Molecular assessment of resistance to amoebic gill disease

James Watkin Wynne

Bachelor of Aquaculture (Hons)

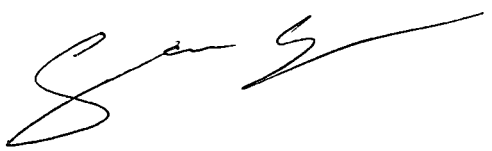
Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Tasmania

October, 2008

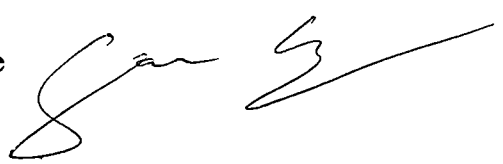
Statement of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of the candidate's knowledge and belief no material previously published or written by another person except where due acknowledgment is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed, James Wynne  15 October 2008

Statement of Authority

The thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Signed, James Wynne  15 October 2008

Abstract

Amoebic gill disease (AGD) is the most significant health issue affecting the culture of Atlantic salmon (*Salmo salar* L.) in Tasmania, Australia. Recent research has suggested that heritable variation in AGD resistance exists within the Tasmanian Atlantic salmon population. Subsequently, enhancing this resistance through selective breeding has become a major research focus in Tasmania. The mechanisms controlling this commercially important trait remain poorly understood. To this end, an investigation of the molecular mechanisms controlling AGD resistance was conducted.

Due to their high polymorphism and important immune function genes of the major histocompatibility complex (MHC) – known as the MH genes in Atlantic salmon – represent some of the best candidate loci with a possible influence upon AGD resistance. With this in mind, the amount of MH variation and its association with resistance to AGD was investigated. In contrast to what has been previously reported at non-coding microsatellite loci, a high level of MH class II diversity has been maintained in the Australian Atlantic salmon population compared to the ancestral population. The use of an AGD challenge test with subsequent MH genotyping demonstrated that the presence of one MH class II alpha allele known as *Sasa-DAA-3UTR* 239 was significantly associated with reduced disease severity. Individuals containing a copy of this allele had 4.0% less gill filaments infected by AGD compared to individuals without this allele.

Next utilising a cDNA microarray with real-time PCR verification the transcriptional changes associated with AGD and AGD resistance were investigated. Comparing the gene expression profiles within the gill, liver and anterior kidney between naïve and AGD affected (at 19 days post inoculation) Atlantic salmon suggests the host response to AGD upon acute first infection is largely suppressive and localised to the site of infection, the gill. Next, the gill transcriptome response between Atlantic salmon deemed putatively resistant and putatively susceptible to AGD following chronic natural infection was investigated. Results suggested that compared to the susceptible individuals, Atlantic salmon resistant to AGD demonstrate an up-regulation of adaptive immune genes and negative regulators of the cycle cell. Further characterisation of the full length mRNA sequence and expression distribution of one unknown transcript which was significantly up-regulated in both previous microarray experiments was investigated.

This research has provided the first molecular assessment of resistance to AGD in Atlantic salmon. The implications of this research in terms of the understanding of the molecular mechanisms of AGD resistance and the ultimate development of genetic markers linked to resistance will be considered.

Acknowledgements

In years to come when I think about my PhD project, I won't remember the difficultly I had with a particular PCR, or even how many papers I had published. Rather I'll remember the generosity and kindness of the many people who helped me through what is often a difficult and stressful process.

First and foremost I must thank my supervisors; Dr Mathew Cook, Associate Professor Barbara Nowak and Dr Nick Elliott. The guidance and encouragement you provided during my candidature was tremendous, and I will always be sincerely grateful. Thanks must also go to researchers at CSIRO Mathematical and Information Sciences, in particular Ms Maree O'Sullivan for her terrific advice and support concerning microarray experimentation.

Molecular biology is rarely straight forward. Often students rely on the advice and support from more experienced researchers, and I was no different. I would therefore like to thank the laboratory staff in the Genomics and Taxonomy group at CSIRO. In particular, I would to thank the laboratory manager Dr Sharon Appleyard as well as Mrs Bronwyn Holmes, Mrs Natasha Carr, Mr Giles Campbell, Dr Jawahar Patil, Dr Stan Robert and Dr Bob Ward. I would also like to thank Dr Melony Sellars from the Cleveland site of CSIRO.

I would also like to acknowledge staff at the School of Aquaculture, University of Tasmania, in particular Mr Michael Attard, Dr Melanie Leef, Dr Phil Crosbie and Dr Mark Adams for their wonderful assistance in fish husbandry and histopathology. I would also like to thank Mr Adam Smolenski from the University of Tasmania.

Thanks must also go to the student contingent at CSIRO. Particularly, Mr Richard Taylor for all the excellent advice and assistance he provided during my field work, and all the past and present students in the block 4 student room for the terrific support, encouragement and a good dose of humour they gave each and every day.

A special thank you must go to the late Professor René Stet. During my candidature Prof Stet hosted a laboratory visit for me to conduct MHC research on Atlantic herring at the University of Aberdeen. I will always be sincerely grateful for this opportunity and for the support and guidance he offered me.

I would also like to thank participants of the TRAITS/SGP projects, in particular Mr Richard Talbot, Dr Samuel Martin and Dr Bjørn Høyheim for supplying and advising on the use of the microarrays.

I would also like to acknowledge the research and industry partners of the Tasmanian Atlantic salmon selective breeding program, in particular, Dr Peter Kube, for supplying animals and general advice on experimentation.

My PhD project was jointly funded by CSIRO Food Futures Flagship and the Aquafin CRC. I would like to thank the Aquafin CRC, and in particular Dr Peter Montague and Ms Emily Downes, for the terrific training and travel support that has been provided to me as an Aquafin CRC student.

I would like to thank the communications unit at CSIRO for the excellent assistance throughout my PhD for posters, presentations and editorial advice.

PhD projects are a 24 hour a day, seven days a week commitment. This means the support and encouragement of family and friends is essential. I have been extremely fortunate to have a family who have always encouraged my studies. I am sincerely indebted to my Mum and Dad for such wonderful support. Finally I would like to thank my best friend Ruth Thomas, whose kindness and generosity has kept me sane during the last three and a half years.

Contents

List of Figures

List of Tables

List of abbreviations

AGD	amoebic gill disease
APP	acute phase proteins
APR	acute phase response
AR	allelic richness
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pairs
C/EBPβ	CCAAT/enhancer-binding protein beta
C1q	C1q-like adipose specific protein
CD	conserved domain
CD4	MHC class II receptor
CD8	MHC class I receptor
CDC14	CDC14 cell division cycle 14 homolog a
CDK	cyclin-dependent kinase
CDKI	p13 cyclin dependent kinase inhibitor
CP	connecting peptide
CT	cycle number at which the signal reaches the thresholds
CTAB	hexadecyltrimethylammonium bromide
CTL	c type lectin receptor A
CVS	central venous sinus
Cy3	cyanine dye 3
Cy5	cyanine dye 5
CYCB	cyclin B1
CYT	cytoplasmic
DEPC	diethylpyrocarbonate
DIG	digoxigenin
DPI	days post inoculation
DRTP	differentially regulated trout protein 1
DRTP1	differentially regulated trout protein 1
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EFAB	elongation factor IAB
EST	expressed sequence tag
GADDβ	growth arrest and DNA-damage-inducible beta
GO	Gene Ontology
GPR	GenePix results
GPX	glutathione peroxidase
GST	glutathione S-transferase
H_e	expected heterozygosity
HIS	histidine

H_o	observed heterozygosity
IFN-γ	interferon gamma
Ig	Immunoglobulin light chain
IL-1β	interleukin-1 beta
IL-6	interleukin-6
iNOS	nitric oxide synthase
IPD	Immuno Polymorphism Database
ISA	infectious salmon anaemia
kDa	kilodalton
LIP	lipase H
LNA	locked nucleic acid
LP	leader peptide
LPS	lipopolysaccharide
MAS	marker assisted selection
MD	Marek's disease
MH	major histocompatibility
MH Ic	MHC class II invariant chain-like protein 2
MHC	major histocompatibility complex
MHC	major histocompatibility complex
NADP	NADP transhydrogenase
NF-κB	nuclear factor kappa beta
nr	non-redundant
NSW	New South Wales
NTC	no template control
Onmy-UBA	rainbow trout major histocompatibility class I chain
ORF	open reading frame
PBR	peptide-binding regions
PBS	phosphate buffered saline
PCATH	procathepsin
PCNA	proliferating cell nuclear antigen
PMT	photo multiplier tube
PSMB	immunoproteasome subunits
qPCR	real-time quantitative RT-PCR
QTL	quantitative trait loci
RACE	rapid amplification of cDNA ends
RNAi	RNA interference
RT-PCR	reverse transcription PCR
SAA	serum amyloid A
Sasa-DAA	Atlantic salmon MH class II alpha chain
Sasa-DAA-3UTR	Atlantic salmon MH class II alpha chain 3' minisatellite
Sasa-DAB	Atlantic salmon MH class II beta chain
Sasa-TAP2B	Atlantic salmon transporter associated with antigen presentation
Sasa-UBA	Atlantic salmon MH class I chain
Sasa-UBA-3UTR	Atlantic salmon MH class I chain 3' microsatellite
SDS	sodium dodecyl sulfate

SE	standard error
SNP	single nucleotide polymorphism
SSC	sodium chloride sodium citrate
TAE	tris-acetate-EDTA
TBS	tris buffered saline
TDX	Thioredoxin
TE	tris-EDTA
TFGβ	transforming growth factor beta
TM	transmembrane
TNF- α	tumor necrosis factor alpha
TSC-22	TFGβ stimulated clone-22
UT	untranslated

Chapter 1: General introduction

Resistance to amoebic gill disease (AGD) is a trait of high commercial importance. Exploiting this resistance through selection is a major breeding goal in Tasmania. The ability to accurately select desirable broods could be significantly enhanced through a better understanding of the mechanisms of resistance, and the ultimate development of marker assisted selection for AGD resistance. To this end, using a candidate gene and transcriptome profiling approach I investigate the molecular mechanisms influencing/controlling AGD resistance.

1.1 Atlantic salmon aquaculture

Aquaculture is one of the world's fastest growing industries (Tidemand-Johnannessen 1999). Initially intended to substitute for shortfalls in wild fisheries, aquaculture has grown rapidly, and is now considered to be the future for seafood production. The culture of Atlantic salmon (*Salmo salar*, L.) is a leading example and is undertaken in many countries, with the majority of production based in Chile, Norway, Scotland and Canada (FAO Fisheries Department, 2006).

In Australia, Atlantic salmon have been commercially cultured since the 1980s with the majority of production in Tasmania. Atlantic salmon are not native to Australia and were introduced in three individual imports of 100,000 ova each from the River Philip, Nova Scotia, Canada, in the mid-1960s (Ward et al. 1994). This founder population was first maintained in a freshwater landlocked hatchery located in New South Wales (Ward et al. 1994). During this time the population underwent a moderate genetic bottleneck

event and, despite poor hatchery records, anecdotal reports suggest breeding numbers reduced significantly. In the mid-1980s three introductions from this founder population were made into Tasmania (Ward et al. 1994). No additional stock introductions have occurred in Tasmania or New South Wales following these initial importations.

In Tasmania Atlantic salmon aquaculture is a major local industry. Annual production exceeds 20,000 tonnes with an estimated net worth of AUD$250M (Australian Bureau of Agriculture and Resource Economics; www.abareconomics.com). In 2003 this industry employed over 3,000 people, making it the largest employer of any aquaculture sector in Australia (Carington Smith and Wadley 2003). Despite continuing growth of Atlantic salmon aquaculture in Tasmania, this industry still faces a number of limitations. Disease, and in particular, a condition known as AGD, is the most significant limitation.

1.2 Amoebic gill disease

1.2.1 The aetiological agent

AGD has been recorded in the Tasmanian Atlantic salmon population since the commencement of its marine culture in the mid-1980s (Munday 1986). Initially, it was believed to be caused by the amphizoic parasite *Neoparamoeba pemaquidensis* (Page) (Kent et al. 1988) and later also *Neoparamoeba branchiphila* Dyková, Nowak, Crosbie, Fiala, Pecková, Adams, Macháčková et Dvořáková (Dyková et al. 2005). However,

more recent research has shown that the most predominant species associated with AGD pathology is a newly described species designated *Neoparamoeba perurans* Young, Crosbie, Adams, Nowak et Morrison (Young et al. 2007; Young et al. 2008c). At present no sub-strains of *Neoparamoeba perurans* have been identified.

1.2.2 Affected species and distribution

The salmonids Atlantic salmon and rainbow trout, *Oncorhynchus mykiss* (Walbaum), appear to be the species most susceptible to AGD (Munday et al. 2001), but this disease can also affect other teleosts such as turbot *Scophthalmus maximus* (L.) (Dyková et al. 1999), coho salmon *Oncorhynchus kisutch* (Walbaum) (Kent et al. 1988) and seabass *Dicentrarchus labrax* (L.) (Dyková et al. 2000). The majority of wild teleost species caught in and around commercial salmon farms in Tasmania do not appear to be susceptible to AGD (Douglas-Helders et al. 2002). One exception was a single specimen of blue warehou, *Seriolella brama* (Günther) which presented significant AGD pathology along with the presence of *Neoparamoeba* spp. (Adams et al. 2008).

While AGD is essentially endemic to Tasmania and certain locations in the USA (Kent et al. 1988), sporadic cases have also been recorded in Ireland, France, Spain, New Zealand, Norway, Chile and Scotland (Dyková et al. 1995; Rodger and McArdle 1996; Findlay and Munday 1998; Steinum et al. 2008; Young et al. 2008c). In Tasmania AGD is most prevalent in the summer months (December to February) coinciding with increasing water temperature.

1.2.3 Signs and symptoms of AGD

Individuals affected by AGD display lethargy, respiratory distress, rapid opercular movement, and ultimately, if not treated, death will result (Munday et al. 1990). Gross examination of AGD affected gill tissue reveals raised white/grey multifocal mucoid lesions on the gill surface (Munday et al. 1990). Microscopically, gross lesions coincide with extensive morphological and cellular alterations in the gill epithelium. AGD is initiated by the attachment of trophozoites to healthy gill epithelial tissue. Following this attachment, localised tissue changes occur including epithelial desquamation and a moderate infiltration of leucocytes within the central venous sinus (Adams and Nowak 2004a). As the disease progresses, more pronounced cellular changes can be seen including lamellar fusion, extensive hyperplasia/hypertrophy, oedema and interlamellae vesicle formation (Adams and Nowak 2003). In such advanced lesions, an infiltration of leucocytes from the central venous sinus into the hyperplastic tissue is observed (Adams and Nowak 2004a). In some instances, leucocytes are found in close proximity to amoeba within the hyperplastic lesions (Bridle et al. 2003).

1.2.4 AGD diagnosis and quantification of disease severity

While a number of diagnostic tests for AGD have been developed (Zilberg et al. 1999; Douglas-Helders et al. 2001; Dyková and Novoa 2001; Young et al. 2008b), clinical diagnosis within commercial salmon farming operations is principally based upon the visual inspection of gross gill lesions. The advantage of this system is that the severity of infection is also estimated simply by quantifying the number of gross lesions on the

gill surface. A score between 0 and 5 is then assigned to each individual, where 0 represents no lesions and 5 represents numerous lesions (Table 1.1). A strong

Infection Level	Score	Gross Description
Clear	0	No sign of infection and healthy red colour
Very Light	1	1 *Neoparamoeba* spp. spot, light scarring or undefined necrotic streaking
Light	2	2-3 spots/small mucus patch
Moderate	3	Established thickened mucus patch or spot groupings up to 20% of gill area
Heavy	4	Established lesions covering up to 50% of gill area
Extreme	5	Extensive lesions covering most of the gill surface

Table 1.1. Gross gill scoring system utilised by the Tasmanian Atlantic salmon farming company, Tassal Group Limited, to estimate the severity of infection by AGD. This table was provided by R. Taylor[1] and D. Cameron[2].

[1] Mr Richard Taylor, University of Tasmania, CSIRO Food Futures Flagship
[2] Mr David Cameron, Tassal Group Limited

relationship between the severity of AGD, estimated by this gross gill score, and fish survival time has been observed (R. Taylor, pers. comm). This non-destructive gross gill scoring method is employed by commercial salmon growers to estimate disease severity and then perform accordant treatment and management strategies.

AGD diagnosis and the severity of infection can also be determined through histopathology (Munday et al. 2001). Diagnosis is based on the presence of both the causative agent, *Neoparamoeba* spp., and associated gill pathology (Adams and Nowak 2004a; Young et al. 2007; Young et al. 2008c). The severity of infection is quantified by counting the proportion of primary lamellae displaying hyperplastic lesions. When destructive sampling is performed (generally only in research operations), histopathology is preferable to gross gill scoring because both the presence of the pathogen and the associated pathology can be observed. Moderate to good agreement between gross gill score and histology based gill score has been reported (Adams et al. 2004; Taylor et al. 2007).

1.3 Host response

The host response, and in particular the immune response, to AGD has been the subject of a number of studies (Akhlaghi et al. 1996; Gross et al. 2004a; Gross et al. 2005; Bridle et al. 2006a; Bridle et al. 2006b; Morrison et al. 2006b; Morrison et al. 2007; Young et al. 2008a). Interestingly, despite the extensive morphological changes associated with AGD, this disease appears to stimulate only a modest local and systemic immune response.

1.3.1 The systemic immune response

Previous studies have shown that total plasma protein and lysozyme levels do not increase following AGD infection in Atlantic salmon (Gross et al. 2005). Furthermore, respiratory burst by phagocytes isolated from the anterior kidney of AGD affected Atlantic salmon is suppressed during advanced stages of disease (Gross et al. 2004b; Gross et al. 2005). Conversely, an increase in the chemotactic and phagocytosis function – albeit variable – of phagocytes derived from AGD affected Atlantic salmon has also been reported (Gross et al. 2005). Respiratory burst of AGD affected derived phagocytes can be enhanced *in vitro* through stimulation with beta (β)-glucans (Bridle et al. 2005). Dietary application of β-glucans, however, does not increase protection against AGD (Bridle et al. 2005).

The adaptive immune response to AGD has also been investigated. While the presence of anti-*Neoparamoeba* spp. antibodies in the serum of both Atlantic salmon and rainbow trout has been demonstrated, evidence of their protective properties remains inconclusive (Findlay et al. 1995; Akhlaghi et al. 1996; Findlay and Munday 1998; Zilberg and Munday 2001; Gross et al. 2004a; Vincent et al. 2006). The presence of anti-*Neoparamoeba* spp. antibodies has been demonstrated in both laboratory induced AGD challenge trials (Vincent et al. 2006) and under commercial culture conditions (Gross et al. 2004a). In the commercially cultured Atlantic salmon, the antibody activity did not appear to be influenced by the infection history or the number of *Neoparamoeba* spp. present on the gills (Gross et al. 2004a). Only one study has found evidence to suggest anti-*Neoparamoeba* spp. antibodies may be protective against AGD in Atlantic

salmon (Vincent et al. 2006). In this study, previously exposed and naïve Atlantic salmon were re-infected with AGD and their serum anti-*Neoparamoeba* spp. antibodies measured in surviving fish. The authors reported that fish previously infected with AGD showed a significant increase in resistance to subsequent re-infection (Vincent et al. 2006). This resistance was believed to be associated with circulating anti-*Neoparamoeba* spp. antibodies which were present in 50% of the surviving individuals previously exposed to AGD, but not detected in any individuals exposed only once (Vincent et al. 2006).

1.3.2 Localised immuno-inflammatory response

Gene expression of candidate immune related genes has been investigated in both rainbow trout (Bridle et al. 2006b) and Atlantic salmon (Bridle et al. 2006a; Morrison et al. 2007; Young et al. 2008a) and, for the most part, the results from these studies agree with the hypothesis that AGD affected salmonids display only a minor immune response. Many immune related genes do not demonstrate increased expression following infection by AGD. Examples of such genes include serum amyloid A, serum amyloid P-like pentraxin (Bridle et al. 2006a), transforming growth factor beta, cyclooxygenase 2, major histocompatibility (MH)[3] (*Sasa-DAB*), T-cell receptor beta (Bridle et al. 2006b) and tumour necrosis factor alpha (Bridle et al. 2006a; Morrison et al. 2007). AGD affected rainbow trout, however, demonstrate an up-regulation of the pro-inflammatory cytokine interleukin-1 beta (IL-1β) and inducible nitric oxide

[3] In regards to Atlantic salmon (*Salmo salar*) the major histocompatibility complex (MHC) genes are designated the "MH" genes throughout this thesis in line with Stet et al. (2002). Where the acronym MHC is used in this thesis it refers to species other than Atlantic salmon.

synthase (iNOS) in the gills at both 7 and 14 days post inoculation (DPI) (Bridle et al. 2006b). Atlantic salmon also display a significant up-regulation of IL-1β at 14 DPI within the gill but there are no changes to iNOS (Bridle et al. 2006a; Morrison et al. 2007). Moreover, the up-regulation of IL-1β appears to be lesion-restricted (Morrison et al. 2007). This research suggests the host immune response to AGD is not only minor, but also extremely localised.

It has become evident that processes other than the immune response are important in AGD pathogenesis. Indeed, transcriptome profiling of both the early and intermediate stages of AGD within the gill has provided valuable insight into the host response. Utilising a 16k salmonid specific microarray produced by the Genomic Research on Atlantic Salmon Project (GRASP), Morrison et al. (2006a) examined the transcriptome response of the gill to AGD at four time points up to 8 DPI. While a number of biologically significant findings were observed, the down-regulation of tumor suppressor p53 and concurrent up-regulation of Atlantic salmon anterior gradient-2 (an inhibitor of p53) and proliferating cell nuclear antigen suggests that genes involved in cellular proliferation are important in AGD pathogenesis.

1.4 Treatment

At present the only commercially feasible treatment for AGD is a freshwater bath for 2-6 h (Clark and Nowak 1999; Munday et al. 2001). Bathing in freshwater significantly reduces the prevalence of gross mucoid lesions and the number of viable *Neoparamoeba* spp. on the gill (Parsons et al. 2001). Generally, this practice is repeated

approximately six times per sea-cage during the Tasmanian industry's two year production cycle (R. Taylor pers. comm). Due to the high labour and financial costs associated with freshwater bathing, this practice, at least in its current design, is not considered a practicable long term solution to AGD. Studies to improve the effectiveness of freshwater bathing are on-going. Water hardness appears to be an important contributing factor to the effectiveness of freshwater bathing, with soft freshwater being more efficacious for the alleviation of AGD compared to hard freshwater (Roberts and Powell 2003).

1.5 Genetic resistance to AGD

Resistance to AGD is undoubtedly a complex trait. Typically, resistance is defined as reduced gill pathology (quantified through either gross gill score or histopathology gill score) and/or increased survival time following infection (with no subsequent treatment). In commercial culture operations, however, resistance may be variously defined as time to first treatment, time between treatments or even the ability to tolerate AGD by demonstrating growth performance during infection. For the purpose of this thesis, resistance is defined as reduced gill pathology unless otherwise stated.

AGD resistance in Atlantic salmon was first documented by Bridle et al. (2005) who inadvertently discovered that a subpopulation of individuals were able to survive for an extended period under experimental challenge conditions. More recently, experiments based on Atlantic salmon family lines have suggested AGD resistance is significantly genetically influenced. This study examined the variation in gill pathology measured

through histology and gross gill scoring between 30 full-sibling families within a laboratory based AGD challenge trial (See appendix 1.1; Taylor et al. 2007). Considerable variation between families was observed. The estimate of broad sense heritability (H^2) ranged between 0.16 ± 0.07 and 0.30 ± 0.09 for the different scoring methods utilised. These results suggest AGD resistance is indeed a genetically influenced trait. Results from the Tasmanian commercial Atlantic salmon selective breeding program continue to show that AGD resistance is a moderately heritable trait (P. Kube pers. comm[4]). Consequently, the prospect of breeding for AGD resistance appears promising, potentially relieving commercial reliance on cost-intensive freshwater bathing.

1.6 Selective breeding and marker assisted selection

The aim of any selective breeding program is to cross one individual that has a desirable trait (i.e. increased disease resistance) with another individual also displaying that trait, so that the trait is passed to the next generation (Fjalestad et al. 2003). Such selection causes phenotypic changes by altering the allele frequencies at loci controlling the desired trait. Genetic gains in progeny are generally retained and improvements in the following generation are additive. Atlantic salmon selective breeding programs are now common in many of the salmon producing countries (Fjalestad et al. 2003). Traits under selection in such programs include growth rate, age at sexual maturity, flesh quality and disease resistance (Fjalestad et al. 2003). In Tasmania a selective breeding program for

[4] Dr Peter Kube, CSIRO Marine and Atmospheric Research.

