting allele sizes and intensities were compared to previous results to ensure consistencies in amplification until all markers were amplified consistently within their respective hexaplex groups.

The initial PCR was carried out in a reaction volume of 12 µl containing: 25 ng of genomic DNA, 2.0 mM $MgCl_2$, 200 µM dNTPs, 1X manufacturer's reaction buffer (Applied Biosystems 10X PCR Buffer), 0.5 unit AmpliTaq Gold *Taq* Polymerase (Applied Biosystems, Foster City, CA, USA) and 1.00 µM of each primer. Amplifications were conducted using a MJ Research DNA Engine thermal cycler (Model PTC200, MJ Research, Waltham, MA) as follows: an initial denaturation at 95°C for 10 min, 29 cycles consisting of 94°C for 60 sec, annealing temperature (Table 1) for 45 sec, 72°C extension for 2 min, followed by a final extension of 72°C for 10 min. PCR amplification was verified on 3% agarose gels. Successful reactions then were genotyped with an ABI 3730 genetic analyzer, and the resulting genotype data were analyzed using GeneMapper 3.7 software (ABI, Foster City, CA, USA).

2.3. Parental analysis

All thirty parents and 95 progeny from both fish-meal and plant-meal intermediate groups previously used by Palti et al. (2006) were genotyped using all twelve markers from the two hexaplex marker systems. The resulting genotypes were used to assign each offspring to a single parental pair by the exclusion methods of PAPA 2.0 (Duchesne et al., 2002). Input files for parents were separated by sex and structured according to the mating design described by Palti et al. (2006). The results of parental assignments were compared to results of the previous assignments to evaluate the accuracy and informative capabilities of the hexaplex marker systems.

2.4. Assessment of genetic variation in the NCCCWA broodstock

All fish from the even (2002, 2004) and odd (2003, 2005) brood-years at the NCCCWA were genotyped using the twelve markers by optimized amplification conditions described in the Results section below labeled, *Multiplex optimization*. Data from successful multiplex amplifications were compared to existing data from the nine markers described in a previous study (Silverstein et al., 2004). If the first amplification attempt using the multiplex was unsuccessful, the data were considered missing and no re-amplifications were attempted. Genotypes from both data sets were formatted in Microsoft Excel spreadsheets, and the Microsatellite Toolkit 3.1 (Park 2001) was used to calculate number of individuals genotyped, *PIC* values (Botstein et al., 1980), allele size range, and number of alleles for each locus by population combination. Population differentiation was quantified based on 10100 random permutations for all populations with both marker sets by the infinite allele model-based F_{st} metric (Wright 1965) using ARLEQUIN 3.01 (Schneider et al., 2006). Inbreeding per individual at the subpopulation level (F_{is}) was estimated for both marker sets using GENEPOP 3.3 following the methodologies of Weir and Cockerham (1984). The program MICROSAT 1.5 (Minch 2003) was employed to estimate pairwise genetic distances among individuals within a brood-year, among a subset of unrelated individuals within BY2004 and BY2005, and between brood-years using the transformed proportion of shared alleles (-ln(*PSA*)) and the transformed kinship coefficient (-ln(*Dkf*)) without adjusting for average distances within groups. Unrooted trees were constructed from both distance matrices using the NEIGHBOR option of the program PHYLIP 3.5c (Felsenstein 1993) for the development of neighbor-joining trees. Tree files created by NEIGHBOR were visualized using the program TREEVIEW 1.5 (Page 1998). A

model-based cluster analysis under a Bayesian framework was implemented using the computer program STRUCTURE 2.1 (Pritchard et al., 2000) to infer the proportion of membership of each unrelated individual from BY2004 and BY2005 to inferred clusters using both marker sets. To infer the number of clusters or populations (K) present in the genomic data set (X), the program STRUCTURE estimates the proportion of the genotypes of each individual having ancestry to each cluster and applied a prior probability of the data [$\Pr(X|K)$]. Here, the prior probability $\Pr(X|K)$ was estimated from a series of independent runs of the Gibbs sampler using an admixture model for each value of K between 1 and 10. The admixture model assumes that individuals may have acquired a fraction of their genetic makeup from different strains depending on parental crosses. The results presented are based on runs of 10^6 Monte Carlo Markov Chain (MCMC) iterations, following a burn-in period of 10^5 iterations with a thinning value of 10 to minimize computational requirements.

3. Results

3.1. Multiplex optimization

Reliable amplification was achieved for the 12 microsatellite markers using the multiplex conditions described in Table 1. Adjustments were made to equalize variation in amplified products between loci by changing primer concentrations. Primer concentrations for loci with high signal intensity were reduced and concentrations of those with low signal intensity were increased until a balance was achieved. Amplifications for hexaplex I were conducted using a 'touchdown' PCR protocol (Rithidech et al., 1997) as follows: an initial denaturation at 95°C for 10 min, 2 cycles for each of three annealing temperatures consisting of 94°C for 60 sec, 60°C (58°C, 54°C) for 45 sec, 72°C extension for 2 min, followed by 29 cycles of amplification at the 'bottom' temperature (54°C) and a final extension of 72°C for 45 min. Amplifications for hexaplex II were as follows: an initial denaturation at 95°C for 10 min, 29 cycles consisting of 94°C for 60 sec, 58°C for 45 sec, 72°C extension for 2 min, followed by a final extension of 72°C for 45 min. Extension steps were increased from 15 min. to 45 min. to reduce "plus A" variation as suggested by Fishback et al. (1999).

3.2. Parentage assignment

Polymorphic information content (*PIC*) values were calculated and used to rank each marker (Table 1) based on genotypes from the Troutlodge, Inc. population. Four markers were excluded from the parental analysis. Locus *OMM5264* was determined to have null alleles by Rexroad et al. (2006) and therefore was likely to introduce errors in the exclusion analysis. Two pairs of markers (*OMM5007* and *OMM5047*; *OMM1303* and *OMM1088*) were shown to be in the same linkage group (Rexroad et al., 2006). Loci *OMM5007* and *OMM1303* had the lower *PIC* values in their respective linkage groups and were excluded from the analysis as they did not improve the overall resolving power of the marker panels. Additionally, the genotypic quality of *OMM1303* as part of the multiplex was questionable, and it was removed to reduce genotyping errors. Most of the *OMM5132* alleles were separated by only one base-pair in this population, which caused difficulty in allele scoring and was likely to increase genotyping errors as well.

Genotypes from the remaining eight microsatellite loci enabled correct parental assignments for 98% of the progeny. Complete assignment (100%) was achieved with the uniform error rate allowing for mutations and genotyping errors of ±2 base-pairs. Using the five-marker data, Palti et al. (2006) were able to assign parents to 92% of the progeny before allowing a ±2 base-pair uniform error rate to achieve complete assignment. Parental assignments matched using both data sets. The same four individuals were identified as triploids based on the occurrence of three or more alleles at two or more loci (Palti et al., 2006).

3.3. Genetic assessment of NCCCWA broodstock

Sample sizes, *PIC* values, allele size ranges, and number of alleles for each locus x population combination estimated from the genotypes of both marker sets are shown in Table 2. On average, sample sizes for loci within the multiplex data set were lower because no re-amplifications were attempted. Data from locus *OMM5132* were excluded from the population analysis due to single base-pair spacing between alleles, and data from *OMM5264* were removed because of the likelihood of null alleles.

Population differentiation (F_{st}) and average inbreeding per marker at the subpopulation level (F_{is}) were estimated for both marker sets. Very low F_{st} values between BY2002 and BY2004 in both marker sets indicated, as expected, that the two populations are closely related. Surprisingly, lower similarity was detected between BY2003 and BY2005. Overall, the multiplex F_{st} values were higher than for the nine-marker set (Table 3). F_{is} values were similar

for both marker sets, with the BY2003 values somewhat higher than those for the other three brood-years (Figure 1). F_{is} indirectly estimates the amount of assortative mating in these broodstocks (Wright 1965); higher F_{is} values indicate that selection or inbreeding is causing a reduction in genetic variation.

Estimates of individual pairwise genetic distances, using both –ln(*PSA*) and –ln(*Dkf*) algorithms, led to similar trees for each brood-year using both marker sets. All trees showed evidence of strain subgrouping, indicating that within-strain substructuring existed both in the founding and in the first generation of the broodstocks. A representative neighbor-joining tree created from the –ln(*PSA*) distance matrix is shown in Figure 2. I report a tree based on individuals without common parents from BY2005 to facilitate interpretation of relationships among individuals based solely on strain substructuring instead of family structuring among progeny. To ensure a random sample of individuals from each family, re-sampling was employed, and similar trees were observed. The branch numbers show partitioning of the three major strains, with branches 1, 2, and 3 corresponding to Trout Lodge (TL), University of Washington (UW), and Shasta (SH) resource strains, respectively.

Assignment tests conducted with STRUCTURE succeeded in assigning each individual fish to a cluster relating to their parents' source subpopulation. The BY2004 assessment exhibited two clusters based on the multiplexed genotypes and four clusters based on the nine-marker set (Table 4). For BY2005, the ten markers from the multiplex system exhibited four clusters and the nine-marker set detected three clusters. From these four data sets, the likelihood support of the data [Pr($X|K$)] and the posterior probability of the inferred K [Pr($K|X$)] were the highest for the multiplexed loci in the BY2005. The resolution of the multiplex markers provided results consistent with the number of strains present in each brood-year, suggesting that these markers are more informative for cluster analysis. All clustering was based on differences in allele frequencies, and each cluster was given a unique color (Figure 3). Bars with red coloration denote membership to the Shasta (SH) strain, yellow and green to Trout Lodge (TL) strain, and blue to University of Washington (UW) strain. Individuals of mixed origin (e.g., TLSH) show membership to more than one cluster group. The near equal proportions of green and yellow within TLTL individuals indicate that two genomic backgrounds exist within this strain, while SHSH and UWUW are more homogenous. The influence of the House Creek (HC) strain was not prominent enough to give rise to a fifth cluster, likely due to the absence of pure HCHC

matings and relatively few individuals contributing to this generation. Both the distance-based method and the model-based approach of cluster analysis provide evidence for population structure and admixture in the NCCCWA broodstocks used in this investigation.

4. Discussion

In this paper, I describe a detailed protocol for multiplexing 12 DNA microsatellite markers in rainbow trout and demonstrate the reliability of the multiplex performance through critical evaluation and comparison with previous markers sets. I used the multiplex system to assign parentage in a mixed-family population and verified the accuracy of the assignments. I removed four markers from the parental analysis, as I found them to contain null alleles, be physically linked to other markers, or subject to genotyping error. Markers that exhibit null alleles evoke false exclusions of heterozygous parents that were mistakenly identified as homozygotes (Dakin and Avise 2004), and therefore they should be avoided in parental allocation studies. Linkage between loci can reduce the number of informative alleles due to a reduction in the frequency of recombination and are likely to present problems in parental studies when closely related individuals are being assigned (Amos et al., 1992). I chose to exclude one marker from each linkage group because success of using multilocus genotypes to allocate parentage is a function of the number and properties of the loci (Bernatchez and Duchesne 2000); therefore, the linked markers with lower *PIC* values were removed. Parentage allocation by exclusion is very sensitive to genotyping errors, which can lead to the false exclusion of a potential parent; therefore, the quality of multiplexed genotypes was critically evaluated. For example, stutter peaks from some alleles at locus *OMM5132* showed overlap with peaks from adjacent alleles differing in length by a single base-pair. In these instances, heterozygous individuals were likely to be misidentified as homozygotes; hence, this marker was removed from the analysis. Percentage of progeny assigned to parents with no genotyping errors allowed in the model was higher with the eight markers from the multiplex system compared to the five markers used in the previous study (98% vs. 92%). Through the utilization of multiplexing, additional markers were used to improve parental allocation while also improving efficiency by reducing the number of required PCR reactions. The eight-marker subset from the multiplex

system that I used for parentage allocation is likely to be useful for this application in most rainbow trout populations, as it fits the criteria outlined by Vandeputte et al. (2006).

