FRDC PROJECT 2000/224

AUGUST 2004

ATLANTIC SALMON AQUACULTURE SUBPROGRAM

Molecular genetic tools for the Tasmanian Atlantic salmon industry

DEVELOPMENT AND APPLICATION

Nicholas G. Elliott

Bronwyn H. Innes

James W. Wynne

Mathew T. Cook

Australian Government

Fisheries Research and Development Corporation

Atlantic Salmon Aquaculture Subprogram

Molecular Genetic Tools for the Tasmanian Atlantic salmon industry

Development and application

Nicholas G. Elliott, Bronwyn H. Innes, James W. Wynne and Mathew T. Cook

FRDC Project 2000/224

© CSIRO Marine Research, and Fisheries Research and Development Corporation (FRDC), Australia, 2004

CSIRO Marine Research
GPO Box 1538
Hobart Tas. 7001
Tel: (03) 6232 5222; Fax: (03) 6232 5000

Fisheries Research and Development Corporation
PO Box 222
Deakin West ACT 2602
Tel: (02) 6285 0400; Fax: (02) 6285 4421

This work is copyright. Except as permitted under the Copyright Act 1968 (Cth), no part of this publication may be reproduced by any process, electronic or otherwise, without written permission from the copyright owners. Neither may information be stored electronically in any form whatsoever without such permission.

The Fisheries Research and Development Corporation plans, invests in and manages fisheries research and development throughout Australia. It is a federal statutory authority jointly funded by the Australian Government and the fishing industry.

National Library of Australia Cataloguing-in-Publication entry

Molecular genetic tools for the Tasmanian Atlantic salmon industry: development and application.

Bibliography.
ISBN 1 876996 60 9.

1. Atlantic salmon - Tasmania - Variation. 2. Atlantic salmon - Tasmania - Molecular genetics. 3. Atlantic salmon - Breeding - Tasmania. 4. Atlantic salmon fisheries - Tasmania. I. Elliott, N. G. (Nicholas Grant), 1953- . II. CSIRO. Marine Research. (Series : FRDC Project ; 2000/224).

572.86175609946

Disclaimer: CSIRO Marine Research and FRDC do not warrant that the information contained in this report is free from errors or omissions. CSIRO Marine Research and FRDC shall not be in any way be liable for any loss, damage or injury suffered by the User consequent upon, or incidental to, the existence of errors or omissions in the information.

TABLE OF CONTENTS

Page

1. NON-TECHNICAL SUMMARY....3

2. ACKNOWLEDGEMENTS....5

3. BACKGROUND....6

4. NEED....8

5. OBJECTIVES....8

6. MICROSATELLITE DNA PEDIGREE ANALYSES....9

6.1 Introduction....9
6.2 General methods - DNA extraction, gel running and data analysis....9
6.3 Initial selection and testing of loci....10
6.4 Pedigree suite A....12
6.5 Pedigree suite B....12
6.6 Selection of final suite....13
6.7 PCR multiplex conditions for selected suite of microsatellite loci....14
6.8 Validation of final test....15
6.9 Assessment of alternative methods for obtaining and extracting DNA....16
6.10 Discussion....16
6.11 Conclusion....19

7. GENETIC CHANGE IN THE TASMANIAN POPULATION....25

7.1 Introduction....25
7.2 Methods....26
7.2.1 Samples....26
7.2.2 Total genomic DNA extraction....26
7.2.3 Microsatellite analysis....27
7.2.4 PCR amplification of microsatellite loci....27
7.2.5 Statistical analysis....28
7.3 Results....28
7.4 Discussion....33

8. MHC VARIATION AND LINK TO AGD....36

8.1 Introduction....36
8.2 MHC loci....37
8.2.1 Four MHC primers....37
8.2.2 PCR amplification conditions....37
8.2.3 MHC gene amplifications....38
8.2.4 Restriction fragment length polymorphism (RFLP)....38
8.2.5 Allele determination by DNA sequencing....39
8.2.6 *MHC* class Iα1 RFLPs and alleles....40
8.2.7 *MHC* class Iα2 RFLP and alleles....41
8.2.8 *MHC* class IIα1 RFLPs and alleles....42
8.2.9 *MHC* class IIβ1 RFLPs and alleles....43
8.3 Population study....45
8.3.1 Samples and statistical analyses....45
8.3.2 Results....46
8.3.3 Discussion....49
8.4 MHC variation and AGD resistance/susceptibility....51
8.4.1 Material and Methods....51
8.4.2 Results....52

8.4.3 Discussion 55
8.5 Conclusion 57

9. AFLP ANALYSIS FOR SEX MARKER 58

9.1 Introduction 58
9.2 Materials and Methods 59
9.2.1 Sibling DNA samples 59
9.2.2 AFLP protocol 59
9.3 Results 61
9.4 Discussion 61

10. BENEFITS AND ADOPTION 63

11. FURTHER DEVELOPMENT 63

12. PLANNED OUTCOMES 64

13. CONCLUSIONS 64

13.1 Objective 1. 64
13.2 Objective 2. 64
13.3 Objective 3. 64

14. REFERENCES 66

APPENDIX 1: Intellectual property 74

APPENDIX 2: Staff 75

APPENDIX 3: Supplementary data 76

I: Extraction methods for total genomic DNA 76
II. Summary genetic statistics 79
III. MHC allele sequences 94

1. NON-TECHNICAL SUMMARY

2000/224 Atlantic Salmon Aquaculture Subprogram: Molecular genetic tools for the Tasmanian Atlantic salmon industry – development and application

PRINCIPAL INVESTIGATOR: Dr Nicholas Elliott
ADDRESS: CSIRO Marine Research
GPO Box 1538, Hobart, Tas 7001
Telephone: 03 6232 5222 Fax: 03 6232 5000

OBJECTIVES:
1. To further develop and apply molecular markers for use in industry breeding programs.
2. To genotype selected broodstock with a suite of microsatellite markers to enable efficient pedigree analysis of progeny.
3. To compare microsatellite DNA variation from archival scale samples from the progenitor Canadian population with past and current samples of the Tasmanian population

NON-TECHNICAL SUMMARY:

OUTCOMES ACHIEVED TO DATE
An outcome to date is the use of molecular markers by a member of the Atlantic salmon industry to pedigree and estimate relatedness of broodstock for the commencement of a selective breeding program. An industry selective breeding program using molecular tools for pedigree and selection assistance is currently under discussion by the industry, and its commencement will be the major outcome and final prove of the value of this project, along with further R and D in molecular technology for the Atlantic salmon industry.

- Archived DNA was successfully extracted from 30-year-old dried scales collected in 1971 and 1972 from wild Atlantic salmon from the River Philip, Canada.
- Genetic variation was assessed at 11 nuclear DNA microsatellite loci (three tetra- and eight di-nucleotide repeats) in two samples from the River Philip, Canada (1971/72 and 1991 year-classes), one from Gaden, NSW (1997) and seven from the Tasmanian population (1992, and 1996 to 2001 year-classes).
- The 1971/72 sample (archived scales) from Canada proved to be a reliable and realistic baseline (compared to the 1991 sample) against which to assess genetic drift in the Tasmanian population.
- A highly significant loss of alleles was observed in all hatchery (Canada, Gaden and Tasmanian) produced samples compared to the wild 1971/72 Canada sample. No difference in heterozygosity was observed.
- Estimates of per-generation effective population sizes for the Tasmanian Atlantic salmon population, based on allele frequency variance with the 1971/72 Canada sample, ranged from 102 to 207 individuals, with the average value for the last three (1999 to 2001) year-classes of 185 individuals. These values are up to 80% higher than previous estimates based on allele variance comparison with the 1991 Canada sample. The values are consistent with the population having experienced a minor

bottleneck early in the introduction to Australia but with subsequent hatchery controlled generations of high effective breeding numbers.
- Despite the observed loss of alleles compared to the wild progenitor population our results vindicate the Tasmanian hatchery protocols that have maintained a sufficiently large breeding population to preserve a genetically healthy population.

- A suite of ten highly polymorphic microsatellite loci has been developed that has been shown to provide over 99% accuracy in unequivocal parentage assignment for progeny from commercial multiple cross spawnings. Generally, human error at various points of the test and poor quality DNA accounts for most of the unassigned individuals.
- Given the specific selection of highly variable microsatellite loci and some technical limitations we have been unable to combine all ten loci into a single multiplex PCR reaction that would provide satisfactory results. Rather, to use our test, three separate PCR reactions and two genotype visualisation runs are required per individual.
- A commercial salmonid-DNA typing system also with ten loci has been recently developed overseas. We have compared the resolving power of this set of loci with the ones included in our system. Both sets of loci have a similar resolving/exclusion power, although all our selected loci have ten or more alleles observed in the local population and so are likely to more useful in separating closely related individuals.

- A male specific or Y-chromosome marker for Atlantic salmon was not identified despite the application of 30 AFLP (amplified fragment length polymorphism) primer combinations to male and female siblings. Such a marker continues to elude ourselves and researchers in other salmonid genetic laboratories.

- For the first time, genetic variation in the Tasmanian Atlantic salmon population was examined at coding regions of the nuclear genome. In particular four regions of the MHC (major histocompatibility complex) immune genes were examined in a pilot study. The results suggest a possible reduction in genetic variation at these loci in the Tasmanian population; but the results require validation in a more comprehensive study.
- A significant association was observed between phenotypes at one of the MHC regions with an individual's tendency for susceptibility and resistance to AGD in a laboratory AGD challenge trial. This very promising result requires further examination, but does suggest a genetic influence to AGD resistance that could be exploitable through a selective breeding program.

KEYWORDS: Atlantic salmon, *Salmo salar*, pedigree, genetic variation, microsatellites, AFLP, AGD resistance, MHC

2. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the following individuals or groups for various forms of assistance and collaboration during this and associated research, and apologise to anyone we have inadvertently omitted.

Harry King, SALTAS
Pheroze Jungalwalla, TSGA and formerly TASSAL
Innes Weir, Adrian Steenholdt and David Mitchell, HAC
Doug Paveley and Guy Westbrook, NORTAS
Vicki Wadley, formerly TSGA
Barbara Nowak, School of Aquaculture, University of Tasmania
CRC for Sustainable Aquaculture of Finfish (Aquafin CRC), in particular the AGD research group and the Health Program
John Ritter and colleagues, Diadromous Fish Division, Department of Oceans and Fisheries, New Brunswick, Canada
Glenn Tong and colleagues, AgGenomics Pty. Ltd., Melbourne
Alastair Hamilton, Landcatch Natural Selection, Scotland
Stephen Pyecroft, Fish Health Unit, Department of Primary Industries, Water and Environment, Tasmania,
Bjorn Hoyheim, Norwegian School of Veterinary Science, Norway
Eric Verspoor, FRS Marine Laboratory, Aberdeen, Scotland
Timothy King, USGC-BRD, West Virginia, USA

We are also extremely grateful to the assistance and support of our colleagues at CSIRO Marine Research in particular Jawahar Patil, Sharon Appleyard and Bob Ward.

3. BACKGROUND

Atlantic salmon production in Australia continues to increase (10 970t HOGG 1999/2000 to 14,292t, worth $111.5 m in 2001/02 ABARE eReport 03.8), and accounts for over half (by weight) of the Australian finfish aquaculture production. To maintain its position in a competitive market place, the industry continually improves production efficiency through changes in management practices and nutrition.

Genetic improvement programs have provided significant production gains for many salmonid industries overseas (e.g. Norway, greater than 10% per generation). The local industry is about to embark on its own program. In 1999 the Tasmanian Salmonid Growers Association (TSGA) instigated a scoping study on a selective breeding program, and this was followed by a second study and business plan in 2003. Molecular genetic technology will be pivotal to the industry's effort to produce superior stock through a selective breeding program. It will be required for pedigree analysis and identification of economically important traits or gene markers (e.g. sex marker, AGD resistance) to assist in broodstock selection.

A previous molecular genetics project on Atlantic salmon (FRDC 95/80 and 96/347) developed and applied new genetic techniques (including microsatellite loci) to locate a sex marker and assess the 'genetic health' (genetic variation) of the local population. The Tasmanian population is descended from ova imported from Canada to NSW in the mid-1960s, and then to Tasmania in the mid 1980s; it has been isolated for over 30 years, with no new material to boost the local gene pool. The previous project suggested some loss of genetic variation in the local population. However, some essential data are needed to refine the analysis: better estimates of the base-line genetic variation in the progenitor population; analyses of additional year-classes of the local population; and research into the effective number of broodstock contributing to each generation.

The advance of molecular genetic techniques over the past few years has made it possible to extract and analyse DNA from archived tissue samples such as fish scales. Analysis of genetic variation in the Canadian progenitor population using such archived material, and additional year-class samples from the local population, would provide a much more robust assessment of the actual loss of genetic variation in the Tasmanian population.

To optimise a breeding program the number of broodstock fish must be optimised to reduce the risk of inbreeding and subsequent detrimental effects on the population. As fish have a high reproductive capacity (thousands of eggs per individual), a small number of individuals could potentially make a large contribution to the genetic pool in successive generations. To avoid this, the pedigrees of individuals are required for broodstock selection and mating design. Unfortunately such information is currently not available for the local population. However, pedigree genotyping and assessment of relatedness is possible through molecular genetic markers such as microsatellites. A £1.5 million genetic fingerprint research project using this technology for Scottish salmon was initiated in 1998 (Fish Farming International Feb. 1998).

An important step in applying genetic markers to selective breeding programs is an understanding of the correlation or linkage between markers, and their relative positions within the genome or on particular chromosomes; this is called genome mapping. Such a

(linkage or genetic marker) map will be a necessary tool in the search for loci (chromosome regions) that have major effects on economically important traits. Such loci, termed Quantitative Trait Loci (QTLs), can mark such traits as growth, disease resistance and product quality. These markers can then be applied to assist in the selection of broodstock with desired characteristics early in their development, rather than having to wait for the phenotypic expression of the desired trait.

Such a genetic marker map for salmonids (Atlantic salmon, rainbow trout and brown trout) was recently completed under a European Commission funded project (SALMAP - five European laboratories). The project produced over 250 highly informative DNA microsatellite markers. We also developed such microsatellite markers during project 95/80 and some are being included on the SALMAP Atlantic salmon map through our collaborative status. In addition our laboratory is now experienced with the application of AFLP (amplified fragment length polymorphism) technology in oysters, abalone and krill, and the application to salmon would quickly increase the number of informative markers available for genetic mapping. A continuation of the European project with a local and/or international QTL mapping project would be invaluable for assisting selective breeding programs.

In addition to the genetic variation study, the previous FRDC project also searched but failed to locate a sex-marker for use in the all-female smolt breeding programs, which produce a significant component of the overall industry production. AFLPs have been used to identify a sex marker in penaeid prawns, and may assist in Atlantic salmon.

This project could be integrated into any selective breeding program for Atlantic salmon. The application of molecular genetic techniques would extend beyond the three-year life of this project, particularly for pedigree analysis and QTL mapping.

4. NEED

The efficiency and effectiveness of selective breeding programs can be greatly enhanced through the use of DNA technology. The application of such technology will be used for pedigree information and identification of markers for economic traits leading to marker-assisted selection. Although various international laboratories and CSIRO have developed a bank of molecular markers, the potential of these and other nuclear DNA markers need evaluating.

Genetic variation within the Tasmanian population remains unclear. In time it may be considered advantageous to import new genetic material to enhance various characteristics of the local population. However, at present that is unlikely and the status of the population, and of the effectiveness of breeding protocols, needs to be better understood. Significant progress has been made with this (project 96/347) and is extended in this project

5. OBJECTIVES

1. To further develop and apply molecular markers for use in industry breeding programs.

2. To genotype selected broodstock with a suite of microsatellite markers to enable efficient pedigree analysis of progeny.

3. To compare microsatellite DNA variation from archival scale samples from the progenitor Canadian population with past and current samples of the Tasmanian population

6. MICROSATELLITE DNA PEDIGREE ANALYSES

The research below was undertaken towards achieving Objective 2.

6.1 Introduction

The initiation of this FRDC project proposal in 1999 coincided with a scoping study on a selective breeding program instigated by the TSGA. One of the issues highlighted in that study in relation to program design was the importance of maintenance of pedigree information. Of particular concern was the high capital cost for a specific family-based breeding facility, where families would be reared in separate tanks until large enough to be tagged. The proposed alternative was for a breeding program utilising DNA markers to pedigree and so assign individuals to family groups. In such a program, the progeny from multiple families would be reared together, thereby reducing environmental influences on early performance and capital costs for replicated family tanks. The alternative program design required an efficient pedigree test.

This aim of this component of the project was to develop and test a set of DNA microsatellite markers that would correctly assign individuals to family groups (i.e. to correct parents).

Microsatellites were the DNA marker of choice as they are neutral, codominant markers with Mendelian inheritance. They can be PCR-amplified from DNA extracted from small amounts of tissue such as a finclip, blood or scale sample, and importantly many microsatellite loci are extremely variable (high number of alleles). Our task was to select a set of microsatellite markers, and evaluate their technical capabilities and efficacy in assigning pedigree in the Tasmanian Atlantic salmon population.

Since the initiation of this work similar research on Atlantic salmon has been reported and incorporated into breeding programs (e.g. Norris et al 2000; Guy and Hamilton 2003).

6.2 General methods - DNA extraction, gel running and data analysis

All samples were stored at – 80°C until required for DNA extraction. DNA was in most instances extracted using either a modified CTAB protocol or a DNeasy™ Tissue Kit (Qiagen). However we also tested the efficacy of two rapid DNA extraction methods. All DNA extraction protocols are listed in Appendix 3.

All PCR products were run on 1.5 to 2% agarose gels to assess amplification quality. PCR products were then diluted, usually 1 in 3, with purified water. 1µl of this diluted PCR product was mixed with 2µl of a size standard mix. The size standard mix consisted of 1.15µl formamide, 0.40µl of blue loading dye and 0.45µl of GeneScan™-500 TAMRA ™ size standard (Applied Biosystems). Samples were denatured at 95°C for 3 minutes and then kept on ice until ready to load. Samples were run on a 4.8% polyacrylamide gel on an ABI Prism 377 DNA Sequencer. The duration of the gel run times depended on the size of the largest PCR product but usually ranged from 2 to 3.5 hours.

The microsatellite gels were analysed using Genescan Analysis Software 3.1 (PE Applied Biosystems) and Genotyper DNA Fragment Analysis Software 1.1.1 (PE Applied Biosystems). A table was created in Genotyper listing the genotypes of the individuals.

This table was then exported into an Excel spreadsheet and translated into the correct format for Probmax (Danzmann 1997), a parentage assignment program. An incorrect genotype caused by experimental errors or a mutation in the DNA would result in a non-assignment of a parent to its progeny. The number of non-assignments could be reduced by lowering the threshold for inclusion in a match. Probmax allows this to be done on a locus by locus basis.

6.3 Initial selection and testing of loci

Isolation of microsatellite loci is a time and labour consuming task, however, several hundred have been isolated by researchers in both Europe and North America either specific to Atlantic salmon or to closely related species (e.g. brown trout *Salmo trutta)* or genera (*Oncorhynchus* and *Salvelinus*). These loci are available either in the literature, on GenBank (a sequence database located on the National Centre for Biotechnology Information website: www.ncbi.nlm.nih.gov), or directly as unpublished loci from colleagues.

We set the following criteria against which to judge a microsatellite locus for suitability in a pedigree test:

- high polymorphisms – a reported large number of alleles, and importantly an observed high number in our local population,
- even allele frequency – ideally loci not dominated by a single common allele, with numerous rare alleles, but rather a number of alleles at a similar frequency (Villanueva et al. 2002)
- ease and accuracy of scoring - a locus with a tetranucleotide repeat motif (e.g. CAGA) was preferable to one with a dinucleotide repeat motif (e.g. CA) due to the reduced level of stutter peaks, higher level of variability and increased distance between each allele, all of which can potentially reduce scoring errors (O'Reilly et al. 1996, O'Reilly et al. 1998, Letcher and King 2001).
- ability to multiplex – to reduce costs it is preferable to amplify multiple loci in one PCR reaction and/or run in one gel analysis lane
- no null alleles – non-amplifying alleles make parentage assignment difficult, and so preference would be to not use loci known to have null alleles.

PCR primers were ordered for 55 microsatellite loci in order to test their suitability when analysed on an ABI Prism 377 DNA Sequencer and with samples from the Tasmanian Atlantic salmon population (Table 6.1). Initially all microsatellite loci were amplified in singleplex reactions (one locus per PCR reaction). Optimum PCR conditions were determined empirically using the information available for each locus. Amplifications were in a 25µL volume containing 1x PCR buffer, 1.5mM $MgCl_2$, 0.2-0.04 µM forward primer (fluorescent labelled), 0.2-0.04 µM reverse primer, 200µM each dNTP, 1 unit Taq polymerase (Fisher Biotech) and 8 µl of 1/10 - 1/20 diluted DNA (approx. 20-40 ng). Cycling parameters were an initial denaturation at 95°C for 3 min, followed by 35 cycles at 95°C for 30 sec, specific annealing temp for 30 sec, at 72°C for 1 min, followed by a final extension at 72°C for 20 mins. Each locus was tested on a minimum of 10 individuals.

Table 6.1. Microsatellite loci tested for suitability as pedigree identification markers for Tasmanian Atlantic salmon. Repeat type D = dinucleotide repeat (e.g. CACACA), T = tetranucleotide repeat (e.g. CAAGCAAGCAAG)

Locus	Genbank accession number	Repeat type	Reference
Ssa 4	unknown	D	McConnell et al. 1995
Ssa 12	U58900	D	Norris et al. 2000
Ssa 13	U58903	D	Norris et al. 2000
Ssa 85	U43692	D	O'Reilly et al. 1996
Ssa 141	U58905	D	Norris et al. 2000
Ssa 171	U43693	T	O'Reilly et al. 1996
Ssa 197	U43694	T	O'Reilly et al. 1996
Ssa 202	U43695	T	O'Reilly et al. 1996
Ssa 289	unknown	D	McConnell et al. 1995
Ssa 293	unknown	D	McConnell et al. 1995
Ssleei 84	U86703	D	Schill and Walker (unpublished)
Ssleer 15	U86708	D	Schill and Walker (unpublished)
Ssosl 85	Z48596	D	Slettan et al. (unpublished)
Ssosl 417	Z48598	D	Slettan et al. (unpublished)
Ssosl 439	Z49996	D	Slettan et al. (unpublished)
Ssosl 444	Z49997	D	Slettan et al. (unpublished)
Ssa0062NVH	unknown	T	Hoyheim B. pers. comm.
Ssa0071NVH	unknown	T	Hoyheim B. pers. comm.
Ssa0090NVH	unknown	T	Hoyheim B. pers. comm.
Ssa0103NVH	unknown	D	Hoyheim B. pers. comm.
Ssa0105NVH	unknown	T	Hoyheim B. pers. comm.
SsaD58	AF525210	T	Eackles and King (unpublished)
SsaD71	AF525211	T	Eackles and King (unpublished)
SsaD190	AF525206	T	Eackles and King (unpublished)
SsaD237	AF525207	T	Eackles and King (unpublished)
SsaD144	AF525203	T	Eackles and King (unpublished)
SsaD170	AF525205	T	Eackles and King (unpublished)
SsaD85	AF525213	T	Eackles and King (unpublished)
Ssa401UOS	AJ402718	T	Cairney et al. 2000
Ssa403UOS	AJ402720	T	Cairney et al. 2000
Ssa404UOS	AJ402721	T	Cairney et al. 2000
Ssa407UOS	AJ402724	T	Cairney et al. 2000
Ssa408UOS	AJ402725	T	Cairney et al. 2000
Ssa410UOS	AJ402727	T	Cairney et al. 2000
Ssa417UOS	AJ402734	T	Cairney et al. 2000
Ssa420/1UOS	AJ402737	T	Cairney et al. 2000
Ssa421UOS	AJ402738	T	Cairney et al. 2000
Ssa422UOS	AJ402739	T	Cairney et al. 2000
SSsp2215	AY081810	T	Paterson et al. 2004
SSsp2216	AY081811	T	Paterson et al. 2004
SSsp1605	AY081812	T	Paterson et al. 2004
SSsp2201	AY081807	T	Paterson et al. 2004
SSsp2210	AY081808	T	Paterson et al. 2004
SSsp2213	AY081809	T	Paterson et al. 2004
SSspG7	AY081813	T	Paterson et al. 2004
Omy 77	unknown	D	Morris et al. 1996
Omy 207	unknown	D	O'Connell et al. 1997
Omy 301	unknown	D	Ward et al. 2002
Omy 325	unknown	D	O'Connell et al. 1997
Omy 335	unknown	D	Jackson et al. 1998
FGT3	unknown	D	Estoup et al. 1998
One 2	U56700	D	Scribner et al. 1996
Ots 1	AF107029	D	Banks et al. 1999
MST 3	AB001060	D	Presa and Guyomard 1996
Sfo 8	U50305	D	Angers et al. 1995

Loci were rejected due to:

- poor quality or inconsistent PCR amplification
- the presence of null (non-amplifying) alleles (progeny homozygous, no parental match, homozygote excess in samples run)
- an extremely uneven distribution of allele frequencies
- low number of alleles observed in sample
- possible preferential amplification of smaller sized alleles.

6.4 Pedigree suite A

Eight loci were selected from the initial group of loci tested. These were *Ssa197, Ssa171, Ssa103, Ssa202, Ssa62, Ssleei84, cmrSs1.22* and *Ssosl85*.

The suitability of this suite of loci for pedigree analysis was tested on a sample of 100 random progeny (collected in September 2001) from a May 2001 NORTAS spawning purported to have comprised 9 females crossed with 3 males. Three sets of potential parents (each 9 females and 3 males) were analysed as at the time we did not know from which of these three mixed spawnings the progeny would be derived. The intention was for the progeny from each spawning to be kept separate.

We found that the 100 sampled progeny actually represented individuals with parents in all three of the spawning sets (14, 37 and 49 individuals respectively). As each individual progeny was only assigned to parents within the same set of spawners, this suggests mixing between the three cohorts was not at fertilization but at a later time in the hatchery.

With the set of eight loci we were able to initially unambiguously assign parentage to 81% of the sampled progeny, i.e. assigned 81 of the 100 progeny exclusively to a particular parental pair and to no other possible pair of parents. Of the remaining 19 individuals, only two were assigned to more than one set of parents. Genotyping errors (11), missing parental data (5; assume non-sampled broodstock involved in spawning) and possible mutations (1) accounted for the remaining unassigned individuals. Re-analysis of the unassigned individuals would permit all but six individuals to be unambiguously assigned to a parental pair.

This same suite of eight loci was also tested on 64 progeny randomly selected from each of three single pair-crosses at the SALTAS hatchery (May 2001). The progeny sampled in this trial were very early hatchings, individuals in some instances still retaining remnants of the yolk sac. Some unusual genotypes were observed including the presence of three discrete alleles at one locus in an individual and two individuals from one family were homozygous at all 8 loci. Ignoring these unusual individuals we were able to unambiguously assign parentage to 100% of the remaining progeny.

6.5 Pedigree suite B

In an effort to improve the assignment rate of the suite A set of markers, four were discarded (*cmrSs1.22, Ssosl85, Ssleei 84* and *Ssa62*) due to inconsistent PCR amplification or stutter peaks on the gels which made it difficult to distinguish heterozygotes from homozygotes. Preferential amplification of the smaller alleles over the larger alleles was also observed at locus *Ssa62* which caused many individuals to be initially incorrectly genotyped as homozygotes when they were really heterozygotes.

Ten new loci (*Ssa403UOS, SSsp2216, SSsp2215, SsaD144, SsaD85, SsaD170, Ssa410UOS, Ssa417UOS, SSsp1605*, and *Ssa401UOS*) were included with the remaining four of the loci of Suite A (*Ssa197, Ssa171, Ssa103,* and *Ssa202*). We retained the dinucleotide repeat locus *Ssa103* because of the small size of its alleles (82 to 128 base pairs) a potential advantage when multiplexing.

This new suite of loci was tested on 100 randomly chosen progeny (collected January 2001) derived from an unknown number of crosses involving up to 52 parents (collected July 2000) in a 2000 NORTAS spawning.

The 14 loci allowed us to initially assign unequivocal parentage to 70% of the individuals. The mis-matches were due to human error and poor sample quality (11%), the presence of a previously unidentified null allele in locus *Ssa401UOS* (12%) and missing parental data (7%). If in the parental assignment analysis we allowed an incorrect genotype match between parent and progeny at one or two of the 14 loci (i.e. reduced the stringency of the test) then 93% of the individuals could be assigned to a specific set of parents.

6.6 Selection of final suite

Mindful of designing a low cost pedigree test, we investigated reducing the number of loci in suite B while improving the efficacy of the suite of markers. As locus *Ssa401UOS* was observed to have apparent null alleles in our samples it was the first locus to be eliminated. We then used the program Probmax and the suite B dataset to determine the assignment potential of particular combinations of loci.

We first trialled the ‘best’ six loci (*Ssa197, Ssa171, Ssa403UOS, SsaD144, SsaD170* and *Ssa410UOS*); all with high allele numbers and relatively even distribution of alleles (Table 6.2 and Figure 6.1). This suite of loci provided unequivocal parental assignment to 87% of the progeny assuming no genotyping errors in the data; 3% of the mismatch was due to poor amplification at one or more loci, 3% mismatched at one locus, 2% were assigned to two possible sets of parents and 5% no match at all. If we lowered the inclusion stringency (i.e. allowed a genotype mis-score) the rate of assignment to one parental-pair dropped to 68%, which was not unexpected due to the small number of available loci allowing for more possible choices of parent pairs.

Adding the two loci *SSaD85* and *Ssa417* resulted in an initial assignment of only 76% to a single parental-pair, this was at the highest stringency. The lower success was due to 12 new mismatches associated with the two additional loci. If the stringency is lowered to allow a genotype mismatch at one locus, then 15 of the previous mismatches are assigned to a parental pair, but another 5 previously assigned to one pair are assigned to two possible pairs. Further combinations of eight or ten loci resulted in initial assignment rates to a single parental-pair of between 76% to 91%.

In all the examinations of the sample set of 100 individuals, there were three individuals that had poor PCR amplification for some loci, and so could not under any scenario be assigned to parents. DNA from such individuals could be re-extracted and/or re-amplified which may allow assignment to parents, there by increasing the assignment rate. Overall only six of the individuals could not be assigned to a set of parents (mutations, missing parents or gel scoring error most likely to account for this).