Atlantic salmon begun in 2004 (P. Kube pers. comm). This program selects for AGD resistance, among other traits.

While simple selection based on phenotype may be adequate for traits that have high heritabilities or easily observed phenotypes, selection for traits in which the phenotype is not easily quantified, or whose quantification is prohibitively expensive, may benefit from the use of molecular technologies. Marker assisted selection (MAS) is one strategy of enhancing selective breeding programs through the integration of molecular technologies (Beaumont and Hoare 2003). In this approach genetic markers linked to or residing within loci that code for the desirable phenotype are identified. These quantitative trait loci (QTL) (i.e. the region of a genome that affects the trait in question) can then be examined within potential broodstock and appropriate broods selected based on their genotype. MAS can also be further enhanced through the identification of the gene in which the variation resides. This allows direct selection to be made on that causative loci and will increase selection accuracy. This type of extension of MAS is known as gene assisted selection or GAS.

MAS has the potential to contribute to many traditional breeding programs because it more easily quantifies desirable, but difficult to measure traits compared to conventional phenotype based selection. While this approach has proven successful with many plant breeding programs (Zhou et al. 2003) its full potential is yet to be realised within many aquaculture sectors. One exception is the recent development of a MAS program for resistance to infectious pancreatic necrosis in Atlantic salmon by Landcatch Natural Selection (www.landcatch.co.uk).

AGD resistance as a trait offers an excellent candidate for marker assisted selection for a number of reasons. Resistance to AGD is not easily measured; its measurement either requires laboratory based challenge trials or large scale field infections, both of which are expensive and labour intensive. Furthermore, Tasmanian biosecurity protocols restrict the transfer of potential broodstock from marine grow-out sites back to the freshwater hatcheries. Thus most broodstock are never actually exposed to the marine environment or to the causative agent of AGD. In this situation selection is based on sibling information rather than broodstock performance. The ability to select broods based on genotype rather than sibling phenotype performance could greatly enhance the accuracy of selection.

1.7 Identifying molecular mechanisms of resistance to AGD

If AGD resistance is to be considered a candidate for marker assisted selection, genes (or linked markers) which influence this trait must be identified. Typically, such a task would be conducted using a genome scan approach by examining hundreds, if not thousands, of polymorphic markers in segregating informative families. However due to the infancy of the Tasmanian Atlantic salmon selective breeding program, large numbers of segregating families were not available at the time of this research. Instead, a candidate gene and transcriptome approach was utilised.

1.7.1 Candidate gene approach

The candidate gene approach examines a small number of genes whose function is known and biologically relevant to the trait in question. Accordingly, the immunologically important major histocompatibility (MH) genes are an excellent candidate for identifying mechanisms of disease resistance. The MH genes are highly polymorphic and encode proteins responsible for the presentation of antigenic peptides.

A large number of associations between variation in the MH genes and disease resistance in salmonids have been reported (Langefors et al. 2001; Palti et al. 2001; Arkush et al. 2002; Lohm et al. 2002; Grimholt et al. 2003; Miller et al. 2004; Kjøglum et al. 2006; Glover et al. 2007; Kjøglum et al. 2008). Langefors et al. (2001) reported one of the first associations between variation in the MH class IIβ chain (*Sasa-DAB*) and disease resistance in Atlantic salmon. This study found broods containing a certain *Sasa-DAB* allele had a 12-fold higher chance of becoming resistant to furunculosis compared to broods without this allele (Langefors et al. 2001). This association was further validated through challenge trials of juveniles with this allele (Lohm et al. 2002). An association between MHC class II heterozygosity and resistance to infectious hematopoietic necrosis virus has also been reported in Chinook salmon *Oncorhynchus tshawytscha* (Walbaum) (Arkush et al. 2002). However, perhaps some of the most compelling studies are those of Grimholt et al. (2003), Kjøglum et al. (2006) and Kjøglum et al. (2008) who reported highly significant associations between variation in the MH class I and class II and increased survival of Atlantic salmon to both furunculosis and infectious salmon anaemia (ISA) following laboratory challenge trials.

1.7.2 Transcriptome profiling

The ability to conduct a genome-wide analysis of the transcriptome has been facilitated by the advent of cDNA microarray technology. Essentially, microarrays allow the simultaneous expression of thousands of genes to be compared between two samples. A number of microarrays have been developed for salmonids, including the 3.7 and 16k GRASP platforms (Rise et al. 2004b; von Schalburg et al. 2005b) and a 17k TRAITS/SGP platform (produced by the Transcriptome Analysis of Important Traits in Salmon (TRAITS) and Salmon Genome Projects) (Martin et al. 2007a). In salmonids, microarrays have been used to examine responses to stress (Krasnov et al. 2005), bacterial infection (Rise et al. 2004a; Ewart et al. 2005), maturation (von Schalburg et al. 2005a), vaccination (Purcell et al. 2006) and cytokine stimulation (Martin et al. 2007a; Martin et al. 2007b).

The use of microarrays to examine the transcriptional changes associated with disease resistance is gaining interest in several animal species. For instance, this strategy has proven successful to identify genes and pathways differentially expressed between lines of sheep resistant and susceptible to gastrointestinal nematode infestation (Diez-Tascón et al. 2005). In a similar study DNA microarrays were used in combination with genetic mapping to identify loci and a putative QTL conferring resistance to Marek's disease in chickens (Liu et al. 2001). Using DNA microarrays to identify the mechanisms of disease resistance in salmonids may provide valuable insight into the genes and pathways controlling this phenotype. Further research can then be directed at identifying the genetic parameters controlling these gene pathways.

1.8 Aims of this thesis

The identification of a genetic basis for AGD resistance (Taylor et al. 2007) and its subsequent inclusion in the Tasmanian Atlantic salmon selective breeding program has increased optimism for controlling AGD in Tasmania. The discovery of the molecular mechanisms influencing AGD resistance may enable a more effective exploitation of this trait through a better identification of the desirable genotype. Identifying these mechanisms is the key aim of my thesis.

Due to their high level of polymorphism and important immune function, genes of the MH represent some of the best candidate loci associated with disease resistance in vertebrates. The MH class II is particularly relevant for ectoparasite resistance because proteins encoded by these loci are responsible for the presentation of exogenously derived peptides (Storni and Bachmann 2004). I therefore conducted two studies which aimed to determine:

1. how much genetic variation at the MH class II has been retained in the Australian Atlantic salmon population; and,
2. was variation at the MH class I or class II genes associated with AGD severity?

The impetus for the first aim was derived from previous reports of reduced genetic variation within the Tasmanian Atlantic salmon population, supposedly caused by a genetic bottleneck event following importation into Australia (Elliott and Reilly 2003; Innes and Elliott 2006). This research is presented in Chapter 2. To address the second

aim, I utilised a laboratory based AGD challenge model and conducted MH genotyping to investigate the association between specific MH alleles/genotypes and resistance to AGD quantified through histopathology. This work is presented in Chapter 3.

Transcriptome profiling through the use of DNA microarrays has become a popular strategy for the identification of genes associated with disease resistance in livestock species (Liu et al. 2001; Diez-Tascón et al. 2005). This strategy can also be used to investigate genes associated with disease resistance in salmonids, made possible by the development of a number of salmonid specific cDNA microarray platforms (Rise et al. 2004b; von Schalburg et al. 2005b; Martin et al. 2007a). Utilising a newly developed Atlantic salmon cDNA microarray, I conducted two studies aimed to determine:

3. which is the most transcriptionally active organ at a time when considerable variation in disease severity is present; and,
4. which genes and pathways show differential expression between AGD resistant and AGD susceptible Atlantic salmon?

To address this third aim I compared the transcriptome response in the gill, liver and anterior kidney between naïve and AGD affected Atlantic salmon. Individuals used in this study were derived from the laboratory based AGD challenge model. This research is presented in Chapter 4. The fourth aim was achieved by comparing the gill transcriptomes of Atlantic salmon deemed to be putatively resistant or susceptible to AGD following natural infections. This is presented in Chapter 5.

In both transcriptome profiling experiments I observed one transcript of unknown identity which demonstrated particularly high differential expression between naïve and AGD affected Atlantic salmon. Due to its significant up-regulation I hypothesise that this transcript may have an important role in AGD pathogenesis. A final study, described in Chapter 6, was conducted to elucidate the role of this transcript. More specifically, this study was aimed to determine:

5. what is the full length mRNA sequence and expression characteristics of this unknown but highly differentially expressed transcript?

Finally, Chapter 7 integrates and discusses the significance of the results obtained from the preceding five research chapters and proposes future research directions.

1.9 Explanatory notes regarding thesis structure

This thesis is structured as a series of separate published manuscripts. All research chapters, with the exception of Chapter 6, comprise an already published or a manuscript in preparation. In consequence, some textual overlap occurs in the chapters. Relevant research published after the publication of these research chapters will be considered in Chapter 7. The referencing style of *Marine Biotechnology* has been adopted and a single bibliography is presented at the end of the thesis.

Chapter 2: Diversity at the MH class II within domesticated Australian Atlantic salmon

Published as: Wynne JW, Cook MT, Holmes BH, Elliott NG (2007) Allelic and haplotypic diversity at the major histocompatibility class II within domesticated Australian Atlantic salmon (*Salmo salar* L.). J Fish Biol 70:45-59

Abstract

Variation within the major histocompatibility (MH) class II alpha gene (*Sasa-DAA*) was compared between domesticated Australian Atlantic salmon and their ancestral Canadian population. The level of *Sasa-DAA* and MH class II beta gene (*Sasa-DAB*) sequence variation was also examined within the Australian population and compared to that published for European Atlantic salmon populations. In contrast to variation previously reported for non-coding microsatellite loci, a high level of MH class II allelic variation has been maintained within the domesticated Australian populations. Furthermore, a high level of *Sasa-DAA* and *Sasa-DAB* sequence diversity was also observed, and exceeded that reported for other cultured Atlantic salmon populations. The number of *Sasa-DAB* allele sequences (14) surpassed the number of *Sasa-DAA* allele sequences (9) to produce 14 unique class II haplotypes. We conclude the Australian Atlantic salmon populations show high MH class II allelic and haplotypic variation compared to both its ancestral Canadian population and other cultured Atlantic salmon populations.

2.1 Introduction

Atlantic salmon (*Salmo salar* L.) were successfully imported from the River Philip, Canada to New South Wales, Australia, in the mid 1960s. Following importation, the population was maintained in a freshwater landlocked hatchery at Gaden, New South Wales. In 1984 a Tasmanian commercial aquaculture industry commenced from this Gaden stock (Ward et al. 1994), and has grown rapidly. Anecdotal comments about population bottlenecks have caused some concern within the industry. Consequently, a number of studies, most of which concentrated on non-coding loci, have been conducted to assess genetic variation within the Australian Atlantic salmon populations (Reilly et al. 1999; Elliott and Reilly 2003; Innes and Elliott 2006). Although some inconsistencies have been observed, these studies have generally demonstrated reductions in both the number of alleles (Innes and Elliott 2006) and heterozygosity (Elliott and Reilly 2003) compared to the ancestral Canadian population. While these results suggest that some loss of genetic diversity has occurred, the direct functional relevance of reduced non-coding variation is unknown. This study posits that coding regions, such as the major histocompatibility complex (MHC) class II genes, may be more biologically relevant regions to examine, given that a reduction in MHC variation can have direct functional importance to disease resistance (O' Brien et al. 1985) and so to survival.

Genes of the MHC represent the most polymorphic genes in the vertebrate genome (Marsh et al. 2000). The classical MHC genes encode cell surface glycoproteins responsible for the presentation of self and non-self peptides to T lymphocytes (T cells).

In most cases, presentation and recognition of foreign peptides elicits a humoral or cell mediated immune response. Polymorphism within the MHC enables different allelic variants to bind and present unique sets of antigenic peptides, functionally resulting in differences in disease resistance.

In contrast to other vertebrates, the teleost MHC has undergone extensive genomic reorganisation (Nonaka et al. 2001). The MHC class I and class II genes form a tightly linked complex in most vertebrates (Nonaka et al. 1997; Kaufman et al. 1999). However, in teleosts the class I and class II genes remain on different linkage groups (Sato et al. 2000; Nonaka et al. 2001; Shum et al. 2001). As the MHC genes in Atlantic salmon do not form a complex, they are often designated the MH genes (Stet et al. 2002). Some linkage between the Atlantic salmon MH genes still exists. For instance, genes encoding the MH class II alpha (*Sasa-DAA*) and beta (*Sasa-DAB*) genes are tightly linked and co-segregate as composite haplotypes (Stet et al. 2002). Furthermore, in some cultured Atlantic salmon populations the same number of alleles are observed at the *Sasa-DAA* and *Sasa-DAB* loci and combine to produce an equivalent number of unique class II composite haplotypes (Stet et al. 2002). Another interesting feature of *Sasa-DAA* is the presence of a polymorphic minisatellite embedded within the 3' untranslated (UT) region (Grimholt et al. 2000; Grimholt et al. 2002; Stet et al. 2002). In some cultured Atlantic salmon populations, each unique MH class II composite haplotype (*Sasa-DAA/Sasa-DAB*) is characterised by a unique minisatellite marker allele from this 3' UT region (Stet et al. 2002).

Pathogen mediated balancing selection through either overdominant (Hughes and Nei 1989) or negative frequency dependent selection (Takahata and Nei 1990) is suggested

to be the major evolutionary force maintaining and promoting polymorphism within the MHC genes (Jeffery and Bangham 2000; Bernatchez and Landry 2003; Sommer 2005). Evidence for both overdominant (heterozygous advantage) and negative frequency dependent selection has been found within salmonid populations. For instance, chinook salmon (*Oncorhynchus tshawytscha*) heterozygous at their MHC class II beta locus have increased survival following challenge by infectious haematopoietic necrosis virus (Arkush et al. 2002). Moreover, specific MH class I and class II alleles have been associated with increased resistance to furunculosis (Langefors et al. 2001; Grimholt et al. 2003) and infectious salmon anaemia (Grimholt et al. 2003). Such results suggest a loss of MH variation either as reduced heterozygosity or allelic diversity may have adverse affects on a population's ability to resist disease.

In the present study we investigated whether previously reported losses in non-coding variation in the Tasmanian Atlantic salmon population are reflected within the coding MH class II genes. This was done first by comparing variation within a *Sasa-DAA* minisatellite repeat (known as *Sasa-DAA-3UTR*) in two domesticated Australian Atlantic salmon populations with their ancestral Canadian population, and then by sequence analysis of the *Sasa-DAA* and *Sasa-DAB* genes within the Tasmanian population, and comparison with European populations.

2.2 Materials and Methods

2.2.1 Population samples for *Sasa-DAA-3UTR* analysis

Three populations of Atlantic salmon were analysed for *Sasa-DAA-3UTR* variation. The first sample comprised archived scale samples (n= 63) from wild Atlantic salmon from the ancestral population in the River Philip, Nova Scotia, Canada. Collection details for these samples, captured in 1971 and 1972, are described by Innes and Elliott (2006). This population was considered to represent natural MH diversity in the population at the time of importation into Australia. The second sample was muscle tissue (n= 78) from the Gaden, New South Wales (NSW) population (Reilly et al. 1999) and the third was muscle tissue (n= 80) from the 2002 year class of the Tasmanian commercial population.

2.2.2 DNA isolations

Genomic DNA (gDNA) isolation from the Canadian archived scale samples is previously described by Innes and Elliott (2006) and followed a modified protocol of Adcock et al. (2000). Briefly, scales (1-2 per sample) were incubated in extraction buffer at 55°C overnight. DNA was then extracted twice with phenol and once with chloroform isoamyl alcohol (24:1, v:v), washed twice with 100% ethanol and resuspended in MilliQ water. gDNA from the Tasmanian and New South Wales muscle samples were isolated using a modified CTAB protocol outlined by Ward et al. (1994).

2.2.3 *Sasa-DAA-3UTR* genotyping

Variation was assessed within the previously described *Sasa-DAA* minisatellite (*Sasa-DAA-3UTR*), which resides in the 3' UT region on the MH class II alpha gene (Grimholt et al. 2000; Stet et al. 2002). *Sasa-DAA-3UTR* was amplified using the fluorescently FAM labelled sense primer (5'-GATGGCAAAGAGGAAAGTGAG-3') and anti-sense primer (5'-TTGTTATGCTCTACCTCTGAA-3') (Stet et al. 2002). PCR reactions were performed in a total volume of 12.5 µL, containing 10-50 ng of gDNA, 1.25 µl of 10X buffer, 200 µM dNTPs, 2.5 mM $MgCl_2$, 200 nM of each primer and 1 unit of *Taq* polymerase (Applied Biosystems, Foster City, USA). Cycling was performed in an ABI 9600 thermal cycler (Applied Biosystems) with the following parameters: 95°C for 10 min, 35 cycles of 94°C for 1 min 52°C for 1 min and 72°C for 1min and a final extension of 72°C for 7 min. Resulting products were diluted (1/5), denatured in HiDi formamide (Applied Biosystems) and analysed using a ABI 3100 capillary sequencer (Applied Biosystems). Allele size was determined using the internal size standard, Genescan-500 LIZ, and the program Genemapper ver3.5 (Applied Biosystems).

2.2.4 *Sasa-DAA* and *Sasa-DAB* sequencing

Full length sequences of expressed *Sasa-DAA* alleles and their co-segregating *Sasa-DAB* alleles were obtained from individuals comprising the 2002 commercial

Tasmanian year class. Sequences were obtained from three to eight individuals presenting a unique S*asa-DAA-3UTR* allele. Total RNA was isolated from anterior kidney using RNAwiz (Ambion, Austin, USA) as per the manufacturer's instructions. Approximately 2-5 μg of RNA was reverse transcribed using Superscript III (Invitrogen, Mount Waverly, Australia) to generate first strand cDNA. Coding regions for the *Sasa-DAA* locus were amplified using the primers DA66F (5'-TGCTGGCAGGTGTATGCAGAA-3') and DA1054R (5'-TTGTTATGCTCTACCTCTGAA-3') (Stet et al. 2002). The *Sasa-DAB* locus was amplified using the primers DB40F (5'-ATGTCGATGTCTATCTTCTG-3') and DB684R (5'-CAGACCAGACGCACCGATGGC-3') (Stet et al. 2002). The PCR reactions were identical for both loci and were performed in a total volume of 50 μL containing approximately 5 μl 10X buffer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 200 nM of each primer, 2 units of *Taq* polymerase (Applied Biosystems) and 1 μL of cDNA template. Amplification was performed in a GeneAmp 9700 thermal cycler (Applied Biosystems) using the following cycling profile: 10 min at 95°C, 25 cycles of 1 min at 94°C, 30 sec at 58°C and 1 min at 72°C, and a final extension step for 7 min at 72°C.

PCR products were purified using the QIAquick PCR purification kit (Qiagen, Doncaster, Australia) and ligated into pGEM-T easy vector (Promega, Armadale, Australia). Constructs were transformed into an electrocompetent strain of DH10β as per the manufacturer's instructions (Promega). Eight to ten clones from each PCR amplification were picked, on-grown and plasmid purified using the QIAprep spin miniprep kit (Qiagen). By sequencing multiple clones (8 to10), PCR errors could be identified and eliminated from analysis. Both strands of each plasmid was sequenced using the ABI Prism Bigdye Terminator Cycle Sequencing Ready Reaction kit (Applied

Biosystems) using the T7 and Sp6 primers. Sequencing products were purified using the CleanSEQ Sequencing Reaction Clean-Up system (Agencourt Bioscience, Beverly, USA). Purified sequencing products were analysed on an ABI 3100 capillary sequencer (Applied Biosystems).

2.2.5 Sequence and data analysis

All sequences were visualised in ChromasPro ver1.2 (Technelysium, Tewantin, Australia) and alleles were defined based on their amino acid sequence as in previous studies (Grimholt et al., 2002; Stet et al., 2002). Alignments were performed using ClustalW in the program MEGA ver3.0 (Kumar et al. 2004) and BioEdit ver7.0.5 (Hall 1999). Phylogenetic analysis was performed on both the *Sasa-DAA* and *Sasa-DAB* loci using the neighbour-joining method in the program MEGA ver3.0 (Kumar et al. 2004). Topology was tested using 1,000 bootstrap iterations and only values over 50% are shown. Expected and observed heterozygosity and subsequent conformance to Hardy-Weinberg equilibrium was tested using Genepop ver3.2 (Raymond and Rousset 1995). Genetic differentiation and pairwise F_{st} values were calculated using Arlequin ver3.01 (Excoffier and Schneider 2005). Allelic richness was calculated using FSAT ver2.9.3.2 (Goudet 1995).

2.3 Results

2.3.1 *Sasa-DAA-3UTR* variation

Eight *Sasa-DAA-3UTR* alleles were identified across the three populations of Atlantic salmon (Table 2.1). In most cases alleles were at 9 to12 base pair (bp) intervals and ranged in size from 207bp to 307bp. With one exception, all *Sasa-DAA-3UTR* alleles were shared between the three populations. The exception was the *Sasa-DAA-3UTR* allele 307, which was only found (at a low frequency) in the Canadian sample. Significant allelic differentiation was observed between all populations. Pairwise F_{st} estimates indicated the greatest differentiation was between Canada and NSW (F_{st} =0.025, $p<0.001$) and the least between NSW and Tasmania (F_{st} =0.013, $p<0.001$). In both the Tasmanian and NSW samples the *Sasa-DAA-3UTR* alleles 229 and 259 were the most common, while in the Canadian sample the *Sasa-DAA-3UTR* allele 239 followed by 229 were the most common. Only a small decrease in *Sasa-DAA-3UTR* allelic richness was observed within the Australian samples compared to the Canadian sample. High heterozygosity was observed in all samples and no significant deviations from Hardy-Weinberg equilibrium were detected.

Locus	Allele (bp)	Canada	New South Wales	Tasmania
Sasa-DAA-3UTR	207	0.159	0.147	0.162
	229	0.206	0.205	0.250
	239	0.302	0.173	0.144
	250	0.016	0.006	0.038
	259	0.167	0.340	0.206
	279	0.063	0.026	0.125
	298	0.048	0.103	0.075
	307	0.040	-	-
	n	63	78	80
	H_o	0.74	0.83	0.83
	H_e	0.81	0.78	0.82
	p	0.512	0.170	0.128
	AR	8.000	6.807	7.000

Table 2.1. *Sasa-DAA-3UTR* allele frequencies. (n) number of individuals scored; (H_o) observed and (H_e) Hardy-Weinberg expected heterozygosity estimates; (p) probability of conformance to Hardy-Weinberg equilibrium; (AR) allelic richness; (-) allele not detected.

2.3.2 *Sasa-DAA* sequences

Sequencing of full length *Sasa-DAA* clones revealed the presence of nine expressed allele sequences, four of which were novel (Figure 2.1). The four novel allele sequences were designated *Sasa-DAA*0802*, *Sasa-DAA*0902*, *Sasa-DAA*1002* and *Sasa-DAA*1102* according to the nomenclature used by Shum et al. (2002). These four sequences were deposited in GenBank and assigned the accession numbers AM259956 to AM259959. The five previously described allele sequences detected were, *Sasa-DAA*1001* (Grimholt et al. 2000) *Sasa-DAA*0201*, *Sasa-DAA*0601* (Stet et al. 2002) *Sasa-DAA*0303* and *Sasa-DAA*1201* (Consuegra et al. 2005a). In all cases the predicted *Sasa-DAA-3UTR* fragment size was found within the expressed *Sasa-DAA* sequence, verifying that the *Sasa-DAA-3UTR* alleles were derived from the expressed *Sasa-DAA* locus and little or no genotyping/sequencing error was evident. An alignment of the amino acid sequences demonstrates the majority of variation between allele sequences resides within the alpha-1 domain. In total, 20 of the 225 possible amino acid residues were polymorphic, with 15 of these positions occurring within the alpha-1 domain. No individuals expressed more than two allele sequences.