The newly developed multiplex system is effective in evaluating genetic diversity and identifying strain subdivision in the even and odd brood-years of the NCCCWA broodstock. I removed two of the twelve markers from this analysis to increase confidence in the results. Null alleles distort levels of heterozygosity and can give false impressions of the Wahlund effect or over-estimations of inbreeding in population genetic analyses (Dakin and Avise 2004); therefore, *OMM5264* was excluded. The removal of locus *OMM5132* was necessary to ensure accuracy of genotypes. With regard to the linked loci, effects of linkage did not substantially affect estimates of population differentiation (data not shown) as reported by Amos et al. (1992); therefore, all linked loci were included in the NCCCWA broodstock assessments. The remaining ten loci showed the ability to detect subpopulation structure at a higher resolution than the previously used set of nine individually amplified markers (Silverstein et al., 2004) and is currently being used to genotype BY2006. The protocol was modified into two pentaplexes and re-optimized by doubling the amount of *Taq* polymerase used per PCR reaction to increase signal intensity. Slight modifications may be necessary when marker combinations are changed or if different thermal cycling equipment is used.

Detection and evaluation of subpopulation structure and admixture portions for each individual is useful in assessing genetic relatedness in new broodstocks (such as the one developed at the NCCCWA) where little or no pedigree information is available for reducing levels of inbreeding in the mating design of the founding stocks. This approach also is useful in determining whether substructure exists in established and closed programs that may be losing heterozygosity and genetic variation faster then planned using solely pedigree information. The microsatellite markers evaluated here show potential for detecting substructuring for many generations. I draw this conclusion from model-based cluster analysis results, which showed the contribution of two distinct genomes within the TLTL individuals. Further investigation on the origins of this TLTL strain revealed that it was created from Kamloops drainage trout crossed with Puget Sound steelhead prior to 1980 (Silverstein et al., 2004). The multiplex marker set can clearly discriminate between well-described domesticated rainbow trout strains and also within strains of mixed origin; therefore, these markers are also useful as a unique strain identification tool for breeders.

Acknowledgments

The author would like to thank Trout Lodge, the University of Washington, Ennis National Fish Hatchery, and the College of Southern Idaho for providing the fish for the NCCCWA broodstock program. I thank Tim King (U.S. Geological Survey) for his advice in conducting the phylogenetic analysis and Mike Eackles (U.S. Geological Survey) for his help in screening microsatellites during initial multiplex optimization. I thank Jeffrey Silverstein (U.S. Department of Agriculture) for providing scientific advice and information regarding relatedness of hatchery broodstocks. I also sincerely appreciate the technical contributions and guidance of Issa Coulibaly, Renee Fincham, Roseanna Long, Fernanda Rodriguez, and Kristy Shewbridge.

References

Amos, W., et al. (1992). DNA fingerprinting: parentage studies in natural populations and the importance of linkage analysis. Proceedings of the Royal Society of London. Series B 249: 157-162.

Bernatchez, L. and Duchesne, P. (2000). Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Canadian Journal of Fisheries and Aquatic Sciences 57: 1-12.

Botstein, D., et al. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32: 314-331.

Chistiakov, D. A., et al. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255: 1-29.

Coulibaly, I., et al. (2005). Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. Animal Genetics 36: 543-550.

Dakin, E. E. and Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. Heredity 93: 504-509.

Danzmann, R. G., et al. (1999). Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. Aquaculture 173: 45-58.

Duchesne, P., et al. (2002). PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. Molecular Ecology Notes 2: 191-193.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) Version 3.5c. Department of Genetics, University of Washington, Seattle, WA, USA., (Distributed by Author).

Fishback, A. G., et al. (1999). Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout (*Oncorhynchus mykiss*). Aquaculture 172: 247-254.

Lerceteau-Kohler, E. and Weiss, S. (2006). Development of a multiplex PCR microsatellite assay in brown trout *Salmo trutta*, and its potential application for the genus. Aquaculture 258: 641-645.

Minch, E. (2003). MICROSAT, version 1.5d, (Online) http://hpgl.stanford.edu/projects/microsat/programs/.

O'Reilly, P. T., et al. (1996). Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Canadian Journal of Fisheries and Aquatic Sciences 53: 2292-2298.

Overturf, K., et al. (2003). Comparison of growth performance, immunological response and genetic diversity of five strains of rainbow trout (*Oncorhynchus mykiss*). Aquaculture 217(1-4): 93-106.

Ozaki, A., et al. (2001). Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). Molecular Genetics and Genomics 265(1): 23-31.

Page, R. D. M. (1998). TREEVIEW, version 1.5, Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, Scotland.

Palti, Y., et al. (2002). Characterization of 38 polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Molecular Ecology Notes 2(4): 449-452.

Palti, Y., et al. (2006). Evaluation of family growth response to fishmeal and gluten-based diets in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 255(1-4): 548-556.

Park, S. D. E. (2001). Trypanotholerance in West African cattle and the population genetic effects of selection., University of Dublin.

Porta, J., et al. (2006). Development of a microsatellite multiplex PCR for Senegalese sole (*Solea senegalensis*) and its application to broodstock management. Aquaculture 256: 159-166.

Pritchard, J. K., et al. (2000). Inference of population structure from multilocus genotype data. Genetics 155: 949-959.

Pruett, C. L., et al. (2005). Microsatellite DNA markers for population genetic studies and parentage assignment in cobia, *Rachycentron canadum*. Molecular Ecology Notes 5: 84-86.

Rexroad, C. E., et al. (2002a). Development of rainbow trout microsatellite markers from repeat enriched libraries. Marine Biotechnology 4: 12-16.

Rexroad, C. E., et al. (2002b). Rapid communication: Thirty-eight polymorphic microsatellite markers for mapping in rainbow trout. Journal of Animal Science 80: 541-542.

Rexroad, C. E., et al. (2006). The NCCCWA genetic map for rainbow trout. San Diego, CA, USA, Plant & Animal Genome XIV, the international conference on the status of plant and animal genome research. Conference Book of Abstracts.

Rexroad, C. E. and Palti, Y. (2003). Development of ninety-seven polymorphic microsatellite markers for rainbow trout. Transactions of the American Fisheries Society 132: 1214-1221.

Rexroad, C. E., et al. (2005). Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). BioMed Central Genomics 6: 54.

Rithidech, K. N., et al. (1997). Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms. Biotechniques 23: 36-44.

Schneider, S., et al. (2006). Arlequin: a software for population genetics data analysis, version 3.01, Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.

Silverstein, J. T., et al. (2004). Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). Aquaculture Research 35: 40-48.

Vandeputte, M., et al. (2006). An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes 6: 265-267.

Weir, B. S. and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Wright, S. (1965). The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution 19: 395-420.

Table 1. Microsatellite marker information for the 12 loci used in the two multiplex systems.

Hexaplex I	GenBank Acc. No.	Primer sequences (5' to 3')	Dye label	Primer conc. (μM)	T_a (ºC)	$MgCl_2$ (m*M*)	*PIC*[a]
OMM5132[b]	**BV211863**	F: CAACTCTAGTCCATGCCA R: CGAGTGCGTATGCATGTA	Fam	1.67	60	1.5	0.62
OMM5047[b]	**CA349175**	F: ACTTTCAGCAGCATCTGGTCA R: CCTGGTCCTCAGCGTTCAT	Fam	0.67	58	2	0.79
OMM1303[b]	**G73543**	F: GGAAGGAAAGGCACTT R: TCTACACCAGGAGAGAGTAAT	Hex	2.33	54	2	0.80
OMM5007[b]	**CO805113**	F: AGATGCCTGTCGAGTGTTG R: GAGGAGCATCATTTAGAGACTACA	Hex	0.33	58	2	0.75
OMM5233	**BV211954**	F: GGATCTCGCATAAGTTCA R: GACAGGGAATTGTTGCAT	Ned	0.33	58	2.5	0.71
OMM1008	**AF346670**	F: GATCCTTTGGGAGATTAACAG R: CACCACAGTTGCTACTGCC	Ned	0.67	58	2	0.67
Hexaplex II							
OMM5177	**BV211903**	F: GCTGTCTGCGCTAGAATC R: CAGAGCCCTATGCCAAAC	Fam	0.67	58	2.5	0.75
OMM1051	**AF346695**	F: CCTACAGTAGGGATTAACAGC R: CATGCCCACACATTACTAC	Fam	2.00	58	2	0.84
OMM5264[b]	**BV211979**	F: AAGCATCATTGCCGTGAC R: TCTCTTCGCAGGGATTCT	Hex	1.33	58	1.5	0.71
OMM1097	**AF352763**	F: CTAGCCATCCGAACACTG R: AGAATAGGGTGCCTGTATCTC	Hex	1.67	58	2	0.88
OMM1088[b]	**AF352757**	F: CTACAGGCCAACACTACAATC R: CTATAAAGGGAATAGGCACCT	Ned	0.67	58	2	0.79
OMM1325	**G73562**	F: TCTCTGCCAATGTGACATGCCT R: TAACTATCACTGCCACTCCTCGTG	Ned	1.00	58	2	0.46

[a] Polymorphic information content calculations are based on genotypes from the Troutlodge, Inc. population (Palti et al., 2006).
[b] See text for suggested modifications to protocol if necessary to remove any of these markers.

Table 2. Number of samples genotyped, polymorphic information content (*PIC*), allele size range, and total number of alleles for the NCCCWA 2002-2005 brood-years for two microsatellite marker sets.