We concluded that a suite of 10 microsatellite loci would provide the most efficient pedigree test (Table 6.2).

Table 6.2. Final suite of 10 microsatellite loci suggested for pedigree assignment. N is the number of individuals from the Tasmanian population examined at each locus.

Locus	Alleles observed	Allele size range (base pairs)	N	GenBank Accession Number	Reference
SsaD144	24	181 to 269	514	AF525203	Eackles and King (unpublished)
SsaD170	29	152 to 328	508	AF525205	Eackles and King (unpublished)
SsaD85	20	261 to 373	488	AF525213	Eackles and King (unpublished)
Ssa410UOS	18	211 to 279	461	AJ402727	Cairney et al. 2000
Ssa417UOS	23	311 to 411	422	AJ402734	Cairney et al. 2000
Ssa171	16	214 to 262	416	U43693	O'Reilly et al. 1996
Ssa197	15	163 to 223	514	U43694	O'Reilly et al. 1996
SSsp2215	11	148 to 199	519	AY081810	Paterson et al. 2004
SSsp2216	14	201 to 317	516	AY081811	Paterson et al. 2004
Ssa403UOS	20	155 to 255	520	AJ402720	Cairney et al. 2000

6.7 PCR multiplex conditions for selected suite of microsatellite loci.

Co-amplifying microsatellite loci in the one PCR reaction (multiplexing) is a recognised method to reduce costs when assessing multiple loci. The main advantages are reduced labour costs and lower volumes of Taq polymerase (an expensive component of a PCR reaction). The disadvantages of multiplexing, apart from time spent in achieving optimal conditions, are different annealing temperatures of loci, failure of some loci to amplify efficiently in the presence of others, and sensitivity of some loci when multiplexed to the quality and quantity of the DNA template. We endeavoured to multiplex a set of ten loci but were unable to achieve a consistent and satisfactory PCR product for all loci in single multiplex. An oligonucleotide extension on the end of each PCR primer can improve the quality and quantity of loci co-amplified in a multiplex reaction (Shuber et al. 1995). We trialled unsuccessfully such primer extensions designed by Ian Franklin (CSIRO, pers. comm.) and a commercial multiplex PCR kit (QIAGEN® Multiplex PCR kit). We were unable to get suitable quality PCR amplification despite varying annealing temperatures (50°C to 71°C).

We were however, able to amplify our 10 selected loci in three multiplex reactions:

Multiplex One (*Ssa403UOS, Ssa197, SSsp2216* and *SSsp2215*)
Amplifications were in 25µl volumes containing 1 X PCR buffer, 1.5mM $MgCl_2$, 0.2 µM each of forward and reverse Ssa403UOS primer, 0.06 µM each of forward and reverse *Ssa197* primer, 0.06 µM each of forward and reverse *SSsp2216* primer, 0.04 µM each of forward and reverse *SSsp2215* primer, 1 unit Taq polymerase (Fisher Biotech) and 8 µl of

1/10 -1/20 diluted DNA (approximately 20-40 ng/µl). Cycling parameters were: 95°C for 3 min, followed by 35 cycles at 95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min, followed by a final extension at 72°C for 10 mins.

Multiplex Two *(SsaD144, SsaD85* and *SsaD170)*
Amplifications were in 25µl volumes containing 1 X PCR buffer, 1.5mM $MgCl_2$, 0.12 µM each of forward and reverse *SsaD144* primer, 0.12 µM each of forward and reverse *SsaD85* primer, 0.12 µM each of forward and reverse *SsaD170* primer, 1 unit Taq polymerase (Fisher Biotech) and 8 µl of 1/10 -1/20 diluted DNA (approximately 20-40 ng/µl). Cycling parameters were: 95°C for 3 min, followed by 35 cycles at 95°C for 30 sec, 58°C for 30 sec, 72°C for 1 min, followed by a final extension at 72°C for 10 mins.

Multiplex Three *(Ssa410UOS, Ssa417UOS* and *Ssa171)*
Amplifications were in 25µl volumes containing 1 X PCR buffer, 1.5mM $MgCl_2$, 0.18 µM each of forward and reverse *Ssa410UOS* primer, 0.18 µM each of forward and reverse *Ssa417UOS* primer, 0.20 µM each of forward and reverse *Ssa171* primer, 1 unit Taq polymerase (Fisher Biotech) and 8 µl of 1/10 -1/20 diluted DNA (approximately 20-40 ng/µl). A touchdown protocol was used to amplify these loci. Cycling parameters were: 95°C for 3 min, followed by 15 cycles at 95°C for 30 sec, 58°C for 30 sec, 72°C for 1 min, followed by 20 cycles at 95°C for 30 sec, 57°C for 30 sec, 72°C for 1 min, followed by a final extension at 72°C for 10 mins.

PCR products were run on 1.5-2% agarose gels to assess amplification quality. PCR products were then diluted, usually 1 in 3, with purified water. 1µl of this diluted PCR product was mixed with 2µl of a size standard mix. The size standard mix consisted of 1.15µl formamide, 0.40µl of blue loading dye and 0.45µl of GeneScan™-500 TAMRA ™ size standard (Applied Biosystems). Samples were denatured at 95°C for 3 minutes and then kept on ice until ready to load. Samples were run on a 4.8% polyacrylamide gel on an ABI Prism 377 DNA Sequencer. The duration of the gel run times depended on the size of the largest PCR product but usually ranged from 2 to 3.5 hours. To reduce the number of gels, multiplexes *one* and *three* can be co-loaded into a single gel although alleles from locus *Ssa171* can be obscured by alleles from locus *SSsp2216*.

6.8 Validation of final test

The suite of 10 selected loci (Table 6.2) and the optimised PCR multiplexes were tested on sample from the 2003 SALTAS spawning. The test sample comprised of:

Parents: 78 (64 females and 14 sex-reversed males). Finclips were collected at the time of spawning (May 2003).
Fertilization cross design: 8 separate batches each consisting of mixed ova from 8 females crossed with mixed milt from either 3 or 4 males.
Incubation of batches: Each of the 8 batches was incubated separately.
Progeny: The progeny from all 8 batches were mixed at first feed (November 2003). A random sample of 300 individuals was collected from each of two tanks containing mixed progeny for all 8 batches.

DNA was extracted from all 78 parents and 299 random progeny. All DNA samples were then PCR amplified for the 10 selected loci (Table 6.2) using the three multiplex

reactions, genotypes were determined, and parental assignment of the progeny determined.

From the initial gel analyses unambiguous assignment of a single parental pair was made for 94% of the progeny. Of the remaining progeny poor sample DNA (both parents and progeny) and amplification accounted for most cases, and re-analyses of these samples would result in an assignment of parentage to 98-99% of the progeny.

6.9 Assessment of alternative methods for obtaining and extracting DNA

We examined three alternative non-lethal methods of obtaining tissue for DNA extraction. These were finclip samples, using either a biopsy punch or a pair of scissors, and scale and mucus samples by running a glass slide along the side of the fish and collecting the residue with filter paper. The quality of extracted DNA from these samples was examined using the final set of 10 microsatellite loci. In addition we examined two alternative DNA extraction protocols (see Appendix 3 for details).

Two rapid and inexpensive DNA extraction techniques (boiling and overnight incubation) were tested on mucus, dorsal fin, tail fin and scale samples (muscle was used as control). Both extraction techniques had a minimal number of steps and used very few reagents (Appendix 3). The quality and quantity of the DNA extracted by both techniques was compared to that obtained with either of our standard methods; a commercially available DNA extraction kit (QIAGEN DNeasy™ tissue kit) or our standard high quality, but more time-consuming, CTAB extraction protocol.

The PCR amplification quantity and quality using DNA from the two rapid methods was comparable to the standard extraction methods, although the boiling method was less effective in terms of quality. None of the extraction methods produced DNA from the mucus sample, and we are not sure whether this was due to a lack of cells in the mucus material or whether there was an inhibiting substance that affected the PCR reaction.

Initial gel runs using PCR product from the overnight incubation method suggested that the quality of the genotype data was not compromised although a slightly higher concentration of DNA was required for a successful multiplex PCR. However following further multiplex PCR reactions and genotyping of DNA extracted using both rapid methods the consistency of the product was poor. Our conclusion is that the quality of DNA obtained by the rapid extraction methods is suitable for single genotype PCR reactions but not reliable for undertaking multiplex PCRs.

6.10 Discussion

We have produced, and tested, a suite of highly polymorphic microsatellite DNA markers that will provide an efficient analysis of both parentage and relatedness within the Tasmanian Atlantic salmon population. The availability of such a test allows for the pedigree (parental assignment) and relatedness of individuals within a mixed family cohort to be determined. As such, progeny from multiple fertilizations can be cultured as a single cohort within a single culture unit prior to reaching a size appropriate for physical tagging. For a selective breeding program this reduces the capital and maintenance costs associated with multiple individual-family culture units. In addition it reduces the variation in environmental influences on early development that may affect variation in later production traits between families. A uniform developmental environment is a distinct advantage in a selective breeding program and critical for estimating trait

heritability (Hartl and Clark, 1997; Villanueva et al. 2002). When families are maintained in separate tanks it can be problematic to differentiate environmental effects from true genetic differences, and early developmental history can influence later performance.

Our suite of markers was tested under current commercial conditions of complex multiple fertilizations (mixed ova from multiple females fertilized with mixed milt from multiple males) and resulted in unequivocal assignment of parents to > 95% of the progeny. This figure is intermediate to those reported by O'Reilly et al. (1998) and Norris et al. (2000) of 81.6% to 99.6%. However, a discrimination or assignment result is dependent upon the number of loci and their exclusion power (number and frequency of alleles in the population) and the level of half-sibs present in the test sample. We have also shown that from single pair crosses we are able to assign parentage to almost 100% of the progeny. Therefore under a selective breeding program with a full-sib family protocol (i.e. single pair matings) we are confident that parental assignment can be made to 99% of the progeny. Achieving 100% assignment will always be difficult given the possibility of human and technical errors at various stages in the hatchery and laboratory, and the possibility of random mutations.

Errors in the genotype data can cause problems for parentage assignment. O'Reilly et al. (1998) reported an error rate of 2-3% per allele scored. Human error can occur at the hatchery and in the laboratory. Tissue samples may not be collected from all of the contributing parents; there may be an unrecorded change in the cross design; there may be a mix up of samples during either the DNA extraction procedure, the PCR reaction or the loading of the samples on the gels; and there may be errors in genotyping. If both hatchery and laboratory staff are aware of these sources of error they can be kept to a minimum. Laboratory errors should decrease as the process becomes more automated. Mutations may also play a small role in causing mismatches between related individuals. The advantage of microsatellites is their high variability, but this is due to their high mutation rate of 10^{-4} (O'Reilly et al. 1998, Villanueva et al. 2002) which can therefore cause errors in parentage assignment. Some non-assignments seen during this study were thought to be caused by mutations although this needs to be confirmed by direct sequencing. The presence of null alleles can also cause errors in parentage assignment. Null alleles occur when there is a mutation in the priming site causing the non-amplification of alleles. Loci with known null alleles were excluded from our suite pedigree loci. Mutations in the PCR priming site may not be present in all individuals so although none of the loci in the current suite are known to have null alleles that is not to say that null alleles will be not be observed in other individuals for these loci.

We have designed a pedigree test that incorporates 10 microsatellite loci shown to be highly polymorphic (>10 alleles observed in the population sub-sample examined to date) with relatively even allele frequency distribution. Under current PCR conditions and primers these loci could be multiplexed only in three separate reactions, with two of these able to be co-loaded on a gel for analysis. Multiplex PCRs are efficient methods when undertaking bulk analyses but we found they can lead at times to technical problems in some instances with individual loci (and not always the same locus) not amplifying. Therefore repeat PCRs are sometime required for some individual fish.

For a commercial breeding program dependent on DNA pedigrees, samples from thousands of progeny would need to be DNA extracted and genotyped per year, therefore a commercial laboratory able to handle such numbers would need to be contracted. No

commercial genetic laboratory in Australia is currently familiar with analysing Atlantic salmon DNA, and would require lead in time for optimisation before undertaking such bulk analyses (they may with further optimisation be able to combine all 10 loci in one multiplex reaction). Quotes obtained in April 2004 from two leading commercial Australian genetic services are:

Australian Genome Research Facility (AGRF) (www.agrf.org.au)

Cost per individual for 10 genotypes (includes PCR, electrophoresis and genotyping) run on two gels – $27.00.

Note – AGRF does not offer a DNA extraction service, so an additional cost per individual of $6 to $9 would need to be added for this to be undertaken in-house or by a sub-contractor.

It is not known whether a reduction in cost would be available for large sample runs (e.g. sample sizes of greater than 1 000 and/or 5 000)

AgGenomics Pty Ltd. (www.aggenomics.com.au)

Cost per individual for 10 genotypes (includes DNA extraction, PCR, electrophoresis and genotyping):

	3PCR reactions/2 gels	3PCR reactions/1 gel
<1 000 samples	$28.60	$23.20
>1 000 samples	$22.90	$18.60
>5 000 samples	$19.00	$15.50

An alternative option that is now available is to use an overseas commercial laboratory specialising in salmonid DNA. Such a service has recently become available through Landcatch Natural Selection Limited (www.landcatch.co.uk). The service provided would use their suite of 10 microsatellite loci developed, optimised and routinely used for pedigree assignment of Atlantic salmon from European and Chilean hatcheries.

In February 2004, Landcatch agreed to genotype a set of DNA samples from Tasmania to enable us to compare the resolving power of their suite of loci compared to our own suite of 10 loci. We sent 96 anonymous DNA samples comprising parents and progeny from various NORTAS and SALTAS cohorts that we had already analysed. As with our loci all the progeny examined were correctly assigned to parents by the Landcatch suite of loci.

Whether the Landcatch and CSIRO suites of markers share loci is unknown as the 10 Landcatch loci are only identified as Ssa1 to Ssa10. Three of the Landcatch loci had low numbers of observed alleles (< 8) in the Tasmanian individuals, and two of these loci are reported to have low allele numbers (< 10) in Northern Hemisphere samples as well but do not restrict the overall exclusion capability of the set of 10 loci (pers. comm. Alastair Hamilton, Landcatch). The other Landcatch loci were all highly polymorphic and have allele frequencies relatively evenly distributed similar to the CSIRO selected loci. The comparative genotyping price is:

Landcatch Natural Selection Limited

Cost per individual for 10 genotypes (includes DNA extraction, PCR, electrophoresis and genotyping) is 12 to 14 Euros (at an exchange rate of $0.65, this is equivalent to $18.00 to $22.00); there is potential cost reduction with samples sizes greater than 4 000 samples.

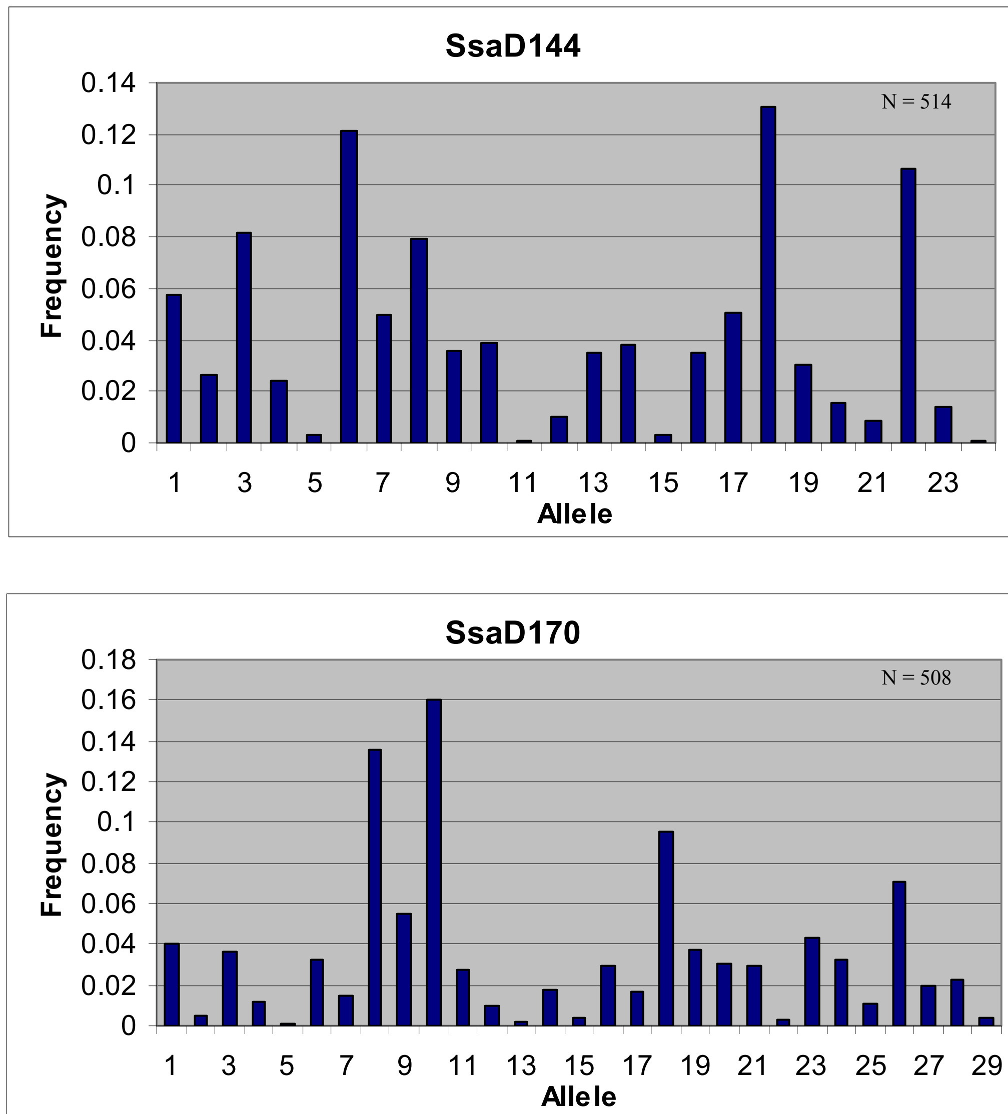
6.11 Conclusion

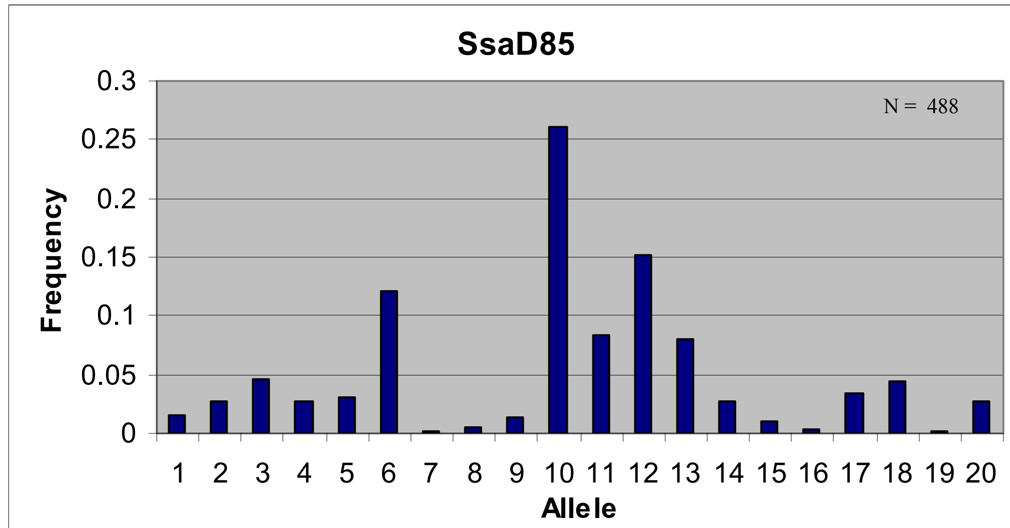
We have selected and tested a suite of loci with a high resolving power for parentage and relatedness analyses in the Tasmanian Atlantic salmon population, these loci can be multiplexed in three PCR reactions (and this may be improved to one reaction with further optimisation).

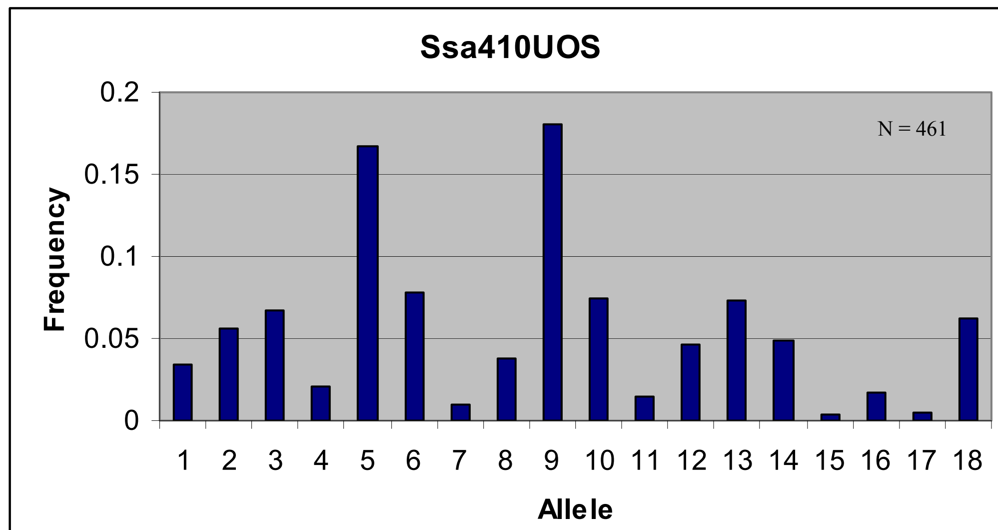
Commercial genetic laboratories in Australia (e.g. AGRF and Genetype) can be contracted to undertake DNA extraction and/or genotyping services. The disadvantage in using these laboratories is the relative inexperience in working with Atlantic salmon DNA and the particular set of loci, although with time this would be overcome.

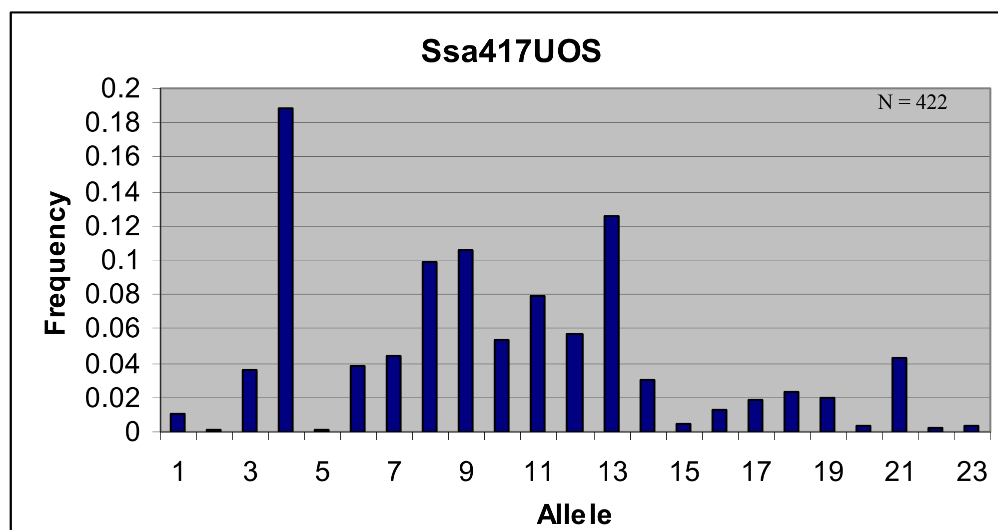
A commercial parentage and relatedness service for salmonids is provided by Landcatch Natural Selection Limited using a different set of loci. This company undertakes DNA extraction and genotyping services for European and Chilean hatcheries. The advantage of this service is that the laboratory is very experienced with salmonids and their particular set of loci. There is little difference between the resolving power of the Landcatch and our suites microsatellite loci, with both capable of providing a high resolution of parentage assignment. A minor advantage to using the Landcatch service is that any individuals that could not be unambiguously assigned to a parental pair using their 10 loci, could potentially be genotyped with one or more of the loci from our suite to provide additional resolving power.

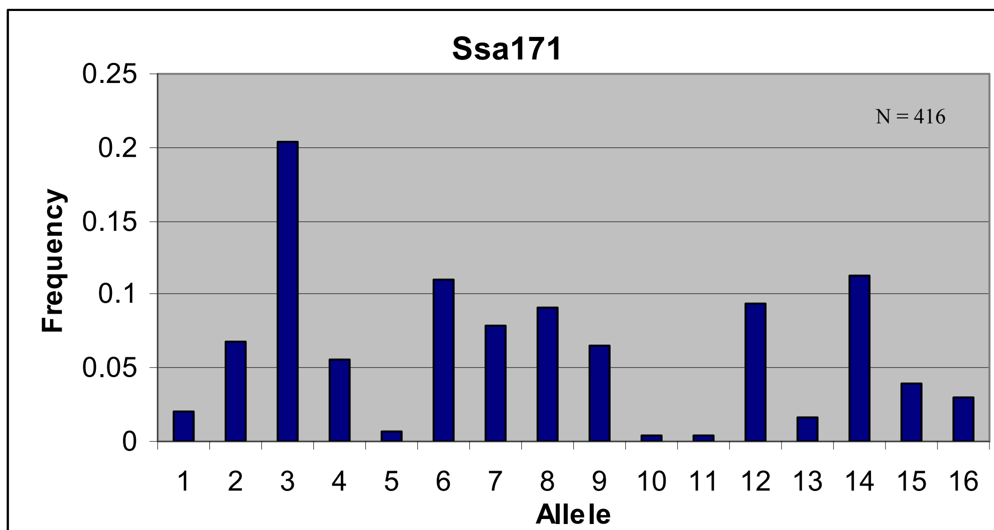
Figure 6.1. Allele frequency distributions for the 14 microsatellite loci considered in suite B. The first 10 are the selected final suite of loci. N is number of individuals in the Tasmanian population genotyped for each locus.