Phylogenetic analysis of the alpha-1 domain (exon 2) nucleotide sequences with other published (European Atlantic salmon populations) *Sasa-DAA* sequences (Grimholt et al. 2000; Stet et al. 2002; Consuegra et al. 2005a) demonstrate the existence of three possible allelic lineages (Figure 2.2). The nine observed sequences in this study were spread across all three lineages. The four unique Tasmanian sequences (*Sasa-*

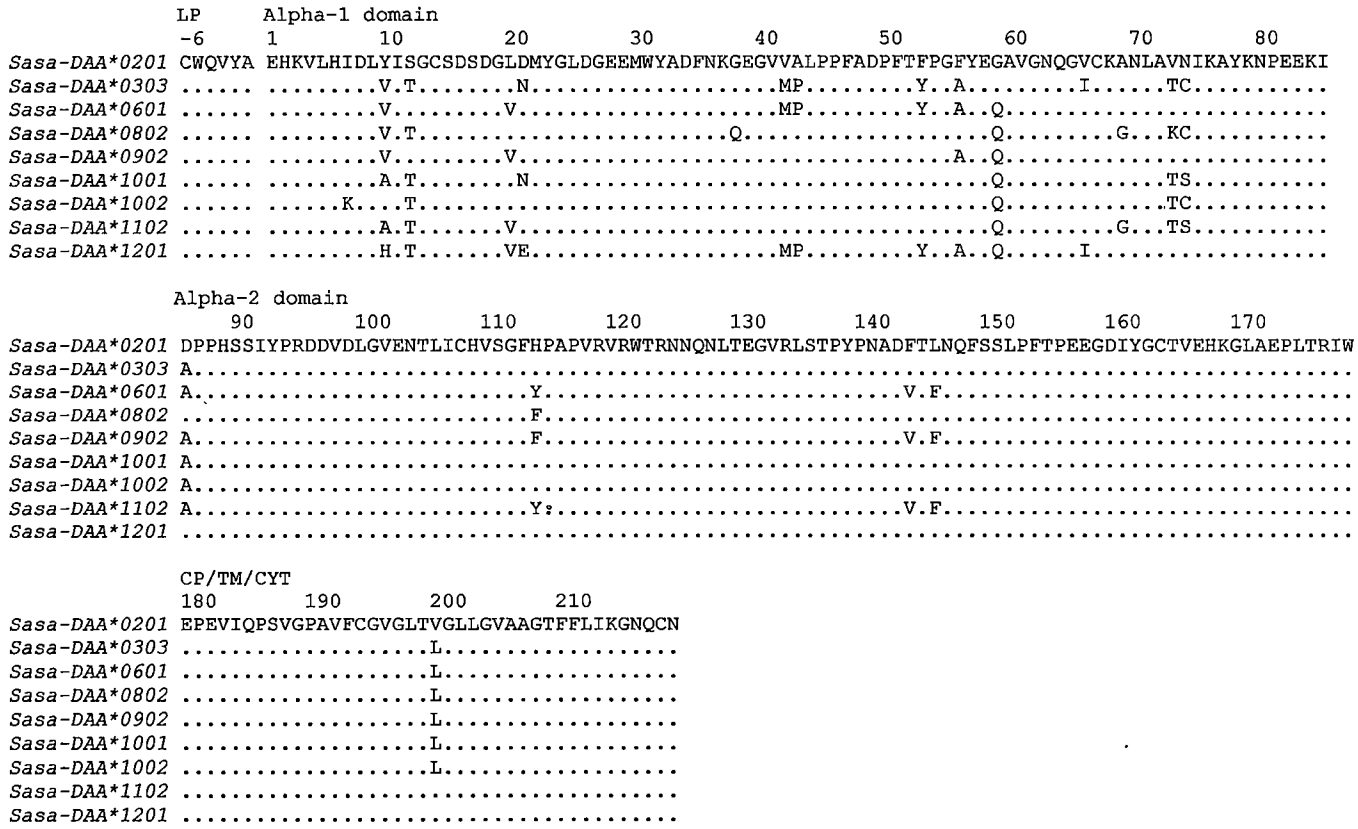


Figure 2.1. Alignment of all MH class II *Sasa-DAA* amino acid sequences detected within our study. *Dots* denote identities to *Sasa-DAA*0201*. (LP) leader peptide, alpha 1-domain, alpha-2 domain, (CP) connecting peptide, (TM) transmembrane and (CYT) cytoplasmic regions indicated according to Stet et al. (2002).

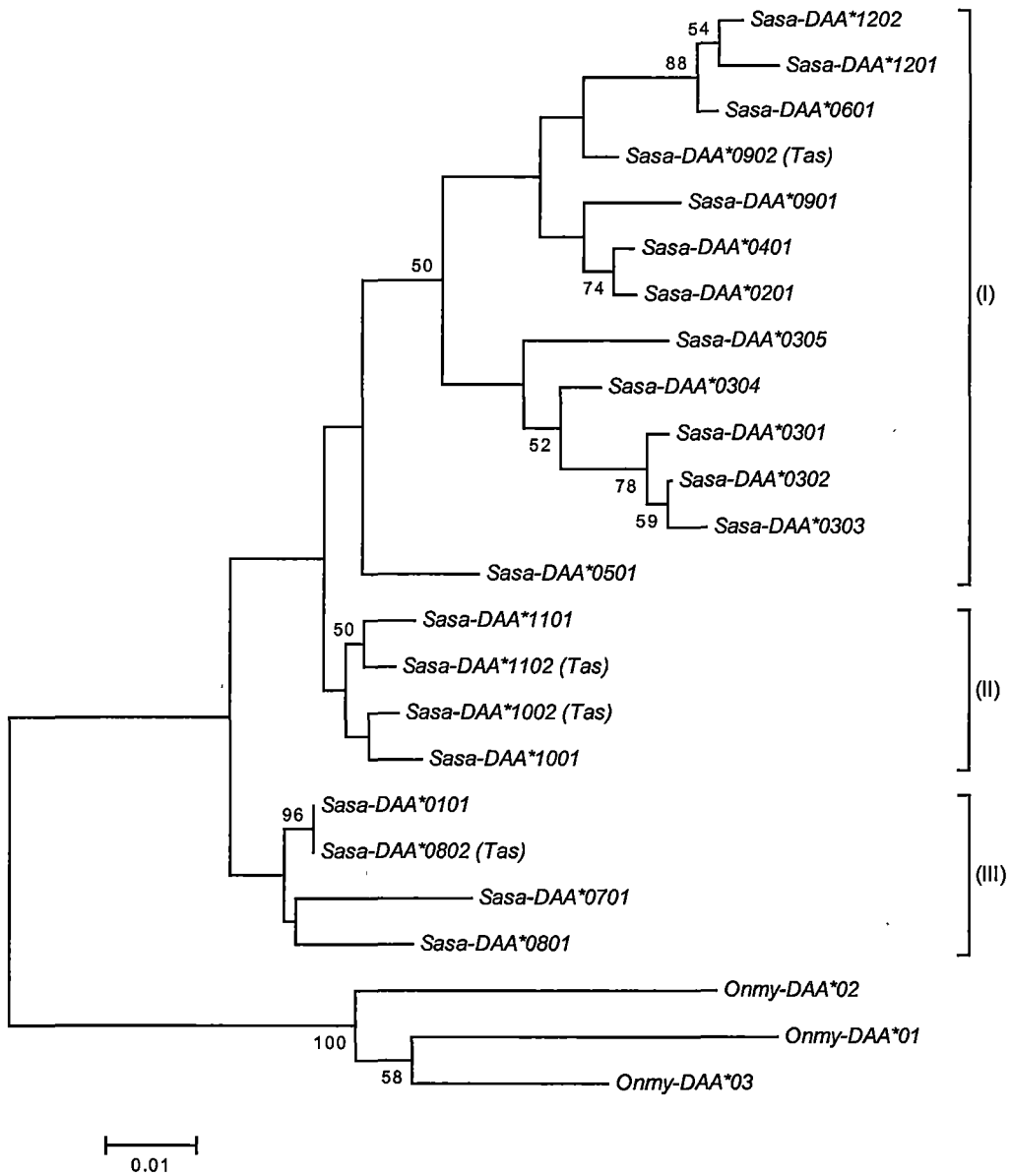


Figure 2.2. Phylogenetic analysis of *Sasa-DAA* allele sequences. Analysis was performed using the neighbour-joining method based on the alpha-1 domain nucleotide sequences with 1,000 bootstrap iterations. Only bootstrap values of 50 or greater are depicted and genetic distance is shown *below*. The three rainbow trout allele sequences, *Onmy-DAA*01* to *Onmy-DAA*03*, were used as out-groups and were obtained from Grimholt et al. (2000). Sequences unique to Tasmania (Tas) are shown in *Brackets*.

*DAA*0802*, *Sasa-DAA*0902*, *Sasa-DAA*1002*, *Sasa-DAA*1102*) clustered within the three allelic lineages.

Genotyping the *Sasa-DAA-3UTR* locus and sequencing expressed full length *Sasa-DAA* clones facilitated the examination of the association between *Sasa-DAA-3UTR* alleles and allele sequences. In most cases each unique *Sasa-DAA* sequence appeared to be characterised by a single *Sasa-DAA-3UTR* allele. For instance, *Sasa-DAA-3UTR* alleles 207, 239, 250, 259 and 279 appeared linked to sequences *Sasa-DAA*0201*, *Sasa-DAA*1001*, *Sasa-DAA*0303*, *Sasa-DAA*1201* and *Sasa-DAA*1102*, respectively (Table 2.2). However, *Sasa-DAA-3UTR* alleles 229 and 298 were each linked to two allele sequences.

2.3.3 *Sasa-DAB* sequences

Sequencing the *Sasa-DAB* locus revealed the presence of 14 allele sequences (Figure 2.3). Two of these sequences were previously described full length sequences, seven were previously described partial sequences (beta-1 domain) and five were novel. The two full length sequence identities were *Sasa-DAB*0801* and *Sasa-DAB*0401* (Stet et al. 2002). The seven partial identities were to alleles Sm11x, Sm5b, Sm12t, Sm6v, Sm9w, Sm2c and Sm4k (Landry and Bernatchez 2001) all of which were only exon 2 sequences (beta-1 domain). For simplicity, our full length sequences of these alleles have been renamed *Sasa-DAB*1704*, *Sasa-DAB*1602*, *Sasa-DAB*0402*, *Sasa-DAB*1502*, *Sasa-DAB*0303*, *Sasa-DAB*1303* and *Sasa-DAB*0403*, respectively, according to the Immuno Polymorphism Database (IPD)

Sasa-DAA-3UTR	*Sasa-DAA* sequence	No. of fish	*Sasa-DAB* sequence	No. of fish
207	*DAA*0201*	5	*DAB*1704*	4
			*DAB*0403*	1
229	***DAA*0802***	3	*DAB*0801*	3
229	***DAA*1002***	3	*DAB*1602*	3
239	*DAA*1001*	8	*DAB*0402*	3*
			*DAB*1502*	3
			DAB*1705	1
250	*DAA*0303*	3	*DAB*0401*	3
259	*DAA*1201*	7	***DAB*0702***	5
			DAB*1706	1
			DAB*0103	1
279	***DAA*1102***	3	*DAB*0303*	3
298	***DAA*0902***	2	***DAB*0603***	2
298	*DAA*0601*	1	*DAB*1303*	1

Table 2.2. *Sasa-DAA-3UTR* alleles linked to *Sasa-DAA* allele sequences and their co-segregating *Sasa-DAB* allele sequence for Tasmanian Atlantic salmon. Allele sequences shown in *bold* are newly described alleles (* note one co-segregating *Sasa-DAB* not detected in the 10 clones sequenced).

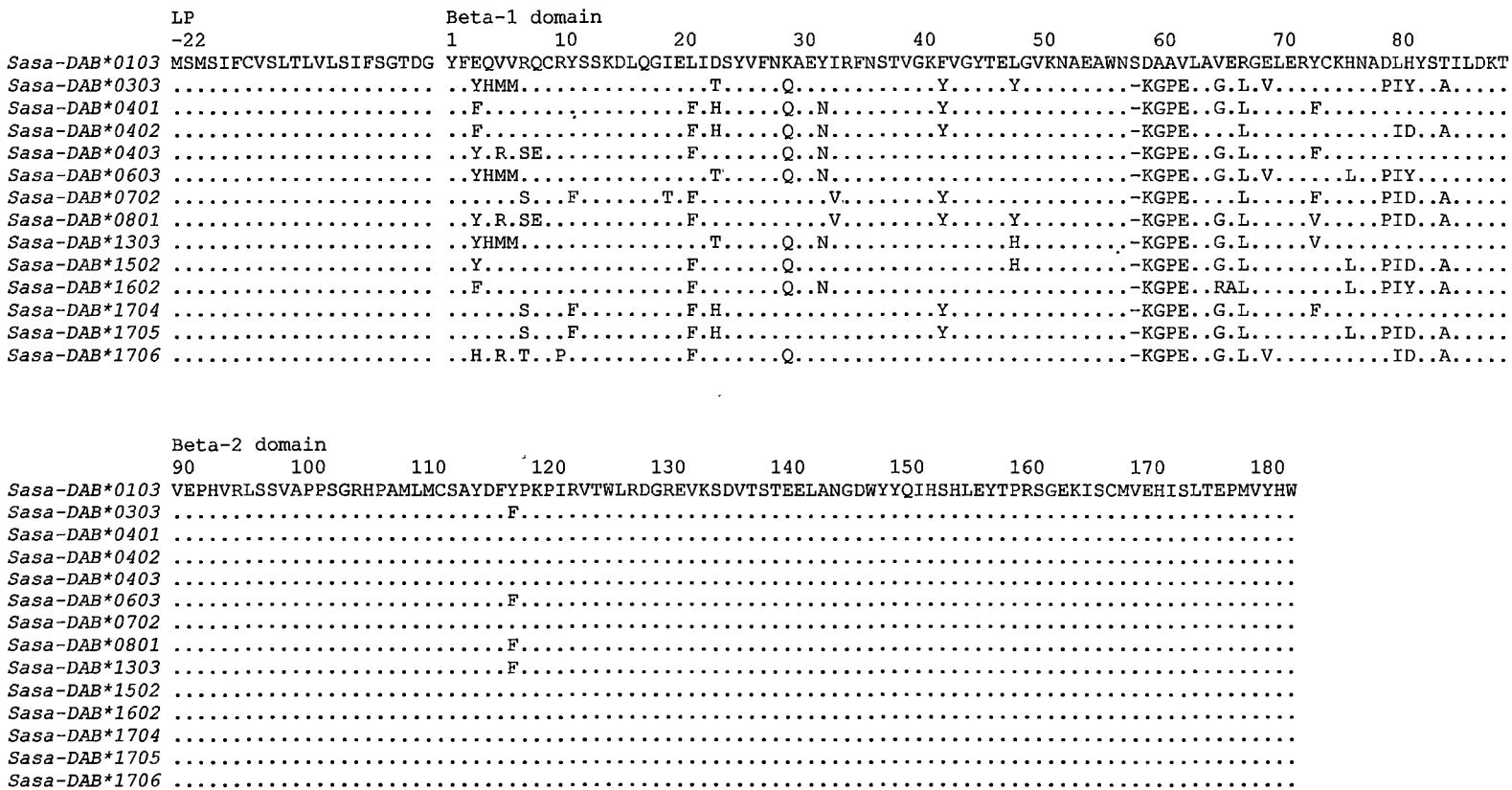


Figure 2.3. Continued over page

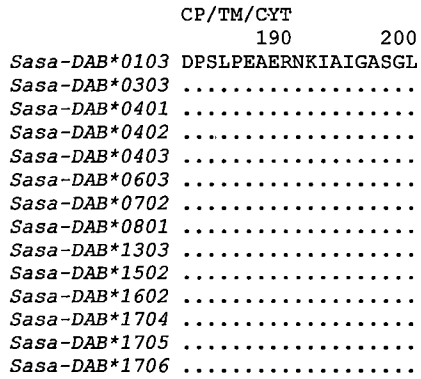


Figure 2.3. Alignment of all MH class II *Sasa-DAB* amino acid sequences detected within our study. *Dots* denote identities to *Sasa-DAB*0103* and *dashes* indicate gaps within the alignment. (LP) leader peptide, beta 1-domain, beta-2 domain, (CP) connecting peptide, (TM) transmembrane and (CYT) cytoplasmic regions indicated according to Stet et al. (2002).

(www.ebi.ac.uk/cgi-bin/ipd/mhc) (Ellis et al. 2006). The five new alleles identified within this study have been designated *Sasa-DAB*0103*, *Sasa-DAB*0603*, *Sasa-DAB*0702*, *Sasa-DAB*1705* and *Sasa-DAB*1706*. Our full sequences for the five novel and seven partially described sequences were deposited in GenBank and assigned the accession numbers AM259944 to AM259955. An amino acid alignment demonstrates that almost all variation at the *Sasa-DAB* locus is restricted to the beta-1 domain. A total of 32 polymorphic amino acid positions were detected, with 31 of these residing within the beta-1 domain. Like the *Sasa-DAA* locus, no individual expressed more than two *Sasa-DAB* allele sequences.

Phylogenetic analysis of the beta-1 domain with other published sequences (Hordvik et al. 1993; Stet et al. 2002) demonstrated the presence of one major allelic lineage containing 18 of the 20 sequences (Figure 2.4). One possible sub lineage containing the sequences *Sasa-DAB*0301*, *Sasa-DAB*0303*, *Sasa-DAB*1303*, *Sasa-DAB*0601* and *Sasa-DAB*0603* was identified within the major lineage and was supported by a high bootstrap value (91). Four of the newly identified Tasmanian sequences (*Sasa-DAB*0603*, *Sasa-DAB*0702*, *Sasa-DAB*1705* and *Sasa-DAB*1706*) were within the main allelic lineage. The other new sequence, *Sasa-DAB*0103*, along with *Sasa-DAB*0101*, showed high divergence from the main lineage. The seven sequences (*Sasa-DAB*1704*, *Sasa-DAB*1602*, *Sasa-DAB*0402*, *Sasa-DAB*1502*, *Sasa-DAB*0303*, *Sasa-DAB*1303* and *Sasa-DAB*0403*) which had been previously described only at the beta-1 domain from Canadian Atlantic salmon (Landry and Bernatchez 2001) showed no distinct clustering, nor did the sequences derived from Norwegian Atlantic salmon (Stet et al. 2002).

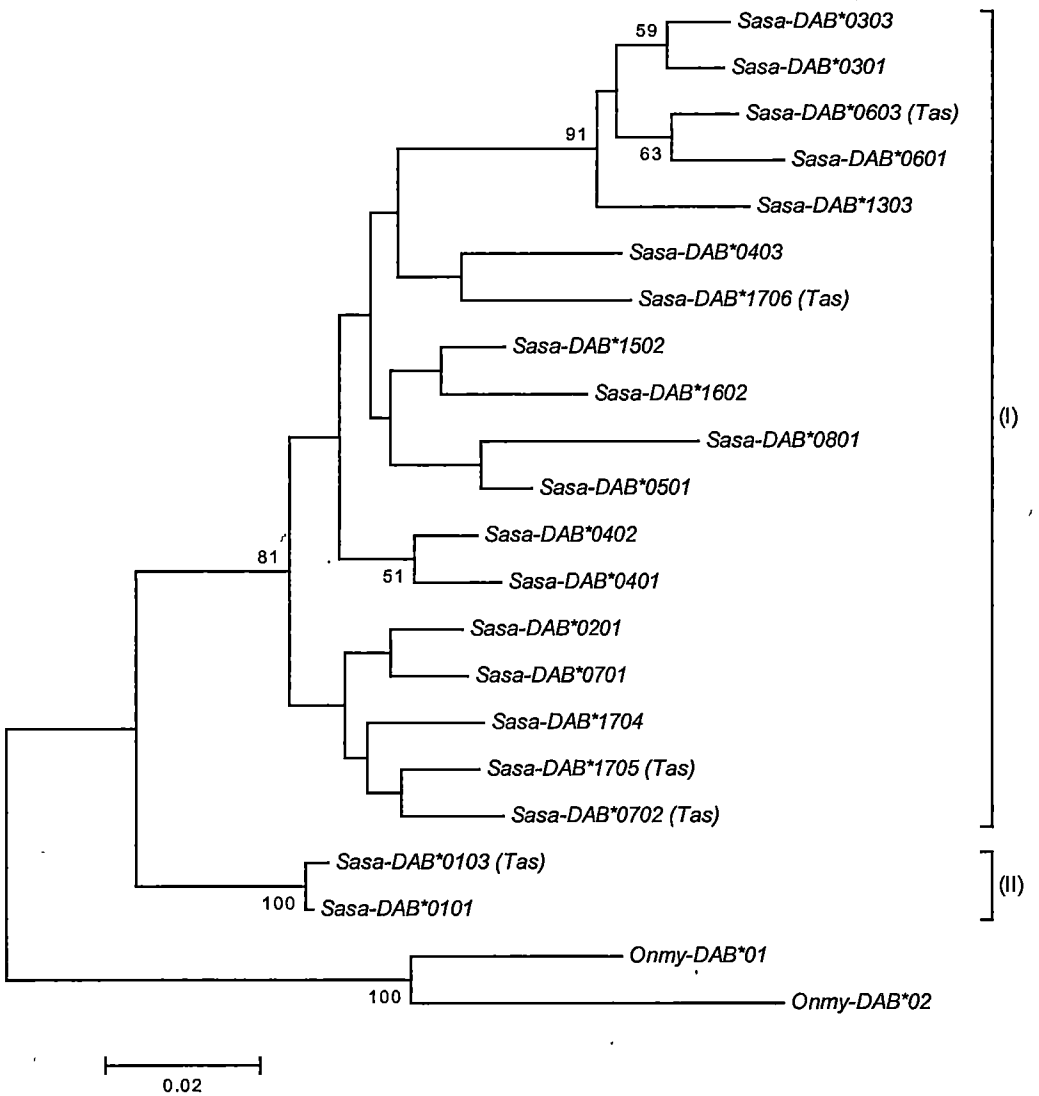


Figure 2.4. Phylogenetic analysis of *Sasa-DAB* allele sequences. Analysis was performed using the neighbour-joining method based on the beta-1 domain nucleotide sequences with 1,000 bootstrap iterations. Only bootstrap values of 50 or greater are depicted and genetic distance is shown *below*. Two rainbow trout sequences, *Onmy-DAB*01* and *Onmy-DAB*02*, were used as out-groups and were obtained from Glamann (1995). Sequences unique to Tasmania (Tas) are shown in *Brackets*.

The 14 *Sasa-DAB* sequences identified within this study combined with the nine *Sasa-DAA* sequences to produce 14 unique MH class II composite haplotypes (Table 2.2). Variations in composite haplotype frequencies were evident. Five composite haplotypes were observed only within single fish while the remaining nine composite haplotypes were detected in two to five fish each. In some cases multiple *Sasa-DAB* allele sequences combined with single *Sasa-DAA* allele sequences (generally the most common alleles). In one individual, heterozygous for *Sasa-DAA*0201/Sasa-DAA*1001*, only one co-segregating *Sasa-DAB* allele (*Sasa-DAB*1704*) was detected within the 10 clones. We suggest the *Sasa-DAA*1001* co-segregating *Sasa-DAB* allele was present, however was not detected within the 10 clones selected for sequencing.

2.4 Discussion

Non-coding genetic variation within the Tasmanian Atlantic salmon population has been the subject of multiple studies (Ward et al. 1994; Reilly et al. 1999; Elliott and Reilly 2003; Innes and Elliott 2006). A general reduction in the number of microsatellite alleles, and to a lesser extent heterozygosity, has been detected within the Tasmanian population compared to its ancestral Canadian population. A moderate genetic bottleneck event during the early years of domestication has been hypothesised as responsible (Elliott and Reilly 2003). The present study was undertaken to test if this decreased non-coding variation was also reflected in coding genes. The MH genes were chosen due to their high polymorphism and direct functional relevance to disease resistance.

Results from both the minisatellite and sequence analysis suggest high variation exists within the MH class II genes of Australian Atlantic salmon. Heterozygosity at the *Sasa-DAA-3UTR* locus was similar for both domesticated Australian populations (Tasmanian and NSW) and their ancestral Canadian population. Furthermore, only one *Sasa-DAA-3UTR* allele was missing within the Australian population, and this might well be attributed to the small sample size. High numbers of *Sasa-DAA* (9) and *Sasa-DAB* (14) allele sequences were observed within the Tasmanian Atlantic salmon population and exceeded that observed for a cultured Atlantic salmon population in Norway (Stet et al. 2002). This Norwegian population, known as G1, was originally founded from wild stock collected in 13 Norwegian rivers in 1973 (Gjedrem 2000). In this population, seven *Sasa-DAA* and seven *Sasa-DAB* alleles sequences were identified across 33 families. Interestingly, these seven *Sasa-DAA* and *Sasa-DAB* alleles co-segregated to

produce seven unique class II composite haplotypes (Stet et al. 2002). In contrast, the nine *Sasa-DAA* and 14 *Sasa-DAB* allele sequences reported within our study combined to produce 14 unique class II composite haplotypes. At the time of conducting this study the Tasmanian Atlantic salmon population had never been selected for any production traits.