9 markers	NCCCWA 2002				NCCCWA 2003				NCCCWA 2004				NCCCWA 2005			
	n	*PIC*	Allele range	No. alleles	*n*	*PIC*	Allele range	No. alleles	*n*	*PIC*	Allele range	No. alleles	*n*	*PIC*	Allele range	No. alleles
OMM1007	69	0.74	151-179	7	157	0.73	151-179	8	131	0.7	151-179	7	139	0.76	151-179	8
OMM1013	67	0.83	254-350	14	156	0.82	254-353	18	131	0.81	254-329	12	141	0.89	254-329	16
OMM1019	66	0.76	173-191	9	159	0.74	181-197	8	132	0.72	181-193	7	139	0.78	181-197	8
OMM1020	61	0.63	178-216	8	158	0.51	196-216	6	133	0.53	192-216	7	144	0.61	196-216	6
OMM1021	61	0.74	191-223	11	145	0.82	191-219	10	112	0.62	191-219	9	122	0.81	191-219	10
OMM1036	69	0.79	226-290	14	157	0.82	226-302	17	135	0.79	226-286	11	144	0.84	226-302	17
OMM1039	69	0.73	129-153	11	160	0.83	125-163	15	134	0.71	129-155	10	143	0.85	113-159	15
OMM1056	66	0.88	97-157	15	157	0.85	95-157	15	135	0.86	93-157	16	143	0.84	95-157	16
OMM1050	69	0.91	257-353	18	159	0.92	257-353	25	134	0.91	257-353	16	141	0.93	257-353	23
Average	66.3	0.78	-	11.9	156.4	0.78	-	13.6	130.8	0.74	-	10.6	139.6	0.81	-	13.2
Multiplex																
OMM1008	63	0.55	256-284	7	156	0.67	256-284	8	134	0.61	256-284	7	139	0.68	256-284	8
OMM1051	62	0.86	219-283	16	126	0.88	219-283	17	102	0.86	221-279	15	139	0.83	221-283	16
OMM1088	66	0.8	97-153	14	146	0.86	113-187	16	116	0.81	97-155	18	141	0.83	99-181	16
OMM1097	63	0.86	231-319	15	136	0.85	213-319	17	117	0.91	219-319	22	142	0.78	221-323	15
OMM1303	60	0.76	278-362	12	156	0.86	278-362	16	126	0.79	278-362	13	140	0.88	260-362	17
OMM1325	64	0.53	278-296	5	132	0.43	254-296	6	125	0.58	268-300	8	139	0.59	248-296	5
OMM5007	64	0.84	152-198	16	154	0.91	152-198	21	126	0.83	150-198	17	141	0.88	150-198	19
OMM5047	63	0.76	260-280	8	155	0.79	260-280	9	132	0.77	260-280	8	139	0.78	260-280	9
OMM5177	67	0.68	114-141	7	152	0.68	96-144	12	122	0.75	96-141	9	140	0.75	114-144	11
OMM5233	62	0.86	116-138	11	156	0.86	112-138	14	129	0.85	114-138	12	139	0.84	112-138	13
Average	63.4	0.75	-	11.1	146.9	0.78	-	13.6	122.9	0.78	-	12.9	139.9	0.78	-	12.9

Table 3. Pairwise F_{st} values for multiplex marker set (above diagonal) and previous marker set of nine microsatellite loci (below diagonal) for four brood-years of rainbow trout broodstock.

Populations	NCCCWA 2002	NCCCWA 2003	NCCCWA 2004	NCCCWA 2005
NCCCWA 2002	-	**0.0417**	0.0000	**0.0253**
NCCCWA 2003	**0.0168**	-	**0.0481**	**0.0273**
NCCCWA 2004	0.0066	**0.0232**	-	**0.0228**
NCCCWA 2005	**0.0164**	**0.0095**	**0.0181**	-

Values in bold denote statistical significance ($p < 0.0001$).

Table 4. Estimated posterior probability of *K* clusters [Pr(*K*|*X*)][a] for nine-marker set loci and ten multiplex (MP) loci typed in BY2004[b] and BY2005[c] samples.

K	Nine loci - BY2004		MP - BY2004		Nine loci - BY2005		MP - BY2005	
	Ln Pr(*X*|*K*)	Pr(*K*|*X*)	Ln Pr(*X*|*K*)	Pr(*K*|*X*)	Ln Pr(*X*|*K*)	Pr(*K*|*X*)	Ln Pr(*X*|*K*)	Pr(*K*|*X*)
1	-934.6	0.039149	-1116.3	0.000000	-1857.6	0.000000	-2062.4	0.000000
2	-1020.8	0.000000	-1077.1	0.845532	-1760.9	0.000000	-1987.4	0.000000
3	-939.1	0.000435	-1078.8	0.154465	-1707.1	0.845534	-1906.5	0.000000
4	-931.4	0.960416	-1089.6	0.000003	-1708.8	0.154465	-1884.2	1.000000
5	-960.2	0.000000	-1108.5	0.000000	-1721.7	0.000000	-1898.8	0.000000
6	-975.6	0.000000	-1130.7	0.000000	-1725.2	0.000000	-1921.3	0.000000
7	-1017.1	0.000000	-1151.6	0.000000	-1737.7	0.000000	-1939.0	0.000000
8	-1031.8	0.000000	-1177.3	0.000000	-1752.6	0.000000	-1963.3	0.000000
9	-1050.4	0.000000	-1190.0	0.000000	-1769.9	0.000000	-1988.9	0.000000
10	-1074.1	0.000000	-1213.1	0.000000	-1789.0	0.000000	-2013.7	0.000000

[a] The STRUCTURE program was run using an admixture model. These results are based on runs of 10^6 iterations, following a burn-in period of 10^5 iterations and a thinning value of 10.
[b] Sample size of N=30 unrelated individuals for BY2004.
[c] Sample size of N=52 unrelated individuals for BY2005.

Figure 1. Bar-plots based on F_{is} estimates (± *S.E.*) comparing the ability of both microsatellite marker sets for detecting relatedness among individuals within the NCCCWA brood-years 2002-2005.

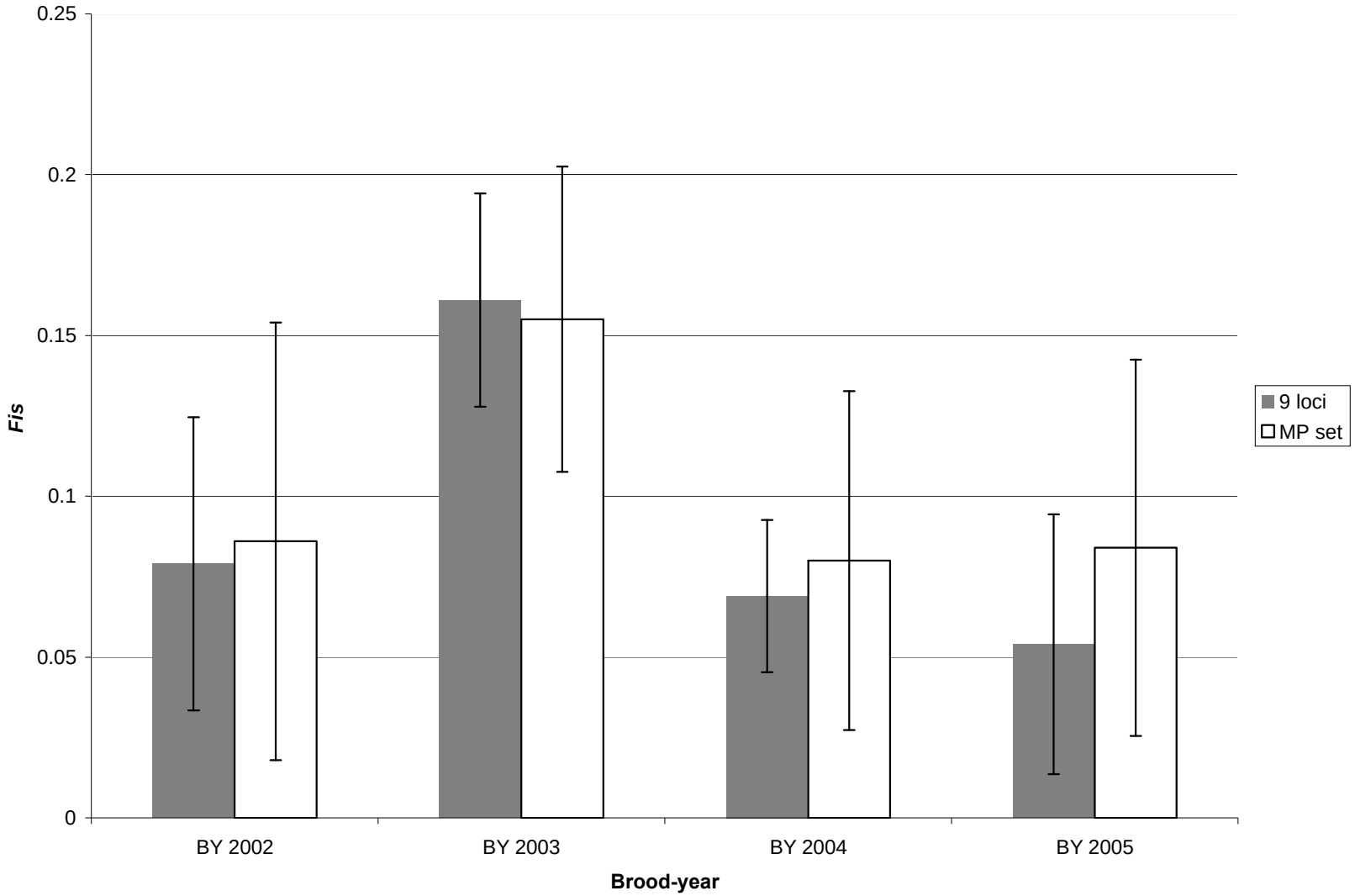


Figure 2. An unrooted neighbor-joining tree showing relationships inferred using the –ln(*PSA*) distance method among unrelated individuals within the NCCCWA 2005 brood-year based on genotypic data from ten multiplexed loci. Major groupings are indicated by numbers and separated by single- and double-line breaks. Note the distribution of Trout Lodge (TL) strain groups below the single-line break (1), University of Washington (UW) between the single- and double-line breaks (2), and Shasta (SH) to the right of the double-line break (3).