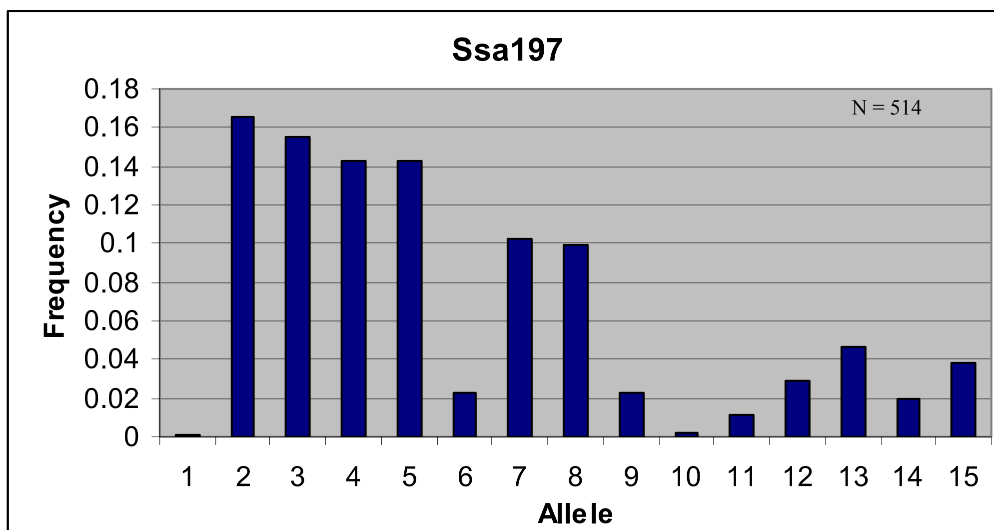


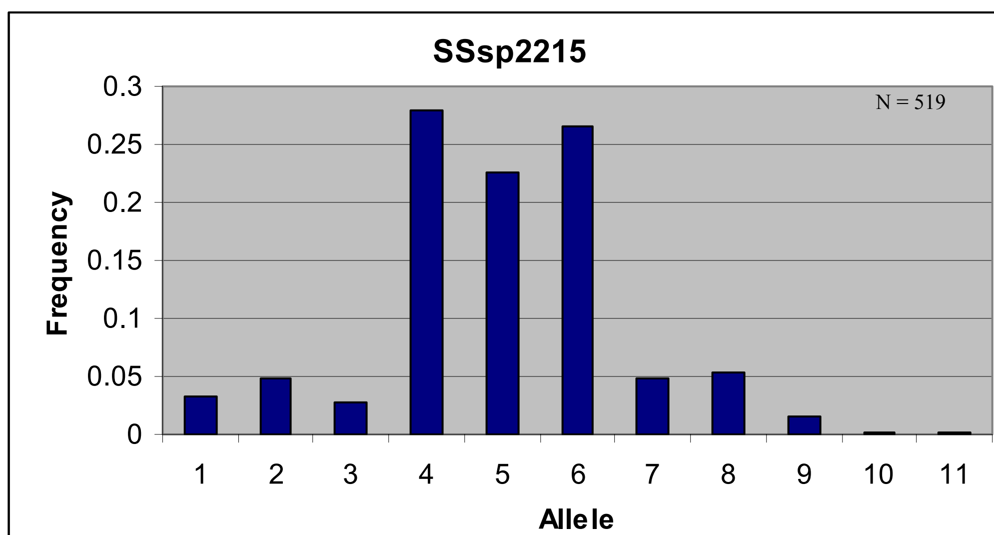

SsaD85
N = 488
Frequency
0.3
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Allele

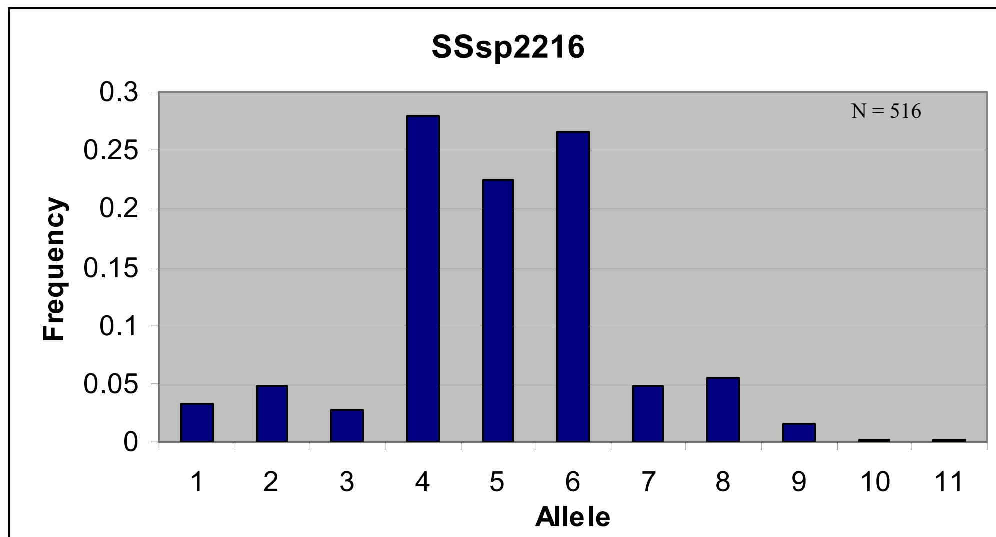

Ssa410UOS
N = 461
Frequency
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Allele

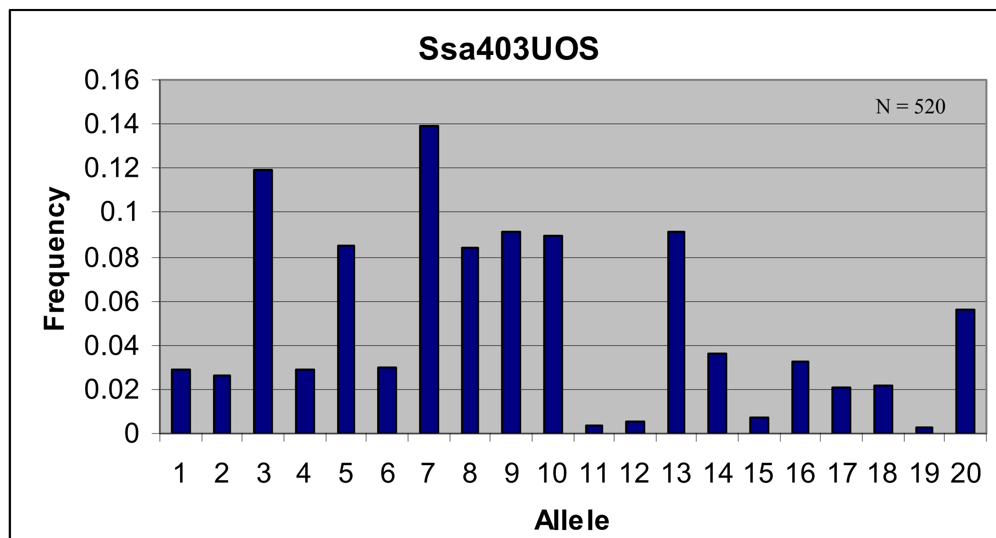

Ssa417UOS
N = 422
Frequency
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23
Allele

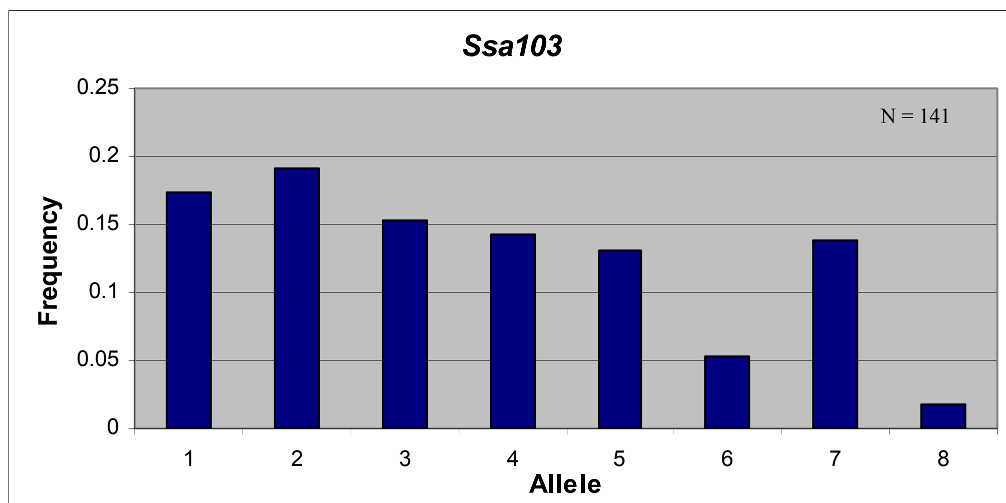

Ssa171
N = 416
Frequency
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Allele

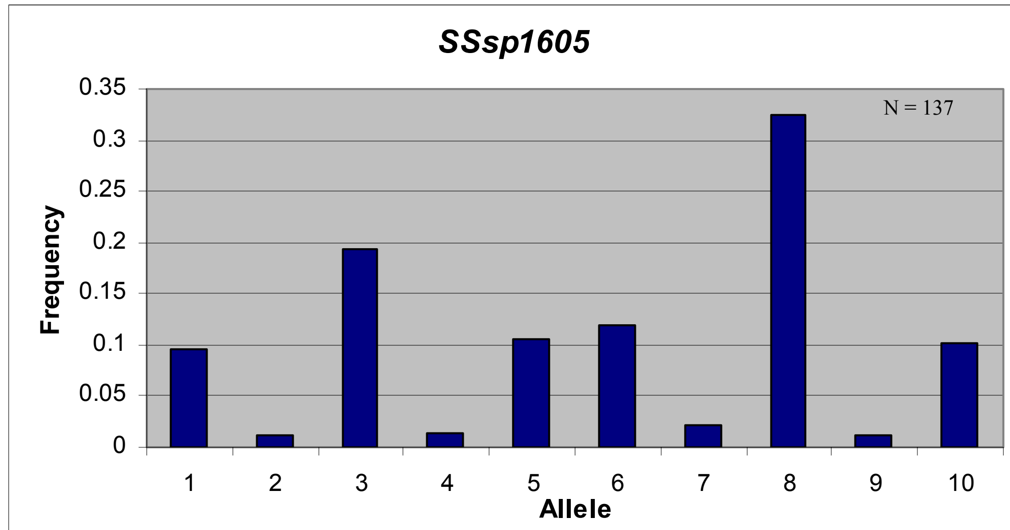

Ssa197
N = 514
Frequency
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Allele

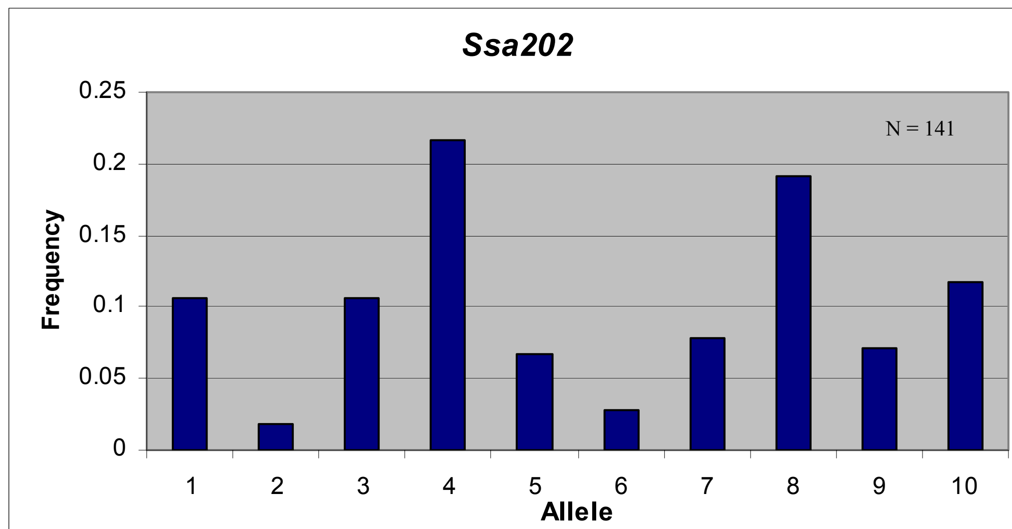

SSsp2215
N = 519
Frequency
0.3
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11
Allele

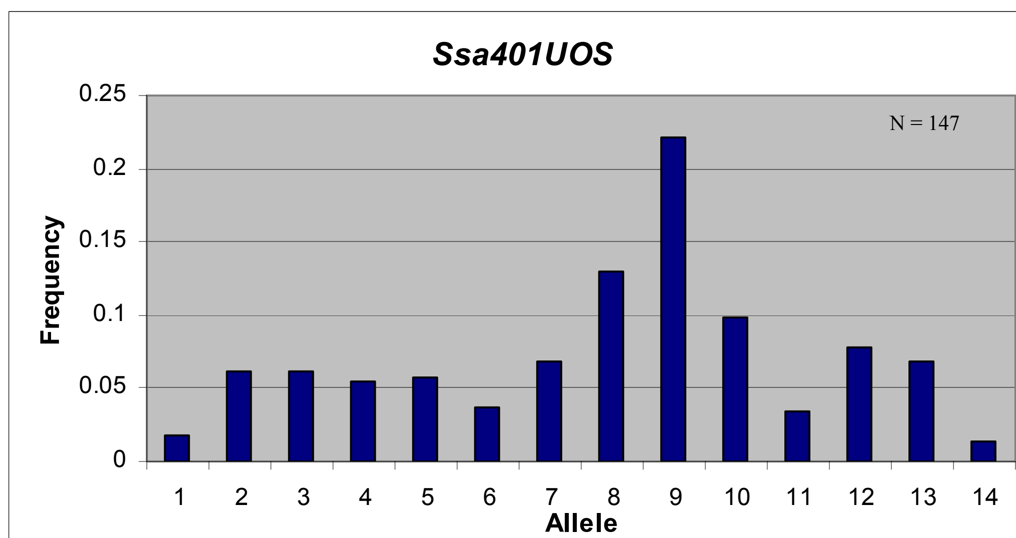

SSsp2216
N = 516
Frequency
0.3
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11
Allele


Ssa403UOS
N = 520
Frequency
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Allele


Ssa103
N = 141
Frequency
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8
Allele


SSsp1605
N = 137
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10
Allele


Ssa202
N = 141
Frequency
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10
Allele


Ssa401UOS
N = 147
Frequency
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Allele

7. GENETIC CHANGE IN THE TASMANIAN POPULATION

The research below was undertaken towards achieving Objective 3.

7.1 Introduction

Genetic changes in hatchery produced populations can occur over time through various factors, including low initial breeding numbers, inappropriate mating designs (low effective broodstock numbers), genetic drift, inadvertent inbreeding, or domestication selection (Verspoor 1988, Hedgecock et al. 1992, Winkler et al. 1999). The Tasmanian Atlantic salmon industry is based on a closed hatchery population that originated in the mid 1980s from Gaden, New South Wales; the location of the original 1960s importation from Canada. Since the original importations to Gaden and subsequently to Tasmania, no further introductions of genetic material have occurred to either population due to strict importation and quarantine guidelines. As the breeding programs for the Australian Atlantic salmon industries are fully reliant on non-pedigreed broodstock, it is important for long-term sustainability that the genetic health of the population is monitored. Any major loss of genetic variation could be commercially costly and difficult to rectify.

Within the Tasmanian Atlantic salmon population, genetic change has been previously examined by comparing variation at allozyme, mitochondrial DNA (mtDNA) and microsatellite loci with that observed in a sample from its progenitor River Philip population in Nova Scotia, Canada (Ward et al., 1994; Elliott and Reilly 1998; Reilly et al., 1999). There was no evidence of any change at the allozyme or mtDNA loci. However, the microsatellite analyses showed significant reduction in heterozygosity, and low estimates of per-generation effective population sizes consistent with a short-term moderate bottleneck. In addition, allele frequency differences suggested that three discrete year-class lines may have been established within the Tasmanian population. The previous microsatellite analyses and other studies (e.g. Norris et al., 1999) have demonstrated that these hypervariable markers are more sensitive than allozyme or mtDNA to changes in genetic variation.

A major assumption in the previous analyses was that the 1991 year-class River Philip sample (collected in 1992) was a random and unbiased representation of the original progenitor population at the time of exportation of ova to Australia in the late 1960s. This was potentially invalid on two grounds. Firstly, the sample was from a hatchery spawning of only 16 wild returning females and 14 wild returning males. Secondly, the variance in alleles between the samples represented an accumulation of change along two divergent lines of descent from the original 1960s progenitor population – one for the River Philip population and the other the Australian population initially in New South Wales and then in Tasmania.

This study therefore had two specific aims:

- to clarify the issue of genetic change within the Tasmanian population using the analysis of DNA extracted from archived scales from River Philip fish collected closer to the time of the original exportation of ova to Australia; and
- to monitor microsatellite variation between six consecutive year-classes of the Tasmanian population.

7.2 Methods

7.2.1 Samples

Microsatellite variation was examined in 10 samples of Atlantic salmon from three locations.

Location: River Philip, Nova Scotia, Canada
The progenitor population for the Australian Atlantic salmon population came from the River Philip. We examined two samples from this location.

- Scales collected from River Philip fish caught by fish traps and line fishing in June-October 1971(n=50) and August-November 1972 (n=51). Dried scales from each fish were stored in individual envelopes. Details about the fish (date of collection, sex, method of capture) were noted on each envelope. These samples were archived at the Department of Oceans and Fisheries, Nova Scotia.
- Muscle samples collected in 1992 (n=65) from the Mersey Federal Fish Hatchery Queen's County, Nova Scotia. These fish were the progeny of 1991 wild returning River Philip fish. Details of these fish are listed in FRDC project report 96/347 (Elliott and Reilly 1998) and Ward et al. (1994).

Location: SALTAS hatchery, Wayatinah, Tasmania.
We examined samples from seven year-class cohorts.

- 1992 year-class: muscle samples collected in 1993 (n=64).
- 1996 year-class: whole fish collected in 1997 (n=99).
- 1997 year-class: finclips collected in 2000 (n=94).
- 1998 year-class: blood samples (female fish only) and finclips (females and males) were collected from 81 individuals in 2001.
- 1999 year-class: muscle samples collected in 2001 (n=100).
- 2000 year-class: whole fish collected in 2001 (n=101).
- 2001 year-class: whole fish collected in 2002 (n=80 from two separate tanks, 40 per tank).

Location: NSW Fisheries hatchery, Gaden, NSW.
This is an intermediary population between the Canadian progenitor population and the Tasmanian population. One sample of whole fish was collected in 1997 from the 1996 year-class (n=100).

The 1992 Canada sample, the 1992 and 1996 Tasmanian year class cohorts and the 1996 NSW sample were previously examined for microsatellite variation at eight loci in FRDC projects 95/80 and 96/347 (Elliott and Reilly 1998). In this study we have strengthened the previous analysis with the inclusion of three new and potentially more informative loci.

The muscle, finclip or whole fish samples were either frozen at -80°C or stored in 95% ethanol at room temperature. The blood samples were frozen at -20°C. The scales were stored dry at room temperature.

7.2.2 Total genomic DNA extraction

DNA was extracted from 10-20 mg of muscle or finclip using one of three methods: a modified CTAB protocol, a DNeasy™ Tissue Kit (Qiagen) or a shortcut extraction method. DNA was extracted from 25-100µL blood using a modified protocol listed in

Appendix 15.2 FRDC Final Report Projects 95/80 and 96/347. DNA was extracted from 1-2 scales using a modified protocol listed in Adcock et al. (2000) and Whitmore et al. (1992). All DNA extraction protocols are listed in Appendix A, and all methods provided suitable quality DNA template for microsatellite analyses.

7.2.3 Microsatellite analysis

Eleven microsatellite loci were used to examine the genetic variation within and between the three locations (Table 7.1). The eight dinucleotide repeat loci were used in the previous FRDC study (FRDC projects 95/80 and 96/347).

Table 7.1. Microsatellite loci used to examine genetic variation.

Locus	Repeat type	Annealing temperature (°C)	Size range (base pairs)	Number of observed alleles	Reference
cmrSs1.22	dinucleotide	60	186 - 264	20	1
cmrSs1.14	dinucleotide	65–60-55	136 - 170	9	1
cmrSs1.10	dinucleotide	60	218 - 222	3	1
D30	dinucleotide	53–52	230 - 244	8	2
20.19	dinucleotide	65–60-55	78 - 100	5	2
F43	dinucleotide	65–60-55	103 - 135	11	2
F49	dinucleotide	58	174 - 176	2	3
5.27	dinucleotide	53–52	124 - 128	2	3
Ssa202	tetranucleotide	59	247 - 321	18	4
Ssa197	tetranucleotide	59	159 - 229	18	4
Ssa171	tetranucleotide	55	210 - 274	26	4

1. FRDC projects 95/80 and 96/347 and Reilly et al. (1999)
2. Sanchez et al. (1996)
3. Clabby (1996) unpublished. Genbank accession numbers U37496/7, U37491
4. O'Reilly et al. (1996)

7.2.4 PCR amplification of microsatellite loci

It was impractical to quantitate the DNA of every individual. DNA was routinely diluted 1/20 for PCR amplifications (approximately 20-40 ng/ul) with the exception of DNA extracted from the archived scales which was diluted 1/5. The DNA was diluted with autoclaved purified water (MilliQ synthesis filtration system, Millipore)

All PCR amplifications were performed in a GeneAmp PCR System 9600 thermal cycler (Perkin Elmer). PCR amplifications using non-archival DNA were in a 25µL volume containing 1x PCR buffer, 1.5mM $MgCl_2$, 0.2µM forward primer (fluorescent labelled), 0.2µM reverse primer, 200µM each dNTP, 1 unit Taq polymerase (Fisher Biotech) and 7 µl of diluted DNA.

Cycling parameters for loci *cmrSs1.22, cmrSs1.10, F49, Ssa202, Ssa197* and *Ssa171* began with an initial denaturation at 95°C for 3 min, followed by 35 cycles of 95°C for 30 sec, the locus specific annealing temperature (see Table 7.1) for 30 sec and an extension at 72°C for 60 sec. This was followed by a final extension at 72°C for 20 min.

A touchdown protocol was used to amplify loci *D30* and *5.27*, with an initial denaturation at 95°C for 3 min followed by 15 cycles of 95°C for 30 sec, 53°C for 30 sec and 72°C for 60 sec and then 20 cycles of 95°C for 30 sec, 52°C for 30 sec and 72°C for 60 sec. This was followed by a final extension at 72°C for 20 min. A touchdown protocol was also

used to amplify loci *cmrSs1.14, 20.19* and *F43*. The initial denaturation was at 95°C for 3 min followed by 5 cycles of 95°C for 30 sec, 65°C for 30 sec and 72°C for 60 sec, then 10 cycles of 95°C for 30sec, 60°C for 30 sec and 72°C for 60 sec, then 20 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 60 sec. This was followed by a final extension at 72°C for 20 min.

Scale DNA amplifications were in a 25µL volume containing 1x PCR buffer, 1.5mM $MgCl_2$, 0.3µM forward primer (fluorescent labelled), 0.3µM reverse primer, 200µM each dNTP, 1 unit Taq polymerase (Fisher Biotech) and 2 µl of diluted DNA. Some modifications were made to the cycling parameters: the total number of cycles was increased from 35 to 40 and the final 72°C extension was increased to 60 minutes. Precautionary measures to avoid cross-contamination of modern salmon DNA with the 30 year old scale DNA included undertaking DNA extractions in a laboratory not used for molecular biology and the sterilisation, by autoclaving or UV irradiation, of the PCR tubes and PCR reagents (not Taq polymerase or primers). Negative controls were used to detect non-specific amplifications.

Amplification products were electrophoresed on 2% agarose gels to check the quality of the PCR reaction. PCR products were diluted 3-fold and run on an ABI377 Prism DNA autosequencer with an internal size standard (Genescan™-500 Tamra™ size standard, Applied Biosystems). Gels were analysed using Genescan Analysis Software 3.1 (PE Applied Biosystems) and Genotyper DNA Fragment Analysis Software 1.1.1 (PE Applied Biosystems). Samples from Reilly et al. (1999), Elliott and Reilly (2003) and FRDC projects 95/80 and 96/347 were run alongside samples collected for this project to ensure consistency in allele sizing between all samples. This exercise resulted in some minor revision of allele scoring and binning in samples run previously.

7.2.5 Statistical analysis

Tests for conformance to Hardy-Weinberg expectations and observed (H_o) and expected (H_e) heterozygosity values were calculated using Arlequin v.2.0 (Schneider et al. 2000). NULLTEST (Bill Amos, Department of Zoology, Cambridge, CB2 3EJ, UK) calculated the frequency of null (non-amplifying) alleles at each locus. Genotypic disequilibrium (the independence of the loci) and genic differentiation (comparison of allele frequencies across all samples) were examined using Genepop v.3.2 (Raymond and Roussett 1995). Pairwise F_{st} values were calculated in Arlequin v 2.0, per locus F_{st} values were calculated in Genepop v.3.2. Estimates of the per-generation (average) effective population size and a test of random genetic drift based on these estimates were made from the microsatellite allele-frequency variances between the Australian (derived) and Canadian (progenitor) samples (Hedgecock and Sly, 1990; full description is available in Elliott and Reilly 1998). When multiple tests were carried out, the significance level of $\alpha = 0.05$ was adjusted by the sequential Bonferroni procedure for multiple tests (Rice, 1989).

7.3 Results

Good quality DNA was successfully extracted from all of the 30-year-old scale samples. The number of microsatellite loci successfully amplified from each scale sample ranged from 7 to 11, with 77% of the scale samples amplifying for all loci.

The three new tetranucleotide repeat loci produced clear peaks with few stutter bands. The allele sizes however for these loci were not always the expected four base pair

increments; locus *Ssa171* for example had numerous alleles that were two base pairs apart as reported by O'Reilly et al. (1996).

The total number of observed alleles per locus ranged from 2 (*5.27* and *F49*) to 26 (*Ssa171*), with a range of 2 to 15.4 for the mean number of alleles per locus (Table 7.2). The mean heterozygosity per locus ranged from 0.096 for locus *cmrSs1*.10 to 0.894 for locus *Ssa171*, with eight loci having a mean level of heterozygosity above 50%. The four most polymorphic loci in terms of numbers of alleles (*Ssa171, Ssa197, Ssa202* and *cmrSs1.22*), also had the highest average heterozygosity (85% to 89%) across all samples (Table 7.2).

Table 7.2. Summary statistics for the 11 microsatellite loci across all samples analysed. Total (Ta) and mean (Ma) number of alleles observed, mean and standard error of the observed (Ho) and expected (He) heterozygosity, and total number of individuals examined (n) at each locus

Locus	Ta	Ma	Ho	He	n
Ssa 171	26	15.4	0.894±0.015	0.896±0.006	849
Ssa 197	18	13.1	0.888±0.013	0.874±0.005	873
Ssa 202	18	10.8	0.871±0.010	0.853±0.010	850
cmrSs1.10	3	2.7	0.096±0.018	0.107±0.018	879
cmrSs1.14	9	4.8	0.678±0.025	0.645±0.019	862
cmrSs1.22	20	13.1	0.854±0.016	0.828±0.013	865
D30	8	4.0	0.505±0.024	0.505±0.016	862
5.27	2	2.0	0.341±0.018	0.359±0.015	860
F49	5	2.0	0.267±0.017	0.290±0.021	867
20.19	2	4.2	0.679±0.027	0.659±0.015	871
F43	11	5.5	0.557±0.060	0.661±0.026	860

Significant linkage disequilibrium across all samples was observed at 15 of the 55 paired tests (data not shown). This is more than would be expected by chance. However, five of these pairs, while significant across all samples, were not significant in any one sample, and no pairs were consistently in disequilibrium in all samples. In only two samples was disequilibrium observed, Canada 1991 and Tasmania 2000. The Canada 1991 sample had the highest number of loci showing linkage disequilibrium (9 pairs); this sample as discussed previously is derived from a small number of parents (n=30), with perhaps an even smaller effective number of parents and is therefore similar to a population which has gone through a bottleneck. It is known that such populations can show an increased level of linkage disequilibrium (Hartl and Clark, 1997). There is no evidence to suggest true linkage between any of the 11 loci.

Allele frequencies per locus per sample are provided in Appendix 3 (Tables A.1 to A.11). In all, 122 alleles were observed over the 11 loci, with 112 of these being observed in the combined Canada 1971/72 scale sample (94 in the 1971 and 100 in the 1972 samples). This compares with 87 in the Canada 1991 sample and 84 in total in all the Tasmanian samples. Of the ten alleles not observed in the Canada 1971/72 sample, four were observed in both the Canada 1991 and Tasmanian samples, four only in the Canada 1991 sample and two alleles, both at locus *Ssa197*, were unique to the Tasmanian samples (Appendix 3. Table A.2).

The lowest mean number of alleles per locus observed in a sample was 5.8 for the Tasmania 1992 sample (n = 64), with the other Tasmanian samples ranging between 6.5 to 7.0, in contrast to 10.2 for the Canada 1971/72 sample (Table 7.3). Sample sizes of 80

to 100 individuals would appear to provide adequate representation of diversity in the Tasmanian population with 84 to 92% of the total alleles seen in over 600 fish being observed in such sample numbers. However, the two Tas2001 samples of 40 individuals each provided a similar level of representation to each other and the combined sample (Table 7.3). Our results also indicate that the sampling protocols for each year-class sample did provide a random unbiased sample of the population.

The mean (across the eleven loci) observed heterozygosity in a sample varied from 0.527 (Tasmania 1992) to 0.649 (Canada 1972 and Tasmania 2001A) (Table 7.3); with no significant difference (t-test) observed between the Canadian and Australian samples.

Table 7.3. Summary statistics for all samples analysed across the 11 microsatellite loci. Total (Ta) and mean (Ma) number of alleles observed, mean and standard error of the observed (Ho) and expected (He) heterozygosity, and total number of individuals in the sample (n). Canada 1971/72 is the combination of the Canada 1971 and 1972 samples, and Tasmania 2001 is the combined Tasmania 2001a and 2001b samples.

Sample	n	Ta	Ma	Ho	He
Canada 1971	50	94	8.6	0.645±0.082	0.649±0.077
Canada 1972	51	100	9.1	0.649±0.083	0.658±0.079
Canada 1971/72	101	112	10.2	0.647±0.082	0.652±0.079
Canada 1991	65	87	7.9	0.626±0.088	0.617±0.081
Tasmania 1992	64	64	5.8	0.527±0.095	0.560±0.085
Tasmania 1996	99	71	6.5	0.547±0.086	0.588±0.078
Tasmania 1997	94	77	7.0	0.614±0.082	0.601±0.080
Tasmania 1998	81	74	6.7	0.586±0.081	0.601±0.079
Tasmania 1999	100	74	6.7	0.606±0.091	0.608±0.088
Tasmania 2000	101	72	6.6	0.581±0.083	0.587±0.078
Tasmania 2001a	40	72	6.6	0.649±0.084	0.610±0.079
Tasmania 2001b	40	71	6.5	0.636±0.086	0.620±0.080
Tasmania 2001	80	77	7.0	0.643±0.085	0.609±0.081
NSW 1997	100	73	6.6	0.566±0.096	0.585±0.084

Genotype frequencies were in agreement with Hardy-Weinberg equilibrium with eight exceptions (Appendix 3. Table A.12). Four samples as previous reported (Elliott and Reilly, 1998 and 2003) exhibited significant disequilibrium at locus *F43* (Canada 1991, NSW 1997, Tasmania 1992 and Tasmania 1996), and the Canada 1991 sample also displayed disequilibrium at four other loci (*Ssa171, Ssa197 Ssa202* and *cmrSs1.22*). As expected, the four samples not in Hardy-Weinberg equilibrium at locus *F43* due to high heterozygote deficiency all had high frequencies of putative null alleles (Appendix 3. Table A.13). However, from our latest results, the cause of the *F43* disequilibrium in the four original samples analyzed was most likely a technical problem (PCR preferential amplification/scoring artefact) as all the new samples run in this study were in equilibrium, and had higher frequencies of the larger alleles (especially allele **125*) and higher observed heterozygosities (Appendix 3 Tables A.11 and A.12). In some of the following analyses the *F43* locus is omitted. The disproportionate number of disequilibrium tests for Canada 1991 is most probably a consequence of this sample originating from only a small number of parents.

There were highly significant differences ($P < 0.001$) in allele frequencies between all samples analyzed at all loci except locus *5.27*. The overall F_{st} value was 0.0216, with individual locus values ranging from 0.0038 for locus *5.27* to 0.0365 for locus *F43* (Appendix 3 Table A.14). The pairwise F_{st} comparisons highlighted small ($F_{st} < 0.050$) but significant differences between most of the samples (Appendix 3 Table A.15). The highest F_{st} values were generally associated with the Tasmania 1992 and Canada 1991 samples. The omission of locus *F43* decreased the F_{st} values associated with these (and the Tasmania 1996) samples slightly, but had no bearing on the significance levels. The pairwise comparisons where no significant difference was observed between allele frequencies were for:

- Canada 1971 and 1972 ($P = 0.283$). The two archived scale samples, which could therefore be combined as one sample for comparative analyses with the Tasmanian samples.
- Tasmania 2001a and 2001b ($P = 0.797$). The two samples could therefore be combined as one sample, and demonstrates that random selection of the samples from one or more tanks did not significantly affect the sample representing that year-class cohort.
- Tasmania 1998 was not significantly different to Tasmania 2001a ($P = 0.316$), 2001b ($P = 0.849$) and 2001 combined ($P = 0.582$). The two cohorts represented here are three years apart, and 87% of the broodstock for the 2001 year-class were 1998 year-class fish; the remaining 13% were 1997 year-class females.
- The relationship between Tasmania 1997 was marginally with Tasmania 1998 ($P = 0.027$), Tasmania 2000 ($P = 0.002$), 2001a ($P= 0.006$) and 2001b ($P = 0.122$) (but significantly different to the combined 2001 sample ($P < 0.001$)).
- The comparisons of the Tasmania 1996 and Tasmania 1999 samples ($P = 0.003$) and the Tasmania 1997 and Tasmania 2000 samples ($P = 0.002$) were also marginal, with non-significance dependent on Bonferroni correction for multiple tests.