The high level of MH class II heterozygosity and allelic/haplotypic diversity reported herein, contrasts the level of genetic diversity reported at non-coding loci. Significant losses of non-coding microsatellite alleles (up to 21-43%) (Innes and Elliott 2006) and to a lesser extent heterozygosity (Elliott and Reilly 2003) has been reported within the Tasmanian Atlantic salmon population compared to the ancestral Canadian population. It seems likely that while a genetic bottleneck event may have reduced variation at many neutral non-coding loci in the Tasmanian stock, diversity at the MH class II has been maintained. A somewhat similar situation is observed for the San Nicolas Island fox, *Urocyon littoralis* (Baird). This species is extremely monomorphic at most non-coding loci, yet displays high levels of MHC variation, presumably maintained and promoted through intense pathogen mediated balancing selection (Aguilar et al. 2004). While, the extent to which balancing selection acts to promote MH variation within an aquaculture environment may be lower than in the natural state, as cultured fish experiencing disease tend to be quickly treated, pathogen mediated selection is still likely to be a significant evolutionary force.

In our study, the seven unique *Sasa-DAA-3UTR* alleles were linked to nine *Sasa-DAA* allele sequences. This demonstrates that unique *Sasa-DAA* sequences are not always characterised by a single *Sasa-DAA-3UTR* allele, in contrast to the findings in a

Norwegian cultured population (Stet et al. 2002). A similar situation was observed for the Atlantic salmon MH class I (*Sasa-UBA*) locus, whereas many as seven *Sasa-UBA* allele sequences are characterised by a single *Sasa-UBA-3UTR* allele (Grimholt et al. 2002). This finding suggests not all variation at the *Sasa-DAA* locus was detected within our study. It is possible that additional *Sasa-DAA* allele sequences characterised by existing *Sasa-DAA-3UTR* alleles were also present. Such a situation would cause an underestimation of both the true number of *Sasa-DAA* alleles, and possibly the true heterozygosity. However, we considered the abundance of such "non-detected" alleles to be even across all populations and therefore did not bias our result.

This study reported the existence of four novel *Sasa-DAA* sequences. The fact that these alleles have not been previously reported within either Irish or Norwegian Atlantic salmon may suggest these alleles are unique to Tasmanian (or Canadian derived) Atlantic salmon. The phylogenetic analysis of these and other published *Sasa-DAA* allele sequences revealed the presence of three possible allelic lineages. The four novel *Sasa-DAA* alleles and the five previously reported alleles observed within our study were distributed across the three lineages. This result may suggest that the divergent *Sasa-DAA* allelic lineages observed within this study predate the separation of North American and European populations of Atlantic salmon. Interesting, in comparison, significant divergence between the North American and European populations have been reported previously by examining supposedly neutral microsatellite variation (King et al. 2001). A previous study by Consuegra et al. (2005a) also found low divergence between Irish and Norwegian Atlantic salmon populations at the MH class II.

In the case where different *Sasa-DAA* sequences were characterised by a single *Sasa-DAA-3UTR* allele (i.e. sequences linked to *Sasa-DAA-3UTR* alleles 229 and 298), no phylogenetic relationship between the sequences could be established. This suggests that the different sequences characterised by single *Sasa-DAA-3UTR* alleles are similar only by state, not necessarily by descent. A similar situation was sometimes observed at the MH class I *Sasa-UBA* locus where multiple allele sequences linked to single *Sasa-UBA-3UTR* alleles sometimes showed little similarity, however this was not always the case (Grimholt et al. 2002). Evidence against this hypothesis can also be observed within our results when compared to other studies. That is, in some instances when sequences were characterised by a similar *Sasa-DAA-3UTR* allele size, a close phylogenetic relationship can be established. For example, in Norwegian Atlantic salmon the *Sasa-DAA-3UTR* allele 232 is linked to allele sequence *Sasa-DAA*0101* (Stet et al. 2002). This allele shows a high similarity to the allele *Sasa-DAA*0802* which in our study was characterised by the *Sasa-DAA-3UTR* allele 229. It therefore seems likely that while some alleles may just happen to be linked to the same minisatellite allele by chance, other alleles may be derived from the same origin.

Our study also found considerable variation in the MH class II beta gene (*Sasa-DAB*). Although most alleles had been previously described, five novel alleles were also detected. In contrast to other studies, the number of *Sasa-DAB* alleles identified within the Tasmanian sample exceeded the number of *Sasa-DAA* alleles (Stet et al. 2002). A total of 14 *Sasa-DAA/Sasa-DAB* composite haplotypes were identified. In some cases multiple *Sasa-DAB* allele sequences appeared to combine with single *Sasa-DAA* allele sequences. For instance, *Sasa-DAA*1001* appeared to combine with sequences *Sasa-DAB*0402*, *Sasa-DAB*1502* and *Sasa-DAB*1705* to produce three diverse composite

haplotypes. Moreover, the allele sequences *Sasa-DAA*0201* and *Sasa-DAA*1201* also appear to combine with two and three *Sasa-DAB* sequences each. One possible reason to explain this high level of haplotypic diversity may be that within a population multiple *Sasa-DAB* alleles can co-segregate with a single *DAA* allele. This leads to the question, has pathogen driven balancing selection promoted variation in the beta chain more so than the alpha chain within Canadian derived Atlantic salmon? It may be interesting to note that in some other teleost species, such as cichlids, the number of beta loci surpasses the number of alpha loci (Málaga-Trillo et al. 1998; Murray et al. 2000). As a result it has been speculated that each class II alpha chain may combine with more than one beta chain to produce mature class II heterodimers (Murray et al. 2000) . In humans, the class II alpha locus is generally less polymorphic than the class II beta locus (Marsh et al. 2000). The *Sasa-DAB* alleles reported within our study generally showed more polymorphic positions when compared to the *Sasa-DAA* locus, and this result was consistent with other studies on Atlantic salmon (Stet et al. 2002).

To conclude, this study found no evidence of reduced variation at the MH class II *Sasa-DAA-3UTR* locus within the domesticated Australian Atlantic salmon population compared to its ancestral Canadian population. This result is in contrast to previous studies on non-coding microsatellites that have reported both significant loss of alleles and some loss of heterozygosity. *Sasa-DAA* and *Sasa-DAB* sequence variation within the Tasmanian Atlantic salmon population was higher to that reported for other cultured populations of Atlantic salmon. Finally, the variation at the *Sasa-DAB* locus exceeded variation at the *Sasa-DAA* locus and combined to produce high MH class II haplotypic diversity.

Chapter 3: MH polymorphism associated with resistance to AGD in Atlantic salmon

Published as: Wynne JW, Cook MT, Nowak BE, Elliott NG (2007) Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (*Salmo salar* L.). Fish Shellfish Immun 22:707-717

Abstract

The association between major histocompatibility (MH) polymorphism and the severity of infection by amoebic gill disease (AGD) was investigated across 30 full-sibling families of Atlantic salmon. Individuals were challenged with AGD for 19 days and then their severity of infection scored by histopathological examination of the gills. Fish were then genotyped for the MH class I (*Sasa-UBA*) and MH class II alpha (*Sasa-DAA*) genes using polymorphic repeats embedded within 3' untranslated regions of the *Sasa-UBA* and *Sasa-DAA* genes. High variation in the severity of infection was observed across the sample material, ranging from 0 to 85% gill filaments infected. In total, seven *Sasa-DAA-3UTR* and ten *Sasa-UBA-3UTR* marker alleles were identified across the 30 families. A significant association between the marker allele *Sasa-DAA-3UTR* 239 and a reduction in AGD severity was detected. There was also a significant association found between AGD severity and the presence of two *Sasa-DAA-3UTR* genotypes. While the associations between MH allele/genotypes and AGD severity reported herein may be statistically significant, the small sample sizes observed for some alleles and genotypes means these associations should be considered as suggestive and future research is required to verify their biological significance.

3.1 Introduction

Amoebic gill disease (AGD) is the most significant disease affecting Atlantic salmon in Tasmania, Australia (Munday et al. 2001). AGD is caused by the protozoan *Neoparamoeba* spp. which, acting as the primary pathogen, infect the gills of marine cultured salmonids (Wong et al. 2004; Dyková et al. 2005). Clinical signs of AGD include lethargy, respiratory distress, rapid opercular movement, and ultimately, if not treated, death (Munday et al. 1990). Histologically, AGD is initially characterised by attachment of amoeba to healthy gill epithelial tissue followed by the progression of lamellar fusion, epithelial desquamation and oedema, epithelial hyperplasia and interlamellar vesicle formation (Adams and Nowak 2003; Adams and Nowak 2004a). An infiltration of leucocytes into the central venous sinus is also observed adjacent to infected regions and often increases with disease progression (Adams and Nowak 2004a). In some cases, leucocytes are found in close association with amoeba within the hyperplastic tissue (Bridle et al. 2003). Significant variation in the severity of AGD infection is observed within the Tasmanian Atlantic salmon population. At present, the molecular mechanism of any resistance is yet to be elucidated. If genetically controlled, AGD resistance may be enhanced by selective breeding. The identification of genes which influence resistance may greatly improve the efficacy of such a breeding strategy.

Genes of the major histocompatibility complex (MHC) represent the most polymorphic genes in the vertebrate genome (Marsh et al. 2000).The classical MHC genes encode cell surface glycoproteins responsible for the presentation of self and non-self peptides

to T lymphocytes (T cells). MHC class I molecules present endogenously derived peptides to cytotoxic $CD8^+$ T cells, while MHC class II molecules present exogenously derived peptides to helper $CD4^+$ T cells (Rammensee 1995). In most cases, presentation and recognition of foreign peptides elicits a humoral or cell mediated immune response. Typically, the highest level of polymorphism observed in the MHC genes is concentrated within the peptide binding regions (PBR). Polymorphism within the PBR enables different allelic variants to bind and present unique sets of antigenic peptides. Such variations in antigen presentations may functionally result in differences in disease resistance.

In contrast to humans, the genomic organisation of the teleost MHC genes is considered relatively 'simple.' The most unusual feature of this simple MHC is the lack of linkage between the classical class I and class II genes (Sato et al. 2000; Shum et al. 2002), possibly facilitating independent divergence between the classes (Consuegra et al. 2005a; Consuegra et al. 2005b). As the MHC genes in Atlantic salmon do not form a complex, they have been renamed the MH genes (Stet et al. 2002). Interestingly, in salmonids at least, the class II alpha and beta loci (*DAA* and *DAB*) have remained linked, and co-segregate as composite haplotypes (Stet et al. 2002). Expression of both the classical class I and class II genes are characterised by a single dominantly expressed locus in Atlantic salmon and rainbow trout (Grimholt et al. 2002; Shum et al. 2002; Stet et al. 2002). In Atlantic salmon, polymorphic repeats have been identified and characterised within the 3' untranslated (UT) regions of both the class I (*Sasa-UBA*) and class II alpha (*Sasa-DAA*) loci (Grimholt et al. 2000; Grimholt et al. 2002; Stet et al. 2002). In many instances, these unique marker alleles have been linked to unique allele sequences. For instance, Stet et al. (2002) showed in a cultured population of

Norwegian Atlantic salmon that each *Sasa-DAA-3UTR* marker allele was linked to a unique class II haplotype. From the same population, most *Sasa-UBA-3UTR* marker alleles were shown to be linked to one or two allele sequences (Grimholt et al. 2002). However, one *Sasa-UBA-3UTR* allele was linked to seven different allele sequences.

Due to the polymorphic nature of the MHC it has been possible to associate certain alleles/haplotypes to increased disease resistance. For instance, in chickens, one MHC haplotype is significantly associated with resistance to Marek's disease (MD) (Bacon 1987). Similarly, in Soay sheep (*Ovis aries* L.) MHC polymorphism was significantly associated with both juvenile survival and nematode parasitism (Paterson et al. 1998). Perhaps the most compelling examples in salmonids come from the studies by Grimholt et al. (2003), which found a highly significant association between MH polymorphism and resistance towards furunculosis and infectious salmon anaemia in Atlantic salmon. Moreover, increased resistance to infectious hematopoietic necrosis in Atlantic salmon has been associated with certain MH alleles (Miller et al. 2004). The aim of our study was to examine the correlation between MH polymorphism and the severity of infection by AGD in Atlantic salmon.

3.2 Materials and Methods

3.2.1 Animals and disease challenge trial

Thirty full-sibling Atlantic salmon families were infected with AGD in a disease challenge. All fish were obtained from a commercial salmon hatchery and comprised the first generation of the Tasmanian Atlantic salmon selective breeding program. All families were unrelated and comprised of single parent crosses. The relatedness of parents was unknown. Briefly, each family was initially represented by 18 pre-smolt fish which were maintained in equal numbers in three 3000 L recirculating systems at the University of Tasmania. Following temperature and salinity acclimation, each tank of fish was inoculated with 500 cells L^{-1} of *Neoparamoeba* spp. as described by Morrison et al. (2004). Fish were monitored daily for AGD mortalities until day 19 post-inoculation when all individuals were euthanised using with 5 g L^{-1} Aqui-S (Aqui-S NZ Ltd, Lower Hutt, New Zealand). The gill cage and a skeletal muscle sample was dissected from all individuals and stored in seawater Davidson's fixative and 95% ethanol, respectively.

3.2.2 Gill histopathological scoring

The second left anterior hemibranch was dissected for each fish, dehydrated, embedded in paraffin wax and sectioned at 5 μm. Sections were stained with haematoxylin and eosin and viewed with a light microscope (Zeiss, Oberkochen, West Germany) at 400X

magnification. The percentage of gill filaments (primary lamellae) displaying hyperplastic lesions were then counted for each section, as previously described by Adams and Nowak (2004b). Filaments were only counted when the central venous sinus was visible in at least two-thirds of the total length. All sections were also examined for the presence of *Neoparamoeba* spp.

3.2.3 *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* marker analysis

Variation within the MH class I and class II loci were examined using previously described polymorphic repeats identified within the 3' UT region of the *Sasa-UBA* and *Sasa-DAA* loci (Grimholt et al. 2000; Grimholt et al. 2002; Stet et al. 2002). Genomic DNA was isolated from approximately 5 mg of muscle tissue using Chelex chelating resin (Bio-Rad, Hercules, USA) as described by Walsh et al. (1991). The *Sasa-UBA* repeat was amplified using the fluorescently FAM labelled sense primer (5' GGAGAGCTGCCCAGATGACTT 3') and anti-sense primer (5' CAATTACCACAAGCCCGCTC 3') (Grimholt et al. 2002). The *Sasa-DAA* repeat was amplified using the fluorescently FAM labelled sense primer (5' GATGGCAAAGAGGAAAGTGAG 3') and anti-sense primer (5' TTGTTATGCTCTACCTCTGAA 3') (Stet et al. 2002). A duplex PCR reaction was performed in a total volume of 12.5 µl containing approximately 10-50 ng gDNA, 1.25 µl of 10X PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 200 nM of each primer and 1 unit of *Taq* polymerase (Applied Biosystems, Foster City, USA). Amplification was performed in a GeneAmp 9700 thermal cycler (Applied Biosystems) using the following cycling profile: 10 min at 95°C, 35 cycles of 1 min at 94°C, 1 min at 52°C

and 1 min at 72°C, and a final extension step for 7 min at 72°C. Resulting products were diluted (1/5), denatured in HiDi formamide (Applied Biosystems) and analysed using the ABI 3100 capillary sequencer (Applied Biosystems). Allele size was determined using the internal size standard, Genescan-500 LIZ, in the program Genemapper (Applied Biosystems).

3.2.4 Statistical analysis

All histopathology scores and MH genotypes were compiled into a database. Due to the restricted number of individuals comprising each family, this study could not examine the segregation of alleles associated with AGD on a family basis. Instead, we examined the marker-trait association at a population level (i.e. across the 30 families). The dependent variable within this study was the severity of infection measured as the percentage of gill filaments showing AGD lesions. The presence/absence of MH alleles and genotypes were analysed as dichotomous independent variables, with individuals coded 1 for the presence of that allele or genotype and 0 for the absence of that allele or genotype. Power analysis was used to determine the minimum number of individuals to be included within the statistical tests. Given a significance threshold of 0.05 and power of 0.80 it was estimated a minimum of 20 individuals were required for both the single allele and genotype analysis. Multiple regression analysis was first used to test for a significant relationship between presence/absence of MH alleles and genotypes with the severity of infection. Next, any allele that was significant within the multiple regression was independently tested using an analysis of variance (ANOVA) model. When multiple tests were performed (i.e. during independent allele analysis) the significance

value (0.05) was adjusted using a standard bonferroni correction (Rice 1989). Mendelian inheritance of both *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* across the 30 families was tested using χ^2 analysis. All tests were performed using SPSS ver12.0.

3.3 Results

3.3.1 AGD histopathology scoring

AGD histopathology scoring and MH genotyping was successful in 474 individuals. Some individuals were unable to be successfully scored for the severity of AGD due to an inadequate number of countable filaments. These individuals were removed from further analysis. Nineteen AGD induced mortalities (with visible gross lesions) were observed prior to day 19 post-inoculation. These individuals were also removed from further analysis. The percentage of gill filaments infected ranged from 0 to 85% across all individuals (Figure 3.1). The distribution of gill scores showed some skewness and kurtosis and was transformed using an arcsine square root transformation for all statistical analyses. Between families the mean percentage of infected filaments ranged from 21 to 40%. The number of individuals within each family varied between 13 and 18 fish.

Significant histopathology characteristic of AGD was observed within the gill sections, including epithelial oedema and hyperplasia and interlamellar vesicle formation. An infiltration of non-specific inflammatory cells including macrophages was typically observed within both the central venous sinus and on the boundaries of lesions. The presence of *Neoparamoeba* spp. was often observed within the margins of hyperplastic lesions.

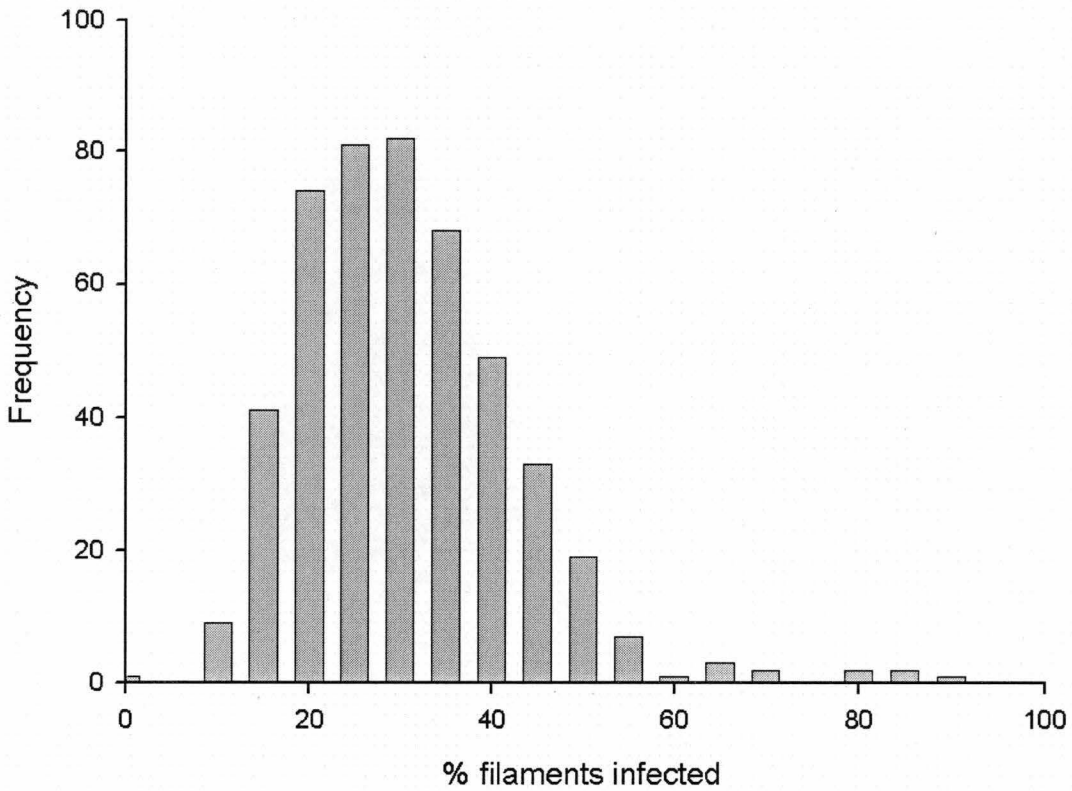


Figure 3.1. Frequency distribution (prior to transformation) of histopathology scores (% filaments infected) in 474 Atlantic salmon deriving from 30 full-sibling families.

3.3.2 MH polymorphism

High variation at both the *Sasa-DAA* and *Sasa-UBA* loci was observed within our study. A total of seven *Sasa-DAA-3UTR* and ten *Sasa-UBA-3UTR* marker alleles were detected across the 30 families. *Sasa-DAA-3UTR* marker alleles were typically observed at 9-12 base pair.(bp) increments, while the *Sasa-UBA-3UTR* marker alleles were at 2bp increments. Marker allele nomenclature was assigned according to the size of each allele determined during capillary electrophoresis. Allele frequencies across the 30 families are presented within Table 3.1. The *Sasa-DAA* locus was dominated by marker alleles 229 and 259, while the *Sasa-UBA* locus was dominated by allele 327. Mendelian inheritance for both the *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* loci was observed across all families analysed (Appendix 2.1 and 2.2).

3.3.3 Association between the MH and AGD severity

The association between the presence and absence of *Sasa-DAA-3UTR* marker alleles and severity of AGD infection was tested by multiple regression analysis. All alleles at *Sasa-DAA-3UTR* locus had sufficient replication to be included within the analysis. The regression was generally a poor fit ($R^2 = 0.05$), however, the overall relationship was significant ($F= 3.54$, $df= 7, 466$, $p<0.05$). Only marker alleles *Sasa-DAA-3UTR* 239 and *Sasa-DAA-3UTR* 298 were found to be significantly associated with severity of infection (Table 3.2). However, following the individual analysis of these alleles, only *Sasa-DAA-3UTR* 239 remained significant ($F= 7.52$, $df= 1, 471$, $p<0.05$) after

Locus	Marker allele (bp)	Frequency	No. of families
Sasa-DAA-3UTR	207	0.164	14
	229	0.336	22
	239	0.086	9
	250	0.051	3
	259	0.252	21
	279	0.085	8
	298	0.025	3
Sasa-UBA-3UTR	309	0.020	3
	313	0.115	12
	315	0.024	2
	317	0.112	9
	321	0.090	10
	323	0.012	1
	327	0.361	28
	329	0.018	2
	333	0.159	14
	335	0.090	11

Table 3.1. Frequencies of *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* marker alleles within the 30 full-sibling families of Atlantic salmon.

Sasa-DAA-3UTR alleles	*b* (95% CI) [a]	(p) ANOVA [b]	No. of fish
207	-0.36 (-2.51 to 1.78)	-	142
229	1.35 (-0.85 to 3.55)	-	266
239	-3.26 (-5.69 to -0.82)*	0.003*	82
250	2.17 (-0.59 to 4.92)	-	47
259.	0.93 (-1.18 to 3.04)	-	211
279	1.57 (-0.89 to 4.03)	-	75
298	6.85 (2.97 to 10.73)*	0.034	24

Table 3.2. [a] Regression coefficients (*b*) with 95% confidence intervals in *brackets* for all single *Sasa-DAA-3UTR* marker alleles included within the multiple regression analysis, * denotes allele significance at $p<0.05$. [b] Individual ANOVA probabilities (p) of only *Sasa-DAA-3UTR* marker alleles significantly associated with severity of AGD infection determined within the multiple regression, * denotes significant difference at $p<0.025$ following bonferroni correction. The number of fish (No. of fish) presenting at least one copy of that allele and therefore included within the analysis is also shown.

bonferroni correction for multiple tests. The negative regression coefficient (b= -3.26) for *Sasa-DAA-3UTR* 239 indicated that the presence of this allele was associated with a decrease in severity of infection and is graphically expressed in Figure 3.2. The mean percentage gill filaments infected for individuals with the presence of marker allele *Sasa-DAA-3UTR* 239 was approximately 4.0% lower compared to individuals without this allele. *Sasa-DAA-3UTR* 239 was present in 82 fish and within 9 of the 30 families analysed.