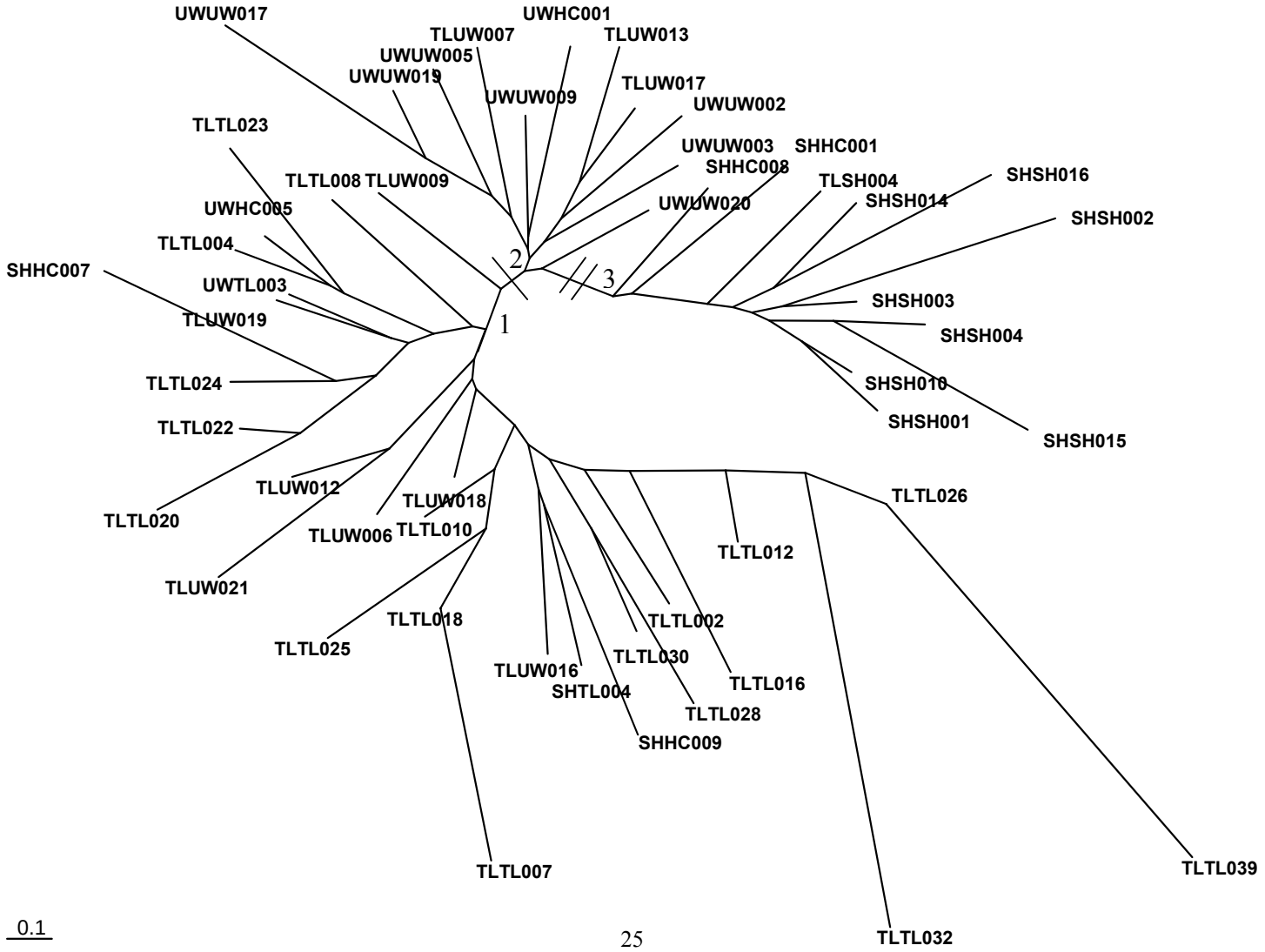
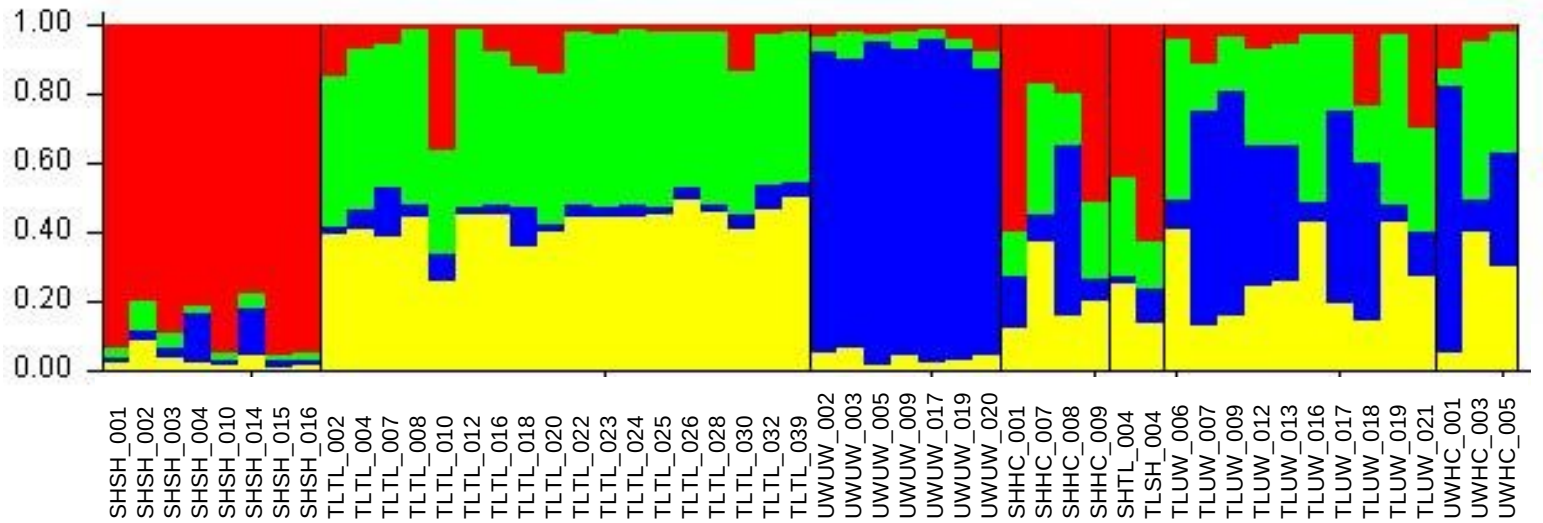


Figure 3. Results of assignment tests performed with STRUCTURE on a sample of unrelated individuals from the NCCCWA 2005 brood-year based on genotypic data of ten multiplex loci. The proportion of membership (*y*-axis) assigned to the inferred genetic clusters is indicated by bars for each individual (*x*-axis). Red, green and yellow, and blue bars denote Shasta (SH), Trout Lodge (TL), and University of Washington (UW) membership, respectively. Individuals of mixed ancestry, including those of House Creek (HC) descent, showed an admixture of membership between the different strains.



Chapter 2

Association of major histocompatibility genetic markers with resistance to bacterial coldwater disease in rainbow trout (*Oncorhynchus mykiss*)

Abstract

Individuals from the 2005 NCCCWA brood-year (75 full-sib families) were challenged with *Flavobacterium psychrophilum*, the causative agent of bacterial coldwater disease (BCWD). The overall mortality rate was 70%, with large variation among families. Resistance to the disease was assessed by monitoring post-challenge days-to-death. Phenotypic variation and additive genetic variation were estimated by developing mixed models of survival data. Microsatellite markers were previously isolated from BAC clones that harbor genes of interest and mapped onto the rainbow trout genetic linkage map. A general relationship between *UBA* gene sequence types and *MH-IA*-linked microsatellite alleles showed that microsatellites mapped near or within specific major histocompatibility (MH) loci are likely to represent sequence variation of MH genes. The parents and grandparents of the 2005 brood-year families were genotyped with markers linked to the four MH genomic regions (*MH-IA*, *MH-IB*, *TAP1*, and *MH-II*) to assess linkage disequilibrium (LD) between those genomic regions and resistance BCWD. Family analysis suggested that *MH-IB* and *MH-II* markers are linked to BCWD resistance. Tests for disease association at the population level substantiated the involvement of *MH-IB* with disease resistance. The impact of MH sequence variation on selective breeding for disease resistance is discussed in the context of aquaculture production.

1. Introduction

The gram-negative bacterium *Flavobacterium psychrophilum* (*Fp*) poses major fish health concerns in trout and salmon hatcheries around the world and is documented to cause mortalities >90% (Cipriano and Holt, 2005). The disease caused by this bacterium is commonly called bacterial coldwater disease (BCWD) in North America or rainbow trout fry syndrome (RTFS) in most European countries, among other names, such as fry mortality syndrome or visceral myxobacteriosis (Cipriano and Holt, 2005). It particularly affects juvenile fish before immune system development, which limits the use of vaccination as a control or treatment. Common symptoms include external lesions, darkening of skin, dorsal swelling, and erratic swimming movements. *Fp* is an opportunistic bacterium transmitted between fish by waterborne exposure and direct contact; abrasion of the skin enhances invasion of the pathogen among fish during disease challenges. In the absence of approved, cost-effective vaccination methods, selective breeding of disease resistant lines is a promising solution (Henryon et al., 2005).

Microsatellites, or simple sequence repeats (SSRs), are genetic markers widely used for a variety of applications. In aquaculture research, they are commonly used in constructing genetic linkage maps and for assessing linkage to quantitative traits for their mapping to a genomic region. Microsatellite markers are co-dominant, which allows scoring of heterozygous genotypes. They are evenly distributed throughout the genome of most animals and plants, and can be surveyed rapidly in many individuals using the polymerase chain reaction (PCR) technique. They are also highly polymorphic, which enables researchers to observe genetic variation in their respective genomic locations.

Genes encoded by the major histocompatibility complex (MHC) are obvious candidates for investigation following bacterial challenges since they are known to play an important role in immune response. Molecules encoded by the human MHC class I and II regions are cell-surface integral membrane glycoproteins responsible for detection and elimination of pathogenic cells. They bind antigens and present them to the adaptive and cellular immune systems, which are essential components of the vertebrate response against invading pathogens. The gene products of the MHC are necessary for the

presentation of antigenic self and non-self peptides to the T-cell receptor-bearing cells produced in the thymus.

MHC class I proteins play a major role in the recognition of intracellular antigens and are expressed on the surface of most cell types. The classical MHC I molecules (Ia) present antigen fragments synthesized inside the cell initiating cellular responses by CD8 T-cells. The major class I gene in rainbow trout, *UBA*, is located in the *MH-IA* locus and encodes the major class I proteins (Kiryu et al., 2005). Class I molecules are composed of three heavy chains (α_1, α_2, and α_3) and a light chain encoded by the non-MHC β_2-microglobulin gene (β_2m) (Elgert, 1996). The *MH-IB* locus in trout is the region encoding non-classical Ib genes (Phillips et al., 2003; Shiina et al., 2005). *TAP1* (aka *ABCB2*) plays a major role in transportation of foreign peptides through the endoplasmic reticulum for processing before binding with β_2m (Elgert, 1996). In rainbow trout, *TAP1* is not linked to other MH genomic regions (Palti et al., 2007; Phillips et al., 2003).

MHC class II molecules present foreign peptides, which are phagocytosed and processed within the host cell and initiate a response by the CD4 T cells. The class II molecules are heterodimers consisting of two different chains (α and β) whose structure forms an antigen-binding groove capable of accommodating a variety of peptides recognized by differences in length and sequence (Elgert, 1996). They are encoded by class IIa and IIb genes and are expressed on the cell surface of professional antigen-presenting cells in mammals.

Class I and II MHC genes are linked in all vertebrates except teleosts (Hansen et al., 1999; Phillips et al., 2003); therefore, MH is the best description of this genomic region in bony fishes. The functional aspects and evolutionary significance of this genomic architecture are not well understood (Hansen et al., 1999). In recent years, much of the research effort on teleost MH genomics has been directed toward comparative studies of the genomic structure of the teleost MHC class I region, which identified a high degree of conserved synteny (Clark et al., 2001; Matsuo et al., 2002; Michalova et al., 2000; Phillips et al., 2003). Interestingly, most of the genes that are involved directly in the classical (Ia) antigen presentation pathway, including *PSMB8*, *PSMB9*, *PSMB9*-like, *PSMB10*, *ABCB3* and *TAPBP*, are linked in the fish MH class I genomic region, forming the class I "core" region, with the exception of *ABCB2* (Palti et al., 2007; Philips

et al., 2003; Shiina et al., 2005). From the structural viewpoint, however, there is nothing in common among the classical MH (class Ia), *PSMB*s and *ABCB2* and *ABCB3* molecules. Therefore, the assumption underlying most comparative studies involving the MHC is that the clustering of these genes is the result of selection pressure requiring co-evolution and co-regulation of the MHC-related genes

Phillips et al. (2003) previously demonstrated that the rainbow trout MH class I core region is duplicated to two different chromosomes, that *TAP1* maps to a separate chromosome, and that the class II region is localized to yet another chromosome. The extent of the class I duplication event was determined by Shiina et al. (2005) via BAC sequencing of the core regions for the class IA and IB regions. More recently, it was determined that *TAPBP* is part of the rainbow trout class Ia core region, and that the rainbow trout tapasin-related gene (*Onmy-TAPBP-R.1*) and *TAP1* are co-localized on the short arm of chromosome 2, while *Omny-TAPBP-R.2* is located on chromosome 3, representing a fifth MH chromosomal region in rainbow trout (Landis et al., 2006).

In this study, I tested the hypothesis that variation of microsatellite markers linked to MH genes in rainbow trout is associated with variation in resistance to BCWD caused by *Fp*. Variability in the *UBA* gene sequence was compared to genotypic variation at *MH-IA* microsatellites to confirm the existence of a linkage relationship between gene sequence and microsatellite alleles. With the knowledge of families with high or low survival rates after bacterial challenge, I screened for correspondence between MH genes and resistance to BCWD, using microsatellite markers linked to four particular MH gene regions (*MH-IA*, *MH-IB*, *TAP1*, and *MH-II*).