The closer relationship of the allele frequencies between samples from year-classes three years apart (1996 and 1999, 1997 and 2000 and 1998 and 2001), but little relationship between other year-classes, suggests that there is still some remnants of three discrete lines in the Saltas population despite the mixing of three-year old and four-old broodstock.

For the determination of the mean effective population sizes of the derived populations (Canada 1991, NSW and Tasmania) and the expected loss of neutral alleles within the derived population samples due to random genetic drift, the two Canada scale samples (1971 and 1972) were combined as a single progenitor Canadian population sample (Canada 1971/72). In addition the Tasmania 2001 samples were treated as a single sample. The omission of locus *F43* for comparisons with the four original samples had little bearing on the overall results, and so is included.

The individual locus allele-frequency variance estimates (F_K) between the Canadian progenitor population sample (Canada 1971/72) and the various samples of the derived populations were: Canada 1991 0.0093 to 0.0500, NSW 0.0001 to 0.0842, and Tasmania 0.0007 to 0.1922 (Appendix 3 Table A.16). The weighted average multi-locus estimates were lowest for the Canada 1991 comparison (0.0334) and highest for the Tasmanian 1992 sample (0.0621). There was no significant difference between any of the nine data sets.

The resulting estimates of the per-generation effective population sizes (N_K) for the seven Tasmanian samples compared to the Canadian progenitor sample (Canada 1971/72) ranged from 102 (1992 sample) to 207 (1997 sample) fish per generation (Table 7.4). These values represent a 44% to 82% increase on the estimates obtained if the Canada 1991 sample was used as a proxy for the progenitor population (Table 7.4); as in the previous study (Elliott and Reilly 1998).

Table 7.4. Estimated genetic drift between the derived Australian and the progenitor Canadian populations based on 11 microsatellite loci. The progenitor population is represented by both the combined 1971 and 1972 sample and the 1991 sample. The assumed numbers of generations of isolation were 10 for the Tas92 sample, 11 for the Tas96 to Tas98 samples and 12 for the Tas99 to Tas01 samples. The mean and standard deviation are provided for the estimated mean effective population size (N_K), as well as the number of observed alleles (Na), the number of new(N) and lost (L) alleles in the derived population, and the number of neutral alleles expected to be lost (expL) and the significance of the actual loss (*P*).

		Can 71 and 72 (Na = 112)						Can 91 (Na = 87)					
		N_K		Alleles				N_K		Alleles			
	Na	mean	SD	N	L	expL	*P*	mean	SD	N	L	expL	*P*
Tas92	64	101.98	17.76	5	53	34.17	0.0001	70.67	13.40	7	30	23.23	0.1288
Tas96	71	128.18	21.87	5	46	31.41	0.0031	86.37	15.95	9	25	21.45	0.4410
Tas97	77	207.08	40.35	3	38	22.15	0.0003	113.34	21.80	12	22	17.00	0.2237
Tas98	74	154.84	28.29	5	43	27.63	0.0011	90.24	17.09	9	24	20.66	0.4749
Tas99	74	180.14	32.69	3	41	26.38	0.0017	112.72	21.35	10	23	18.45	0.2878
Tas00	72	181.46	33.30	1	41	26.24	0.0015	106.25	20.15	8	23	19.40	0.4248
Tas01	77	191.90	35.88	6	41	25.17	0.0005	105.21	20.00	10	20	19.56	0.9872
NSW97	73	142.86	25.11	4	43	29.22	0.0043	92.46	17.41	8	22	20.26	0.7528

The 1971/72 scales sample appears to be a good representative of the progenitor River Philip population. It has the highest number of alleles observed in any sample, and importantly, only six 'new' alleles were observed in all the Tasmanian samples. In contrast there were 15 'new' alleles in the Tasmanian samples compared to the Canada 1991 sample, which had four 'new' alleles not observed in the scale samples (Appendix 3 and Table 7.4). This and the higher N_K values suggest that the new 1971 and 1972 archive scales are more representative of the original progenitor population than the Canada 1991 sample. Despite the lower number of new alleles observed, there was still a high overall loss of alleles in the Tasmanian samples, 41 to 53 of a total of 112 observed in the Canada 1971/72 sample (Table 7.4). In all instances this loss was highly significant compared to the expected loss of neutral alleles due to random genetic drift and the estimated effective population sizes. This result is in contrast to the previous study where the observed loss of alleles was not significant, and this would still be the case despite the inclusion of the three new loci if the Canada 1991 sample was, as before, used to represent the progenitor population (Table 7.4).

For the 20 years between the collection of the scale samples and the sampling of the hatchery population in 1991, the estimated per-generation effective population size for the River Philip population is 147.057 (± 32.85) over an estimated six generations of isolation. Given only a maximum of 30 individuals contributed to the 1991 sample, this would suggest that the effective River Philip population in the other five generations was around 500 individuals (N_K is equivalent to the harmonic mean and so heavily influenced by a low value). A loss of 33 alleles in the Canada 1991 sample compared to the

combined Canada 1971/72 sample was also a significant loss (P = 0.001) compared to that expected of neutral alleles under random genetic drift and with the estimated effective population size.

7.4 Discussion

We have presented a comprehensive assessment of the level of genetic variation present in the Tasmanian Atlantic salmon population. This has been provided through the study of samples from six consecutive year-class cohorts and their comparison to a 1971/72 sample from the progenitor River Philip population in Canada from which the Australian population was derived in the mid 1960s.

It is not possible to determine loss of genetic variation in a population without first knowing its original levels of variation. The earlier studies on this issue used a 1991 hatchery sample from the River Philip; some 25 years after salmon were exported to Australia. It was not known how well this sample represented the mid-1960s population due to the lack of information on the temporal genetic stability of the River Philip population. However, it is clear from our current results and those of other workers that archived scales are a very useful historical DNA record for a fish population. Scales, some up to 60 years old, have been used for population analyses of salmonids: Atlantic salmon (Nielsen et al. 1997), steelhead trout *Oncorhynchus mykiss* (Heath et al. 2002) and brown trout *Salmo trutta* (Hansen 2002). We have shown that the 1971 and 1972 Canadian scale samples are a far more realistic representation of the 1960s progenitor population by virtue of the higher number of alleles observed, the lower number of ‘new’ alleles observed in the Tasmanian samples, a similar decrease in genetic variation seen in the Canadian 1991 sample to that in the Tasmanian samples, and more appropriate (based on hatchery practices) estimates of the per-generation effective population size for the Tasmanian population.

The loss of alleles in a population is usually the first indicator of a loss of genetic variation. The mean number of alleles observed is influenced by the sample size with fewer alleles expected in a smaller sample size (Verspoor, 1988). The method we have used to determine the statistical significance of the loss of alleles takes sample size into account, but importantly most of our samples (with the exception of the 1991 and 1992 samples) were of a similar size. The previous FRDC study found a non-significant loss of alleles in the two Tasmanian cohorts compared to the Canada 1991 sample which was used to represent the progenitor population (Elliott and Reilly, 1998 and 2003). In our current study, however, we have found a significant reduction in the number of alleles in all the hatchery samples compared to the wild 1971/72 Canadian samples.

This reduction in genetic variation as measured by number of alleles is significant compared to the expected loss of neutral alleles due to random genetic drift in a natural population with random mating and no migration or mutation. Mutation would not be expected to play a major role over the 10 to 12 generations of isolation for the Australian population, and there has been only the one introduction of genetic material to form the population. Therefore it is more likely that the loss of alleles is a consequence of low breeding numbers at some stage in the population’s history. Given the consistency of alleles observed in the Tasmanian and NSW samples this loss is most likely to have occurred at or soon after the original introduction to Australia

Heterozygosity values are another indicator of the genetic health of a population, but are less sensitive to a loss of variation as a population can lose significant numbers of rare alleles but still maintain high levels of heterozygosity. A significant reduction in mean microsatellite loci heterozygosity was observed in the previous FRDC study; however, we did not observed such a situation in this more expansive study.
The previous FRDC study (Elliott and Reilly 1998) hinted at some slight differences between the two Tasmanian year-classes examined suggesting that there may be distinct breeding lines forming in the Tasmanian population as a result of the three-year breeding cycle. Our current work has shown that there is more similarity between samples that are three years apart than between other pairs of samples. This suggests that despite the mixing of three-year old and four-year old broodstock there are still some remnants of three cohorts within the Saltas population.

The estimated per-generation effective population size for the Tasmanian population is 180 to 200 individuals over the ten to twelve generations since importation of the population from Canada. These estimates assume the Canada 1971/72 samples are representative of the River Philip population at the time of the mid 1960s importation of ova to Australia. This value is a summary of the population's breeding history over the successive generations. It is the harmonic mean of the effective breeding sizes, and so is heavily dependent upon the lowest breeding number over this time.

Our latest mean effective population size estimates are twice that predicted from the previous study (Elliott and Reilly, 1998 and 2003) but it was speculated at that time that a more realistic estimate was 160 to 180, based on the unknown history of the River Philip population over the intervening years between the 1960s and 1991, and that the Canada 1991 sample was not a truly random and unbiased representation of the progenitor population.

The higher effective population size estimates do reduce the circumstantial evidence of the previous suggestion that a short but moderate bottleneck occurred in the Australian population early in its introduction; however a bottleneck event cannot be dismissed. Our estimated per-generation effective population size is in agreement with recent broodstock numbers (1998 to 2001) of 900 to 1500, and previous numbers of 250 to 350 in Tasmania and 100 to 150 in NSW since introduction. The current estimate, based on a more appropriate base line for the progenitor population, would suggest that the minimum effective breeding number over the 10 to 12 generations would be in the order of 50 to 100 individuals.

In conclusion, the Atlantic salmon population in Tasmania has lost a significantly greater number of alleles than would be expected due to random genetic drift; however the number of remaining alleles is consistent between recent year-class samples suggesting the loss occurred at or early in the introduction to Australia. The measure of the mean effective population size is in accordance with recent hatchery practices and a possible small bottleneck in the initial generation(s) in Australia. The level of heterozygosity has remained quite high and consistent between year-classes. There is some minor evidence of three cohorts (based on three introductions of ova and three-year spawning cycle) within the population despite recent crossing of three-year old and four-year old broodstock. Despite the observed significant loss of alleles compared to the wild progenitor population our results vindicate the Tasmanian hatchery protocols that have

maintained a sufficiently large breeding population to preserve a genetically healthy population.

8. MHC VARIATION AND LINK TO AGD

The research below was undertaken towards achieving Objective 1.

The research was undertaken by James Wynne, and submitted in March 2004 as part of his fulfillment of the requirements for a Bachelor of Aquaculture Honours Degree. Reproduced below is a condensed version of the thesis: J. W. Wynne 2004 'MHC polymorphism in Atlantic salmon (*Salmo salar* L.): implications for an AGD resistance marker', School of Aquaculture, University of Tasmania.

The research was also supported through the Health Program of the CRC for Sustainable Aquaculture of Finfish (Aquafin CRC), in particular FRDC projects 2001/244 AGD Host-pathogen interaction and 2002/251 Development of a vaccine for amoebic gill disease: genomic and cDNA library screening for antigen discovery.

8.1 Introduction

We have demonstrated (Chapter 7) a significant loss in genetic variation at non-coding microsatellite DNA loci (as measured by number of alleles) in the Tasmanian Atlantic salmon population compared with its ancestral Canadian population. Irrespective of the loss mechanism, the observed change at the microsatellite loci is assumed to be representative of all loci (coding and non-coding) in the genome. The major histocompatibility complex (MHC) genes in teleosts and Atlantic salmon in particular, are characterised by their high level of polymorphism (Grimholt et al., 2002). Therefore we considered the MHC genes to be suitable coding regions for validating the degree of genetic change observed at microsatellite loci.

The MHC genes are known to play an important immunological role in many vertebrate species, and have been shown to be closely associated with the immune response and disease resistance in salmonids (Landry and Bernatchez 2001; Langefors et al., 2001a; Xia et al., 2002; Grimholt et al., 2003). Inbred rainbow trout populations in Japan have been shown to have a significantly lower level of MHC polymorphism compared to outbred populations (Xia et al., 2002), and Langefors et al. (1998) demonstrated that Atlantic salmon populations with limited MHC polymorphism have a higher incidence of disease (M74 syndrome) compared to populations with high MHC polymorphism.

The Tasmanian Atlantic salmon population had to survive two major importation events (from Canada to New South Wales and then to Tasmania) and has a demonstrated loss of genetic variation (microsatellite loci). We wished to determine whether the Tasmanian population has a similarly reduced level of MHC variation, and whether this might be associated with the observed high incidence of amoebic gill disease (AGD) in the Tasmanian population.

Anecdotal reports suggest there is some individual variation in resistance to AGD within the Tasmanian population, of which a portion is likely to be genetic, although no hard evidence exists to date. While several immune related genes may be possible candidates, the MHC genes, reported to be associated with resistance to bacterial and viral disease in Atlantic salmon, are candidates worthy of examination. Grimholt et al. (2003) reported that full-sibling families of Atlantic salmon resistant to ISA had a significantly higher frequency of a specific *MH* (sic) class IIα1 allele. Similarly, an *MHC* class IIβ1 allele

appears to be significantly more prevalent in families resistant to furunculosis compared to susceptible families (Langefors et al., 2001a).

This pilot study therefore had two aims:

- to compare MHC polymorphism in samples from the Tasmanian population with one from its ancestral Canadian population, and
- to test for any association between MHC alleles and AGD resistance/susceptibility.

8.2 MHC loci

In teleosts there are *MHC* class I and class II genes, with different biochemical and functional characteristics. The *MHC* class I is encoded by a single α locus that is composed of eight exons (protein coding sequences) with intervening introns (non-coding sequences), while the class II has two loci, α (with four exons) and β (six exons). We concentrated our efforts on the reported (Grimholt et al., 2000) highly polymorphic DNA sequences in the class Iα1, class Iα2, class IIα1 and class IIβ1 exons.

8.2.1 Four MHC primers

MHC class I. Oligonucleotide 22-mer primers were designed to amplify regions of the class Iα1 and Iα2 exons. Primers were designed based on complete messenger RNA (mRNA) *MHC* class I sequences obtained through Genbank (www.ncbi.nlm.nih.gov) (Table 8.1).

MHC class II. Published 18-mer primers were used to amplify the class IIα1 (Stet et al., 2002) and class IIβ1 (Langefors et al., 2000) exons (Table 8.1).

Table 8.1. PCR oligonucleotide primer sequences for four regions of the Atlantic salmon (*Salmo salar*) MHC genes. A, T, C, G represent the nucleotides adenine, thymine, cytosine and guanine respectively. Orientation; A and S corresponds to the antisense and sense primer sequences. Genbank accession numbers are indicated for mRNA sequences used to design primers.

MHC domain	Orientation	Sequence	Reference or accession number
Class Iα1	S	5'-TTCTACACTGCGTCTTCTAAAG-3'	AF104548-AF104553
	A	5'-CTGCTTTAGAATATTGACCTCG-3'	
Class Iα2	S	5'-GATGTGAGTGGGATGATGAGGC-3'	AF1044595-AF104414
	A	5'-GCTTCTTCAGCCACTCAATGCA-3'	
Class IIα1	S	5'-CCAGATGTAATATACTAAC-3'	Stet et al., 2002
	A	5'-CTATTTTCTCTTCTGGGTT-3'	
Class IIβ1	S	5'-TAATCTGACAAAACAATGA-3'	Langefors et al., 2000
	A	5'-CACCTGTCTTGTCCAGTATG-3'	

8.2.2 PCR amplification conditions

The PCR was performed in a total volume of 50µl, containing 5µl (2.5mM) of 25mM $MgCl_2$, 5µl 10X PCR Gold^{TM} buffer (PE Applied Biosystems, Foster City), 1.5µl (0.3mM) of each 10µM primer, 1µl (0.2mM) of each dNTP (Promega, Madison), 0.3µl (1.5U) of Amplitaq Gold (PE Applied Biosystems), 5µl (0.2ng μl^{-1}) of gDNA and 30.7µl of MilliQ water (Millipore Corp, Bedford). Amplification was performed in a GeneAmp 9700 thermalcycler (PE Applied Biosystems) using the following thermal profile: enzyme

activation for 6 min at 94ºC, followed by 35 cycles of denaturation for 1 min at 94ºC, annealing for 1 min at 50ºC for *MHC* class Iα1 and *MHC* class IIα1 and 55ºC for *MHC* class Iα2 and *MHC* class IIβ1, and primer extension for 1.30 min at 72ºC. Final primer extension was for 10 min at 72ºC. PCR product was stored at -20ºC.

To confirm amplification, 5µl of PCR product was combined with 2µl of 6X loading dye (Promega) and electrophoresed on a 2% TAE (Tris-Acetate-EDTA) agarose gel at a constant 100 V for 40 min in 1X TAE buffer. The gel was stained with 0.5µg ml^{-1} ethidium bromide for 30 min, visualised under UV transillumination and photographed with a digital camera (Nikon, Coolpic). Fragment size was determined using a 100 bp size standard (Promega).

8.2.3 MHC gene amplifications

Following optimisation, the standard PCR produced a single amplicon for all four PCR primer sets. The *MHC* class Iα1, class Iα2, class IIα1 and class IIβ1 primer sets amplified a 222bp, 216bp, 294bp and 315bp fragment, respectively (Fig 8.1). Direct DNA sequencing and BLAST analysis (www.ncbi.nlm.nih.gov) of these PCR products confirmed that the primer sets were specific for the four Atlantic salmon MHC domains.

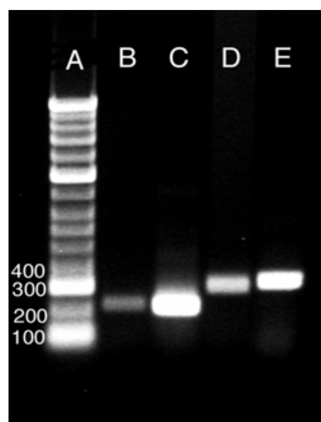


Fig 8.1. PCR product for four MHC domains. *MHC* class Iα1 (lane B), class Iα2 (lane C), class IIα1 (lane D) and class IIβ1 (lane E) domains on 2% agarose gel, stained in ethidium bromide and photographed under UV transillumination. Lane A, 100bp molecular marker (Promega).

8.2.4 Restriction fragment length polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) of the PCR fragments was utilised to detect nucleotide variation at the four MHC domains. A full theoretical explanation of this procedure is available from Beaumont and Hoare (2003).

To produce RFLP phenotypes, the PCR products were digested with specific restriction endonucleases (REs). The REs for each gene were selected based on restriction maps produced by analysing known Atlantic salmon MHC allele sequences (Genbank www.ncbi.nlm.nih.gov) for RE recognition sites using the freeware program WebCutter (www.firstmarket.com). Restriction endonucleases (New England Biolabs, Beverly) selected were: *Mbo*I for class Iα1, *Mbo*I and *Dde*I for class Iα2, and *Alu*I and *Rsa*I for class IIα1 and class IIβ1.

Digestion was performed in a 25µl reaction, containing 10µl PCR product, 2.5µl enzyme buffer (New England Biolabs), 0.2µl (1U) of each restriction endonuclease and MilliQ (Millipore Corp) water to a total volume of 25µl. PCR product was digested in a water bath maintained at 37ºC for 3 h. Following digestion, enzyme activity was inactivated by incubating at 65ºC for 20 min. Ten microliters of the digest was combined with 2µl of 6X loading dye (Promega) and subjected to electrophoresis on an 8% non-denaturing polyacrylamide gel in 1X TBE (Tris-Borate-EDTA) buffer for 3 h at 80 volts and 4 milliamps. Gels were then stained with ethidium bromide (0.5µl ml^{-1}) for approximately 20 min. Gels were visualised under UV transillumination and photographed with a digital camera (Nikon, Coolpic). A 25bp size standard (Promega) was included to allow the size of bands to be determined. Variant RFLP phenotypes were identified and frequencies quantified. This process was repeated for each of the four MHC domains.

A RFLP phenotype is made up from the two alleles that an individual carries (one from each parent). It was assumed that identical RFLP phenotypes produced with particular REs contained the same alleles. This is correct based on presence and absence of restriction sites, however nucleotide variation at non-restriction sites may exist such that some allelic variation is missed using the RFLP method compared to the direct DNA sequencing of all individuals (the latter being beyond the resources of this pilot study).

8.2.5 Allele determination by DNA sequencing

In order to determine which MHC allele/s are represented within each RFLP phenotype, the DNA sequence of both alleles for each phenotype was determined.

This was performed by ligating the selected PCR product into a pCR®2.1-TOPO® (Invitrogen, The Netherlands) cloning vector. Briefly a 6µl ligation reaction, consisting of 4µl of PCR product, 1µl of salt solution (Invitrogen) and 1µl of TOPO® vector was incubated on ice for 30 min. Following ligation, the TOPO® construct was transformed into competent TOP10F' *Escherichia coli* cells (Invitrogen) by heat shock for 30 seconds at 42ºC. Each transformation reaction was inoculated into 250µl of SOC media (Invitrogen) and incubated at 37ºC for 1 h. Following incubation, 100µl of each transformation was spread on prewarmed (37ºC) LB agar plates containing 50µg ml^{-1} ampicillin (Invitrogen) and 2mg ml^{-1} X-gal (Promega) and incubated overnight at 37ºC. Following incubation, 10 white colonies from each plate were selected and inoculated into 5ml of LB broth, containing 50µg ml^{-1} ampicillin (Invitrogen) and grown overnight at 37ºC. Plasmid DNA was purified from the clones using a modification of the QIAprep spin miniprep kit (QIAGEN, The Netherlands). Following purification, plasmid DNA was resuspended in 200µl MilliQ water (Millipore Corp) quantified using a DU®530 spectrophotometer (Beckman, Fullerton) and diluted to 200ng μl^{-1}.

To confirm the ligation of the PCR insert, plasmid DNA was subjected to restriction digest with *Eco*RI. Digestion was performed in a 20µl reaction containing 5µl of plasmid DNA, 2µl of enzyme buffer (New England Biolabs), 0.5µl of RNaseA (Invitrogen), 0.4µl of *Eco*RI (New England Biolabs) and 12.1µl MilliQ water (Millipore Corp). The reaction was incubated at 37ºC for 3 h. Following incubation, enzyme activity was inactivated by incubation at 65ºC for 20 min. A 10µl aliquot of each digest was then combined with 2µl of 6X loading dye (Promega) and electrophoresed on a 2% TAE agarose gel, in 1X TAE buffer for 40 min at 100V. Following staining in ethidium bromide (0.5µl ml^{-1}) for approximately 20 min the gel was visualised under UV transillumination. A 100bp size standard (Promega) was included on each gel to enable insert size determination.

Plasmids positive for the PCR insert were observed by the presence of one distinct band slightly larger than the original PCR fragment. Plasmids negative for the PCR insert, were discarded and colony selection repeated so that each RFLP phenotype was represented by 10 separate clones.

The 10 positive clones for each RFLP phenotype were used as templates in DNA sequencing reactions using the ABI Prism BigDye™ Teminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems). Sequencing reaction was performed in a total volume of 10µl, containing 4µl BigDye™ (PE Applied Biosystems), 1µl T7 promoter primer (Invitrogen) and 5µl of plasmid template. The sequencing reaction was performed in a GeneAmp 9700 thermalcycler, with the following cycling parameters: initial activation for 1min at 96ºC, followed by 25 cycles of denaturing for 10 sec at 96ºC, annealing for 5 sec at 50ºC, and primer extension for 4 min at 60ºC.

Prior to submitting the BigDye product for sequencing, the reaction was precipitated with 40µl of 75% isopropanol and 1µl of glycogen, incubated at room temperature for 15 min and centrifuged at 13,500 rpm for 20 min. The supernatant was aspirated and the pellet washed in 175µl of 75% isopropanol at 13,500 rpm for 10 min at room temperature. The supernatant was aspirated and the pellet dried in a vacuum centrifuge (Speedvac). The resultant DNA was denatured with formamide for 2 min at 95ºC and run on a 12% denaturing polyacrylamide sequencing gel on an ABI Prism 377 DNA sequencer (PE Applied Biosystems) for 10 h.

Completed sequence lanes were tracked and extracted using Sequencing Analysis™ version 1.0.1 (PE Applied Biosystems) and sequences analysed using Sequence Navigator™ version 1.0.1 (PE Applied Biosystems). Obtained sequences were subjected to a BLAST similarity search via the National Centre for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov). RE recognition sites with the allele sequences were identified using the freeware program WebCutter (www.firstmarket.com).

8.2.6 *MHC* class Iα1 RFLPs and alleles

Three unique *MHC* class Iα1 RFLP phenotypes were identified during this study, and designated *a*, *b*, and *c* (Fig 8.2). RFLP *a* consisted of two single bands, RFLP *b* one single and one doublet band, and RFLP *c* three single bands.

DNA sequencing was performed on each of the three RFLP phenotype variants, and revealed the presence of three previously undescribed *MHC* class Iα1 alleles (full nucleotide sequences in Appendix 3. Figure A.1). These alleles were designated *Iα1*1*, *Iα1*2* and *Iα1*3*. Allele *Iα1*2* contained a single *Mbo*I restriction site at 75bp (from 3' end) and thus produced two fragments, 75bp and 147bp in length. Allele *Iα1*1* also contained a single *Mbo*I restriction site however due to a five base polymorphic deletion (within the primer site) this restriction site was located at 70bp and produced two fragments, 70bp and 147bp in length. Finally allele *Iα1*3* did not contain a *Mbo*I restriction site and thus a single band at 222bp was maintained. *Mbo*I restriction sites and predicted fragment sizes appeared to correspond with observed RFLP banding patterns to within 5-10bp. Small discrepancies may be a result of inaccuracy within the 25bp molecular weight marker as described by the manufacturer (Promega).

Therefore, RFLP phenotype *a* represented a homozygote individual for allele *Iα1*2*, RFLP *b* represented a heterozygote individual for alleles *Iα1*1* and *Iα1*2*, and RFLP *c* represented a heterozygote individual for alleles *Iα1*2* and *Iα1*3*.

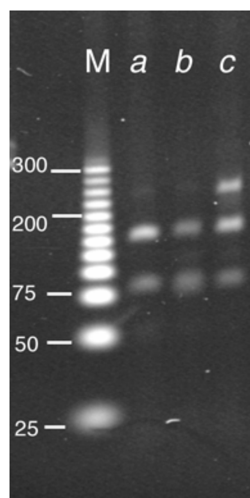


Fig 8.2. Three *MHC* class Iα1 RFLP phenotypes. Restriction digest with *Mbo*I on 8% non-denaturing polyacrylamide gel stained in ethidium bromide and photographed under UV transillumination. RFLP *a* bands are 75bp/147bp, RFLP *b* bands 70bp/75bp/147bp and RFLP *c* bands are 75bp/147bp/222bp (fragment sizes determined from DNA sequencing). M, represents a 25bp molecular weight marker (Promega).

8.2.7 *MHC* class Iα2 RFLP and alleles

RFLP analysis in this study revealed only a single phenotype within the *MHC* class Iα2 domain. The single RFLP phenotype identified, designated *d* (Fig 8.3) consisted of one double and two discrete single bands, plus several smaller bands.

DNA sequencing revealed the presence of two previously undescribed *MHC* class Iα2 alleles, designated *Iα2*1* and *Iα2*2* (Appendix 3. Figure A.2). *Mbo*I and *Dde*I restriction sites produced predicted fragments of 109 bp, 107bp, 100bp and 66bp (or 69bp depending on which enzyme cut first) in length, smaller fragments ranging from 26-21bp were also predicted. Specifically, the allele *Iα2*1* produced the two major bands at 100bp and 66bp (or 69 bp), and *Iα2*2* produced the major bands at 109bp and 107bp. These fragment sizes appeared to correspond to the RFLP banding patterns, with the exception of a smaller 40bp band, which was concluded to be either a primer dimer or extraneous PCR product. Sequencing was replicated three times on three separate but identical RFLPs and consistently revealed the presence of the same two *Iα2*1* and *Iα2*2* alleles.

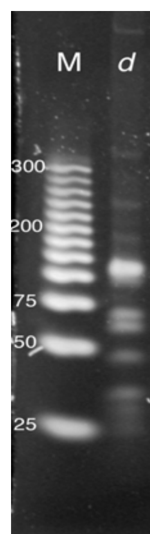


Fig 8.3. Single *MHC* class Iα2 RFLP phenotype. Restriction digest with *Mbo*I and *Dde*I on 8% non-denaturing polyacrylamide gel stained in ethidium bromide and photographed under UV transillumination. M represents a 25bp molecular weight marker (Promega)

8.2.8 *MHC* class IIα1 RFLPs and alleles

In our study four unique *MHC* class IIα1 RFLP phenotypes were identified and designated *e*, *f*, *g* and *h* following double digestion with *Alu*I and *Rsa*I (Fig 8.4). RFLP *e* appeared to consist of four single bands, RFLP *f* of three bands (two high molecular weight bands also evident in *f* were predicted to be extraneous PCR product), RFLP *g* of five bands and RFLP *h* of four bands.

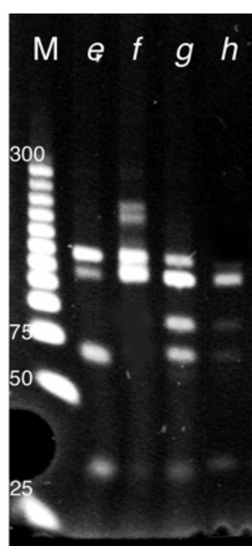


Fig 8.4. Four *MHC* class IIα1 RFLP phenotypes. Restriction digest with *Alu*I and *Rsa*I on 8% non-denaturing polyacrylamide gel stained in ethidium bromide and photographed under UV transillumination. M, represents a 25bp molecular weight marker (Promega).

DNA sequencing of each of the four RFLP phenotypes revealed the presence of one previously described allele (Genbank accession number: 779040; Grimholt, unpublished) and two undescribed *MHC* class IIα1 alleles. The previously described allele was

renamed *IIα1*2* and the two novel alleles *IIα1*1* and *IIα1*3* (Appendix 3. Figure A.3). Allele *IIα1*1* contained a single cut site for each restriction enzyme resulting in three bands at 135bp, 132bp and 27bp. Alleles *IIα1*2* and *IIα1*3* both had a single *Rsa*I cut site and two *Alu*I cut sites, although in different regions. The resulting fragments were 135bp, 66bp (double) and 27bp for *IIα1*2* and 132bp, 72bp, 63bp and 27bp for *IIα1*3*.