At the *Sasa-UBA-3UTR* locus three alleles had insufficient replication to be included within the multiple regression analysis. These were *Sasa-UBA-3UTR* alleles 309, 323 and 329. Analysis of the remaining *Sasa-UBA-3UTR* alleles indicated the regression was a poor fit (R^2= 0.03) yet statistically significant (F= 2.14, df= 7,466, $p<0.05$). However, while the *Sasa-UBA-3UTR* alleles 321 and 327 both had regression coefficients significantly different from zero, neither allele remained significant following independent ANOVA verification.

Analysis of variance was used to compare the severity of infection between fish heterozygous and homozygous for the *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* loci. In total, 101 fish were homozygous for the *Sasa-DAA-3UTR* locus and 90 fish were homozygous for the *Sasa-UBA-3UTR* locus. Only 24 fish were homozygous for both the *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* locus. No significant difference in the severity of infection was observed between heterozygotes and homozygotes for either the *Sasa-DAA-3UTR* (F= 0.51, df= 1,471, $p>0.05$) or the *Sasa-UBA-3UTR* locus (F= 2.50, df= 1,471, $p>0.05$).

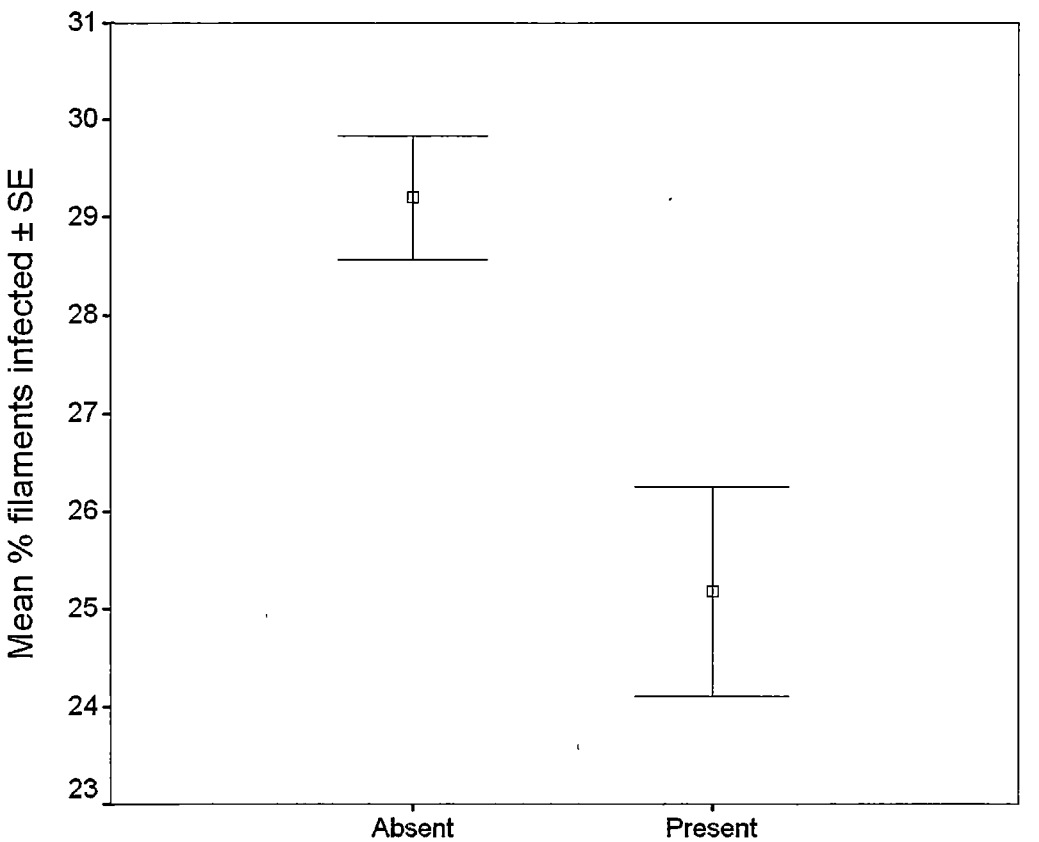


Figure 3.2. Mean % of gill filaments (± standard error) displaying AGD lesions for individuals with the presence and absence of marker allele *Sasa-DAA-3UTR* 239.

The association between combined *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* marker alleles (genotypes) and severity of AGD infection was also investigated. A total of 20 different *Sasa-DAA-3UTR* genotypes and 31 different *Sasa-UBA-3UTR* genotypes were observed within the sample material (Table 3.3). Genotype frequencies varied greatly, ranging from 0.004 up to 0.181. As a consequence only a small number of genotypes had adequate replication to be included within the regression analysis. At the *Sasa-DAA-3UTR* locus ten genotypes were represented by greater than 20 individuals and were therefore used as the predictors within the multiple regression (Table 3.4). The overall regression fit was low (R^2= 0.06), however, statistically significant (F= 2.78, df = 10,463, $p<0.05$). Genotypes *Sasa-DAA-3UTR* 239-259 and 259-259 were both significantly associated with severity of infection. Both genotypes *Sasa-DAA-3UTR* 239-259 (F= 10.29, df= 1,471, $p<0.05$) and 259-259 (F= 6.86, df= 1,471, $p<0.05$) remained significant following independent analysis and bonferroni correction. The presence of *Sasa-DAA-3UTR* genotypes 259-259 and 239-259 were associated with a decrease in disease severity by approximately 4.9 and 5.4%, respectively (Figure 3.3).

At the *Sasa-UBA-3UTR* locus, only eight genotypes were represented by greater than 20 individuals. These *Sasa-UBA-3UTR* genotypes included in the regression analysis were 313-317, 333-335, 321-333, 321-327, 313-327, 317-327, 327-327 and 327-333. The regression fit for this locus was poor (R^2= 0.02) and not statistically significant (F= 1.36, df= 8,465, $p>0.05$).

Locus	Genotype	Frequency	No. of families
Sasa-DAA-3UTR	207-207	0.027	3
	207-229	0.116	8
	207-239	0.019	1
	207-259	0.074	7
	207-279	0.057	4
	207-298	0.006	1
	229-229	0.112	8
	229-239	0.046	4
	229-250	0.038	3
	229-259	0.181	13
	229-279	0.063	7
	229-298	0.004	1
	239-250	0.011	1
	239-259	0.057	6
	239-298	0.04	2
	250-250	0.002	1
	250-259	0.049	4
	259-259	0.059	6
	259-279	0.025	3
	279-279	0.013	2
Sasa-DAA-3UTR	309-313	0.004	1
	309-327	0.023	3
	309-333	0.004	1
	309-335	0.008	1
	313-313	0.011	1

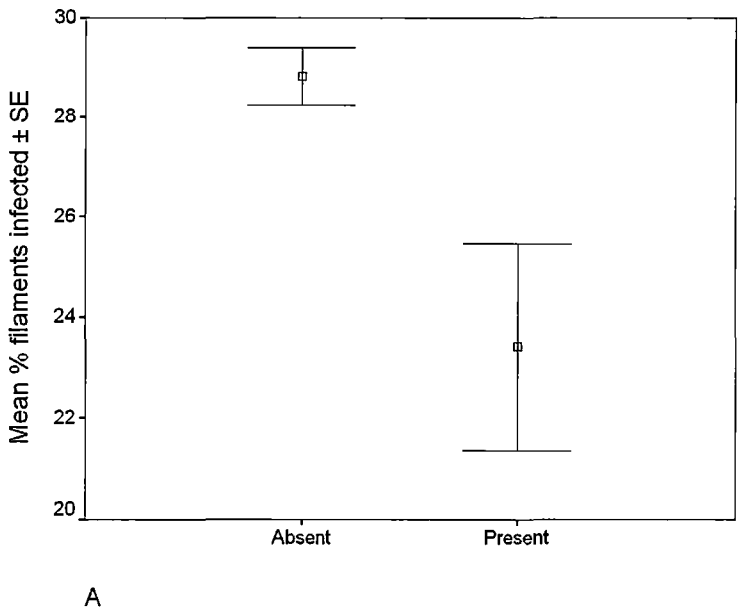
Table 3.3. Continued over page

313-315	0.008	1
313-317	0.042	3
313-321	0.006	1
313-323	0.015	1
313-327	0.078	8
313-333	0.027	4
313-335	0.027	3
315-321	0.017	1
315-327	0.023	2
317-317	0.011	1
317-321	0.03	2
317-327	0.08	7
317-329	0.017	1
317-335	0.034	4
321-327	0.063	6
321-333	0.044	5
321-335	0.019	2
323-327	0.008	1
327-327	0.135	12
327-329	0.006	1
327-333	0.133	12
327-335	0.036	7
329-333	0.013	2
333-333	0.027	1
333-335	0.042	4
335-335	0.006	1

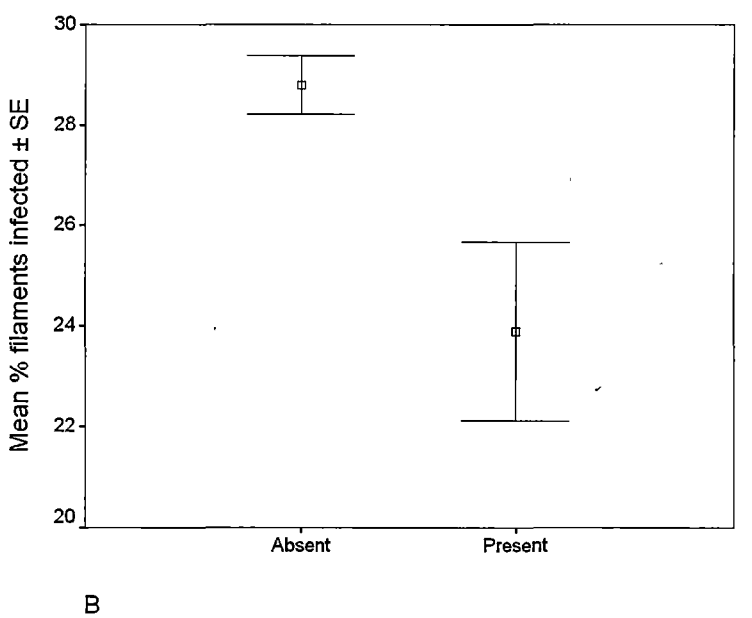
Table 3.3. Frequencies of *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* genotypes within the 30 full-sibling families of Atlantic salmon

Sasa-DAA-3UTR genotypes	*b* (95% CI)[a]	(p) ANOVA[b]	No. of fish
207-229	-2.64 (-5.27 to -0.001)	-	55
207-259	-2.64 (-5.70 to 0.43)	-	35
207-279	-0.02 (-3.39 to 3.35)	-	27
229-229	-0.17 (-2.84 to 2.48)	-	53
229-239	-1.05 (-4.71 to 2.60)	-	22
229-259	0.95 (-1.37 to 3.28)	-	86
229-279	-0.06 (-3.25 to 3.24)	-	30
239-259	-4.45 (-7.82 to -1.07)*	0.007*	27
250-259	3.42 (-0.17 to 7.01)	-	23
259-259	-3.72 (-7.05 to -0.39)*	0.011*	28

Table 3.4. [a] Regression coefficients (*b*) with 95% confidence intervals in *brackets* for *Sasa-DAA-3UTR* genotypes included within the multiple regression analysis, * denotes genotype significance at $p<0.05$. [b] Individual ANOVA probabilities (p) of only *Sasa-DAA-3UTR* genotypes significantly associated with severity of AGD infection determined within the multiple regression, * denotes significant difference at $p<0.025$ following bonferroni correction. The number of fish (No. of fish) with each genotype is also shown.



A



B

Figure 3.3. Mean % of gill filaments (± standard error) displaying AGD lesions for individuals with the presence and absence of *Sasa-UBA-3UTR* genotypes 239-259 (A) and 259-259 (B).

3.4 Discussion

To our knowledge this is the first study to undertake a molecular assessment of resistance to amoebic gill disease in Atlantic salmon. We examined the MH genes due to their high polymorphism, important immune function and previous associations with increased resistance to other diseases. The results demonstrate the overall effect of the *Sasa-DAA* and *Sasa-UBA* genes, as assessed by studying the embedded markers, on the severity of infection was small at the *Sasa-DAA*, yet statistically significant. The presence of marker allele *Sasa-DAA-3UTR* 239 was significantly associated with a decrease in the severity of AGD infection. Individuals containing *Sasa-DAA-3UTR* 239 had approximately 4.0% less gill filaments infected compared to individuals without this allele. Furthermore the presence of *Sasa-DAA-3UTR* genotypes 239-259 and 259-259 were also associated with a decrease in disease severity. Individuals with these genotypes had approximately 5.4 and 4.9% less gill filaments infected compared to individuals without these genotypes, respectively.

Although encouraging, the associations presented herein should only be viewed as suggestive at this time and require further research to verify their biological significance. The high number of alleles and genotypes observed at the MH (often at a low frequency) means finding a strong statistically significant association between alleles/genotypes and disease resistance in a small sample is difficult (Hill 1998). Furthermore high polymorphism of the MH also results in divergent sample sizes between groups of individuals with different alleles or genotypes. This was observed in our study and resulted in an unbalanced design, often with small sample sizes.

Consequently the statistical power to reliably detect differences in disease resistance was reduced.

While a significant association between MH alleles or genotypes and AGD severity was identified within our study, individuals containing a copy of these alleles or genotypes appeared to demonstrate only a small decrease in the severity of infection compared to individuals without these alleles or genotypes. Preliminarily research has shown AGD resistance quantified through histopathological scoring has a moderate heritability (Taylor et al. 2007). It therefore seems possible that while reductions in gill scores of 4 to 5% may be low in regard to the overall variation in resistance, it may actually account for a moderate proportion of the genetic variance. The variation in disease severity between the most resistant and susceptible families was 19%. It is also highly likely that AGD resistance is under polygenic control and thus additional non MH genes may also be contributing. A similar situation was observed for MD rèsistance within chickens following the initial association with the MHC B haplotype (Bacon 1987). Seven additional quantitative trait loci (QTL) were significantly associated with resistance each explaining 2-10% of the variation (Yonash et al. 1999). In rainbow trout, two QTL were found to be associated with resistance towards infectious pancreatic necrosis (Ozaki et al. 2001).

Other associations between MH polymorphism and resistance to salmonid diseases including furunculosis (Langefors et al. 2001; Grimholt et al. 2003) infectious salmon anaemia (Grimholt et al. 2003) and infectious hematopoietic necrosis (Miller et al. 2004) have been reported. Our findings, along with these previous studies, highlight the importance of MH polymorphism in disease resistance. Although speculative, one may

hypothesise that the association between the *Sasa-DAA-3UTR* alleles/genotypes and reduced severity of infection was the result of an enhanced immune response triggered by a more effective presentation of exogenous derived peptide ligands by these MH class II molecules. Studies have indicated MH class II^+ cells are present within AGD affected lesions and show varying levels of expression (Morrison et al. 2006b). The infiltration of leucocytes, with possible antigen presenting capacity, into the AGD affected regions observed within this and previous studies (Adams and Nowak 2004a) may provide further evidence of such an immune response.

Our study could not exclude the possibility that the MH alleles associated with the lower severity of infection were actually in linkage disequilibrium with another locus that may be causing the observed associations. In humans, the MHC is a gene rich region spanning over 4Mb in length (Campbell and Trowsdale 1993; Newell et al. 1996) with an MHC class III region encoding many cytokine and complement factors (Hauptmann and Bahrarn 2004). However, in Atlantic salmon no such class III region has been identified and the only reported linkage between MH genes is between the *Sasa-UBA* and *Sasa-TAP2B* loci (Grimholt et al. 2002) and between the *Sasa-DAA* and *Sasa-DAB* loci (Stet et al. 2002). In rainbow trout a number of physical mapping studies have examined the genomic organisation of the MH class I region using large insert BAC or phage libraries (Shum et al. 2002; Phillips et al. 2003; Shiina et al. 2005). In general a number of loci associated with antigen processing and binding, such as the immunoproteasome subunits (*PSMB)*, are closely linked to the *Onmy-UBA* locus (Shum et al. 2002; Phillips et al. 2003; Shiina et al. 2005).

In a previous study the marker allele *Sasa-DAA-3UTR* 239 was found to be linked to allele sequence *Sasa-DAA*1001* (Wynne et al. 2007a). However, in contrast to European cultured populations of Atlantic salmon (Stet et al. 2002), not all Tasmanian *Sasa-DAA-3UTR* marker alleles are found to be associated with a single allele sequence. For instance, *Sasa-DAA-3UTR* 229 and *Sasa-DAA-3UTR* 298 were each found to be associated with two allele sequences (Wynne et al. 2007a). This suggests that not all variation at the *Sasa-DAA* locus was detected by using the 3' minisatellite marker within our study, and it is therefore possible that additional associations between these non detected alleles and the severity of infection may have been present. Furthermore, in the Tasmanian Atlantic salmon population, the link between specific *Sasa-DAA* and *Sasa-DAB* (MH class II beta chain) sequences remains complex. That is, in some cases multiple *Sasa-DAB* allele sequences appear to combine with a single *Sasa-DAA* allele sequence to produce diverse class II composite haplotypes (Wynne et al. 2007a). As a result, we were unable to determine the *Sasa-DAA/Sasa-DAB* composite haplotype by genotyping the 3' minisatellite marker and could not assess the possible association between *Sasa-DAB* polymorphism and AGD resistance. At present, no examination of the link between *Sasa-UBA-3UTR* marker alleles and allele sequences has been undertaken within the Tasmanian Atlantic salmon population. Combining the *Sasa-DAA-3UTR* marker analysis with allele specific *Sasa-DAB* PCR in order to resolve all alleles is a desirable future research direction. This type of analysis may allow more of the genetic effects of the MH on AGD resistance to be quantified. Unfortunately however, this was beyond the scope of the current study.

The combined allele analysis aimed to identify *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* genotypes which were associated with severity of AGD infection. This approach has

been used successfully within other studies to further examine the association of an allele or genotype with disease resistance (Grimholt et al. 2003). Our study found significant associations between the *Sasa-DAA-3UTR* genotypes 239-259 and 259-259 and deceased disease severity. It must however be acknowledged that due to the high number of MH genotypes observed at the *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* loci, only a small number of genotypes had adequate replication to be included in the analysis. The *Sasa-DAA-3UTR* genotype 239-259 interestingly contained the same allele (*Sasa-DAA-3UTR* 239) that was identified as significant within the single allele analysis, thus strengthening this association. The *Sasa-DAA-3UTR* homozygote genotype 259-259 was also associated with a decrease in disease severity. Interestingly both these genotypes contained the *Sasa-DAA-3UTR* 259 allele, which in the single allele analysis was not significantly associated with resistance. This uncertainty between the effects of *Sasa-DAA-3UTR* 259 in a heterozygote and homozygote form may highlight the limitation of treating alleles as an individual entity rather than a genotype within the single allele analysis. That is, when alleles associated with resistance and susceptibility are combined as a single heterozygote their effects could be diminished. However when present as homozygotes each allele may be more capable of displaying a significant effect because it is not restricted by the other allele.

To conclude, the associations between MH alleles and AGD severity reported within this study should be viewed as preliminary findings only. A second study, aimed to verify the statistical and biological significance of the suggestive associations reported herein is required. Additional research is also currently underway in an attempt to elucidate additional genes that may affect AGD resistance.

Chapter 4: Transcriptome analyses of AGD affected Atlantic salmon reveals localised host gene suppression

Published as: Wynne JW, O' Sullivan MG, Cook MT, Stone G, Nowak BF, Lovell DR, Elliott NG (2008) Transcriptome analyses of amoebic gill disease affected Atlantic salmon (*Salmo salar*) tissues reveal localised host gene suppression. Mar Biotechnol *in press.*

Abstract

The transcriptome response of Atlantic salmon (*Salmo salar*) displaying advanced stages of amoebic gill disease (AGD) was investigated. Naïve smolt were challenged with AGD for 19 days, at which time all fish were euthanised and their severity of infection quantified through histopathological scoring. Gene expression profiles were then compared between heavily infected and naïve individuals using a 17K Atlantic salmon cDNA microarray with real-time quantitative RT-PCR (qPCR) verification. Expression profiles were examined in the gill, anterior kidney and liver. Twenty-seven transcripts were significantly differentially expressed within the gill, with 20 of these transcripts being down-regulated in the AGD affected individuals compared to naïve individuals. In contrast, only nine transcripts were significantly differentially expressed within the anterior kidney, and five within the liver. Again the majority of these transcripts were down-regulated within the diseased individuals. A down-regulation of transcripts involved in apoptosis (procathepsin L, cathepsin H precursor and cystatin B) was observed in AGD affected Atlantic salmon. Four transcripts encoding genes with antioxidant properties were also down-regulated in AGD affected gill tissue according to qPCR analysis. The most up-regulated transcript within the gill was an unknown expressed sequence tag (EST) whose expression was 218 fold (± SE 66) higher within the AGD affected gill tissue. Our results suggest Atlantic salmon experiencing advanced stages of AGD demonstrate general down-regulation of gene expression, which is most pronounced within the gill. We propose that this general gene suppression is parasite mediated, thus allowing the parasite to withstand or ameliorate the host response.

4.1 Introduction

AGD is the most important health issue affecting the culture of Atlantic salmon (*Salmo salar*) in Tasmania, Australia (Munday et al. 2001). This disease is caused by the protozoan *Neoparamoeba perurans* which, acting as the primary pathogen, infects the gills of marine cultured Atlantic salmon (Young et al. 2007). Fish infected with AGD often display lethargy, respiratory distress, rapid opercular movement, and ultimately, if not treated, death (Munday et al. 1990). AGD affects species other than Atlantic salmon: cases have also been reported in rainbow trout (*Oncorhynchus mykiss*) (Munday et al. 1990); turbot (*Scophthalmus maximus*) (Dyková et al. 1999); coho salmon (*Oncorhynchus kisutch*) (Kent et al. 1988); and seabass (*Dicentrarchus labrax*) (Dyková et al. 2000).

AGD is characterised initially by attachment of trophozoites to healthy gill epithelial tissue (Adams and Nowak 2003). Following attachment, localised host cell alterations are observed, including hypertrophy and desquamation of epithelial cells followed by hyperplasia and oedema (Adams and Nowak 2003). Hyperplasia and oedema of the epithelial cells generally increase with disease progression, as does the formation of interlamellar vesicles (Adams and Nowak 2001; Adams and Nowak 2003). At later stages, moderate infiltration of leucocytes is observed within the central venous sinus adjacent to hyperplastic lesions and sometimes within the interlamellar vesicles (Adams and Nowak 2001; Adams and Nowak 2004a). In some cases, leucocytes are also found in close association with amoeba within the hyperplastic tissue (Bridle et al. 2003).

Atlantic salmon infected with AGD demonstrate an up-regulation of the pro-inflammatory cytokine interleukin-1β (IL-1β) within affected gill tissue (Bridle et al. 2006a; Morrison et al. 2007). Recombinant IL-1β has been shown to induce expression of a number of genes within the rainbow trout macrophage cell line RTS-11 (Martin et al. 2007b), many of which are involved with the acute phase response. AGD affected Atlantic salmon however, do not display an up-regulation of genes encoding the acute phase proteins serum amyloid A or serum amyloid P-like pentraxin (Bridle et al. 2006a). Furthermore the expression of tumor necrosis factor alpha (TNF- α), interferon gamma (IFN-γ) (Morrison et al. 2007) and inducible nitric oxide synthesis (iNOS) (Bridle et al. 2006a) appears unaffected by AGD in Atlantic salmon. These results suggest, while IL-1β is up-regulated within AGD affected gill tissue its ability to induce an acute phase response is restricted.

At present, the only successful treatment available for AGD is freshwater bathing (Parsons et al. 2001). Bathing for three hours in freshwater has been shown to reduce gill pathology and the number of attached trophozoites (Adams and Nowak 2004b). Freshwater bathing, however, is not a viable long term solution due to the cost associated with this practice. Recent research has shown a significant genetic component of variability in AGD resistance or susceptibility within the Tasmanian Atlantic salmon population (Taylor et al. 2007). While a small proportion of this resistance may be attributed to variation within the major histocompatibility (MH) class II genes (Wynne et al. 2007b), the molecular mechanisms controlling the majority of this resistance are yet to be elucidated. The identification of genes differentially expressed between AGD affected and naïve individuals - at a time when considerable

variation in resistance is observed - may assist in identifying candidate genes associated with genetic resistance.