2. Materials and methods

2.1. Fish and bacterial challenge

All rainbow trout used in this study were located at the USDA-ARS-National Center for Cool and Coldwater Aquaculture (NCCCWA) facility in Leetown, WV. Founding strains contributing to the NCCCWA 2003 brood-stock (G_0) were: 1) University of Washington, Donaldson, 2) Troutlodge, Inc. Kamloops/Puget Sound steelhead cross, 3) College of Southern Idaho, House Creek, 4) Arlee Strain, originating

from Montana, 5) Ennis National Fish Hatchery Shasta strain and 6) Kamloops strain originating from the Kamloops region in British Columbia, Canada. The parents were determined to be free of viral hemorrhagic septicemia (VHS), infectious hematopoietic necrosis virus (IHN), infectious pancreatic necrosis virus (IPNV), and bacterial kidney disease (BKD). Paired-mating of G_0 fish supplied the focal populations for this experiment, who were hatched in 2003 (G_1) and 2005 (G_2). A schematic illustration of the experimental design is shown in Figure 1.

The 75 full-sib G_2 families evaluated in this study were produced from 56 sires and 65 dams. G_2 fish were challenged by intraperitoneal (IP) injection of approximately 8.8 x 10^6 CFU/fish *Flavobacterium psychrophilum* of the strain CSF-259-93. Approximately 40 fish each from all 75 families were challenged, and 45 of the families were replicated, for a total of 120 lots and about 4,800 fish. The disease trial began on May 16, 2005 and lasted for 21 days. Fish were kept at 13.0±0.5 C under flow-through conditions and fed a standard daily diet. Average weight of fish on May 9, 2005 was 2.8 grams, and the standard tank volume was 2,670 mL. Dead fish were removed from the tanks daily post-injection and examined for symptoms of BCWD. The number of mortalities was recorded daily from each tank. After day 21 post-injection, surviving fish were counted and euthanized by lethal dose of MS-222 at 200 mg/liter. Fin tissues from all grandparents (G_0), parents (G_1), and progeny (G_2) were stored individually in 95% ethanol prior to genomic DNA isolation following methods described in Palti et al. (2006). Additional G_2 fin material was stored separately in RNA*later* (Applied Biosystems, Foster City, CA) at -20ºC pending RNA isolation for cDNA synthesis.

2.2. Survival analysis and model selection

Resistance was evaluated as a longitudinal trait. The number of days from challenge to death for the *i*th fish was recorded, and average days until death (*ADTD*) values were calculated for each family. It is assumed that the longer a fish survives, the greater its resistance. Fish that did not die during the 21 day challenge were recorded as a censored observation.

Mortality was presented as the number of fish in a family that died within 21 days of injection. The rate of mortality is presented by plotting the Kaplan Meier survival function for the 21 days following challenge.

Proportional hazard models often assume that failure times are expressed on a continuous time scale (Cox, 1972; Kalbfleisch and Prentice, 2002). However, there are experimental situations where this assumption is violated. For example, survival after challenging experimental animals with a disease-causing pathogen is evaluated for relatively short periods of time (e.g., time <20 days). In these cases, the time scale is discrete with very few classes. For this discrete data, a survival analysis using standard proportional hazard models is *a priori* incorrect, as this approach assumes continuity of the baseline hazard distribution and/or absence of ties between ordered failure times.

In this study, there were many ties between ordered failure times because the survival data were collected for 21 days after exposure to the disease-causing pathogen. Furthermore, when I plotted the natural logarithm $\ln(-\ln[S(t)])$ against $\ln(t)$, where $S(t)$ is the Kaplan-Meier survivor function, the plot of this relationship did not produce a straight line, which suggested that the Weibull distribution was not a good fit for the survival data. Thus, I decided to use the "grouped data model" of survival analysis (Prentice and Gloeckler, 1978). The Prentice and Gloeckler model is a fully parametric exponential model for which the baseline is estimated at every discrete time-point and for which the definition of hazard is modified to take into account the discrete time scale without rejecting the proportional hazard model (Ducrocq, 1999).

The survival times of *Fp*-challenged animals were described with a discrete proportional hazards model (Cox, 1972; Prentice and Gloeckler, 1978) for which the hazard function $h(\tau_i;\mathbf{x}_j)$ within the interval $\tau_i = (t_{i-1}, t_i)$ of a particular animal *j*, characterized by a vector of explanatory $\mathbf{x}_j$ variables, can be written as:

$$h(\tau_i;\mathbf{x}_j) = h_0(\tau_i)\exp\{\mathbf{x}_j\boldsymbol{\beta}\}$$

where $h_0(\tau_i)$ is the baseline hazard function and $\boldsymbol{\beta}$ is a vector of regression coefficients.

An extension of this model includes the addition of random effects (Ducrocq and Casella, 1996) and this mixed model can be written as:

$$h(\tau_i;\mathbf{x}_j,\mathbf{z}_j) = h_0(\tau_i)\exp\{\mathbf{x}_j\boldsymbol{\beta} + \mathbf{z}_j\mathbf{u}\}$$

where $\mathbf{u}$ is a vector of random variables and $\mathbf{z}_j$ is the incidence matrix.

The fitting of linear mixed-effects models and the assessment of the statistical significance of the random effects such as replication, tank, and the average family weight covariate on the response variable SURV was performed via restricted maximum likelihood (REML) algorithms and likelihood ratio tests implemented in the R Statistical Package from the Bioconductor Software Project (Gentleman et al., 2004). To perform these analyses, we developed an R script using statistical methods outlined by Pinheiro and Bates (2000); this R script is available for distribution from the authors. Akaike information criterion (*AIC*), Bayesian information criterion (*BIC*), and the fixed and random variables with significant effects and models with best log likelihood were considered for use in the survival analysis.

Bayesian analysis of frailty models developed by Ducrocq and Casella (1996) can be applied to the Prentice and Gloeckler's model for discrete data. This method and the survival analysis were implemented using the SURVIVAL KIT version 3.12 program (Ducrocq and Solkner, 1998) and the following Sire-Dam model:

$$h(\tau_i; \mathbf{x}_j, \mathbf{z}_{1j}, \mathbf{z}_{2k}, \mathbf{z}_{3l}) = h_0(\tau_i) \exp\left\{\mathbf{x}_j^{'}\boldsymbol{\beta} + \mathbf{z}_{1j}^{'}\mathbf{s} + \mathbf{z}_{2k}^{'}\mathbf{d} + \mathbf{z}_{3l}^{'}\mathbf{r}\right\}$$

Where $\mathbf{s}$ is the vector of sire genetic effects, $\mathbf{d}$ is the vector of dam genetic effects, $\mathbf{r}$ is the vector of random replication/family effects, and $\mathbf{z}_{1j}, \mathbf{z}_{2k}$ and $\mathbf{z}_{3l}$ are incidence matrices.

The breeding values (*BV*) of sires and dams were estimated using the Bayesian approach outlined by Ducrocq and Casella (1996). Their conversion to components of genetic variance was performed done using standard expressions of genetic analysis (Falconer and Mackay, 1996; Lynch and Walsh, 1997).

2.3. Microsatellite markers

All microsatellites were screened for duplication, null alleles, and genotyping errors. Microsatellites from both IA and IB regions were considered duplicated when more than two alleles for several individuals amplified. Additionally, the highly polymorphic nature of the *MH-IA* region led to the high occurrence of null alleles. Null alleles were first detected when amplification success rates were low for multiple

individuals after three PCR attempts; suggesting the presence of double-null genotypes in a particular marker. The DNA quality of these samples was verified by observing successful amplifications for other markers. The second criterion for a null allele relied on the observation of a heterozygote deficit, as individuals heterozygous for a detectable allele and a null allele would be called as homozygous for the detectable allele. Finally, pedigree information was used to align parental and progeny genotypes to confirm Mendelian inheritance and to reduce the probability of genotyping errors.

Ten qualifying microsatellite markers representing four distinct MH gene regions were utilized in this study to genotype G_0 (*N*=42) and G_1 (*N*=121) fish (Table 1). G_2 fish representing 18 families (*N*=90) were genotyped with *MH-IA* microsatellites to confirm a relationship with *UBA* sequence variation. Marker *OMM3084_new* was previously mapped to intron 2 between α1 and α2, and marker *UBA_R1_3* was located within the 3'-UTR of *UBA* (Shiina et al., 2005). Marker *OMM3089* was the only repeat that was free of duplication and null alleles isolated by sequencing BACs containing *MH-IB* genes (Hansen and Palti, unpublished data). Both *OMM1189* (Rexroad and Palti, 2003) and *BHMS429* (Thorsen et al., 2005) were used to supplement screening of this region because they were found within the same linkage group flanking *OMM3089* by < 5cM on each side (Rexroad et al., 2006). The pair of microsatellites representing the *TAP1* region (*OMM3079* and *OMM3080*) were designed from a sequence previously reported by Palti et al. (2007). Three microsatellites (*OMM3024*, *OMM3026*, and *OMM3028*) were isolated from the partial sequence of a BAC containing *MH-II* genes following the methods described by Rodriguez et al. (2006). The quality of *OMM3028* and *OMM3079* genotypes was questionable due to alleles separated by only one base-pair in the population, and both were removed from the statistical analysis to reduce errors. Johnson et al. (2007) had identical genotyping difficulties with this population using a different set of microsatellites.

PCR reactions included 25 ng of template DNA, 1.5 to 2 mM $MgCl_2$, 10X reaction buffer, forward primers (directly labeled with NED or HEX and tail-labeled with FAM) and reverse primers, along with 0.05 units *Taq*. The tailed reactions followed the protocol described by Rodriguez et al. (2003). Directly labeled reactions were thermocycled as follows: 15 min at 95°, 1 min at 94°, annealing temperatures for 45 sec,

29 cycles of 45 sec at 72°, 10 min at 72°, 1 hr at 4°, and 12° hold. Tailed reactions were thermocycled as follows: 15 min at 95°, 1 min at 94°, annealing temperatures for 45 sec, 29 cycles of 45 sec at 72°, 10 min at 72°, 1 min at 94°, 45 sec at 53°, 45 sec at 72° for 5 cycles, 10 min of 72°, 1 hr at 4°, and 12° hold. PCR reactions were conducted in MJ Research DNA Engine thermal cyclers (MJ Research, Waltham, MA). An ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA) was used to genotype the PCR reactions. Samples were prepared for this analysis by the following means: dilution plates were made, using 3 µl PCR products and 20 µl water. One µl of this dilution was added to 12 µl Hi Di formamide (Applied biosystems, Foster City, CA) and 0.12 µl Rox size-standard. The DNA samples then were denatured at 95°C for 5 min and placed on ice prior to loading onto the ABI 3730 Genetic Analyzer. Successful amplifications were analyzed using GeneMapper 3.7 (Applied Biosystems).

2.4. Haplotype designation

Pairwise linkage disequilibrium (LD) tests of Slatkin and Excoffier (1996) were conducted in Arlequin 3.01 (Schneider et al., 2006) following 100,000 random permutations for two sets of microsatellite markers. The eight MH microsatellite markers were tested for LD within a set of 51 unrelated G_1 individuals to determine whether markers linked to a particular MH region showed linkage. Microsatellite haplotypes were identified for specific MH regions using PHASE 2.0 following methodologies described in Stephens et al. (2001) and Stephens and Scheet (2005). A previous study by Johnson et al. (2007) detected subpopulation structure in the G_1 population; therefore LD also was tested using those eight "unlinked" markers to account for base-line LD in the population.