Table 8.2. Predicted fragment sizes for the four *MHC* class IIα1 alleles following the *Alu*I and *Rsa*I double digestion.

RFLP phenotype	Alleles	Predicted composite RFLP band sizes
e	*IIα1*1 / IIα1*2*	135/135/132/66/66/27/27
f	*IIα1*1 / IIα1*1*	135/132/27
g	*IIα1*1 / IIα1*3*	135/132/132/72/63/27/27
h	*IIα1*3 / IIα1*3*	132/72/63/27

There appeared to be some discrepancy between predicted and observed banding patterns (Table 8.2 and Figure 8.4). For instance, sequencing shows a difference of only three base pairs between a 132 and a 135bp fragment. However visual examination of the gels suggested this difference to be more likely in the magnitude of 20bp. Considering that this study utilised non-denaturing gels it is likely folding or secondary structures may have altered the appearance of particular fragments. For instance fragments of a similar size (in this instance 132bp and 135bp) but with different nucleotide sequences can form different secondary structures, which may ultimately influence the speed at which they migrate under electrophoresis. It was assumed secondary structures of a particular fragment are produced at the same proportion within all individuals and therefore are unlikely to affect RFLP frequencies. The 72bp and 63bp bands in *g* and *h* also appear to be of high weight on the gel images, but the size was confirmed by sequencing. The two additional high molecular weight bands observed in phenotype *f* were not predicted by sequencing and were concluded to be extraneous PCR product.

8.2.9 *MHC* class IIβ1 RFLPs and alleles

Extensive polymorphism was observed at the *MHC* class IIβ1 domain and so two individual restriction enzyme digests (*Alu*I and *Rsa*I) rather than a single double digest were performed in an attempt to separate RFLP phenotypes more effectively. The single restriction enzyme phenotypes were then combined to form a composite multiple restriction enzyme phenotype.

From all individuals examined in this study the enzyme *Rsa*I resolved three RFLP phenotypes (designated *Ra*, *Rb* and *Rc*; Figure 8.5A) and *Alu*I four phenotypes (designated *Aa*, *Ab*, *Ac* and *Ad*; Figure 8.5B).

The phenotype *Rc* has a single band (at 315bp confirmed by sequencing) and represents a homozygous individual with both alleles uncut by *Rsa*I. Phenotype *Rb* with two visual bands (214bp and 101bp) represents a homozygous individual with both alleles containing the same single *Rsa*I cut site (Appendix 3 Figure A.4), and phenotype *Ra* has three bands, an uncut fragment (315bp) and two fragments (214bp and 101bp) from an

allele with a single cut site; therefore representing a heterozygote individual. Additional doublet bands were also present in *Ra* and *Rc* but due to being larger than the initial amplicon, they were assumed to be extraneous PCR product.

The phenotypes *Aa* and *Ab* each consisted of four bands, *Ac* of three bands and *Ad* of two single bands (Fig 8.5B). The predicted RFLP banding patterns from the DNA sequences (Appendix 3 Figure A.4) were consistent with those observed on the gels, with the exception of phenotype *Aa*. This phenotype represented a heterozygote individual with multiple cut sites and predicted fragment sizes of 230bp, 188bp, 73bp, 64bp 42bp, 12bp and 9bp. The smaller bands at 12bp and 9bp due to their small size were not visible in this or the other phenotypes. The 64bp band for this phenotype however was also not observed on a gel. While this band could not be visualized, perhaps due to low concentration of DNA, it was assumed to be present. Phenotype *Ab* was also a heterozygote with predicted fragments of 230bp, 188bp, 64bp, 42bp, 12bp and 9bp, all of which were seen, with the exception of the two smaller fragments. The three observed fragments for the *Ac* phenotype were 188bp, 64bp and 42bp, with 12bp and 9bp fragments not visualised. This phenotype was for a homozygote individual, as was phenotype *Ad* that had observed bands at 230bp and 64bp, with unseen 12bp and 9bp fragments. The additional observed doublet band (larger then the initial amplicon) present in *Ab*, a weaker doublet band in *Aa* and one single weak band in *Ad* were not predicted by DNA sequencing and thus were assumed to be excess non-template DNA.

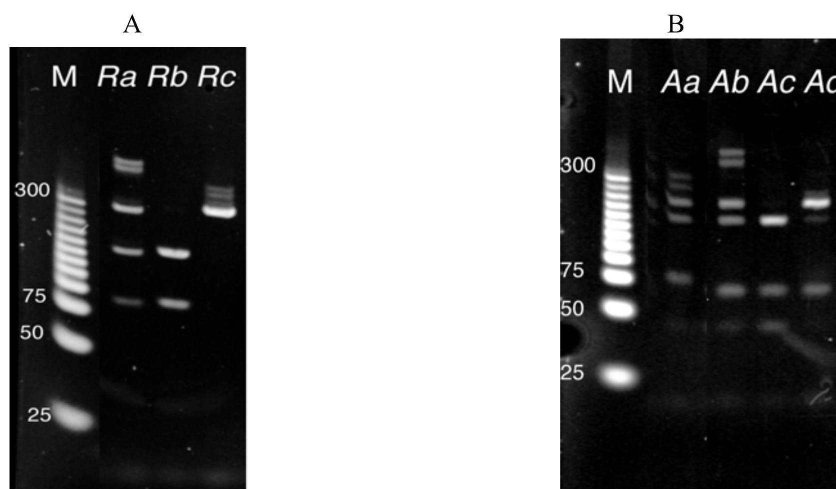


Fig 8.5. *MHC* class IIβ1 composite RFLP phenotypes. A; shows three *MHC* class IIβ1 RFLP phenotypes following restriction digest with *Rsa*I and B; shows four *MHC* class IIβ1 RFLP phenotypes following restriction digest with *Alu*I. On 8% non-denaturing polyacrylamide gel stained in ethidium bromide and photographed under UV transillumination.

Combining the *Alu*I and *Rsa*I RFLPs produced seven composite RFLP patterns and were designated *i* to *o*. DNA sequencing of each composite phenotype revealed the presence of six unique *MHC* class IIβ1 alleles (Appendix 3, Figure A.4). Four of these alleles, designated *IIβ1*1* to *IIβ1*4,* had been previously described (*Sm7u*, *Sm4k*, *Sm2c* and *Sm6v*, with Genbank accession numbers AF373703, AF373705, AF373707 and AF373692, respectively) by Landry and Bernatchez (2001). The two new alleles were designated *IIβ1*5* and *IIβ1*6*.

Restriction sites and predicted fragment sizes were determined for each allele (Table 8.3). Due to several of the alleles containing identical restriction sites (but unique DNA sequences between, e.g. alleles *IIβ1*1* and *IIβ1*3*), the assumption that all identical RFLPs will contain identical alleles cannot be made with any validity for this locus and is taken into account in the following analyses.

Table 8.3. *Alu*I and *Rsa*I single digestion predicted fragment sizes for the six *MHC* class IIβ1 alleles and allele combination for the seven composite RFLPs.

	Fragment sizes	
	***Alu* I**	***Rsa* I**
Allele		
IIβ1*1	230/64/12/9	315
IIβ1*2	188/64/42/12/9	315
IIβ1*3	230/64/12/9	315
IIβ1*4	230/73/12	315
IIβ1*5	188/64/42/12/9	214/101
IIβ1*6	188/64/42/12/9	315
RFLP	**Composite**	**Alleles**
i	***Ab / Rc***	***IIβ1*3 / IIβ1*2***
j	***Aa / Rc***	***IIβ1*4 / IIβ1*2***
k	***Ac / Rc***	***IIβ1*6 / IIβ1*6***
l	***Ab / Ra***	***IIβ1*1 / IIβ1*5***
m	***Ac / Rb***	***IIβ1*5 / IIβ1*5***
n	***Ad / Rc***	***IIβ1*1 / IIβ1*1***
o	***Ac / Ra***	***IIβ1*5 / IIβ1*2***

8.3 Population study

8.3.1 Samples and statistical analyses

To examine the extent of MHC polymorphism within the Tasmanian Atlantic salmon population, RFLP-phenotype and allele frequencies were compared between two Tasmanian samples and one Canadian sample from the ancestral progenitor population. The Canadian sample was a random 30 fish collected in November 1992 from a hatchery spawning of wild River Philip fish. The same individuals were examined for allozyme and microsatellite variation in previous projects (FRDC 1992/252, 1995/80 and 1996/347; Ward et al., 1994; Reilly et al. 1999; Elliott and Reilly 1998 and 2003). While we have shown in Chapter 7 that this sample is far from ideal as a baseline sample for the progenitor Canadian population, the resources and time available for this study precluded examination of MHC polymorphism in the archival 1971/72 scale samples. The two Tasmanian samples consisted of a random 30 individuals each from the SALTAS 1997 and 1999 year-classes that had been previously examined for microsatellite variation analysis (Chapter 7). Genomic DNA had been extracted as previously described.

Allelic frequency differences between samples were calculated using the F-statistic, F_{st}, in ARLEQUIN version 2.00. Deviations between observed (H_o) and expected (Hardy-Weinberg H_e) heterozygosity were tested using GENEPOP version 3.2. Significance levels were determined after 500 batches of 5000 iterations each of a Markov chain. When the presence of null alleles (non-amplification of a PCR product due possibly to mutation in the PCR primer region) was expected the manually calculated human ABO blood group model was applied (CaValli-Storza and Bodmer, 1971). *P* values less than 0.05 were considered significant and sequential Bonferroni correction for multiple tests was then applied to all significance levels.

8.3.2 Results

PCR amplification was achieved in all 90 individuals for *MHC* class Iα2 and class IIβ1. *MHC* class IIα1 amplification was successful for all individuals within the two Tasmanian samples but only for 27 individuals within the Canadian sample. The three missing individuals were excluded from the analysis of this locus, with non-amplification assumed to be due to inhibitory compounds in the PCR reaction. *MHC* class Iα1 amplification was achieved in only 24, 22, and 19 individuals in the Canadian, and the Tasmanian 1997 and 1999 samples, respectively. This non-amplification was assumed to be due to null alleles.

No polymorphism within the *MHC* class Iα2 domain was observed either within or between the samples. All individuals were represented by the single *d* RFLP phenotype and so were heterozygous for alleles *Iα2*1* and *Iα2*2*.

8.3.2.1 *MHC* class Iα1

The RFLP phenotype *c*, representing heterozygous individuals for alleles *Iα1*2* and *Iα1*3,* was the most frequent within the Canadian sample. This phenotype was not observed in any of the Tasmanian individuals in which phenotype *a*, representing homozygous individuals for allele *Iα1*2* was the most frequent (Table 8.4). The RFLP phenotype *b* (heterozygous individuals for alleles *Iα1*1* and *Iα1*2*) was at a low frequency in all samples, although slightly more prominent within the Tasmanian 1999 sample.

Allele frequencies for each sample were compiled according to the RFLP frequencies, and significant differentiation was observed between both Tasmanian samples and the Canadian sample (1997 and Canadian F_{st}=0.159, P<0.001 and 1999 and Canadian F_{st}=0.125, P=0.002). No significant allele frequency differentiation was observed between the two Tasmanian samples (F_{st}=0.016, P=0.316). The Canadian sample had two equally dominant alleles (*Iα1*2* and *Iα1*3*) while the Tasmanian samples had only one dominant allele (*Iα1*2*).

Observed heterozygosity was highest in the Canadian sample where all amplified individuals were heterozygotes (Table 8.4). Levels in the two Tasmanian samples were both much lower, however, heterozygotes with a null allele would not have been detected. Observed heterozygosity (H_o) in the 1997 Tasmanian sample was in agreement with predicted Hardy-Weinberg equilibrium but significantly greater than expected in the Canadian (χ^2=484.9, df=1, P<0.001) and 1999 Tasmanian (χ^2=14.14, df=1, P<0.001) samples.

Table 8.4. *MHC* class Iα1 phenotype and allele frequencies for the Canadian, 1997 Tasmanian and 1999 Tasmanian samples. Assumed null represents individuals that did not PCR amplify and so were assumed to contain null alleles. * indicates significant difference between observed and Hardy-Weinberg expected allelic heterozygosity.

	Canadian	Tas 1997	Tas 1999
Phenotypes			
a	0.00	0.63	0.43
b	0.03	0.10	0.20
c	0.77	0.00	0.00
Assumed Null	0.20	0.27	0.37
Alleles			
Iα1*1	0.02	0.05	0.10
Iα1*2	0.40	0.68	0.53
Iα1*3	0.38	0.00	0.00
Assumed Null	0.20	0.27	0.37
Observed heterozygosity	0.800*	0.100	0.200*

8.3.2.2 *MHC* class IIα1

The RFLP phenotype *e* (heterozygous individuals for alleles *IIα1*1* and *IIα1*2)* appeared most frequently in both the Canadian and 1997 Tasmanian samples, with phenotypes *f* (homozygous individuals for alleles *IIα1*1)* and *g* (heterozygous individuals for alleles *IIα1*1* and *IIα1*3)* both equally less frequent (Table 8.5). Conversely the 1999 Tasmanian sample had a more even frequency of all three phenotypes. The allele *IIα1*1* appeared most frequent in all populations.

Significant differentiation in allele frequencies was observed between the Tasmanian 1999 sample and the Canadian sample (F_{st}=0.050, P=0.006). No significant difference was observed between the 1999 and 1997 Tasmanian samples (F_{st}=0.020, P=0.058) or between the 1997 Tasmanian and the Canadian sample (F_{st}=0.011, P=0.541). Observed heterozygosity was highest in the Canadian sample and lowest in the 1999 Tasmanian sample (Table 8.5), and was significantly greater than expected in both the 1997 Tasmanian (P= 0.003) and Canadian samples (P=0.007).

Table 8.5. *MHC* class IIα1 phenotype and allele frequencies for the Canadian, 1997 Tasmanian and 1999 Tasmanian samples. * indicates significant difference between observed and Hardy-Weinberg expected allelic heterozygosity.

	Canadian	Tas 1997	Tas 1999
Phenotypes			
e	0.63	0.56	0.26
f	0.13	0.23	0.36
g	0.13	0.20	0.26
h	0.00	0.00	0.10
Non Amp	0.10	0.00	0.00
Alleles			
IIα1*1	0.57	0.60	0.58
IIα1*2	0.35	0.28	0.13
IIα1*3	0.07	0.12	0.28
Oberved heterozygosity	0.852*	0.767*	0.533

8.3.2.3 *MHC* class IIβ1

Of the seven identified composite RFLP phenotypes, four were present within the Canadian sample and all seven were present in both the Tasmanian samples (Table 8.6). Furthermore within the Canadian sample one RFLP (*i* representing heterozygous individuals for alleles *IIβ1*2* and *IIβ1*3*) was considerably more dominant, while a more even distribution was observed within the Tasmanian samples.

As some alleles contained identical restriction sites but had unique DNA sequences we cannot assume that all identical RFLPs contain identical alleles for this locus. Alleles with identical restriction sites were therefore pooled to produce four composite alleles (i.e different alleles by DNA sequencing that have the same restriction sites). These were designated A (consisting of *IIβ1*2* and *IIβ1*6*), B (consisting of *IIβ1**5), C (consisting of *IIβ1*1* and *IIβ1*3*) and D (*IIβ1*4*). While not optimal, under the circumstances of this pilot study this approach does allow a more valid comparison between samples.

The Canadian sample was dominated by the composite alleles A and C, and although both alleles were at a high frequency in the Tasmanian samples, composite allele B was also common despite not being observed in the Canadian sample (Table 8.6); allele D was at a low frequency in all samples. The observed difference between the Canadian and Tasmanian samples was however not significant at the sample sizes used in this pilot study (F_{st}=0.028, P=0.073 and F_{st}=0.033, P=0.071 for the 1997 and 1999 samples respectively), nor was the difference between the two Tasmanian samples (F_{st}=0.015, P=0.98).

MHC class IIβ1 heterozygosity was greatest in the Canadian sample and lowest in the 1999 Tasmanian sample (Table 8.6). Indeed, observed *MHC* class IIβ1 heterozygosity was significantly lower than expected for the 1999 Tasmanian sample (P<0.001) but in agreement with Hardy-Weinberg equilibrium for the other two samples.

Table 8.6. *MHC* class IIβ1 RFLP phenotype, allele and composite allele frequencies for the Canadian, 1997 Tasmanian and 1999 Tasmanian samples. * indicates significant difference between observed and Hardy-Weinberg expected allelic heterozygosity. Letters in parenthesis after the alleles refer to composite alleles.

	Canadian	Tas 1997	Tas 1999
Phenotypes			
i	0.53	0.20	0.13
j	0.10	0.13	0.13
k	0.20	0.27	0.23
l	0.00	0.13	0.03
m	0.00	0.10	0.13
n	0.17	0.13	0.23
o	0.00	0.03	0.10
Alleles			
***IIβ1*1* (C)**	0.17	0.20	0.25
***IIβ1*2* (A)**	0.32	0.18	0.18
***IIβ1*3* (C)**	0.27	0.10	0.07
***IIβ1*4* (D)**	0.05	0.07	0.07
***IIβ1*5* (B)**	0.00	0.18	0.20
***IIβ1*6* (A)**	0.20	0.27	0.23
Composite alleles			
A	0.51	0.45	0.41
B	0.00	0.18	0.20
C	0.43	0.30	0.31
D	0.05	0.06	0.06
Observed heterozygosity	0.693	0.500	0.400*

8.3.3 Discussion

This pilot study has shown differences exist at the MHC genes between samples from the Tasmanian population and one from its ancestral Canadian (River Philip sampled in 1992) population. The differences are expressed as change in allele frequencies and decrease in heterozygosity, but unlike the microsatellite studies (and more like the allozyme study) no major loss of alleles was detected. Caution is advised in interpreting these differences, as we have demonstrated (Chapter 7) that the particular Canadian sample used in this study is not an ideal representation of the River Philip population.

A further caution with these results is that it must be remembered that we have used RFLP technology which does not detect all nucleotide diversity present at these genes. It is possible that additional MHC alleles may be present (same RFLP cut site but different nucleotide sequence), and a greater change between the ancestral and derived populations may be detected. To confirm or deny this, DNA sequencing would need to be performed on every individual. The sample sizes used in this study were small; however the fact that

the same alleles were detected in all three samples (and the forty fish in the challenge test Section 8.4) would suggest that this was possibly not a limiting factor in this study.

The number of *MHC* class IIβ1 alleles we observed is similar to the number of alleles (7 to 16) found (through RFLP) in four Atlantic salmon hatchery populations in Sweden (Langefors et al., 1998). In contrast we observed considerably less alleles in the *MHC* class IIα1 and class Iα1 domains than previous studies. For instance from four populations in Norway, Stet et al. (2002) identified 8 *MHC* class IIα1 alleles among 42 sires and 42 dams; and Grimholt et al. (2002) identified 14 *MHC* class Iα1 alleles among 41 sires and 41 dams. The lack of polymorphism observed within the *MHC* class Iα2 locus is in contrast to a recent study that identified extensive polymorphism within this (exon 3) region of the *MHC* class I locus (Grimholt et al., 2002). The observed disparity between studies may be due to differences between western and eastern Atlantic stocks of Atlantic salmon, an inherent factor of the River Philip population, or to the different techniques applied. We used RFLP of genomic DNA, which will only detect a polymorphism if it resides within an enzyme restriction site, whereas Grimholt et al. (2002) solely utilised cDNA sequencing which is known to detect single nucleotide polymorphisms regardless of their allelic position. This technique greatly improves the efficacy of detection and quantification of DNA polymorphisms, but was beyond the resources of this study, in addition, because of their archived nature, the population samples were unsuitable for RNA extraction.

It has been observed in mammalian studies that heterozygosity at MHC loci often does not conform to Hardy-Weinberg equilibrium, which assumes that mating is random and no selection or mutations occurs. Genes of the MHC however represent a chaotic version of this model, because both selection (Jeffery and Bangham, 2000) and mutation (Miller and Withler, 1996; Langefors et al., 2001b) can occur with respect to these important immunological genes. The theory of over-dominant selection has been presumed to act on genes of the MHC (Jeffery and Bangham, 2000), with heterozygous advantage the most common form (Bernatchez and Landry, 2003). The hypothesis of heterozygous advantage proposes that such individuals are able to present a greater variety of peptide ligands than homozygotes, ultimately resulting in a more productive immune response to a diverse array of pathogens (Carrington et al., 1999). Very few studies have confirmed heterozygous advantage in salmonids, although Arkush et al. (2002) reported that chinook salmon (*Oncorhynchus tshawytscha*) heterozygous at the *MHC* class IIβ locus were significantly more resistant to IHN than homozygotes. The generally higher heterozygosity observed in our Canadian fish may be a consequence of the broodstock for this sample being wild returning individuals likely to have survived in the presence of a diverse array of both marine and freshwater pathogens and therefore heterozygotes may have had a select advantage. The Tasmanian fish on the other hand not only are progeny of hatchery reared broodstock but also their ancestors survived a population bottleneck during introduction to Australia and spent six generations exposed only to freshwater pathogens in New South Wales.

Further examination of variation at the MHC and other important genes in the Tasmanian population is warranted to determine the significance of the observed lower polymorphism (number alleles and heterozygosity) compared to other studies.

8.4 MHC variation and AGD resistance/susceptibility

8.4.1 Material and Methods

8.4.1.1 Experimental AGD challenge trial

Naïve Atlantic salmon were challenged with AGD in a cohabitation challenge model at the University of Tasmania. The model consisted of a 4000L recirculation system equipped with a swirl separator, biological filter, heat exchange and aeration. Twenty five Atlantic salmon were randomly selected as the 'infective cohabitants' from 500 SALTAS 2002 pre-smolts, and acclimated to 35ppt over 7 days in the 4000L recirculation system. To achieve acclimation, partial replacements with 35ppt seawater were made daily, increasing the salinity by approximately 5ppt each day. The 25 cohabitants were bath infected with 3000 cells L^{-1} *N. pemaquidensis* isolated from the gills of clinical AGD infected Atlantic salmon following the protocol by Zilberg et al. (2001). Following 7 days, the 25 individuals were anaesthetised with clove oil (0.02%) and AGD confirmed by visual examination of gills for gross lesions (Adams and Nowak, 2001). The fish were then tattooed by injecting 10µl of alcian blue (Sigma-Aldrich, Castle Hill) under the skin (approximately 1cm in front of the pelvic fin) using a 1ml syringe and 25-gauge needle, and placed in 200L of highly oxygenated 5µm filtered seawater for approximately 30 min to allow recovery from the anaesthetic. The 25 tattooed individuals were then transferred back to the recirculation system and allowed to further recover for 48h.

For the challenge test, 107 Atlantic salmon (mean weight of 69.6 ± 17.6g) were randomly selected from the remaining SALTAS pre-smolts. Although the exact numbers of parents are unknown, assuming current spawning protocols employed at SALTAS, it is likely the original 500 SALTAS pre-smolts were the progeny of in excess of 75 males and 200 females (H. King, personal communication 2004).

The 107 individuals were acclimated to 35ppt in a separate 4000L recirculation system as described above for 7 days prior the challenge. Following acclimation the test fish were transferred into the recirculation system containing the 25 AGD 'infective cohabitants'. The challenge trial was performed for 37 days with the temperature and salinity maintained at 16.2 ± 0.1ºC and 35.3 ± 0.8ppt, respectively. Every 7 days, 1000L of fresh filtered seawater (5µm, 35ppt) was replenished to maintain water quality. Fish were fed to satiation daily, with a commercial 5mm salmon diet (Skretting, Hobart).

Moribund individuals were collected daily and visually examined for gross gill lesions. Tissue samples (200mg) from each individual were aseptically taken from the left dorsal skeletal muscle, and immediately stored at -80ºC. AGD resistance for each individual was measured using the variable "time to morbidity". Following termination of the challenge trial, surviving individuals were euthanased, gill arches dissected for histological examination and tissue samples collected as above. All tissue samples were transported on dry ice to the CSIRO Marine Research Laboratories, and stored at -20ºC prior to gDNA extraction using standard protocols described previously.

8.4.1.2 Gill histology of surviving individuals

Following dissection, gills were fixed for 6 h in Davidson's fixative and preserved in 70% ethanol. Histology was performed on the second left gill arch for all surviving individuals. The gill arch was excised, placed in a histology cassette and processed using an automatic tissue processor (Tissue-tek II). Processed gills were embedded in paraffin

wax and one 5µm section taken using a microtome (Microm, Heidelberg). Sections were attached to a slide, stained with haematoxylin and eosin and viewed under 400X magnification (Olympus, Germany). AGD was confirmed by observing the presence of *Paramoaeba* spp. (amoeba with parasome/s) together with characteristic histological changes in gill tissue as described by Adams and Nowak (2003). Level of infection was determined through quantifying the proportion (as a percentage) of gill filaments (primary lamella) displaying hyperplastic lesions.

8.4.1.3 Analyses

For the comparative MHC analysis the 20 surviving individuals ('resistant') and the first 20 moribund individuals ('susceptible') were examined. RFLP phenotype frequencies were compared between the two groups of fish using a standard Monte-Carlo chi-square analysis in the program CHIRXC with 10,000 steps of a Markov chain. Significance levels were subjected to standard Bonferroni correction for multiple tests. Odds ratios were calculated for significant variations according to Quinn and Keough (2002).

8.4.2 Results

8.4.2.1 Challenge trial

The first moribund individual was observed 4 days post exposure (DPE) and subsequent moribund individuals were recorded daily until 32 DPE (Fig 8.6). The daily morbidity rate appeared to fit a normal distribution with the most moribund individuals recorded on days 16 (7 individuals) and 21 (7 individuals) post exposure. No moribund individuals where observed after day 32 and the challenge was terminated on day 36 post exposure. At termination the fraction surviving was 0.19 (20 individuals).

Gross gill lesions typical of an AGD infection were observed in 100% of the moribund individuals and it was assumed AGD was the cause of morbidity for all individuals.

Histological gill examination of the 20 individuals from the resistant (surviving) group identified AGD associated epithelial hyperplasia and interlamellar vesicle formation (crypts) in all fish. However the presence of paramoeba and subsequent confirmation of clinical AGD was only possible in 18 of the 20 individuals. The two individuals negative for AGD were positive for hyperplastic lesions but the presence of paramoeba was not confirmed. Considerable variation in the level of infection (quantified as % gill filaments displaying hyperplastic lesions) was observed in the surviving individuals. The percentage of filaments infected ranged from 14 to 98% (Fig 8.7).

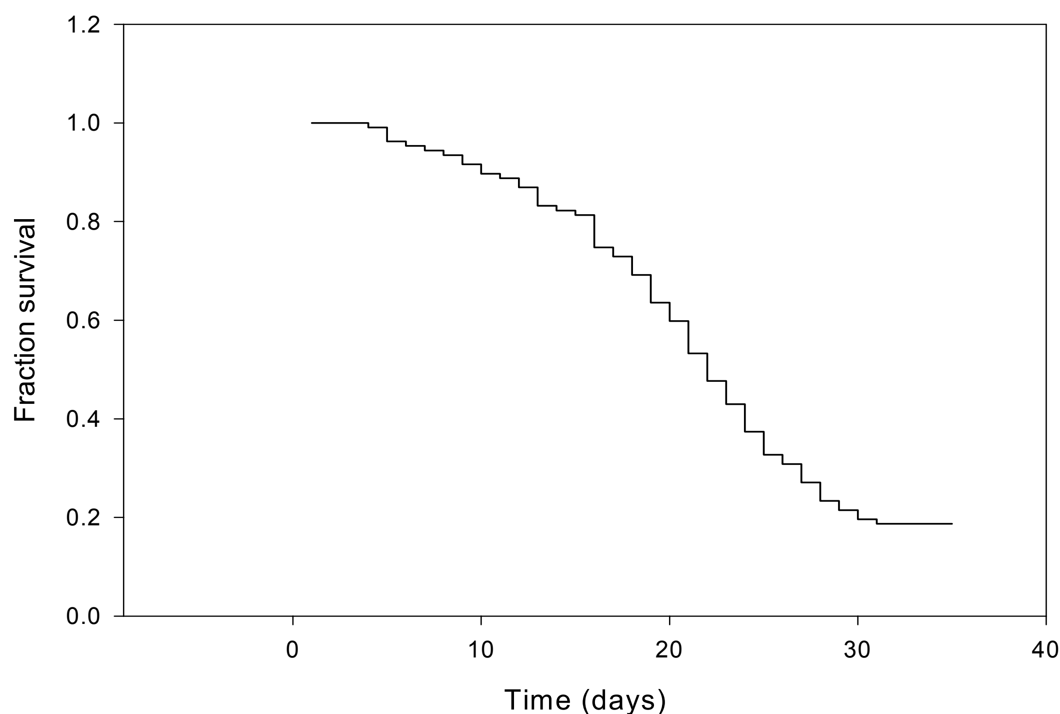


Fig 8.6. Survival in AGD challenge. Kaplan-Meier plot showing the fraction of Atlantic salmon survivors in the AGD challenge experiment from day of exposure (cohabitation) until study termination.

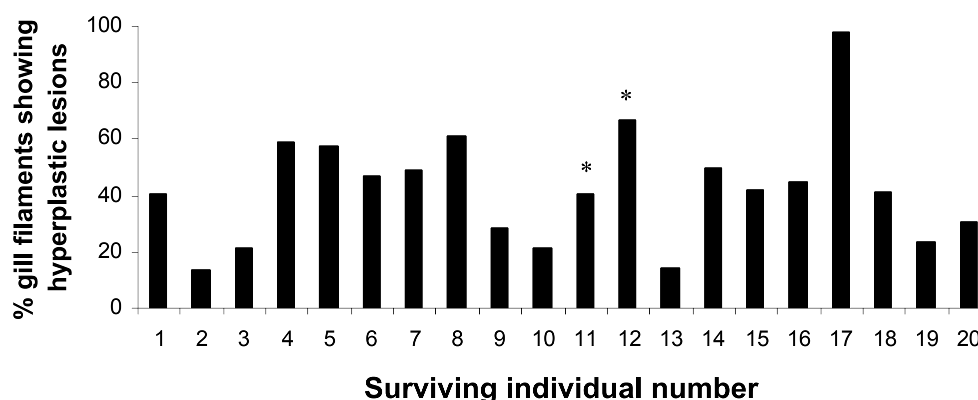


Fig 8.7. Lesion variation on gill filaments. Percentage of gill filaments (on second left gill arch) showing hyperplastic lesions for individuals surviving the AGD challenge trial,* represents individuals where the presence of paramoeba was not confirmed in histological sections.