DNA microarrays have become an important research tool for assessing transcriptional changes in a range of aquatic organisms, including microalgae (Lidie et al. 2005), crustaceans (Wang et al. 2006), molluscs (Jenny et al. 2007) and teleosts (von Schalburg et al. 2005b). In particular, in salmonoids, a number of microarray platforms have been developed (Rise et al. 2004b; Krasnov et al. 2005; von Schalburg et al. 2005b; Martin et al. 2007a). These platforms have been utilised to examine the response to such processes as stress (Krasnov et al. 2005), bacterial infection (Rise et al. 2004a; Ewart et al. 2005), maturation (von Schalburg et al. 2005a), vaccination (Purcell et al. 2006) and cytokine stimulation (Martin et al. 2007a; Martin et al. 2007b).

While previous studies have provided valuable insight into the early and intermediate gene expression response to AGD (Bridle et al. 2006a; Bridle et al. 2006b; Morrison et al. 2006a), analysis of the transcriptome at later stages of disease and in other organs is needed. The present study aimed to address that need by examining the transcriptome response in the gill, liver and anterior kidney of AGD affected Atlantic salmon at 19 days post inoculation (DPI).

4.2 Materials and Methods

4.2.1 Animals and disease challenge trial

This study used individuals from an AGD challenge trial (Taylor et al. 2007) (Chapter 3). Atlantic salmon smolts derived from 30 full-sibling families were sampled within the study. These families had never been selected for AGD resistance or any other production traits. Each family was initially represented by 18 individuals equally divided across three replicate 3000 L recirculating systems, equipped with primary and biological filtration, a heat exchanger and aeration. All individuals were initially maintained in freshwater and then acclimated to a salinity of 35 ppt and a temperature of 16°C over a two week period. Following acclimation and just prior to challenge, two individuals were randomly selected from each tank and euthanised with 5 g L^{-1} Aqui-S (Aqui-S NZ Ltd, Lower Hutt, New Zealand). These individuals represented the naïve samples within our experimental design. Samples were dissected from the anterior kidney, liver and the first right gill hemibranch and stored in RNA*later* (Ambion, Austin, USA) at -80°C. The remaining gill cage was then dissected and fixed in seawater Davidson's fixative for histology.

Following acclimation, each tank was inoculated with 500 cells L^{-1} of *Neoparamoeba* spp. as described by Morrison et al. (2004). Individuals were monitored daily for AGD mortalities until 19 DPI when all individuals were euthanised with 5 g L^{-1} Aqui-S (Aqui-S NZ Ltd) and samples taken as described above.

4.2.2 Gill histopathology

The second left anterior hemibranch of each individual was dissected, dehydrated, embedded in paraffin wax and sectioned at 5 μm. Sections were stained with haematoxylin and eosin and viewed with a light microscope (Zeiss, Oberkochen, West Germany) at 400X magnification. The severity of AGD infection was quantified for each individual by counting the percentage of gill filaments (primary lamellae) displaying hyperplastic lesions as previously described by Adams and Nowak (2004b). All sections were also examined for the presence of *Neoparamoeba* spp. in order to diagnose clinical AGD.

4.2.3 Microarray experimental design

This study used a 16,950 feature Atlantic salmon cDNA microarray developed by the TRAITS (Transcriptome Analysis of Important Traits in Salmon) and Salmon Genome Projects (Martin et al. 2007a) and was designed to follow MIAME guidelines (Brazma et al. 2001). Our study made direct comparisons between AGD affected and naïve Atlantic salmon, in an attempt to identify genes differentially expressed between these two states.

For all three organs (gill, anterior kidney and liver), one naïve individual (6 total) was directly compared with two AGD affected individuals (12 total) in two independent hybridisations. The naïve individual was labelled with Cy3 for the first hybridisation and then Cy5 for the second hybridisation. Thus, dye-swap technical replicates were

performed for naïve individuals only. This 1:2 comparison was performed for all six naïve and 12 AGD affected individuals, therefore utilising 12 arrays for each organ. No RNA samples were pooled and comparisons were performed on an individual tank basis. Furthermore, individuals examined in the microarray analysis were derived from six of the 30 full-sibling families. All hybridisations and protocols have been deposited into the ArrayExpress database (www.ebi.ac.uk/microarray-as/aer) and assigned the accession number E-MEXP-1286.

4.2.4 RNA isolation

Total RNA was isolated from approximately 100 mg of tissue using the RiboPure™ kit according to the manufacturer's protocol (Ambion). Following isolation, RNA was treated with rDNase I from the DNA-*free*™ kit (Ambion) to remove possible genomic DNA contamination. The ratio of absorbance at 260 nm and 280 nm was then quantified using a Beckman Coulter DU®530 spectrophotometer (Beckman Coulter, Fullerton, USA). Next, 2 µg of total RNA was denatured in 15 µl of NorthernMax® Formaldehyde Load Dye (Ambion) and 20 µg ml^{-1} ethidium bromide at 65°C for 15 min. Samples were then run on a native TAE agarose gel and visualised under UV transillumination.

4.2.5 Microarray hybridisation

Total RNA was quantified using the Qubit fluorescence quantification system as per the manufacturer's protocol (Invitrogen, Mount Waverly, Australia). Next, 3 µg of total RNA was reverse transcribed using Superscript II Reverse Transcriptase (Invitrogen) and the 3DNA Array 350™ kit (Genisphere, Hatfield, USA) as per the manufacturer's protocol. cDNA synthesis was terminated by addition of 3.5 µl of 0.5 M NaOH/50 mM EDTA and incubation at 65°C for 15 min. Reactions were neutralised with 5 µl of 1 M tris-HCl (pH 7.5). Appropriate Cy3 and Cy5 reactions were then combined and purified using the QIAquick PCR purification kit (Qiagen, Doncaster, Australia) with one additional wash of 500 µl of 80% ethanol and cDNA was eluted in 30 µl of nuclease-free water.

All hybridisations were performed with the Maui® hybridisation system (BioMicro® Systems, Salt Lake City, USA) using the FL hybridisation mixing chambers (BioMicro® Systems). First, concentrated cDNA was combined with 2 µl LNA dT blocker (Genisphere) and 30 µl of 2X formamide-based hybridisation buffer (Genisphere) and incubated at 80°C for 10 min and then at 49°C. Next, 50 µl of the cDNA hybridisation mix was applied to each pre-warmed microarray slide. The hybridisation was then performed for 15 hours at 49°C. Following hybridisation, the chambers were removed while submerged in 2X SSC/0.2% SDS at 49°C. Next, slides were washed once with 2X SSC/0.2% SDS at 49°C for 10 min, twice with 2X SSC/0.2% SDS at room temperature for 5 min, twice with 2X SSC at room temperature for 5 min, then twice with 0.2X SSC at room temperature for 5 min. All washes were performed with agitation. Slides were centrifuged dry at 510*g* for 5 min at 10°C.

The Cy3 and Cy5 cyanine capture reagents (Genisphere) were hybridised to the microarray slides by combining 2.5 µl of both the Cy3 and Cy5 capture reagents with 20 µl nuclease free water and 25 µl of 2X formamide-based hybridisation buffer (Genisphere) and incubated at 80°C for 10 min and then at 49°C. Fifty microlitres of the cyanine hybridisation solution was then applied to the microarray slides and hybridised at 49°C for 2 hours. Following hybridisation, the chambers were removed in 2X SSC/0.2% SDS at 49°C and slides then were washed and dried as described above.

4.2.6 Image and data analysis

Each slide was immediately scanned at a resolution of 5 µm with an Axon 4000B scanner (Axon Instruments, Union City, USA) using the software GenePix Pro ver3.0 (Molecular Devices Corporation, Sunnyvale, USA). The Cy3 and Cy5 fluorphores where excited at 532 and 635 nm respectively. Power was consistently maintained at 100% and the photo multiplier tube (PMT) balance was identical for all slides. PMT gain for the 532 nm channel was 574 and for the 635 nm channel was 663. GAL files were aligned automatically and fluorescence signal intensity data for the 532 and 635 nm channels extracted. Results were saved as GenePix results (GPR) files.

GPR files were read into the statistical software package R (ver2.3.1) and analysed with Limma (Smyth 2005). Prior to statistical analysis the data were normalised with the NormalizeWithinArrays function using the default print-tip loess method (Smyth and Speed 2003). The moderated t-test statistic, in conjunction with the 'holm' multiple testing adjustment, was used to detect differential expression between heavily infected

and naïve individuals (Speed 2003). Background correction was performed on the data using the default method within Limma (Smyth 2005) . Filtering, however, was applied to remove all features with an average intensity < 97 (i.e., features where both channels had low intensity) – this resulted in approximately 20% of features being removed.

The experimental design used was complex as it incorporated technical replication for the naïve fish only: each naïve fish was compared to two AGD affected individuals. The analysis of such a design is not straightforward and, depending on assumptions made, different analyses would be appropriate. A decision was made to fit three separate models to accommodate different scenarios, each with a slightly different approach to the treatment of replication of naïve individuals. Model 1 ignores technical replication of naïve fish; model 2 incorporates correlation for replicates (utilises the duplicateCorrelation function (Smyth et al. 2005)); and model 3 uses an averages over fixed effects approach (treats the naïve individuals as a fixed effect - no assumption of same correlation for all spots and estimates the residual errors differently).

For each model, transcripts were deemed to be differentially expressed if both replicate features had significant t-statistics (adjusted $p<0.05$). The mean adjusted p-value and fold-change were then calculated. In the case where the same transcripts were significant across multiple models, the mean adjusted p-value for lowest model number is presented within the Table (the fold-change was unaffected by the model).

4.2.7 Transcript annotation

Sequences of differentially expressed transcripts were first subjected to Basic Local Alignment Search Tool (BLASTx) against the NCBI non-redundant protein databases in the program Blast2GO (Conesa et al. 2005). Sequences with *E* values $<1\times10^{-10}$ were deemed to be orthologous. If no significant homology was detected, the sequence was further analysed (using BLASTn) against the NCBI expressed sequence tag database (dbEST) in an attempt to identify UniGene sequence clusters (www.ncbi.nlm.nih.gov/sites/entrez?db=unigene). Functional proteins associated with UniGene clusters were then used for annotation. In the case where no proteins were associated with the UniGene clusters, the UniGene code and the annotation "unknown" was assigned. Where no significant BLAST or UniGene sequence was identified the transcript was designated "no BLAST hit."

Identified transcripts were further annotated using Gene Ontology (GO) within the program Blast2GO (Conesa et al. 2005). When possible, the upper level parent GO term was assigned to transcripts for each of the following ontologies: cellular component, biological process and molecular function. Transcripts differentially expressed within the gill were then grouped based on these upper level GO terms for cellular component, biological process and molecular function.

4.2.8 Real-time quantitative RT-PCR

Two independent real-time quantitative RT-PCR (qPCR) experiments were conducted. The first aimed to verify the microarray results and the second aimed to compare the

expression of a sub-sample of genes between fish with differing severities of AGD infection.

To substantiate the microarray results a selection of transcripts from the gill, anterior kidney and liver gene lists were selected for qPCR analysis. Eight transcripts in total were selected. The expression of the eight transcripts was compared between the six naïve and 12 AGD affected Atlantic salmon previously compared in the microarray analysis. The RNA was obtained from the previous extraction.

The expression of a sub-sample of genes within the gill was also compared between fish displaying differing severities of AGD infection. These individuals represented the extremes of the severity distribution and were derived from a range of families. Eleven individuals with low AGD severity and 12 individuals with high AGD severity were chosen for analysis. The mean percentage of gill filaments infected with AGD lesions was 13.05 ± 0.77 and 58.97 ± 5.32 for the groups considered to have low and high disease severities, respectively.

Primers were designed from the expressed sequence tags using PrimerExpress (Applied Biosystems). When the genomic organisation could be determined (through PCR with genomic DNA), primers that crossed exon-exon junctions were designed. All primers used for the qPCR are shown in Table 4.1. The housekeeping gene beta (β)-actin was chosen as an endogenous control using the primers obtained from Bridle et al. (2006a). cDNA was generated from total RNA using Superscript III Reverse Transcriptase. Briefly, 2 µg of total RNA was quantified and reverse transcribed in a 20 µl reaction

TRAITS clone ID	Identity	Organ	Amplicon size (bp)	Primer sequence (5'-3')
CK880278	UniGene-Ssa.25836	Gill	92	AATACCATCACTCATACCGCTT CCACTTTGCCACCTGTTATGT
AM049499	Differentially regulated trout protein 1	Gill	105	AGCCTTTGCCAACACAACG CCGAAGAAAGGAAAGATGAAAGC
EG648298	Glutathione S-transferase	Gill, Liver	63	CGCTGTGGCGCTATGAGAA CACCACGATCTCCTTCCATTC
EG647964	Glutathione peroxidase	Gill	64	CCGGGAAATGGCTTTGA GCGTCCTTCCCATTCACAT
CK877314	Lipase H	Gill	90	CAGGCCTCTCTGCCGTTATG GTCTCAGAAGTTGGACTCCTCACA
AM049784	Thioredoxin	Gill	62	GCCACGCTGGAGGAGAAG CAGCGTCAGTGGCAGCATTA
BI468027	Procathepsin L	Gill	68	TGGGCATGAACCACTTTGGT CTGCTTGTAGCCGTTCATCGT
CK887101	NADP transhydrogenase	Gill	60	CCATCGGTGGTGCAGACAT CCCAGCCGGAGTAGCTGTTA
AJ424551	CCAAT/enhancer-binding protein β	Anterior kidney	71	CGGAGACGCAACACAAAGTG CAGCTGCTCCACGCGTTT
-	β-actin (endogenous control)	All	113	TCTCTGGAGAAGAGCTAC CAAGACTCCATACCGAGGAA

Table 4.1. List of transcripts chosen for microarray validation using real-time quantitative RT-PCR. The clone sequence ID used to design primers is indicated along with the identity of the most significant BLAST result. The organ in which the transcript was examined is shown. PCR and cycling parameters were constant for all transcripts and are described within the text. The sense and anti-sense primer sequences are indicated along with amplicon size in base pairs (bp).

containing 4 µl of 5X buffer, 200 nM of oligo $dT_{(17)}$, 0.5 mM of dNTPs, 5 mM of DTT, 40 units of RNaseOUT and 200 units of Superscript III reverse transcriptase. Following the addition of RNA, oligo dT and dNTPs, the reaction was heated to 65°C for 5 min. The buffer, DTT, RNaseOUT and reverse transcriptase were then added and reactions were incubated at 50°C for 1 hour. Two units of RNase H were finally added and reactions were incubated for 20 min at 37°C.

All reactions were performed in triplicate in a total volume of 25 µl containing 12.5 µl of 2X SensiMix (Quantace, London, UK), 0.5 µl of SYBR® Green I, 200 nM of forward and reverse primer, 5 µl of cDNA template (equivalent to 20 ng starting RNA) and 5 µl of water. Reactions were subjected to 95°C for 10 min, followed by 40 cycles of 94°C for 20 sec, 56°C for 20 sec and 72°C for 20 sec using a Rotor-Gene™3000 real-time thermal cycler (Corbett Research, Mortlake, Australia). To verify amplification specificity, melt curves for each gene were examined from 72-95°C with a heating rate of 1°C per 5 sec. In addition, amplicons were examined by agarose gel electrophoresis and also cloned into pGEM-T-easy (Promega, Annandale, Australia) and sequenced using T7 and Sp6 primers. To check for contaminating genomic DNA, a sub-sample of reverse transcription negative reaction was performed, where Superscript III Reverse Transcriptase (Invitrogen) was omitted from the reaction and PCR performed as above.

4.2.9 qPCR data analysis

Raw fluorescence data were obtained and analysed using the Rotor-Gene Analysis Software ver6.1 (Corbett Research). Thresholds were set manually above the point

where the fluorescence exceeds the background signal. The C_T (cycle number at which the signal reaches the thresholds) were then calculated accordingly. Amplification efficiency (E) for each transcript was determined from a pooled cDNA sample serially diluted two-fold five times and calculated as $E=10^{(-1/S)}$ where s is the slope of the standard curve of serial dilutions according to Pfaffl (2001). Differences in gene expression between AGD naïve and affected Atlantic salmon were analysed using relative expression, with the software REST-XL (Pfaffl et al. 2002) according to the following equation:

$$\text{Ratio}=(E_{\text{target}})^{\Delta CT[\text{target(control-sample)}]}/(E_{\text{ref}})^{\Delta CT[\text{ref(control-sample)}]}$$

β-actin served as the endogenous reference gene (ref). Naïve Atlantic salmon were considered the control group for the microarray validation experiment. The fold-change therefore represents the expression change in AGD affected individuals relative to naïve individuals. When gene expression was compared between individuals with high and low AGD severities, the high severity group was considered the control. In this experiment, the fold-change thus represents the expression change in low AGD severity individuals relative to high AGD severity individuals. The REST-XL software calculates the relative expression ratio of a target gene, while adjusting for differences in amplification efficiency. The mean ratio of each transcript is presented along with standard errors.

4.3 Results

4.3.1 Challenge trial and gill histopathology

The challenge trial was terminated at 19 DPI in an attempt to gain as much variation in the severity of infection as possible, without allowing the most susceptible individuals to succumb to AGD. The severity of infection, determined by histopathology scoring, ranged from 0 to 85% of gill filaments infected. Significant variation in the severity of infection was observed between the families as reported by Taylor et al. (2007) (Appendix 1.1).

Histopathological examination of AGD affected gill tissue confirmed pathology characteristic of AGD including epithelial hyperplasia, oedema and interlamellar vesicle formation (Figure 4.1). The presence of amoebae with the symbiont (parasome) was observed within the margins of hyperplastic lesions. Histopathological examination of naïve gill tissue found no pathology characteristic of AGD or of any other diseases.

4.3.2 Microarray analysis

Following analysis of gene expression profiles within the gill, 36 transcripts were significantly differentially expressed between AGD affected and naïve individuals using the three statistical models (Table 4.2). These represented 29 down-regulated and seven up-regulated transcripts. Nine transcripts were significantly differentially expressed

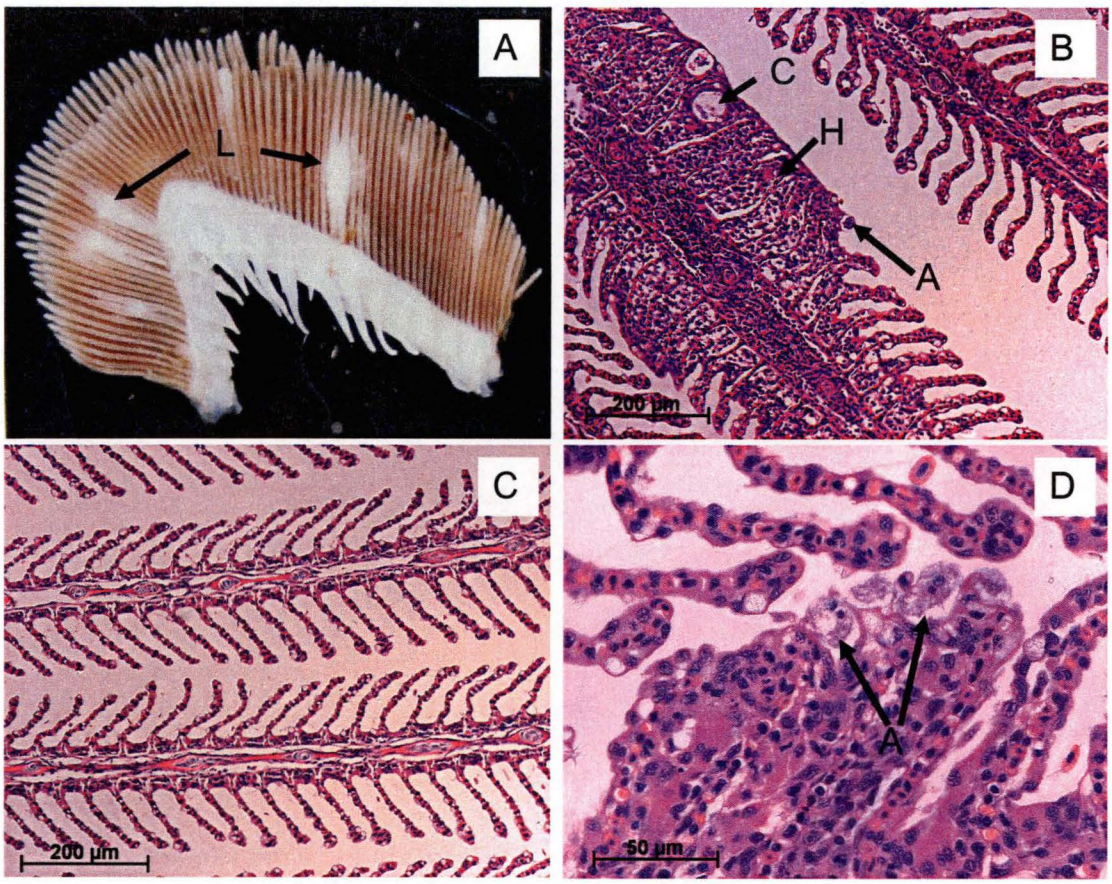


Figure 4.1. Gross and histopathological examination of AGD affected and naïve Atlantic salmon at 19 DPI. A: seawater Davidson's fixed AGD affected gill arch, multifocal mucoid lesions (L) are evident. B: histopathology of AGD affected gill tissue, interlamellar crypts (C), hyperplastic lamellar fusion (H) and the parasite *Neoparamoeba* spp. (A) are clearly visible. C: histology from gill of AGD naïve Atlantic salmon representing normal gill morphology. D: histopathology of AGD affected Atlantic salmon, *Neoparamoeba* spp. (A) are associated with the margins of hyperplastic lesions.