2.5. UBA (MH-Ia) gene typing

The relationship between *MH-IA* gene variation and *MH-IA* marker haplotypes was assessed using the *UBA* lineage classification of Kiryu et al (2005). A total of 41 individuals from 13 G_2 families representing all *MH-IA* marker haplotypes in this population were selected. Fin clips that were stored in RNA*Later* were chopped and total RNA was isolated following the TRI-reagent (Sigma-Aldrich Corp., St. Louis, MO) modification of the guanidine isothiocyanate/phenol-chloroform method (Chomczynski

and Sacchi, 1987). The RNA was dissolved in 20-50 µL of nuclease-free water and subjected to DNAse treatment following the manufacturer's protocol (DNAse RQ-1, Promega, Madison, WI) to remove any genomic DNA contamination. The DNAse treatment was followed with a re-extraction with TRI-Reagent to remove all residual DNAse activity. The quantity of RNA was estimated using spectrophotometry (ND-1000, NanoDrop Technologies, Wilmington, DE). The quality of the RNA was assessed by agarose gel electrophoresis with the visualization of the 28S and 18S rRNA bands. All RNA samples were stored at -80ºC.

Complementary DNA (cDNA) was synthesized using two micrograms of total RNA mixed with 1 µg of random hexamer primers (Promega, Madison, WI) and heated at 70ºC for 5 min. This mixture was cooled on ice before addition of the reverse transcription cocktail containing 1 x RT Buffer, 0.5 mM of each dNTP, 25 U of rRNAse inhibitor (Promega, Madison, WI) and nuclease-free water to a total reaction volume of 25 µL. First strand cDNA synthesis was carried out at 37ºC for 60 min followed by heating the sample to 95ºC for 5 min. The PCR strategy of Kiryu et al. (2005) was used to amplify and type the *UBA* cDNA sequences in the population. PCR products were separated on 3% agarose gels stained with ethidium bromide. Gel images were visualized and photographed on an Alpha Innotech ChemiImager 4400 (Alpha Innotech Corporation, San Leandro, CA).

2.6. Statistical analyses

Markers from each of the MH gene regions were analyzed both independently and as haplotypes using *ADTD* (phenotypic variation) or *BV* (genetic variation). Family-based linkage and association tests adjusting for population admixture were implemented using FBAT (Laird et al., 2000) across 21 G_1 families, including 42 G_0 parents and 71 G_1 progeny. The power of the family-based association test statistic ((Laird et al., 2000) was calculated by numerical integration (precision of 0.01) and 100,000 Monte-Carlo simulations using PBAT (Lange et al., 2003). Analysis of variance (ANOVA) was conducted in SAS 8.0 (SAS Institute Inc., 2000) using proc GLM to test for associations between the MH genotypes and each trait measurement (*ADTD* or *BV*) among a subset of unrelated G_1 individuals (N=51) that was re-sampled twice.

3. Results

The average mortality rate 21 days post-*Fp* infection in the G_2 population was 70%. The range of individual family mortality was 28% to 99%. Family-based variation in response to infection by *Fp* is illustrated in the Kaplan Meier survival function plot (Figure 2). The large variation in family-based response implies a large genetic component to immune responsiveness as described in greater detail by Silverstein et al. (in preparation).

I identified 28 *MH-IA*, 46 *MH-IB*, and 3 *MH-II* marker haplotypes among the 71 G_1 samples from the family analysis of disease association. Linkage disequilibrium between markers was estimated to support the use of marker haplotypes in the disease resistance association analysis. Tests for LD within the two subsets of 51 unrelated G_1 individuals using a random set of unlinked microsatellite markers resulted in *P* values as low as 0.0002 (Table 2); requiring compensation for effects due to population substructuring that were detected by Johnson et al. (2007). Therefore, I applied a minimum threshold of $P < 0.00002$ to the MH markers LD analysis to reduce the effect of the sub-population structure on LD estimates. Significant LD was found using that threshold between marker pairs within *MH-IA* and *MH-II*, and between two of the three *MH-IB* markers pairs (Table 3).

A general relationship was observed between *UBA* sequence variation and *MH-IA* marker haplotypes (Table 4). Nine *UBA* types and sixteen *MH-IA* marker haplotypes were identified in the 41 G_2 fish sampled. A qualitative comparison showed that most *UBA* types were represented by more than one *MH-IA* marker haplotype.

The small population sample for the G_1 family design resulted in reduced power for the statistical analysis to examine associations between MH markers or marker-haplotypes and *Fp* resistance. Using numerical integration or Monte-Carlo simulations, the PBAT power calculation of the family-based association test statistic was approximately 5%. Therefore, I considered all associations with *P* value ≤ 0.10 to be suggestive (Table 5). Disease resistance measured as *ADTD* was linked to *OMM3026* (P=0.048) and *OMM3089* (P=0.048) after employing a constant offset (11.576) to

normalize the data around the mean. Linkage to *OMM3089* was detected with *BV* (*P*=0.052), and the *P*-value increased to *P*=0.081 when including an offset of 0.046.

Association analysis results were similar for both unrelated subsets, indicating that efforts to identify sampling bias by re-sampling were effective (Table 6). *OMM1189* showed association with disease resistance measured by *ADTD*, and *MH-IB* region haplotypes showed association to both *ADTD* and *BV*. Three marker haplotypes were tested (*MH-IA*, *MH-IB*, and *MH-II*); therefore, the significant *P*-value following Bonferroni's method (Milliken and Johnson, 1984) to control for type I error in multiple comparisons between haplotypes is 0.0017. However, all associations with *P* value ≤ 0.10 are highlighted and were considered suggestive due to power limitations caused by a limited number of unrelated fish in the G_1 population.

4. Discussion

One of my research goals is to identify genes in rainbow trout associated with resistance to *Flavobacterium psychrophilum* (*Fp*), the causative agent of bacterial cold-water disease (BCWD). BCWD has a devastating economic impact on trout and salmon farming worldwide, and the identification of genes or DNA markers that are associated with resistance can aid a selective breeding program for improved BCWD survivability. I found evidence that DNA markers and haplotypes associated with survivability to BCWD could be identified using microsatellites, despite the low power of the experimental design, which suffered from small population size, which limited the detection likelihood to loci with major effects. The family analysis revealed an *MH-IB* marker that may be linked to BCWD survivability as measured by both *ADTD* and *BV* and *MH-II* marker and haplotypes that may be linked with *ADTD*. Tests for association at the population level substantiated the involvement of *MH-IB* with *ADTD* and *BV*. More significant *P*-values for *ADTD* are likely due to differences in calculations for *ADTD* and *BV*. *ADTD* captures the phenotypic variation of the trait, while *BV* represents the genetic variation and excludes common environment effects (e.g. higher pathogen load in tanks of susceptible families, which is likely to cause overrepresentation of the family effect on

phenotypic variation). In addition, *BV* estimates were adjusted for average family weight and survivorship, which was treated as a censor in the model I used.

In humans, microsatellites offer an economic alternative for typing of MHC class II alleles (Lee et al., 2006). Previous studies have reported associations between MH loci and disease resistance in salmonids (Grimholt et al., 2003; Langefors et al., 2001; Miller et al., 2004; Ozaki et al., 2001; Palti et al., 2001). To our knowledge, this is the first study that used microsatellite markers and haplotypes from the four main genomic regions known to contain MH genes to examine associations with disease resistance in salmonids and the first study that explored the relationship between *UBA* gene types and *MH-IA* microsatellite marker haplotypes in rainbow trout. I hypothesized that microsatellites mapped near or within specific MH loci would represent sequence variation of MH genes. My results support this hypothesis, as I detected a relationship between *UBA* gene sequence types and *MH-IA*-linked microsatellite alleles, although a strict relationship was not observed. The lack of a more strict relationship in the *MH-IA* region is likely in large part due to the unusually high rate of exon shuffling caused by a recombination "hot-spot" in *UBA* intron 2 (Kiryu et al., 2005). The limited options in microsatellite selection due to MH-I duplication in salmonids, high variability of *MH-IA*-linked microsatellites and occurrence of null alleles complicated my ability to define a strict relationship with *UBA* types. I used the *UBA* typing approach, as it is the major MHC class I molecule in rainbow trout, and a well defined typing system is available (Kiryu et al., 2005). However, a more strict relationship may be found between microsatellites and genes in less polymorphic regions (e.g. *MH-II*) or even between microsatellites and less polymorphic genes within the *MH-IA* region.

We challenged young rainbow trout, because BCWD has the largest effect on juvenile fish, which lack immune memory responses. Therefore, my findings of association between *MH-IB* markers and BCWD survivability and the lack of correlation with *MH-IA* markers are not surprising. Interestingly, similar results were found for viral disease resistance in Atlantic salmon and rainbow trout (Miller et al. 2004; Ozaki and Dijkstra, personal communication, respectively) implicating a common mechanism of innate immunity. Although the role of the Ib molecules in the fish immune response is largely unknown, their functions in mammals include non-specific response activities

such as regulation of natural killer cells activity and iron metabolism (Braud et al., 1999; Khakoo et al., 2000). Class Ia molecules present intracellular peptides to cytotoxic (CD8) T-cells and therefore may not be involved in resistance of juvenile trout to *Fp,* which is an extracellular pathogen.

The suggestive weak linkage that I found between *MH-II* markers and BCWD survivability may be related to the role of classical MHC II molecules in presenting exogenous antigens to CD4 T-cells to activate the adaptive response. However, since the memory response is not thought to be activated in juvenile rainbow trout it is more likely that this linkage was caused by other genes, which are physically linked to MHC II genes and are involved in the non-specific defense mechanisms. Weak linkage and associations of *MH-II* loci in salmonids to both bacterial and viral disease resistance were found in previous studies (Langefors et al., 2001; Miller et al., 2004; Palti et al., 2001).

Loss of variation in MH regions should be closely monitored in a selective breeding program for disease resistance as ample evidence exists for balancing selection on classical MHC genes at the population level, which is thought to be the result of contrasting selection pressures from a variety of parasites and pathogens and from ever-changing and evolving pathogens (e.g. Cohen et al., 2006; Hedrick and Thomson, 1983; Hedrick and Black, 1997). Therefore, the possible linkage of *MH-II* to genes involved in *Fp* resistance should be monitored as it can cause unwanted loss of MHC class II variation, which may lead to increased susceptibility of the selected broodstock to other pathogens and parasites. My main finding in this study is the association of *MH-IB* markers with *Fp* resistance. *MHC Ib* loci are not as variable as *Ia* loci and are not as ubiquitously expressed, but they do display moderate allelic variation, polymorphism in transcription and possible variation in the efficiency of protein transcription (Dijkstra et al., 2006). Therefore, the *MH-IB* is a good candidate for QTL mapping of *Fp* resistance. The loss of polymorphism at *Ib* loci and its effect on susceptibility to other pathogens should be examined in future generations of our breeding program.

Acknowledgements

The author thanks M. Renee Fincham and Melissa Johnson for their technical contribution and assistance, Dr. Scott Gahr for his guidance during RNA extraction and cDNA synthesis, Dr. Klaus Elgert for advice regarding MHC nomenclature and Dr. John Hansen for sharing unpublished sequences of the *MH-IA* and *MH-IB* genomic regions in rainbow trout.

References

Braud, V.M., Allan, D.S., McMichael, A.J., 1999. Functions of nonclassical MHC and non-MHC-encoded class I molecules. Current Opinion in Immunology 11, 100-108.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162, 156-159.

Cipriano, R.C., Holt, R.A., 2005. *Flavobacterium psychrophilum*, cause of bacterial cold-water disease and rainbow trout fry syndrome. Fish Disease Leaflet No. 86. United States Dept. of the Interior. U.S. Geological Service, National Fish Health Research Laboratory, Leetown, WV.