8.4.2.2 RFLP analysis of AGD susceptible and resistant groups

To test for an association between MHC polymorphism and AGD resistance, RFLP frequencies were compared between the resistant (surviving) and susceptible (first moribund fish) groups. This approach offers a more valid test for the effects of MHC polymorphism, than comparing allele frequencies that would require (without DNA sequencing all fish) the assumption that identical RFLPs contain the same alleles.

MHC class Iα1
Six individuals from both the resistant and susceptible groups did not amplify for the *MHC* class *Iα1* domain, and it was assumed this was due to the presence of null alleles. As the frequency was similar in both groups it is assumed the null allele frequency did not vary between the susceptible and resistant groups.

From the successfully amplified individuals the two previously identified RFLP phenotypes (*a* and *b*) in the Tasmanian fish were observed, with phenotype *a* as expected the more common. The phenotype frequencies (Fig. 8.8) did not significantly differ between the resistant or susceptible groups (χ^2=0.243, df=1, *P*=1.00).

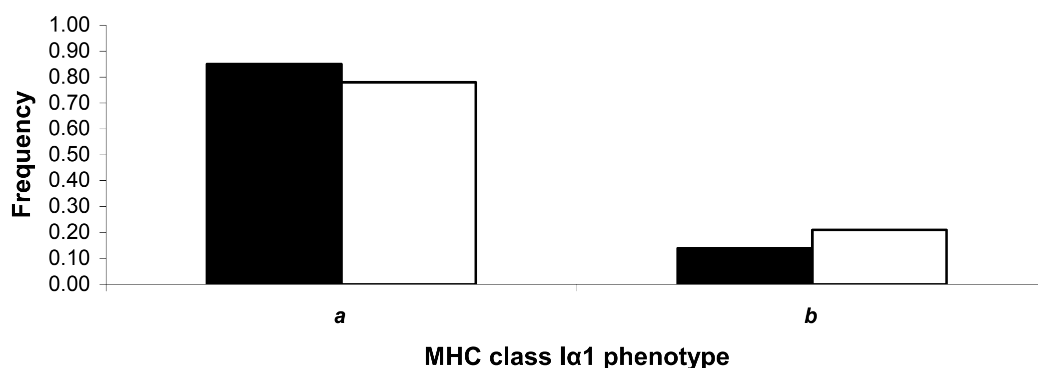


Fig 8.8. Frequency distribution of *MHC* class Iα1 phenotypes in AGD susceptible (*black bars*) and AGD resistant (*white bars*) groups.

MHC class Iα2
As discussed above only the single phenotype was observed at this domain in all the individuals examined.

MHC class IIα1
At this domain the three more common RFLP phenotypes (*e, f,* and *g*) identified in the population analysis were observed within the resistant and susceptible groups (Fig 8.9). The minor difference observed at phenotype *f* and *g* was not statistically different (χ^2=0.619, df=2, *P*=0.798).

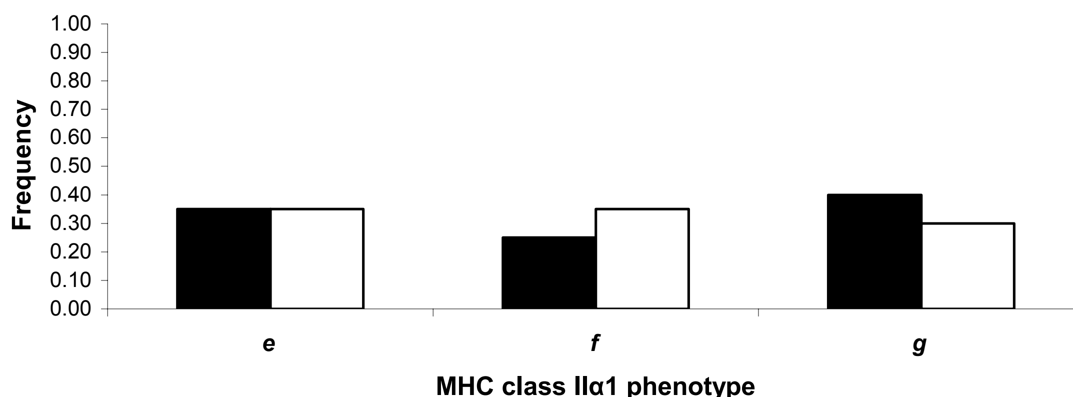


Fig 8.9. Frequency distribution of *MHC* class IIα1 phenotypes in AGD susceptible (*black bars*) and AGD resistant (*white bars*) groups.

MHC class IIβ1
Three individuals within the resistant group failed to PCR amplify at this domain and were excluded from the comparative analysis. It was assumed a PCR inhibitor rather than the presence of null alleles was responsible.

The six more common phenotypes identified in the population analysis were observed in the AGD resistant group of fish while only four phenotypes were observed in the susceptible group of fish (Fig. 8.10). Phenotype *n* was observed at a significantly higher frequency in individuals from the susceptible group than in individuals from the resistant group and the phenotype *j* was significantly more frequent in individuals from the resistant group (χ^2=13.91, df=5, *P*=0.003). Indeed the odds of being susceptible were 2.6 times greater for individuals with the *n* phenotype. Frequencies of the remaining phenotypes did not differ significantly between the two groups.

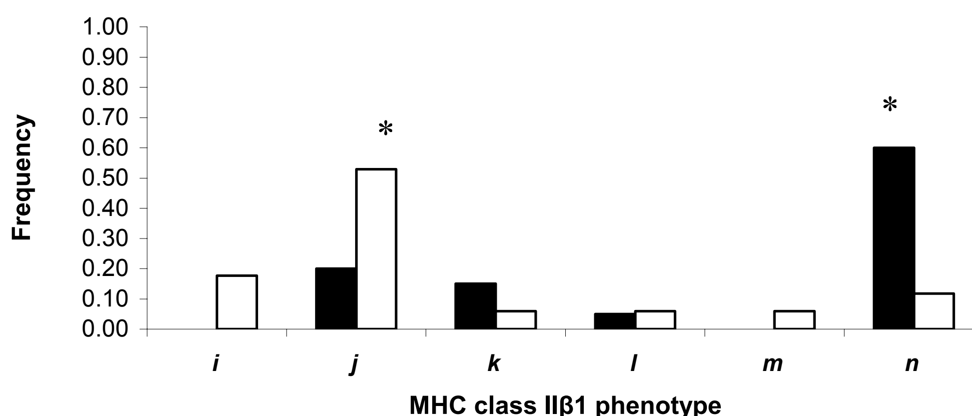


Fig 8.10. Frequency distribution of *MHC* class IIβ1 phenotypes in AGD susceptible (*black bars*) and AGD resistant (*white bars*) groups of Atlantic salmon. * indicate frequencies significantly greater than expected.

Phenotype *n* that is more common in susceptible fish represents a homozygous individual for allele *IIβ1*1*, while phenotype *j* represents a heterozygous individual for alleles *IIβ1*2* and *IIβ1*4* (Table 8.2, and Appendix 3 Fig A.4).

Infection variation in resistant fish
Variation in infection intensity (% gill filaments showing hyperplastic lesions, Fig. 8.7) identified through histology, did not appear to be linked with any particular *MHC* class Iα1, class IIα1 or class IIβ1 RFLP phenotypes.

8.4.3 Discussion

The AGD challenge results support the hypothesis of inherent variation in AGD resistance, quantified in this case as both level of infection and survival. The fact that all surviving individuals showed characteristic AGD hyperplastic lesions suggests these individuals had superior resistance to the effects of AGD but not resistance to the paramoeba itself. This finding is consistent with recent studies by Gross et al. (2004) who reported no significant difference in infection intensity (quantified as % gill filaments showing hyperplastic lesions) between AGD susceptible (first mortalities) and surviving individuals. The variation in level of infection indicates resistance to AGD may not be

absolute, yet some level of resistance is evident and this may increase survival time of infected individuals.

The observed association between polymorphism at the *MHC* class IIβ1 domain and AGD resistance/susceptibility is in agreement with previous studies showing associations between this gene and disease resistance. Langefors et al. (2001a) observed that specific *MHC* class IIβ alleles were associated with both resistance and susceptibility to furunculosis in Atlantic salmon. The authors concluded a single nonsynonymous (coding) nucleotide polymorphism with the resistant allele was conferring resistance. In a similar study of the *MHC* class IIβ locus, Palti et al. (2001) reported one allele differing by three polymorphic positions was significantly more frequent in individuals (rainbow x cutthroat trout backcrosses) resistant to infectious hematopoietic necrosis than in susceptible individuals. Accordingly it can be speculated that the minor nucleotide polymorphisms differentiating the *MHC* class *IIβ1* RFLP phenotypes *n* and *j* are altering the amino acid sequence of the peptide-binding cleft and therefore may be functionally conferring AGD resistance/susceptibility. Alternatively, it may be a consequence of homozygous individuals for a particular allele (RFLP *n* was homozygous for allele *IIβ*1*) being more susceptible.

The *MHC* class I molecules are responsible for the presentation of endogenously derived self and viral peptides (Stet and Egberts, 1991). Considering *Neoparamoeba* antigenic peptides are unlikely to derive endogenously it seems appropriate that no association was found between AGD resistance/susceptibility and polymorphism of the *MHC* class I locus. However, the fact that no RFLP variation was observed within the *MHC* class Iα2 locus requires further investigation before confirming this view.

The association between AGD resistance and polymorphism of the *MHC* class IIβ1 domain agrees with the hypothesis that *MHC* class II genes may be responsible for the presentation of exogenously derived peptides. While yet to be investigated in AGD infected Atlantic salmon, *MHC* class II antigen presentation has been found to be essential for resistance to both endoparasitic and ectoparasitic diseases in mammals. For instance, parasitic challenge of *MHC* class II knockout and control mice found the *MHC* class II molecules are essential for an effective immune response against the blood-stage malarial endoparasites, *Plasmodium yoelii* and *Plasmodium chabaudi* (Cigel et al., 2003). Moreover, antigen presentation by the *MHC* class II molecules is essential for increased resistance to some gastrointestinal ectoparasites (i.e. *Schistosoma mansoni*) in humans (Quinnell, 2003).

Correlating MHC polymorphism with AGD resistance suggests the MHC genes could be appropriate candidates as quantitative trait loci (QTL) for disease resistance in a selective breeding program. The MHC genes have recently been considered as QTLs for resistance to furunculosis and ISA in Atlantic salmon (Grimholt et al., 2003) and resistance to parasite infection (*Trypanoplasma borreli*) in carp (Wiegertjes et al., 1995, Van Muiswinkel et al., 1999). While the MHC genes appear to play an important role in disease resistance in teleosts (Palti et al., 2001; Langefors et al., 2001a; Grimholt et al., 2003), it is likely additional genes are also involved. The association between a gene and disease resistance may be the result of the gene itself or the gene in question may be in linkage disequilibrium with another gene that is causing the observed resistance. The *MHC* class I and class II genes in teleosts reside on different linkage groups (Grimholt et al., 2002; Stet et al., 2002) and genes encoding the TAP transporter proteins (Hansen et

al., 1999) and natural killer cells (Sato et al., 2003) are only in linkage with genes of the *MHC* class I. Very few studies have addressed linkage of *MHC* class IIβ genes in Atlantic salmon. As a result the possibility that other linked genes are contributing to AGD resistance cannot be excluded.

The possible inherent resistance to AGD observed in this study is prevented from expressing itself under current management practices. AGD treatment has always involved the freshwater bathing of infected fish. While this treatment has successfully reduced AGD related mortalities, it also ensures the continued susceptibility of the population (Munday et al., 2001). In the wild susceptible individuals carrying unfavourable genetic alleles would succumb to a particular disease, and are unlikely to reproduce and pass the unfavourable alleles on to subsequent generations. This mode of selection ensures resistant alleles are maintained and possibly enhanced within a population. AGD treatment in Tasmania however is applied according to an on-farm assessment of gross gill lesions often before any mortality is observed (Clark and Nowak, 1999). As a result susceptible individuals possibly carrying unfavourable alleles will not succumb to AGD and may therefore pass these alleles on to subsequent generations. The frequency of favourable alleles is therefore unlikely to vary between year classes and as a result the level of susceptibility will remain constant across generations. Considering that the frequency of the *MHC* class IIβ1 RFLP (*j*) associated with AGD resistance was consistent and low across both population analysis samples (Table 8.6), it can be assumed the RFLP (or more specific allele) is providing little advantage to the industry under current culture conditions.

While this study provides evidence that MHC polymorphism is linked with AGD resistance, limitations of the study must be considered. For instance, while the association has been made between MHC polymorphism and experimental AGD it remains to be established whether the same RFLP phenotypes convey AGD resistance or susceptibility in a field infection. Furthermore this study was limited by its small sample sizes and caution must therefore be taken when making inferences pertaining to a larger population. Nevertheless, these results strongly advocate utilising the polymorphism of the *MHC* class IIβ1 locus, identified through RFLP, to further investigate AGD resistance and susceptibility at the molecular level.

8.5 Conclusion

Despite the limitations of this pilot study, there are four major observations from this initial investigation of MHC variation and its association with AGD:

- allele frequency differentiation between two samples from the Tasmanian population and one from its ancestral population;
- a possible reduction in MHC heterozygosity within the Tasmanian population;
- perhaps a lower genetic variation in samples with a western Atlantic origin compared to those with an eastern Atlantic origin; and
- an association between *MHC* class IIβ1 polymorphism and AGD resistance/susceptibility.

These observations suggest that further investigation of the MHC genes and other genes of potential commercial importance should be undertaken and in particular in relation to AGD where some evidence of a genetic basis to resistance/susceptibility has been demonstrated.

9. AFLP ANALYSIS FOR SEX MARKER

The research below was undertaken towards achieving Objective 1.

9.1 Introduction

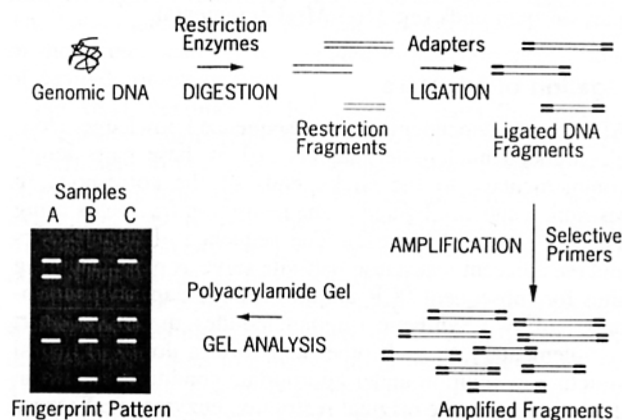
Development of a sex marker in Atlantic salmon is required to improve the efficiency of all-female production through the unambiguous identification of sex reversed neo-males from normal males. A neo-male being a genotypic female treated at an early life stage with a male hormone (e.g. androgen) resulting in an individual carrying female genetic characteristics but phenotypic male characters (e.g. produces milt). A cross between a neo-male and a normal female results in all-female progeny. Neo-males and normal males are phenotypically very similar, and often used discerning characteristics such as an incomplete sperm duct are not wholly reliable.

Although Devlin et al. (1991) identified a Y-chromosome marker for the Chinook salmon (*Oncorhynchus tshawytscha*) none has yet been acknowledged for Atlantic salmon. Elliott and Reilly (1998) used a subtractive hybridization technique (RDA) similar to that applied by Devlin et al (1991), and McGowan and Davidson (1998) applied the RAPD (random amplification of polymorphic DNA) technique but neither study located a reliable marker in Atlantic salmon. In this study we have applied the AFLP™ (Amplified Fragment Length Polymorphism) technique. This approach has recently been successful in isolating sex markers (e.g. penaeid prawns, Li et al., 2003).

AFLP is a powerful DNA fingerprinting technique, first described by Vos et al. (1995), based on the selective amplification of restriction fragments from a total DNA digest of genomic DNA. It does not require any prior knowledge of the organism's DNA sequence. The technique involves restriction of the genomic DNA and selective amplification of a subset of all the fragments. This method is described as a cost-effective way to generate marker-assisted genetic maps for discriminating between closely related lines that differ for a particular trait of interest.

The technique involves digesting the genomic DNA with two restriction enzymes: a frequent base cutter such as Mse I and a rare base cutter such as EcoRI (Figure 9.1). Adapters which recognise the restriction cut sites of Mse I and EcoRI are attached to the ends of these DNA fragments. These adapters allow a proportion of the DNA fragments to be amplified using the polymerase chain reaction (PCR). Primers designed to recognise the adapter sequence with 1-3 additional bases of the researcher's choice at the 3' end are used to selectively amplify a proportion of the DNA fragments from the total pool of DNA fragments. The number of DNA fragments can be further reduced by doing a second round of PCR amplification using the previous amplification product as a template. These amplification products are then separated by size on either an agarose or polyacrylamide gel and visualised using either a stain (ethidium bromide, silver staining) or by labelling the PCR primers with a radio or fluorescent label. The different banding patterns are then compared between individuals.

Figure 9.1. An outline of the AFLP protocol (reproduced from Blears et al. 1998)



For our purpose of identifying a potential sex marker we aimed to compare AFLP fragment patterns between female and male full-siblings, thereby reducing the number of genetic differences likely to be observed.

9.2 Materials and Methods

9.2.1 Sibling DNA samples

Our main attempt to obtain DNA from full-sibs involved the sampling of 100 juvenile fish from the Nortas Hatchery in September 2001. These were the progeny from a May 2001 spawning for which tissue samples had been collected from the broodstock. The caudal fin was removed from each juvenile fish and frozen at -80°C. The remainder of the fish was stored in 10% formalin for sex identification by histology. DNA was extracted from all juvenile fish that were then genotyped and assigned to familial groups using the pedigree suite A microsatellites (Chapter 6).

Fifteen whole fish (less caudal fin) representing three families (five from each) were sent to the Tasmanian Department of Primary Industries, Water and Environment, Mt Pleasant Laboratories, for sex determination by histology. Unfortunately, unambiguous determination of the sex of these sibling individuals was not achieved. No further samples from this collection were sent for sex determination.

To undertake our AFLP analysis we obtained DNA samples from 12 confirmed full-siblings (6 males and 6 females) supplied by Dr Eric Verspoor, FRS Marine Laboratory, Aberdeen, Scotland. The samples were stored at -20°C.

9.2.2 AFLP protocol

Approximately 300-500ng DNA was digested with two restriction enzymes (New England Biolabs). The DNA was digested with either 5U MseI and 5U EcoRI at 37°C for 3 hours or with 5U TaqI for 2 hours at 65°C followed by the addition of 5U EcoRI and incubated at 37°C for 2 hours. All samples were then heated to 80°C for 15 minutes to denature the enzymes. Adapters were attached to the digested DNA by incubating 5pmol of EcoRI adapter and 50pmol of MseI or Taq I adapters with 1U T4 DNA ligase at 37°C for 3 hours. The adapter ligated samples were diluted 1 in 10 in autoclaved purified water for the preselective amplification step. The preselective amplification step involved one round of PCR amplification using one primer specific to the EcoRI adapter sequence and

one primer specific to the MseI sequence. The PCR reaction was in a 20ul volume which consisted of 1x PCR Buffer, 1.5mM $MgCl_2$, 2mM of each dNTP, 0.75µM of Mse or Taq primer with one selective base (A) and 0.75µM of Eco primer with one selective base (A), 1U of Taq ploymerase (FisherBiotech) and 5ul of diluted adapter ligated sample. The cycling parameters were as follows: 30 cycles of 94°C/30sec, 56°C/1 min, 72°C/1 min, followed by a final extension at 72°C for 10 mins. The amplification products from the preselective amplification were diluted 1 in 10 in autoclaved purified water and used as template for a second round of amplification. This second PCR reaction consisted of 1x PCR buffer, 1.5mM $MgCl_2$, 2mM each dNTP, 0.5uM of MseI or TaqI primer with 3 selective bases, 0.5µM of EcoRI primer with 3 selective bases, 1 unit of Taq polymerase (Fisher Biotech) and 2.5 ul of diluted preselective amplification. Cycling parameters began with 13 cycles of 94°C/30sec, 65°C/1 min decreasing by 0.7°C per cycle, 72°C/1 min followed by 23 cycles of 94°C/30sec, 56°C/1 min and 72°C/1 min.

The adapters and the preselective and selective (additional bases in bold) primers used in this study were:

Oligo Name	Sequence (5'-3')
Taq I adapter A	GAC GAT GAG TCC TGA G
Taq I adapter B	CGC TCA GGA CTC AT
Mse I adapter A	GAC GAT GAG TCC TGA G
Mse I adapter B	TAC TCA GGA CTC AT
EcoR I adapter A	CTC GTA GAC TGC GTA CC
EcoR I adapter B	AAT TGG TAC GCA GTC TAC
Taq I + A	GAT GAG TCC TGA GCG A**A**
Mse I + A	GAT GAG TCC TGA GTA A**A**
EcoR I + A	GAC TGC GTA CCA ATT C**A**
EcoR I + ACG	GAC TGC GTA CCA ATT C**ACG**
EcoR I + AGC	GAC TGC GTA CCA ATT C**AGC**
EcoR I + ATG	GAC TGC GTA CCA ATT C**ATG**
EcoR I + AAT	GAC TGC GTA CCA ATT C**AAT**
Taq I + ACG	GAT GAG TCC TGA GCG A**AC G**
Taq I + AGT	GAT GAG TCC TGA GCG A**AG T**
Taq I + ACA	GAT GAG TCC TGA GCG A**AC A**
Taq I + ACT	GAT GAG TCC TGA GCG A**AC T**
Taq I + ATG	GAT GAG TCC TGA GCG A**AT G**
Mse I + ACG	GAT GAG TCC TGA GTA A**AC G**
Mse I + AGT	GAT GAG TCC TGA GTA A**AG T**
Mse I + ACA	GAT GAG TCC TGA GTA A**AC G**
Mse I + ACT	GAT GAG TCC TGA GTA A**AC T**
Mse I + ATG	GAT GAG TCC TGA GTA A**AT G**

AFLP products were visualised on 8% non-denaturing polyacrylamide gels (8cm x 9cm). These gels consisted of 8mL Page Plus (Amersco), 4mL 10x TBE and 28mL water. The gel mix was degassed for 2 mins and 300ul freshly made 10% ammonium persulfate and 30µl TEMED was added. The gels were poured using a syringe and needle. Gels were left to polymerise for 2 hours.

5µl of PCR product was mixed with 2.5µl of 3x loading dye (Amersco) and loaded onto the gel using duckbill tips. A 100bp DNAladder (Gibco BRL) was run on all gels. The gels were run in 1XTBE buffer at 110V, 20-25mA for 1 hour 30 mins or until the blue dye got close to end of the gel. Gels were stained with ethidium bromide for about 30 - 45 minutes and then photographed with a digital camera.

Ethidium bromide was used to stain the gels because it is a fast, cheap and easy to use staining method. Ethidium bromide can detect down to 10ng of DNA (Scott et al., 1998) which is not as sensitive as some other methods but this was considered an advantage as it could potentially reduce the complexity of the banding pattern seen on the gel. A small number of samples were PCR-amplified using fluorescent labelled primers and run on a ABI Prism™ 377 DNA Sequencer. The high sensitivity of this equipment made it difficult to interpret the gel data, and so we concentrated on the ethidium bromide method. Gel percentages ranging from 5-15% were trialled. A concentration of 8% was found to be the most suitable for this study.

9.3 Results

AFLP bands observed ranged in size from 20 to 1000 base pairs, and the number of bands produced by each primer combination ranged from 11 to 25 (Table 9.1). These numbers are only approximate due to the less sensitive nature of the ethidium bromide staining. Initially samples were screened as two groups: male and female. This is known as bulked segregant analysis (Michelmore et al., 1991). Female samples and male samples were separately pooled and then compared for differences in the banding patterns. This is known as a primary screen and is a fast way to find potential differences in AFLP banding patterns between the sexes. When a difference was detected (the presence or absence of a band) a secondary screen was done where all 12 individuals were analysed separately to see if the observed difference was consistent in either all females or all males. Our initial primary screens did show differences between the sexes. However in the secondary screens using individual fish a completely different banding pattern was observed and no female or male specific bands were seen.

The difference in the banding patterns between the pooled and individual samples was disappointing, but not totally unexpected. Li et al. (2002) report finding putative sex-specific markers for the green spotted pufferfish (*Tetraodon nigroviridis*) during their primary screen but none observed during their secondary screen; however they make no mention of observing different banding patterns as we have observed. Due to resources restriction we have been unable to follow-up this problem.

9.4 Discussion

The only teleost that has had its major sex-determining gene characterised is the medaka (Woram et al., 2003). The isolation of sex-specific markers in salmonids has met with mixed success (see Woram et al., 2003; Devlin and Nagahama, 2002). The sex-determining region seems to be largely undifferentiated and linkage studies have shown a lack of gene conservation between salmonids (Woram et al., 2003). Three species of salmonids can be sexed based on their karyotype: lake trout (*Salvelinus namaycush*), sockeye salmon (*Oncorhynchus nerka*) and some strains of rainbow trout (*O.mykiss*) (Iturra et al. 2001). Sex chromosomes are not apparent in the Atlantic salmon karyotype so it is thought that a sex-determining region is likely to be located on an autosome (McGowan and Davidson, 1998). A sex-linked growth hormone sequence has been isolated in four species of Pacific salmon: chinook, coho, chum and pink salmon (Devlin et al., 2001).

AFLP markers linked to a sex-determining region have been successfully isolated in penaeid prawns (Li et al., 2003), three-spined stickleback (Griffiths et al., 2000), rainbow trout (Young et al., 1998) and ostriches (Griffiths and Orr, 2000). Attempts to find AFLP

markers linked to sex have been unsuccessful in pufferfish (Li et al., 2002). In this study we have examined 30 AFLP primers combinations on sibling DNA and in all cases, no sex-linked markers were identified.

Table 9.1. Total number of bands observed for male and female samples at each primer combination

Selective primer combination		Number of bands observed females	Number of bands observed males
Eco-AGC	+ Mse-AGT	20	21
	+ Mse-ACA	16	16
	+ Mse-ACT	17	18
	+ Mse-ACG	17	15
	+ Mse-ATG	14	13
Eco-ACG	+ Mse-ATG	17	18
	+ Mse-ATG	14	16
	+ Mse-AGT	14	14
	+ Mse-ACA	14	16
	+ Mse-ACT	15	19
Eco-AGC	+ Taq-AGT	13	15
	+ Taq-ACA	7	9
	+ Taq-ACT	15	16
	+ Taq-ACG	18	14
	+ Taq-ATG	13	11
Eco-ACG	+ Taq-ATG	23	21
	+ Taq-ACG	21	23
	+ Taq-AGT	23	22
	+ Taq-ACA	16	20
	+ Taq-ACT	11	14
Eco-ATG	+ Taq-ACA	20	13
	+ Taq-ACG	21	24
	+ Taq-ACT	16	18
	+ Taq-AGT	20	23
	+ Taq-ATG	20	19
Eco-AAT	+ Taq-ACA	17	22
	+ Taq-ACG	14	15
	+ Taq-ACT	18	15
	+ Taq-AGT	25	24
	+ Taq-ATG	20	22

10. BENEFITS AND ADOPTION

The Tasmanian Atlantic salmon industry is the principal beneficiary of the molecular genetics research undertaken during this project. This benefit arises through the provision of:

- an effective and efficient DNA pedigree system for use in selective breeding programs for parentage and relatedness analyses;
- a more reliable assessment of the genetic status of the Tasmanian population, and the provision of a more realistic genetic baseline for the population for future reference;
- evidence that current breeding protocols have been satisfactory to maintain a healthy population; and
- preliminary evidence of a genetic influence on AGD resistance/susceptibility

A secondary beneficiary is the Australian aquaculture industry. The experience gained in applying molecular genetic techniques to assist breeding programs, and understanding genetic variation in the population and in relation to specific traits will have indirect benefit for other aquaculture species.

This research will prove to be cost effective when the industry commences a selective breeding program that uses genetic makers for relatedness and pedigree analyses, and the inclusion of AGD resistance and other selection traits with or without the use of marker assisted selection. Such a program has been recommended to the industry following two consultancies for the TSGA (1999 and 2003), and is under serious consideration in early 2004. The application of molecular markers for relatedness analysis of broodstock has been used by an industry member in March 2004. It is estimated that a selection program would return at a minimum production gains of 5% within two generations, and the annual cost to FRDC and the industry of this project has been less than 0.2% of the current value of the industry.

11. FURTHER DEVELOPMENT

The research undertaken has provided a suite of ten highly informative markers for pedigree analysis, however, within the resources available we have been unable to fully refine the protocol to run as a single multiplex PCR reaction. Further refinement is recommended, but with due regard to the commercial genetics laboratory contracted to run future bulk genotype analyses for the selective breeding program.

A genetic component to AGD infection (resistance/susceptibility) has been suggested from the pilot study undertaken. It is recommended that this research is expanded into a more detailed study of the molecular assessment of both the AGD infection process (to assist improvements to and/or development of new treatments) and the reaction of individuals to infection (to develop resistance/susceptibility genetic markers for selective breeding).

In addition a more detailed comparison of MHC polymorphism within the local population to other hatchery and wild populations is recommended.

Further, growing international experience and gains from selective breeding, and our preliminary evidence of a genetic component to AGD resistance strengthens the argument

in favour of commencing a selective breeding program to ensure the long-term sustainability of the Tasmanian Atlantic salmon industry.

12. PLANNED OUTCOMES

The project has successfully led to the development of a reliable and informative DNA pedigree protocol, a better understanding of the genetic health of the population and preliminary evidence for a genetic component to AGD resistance/susceptibility. All these outputs ultimately add further evidence to the growing justification and support for an industry selective breeding program to ensure the long-term sustainability of the industry.

As this Final Report is being prepared the DNA pedigree for approximately 400 prospective 2004 broodstock is being provided to SALTAS. Based on these genotypes, breeding advice on these individuals in relation to matings to minimize inbreeding and reduce overrepresentation of particular family groups will be provided. These matings it is hoped will produce the first generation of a selective breeding program.