Accession No / Clone ID	Identity	Align / (%)	Fold-change	p value	Model	Upper level GO terms
CK880278	Unknown UniGene-Ssa.25836	N/A	19.34	2.5435E-06	1,2,3	N/A
AM049530	Differentially regulated trout protein 1 [*Oncorhynchus mykiss*] (AAG30030)	58 / 94%	-4.20	1.0164E-05	1,2,3	Immune-relevant gene (Bayne et al. 2001)
AM049483	Differentially regulated trout protein 1 [*Oncorhynchus mykiss*] (AAG30030)	88 / 95%	-4.08	2.3907E-05	1,2,3	Immune-relevant gene (Bayne et al. 2001)
CO472444	Unknown UniGene-Ssa.26266	N/A	11.88	3.3783E-05	1,2,3	N/A
AM049499	Differentially regulated trout protein 1 [*Oncorhynchus mykiss*] (AAG30030)	88 / 97%	-4.08	2.4468E-04	1,2,3	Immune-relevant gene (Bayne et al. 2001)
gil_cgi_D1E08_car_tra_sub_0p	Differentially regulated trout protein 1 [*Oncorhynchus mykiss*] (AAG30030)	88 / 95%	-3.78	0.0012	1,2,3	Immune-relevant gene (Bayne et al. 2001)
AM042205	PREDICTED: similar to es1 protein [*Danio rerio*] (XP_693678)	186 / 74%	-2.65	0.0012	1,2,3	GO:0005622 (1) intracellular
EG355156	Carcinoembryonic antigen-related cell adhesion molecule 1 [*Canis lupus familiaris*] (ABJ53204)	187 / 34%	-3.16	0.0013	1,2,3	GO:0016020 (1) membrane; GO:0007155 (2) cell adhesion
CK891045	Unknown UniGene-Ssa.24730	N/A	-3.27	0.0016	1,2,3	N/A
EG648298	Glutathione S-transferase [*Oncorhynchus nerka*] (BAA76974)	202 / 91%	-2.07	0.0023	1,2,3	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
EG647964	Glutathione peroxidase [*Danio rerio*] (AAO86703)	188 / 80%	-2.00	0.0026	1,2,3	GO:0005622 (1) intracellular; GO:0042221 (2) response to chemical stimulus; GO:0016491 (3) oxidoreductase activity
CK890208	Unnamed protein product [*Tetraodon nigroviridis*] (CAG06489)	66 / 65%	-3.31	0.0048	1,2	N/A
CK885623	Similar to villin 2 UniGene-Ssa.22680	N/A	-1.78	0.0056	1,2	GO:0016020 (1) membrane; GO:0016043 (2) cell organization and biogenesis; GO:0005515 (3) protein binding

Table 4.2. Continued over page

CK877314	Lipase member H [*Danio rerio*] (NP_001003499)	97 / 57%	2.17	0.0059	1,2,3	GO:0005576 (1) extracellular region; GO:0044238 (2) primary metabolic process; GO:0016787 (3) hydrolase activity
CK889792	PREDICTED: similar to es1 protein [*Danio rerio*] (XP_693678)	196 / 76%	-2.10	0.0064	1,2,3	GO:0005622 (1) intracellular
EG649354†	Glutathione S-transferase [*Oncorhynchus nerka*] (BAA76974)	176 / 96%	-2.26	0.0065	1,2,3	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
CK888444	Differentially regulated trout protein 1 [*Oncorhynchus mykiss*] (AAG30030)	88 / 95%	-2.52	0.0067	1,2	Immune-relevant gene (Bayne et al. 2001)
CK878970	Dynein, cytoplasmic 1, heavy chain 1 [*Homo sapiens*] (EAW81764)	149 / 85%	2.22	0.0081	1,2	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016787 (3) hydrolase activity
CK879849	Tryptase gamma 1 [*Mus musculus*] (NP_036164)	208 / 36%	-1.97	0.0111	1,2,3	GO:0016020 (1) membrane; GO:0016787 (3) hydrolase activity
CO470395	Unknown UniGene-Ssa.35702	N/A	7.08	0.0133	1,2	N/A
AM049784	Thioredoxin [*Ictalurus punctatus*] (AAG00612)	99 / 69%	-1.50	0.0153	1,2	GO:0006118 (2) electron transport; GO:0016491 (3) oxidoreductase activity
CK877639	Cystatin B [*Oncorhynchus mykiss*] (AAT98592)	42 / 92%	-1.71	0.0193	1,2	GO.0005622 (1) intracellular; GO:0050789 (2) regulation of biological process; GO:0004857 (3) enzyme inhibitor activity
CK884070	Glutathione peroxidase type 2 [*Oncorhynchus mykiss*] (AAV32968)	88 / 84%	-1.89	0.0250	1,2	GO:0005622 (1) intracellular; GO:0042221 (2) response to chemical stimulus; GO:0016491 (3) oxidoreductase activity
BI468027	Procathepsin L [*Oncorhynchus mykiss*] (AAK69706)	174 / 98%	-1.55	0.0277	1,2	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016787 (3) hydrolase activity
CK895896	Cystatin B [*Oncorhynchus mykiss*] (AAT98592)	98 / 61%	-1.67	0.0323	1,2	GO:0005622 (1) intracellular; GO:0050789 (2) regulation of biological process; GO:0004857 (3) enzyme inhibitor activity
CK887577	Cathepsin H precursor [*Fundulus heteroclitus*] (AAO64473)	167 / 72%	-1.62	0.0102	2	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016787 (3) hydrolase activity

Table 4.2. Continued over page

AM042158	C1q-like adipose specific protein [*Danio rerio*] (XP_689670)	100 / 47%	3.62	0.0261	2	GO:0016020 (1) membrane; GO:0002376 (2) immune system process GO:0005515 (3) protein binding
CK898222	Argininosuccinate synthetase [*Danio rerio*] (NP_001004603)	177 / 86%	-1.54	0.0279	2	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016874 (3) ligase activity
CK885691	Foxk2 protein [*Danio rerio*] AAI29152	107 / 64%	-1.75	0.0414	2	N/A
DW590853‡	No BLAST hit	N/A	-3.35	0.0583	3	N/A
CK880525‡	Unknown UniGene-Ssa.33252	N/A	-4.52	0.0190	3	N/A
CK884492‡	Similar to mitogen-activated protein kinase UniGene-Ssa.709	N/A	-4.37	0.0267	3	GO:0050790 (2) regulation of catalytic activity; GO:0005515 (3) protein binding
CK875582‡	Unknown UniGene-Ssa.589	N/A	-3.76	0.0209	3	N/A
DW591218‡	Monoacylglycerol O-acyltransferase 1 [*Xenopus tropicalis*] (NP_001039202) .	111 / 70%	-4.43	0.0350	3	GO:0016020 (1) membrane; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
CK879686	Glutathione S-transferase P [*Cricetulus longicaudatus*] (P46424)	131 / 61%	-2.30	0.0402	3	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
CK878339	Myelin and lymphocyte protein [*Mus musculus*] (NP_034892)	101 / 52%	3.41	0.0477	3	GO:0016020 (1) membrane; GO:0050789 (2) regulation of biological process; GO:0005215 (3) transporter activity

Table 4.2. Legend over page

Table 4.2. List of transcripts significantly differentially expressed within the gill of AGD affected Atlantic salmon. Clone ID (accession number or TRAITS ID) is indicated, along with the identity of the most significant BLAST result, the accession number and the [species]. The number of aligned amino acids and similarity (%) between the transcript and its BLAST hit is shown for those transcripts with a significant ($E<1\times10^{-10}$) BLASTx result. The fold-change and adjusted p value is also indicated. The model number in which the transcript was significant (adjusted $p<0.05$) is displayed. ‡ indicates transcripts also significantly differentially expressed within the anterior kidney, † indicates transcripts also significantly differentially expressed within the liver. Gene Ontology (GO) terms represent (1) cellular component, (2) biological progress and (3) molecular function

within the anterior kidney of AGD affected salmon, with eight transcripts down-regulated and one up-regulated (Table 4.3). A total of five transcripts were significantly differentially expressed within the liver of AGD affected salmon, with all but one transcript down-regulated (Table 4.4).

BLAST analysis of the 36 transcripts from the gill list suggests that these sequences represent 27 unique transcripts, one of which had no significant sequence homology (designated no BLAST hit). Of the 26 identified transcripts, eight were recognised through UniGene clusters, however, six of these clusters had no functional protein assigned to them (transcripts were therefore designated unknown) (Table 4.5). The most significantly up-regulated transcript identified within the gill was the UniGene cluster Ssa.25836. One of most highly down-regulated transcripts in the gill was differentially regulated trout protein 1 (DRTP1). This gene was represented by five transcripts and showed a down-regulation between 2.52 and 4.20 fold within AGD affected gill. Other significant transcripts represented by more than one clone were sequences displaying strong homology to glutathione S-transferase, glutathione peroxidase, a gene coding for es1 protein and cystatin B.

Of the nine transcripts differentially expressed within the anterior kidney, two had no significant sequence homology using BLAST and four had homology to non-annotated UniGene clusters. The only transcript significantly up-regulated appeared to be homologous to the CCAAT/enhancer-binding protein beta (C/EBPβ). As observed in the gill, transcripts with homology to mitogen-activated protein kinase and monoacylglycerol O-acyltransferase were also down-regulated.

Accession No / Clone ID	Identity	Align / (%)	Fold-change	p value	Model	Upper level GO terms
AJ424551	CCAAT/enhancer-binding protein β [*Oncorhynchus mykiss*] (AAN41660)	71 / 98%	2.39	0.0019	1,2	GO:0005622 (1) intracellular; GO:0006952 (2) defence response; GO:0003700 (3) transcription factor activity
CK888465	Unknown UniGene-Ssa.2141	N/A	-2.70	0.0033	1,2,3	N/A
CK880525‡	Unknown UniGene-Ssa.33252	N/A	-6.09	0.0260	1,2	N/A
DW591218‡	Monoacylglycerol O-acyltransferase 1 [*Xenopus tropicalis*] (NP_001039202)	111 / 70%	-5.92	0.0293	1,2	GO:0016020 (1) membrane; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
CK884492‡	Similar to mitogen-activated protein kinase UniGene-Ssa.709	N/A	-5.93	0.0324	1,2	GO:0050790 (2) regulation of catalytic activity; GO:0005515 (3) protein binding
DW590853‡	No BLAST hit	N/A	-4.88	0.0422	1,2	N/A
CK875582‡	Unknown UniGene-Ssa.589	N/A	-5.35	0.0484	2	N/A
CK888829	Unknown UniGene-Ssa.31031	N/A	-3.42	0.0412	3	N/A
CK887535	No BLAST hit	N/A	-2.52	0.0410	3	N/A

Table 4.3. List of transcripts significantly differentially expressed within the anterior kidney of AGD affected Atlantic salmon. Clone ID (accession number or TRAITS ID) is indicated, along with the identity of the most significant BLAST result, the accession number and the [species]. The number of aligned amino acids and similarity (%) between the transcript and its BLAST hit is shown for those transcripts with a significant ($E<1\times10^{-10}$) BLASTx result. The fold-change and adjusted p value is also indicated. The model number in which the transcript was significant (adjusted $p<0.05$) is displayed. ‡ indicates transcript also significantly differentially expressed within the gill. Gene Ontology (GO) terms represent (1) cellular component, (2) biological progress and (3) molecular function.

Accession No / Clone ID	Top BLAST gene name	Align / (%)	Fold-change	p value	Model	Upper level GO terms
CK896995	No BLAST Hit	N/A	-2.98	0.0261	1,2	N/A
BM413998	C1q-like protein [*Dissostichus mawsoni*] (ABN45966)	104 / 57%	-2.19	0.0351	1,2	GO:0016020 (1) membrane; GO:0002376 (2) immune system process GO:0005515 (3) protein binding
gil_cgi_E3H04_car_tra_sub_0p	Translation initiation factor-3 subunit 5 [*Mus musculus*] (AAL38054)	43 / 95%	2.14	0.0268	1,2	GO:0006413 (2) translational initiation
EG649354†	Glutathione S-transferase [*Oncorhynchus nerka*] (BAA76974)	176 / 96%	-1.75	0.0160	2	GO:0005622 (1) intracellular; GO:0044237 (2) cell metabolism; GO:0016740 (3) transferase activity
CK896685	Uncharacterised protein C6orf58 [*Oncorhynchus mykiss*] (Q5QT17)	193 / 86%	-2.21	0.0288	2	N/A

Table 4.4. List of transcripts significantly differentially expressed within the liver of AGD infected Atlantic salmon. Clone ID (accession number or TRAITS ID) is indicated, along with the identity of the most significant BLAST result, the accession number and the [species]. The number of aligned amino acids and similarity (%) between the transcript and its BLAST hit is shown for those transcripts with a significant ($E<1\times10^{-10}$) BLASTx result. The fold-change and adjusted p value is also indicated. The model number in which the transcript was significant (adjusted $p<0.05$) is displayed. † indicates transcripts also significantly differentially expressed within the gill. Gene Ontology (GO) terms represent (1) cellular component, (2) biological progress and (3) molecular function.

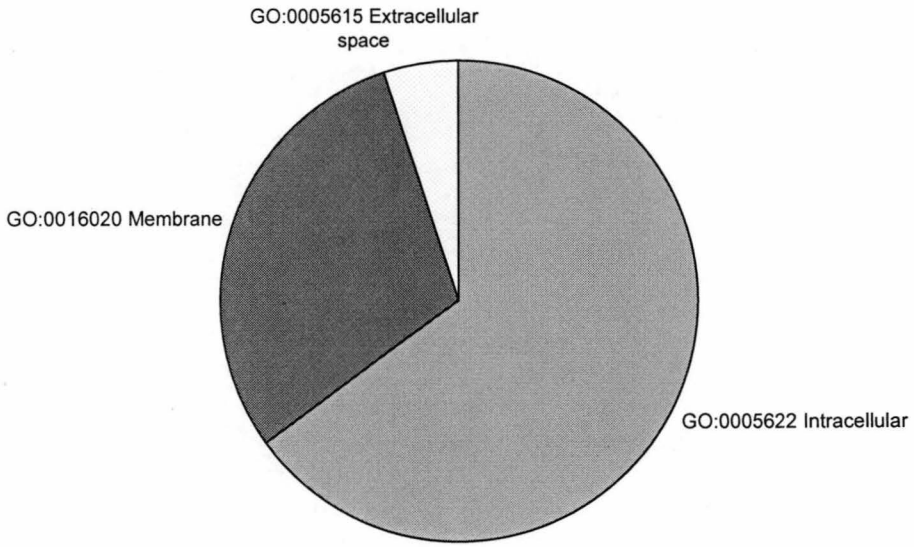
	Total		BLAST hit ($<1\times10^{-10}$)	UniGene cluster	No BLAST hit
	Up-regulated	Down-regulated			
Gill	7	20	26	8	1
Anterior kidney	1	8	7	5	2
Liver	1	4	4	0	1

Table 4.5. The proportion of transcripts up and down-regulated within the gill, anterior kidney and liver. The number of transcripts with significant BLAST identities, UniGene clusters or no BLAST hits are summarised.

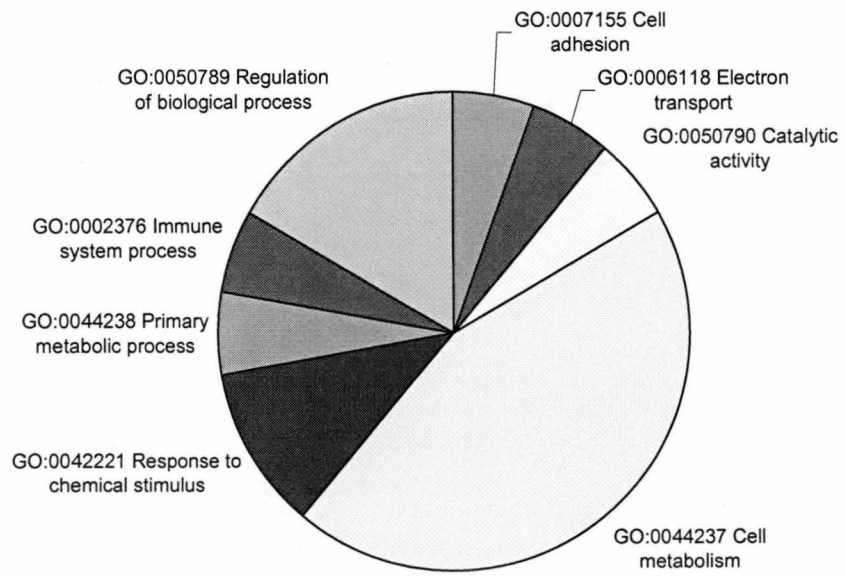
The liver showed the least number of differentially expressed transcripts between AGD affected and naïve individuals. The degree of differential expression was also lower. BLAST analysis of the five significantly differential expressed transcripts indicated that they represent five unique genes with one no BLAST hit. Only a single transcript that showed high homology with translation initiation factor 3 subunit 5 was significantly up-regulated. As observed within the gill, a transcript with strong homology to glutathione S-transferase was down-regulated.

Grouping the gill differentially expressed transcripts by parental gene ontology terms suggests that the majority of informative transcripts assigned cell component terms served as intracellular (GO:0005622) cell components (Figure 4.2). A smaller number of transcripts were associated with extracellular (GO:0005615) and membrane (GO:0016020) cell components. Grouping the informative transcripts' assigned biological processes resulted in a diverse range of parent terms: the most common processes being cell metabolism (GO:0044237) and regulation of biological process (GO:0050789). Grouping parent terms of informative transcripts' assigned molecular functions indicates that the majority of these are involved in activities such as oxidoreductase (GO:0016491), transferase (GO:0016740) and hydrolase (GO:0016787). A large number of transcripts were also grouped by protein binding activities (GO:0005155).

A


GO:0005615 Extracellular space
GO:0016020 Membrane
GO:0005622 Intracellular

B


GO:0007155 Cell adhesion
GO:0050789 Regulation of biological process
GO:0006118 Electron transport
GO:0050790 Catalytic activity
GO:0002376 Immune system process
GO:0044238 Primary metabolic process
GO:0042221 Response to chemical stimulus
GO:0044237 Cell metabolism

C

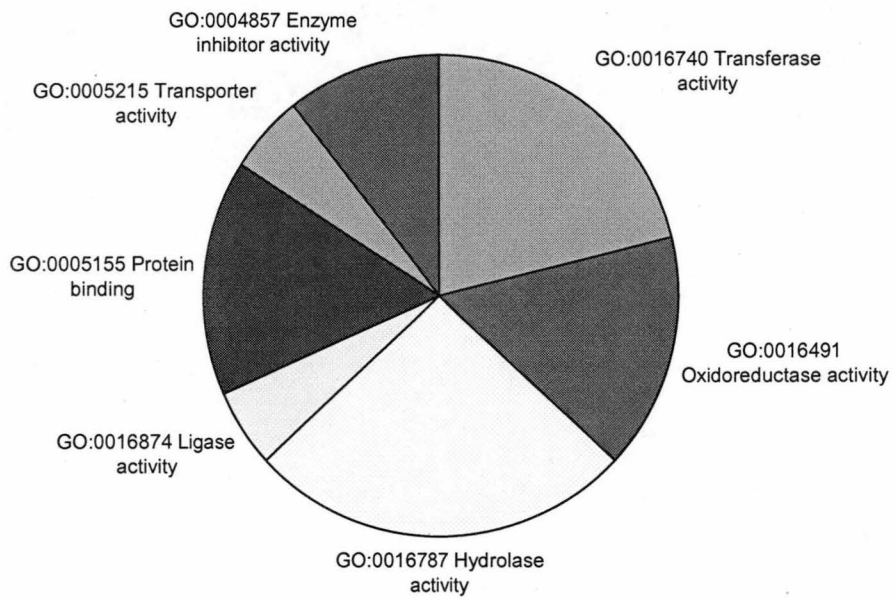


Figure 4.2. Gene Ontology annotation of transcripts differentially expressed within the gill. Gene Ontology assignment was possible for 61% of transcripts. A: frequency of parent terms from transcripts assigned cell component GO terms. B: frequency of parent terms from transcripts assigned biological process GO terms. C: frequency of parent terms from GO transcripts assigned molecular function child terms.

4.3.3 qPCR verification of microarray results

Nine transcripts identified from the microarray analysis were subjected to qPCR to further investigate their expression within our sample material (Figure 4.3). While NADP transhydrogenase was not identified as significantly differentially expressed within the microarray experiment, its involvement in the thioredoxin and glutathione pathways justified its additional examination with qPCR.

Melt analysis demonstrated single melt curves for all amplicons confirming amplification of a single product. Moreover, cloning and sequencing of qPCR products demonstrated that the transcript of interest was being amplified. As reported previously (Bridle et al. 2006b) expression of the endogenous control, β-actin, did not appear to be influenced by AGD. No amplification was observed within any RT negative reactions.

The fold-change (± standard errors) determined through microarray and qPCR analysis for all eight transcripts examined in the gill is depicted in Figure 4.3. The transcript showing homology to Lipase H showed disagreement according to qPCR and microarray analysis. It appears when measured with qPCR, this transcript is actually down-regulated in AGD affected individuals – rather than up-regulated – as indicated by the microarray analysis. This discrepancy may indicate cross-hybridisation effects on this feature. All other transcripts examined in the gill with qPCR showed agreement in terms of up and down-regulation.

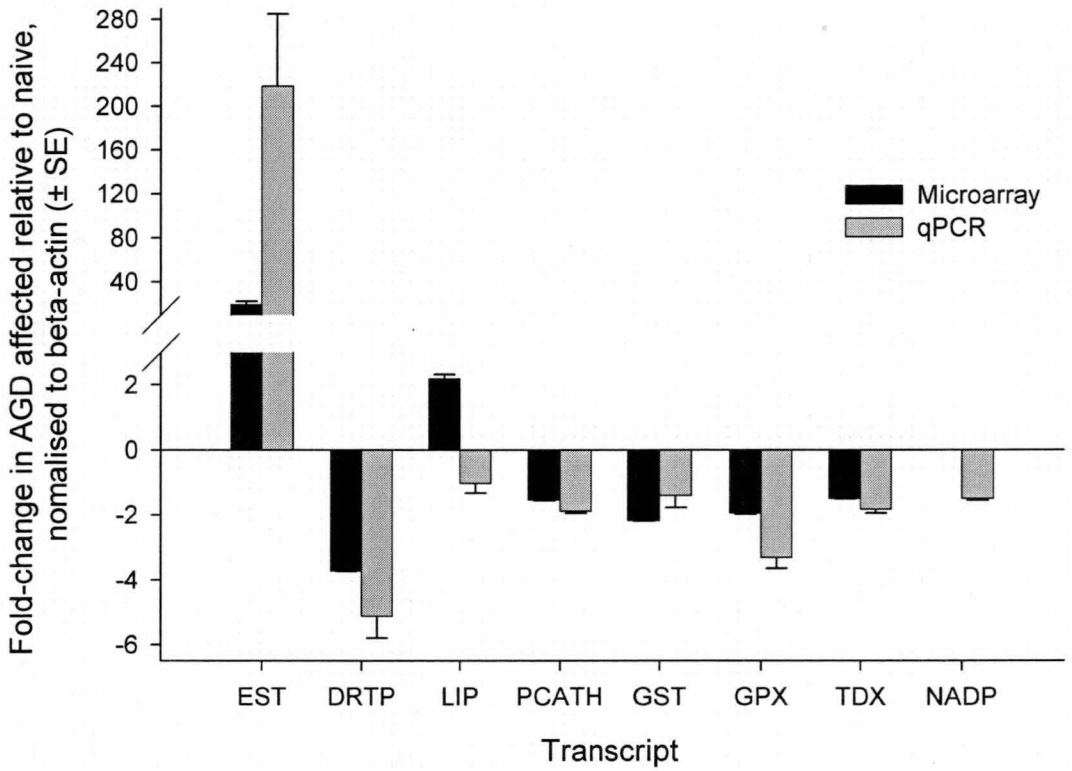


Figure 4.3. Fold-change (± standard errors) values relative to naïve and normalised to β-actin of all transcripts verified in the gill by real-time quantitative RT-PCR (qPCR) in AGD affected Atlantic salmon. EST = expressed sequence tag UniGene-Ssa.25836, DRTP = differential regulated trout protein 1, LIP = lipase H, PCATH = procathepsin L, GST = glutathione S-transferase, GPX = glutathione peroxidase, TDX = thioredoxin, NADP = NADP transhydrogenase. Values between -1 and +1 indicate no differential expression.

With the exception of Lipase H the only other transcript to show disagreement between qPCR and microarray analysis was glutathione S-transferase when examined in the liver. This transcript was found to be up-regulated by 1.07 (± SE 0.30) fold according to the qPCR analysis, however, it was down-regulated by -1.75 (± SE 0.02) in the microarray analysis. There were some discrepancies between the magnitude of the fold-change between the microarray and qPCR results. For instance, C/EBPβ within the anterior kidney of infected fish was up-regulated 2.39 (± SE 0.18) fold in the microarray analysis, but only 1.57 (± SE 0.45) fold up-regulated when examined with qPCR. The unknown EST (UniGene-Ssa.25836) was the most up-regulated transcript in regard to both the microarray and the qPCR results. Data from the microarray analysis suggest that this transcript was up-regulated 19.34 (± SE 3.04) fold, however closer examination with qPCR demonstrated this transcript was 218 (± SE 66) fold up-regulated in AGD affected gill tissue (Figure 4.3). The reason for these discrepancies is probably due to the greater dynamic range displayed by qPCR compared to microarrays.

4.3.4 qPCR analysis of fish with different levels of AGD severity

The gill gene expression of four transcripts was also compared between fish showing high and low severities of AGD infection, as quantified through histopathological scoring. Compared to the heavily infected group, expression of the unknown EST (UniGene-Ssa.25836) showed a significant down-regulation within lower severity group ($p<0.001$). Fish with less severe cases of AGD showed a suppression of this transcript by 4.57 (± SE 0.72) fold compared to fish with high disease severities. No

significant difference in expression between groups was observed for DRTP, thioredoxin or procathepsin L ($p>0.05$).

4.4 Discussion

This study aimed to examine the transcriptome response to AGD at a time when considerable variation in the severity of infection was present between individuals. An unexpected finding was the general down-regulation of genes within the gill – and to a lesser extent within the anterior kidney and the liver – of AGD affected Atlantic salmon. This result is in contrast to the gill transcriptome response during the early stages of AGD infection (44 to 189 hours), where, at all time points, a greater number of genes were up-regulated than down-regulated (Morrison et al. 2006a). The late stages of AGD are typically characterised by extensive lesion formation on the gill, with numerous amoeba associated with the margins of these lesions (Adams and Nowak 2003). Our study examined fish at 19 DPI, a time when advanced signs of AGD were present. The contrast between the transcriptome response reported in this and previous studies (Morrison et al. 2006a) possibly reflects the different time points examined.

The present study demonstrated that once individuals become highly infected with AGD and thus approach the terminal stages of disease, a down-regulation of many genes occurs. We believe this gene suppression may be mediated by the parasite in an attempt to restrict the immune response and thus increase parasite survival. The fact that the most pronounced gene suppression was at the site of infection – the gill – rather than the liver or anterior kidney adds strength to this hypothesis. For instance, if the gene suppression was simply a consequence of the individual approaching death we would expect to see systemic gene suppression, rather than this demonstrated local response. Nevertheless, it must be conceded that the extreme level of parasitism would have

resulted in a extreme stress response which may have influenced the results obtained. Indeed if this stress response affected oxygen consumption the down-regulation of gene expression may indeed be most pronounced in the gill compared to the anterior kidney and liver. Future transcriptome profiling experiments aimed at assessing the generally stress response in Atlantic salmon – especially in regard to respiratory stress – may provided valuable information concerning this hypothesis.