Clark, M.S., Shaw, L., Kelly, A., Snell, P., Elgar, G., 2001. Characterization of the MHC class I region of the Japanese pufferfish (Fugu rubripes). Immunogenetics 52, 174-185.

Cohen, S., Tirindelli, J., Gomez-Chiarri, M., Nacci, D., 2006. Functional implications of major histocompatibility (MH) variation using estuarine fish populations. Integrative and Comparative Biology. 46, 1016-1029.

Cox, D.R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B, 187.

Dijkstra, J., Kiryu, I., Yoshiura, Y., Kumanovics, A., Kohara, M., Hayashi, N., Ototake, M., 2006. Polymorphism of two very similar MHC class Ib loci in rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 58, 152-167.

Ducrocq, V., Casella, G., 1996. A Bayesian analysis of mixed survival models. Genetics of Selective Evolution. 28, 505-529.

Ducrocq, V., Solkner, J., 1998. The SURVIVAL KIT - a FORTRAN package for the analysis of survival data., Proc. 6th World Congress on Genetics Applied to Livestock Production, pp. 447-448.

Ducrocq, V., 1999. Extension of survival analysis to discrete measures of longevity. Fourth International Workshop on Genetic Improvement of Functional Traits in Cattle - Longevity. Interbull, Jouy-en-Josas, France, pp. 41-47.

Elgert, K.D., 1996. Immunology: understanding the immune system. Wiley-Liss, Inc., New York, NY, 441 pp.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics, fourth edition. Pearson Prentice Hall, Harlow, UK.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for comuptational biology and bioinformatics. Genome Biology. 5, R80.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., Stet, R.J.M., 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55, 210-219.

Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P., Du Pasquier, L., 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. Journal of Immunology 163, 774-786.

Hedrick, P.W., Thomson, G., 1983. Evidence for balancing selection at HLA. Genetics 104, 449-456.

Hedrick, P.W., Black, F.L., 1997. HLA and mate selection: no evidence in South Amerindians. American Journal of Human Genetics 61, 494-496.

Henryon, M., Berg, P., Olesen, N.J., Kjaer, T.E., Slierendrecht, W.J., Jokumsen, A., Lund, I., 2005. Selective breeding provides an approach to increase resistance of rainbow trout (*Onchorhynchus mykiss*) to the diseases, enteric redmouth disease, rainbow trout fry syndrome, and viral haemorrhagic septicaemia. Aquaculture 250, 621-636.

Johnson, N.A., Rexroad III, C.E., Hallerman, E.M., Vallejo, R.L., Palti, Y., 2007. Development and evaluation of a new microsatellite multiplex system for parental allocation and management of rainbow trout (*Oncorhynchus mykiss*) broodstocks. Aquaculture 266, 53-62.

Kalbfleisch, J.D., Prentice, R.L., 2002. The statistical analysis of failure and time data. John Wiley & Sons, Englewood Cliffs, New Jersey, USA.

Khakoo, S.I., Rajalingam, R., Shum, B.P., Weidenbach, K., Flodin, L., Muir, D.G., Canavez, F., Cooper, S.L., Valiante, N.M., Lanier, L.L., Parham, P., 2000. Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. Immunity 12, 687-698.

Kiryu, I., Dijkstra, J.M., Sarder, R.I., Fujiwara, A., Toshiura, Y., Ototake, M., 2005. New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. Fish and Shellfish Immunology 18, 243-254.

Laird, N.M., Horvath, S., Xu, X., 2000. Implementing a unified approach to family-based tests of association. Genetic Epidemiology 19, S36-S42.

Landis, E.D., Palti, Y., Dekoning, J., Drew, R., Phillips, R.B., Hansen, J.D., 2006. Identification and regulatory analysis of rainbow trout tapasin and tapasin-related genes. Immunogenetics 58, 56-69.

Lange, C., Silverman, E.K., Xu, X., Weiss, S.T., Laird, N.M., 2003. A multivariate transmission disequilibrium test. Biostatistics 71, 195-206.

Langefors, A., Lohm, J., Grahn, M., Andersen, O., Schantz, T.v., 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proceedings of the Royal Society of London. 268, 479-485.

Lee, H.S., Li, W., Lee, A., Rodine, P., Graham, R.R., Ortmann, W.A., Batliwalla, F., Lee, K.W., Bae, S.C., Behrens, T.W., Gregersen, P.K., 2006. Microsatellite typing for *DRB1* alleles: application to the analysis of HLA associations with rheumatoid arthritis. Genes and Immunity 7, 533-543.

Lynch, M., Walsh, B., 1997. Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

Matsuo, M.Y., Asakawa, S., Shimizu, N., Kimura, H., Nonaka, M., 2002. Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). Immunogenetics 53, 930-940.

Michalova, V., Murray, B.W., Sultmann, H., Klein, J., 2000. A contig map of the MHC class I genomic region in the zebrafish reveals ancient synteny. Journal of Immunology. 164, 5296-5305.

Miller, K.M., Winton, J.R., Schulze, A.D., Purcell, M.K., Ming, T.J., 2004. Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. Environmental Biology of Fishes 69, 307-316.

Milliken, G.A., Johnson, D.E., 1984. Analysis of Messy Data: Designed Experiments, Vol. 1. Lifetime Learning Publications, Belmont, CA.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M.R.M., Akutsu, T., Okamoto, N., 2001. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). Molecular Genetics and Genomics 265, 23-31.

Palti, Y., Nichols, K.M., Waller, K.I., Parsons, J.E., Thorgaard, G.H., 2001. Association between DNA polymorphisms tightly linked to MHC class II genes and IHN virus resistance in backcrosses of rainbow and cutthroat trout. Aquaculture 194, 283-289.

Palti, Y., Silverstein, J.T., Wieman, H., Phillip, J.G., Barrows, F.T., Parsons, J.E., 2006. Evaluation of family growth response to fishmeal and gluten-based diets in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 255, 548-556.

Palti, Y., Rodriguez, M.F., Gahr, S.A., Hansen, J.D., 2007. Evolutionary history of the *ABCB2* genomic region in teleosts. Developmental and Comparative Immunology 31, 483-498.

Phillips, R.B., Zimmerman, A., Noakes, M.A., Palti, Y., Morasch, M.R.W., Eiben, L., Ristow, S.S., Thorgaard, G.H., Hansen, J.D., 2003. Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. Immunogenetics 55, 561-569.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects models in S and S-Plus. Springer, New York, NY, USA.

Prentice, R.L., Gloeckler, L.A., 1978. Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34, 57-67.

Rexroad, C.E., Palti, Y., 2003. Development of ninety-seven polymorphic microsatellite markers for rainbow trout. Transactions of the American Fisheries Society 132, 1214-1221.

Rexroad, C.E., Danzmann, R.G., Palti, Y., Vallejo, R.L., 2006. The NCCCWA genetic linkage map for rainbow trout. Plant & Animal Genome XIV, the international conference on the status of plant and animal genome research, San Diego, CA, USA, pp. 261.

Rodriguez, F., Rexroad, C.E., Palti, Y., 2003. Characterization of twenty-four microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Molecular Ecology Notes 3, 619-622.

Rodriguez, M., Gahr, S., Rexroad, C., Palti, Y., 2006. A polymerase chain reaction screening method for rapid detection of microsatellites in bacterial artificial chromosomes. Marine Biotechnology 8, 346-350.

Schneider, S., Roessli, D., Excoffier, L., 2006. Arlequin: a software for population genetics data analysis, version 3.01, Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva. http://lgb.unige.ch/arlequin/software/.

Shiina, T., Dijkstra, J.M., Shimizu, S., Watanabe, A., Yanagiya, K., Kiryu, I., Fujiwara, A., Nishida-Umehara, C., Kaba, Y., Hirono, I., Yoshiura, Y., Aoki, T., Inoko, H., Kulski, J.K., Ototake, M., 2005. Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. Immunogenetics 56, 878-893.

Slatkin, M., Excoffier, L., 1996. Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76, 377-383.

Slettan, A., Olsaker, I., Lie, O., 1997. Segregation studies and linkage analysis of Atlantic salmon microsatellites using haploid genetics. Heredity 78, 620-627.

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68, 978-989.

Stephens, M., Scheet, P., 2005. Accounting for decay of linkage disequilbrium in haplotype inference and missing-data imputation. American Journal of Human Genetics 76, 449-462.

Thorsen, J.B., Zhu, E., Frengen, K., Osegawa, P.J., deJong, P.J., Koop, B.F., Davidson, W.S., Hoyheim, B.A., 2005. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BioMed Central Genomics 6, 50.

Table 1. Information on the ten MH-linked microsatellite markers utilized in this study.

Marker	Gene Region	Forward Primer	Reverse Primer	Label*	T_A (C)	$MgCl_2$ (mM)	GenBank
OMM3084_New	*MH-IA*	AAGAGGGAAATGGTGTAC	ACACATACCTACGCATATA	T-FAM	60	2	**BV718452**
UBA_3_F1_R1	*MH-IA*	GAGGAACTTCTCAAGAACATAACAC	ATTACCACAAGCCCGCTCT	T-FAM	54	2	**AB162342**
0MM3089	*MH-IB*	CTCACTGTCCCATCAGATTAG	CAGGTTGGTGAATGTCAGA	T-FAM	58	1.5	**BV718454**
BHMS429	*MH-IB*	CCCCTGTCAAACGTCTTC	AGCACACTGGATTCAAGG	DL-HEX	54	2	**AF256719**
OMM1189	*MH-IB*	GGCGATGATCTCACAGTC	TCGAAGCCTTGCATACAG	DL-NED	58	2	**AF469974**
OMM3024	*MH-II*	GTATAGAACAGCATATGGGA	GAATAGAGTCTCACCTTGTTGCTA	DL-HEX	54	2	**BV012584**
OMM3026	*MH-II*	CGACGTGAGACAAATGACTAT	AACAGCATATGGGATTGTGT	DL-HEX	58	2	**BV012587**
OMM3028	*MH-II*	CCCGTTGCTATGAGACTCGAA	GGGCAGGCTGTCATTGTAAAG	DL-HEX	58	2	**BV718455**
OMM3079	*TAP1*	GTGTATCAGAGTAGCGGGAAG	GGGGTCTGTGTGCATTATG	T-FAM	56	1.5	**BV678007**
OMM3080	*TAP1*	TCAGGGACATCATAGGGA	TACCCACCCACACTGACTTAT	T-FAM	56	1.5	**BV678009**

***** "T" preceding dye type indicates tailed-reaction and "DL" signifies directly-labeled.

Table 2. *P* values for linkage disequilibrium estimates of all pairs of unlinked microsatellites in a subset of unrelated G_1 individuals (*N*=51).

Markers	*OMM1008*	*OMM1051*	*OMM1088*	*OMM1097*	*OMM1325*	*OMM5047*	*OMM5177*	*OMM5233*
OMM1008	-							
OMM1051	0.29741	-						
OMM1088	0.06489	0.26641	-					
OMM1097	0.03928	0.00002	0.26352	-				
OMM1325	0.01356	0.00266	0.04614	0.04987	-			
OMM5047	0.01024	0.00270	0.04077	0.00724	0.33941	-		
OMM5177	0.24682	0.00027	0.31057	0.00280	0.05567	0.15587	-	
OMM5233	0.68856	0.01440	0.02048	0.00319	0.08085	0.00248	0.00309	-

Table 3. *P* values for linkage disequilibrium estimates of all pairs of MH-linked microsatellites in a subset of unrelated G_1 individuals (*N*=51).