13. CONCLUSIONS

The project has successfully achieved its objectives.

13.1 Objective 1.

To further develop and apply molecular markers for use in industry breeding programs.

To achieve this objective a suite of over 50 microsatellite markers were sourced, optimised and evaluated for use in population analyses and DNA pedigree analyses. Three new highly informative loci were added to the original eight for population analyses, and a suite of 10 informative markers were selected and tested for pedigree analysis.

In addition specific regions of selected immune genes (MHC) were targeted, PCR conditions were optimised and regions sequenced for use in a pilot study examining population variation and association with AGD susceptibility/resistance.

13.2 Objective 2.

To genotype selected broodstock with a suite of microsatellite markers to enable efficient pedigree analysis of progeny.

A refined set of 10 highly informative DNA markers has been developed that provides a high efficiency in assigning parentage and family relationships. In addition this protocol has been compared with a newly available commercial protocol. Through the use of both protocols, discounting any human errors and those that may occur at the hatchery, almost 100% efficiency in DNA pedigree analysis is available.

13.3 Objective 3.

To compare microsatellite DNA variation from archival scale samples from the progenitor Canadian population with past and current cohorts of the Tasmanian population.

DNA was successfully extracted from the 30-year-old archived scales from the River Philip population, and variation at 11 microsatellite markers in those samples was

compared with that in a 1991 hatchery sample from the River Philip, a 1997 sample from NSW and seven (six consecutive) year-class from the Tasmanian population. Despite a significant loss of variation in terms of numbers of alleles, the local population appears stable and the estimated mean effective population size (ca. 180 individuals) over the past 30 years is very realistic given the history of the population over this time.

14. REFERENCES

Adams, M.B., and Nowak, B.F., 2001. Distribution and structure in the gills of Atlantic salmon, *Salmo salar* L., affected with amoebic gill disease. J. Fish Dis. 2: 535-542.

Adams, M.B., and Nowak, B.F., 2003. Amoebic gill disease: sequential pathology in cultured Atlantic salmon, *Salmo salar* L. J. Fish Dis. 26: 601-614.

Adcock, G.J., Bernal Ramirez, J.H., Hauser, L., Smith, P. and Carvalho, G.R. 2000. Screening of DNA polymorphisms in samples of archived scales from New Zealand snapper. Journal of Fish Biology 56:1283-1287

Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D. and Hedrick, P.W., 2002. Resistance to three pathogens in the endangered winter-run Chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Can. J. Fish. Aquat. Sci. 59: 966-975.

Angers, B., Bernatchez, L, Angers, A. and Desgroseillers, L. 1995. Specific microsatellite loci for brook charr reveal strong population subdivision on a microgeographic scale. J.Fish.Biol 47(suppl A):177-185.

Banks, M.A., Blouin, M.S., Baldwin, B.A., Rashbrook, V.K., Fitzgerald, H.A., Blankenship, S.M. and Hedgecock, D. 1999. Isolation and inheritance of novel microsatellites in Chinook salmon. J.Hered. 90 (2):281-288.

Bernatchez, L. and Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J. Evo Biol. 16: 363-377.

Beaumont, A.R., Hoare, K., 2003. Biotechnology and genetics in fisheries and aquaculture. Blackwell Science Ltd, Osney Mead, Oxford, pp. 32 –34.

Blears, M.J., De Grandis, S.A., Lee, H. and Trevors, J.T. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology and Biotechnology. 21: 99-114.

Cairney, M., Taggart, J.B. and Hoyheim, B. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Molecular Ecology 9 (12) 2175-2178.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K. and O'Brien, S. J. 1999. *HLA* and HIV-1: heterozygote advantage and B*35-C*04 disadvantage. Science. 283: 1748-1752.

CaValli-Storza, L.L. and Bodmer, W.F. 1971. The genetics of human populations. W.H. Freeman and Company, San Francisco, pp 54-57.

Cigel, F., Batchelder, J., Burns, J.M., Yańez, D., van der Heyde, H., Manning, D.D. and Weidanz, W.P. 2003. Immunity to blood-stage murine malarial parasites is MHC class II dependent. Immunol. Letters. 89: 243-249.

Clark, A. and Nowak, B.F. 1999. Field investigations of amoebic gill disease in Atlantic salmon, *Salmo salar* L., in Tasmania. J. Fish. Dis. 22: 433-443.

Danzmann, R.G. 1997. PROBMAX: A computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. J. Hered. 88: 333

Devlin, R.H., McNeil, B.K., Groves, T.D.D. and Donalson, E.M. 1991. Isolation of Y-chromosomal DNA probe capable of determining genetic sex in Chinook salmon (*Oncorhynchus tshawytscha*). Can, J. Fish. Aquaa. Sci. 48: 1606-1612.

Devlin, R.H., Biagi, C.A. and Smailus, D.E. 2001. Genetic mapping of Y-chromosomal DNA markers in Pacific salmon. Genetica 111:43-58.

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

Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Pytochemistry Bulletin 19:11-15.

Elliott, N.G. and Reilly, A. 1998. Microsatellite variation and identification of a Y-chromosome marker in Atlantic salmon. Final report FRDC projects 95/80 and 96/347.

Elliott, N.G. and Reilly, A. 2003. Likelihood of bottleneck event in the history of the Australian population of Atlantic salmon (*Salmo salar* L.). Aquaculture 215: 31-44.

Estoup, A., Gharbi, K., SanCristobal, M.., Chevalet, C. Haffray, P. and Guyomard, R. 1998. Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. Can.J.Fish.Aquat.Sci 55:715-725.

Gjetvaj, B., Ball, R.M., Burbridge, S., Bird, C.J., Kenchington, E. and Zouros, E. 1997. Amounts of polymorphism at microsatellite loci in the sea scallop *Placopecten magellanicus*. J. Shellfish Res. 16 (2): 547-553.

Griffiths, R., Orr, K.J., Adam, A. and Barber, I. 2000. DNA sex identification in the three-spine stickleback. J. Fish Biol. 57: 1331-1334.

Griffiths, R. and Orr, K. 2000. The use of amplified fragment length polymorphism (AFLP) in the isolation of sex-specific markers. Molecular Ecology 8: 671-674.

Grimholt, U., Getahun, A., Trudi, H. and Stet, R.J.M. 2000. The major histocompatibility class II alpha chain in salmonid fishes. Dev. Comp. Immunol. 24: 751-763.

Grimholt, U., Drabløs, F., Jørgensen, S.M., Høyheim, B. and Stet, R.J.M. 2002.The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics. 54: 570-581.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S. and Stet, R.J.M. 2003. MHC polymorphisms 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.

Gross, K.A., Morrison, R.N., Butler, R. and Nowak, B.F. 2004. Atlantic salmon, *Salmo salar* L., previously infected with *Neoparamoeba* sp. are not resistant to re-infection and have suppressed phagocyte function. J. Fish Dis. 27: 47-56.

Guy, D.R. and Hamilton, A. 2003. Genetic parameters for harvest and processing traits in Landcatch strains of Atlantic salmon *Salmo salar*. Genetics in Aquaculture VIII, Puerto Varas, Chile. Book of Abstracts Pg79.

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

Hansen, M.M. 2002. Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. Molecular Ecology 11:1003-1015

Hartl, D.L. and Clark, A.G. 1997. Principals of population genetics. Sinauer Associates Inc Publishers, Massachusetts, pp. 118-119.

Heath, D.D., Busch, C., Kelly, J. and Atagi, D. 2002. Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). Molecular Ecology 11:197-214.

Hedgecock, D., Chow, V. and Waples, R.S. 1992. Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. Aquaculture 108: 215-232.

Hedgecock, D. and Sly, F. 1990. Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. Aquaculture 88: 21-38.

Iturra, P., Lam, N., de la Fuente, M, Vergara, N. and Medrano, J.F. 2001. Characterization of sex chromosomes in rainbow trout and coho salmon using fluorescence *in situ* hybridization (FISH). Genetica 111:125-131.

Jackson, T.R., Ferguson, M.M., Danzmann, R.G., Fishback, A.G., Ihssen, P.E., O'Connell, M. and Crease, T.J. 1998. Identification of two QTL influencing upper temperature tolerance in rainbow trout (*Oncorhynchus mykiss*). Heredity 80: 143-151.

Jeffery, K.J.M. and Bangham, C.R.M. 2000. Do infectious diseases drive MHC diversity? Microbes and Infection. 2: 1335-1341.

Kruiswijk, C.P. 2000. Evolution of major histocompatibility genes in cyprinid fish, Molecular analysis and phylogenies. PhD thesis, Wageningen Institute of Animal Sciences, Wageningen University, The Netherlands.

Landry, C. and Bernatchez, L. 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibiity complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Mol. Ecol. 10: 2525-2539.

Langefors, Å., Von Schantz, T. and Widegren, B. 1998. Allelic variation of MHC class II in Atlantic salmon; a population genetic analysis. Heredity. 80: 568-575.

Langefors, Å., Lohm, J., Von Schantz, T. and Grahn, M. 2000. Screening of MHC variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. Mol. Ecol. 9: 215-219.

Langefors, Å., Lohm, J., Grahn, M., Andersen, Ø. and Von Schantz, T. 2000b. Association between major histocompatibility complex class IIB and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proc. R. Soc. Lond. 268: 479-485.

Langefors, Å., Lohm, J. andVon Schantz, T. 2001. Allelic polymorphism in MHC class II *B* in four populations of Atlantic salmon (*Salmo salar*). Immunogenetics. 53: 329-336.

Letcher, B.H. and King, T.L. 2001. Parentage and grandparentage assignment with known and unknown matings: application to Connecticut River Atlantic salmon restoration. Can.J.Fish.Aquat.Sci. 58:1812-1821.

Li, Y., Hill, J.A., Yue, G.H., Chen, F. and Orban, L. 2002. Extensive search does not identify genomic sex markers in *Tetraodon nigroviridis*. Journal of Fish Biology. 61:1314-1317.

Li , Y., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., Crocos, P., Preston, N. and Lehnert, S. 2003. Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. Aquaculture 219:143-156.

McConnell, S.K., O'Reilly, P., Hamilton, L., Wright, J.M. and Bentzen, P. 1995 Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. Can.J.Fish.Aquat.Sci 52:1863-1872.

McGowan, C. and Davidson, W.S. 1998. The RAPD technique fails to detect a male-specific genetic marker in Atlantic salmon. Journal of Fish Biology 53:1134-1136.

Michelmore, R.W., Paran, I. and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect

markers in specific genomic regions by using segregating populations. Proc.Natl.Acad.Sci. USA 88:9828-9832.

Miller, K.M. and Withler, R.E. 1996. Sequence analysis of a polymorphic MHC class II gene in Pacific salmon. Immunogenetics. 43: 337-351.

Moran, J.D.W., Margolis, L., Webster, J.M. and Kent, M.L. 1999. Development of *Kudoa thyrsites* (Myxozoa:Myxosporea) in netpen-reared Atlantic salmon determined by light microscopy and a polymerase chain reaction test. Dis.Aquat.Org., 37: 185-193.

Morris, D.B., Richard, K.R. and Wright, J. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Can. J.Fish.Aquat. Sci 53:120-126

Munday, B.L., Zilberg, D. and Findlay, V. 2001. Gill disease of marine fish caused by infection with *Neoparamoeba pemaquidensis*. J. Fish Dis. 24: 497-507.

Nielsen, E.E., Hansen, M.M. and Loeschcke, V. 1997. Analysis of microsatellite DNA from old scale samples of Atlantic salmon *Salmo salar* : a comparison of genetic composition over 60 years. Molecular Ecology 6: 487-492.

Norris, A.T., Bradley, D.G., and Cunningham, E.P. 1999. Microsatellite genetic variation between and within farmed and wild Atlantic salmon *Salmo salar* L. populations. Aquaculture. 180: 247-264.

Norris, A.T., Bradley, D.G. and Cunningham, E.P. 2000. Parentage and relatedness determination in farmed Atlantic salmon (Salmo salar) using microsatellite markers. Aquaculture. 182: 73-83.

O'Connell, M., Danzmann, R.G., Cornuet, J-M., Wright, J.M. and Ferguson, M.M. 1997. Differentiation of rainbow trout (*Oncorhynchus mykiss*) populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. Can.J.Fish.Aquat.Sci 54:1391-1399.

O'Reilly, P.T., Hamilton, L.C., McConnell, S.K. and Wright, J.M. 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Can. J. Fish. Aquat. Sci. 53: 2292-2298.

O'Reilly, P.T., Herbinger, C. and Wright, J.M. 1998. Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. Animal Genetics 29: 363-370.

Palti, Y., Nichols, K.M., Waller, K.I., Parson, J.E. and 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.

Paterson, S., Piertney, S.B., Knox, D., Gilbey, J. and Verspoor, E. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Molecular Ecology Notes 4: 160-162.

Presa, P. and Guyomard, R. 1996. Conservation of microsatellites in three species of salmonids. Journal of Fish Biology 49: 1326-1329.

Quinn, G.P. and Keough, M.J. 2002. Experimental design and data analysis for biologists. Cambridge University Press. Cambridge. pp. 386-387.

Quinnell, R.J. 2003. Genetics of susceptibility to human Helminth infection. Int. J. Parasitol. 33: 1219-1231.

Reilly, A., Elliott, N.G., Grewe, P.E., Clabby, C., Powell, R. and Ward, R.D. 1999. Genetic differentiation between Tasmanian cultured Atlantic salmon (*Salmo salar* L.) and their ancestral Canadian population: comparison of microsatellite DNA and allozyme and mitochondrial DNA variation. Aquaculture. 173: 459-469.

Raymond, M. and Rousset, F. 1995. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. J.Heredity 86: 248-249

Rice, W.R. 1989. Analysing tables of statistical tests. Evolution 43: 223-225.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. and Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. USA. 81: 8014-8018.

Sanchez, J.A., Clabby, C., Ramos, D., Blanco, G., Flavio, F., Vazquez, E. and Powell, R. 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). Heredity 77:423-432.

Sato, A., Mayer, W.E., Overath, P. and Klein, J. 2003. Genes encoding putative natural killer cell C-type lectin receptors in teleosts. PNAS. 100, 7779-7784.

Schneider, S., Roessli, D. and Excoffier, L. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scribner, K.T., Gust, J.R. and Fields, R.L. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. Can.J.Fish.Aquat.Sci. 53: 833-841.

Scott, D., Walker, M., Clarck, C., Prakash, C., and Deahl, K. 1998. Rapid Assessment of Primer Combinations and Recovery of AFLP™ Products using Ethidium Bromide Staining. Plant Molecular Biology Reporter 16: 41-47.

Shuber, A., Grondin, V.J. and Klinger, K.W. 1995. A simplified procedure for developing multiplex PCRs. Genome Research 3: 488-493

Stet, R.J.M. and Egberts, E. 1991. The histocompatibility system in telostean fishes: from multiple histocompatibility loci to a major histocompatibility complex. Fish. Shellfish Immunol. 1: 1-16.

Stet, R.J.M., de Vries, B., Muddle, K., Hermsen, T., van Heerwaarden, J., Shum, B.P. and Grimholt, U. 2002. Unique haplotypes of co-segregating major histocompatibility class II *A* and class II *B* alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics. 54: 320-331.

Valsecchi, E. 1998. Tissue-boiling: a short-cut in DNA extraction for large-scale population screenings. Mol. Ecol., 7: 1243-1245.

Van Muiswinkel, W.B., Wiegertjes, G.F.and Stet, R.J.M. 1999. The influence of environmental and genetic factors on the disease resistance of fish. Aquaculture, 172: 103-110.

Verspoor, E. 1988. Reduced genetic variability in first-generation hatchery populations of Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquatic Sciences 45: 1686-1690.

Villanueva, B., Verspoor, E. and Visscher, P.M. 2002. Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. Animal Genetics 33: 33-41.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, S., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21): 4407-4414.

Ward, R.D., Grewe, P.M. and Smolenski, A.J. 1994. A comparison of allozymes and mitochondrial DNA in Atlantic salmon from Tasmania and from the ancestral population in Canada. Aquaculture 126: 257-264.

Ward, R.D., Jorstad, K.E. and Maguire, G.B. 2002. Microsatellite diversity in rainbow trout (*Oncorhynchus mykiss*) introduced to Western Australia. Aquaculture 219:169-179.

Whitmore, D.H., Thai, T.H., and Craft, C.M. 1992. Gene amplification permits minimally invasive analysis of fish mitochondrial DNA. Transactions of the American Fisheries Society 121: 170-177.

Winkler, F.M., Bartley, D. and Diaz, N.F. 1999. Genetic differentiation among year classes in a hatchery population of coho salmon (*Oncorhynchus kisutch* (Walbaum)) in Chile. Aquaculture 173: 423-431.

Wiegertjes, G.F., Groeneveld, A. and Van Muiswinkel, W.B. 1995. Genetic variation in susceptibility to *Trypanoplasma borreli* infection in common carp (*Cyprinus carpio* L.). Vet. Immunuol. Immunopathol. 47: 153-161.

Woram, R., Gharbi, K., Sakamoto, T., Hoyheim, B., Holm, L., Naish, K., McGowan, C., Ferguson, M., Phillips, R., Stein, J., Guyomard, R., Cairney, M., Taggart, J.,

Powell, R., Davidson, W. and Danzmann, R. 2003. Comparative Genome Analysis of the Primary Sex-Determining Locus in Salmonid Fished. Genome Research. 13:272-280.

Xia, C., Kiryu, I., Dijkstra, J.M., Azuma, A., Nakanishi, T. and Ototake, M. 2002. Differences in MHC class I genes between strains of rainbow trout (*Oncorhynchus mykiss*). Fish. Shellfish Immunol. 12: 287-301.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P., Thorgaard, G.H. 1998. A Detailed Linkage Map of Rainbow Trout Produced Using Doubled Haploids. Genetics 148: 839-850.

Zilberg, D., Gross, A. and Munday, B.L. 2001. Production of salmonid amoebic gill disease by exposure to *Paramoeba* sp. harvested from the gills of infected fish. J. Fish Dis. 24: 79-82.

APPENDIX 1: Intellectual property

The intellectual property and valuable information arising from this research lies with copyright in this Report. At this time no specific intellectual property is identified as being generated from the results of the research.

APPENDIX 2: Staff

Staff engaged on the project:

Principal Investigator	
Dr Nicholas Elliott	CSIRO Marine Research
Research Assistants	
Ms Bronwyn Innes	CSIRO Marine Research
Dr Mathew Cook	CSIRO Marine Research
Honours Student	
Mr James Wynne	University of Tasmania

APPENDIX 3: Supplementary data

I: Extraction methods for total genomic DNA

CTAB extraction

This method has been modified from Saghai-Maroof et al. (1984) and Doyle and Doyle (1987).

1. Place 20mg tissue in a microcentrifuge tube and add 200µL of CTAB Buffer. Grind the tissue then add a further 500µL of CTAB buffer. Grind again if required.
2. Add 5µL proteinase K (20mg / mL). Vortex briefly.
3. Incubate tubes at 65°C in a heating block for 3-4 hours or until the tissue is well digested.
4. Add 600µL of chloroform:isoamyl alcohol (24:1). Mix tubes well by inversion. Centrifuge at 13,500 rpm for 20 minutes.
5. Use a pipette to remove the supernatant (upper layer) and transfer this layer to a fresh tube containing 600µL of phenol:chloroform:isoamyl alcohol (25:24:1). Mix tubes well by inversion and centrifuge at 13,500 rpm for 10 minutes.
6. Transfer the supernatant to a fresh tube containing 600µL of chloroform:isoamyl alcohol (24:1). Mix tubes well by inversion and centrifuge at 13,500 rpm for 5 minutes.
7. After the spin, transfer the supernatant to a fresh tube containing 600µL ice-cold (-20°C) isopropanol. Mix gently by inversion. A precipitate may be seen at this point.
8. Incubate the tubes at –20°C for at least 1 hour or overnight.
9. Centrifuge the tubes at 13,500 rpm (4°C) for 30 minutes. Gently decant the supernatant.
10. Add 1mL 70% ice-cold (–20°C) ethanol. Mix well by inversion and centrifuge at 13,500 rpm for 10 minutes. Gently decant the supernatant.
11. Dry the pellet and resuspend in 200µL of autoclaved water (purified).

Store samples in the fridge (short-term) or in the freezer (–20°C or –80°C long-term).

2X CTAB buffer

50mL 1M Tris-HCl pH 8.0
175 mL 4 M NaCl (or 40.9g)
20 mL 0.5 M EDTA
10g CTAB (hexadecyltrimethylammonium bromide)
1 mL of β-mercaptoethanol (add to working solution not stock solution)
dd H_2O to 500 mL

Extraction of DNA from scales

This is a modified version of the protocol detailed in Adcock et al. (2000). See also Whitmore et al. (1992).

1. Digest 1–2 scales in 200µL extraction buffer and 5µL proteinase K (20mg / ml) overnight at 55°C.
2. Extract this solution twice with 200µL phenol. After each addition of phenol, mix the tubes by inversion and centrifuge for 10 min at 13,500 rpm.

3. Extract this solution once with 200µL chloroform:isoamyl alcohol (24:1, v:v). After addition of the chloroform:isoamyl alcohol (24:1, v:v), mix the tubes by inversion and centrifuge for 10 min at 13,500 rpm.
4. Transfer the aqueous phase to a fresh tube and to this add 2 volumes of 100% ice-cold ethanol.
5. Incubate overnight at –20°C.
6. Centrifuge the samples for 10 min at 10,000 rpm. Decant the supernatant.
7. Resuspend the DNA pellet in 30µL MilliQ purified water.

Extraction buffer

100 mM Tris-HCl	20mL 1M Tris-HCl
10 mM EDTA	4mL 0.5M EDTA
0.1% SDS	2mL 10% SDS
0.05M DTT	10mL 1M DTT
	200mL lysis buffer

Extraction of DNA using a DNeasy Kit (Qiagen)

This kit (catalogue number 69506) was bought from Qiagen Pty Ltd Clifton Hill, Victoria. http://www.qiagen.com

Genomic DNA was extracted following the manufacturer's instructions (DNeasy protocol for animal tissues) with the following modifications: 10 - 20 mg of tissue was used, the ATL buffer/proteinaseK incubation step was lengthened to overnight at 55°C, only one elution step was done using 200µl of autoclaved purified water. Samples were incubated at room temperature for 5 min before the final elution spin.

Extraction of DNA from blood

Two methods were combined to create this protocol.
See FRDC projects 95/80 and 96/347 Appendix 15.2.1 (SDS lysis genomic DNA extraction from blood) and Moran et al. 1999.

Upon collection of blood it is advisable to use an anticoagulant such as EDTA or heparin. Heparin may inhibit PCR. Extraction of DNA from blood was problematic, most likely due to not using an anti-coagulent.

Lysis buffer

10mM TrisHCl	1mL 1M TrisHCl
10mM EDTA	2mL 0.5M EDTA
1% SDS	10mL 10% SDS
400mM NaCl	40mL 1M NaCl
	Q.V. MilliQ water
	100mL lysis buffer

1. To 25µl-100µL of blood add 500µL of lysis buffer and 10µL proteinase K (20mg/mL)
2. Incubate at 37°C overnight.
3. Add 600ul chloroform:isoamyl alcohol (24:1, v:v). Mix by inversion. Centrifuge at 13,500 rpm for 20 minutes.
4. Remove upper aqueous layer and transfer to a fresh tube. Add 600µL phenol:chloroform:isoamyl alcohol (25:24:1, v:v). Mix by inversion. Centrifuge at 13,500 rpm for 20 minutes.

5. Remove upper aqueous layer and transfer to a fresh tube. Add 600 µl chloroform:isoamyl alcohol (24:1, v:v). Mix by inversion. Centrifuge at 13,500 rpm for 20 minutes.
6. Transfer upper aqueous layer into a fresh tube and add 600 µl ice cold (–20°C) isopropanol. Mix gently by inversion. A precipitate may be seen at this point.
7. Incubate the tubes at –20°C for at least 1 hour or overnight.
8. Centrifuge the tubes at 13,500 rpm (4°C) for 30 minutes. Gently decant the supernatant.
9. Add 1mL 70% ice-cold (–20°C) ethanol. Mix well by inversion and centrifuge at 13,500 rpm for 10 minutes. Gently decant the supernatant.
10. Dry the pellet and resuspend in 200µL of autoclaved water (purified).

Shortened protocol for extraction of DNA

a) Tissue boiling method (Valsecchi, 1998)

1. Add 1ml buffer to each tube.
2. Perforate the lids or use a clip to prevent lids from bursting open
3. Immerse up to 2/3 depth in boiling water. Boil for 15 mins.
4. Cool down on ice before storing in the freezer (-20°C).

Extraction buffer

20mM Tris HCl pH 8.0	5mL 1M Tris HCl pH 8.0
2mM EDTA	1mL 0.5M EDTA
1% Triton-X100	2.5mL Triton-X100 *
	Q.V. MilliQ water
	250mL buffer

*make up solution minus the Triton-X100. Squirt the Triton-X100 on top. Triton-X100 takes 10-15 minutes to go into solution.

b) Overnight Incubation (Gjetvaj et al. 1997)

1. Use 200µL extraction buffer for scale samples or 500 µL extraction buffer for finclip and muscle samples.
2. Add 5µL proteinase K (20mg/mL) to each tube.
3. Briefly vortex tubes and incubate at 55°C overnight.
4. Inactivate the proteinase K by heating tubes at 95°C for 5 min.

Extraction buffer

10mM TrisHCl pH 8.3	2.5mL 1M TrisHCl pH 8.3
50mM KCl	50mL 0.25M KCl
0.5% Tween-20	1.25mL Tween-20
	Q.V. MilliQ water
	250mL buffer

II. Summary genetic statistics

Tables A1 to A11 provide the allele frequencies for all eleven microsatellite loci for each sample. The Canada 1971 and 1972 samples and the two Tasmania 2001 samples are shown as separate samples. Table A12 provides the observed and expected heterozygosity estimates for each locus in each sample, Table A13 the results of the NULLTEST, Tables A14 and A15 the F_{ST} results and Table A16 the Nk results.