	MH-II		*TAP1*	*MH-IA*		*MH-IB*		
Markers	*OMM3024*	*OMM3026*	*OMM3080*	*OMM3084*	*UBA_3'UTR*	*OMM1189*	*OMM3089*	*BHMS429*
OMM3024	-							
OMM3026	0.00000*	-						
OMM3080	0.46681	0.4789	-					
OMM3084	0.00632	0.00353	0.45761	-				
UBA_3'UTR	0.07945	0.11818	0.29818	0.00000*	-			
OMM1189	0.53968	0.63574	0.43866	0.32914	0.01148	-		
OMM3089	0.96410	0.98879	0.00453	0.00745	0.00442	0.01330	-	
BHMS429	0.03888	0.07091	0.12994	0.00133	0.00065	0.00000*	0.00000*	-

* $P \leq 0.00002$

Table 4. Relationships between *MH-IA* microsatellite haplotypes and *UBA* gene types.

***MH-IA* Microsatellite Haplotype**[a]	***UBA* Gene Type**[b]
2 (153,112)	I,E
5 (155,112)	P,I
7 (155,139)	D,E
8 (160,112)	I
9 (162,112)	D,H
10 (162,124)	H
11 (162,135)	R
14 (166,112)	B
15 (167,139)	D,H
19 (169,135)	D
20 (169,139)	D
21 (171,110)	F
23 (171,139)	D,F
24 (null,139)	D,F
25 (null,null)	H,C
27 (171,null)	F,B

[a] Alleles in parentheses are from *OMM3084_new* and *UBA_3_R1*, respectively.
[b] Gene type designations adopted from Kiryu et al. 2005.

Table 5. Suggestive linkage between MH regions markers and marker-haplotypes and *Fp* resistance measured as average days to death (*ADTD*) or breeding values (*BV*) assessed in a family design among 21 G_1 families.

MH region or marker	*ADTD*		*BV*	
	Offset	No Offset	Offset	No Offset
OMM3026	0.608	0.048*	0.740	0.899
OMM3089	0.048*	0.142	0.081*	0.052*
MH-II	0.495	0.100*	0.552	0.567

* $P \leq 0.10$

Table 6. Suggestive association between MH regions markers and marker-haplotypes and *Fp* resistance measured as average days to death (*ADTD*) or breeding values (*BV*) assessed in two sets of unrelated G_1 individuals (*N*=51).

MH region or marker	Replicate	Trait	Fstat	DF	*P*-value
OMM1189	1	*ADTD*	2.03	23	0.0449*
	2		2.06	23	0.0387*
	1	*BV*	1.51	23	0.1623
	2		1.64	23	0.1118
MH-IB haplotype	1	*ADTD*	122.69	48	0.0081**
	2		141.46	48	0.0070**
	1	*BV*	88.67	48	0.0112*
	2		81.04	48	0.0123*

* $P \leq 0.10$
** $P \leq 0.01$

Figure 1. Schematic diagram describing which fish were phenotyped for BCWD survivability and which fish were genotyped with MH markers.

G_0 = born 2001, spawned 2003:
MH marker genotypes were determined for 21 parental pairs.

G_1 = born 2003, spawned 2005:
MH marker genotypes were determined for 121 fish that were parents of 75 G_2 full-sib families and included two sets of 51 unrelated fish for population analysis and 71 progeny of the 21 G_0 parental pairs for family-based association analysis.

G_2 = born 2005, spawned 2007:
Five progeny from each family were genotyped with *MH-IA* markers to determine haplotypes. *UBA* typing was determined for 41 fish representing 13 families.

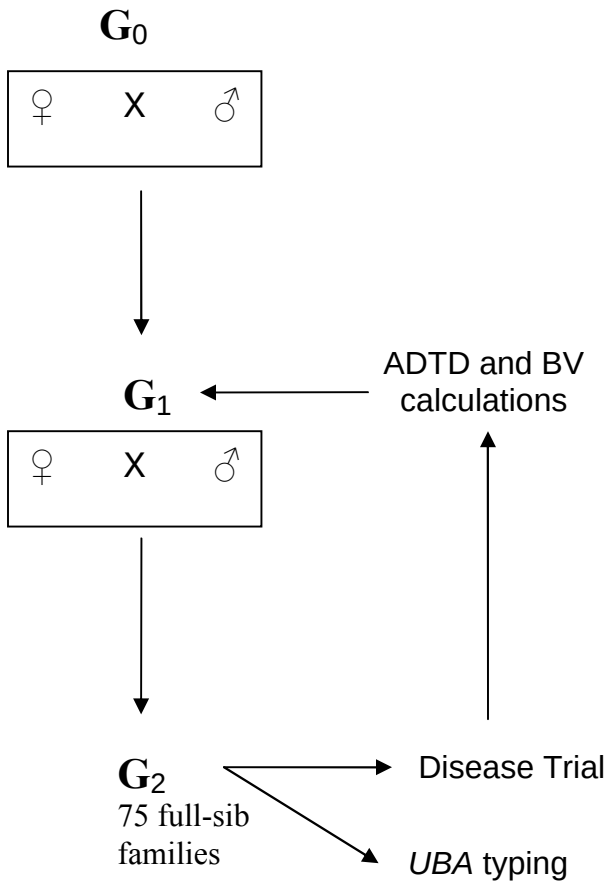
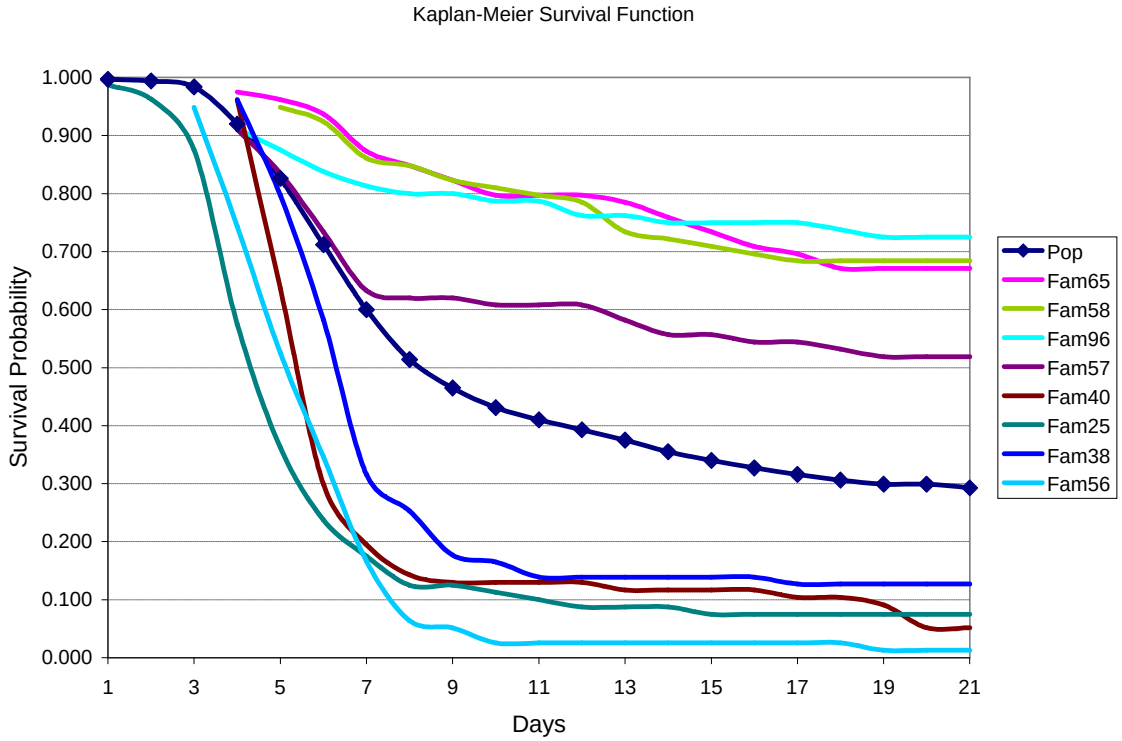


Figure 2. Plot of the Kaplan Meier survival function illustrating the rate of mortality and variation among high and low surviving G_2 families in response to *Fp* challenge.



Summary

The development of genomic tools for rainbow trout research has increased dramatically during the current decade. These advancements are often applicable in closely related species, such as Atlantic salmon and Pacific salmon through cross-species amplifications of molecular markers and comparative mapping. A high-density genetic map for rainbow trout recently was developed at the National Center for Cool and Coldwater Aquaculture (NCCCWA) and much of the current effort is directed toward mapping quantitative trait loci (QTL) involved with immune response, growth rate, and feed conversion. The findings of my thesis research have contributed to these efforts and will benefit the selection program at the NCCCWA.

The multiplex system has improved genotyping efficiency and reduced the costs of pedigree assignments in this program and in large-scale commercial breeding programs. A cost-benefit analysis for genotyping the 2007 NCCCWA broodstock revealed that utilizing the multiplex system reduced costs by over 75% and saved about $300 per PCR plate. It also promoted the practicality of testing family performance in mixed-family tanks, which allows evaluation of genotype x environment effects without confounding common environment (or common garden) effects (Palti et al., 2006).

Diseases are the major cause of fish losses in rainbow trout aquaculture. Thirty-three million rainbow trout were lost to diseases in the US segment of the trout products industry in 2001, causing a $33 million loss (Thorgaard et al., 2002). My progress on identifying genes in rainbow trout associated with resistance to *Flavobacterium psychrophilum* (*Fp*), the causative agent of bacterial cold-water disease (BCWD) could help reduce these losses. The identification of genes or DNA markers that are associated with resistance can aid in the NCCCWA selective breeding program for improved BCWD survivability.

The intent of my research was to develop and evaluate the applicability of various molecular markers in the context of rainbow trout selective breeding programs. This large investment was justified in order to provide future researchers with quality, cost-efficient molecular tools capable of addressing a wide range of immunological and selective breeding issues within rainbow trout, and perhaps other salmonid species.

References

Palti, Y., et al. (2006). Evaluation of family growth response to fishmeal and gluten-based diets in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 255: 548-556.

Thorgaard, G. H., et al. (2002). Status and opportunities for genomics research with rainbow trout. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 133: 609-646.

Vita

Nathan A. Johnson was born in Salem, Ohio in August of 1979; the first of three children of John Lewis Johnson, Jr. and Karen May Johnson. He graduated with honors from Highland Springs High School outside Richmond, Virginia in 1997 and then enrolled in the Department of Fisheries and Wildlife Sciences at Virginia Tech. In December of 2001, he graduated with a Bachelor of Science degree in Fisheries Science with a minor in Biology. From 2002 to 2003, Nathan researched under the supervision of Dr. Richard Neves at the Freshwater Mollusk Conservation Center in Blacksburg, Virginia conducting freshwater biological surveys and propagating endangered freshwater mussels for recovery of populations throughout the Southeastern United States. From 2003-2005, he worked full-time in the Conservation Genetics Laboratory of Dr. Eric Hallerman at Virginia Tech on a variety of population genetics projects focusing on conservation of native freshwater fish and shellfish. During the summer of 2005, he conducted an internship at the National Center for Cool and Coldwater Aquaculture (NCCCWA) in Kearneysville, WV under the direction of Dr. Yniv Palti. Beginning in the fall of 2005, he enrolled at Virginia Tech as a graduate research assistant, where he pursued a Master of Science degree in Fisheries Science. He received his Master of Science in August of 2007 and moved to Gainesville, FL to pursue his doctorate studying the phylogeography and conservation genetics of freshwater mussels at the University of Florida.