Table A1. Allele frequencies for locus *Ssa171*

Allele	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
210	0.020	-	-	-	-	-	-	-	-	-	-	-
214	0.010	-	-	0.114	0.180	0.142	0.048	0.037	0.125	0.025	0.038	0.050
218	0.040	0.067	0.016	0.068	0.047	0.100	0.108	0.056	0.070	0.131	0.087	0.038
222	-	0.011	-	-	-	-	-	-	-	-	-	-
228	0.020	0.011	-	-	-	-	-	-	-	-	-	-
230	-	0.011	-	-	-	-	-	0.006	-	-	-	-
232	0.190	0.100	0.063	0.159	0.141	0.263	0.199	0.130	0.240	0.232	0.100	0.175
234	0.010	0.011	0.086	0.011	-	-	0.060	0.056	0.040	0.005	0.013	0.013
236	0.100	0.189	0.109	0.023	0.016	0.016	0.030	0.006	-	0.005	0.038	0.025
238	0.130	0.111	0.148	0.074	0.094	0.095	0.145	0.117	0.125	0.167	0.162	0.100
240	0.090	0.100	0.047	0.097	-	0.037	0.024	0.049	0.035	0.030	0.075	0.038
242	0.030	0.078	0.023	0.017	0.039	0.032	0.018	0.049	0.070	0.030	0.050	0.087
244	0.070	0.078	0.031	0.074	-	0.005	0.036	0.068	0.040	0.061	0.050	0.063
246	0.040	0.056	0.086	-	-	-	-	-	-	-	-	-
248	-	-	0.023	0.011	0.016	0.005	-	0.006	-	-	0.013	0.013
250	0.030	0.011	0.023	0.028	-	-	0.006	-	-	0.015	-	-
252	0.010	0.089	0.078	0.136	0.141	0.047	0.151	0.191	0.065	0.197	0.162	0.213
254	-	-	0.039	0.080	0.047	0.047	0.042	0.049	0.050	0.030	-	0.025
256	0.070	0.022	0.070	0.045	0.102	0.158	0.060	0.130	0.085	0.020	0.038	0.087
258	0.030	-	0.039	-	-	-	-	-	-	-	-	-
260	0.040	-	0.008	0.063	0.172	0.032	0.054	0.049	0.035	0.035	0.175	0.063
262	0.020	0.011	-	-	0.008	0.021	0.018	-	0.020	0.015	-	0.013
264	-	-	0.078	-	-	-	-	-	-	-	-	-
266	0.010	-	0.008	-	-	-	-	-	-	-	-	-
270	0.020	-	0.016	-	-	-	-	-	-	-	-	-
274	0.020	0.044	0.008	-	-	-	-	-	-	-	-	-
Total fish	50	45	64	88	64	95	83	81	100	99	40	40

Table A2. Allele frequencies for locus *Ssa197*

Allele	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
159	-	-	0.054	-	-	-	-	-	-	-	-	-
163	0.030	0.010	0.023	0.010	-	0.068	0.011	0.019	0.020	0.015	0.013	0.038
167	0.090	0.073	0.077	0.258	0.078	0.281	0.176	0.095	0.215	0.220	0.112	0.063
171	0.100	0.250	0.215	0.196	0.109	0.089	0.160	0.171	0.210	0.120	0.250	0.175
175	0.170	0.156	0.208	0.139	0.125	0.078	0.144	0.228	0.075	0.290	0.188	0.213
179	0.180	0.135	0.031	0.052	0.195	0.115	0.074	0.114	0.085	0.110	0.038	0.050
183	0.170	0.063	0.108	0.077	0.070	0.047	0.053	0.044	0.035	0.055	0.075	0.075
187	0.080	0.094	0.054	0.082	0.055	0.109	0.074	0.044	0.105	0.025	0.087	0.100
191	0.080	0.094	0.062	0.093	0.109	0.089	0.128	0.101	0.085	0.095	0.087	0.112
195	0.070	0.031	0.015	0.005	0.023	0.021	0.064	0.063	0.040	0.020	-	0.050
199	0.010	0.010	0.077	0.052	-	0.021	-	-	0.020	0.010	0.013	-
203	0.010	0.052	0.015	0.031	0.008	-	0.016	0.025	0.005	-	0.013	0.013
207	0.010	-	-	-	-	0.010	0.027	-	0.015	0.020	0.013	0.025
211	-	0.021	0.015	-	0.141	0.036	0.064	0.057	0.070	0.020	0.050	0.063
217	-	-	-	-	0.047	0.005	-	0.019	0.020	-	0.050	0.025
221	-	0.010	0.015	-	-	-	-	-	-	-	-	-
225	-	-	-	0.005	0.039	0.031	0.011	0.019	-	-	0.013	-
229	-	-	0.031	-	-	-	-	-	-	-	-	-
Total fish	50	48	65	97	64	96	94	79	100	100	40	40

Table A3. Allele frequencies for locus *Ssa202*

Allele	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
247	-	0.022	-	-	-	-	-	-	-	-	-	-
251	0.013	-	-	-	-	-	-	-	-	-	-	-
255	0.039	0.044	0.031	0.160	0.211	0.208	0.133	0.229	0.170	0.177	0.188	0.200
259	-	-	0.008	-	-	-	-	-	-	-	-	0.013
263	0.013	-	-	-	-	-	-	-	-	-	-	-
271	0.013	0.011	0.123	-	-	-	-	-	-	-	-	-
275	0.066	0.067	0.054	0.015	0.039	-	-	0.014	0.005	-	0.013	-
279	0.132	0.100	0.100	0.005	-	0.026	0.080	0.007	0.055	0.076	0.075	0.075
283	0.105	0.111	0.054	0.160	0.094	0.130	0.186	0.306	0.155	0.202	0.250	0.162
285	0.026	0.044	0.008	-	-	-	-	-	-	-	-	-
289	0.053	0.022	-	0.005	-	-	-	-	-	-	-	-
293	0.079	0.089	0.223	0.031	0.070	0.036	0.053	0.042	0.055	0.121	0.063	0.075
297	0.079	0.078	0.038	0.062	0.055	0.063	0.069	0.090	0.090	0.066	0.013	0.100
301	0.224	0.133	0.108	0.294	0.180	0.318	0.324	0.201	0.240	0.268	0.250	0.213
305	0.066	0.144	0.169	0.180	0.211	0.125	0.074	0.042	0.135	0.025	-	0.087
309	0.079	0.089	0.062	0.052	0.055	0.031	0.059	0.049	0.035	0.040	0.100	0.038
313	0.013	0.033	0.023	-	-	-	-	-	-	-	-	-
321	-	0.011	-	0.036	0.086	0.063	0.021	0.021	0.060	0.025	0.050	0.038
Total fish	38	45	65	97	64	96	94	72	100	99	40	40

Table A4. Allele frequencies for locus *cmrSs1.22*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
186	0.040	0.010	-	0.126	0.016	0.035	0.016	0.021	0.045	0.016	0.038	0.025
190	0.010	0.020	0.024	0.035	0.008	0.040	0.021	0.021	0.040	0.032	0.025	0.038
192	0.030	0.080	0.105	0.030	0.008	0.106	0.021	0.027	0.035	0.005	0.050	0.063
198	0.040	0.050	0.040	-	-	-	-	-	-	-	-	-
202	0.120	0.050	-	0.015	-	0.005	0.005	-	0.025	0.011	-	-
204	-	-	0.016	0.020	0.048	0.056	0.005	0.034	0.005	-	0.025	-
206	-	0.010	0.048	0.106	0.063	0.045	0.128	0.062	0.070	0.121	0.013	0.025
208	0.120	0.110	0.081	0.313	0.508	0.303	0.303	0.342	0.250	0.232	0.338	0.412
210	0.300	0.270	0.258	0.202	0.127	0.217	0.266	0.178	0.240	0.337	0.112	0.100
212	0.080	0.030	0.105	0.025	0.024	0.020	0.037	0.027	0.055	0.026	0.150	0.063
214	0.120	0.130	0.169	0.020	0.048	0.045	0.090	0.062	0.080	0.058	0.050	0.100
216	0.020	0.040	0.089	0.005	-	-	0.005	0.021	-	0.026	-	-
218	-	0.010	-	-	-	-	-	-	-	-	-	-
224	0.060	0.010	-	0.061	0.016	0.045	0.053	0.041	0.040	0.042	0.038	0.063
228	0.040	0.010	0.008	0.035	0.135	0.081	0.037	0.110	0.115	0.095	0.112	0.100
230	-	0.020	-	0.005	-	-	0.011	0.055	-	-	0.050	0.013
232	0.010	0.030	-	-	-	-	-	-	-	-	-	-
238	0.010	0.020	0.040	-	-	-	-	-	-	-	-	-
240	-	0.050	0.016	-	-	-	-	-	-	-	-	-
264	-	0.050	-	-	-	-	-	-	-	-	-	-
Total fish	50	50	62	99	63	99	94	73	100	95	40	40

Table A5. Allele frequencies for locus *cmrSs1.14*

Allele	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
136	0.120	0.206	0.180	0.242	0.210	0.182	0.333	0.307	0.224	0.352	0.321	0.338
148	-	0.020	0.016	-	-	-	-	-	-	-	-	-
156	0.010	-	-	-	-	-	-	-	0.005	-	-	-
158	0.130	0.147	0.125	0.131	0.113	0.146	0.057	0.053	0.153	0.138	0.115	0.087
160	0.480	0.451	0.500	0.525	0.653	0.515	0.489	0.600	0.408	0.429	0.551	0.488
162	0.200	0.127	0.180	0.101	0.024	0.157	0.115	0.040	0.209	0.082	0.013	0.087
164	0.060	0.020	-	-	-	-	0.006	-	-	-	-	-
166	-	0.020	-	-	-	-	-	-	-	-	-	-
170	-	0.010	-	-	-	-	-	-	-	-	-	-
Total fish	50	51	64	99	62	99	87	75	98	98	39	40

Table A6. Allele frequencies for locus *cmrSs1.10*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
218	0.920	0.951	0.913	1.000	0.992	0.934	0.957	0.963	0.985	0.896	0.950	0.950
220	0.040	0.020	0.024	-	-	0.051	0.032	0.037	0.005	0.094	0.025	0.025
222	0.040	0.029	0.063	-	0.008	0.015	0.011	-	0.010	0.010	0.025	0.025
Total fish	50	51	63	97	63	99	94	81	100	101	40	40

Table A7. Allele frequencies for locus *D30*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
230	-	-	0.023	-	-	-	-	-	-	-	-	-
232	-	0.039	-	-	-	-	-	-	-	-	-	-
234	0.323	0.225	0.273	0.198	0.403	0.278	0.228	0.222	0.335	0.161	0.188	0.225
236	-	0.039	0.055	0.031	-	-	0.006	0.006	0.005	0.005	0.013	-
238	-	0.010	0.016	-	-	-	-	-	-	-	-	-
240	0.635	0.627	0.625	0.677	0.597	0.603	0.700	0.710	0.530	0.774	0.688	0.650
242	0.042	0.049	0.008	0.094	-	0.119	0.067	0.062	0.130	0.059	0.112	0.125
244	-	0.010	-	-	-	-	-	-	-	-	-	-
Total fish	48	51	64	96	62	97	90	81	100	93	40	40

Table A8. Allele frequencies for locus *5.27*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
124	0.780	0.735	0.844	0.762	0.726	0.793	0.780	0.731	0.745	0.856	0.762	0.775
128	0.220	0.265	0.156	0.238	0.274	0.207	0.220	0.269	0.255	0.144	0.237	0.225
Total fish	50	51	64	82	62	99	91	80	100	101	40	40

Table A9. Allele frequencies for locus *F49*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
174	0.740	0.750	0.817	0.783	0.914	0.864	0.828	0.800	0.908	0.842	0.850	0.863
176	0.260	0.250	0.183	0.217	0.086	0.136	0.172	0.200	0.092	0.158	0.150	0.138
Total fish	50	50	63	99	64	99	90	80	98	98	40	40

Table A10. Allele frequencies for locus *20.19*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
78	0.170	0.167	0.047	0.150	0.095	0.025	0.161	0.154	0.045	0.119	0.162	0.188
84	0.350	0.275	0.359	0.230	0.294	0.182	0.211	0.160	0.256	0.342	0.175	0.175
96	0.060	0.039	0.055	0.160	0.111	0.157	0.100	0.210	0.170	0.094	0.237	0.225
98	0.400	0.500	0.539	0.460	0.500	0.636	0.528	0.475	0.528	0.446	0.425	0.412
100	0.020	0.020	-	-	-	-	-	-	-	-	-	-
Total fish	50	51	64	100	63	99	90	81	88	101	40	40

Table A11. Allele frequencies for locus *F43*

Alleles	Can 71	Can 72	Can 91	NSW 97	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas01A	Tas01B
103	0.010	-	-	-	-	-	-	-	-	-	-	-
109	0.050	0.140	0.081	0.039	0.071	0.071	0.016	0.019	0.026	0.015	0.025	-
111	0.010	0.010	-	-	-	-	-	-	-	-	-	-
115	0.410	0.330	0.460	0.702	0.667	0.505	0.410	0.403	0.422	0.460	0.438	0.387
121	0.210	0.140	0.258	0.230	0.143	0.362	0.234	0.273	0.333	0.213	0.138	0.225
123	0.010	0.010	-	-	-	-	-	-	-	-	-	-
125	0.240	0.250	0.202	0.028	0.119	0.061	0.261	0.279	0.198	0.257	0.350	0.325
129	0.010	0.080	-	-	-	-	0.074	0.026	0.021	0.054	0.050	0.063
131	0.030	0.020	-	-	-	-	0.005	-	-	-	-	-
133	0.010	-	-	-	-	-	-	-	-	-	-	-
135	0.010	0.020	-	-	-	-	-	-	-	-	-	-
Total fish	50	50	62	89	63	98	94	77	96	101	40	40

Table A12. Observed (H_o) and Hardy-Weinberg expected heterozygosity (H_e) estimates for the genetic variation samples at 11 microsatellite loci

	Canada			NSW		Tasmania								
Locus	**1971**	**1972**	**1970s**	**1991**	**1997**	**1992**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001A**	**2001B**	**2001**
Ssa 171														
Ho	0.920	0.911	0.916	0.984	0.909	0.844	0.811	0.952	0.840	0.850	0.859	0.925	0.925	0.925
He	0.922	0.911	0.919	0.929	0.911	0.880	0.865	0.893	0.903	0.886	0.857	0.899	0.895	0.900
P	0.215	0.724	0.507	0.000*	0.163	0.068	0.127	0.101	0.019	0.166	0.027	0.794	0.647	0.821
N	50	45	95	64	88	64	95	83	81	100	99	40	40	80
Nallele	21	17	23	20	15	12	14	15	15	13	15	13	15	15
Ssa 197														
Ho	0.860	0.813	0.837	0.862	0.876	0.984	0.885	0.862	0.911	0.870	0.880	0.950	0.900	0.925
He	0.882	0.873	0.883	0.882	0.852	0.891	0.868	0.888	0.879	0.874	0.832	0.872	0.893	0.880
P	0.032	0.100	0.029	0.000*	0.307	0.619	0.351	0.253	0.499	0.930	0.583	0.064	0.239	0.658
N	50	48	98	65	97	64	96	94	79	100	100	40	40	80
Nallele	12	13	14	15	12	12	14	13	13	14	12	14	13	15
Ssa 202														
Ho	0.895	0.844	0.867	0.923	0.856	0.906	0.823	0.851	0.833	0.840	0.879	0.900	0.900	0.900
He	0.900	0.922	0.911	0.879	0.825	0.857	0.819	0.823	0.804	0.857	0.837	0.840	0.867	0.855
P	0.270	0.097	0.540	0.000*	0.075	0.332	0.357	0.782	0.500	0.152	0.254	0.131	0.689	0.377
N	38	45	83	65	97	64	96	94	72	100	99	40	40	80
Nallele	15	15	17	13	11	9	9	9	10	10	9	9	10	11
cmrSs1.10														
Ho	0.160	0.098	0.129	0.175	-	0.016	0.111	0.085	0.074	0.030	0.208	0.100	0.100	0.100
He	0.170	0.114	0.132	0.178	-	0.032	0.134	0.093	0.084	0.040	0.198	0.121	0.121	0.109
P	1.000	1.000	1.000	1.000	-	1.000	0.012	1.000	1.000	1.000	0.675	1.000	1.000	1.000
N	50	51	101	63	97	63	99	94	81	100	101	40	40	80
Nallele	3	3	3	3	1	2	3	3	2	3	3	3	3	3

P = probabilities of conformance to Hardy-Weinberg equilibrium , * = significant after sequential Bonferroni correction for number of loci,
N = number of individuals, Nallele = number of alleles.

Table A.12. cont.

	Canada				NSW			Tasmania						
Locus	**1971**	**1972**	**1970s**	**1991**	**1997**	**1992**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001A**	**2001B**	**2001**
cmrSs1.14														
Ho	0.740	0.824	0.782	0.688	0.677	0.565	0.697	0.701	0.480	0.735	0.663	0.692	0.675	0.684
He	0.711	0.722	0.713	0.675	0.641	0.520	0.659	0.643	0.545	0.719	0.675	0.587	0.641	0.613
P	0.590	0.721	0.428	0.078	0.174	0.425	0.094	0.442	0.319	0.482	0.276	0.639	0.760	0.436
N	50	51	101	64	99	62	99	87	75	98	98	39	40	79
Nallele	6	8	8	5	4	4	4	5	4	5	4	4	4	4
cmrSs1.22														
Ho	0.880	0.900	0.890	0.919	0.889	0.730	0.879	0.809	0.822	0.920	0.853	0.850	0.800	0.825
He	0.864	0.885	0.873	0.869	0.829	0.713	0.835	0.810	0.833	0.850	0.806	0.837	0.805	0.816
P	0.877	0.232	0.871	0.002*	0.405	0.121	0.421	0.400	0.423	0.371	0.704	0.424	0.683	0.310
N	50	50	100	62	99	63	99	94	73	100	95	40	40	80
Nallele	14	19	19	13	14	11	12	14	13	12	12	12	11	12
D30														
Ho	0.563	0.588	0.576	0.578	0.510	0.419	0.443	0.456	0.494	0.570	0.312	0.575	0.550	0.563
He	0.495	0.560	0.530	0.535	0.496	0.485	0.555	0.464	0.454	0.598	0.383	0.503	0.534	0.508
P	0.343	0.983	0.849	0.582	0.008	0.296	0.076	0.179	0.414	0.034	0.172	0.456	1.000	0.532
N	48	51	99	64	96	62	97	90	81	100	93	40	40	80
Nallele	3	7	7	6	4	2	3	4	4	4	4	4	3	4
5.27														
Ho	0.280	0.333	0.307	0.313	0.329	0.355	0.293	0.396	0.363	0.330	0.228	0.475	0.400	0.438
He	0.362	0.407	0.377	0.266	0.374	0.413	0.330	0.353	0.405	0.382	0.256	0.386	0.372	0.367
P	0.209	0.289	0.102	0.337	0.374	0.352	0.352	0.224	0.565	0.191	0.426	0.082	0.652	0.057
N	50	51	101	64	82	62	99	91	80	100	101	40	40	80
Nallele	2	2	2	2	2	2	2	2	2	2	2	2	2	2

P = probabilities of conformance to Hardy-Weinberg equilibrium , * = significant after sequential Bonferroni correction for number of loci,
N = number of individuals, Nallele = number of alleles.

Table A.12. cont.

	Canada			NSW		Tasmania								
Locus	**1971**	**1972**	**1970s**	**1991**	**1997**	**1992**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001A**	**2001B**	**2001**
F49														
Ho	0.320	0.340	0.330	0.302	0.293	0.141	0.253	0.300	0.300	0.184	0.296	0.250	0.225	0.238
He	0.403	0.394	0.389	0.314	0.350	0.173	0.245	0.287	0.322	0.177	0.268	0.279	0.262	0.258
P	0.270	0.471	0.193	1.000	0.225	0.374	0.690	1.000	0.502	1.000	0.448	1.000	0.546	0.655
N	50	50	100	63	99	64	99	90	80	98	98	40	40	80
Nallele	2	2	2	2	2	2	2	2	2	2	2	2	2	2
20.19														
Ho	0.720	0.686	0.703	0.766	0.690	0.524	0.535	0.622	0.741	0.682	0.604	0.750	0.825	0.788
He	0.693	0.660	0.677	0.580	0.691	0.660	0.546	0.645	0.685	0.634	0.665	0.715	0.733	0.714
P	0.110	0.588	0.185	0.010	0.440	0.050	0.789	0.150	0.020	0.266	0.125	0.112	0.715	0.627
N	50	51	101	64	100	63	99	90	81	88	101	40	40	80
Nallele	5	5	5	4	4	4	4	4	4	4	4	4	4	4
F43														
Ho	0.760	0.800	0.780	0.371	0.202	0.317	0.286	0.723	0.584	0.656	0.614	0.675	0.700	0.688
He	0.734	0.791	0.765	0.680	0.463	0.532	0.612	0.707	0.692	0.674	0.677	0.672	0.698	0.684
P (A)	0.495	0.658	0.470	0.000*	0.000*	0.000*	0.000*	0.883	0.015	0.428	0.065	0.475	0.388	0.306
N	50	50	100	62	89	63	98	94	77	96	101	40	40	80
Nallele	11	9	11	4	4	4	4	6	5	5	5	5	4	5

P = probabilities of conformance to Hardy-Weinberg equilibrium , * = significant after sequential Bonferroni correction for number of loci,
N = number of individuals, Nallele = number of alleles.

Table A.13. Frequency of null alleles per genetic variation sample per locus (Values calculated by NULLTEST.)

	Ssa 171	*Ssa 197*	*Ssa 202*	*cmrSs1.10*	*cmrSs1.14*	*cmrSs1.22*	*D30*	*5.27*	*F49*	*20.19*	*F43*
Can 71	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Can 72	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Can 91	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.289
NSW 97	0.000	0.000	0.000	-[1.]	0.000	0.000	0.000	0.000	0.000	0.000	0.376
Tas 92	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.232
Tas 96	0.000	0.000	0.000	0.117	0.000	0.000	0.105	0.000	0.000	0.000	0.345
Tas 97	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 98	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 99	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 00	0.000	0.000	0.000	0.000	0.000	0.000	0.099	0.000	0.000	0.000	0.000
Tas 01A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 01B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

[1.] NSW 1997 was monomorphic at locus cmrSs1.10

Table A14. F_{st} values for each locus and overall loci across all samples.

Locus	F_{st} value
Ssa171	0.0216
Ssa197	0.0191
Ssa202	0.0227
cmrSs1.10	0.0188
cmSs1.14r	0.0177
cmrSs1.22	0.0293
D30	0.0174
5.27	0.0038
F49	0.0134
20.19	0.0186
F43	0.0365
Overall	0.0216

Table A15. Population pairwise F_{st} values. F_{st} values below diagonal, probability values above diagonal. * indicates values not significant with Bonferonni correction for multiple tests. Shaded cells indicate comparisons of year-classes three years apart.

	Can71	Can72	Can91	NSW97	Tas92	Tas96	Tas97	Tas98	Tas99	Tas00	Tas01a	Tas01b
Can 71	*****	0.283*	0.009*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Can 72	0.001	*****	0.010*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Can 91	0.007	0.006	*****	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NSW 97	0.022	0.026	0.033	*****	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 92	0.040	0.045	0.049	0.020	*****	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tas 96	0.028	0.035	0.036	0.012	0.026	*****	0.000	0.000	0.001*	0.000	0.000	0.000
Tas 97	0.012	0.014	0.025	0.011	0.029	0.015	*****	0.060*	0.000	0.002*	0.006*	0.122*
Tas 98	0.024	0.020	0.033	0.015	0.026	0.023	0.002	*****	0.000	0.000	0.316*	0.849*
Tas 99	0.015	0.020	0.026	0.013	0.026	0.006	0.007	0.015	*****	0.000	0.000	0.000
Tas 00	0.018	0.023	0.028	0.019	0.043	0.026	0.005	0.014	0.019	*****	0.000	0.000
Tas 01a	0.024	0.021	0.035	0.017	0.028	0.031	0.007	0.000	0.020	0.016	*****	0.797*
Tas 01b	0.020	0.019	0.031	0.016	0.023	0.024	0.002	-0.003	0.011	0.014	-0.003	*****

Table A16. Estimates of allele-frequency variance (F_K) between the Canada 1971/72 sample representing the progenitor population for the different year-class samples for the derived Australian and Canadian populations The weighted average multilocus estimate of F_K (Wt. av. F_K) was calculated as the average over loci, weighted by the number of independent alleles per locus.

Loci	Can1991	NSW	Tas 92	Tas 96	Tas 97	Tas 98	Tas 99	Tas 00	Tas 01
Ssa 171	0.0334	0.0447	0.0724	0.0551	0.0444	0.0472	0.0495	0.0459	0.0390
Ssa 197	0.0374	0.0360	0.0411	0.0448	0.0269	0.0213	0.0374	0.0363	0.0252
Ssa 202	0.0369	0.0544	0.0603	0.0628	0.0430	0.0686	0.0471	0.0531	0.0466
cmrSs 1.14	0.0162	0.0259	0.0533	0.0185	0.0390	0.0609	0.0212	0.0418	0.0486
cmrSs 1.22	0.0373	0.0632	0.0878	0.0570	0.0486	0.0598	0.0434	0.0532	0.0621
cmrSs 1.10	0.0093	0.0665	0.0480	0.0132	0.0124	0.0356	0.0325	0.0475	0.0020
D30	0.0276	0.0210	0.0407	0.0280	0.0165	0.0168	0.0317	0.0282	0.0266
5.27	0.0466	0.0001	0.0051	0.0073	0.0029	0.0036	0.0008	0.0615	0.0007
F49	0.0301	0.0080	0.1922	0.0881	0.0406	0.0171	0.1771	0.0566	0.0761
20.19	0.0500	0.0460	0.0336	0.1242	0.0315	0.0849	0.0847	0.0219	0.0888
F43	0.0292	0.0842	0.0501	0.0630	0.0265	0.0302	0.0351	0.0286	0.0303
Wt. av. F_K	0.0334	0.0409	0.0621	0.0533	0.0373	0.0472	0.0436	0.0434	0.0428

III. MHC allele sequences

Figure A.1. Nucleotide sequence of MHC class Iα1 alleles.
Nucleotide sequence of the three observed Atlantic salmon MHC class Iα1 alleles. A, T, C, G represent the nucleotides adenine, thymine, cytosine and guanine respectively. *Dashes* indicate identity with the top sequence, *asterisks* indicate substitutions. *Mbo*I restriction sites are indicated in *shaded box* and sense and reverse complement antisense primer recognition sequences are *underscored*.

```
           1                                                50
Iα1*01     TTCT*****GCGTCTTCTAAAGTTACCAACTTCCCAGAGTTTATGGTTGT
Iα1*02     ----ACACT---------------C-----------------G-------
Iα1*03     ----ACACT-----------------------------------------

           511                                             100
Iα1*01     GGGGATGGTGGATGGTGTTCAGATTGATCACTATGACAGCAACATCCAGA
Iα1*02     --------------------------------------------G-----
Iα1*03     ------------------------G-T-----------------G-----

            101                                            150
Iα1*01     AAGTAGTGCCCGAACAGGTCTGGATGAACAAGCAGACAGACGCAGAGTAC
Iα1*02     G-A-G------A------A-------------------------------
Iα1*03     G-A-G------A------A-------------------------------

           151                                             200
Iα1*01     TGGGAGAGGGAGACAGGGATTGCCTTTGATTCCCAGCAGGTTTTCAAAGA
Iα1*02     --------------------------------------------------
Iα1*03     --------------------------------------------------

           201                 222
Iα1*01     CGAGGTCAATATTCTAAAGCAG
Iα1*02     ----------------------
Iα1*03     ----------------------
```

Figure A.2. Nucleotide sequence of MHC class Iα2 alleles.
Nucleotide sequence of the two observed Atlantic salmon MHC class Iα2 alleles A, T, C, G represent the nucleotides adenine, thymine, cytosine and guanine respectively. *Dashes* indicate identity with the top sequence, *asterisks* indicate substitutions. *Mbo*I, *Dde*I restriction sites are indicated in *shaded boxes* (darker and lighter shade respectively), and sense and reverse complement antisense primer recognition sequences are *underscored.*

```
           1                                                50
Iα2*1      GATGTGAGTGGGATGATGAGGCTGGACAGACAGGTGGGTTTACACAGCAT
Iα2*2      ----------------------A---TT-----AA------CAC--TAT-

           51                                              100
Iα2*1      GGTTATGATGGGGAGGACTTCCTTGGGTACGACATGAAGACGTTCACTTG
Iα2*2      --A--------AC-A---C------TA-TT---C------GAGCG--A--

           101                                             150
Iα2*1      GATCGCTCCTAAACAGCAAGCTGAGATCACTCGGCTCAAGTGGAACCATG
Iα2*2      ---T--CT-AGTC-C---G---CTCCA-T-CAAAA-G------G-AGGG-

           151                                             200
Iα2*1      ACCAAGCTGGACTGGCATTCCGGCATAACTACCTCACTCAGATCTGCATT
Iα2*2      ---CGT--A-CA-T-AGAG-GA-A-ACG---------C----A-------

           201            216
Iα2*1      GAGTGGCTGAAGAAGC
Iα2*2      ----------------
```

Figure A.3. Nucleotide sequence of three MHC class IIα1 alleles.
Nucleotide sequence of three Atlantic salmon MHC class IIα1 alleles. A, T, C, G represent the nucleotides adenine, thymine, cytosine and guanine respectively. Dashes indicate identity with the top sequence. AluI and RsaI restriction sites are indicated in shaded boxes (darker and lighter shade respectively). Sense and the reverse complement antisense primer recognition sequences are underscored.

```
           1                                                50
IIα1*1     CCAGATGTAATATACTAACAGTGTGTTTTATTGGGTTTCTTTTCTCAGTT
IIα1*2     --------------------------------------------------
IIα1*3     --------------------------------------------------

           51                                              100
IIα1*1     CTGCATATTGATTTATATATTAGTGGATGCAGTGATTCAGATGGACTGGA
IIα1*2     ---------------GC-----C---------------------------
IIα1*3     ---------------C------C----------------------G----

           101                                             150
IIα1*1     CATGTATGGACTGGATGGGGAAGAGATGTGGTACGCAGACTTCAACAAGG
IIα1*2     -------------------------------------------------C
IIα1*3     A-------------------------------------------------

           151                                             200
IIα1*1     GGGAGGGAGTGGTGGCCCTGCCTCCGTTTGCAGATCCATTTACCTTCCCT
IIα1*2     A-------------------------------------------------
IIα1*3     -----------A--C-A----------------------------A----

           201                                             250
IIα1*1     GGATTTTATGAAGGGGCTGTAGGTAACCAGGGGGTATGCAAAGCAAACCT
IIα1*2     ------------CA------------------------------------
IIα1*3     ---GC-------CA-------------------A----------------

           251                                      294
IIα1*1     GGCCGTAAACATAAAAGCTTACAAGAACCCAGAAGAGAAAATAG
IIα1*2     ----AA-TGT---------------------A------------
IIα1*3     --------------------------------------------
```

Figure A.4. Nucleotide sequence of six MHC class IIβ1 alleles.
Nucleotide sequence of six Atlantic salmon MHC class IIβ1 alleles. A, T, C, G represent the nucleotides adenine, thymine, cytosine and guanine respectively. Dashes indicate identity with the top sequence. AluI and RsaI restriction sites are indicated in shaded boxes (darker and lighter shade respectively). Sense and the reverse complement antisense primer recognition sequences are underscored. A, B, C, D indicate the alleles that form the composite alleles A, B, C, and D respectively.

```
               1                                                50
IIβ1*1[C]      TAATCTGACAAAACAATGATCTGTATTATGTTTTCCTTCCAGATGGATAT
IIβ1*2[A]      --------------------------------------------------
IIβ1*3[C]      -T------------------------------------------------
IIβ1*4[D]      --------------------------------------------------
IIβ1*5[B]      --------------------------------------------------
IIβ1*6[A]      --------------------------------------------------

               51                                              100
IIβ1*1         TTTTATCAGAGGGTGTCAGAGTGCCGATACTCCTCAAAGGACCTGCAGGG
IIβ1*2         --------------------------------------------------
IIβ1*3         --------T-T-A--AG-C-------------------------------
IIβ1*4         --------T-T-A--AG-C-------------------------------
IIβ1*5         ---G-A---GTT---AGTC---------T---------------------
IIβ1*6         --------T-T-A--AG-C-------------------------------

               101                                             150
IIβ1*1         TATAGAGTTTATAGACTCTTATGTTTTCAATAAGGCTGAATATGTCAGAT
IIβ1*2         -------------------------------C--------A--A------
IIβ1*3         ---------G--T-C---G------------C--------A--A------
IIβ1*4         ---------G--TAC---G------------C-----------A------
IIβ1*5         --C-----------------------------------------------
IIβ1*6         ---------G--TAC---G------------C--------A--A------

               151                                             200
IIβ1*1         TCAACAGCACTGTGGGGAAGTATGTTGGATACACTGAGTATGGAGTGAAG
IIβ1*2         ---------------------T----------------CTG---------
IIβ1*3         ---------------------T----------------C-----------
IIβ1*4         --------------------------------------------------
IIβ1*5         --------------------------------------CTG---------
IIβ1*6         ---------------------T----------------CTG---------

               201                                             250
IIβ1*1         AATGCAGAAGCCTGGAACAAAGGTCCTGAGCTGGCTGGAGAGCTAGGGGA
IIβ1*2         --------------------------------------------------
IIβ1*3         --------------------------------------------------
IIβ1*4         -------------------------------------------------T
IIβ1*5         -------------------------------------T------------
IIβ1*6         --------------------------------------------------

               251                                             300
IIβ1*1         GCTGGAGCGTGTCTGTAAGCATAACGCTCCTATCGACTACAGCGCCATAC
IIβ1*2         ----------T------------------A-C--C--------A------
IIβ1*3         ----------------------------GA-C--C--------A------
IIβ1*4         ----------TA----------------------T---------------
IIβ1*5         ----------T---------------------------------------
IIβ1*6         ----------TA----------------------T--------A------

               301          315
IIβ1*1         TGGACAAGACAGGTG
IIβ1*2         ---------------
IIβ1*3         ---------------
IIβ1*4         ---------------
IIβ1*5         ---------------
IIβ1*6         -----------A---
```