Suppression of the immune response to AGD has also been reported in other studies. Indeed, Atlantic salmon have been shown to display decreased phagocyte function in the form of suppressed respiratory burst response during advanced stages of AGD (Gross et al. 2004b; Gross et al. 2005). Immunosuppression caused by both intracellular and ectoparasitism has been reported previously and is suggested as an important process contributing to the interaction and coevolution between parasites and their hosts (Yang and Cox-Foster 2005). Indeed, a general down-regulation in gene expression is observed within macrophages infected with *Leishmania donovani* (Ross), the causative agent of human leishmaniasis (Buates and Matlashewski 2001). Moreover, varroa mites are capable of causing immunosuppression in the honey bee, *Apis mellifera* (L.) (Yang and Cox-Foster 2005). The tick *Ixodes scapularis* (Say) is also capable of causing immunosuppression by releasing a cystatin within the parasite's saliva, which inhibits proliferation of cytotoxic T lymphocytes and decreases cathepsin L activity (Kotsyfakis et al. 2006). Two cathepsin transcripts (procathepsin L and cathepsin H precursor) were also down-regulated within our study.

The gill is the primary and only site of infection by *Neoparamoeba perurans* and displays an extensive cellular response to the pathogen. While a large transcriptome

response was observed at the site of infection – albeit gene suppression – an interesting finding was the lack of differential expression within the immunologically important organs of the anterior kidney and liver. The anterior kidney is postulated as the major source of B lymphocyte development within teleosts and therefore plays a pivotal role within the immune response (Hansen and Zapata 1998). Likewise the liver also plays an important role within the immune response as it is the major source of acute phase proteins (Streetz et al. 2001). The lack of transcriptome response observed in the anterior kidney and liver within the current study is in agreement with two previous studies which examined the differential expression of a selection of immune response genes within gill, liver and anterior kidney. It appears that the majority of transcriptional variation occurs within the gills of both AGD affected Atlantic salmon (Bridle et al. 2006a) and rainbow trout (Bridle et al. 2006b) compared to the other organs. The fact that the liver and anterior kidney had considerably fewer differentially expressed genes highlights the localised nature of the host-pathogen interaction.

The liver and anterior kidney had no significantly differentially expressed transcripts in common. The gill, however, had five transcripts in common with the anterior kidney and two transcripts in common with the liver. Although the degree to which these transcripts were differentially expressed varied between the organs, all but one transcript displayed the same direction of expression. That is, when down-regulated in the gill, the transcript was also down-regulated in the anterior kidney. For instance, a transcript with sequence homology to mitogen-activated protein kinase was significantly down-regulated 5.93 fold within the anterior kidney and 4.37 fold within the gill. A similar situation was observed for Atlantic salmon infected with *Aeromonas salmonicida* Lehmann et Neumann (Ewart et al. 2005). This study found many of the

same genes were differentially expressed in the liver, spleen and anterior kidney following *A. salmonicida* challenge (Ewart et al. 2005). One interesting exception in the present study was the two different transcripts showing homology to the complement C1q-like protein. In the gill, a transcript representing the C1q-like protein was up-regulated, while in the liver a different transcript representing C1q-like protein was down-regulated. A pair-wise sequence alignment of these two transcripts revealed significant dissimilarity, suggesting they may represent different C1q loci.

In order to maintain homeostasis in terms of constant cell numbers, a balance between physiological cell death and cellular proliferation must be achieved. Apoptosis plays an important role in maintaining this balance. Two transcripts capable of triggering apoptotic cell death – procathepsin L and cathepsin H precursor – were down-regulated within AGD affected tissue. Cathepsins are a family of proteases that reside within the lysosome. In response to certain stimuli, cathepsins are released into the cytoplasm and are capable of triggering apoptotic cell death (Chwieralski et al. 2006). Involved in the same pathway, cystatin B has a protective role by binding tightly to cathepsins L and H and preventing inappropriate proteolysis (Pol and Björk 2001; Kopitar-Jerala 2006). Down-regulation of the cathepsin and cystatin B pathway may reduce apoptosis within the gill, which in turn, may promote the cellular proliferation in response to AGD.

An interesting finding from our study was the down-regulation of at least four genes with antioxidant properties. These were glutathione S-transferase, glutathione peroxidase, thioredoxin and NADP transhydrogenase. In general, such antioxidants function mainly to protect an organism from oxidative damage caused by reactive oxygen and nitrogen intermediates. In Atlantic salmon, genes of glutathione S-

transferase family have been shown to be down-regulated in whole fish in response to the fungal disease saprolegniasis (Roberge et al. 2007) and within the head kidney of individuals infected with *Piscirickettsia salmonis* Fryer, Lannan, Giovannoni et Wood (Rise et al. 2004a). However, in contrast, a number of studies have also reported an up-regulation of antioxidants in response to infection. Atlantic salmon infected with *A. salmonicida* display an up-regulation of glutathione peroxidase within the spleen and liver (Ewart et al. 2005). *P. salmonis* infected macrophages also show an up-regulation of antioxidants (Rise et al. 2004a). Down-regulation of antioxidant genes at the terminal stages of AGD may suggest fish are suffering oxidative stress. Such a physiological state combined with the high level of parasitism may be an important factor contributing to AGD induced death.

Thioredoxin is a multifunctional protein. Along with thioredoxin reductase and NADPH this enzyme is capable of catalysing dithiol-disulfide oxidoreductions in what has been collectively termed the thioredoxin system (Arnér and Holmgren 2000) . Due to its powerful redox potential, thioredoxin is also a strong intercellular antioxidant and when over expressed, protects the cell against oxidative stress. In addition, thioredoxin acts as a chemoattractant for monocytes, neutrophils and T-cells (Bertini et al. 1999). Within the gill tissue of AGD affected Atlantic salmon a decreased expression of both thioredoxin and NADP transhydrogenase was observed, suggesting this system has been compromised as either a result of host response or pathogen mediation. As thioredoxin is a chemoattractant, down-regulation of the thioredoxin system may prevent or at least restrict the infiltration of inflammatory cells within the gill. This may compromise the ability of individuals to mount an effective immune response against AGD.

Inflammation is generally characterised by a dramatic increase of acute phase proteins, collectively known as the acute phase response (Gabay and Kushner 1999). This response is stimulated initially by the pro-inflammatory cytokines (such as IL-1, interleukin-6 (IL-6), and TNFα), which leads to the translocation of two important transcription factors, NF-κB and C/EBPβ). The latter then binds to the promoter regions on many acute phase proteins and in so doing increases the expression of these proteins within the plasma (Poli 1998). In mice C/EBPβ is involved in the regulation of serum amyloid A, serum amyloid P and complement component C3 (Poli 1998). A number of other acute phase proteins (C-reactive protein), cytokines (IL-1β, IL-6, TNFα, interleukin-8, interleukin-12) as well as iNOS and lysozyme also contain the functional C/EBPβ binding motif, and are thus believed to be regulated to some degree by C/EBPβ (Agrawal et al. 2001). Recently, transfection of a rainbow trout macrophage cell line with recombinant IL-1β has demonstrated this cytokine will indeed induce expression of C/EBP within salmonid cells (Martin et al. 2007b). C/EBPβ must undergo sequential phosphorylation by mitogen-activated protein kinase and glycogen synthase kinase to acquire DNA binding ability (Tang et al. 2005; Kim et al. 2007).

The up-regulation of IL-1β reported previously (Bridle et al. 2006a; Bridle et al. 2006b; Morrison et al. 2007) and the up-regulation of C/EBPβ within the current study would suggest fish are attempting to mount an acute phase response to AGD. However, no such response is observed in our study, at least at the transcriptome level. Interestingly,. mitogen-activated protein kinase which is essential for C/EBPβ binding activity was down-regulated in both the gill and anterior kidney of AGD affected Atlantic salmon. It therefore seems likely that while the expression of C/EBPβ is increased, its activity as a

mediator of transcription is impaired. It may be possible that the pathogen itself is somehow capable of blocking the activity of mitogen-activated protein kinase and in turn C/EBPβ, which could severely compromise the host response.

The most highly up-regulated transcript observed within AGD affected Atlantic salmon was localised to the gill, and displayed a fold-change of 19.34 (± SE 3.04) and 218 (± SE 66) following microarray and qPCR analysis, respectively. BLAST against the NCBI dbEST suggests this transcript belongs to the UniGene cluster Ssa.25836 which currently has no functional protein assigned. Compared to heavily infected individuals, fish displaying less severe cases of AGD display significantly lower expression of this transcript by 4.57 (± SE 0.72) fold. This result may suggest that the expression of this transcript is positively correlated to the severity of infection. Characterising this unknown EST was beyond the scope of this study. However future research is underway to elucidate if it encodes a functional protein.

In conclusion, this study profiled the transcriptome response in the gill, liver and anterior kidney at the advance stages of an acute AGD infection in Atlantic salmon. We reported a significant down-regulation of many transcripts within the gill, and to a lesser extent within the anterior kidney and the liver. We hypothesise that this gene suppression may be parasite mediated, and involved in a mechanism by which the amoeba avoids or withstands the host response.

Chapter 5: Transcriptome profiling AGD resistant Atlantic salmon

Submitted as: Wynne JW, O' Sullivan MG, Stone G, Cook MT, Nowak B, Lovell DR, Taylor R, Elliott NG (2008) Resistance to amoebic gill disease (AGD) is characterised by the transcriptional dysregulation of immune and cell cycle pathways. *in preparation.*

Abstract

AGD is a parasite mediated proliferative gill disease capable of affecting a range of teleost hosts. While a moderate heritability for AGD resistance in Atlantic salmon has been reported previously, the mechanisms by which individuals resist the proliferative effects remain poorly understood. To gain more knowledge of this commercially important trait, we compared gill transcriptomes of two groups of Atlantic salmon, one designated putatively resistant, and one designated putatively susceptible to AGD. Utilising a 17k Atlantic salmon cDNA microarray we identified 196 transcripts that were differentially expressed between the two groups. Expression of 11 transcripts was further examined with real-time quantitative RT-PCR (qPCR) in the AGD resistant and susceptible animals, as well as non-infected naïve fish. Gene expression determined by qPCR was in strong agreement with the microarray analysis. A large number of differentially expressed genes were involved in immune and cell cycle responses. Resistant individuals displayed significantly higher expression of genes involved in adaptive immunity and negative regulation of the cell cycle. In contrast, AGD susceptible individuals showed higher expression of acute phase proteins and positive regulators of the cell cycle. Combined with the gill histopathology, our results suggest AGD resistance is acquired rather than innately present, and that this resistance is for the most part associated with the dysregulation of immune and cell cycle pathways.

5.1 Introduction

Many pathological conditions in vertebrates result in cell proliferation and changes in cellular architecture. Sometimes such conditions are initiated by a pathogen, as is the case with the parasite mediated proliferative condition: amoebic gill disease (AGD). Proposed to be caused by the protozoan *Neoparamoeba perurans* (Young et al. 2007; Young et al. 2008c), AGD affects cultured teleost species including Atlantic salmon (*Salmo salar*) (Adams and Nowak 2001; Young et al. 2008c), rainbow trout (*Oncorhynchus mykiss*) (Munday et al. 1990); turbot (*Scophthalmus maximus*) (Dyková et al. 1999); coho salmon (*Oncorhynchus kisutch*) (Kent et al. 1988); and seabass (*Dicentrarchus labrax*) (Dyková et al. 2000). Following initial infection, AGD causes extensive alterations in gill morphology: severe epithelial hyperplasia, hypertrophy, oedema and interlamellar vesicle formation (Adams and Nowak 2001). At present, the only successful treatment for fish affected by AGD is freshwater bathing (Parsons et al. 2001). However, as fish are continuously being re-infected, bathing needs to be repeated up to 12 times in a production cycle. Due to the high financial cost associated with this practice, bathing is not considered a viable long-term management solution.

Previous studies have suggested moderate heritability for AGD resistance within Atlantic salmon (Taylor et al. 2007). Enhancing this genetic resistance through selective breeding has become a major research focus. While improvements are possible using traditional phenotype-based selection, the use of marker-assisted selection (MAS) is preferable. MAS is most appropriate for traits in which the phenotype is difficult or expensive to measure, such as disease resistance. To facilitate a MAS program, genes (or their correlated markers) associated with resistance need to be identified. Such an

undertaking is complicated by the fact that disease resistance – as a trait – is often complex and usually under polygenic control. For example, resistance to the parasite *Gyrodactylus salaris* in Atlantic salmon is associated with multiple genomic regions, presumably spread over a number of genes (Gilbey et al. 2006). A similar situation is observed for the loci controlling resistance to *Ceratomyxa shasta* in rainbow trout (Nichols et al. 2003).

Mechanisms controlling AGD resistance in Atlantic salmon remain poorly understood. Studies have demonstrated that some Atlantic salmon previously infected with AGD will develop resistance upon subsequent re-infection and that this resistance is associated with the presence of anti-*Neoparamoeba* spp. antibodies (Vincent et al. 2006). Furthermore, a significant association between AGD resistance and allelic variation within the major histocompatibility (MH) class II alpha (*Sasa-DAA*) chain has also been described (Wynne et al. 2007b). Despite these encouraging associations, the main molecular mechanisms controlling AGD resistance are yet to be identified.

Host response to AGD has been studied extensively. At the transcriptome level, Atlantic salmon infected with AGD show up-regulation of the pro-inflammatory cytokine interleukin-1β (IL-1β) within the gill tissue (Bridle et al. 2006a; Morrison et al. 2007), and the transcription factor CCAAT/enhancer binding protein β (C/EBPβ) within the anterior kidney (Wynne et al. 2008). While suggestive of an acute phase response (APR), previous studies have reported either no, or only modest inductions of acute phase proteins following AGD infection (Bridle et al. 2006a; Bridle et al. 2006b; Morrison et al. 2006a; Wynne et al. 2008; Young et al. 2008a). Furthermore, upon first infection by AGD Atlantic salmon demonstrate a localised host immunosuppression

(Wynne et al. 2008), including down-regulation of genes involved in the MH class I and class II pathways (Young et al. 2008a). More recently, it has become apparent that genes involved in apoptosis and cellular proliferation pathways may have an important role in the host response to AGD, at least upon first infection (Morrison et al. 2006a; Wynne et al. 2008)

DNA microarrays are an important tool for investigating transcriptional changes within many aquatic organisms. In salmonids, microarrays have been used to examine responses to stress (Krasnov et al. 2005), bacterial infection (Rise et al. 2004a; Ewart et al. 2005), maturation (von Schalburg et al. 2005a), vaccination (Purcell et al. 2006) and cytokine stimulation (Martin et al. 2007a; Martin et al. 2007b). Microarrays have also become popular for identifying genes associated with disease resistance (Liu et al. 2001; Diez-Tascón et al. 2005). The present study has used a recently developed Atlantic salmon cDNA microarray (Martin et al. 2007a) to compare the transcriptome response of AGD resistant and AGD susceptible Atlantic salmon following natural infection.

5.2 Materials and Methods

5.2.1 Field AGD challenge

On 17 August 2006 a total of 2375 PIT (passive integrated transponder) tagged Atlantic salmon smolts from half-sibling families were stocked into a single (10x10m) sea-cage located within a commercial Atlantic salmon farm in Southern Tasmania. These fish were allowed to become naturally infected with AGD until 26 September 2006. The entire population was then scored for severity of AGD using the standard industry scoring method, which estimates the number of visible gross lesions on the gill surface (Powell et al. 2000). A score between 0 and 5 is then assigned to each individual, where 0 represents no visible lesions and 5 represents heavily infected gills. Following scoring, all fish were bathed in freshwater using standard industry protocols. After bathing, fish were again allowed to become naturally infected with AGD to a severe commercial score (10% of the cage population scoring over gill score 5) until 6 December 2006 when the gill scoring and bathing process was repeated. Following the second bathing, the remaining 1822 fish were allowed to become re-infected for the third time and their severity of AGD infection scored after 50 days. The fish, however, were not bathed and were allowed to become more severely infected with AGD and eventually succumb to the disease. The trial was then terminated on 16 March 2007 when the total mortality had reached 57% and significant variation in the severity of infection was still present on survivors.

Upon termination all surviving individuals were scored for AGD severity and a sample of 28 fish was selected for further analysis. These animals consisted of 14 individuals with either few or no gross signs of AGD (industry gill score of 0) and 14 individuals with severe gross signs of AGD (industry gill score of 4 or 5). From here on, these individuals are considered "resistant" and "susceptible", respectively. However, it must be acknowledged that, by surviving until trial termination, all individuals have demonstrated some degree of resistance. The 28 individuals were derived from over 14 random half-sibling families. Only one of the families was represented by over two fish. The family origin of some individuals (7 in total) was not known as these individuals had not been selected for pedigree analysis by the Tasmanian Atlantic salmon selective breeding program. Fish were euthanised with 5 g L^{-1} Aqui-S (Aqui-S NZ Ltd, Lower Hutt, New Zealand) and tissue samples dissected from the first right anterior gill hemibranch and stored in RNA*later* (Ambion, Austin, USA) at room temperature for 24 hours then -80°C until used. The first and second left anterior gill hemibranch was also dissected and fixed in seawater Davidson's fixative for 48 hours, then 70% ethanol until processed. All animal procedures were approved by the University of Tasmania Animal Ethics Committee under the guidelines of the Australian Code of Practice.

5.2.2 RNA isolation

Approximately 100 mg of gill tissue was homogenised in TRI-reagent (Ambion) using a Quicklyse tissue homogeniser (Qiagen, Doncaster, Australia). Total RNA was then isolated using the RiboPure™ kit according to the manufacturer's protocol (Ambion). Following isolation, RNA was treated with rDNase I from the DNA-*free*™ kit

(Ambion). The quantity and quality of total RNA was estimated by examining the 260:280 nm absorbance ratio (Nanodrop Technologies, Delaware USA) and visualising 2 μg of total RNA on a native 2% agarose gel. RNA was denatured in 15 μl of NorthernMax® Formaldehyde Load Dye (Ambion) and 20 μg ml^{-1} ethidium bromide at 65°C for 15 min prior to gel electrophoresis.

5.2.3 Microarray experimental design

This study used a 16,950 feature (so-called "17k") Atlantic salmon cDNA microarray developed by the TRAITS (Transcriptome Analysis of Important Traits in Salmon) and Salmon Genome Projects (Martin et al. 2007a). The study was designed to follow MIAME guidelines (Brazma et al. 2001). The experiment aimed to compare the gill gene expression profile between AGD resistant and AGD susceptible Atlantic salmon to identify genes differential expressed between these two states. The microarray experimental design consisted of direct comparisons between single resistant and single susceptible fish. No sample pooling was conducted within this microarray study. A total of 14 single fish comparisons were made, each with a dye-swap replicate. Therefore a total of 28 microarray hybridisations were conducted. The individual comparisons between resistant and susceptible fish were randomised, as was the order of slide processing. All hybridisations and protocols have been deposited into the ArrayExpress database and assigned the accession number E-MEXP-1396.

5.2.4 Microarray processing

Total RNA was quantified using the Qubit fluorescence quantification system as per the manufacturer's protocol (Invitrogen, Mount Waverly, Australia). Next, 5 µg of total RNA was reverse transcribed using Superscript II reverse transcriptase (Invitrogen) and the 3DNA Array 350™ kit (Genisphere, Hatfield, USA) as per the manufacturer's protocol. cDNA synthesis was terminated by the addition of 3.5 µl of 0.5 M NaOH/50 mM EDTA and incubation at 65°C for 15 min. Reactions were neutralised with 5 µl of 1 M tris-HCl (pH 7.5). Appropriate Cy3 and Cy5 reactions were then combined and purified using the QIAquick PCR purification kit (Qiagen) with one additional wash of 500 µl of 80% ethanol. cDNA was eluted in 30 µl of nuclease-free water.

All hybridisations were performed with the Maui® hybridisation system (BioMicro® Systems, Salt Lake City, USA) using the FL hybridisation mixing chambers (BioMicro® Systems). First, concentrated cDNA was combined with 2 µl LNA dT blocker (Genisphere) and 30 µl of 2X formamide-based hybridisation buffer (Genisphere) and incubated at 80°C for 10 min and then at 49°C until loaded onto the slide. Second, 50 µl of the cDNA hybridisation mix was applied to each pre-warmed microarray slide. The hybridisation was then performed for 15 hours at 49°C. Following hybridisation, the chambers were removed while submerged in 2X SSC/0.1% SDS at 49°C. Third, slides were washed once with 2X SSC/0.1% SDS at 49°C for 10 min, twice with 2X SSC/0.1% SDS at room temperature for 5 min, twice with 2X SSC at room temperature for 5 min, then twice with 0.1X SSC at room temperature for 5 min. Slides were centrifuged dry at 510g for 5 min at room temperature.

The Cy3 and Cy5 cyanine capture reagents (Genisphere) were hybridised to the microarray slides by combining 2.5 µl of both the Cy3 and Cy5 capture reagents with 20 µl nuclease free water and 25 µl of 2X formamide-based hybridisation buffer (Genisphere), incubating at 80°C for 10 min, and then at 49°C until loaded on the slide. Fifty microlitres of the cyanine hybridisation solution was then applied to the microarray slides and hybridised at 49°C for 4 hours. Following hybridisation, the chambers were removed in 2X SSC/0.1% SDS at 49°C and slides were washed and dried as described above.

5.2.5 Microarray scanning and image processing

All slides were scanned using an Axon 4000B (Axon Instruments, Union City, USA) with the software GenePix Pro ver3.0 (Molecular Devices Corporation, Sunnyvale, USA). The Cy3 and Cy5 fluorophores were excited at 532 and 635 nm respectively. The photo multiplier tube (PMT) balance was determined empirically through a series of preview scans on the first eight slides. The PMT gain for the 532 nm channel was 500 and for the 635 nm channel was 710. Laser power was maintained at 100% and, along with the PMT balance, remained constant for all slides. GAL files were aligned automatically and checked by eye. Intensity data for the 532 and 635 nm channels were extracted and saved as GenePix results files.

5.2.6 Statistical analysis of microarray data

GPR files were read into the statistical software package R (ver2.3.1) and analysed with Limma (Smyth 2005). Prior to statistical analysis, data were normalised with the NormalizeWithinArrays function using the default print-tip loess method (Smyth and Speed 2003). The moderated *t*-test statistic was computed for each feature and was used to detect differential expression between AGD resistant and susceptible Atlantic salmon. When a two group comparison is conducted this *t*-test is equivalent to an analysis of variance (ANOVA). The moderated *t*-statistic was chosen over an ordinary *t*-statistic because it moderates the standard errors across genes and increases the degrees of freedom and thus reliability of analysis (Smyth 2004) The false discovery rate (FDR) was controlled by a Bonferroni-type FDR approach previously described by Benjamini and Hochberg (1995). Only transcripts which had a significant *t*-statistic with FDR <0.05 (Benjamini and Hochberg 1995) and were greater than 2-fold differentially expressed were considered statistically significant within our study. The mean fold-change is present for each transcript deemed significant within our study.

5.2.7 Transcript annotation

Sequences of differentially expressed transcripts were first subjected to Basic Local Alignment Search Tool (BLAST) against the NCBI non-redundant protein database in the program Blast2GO (Conesa et al. 2005). Sequences with *E* values $<1\times10^{-10}$ were considered homologous. If no significant homology was detected the sequence was further analysed (using BLASTn) against the NCBI expressed sequence tag database

(dbEST) in an attempt to identify UniGene sequence clusters (www.ncbi.nlm.nih.gov/sites/entrez?db=unigene). Functional proteins associated with UniGene clusters were then used for annotation. In the case where no proteins were associated with the UniGene clusters, the UniGene code is presented. When no significant BLAST or UniGene cluster was identified the transcript was designated "unknown."

5.2.8 Real-time quantitative RT-PCR

A selection of genes was chosen for real-time quantitative RT-PCR (qPCR) to verify the microarray results. These genes were selected based on their involvement in the immune and cell cycle pathways. The remaining RNA samples from the microarray analysis (see above) were utilised for qPCR assays. No samples were pooled and assays were run on individual fish (14 resistant and 14 susceptible). In addition, six non-infected naïve smolts were also included within the qPCR analysis (again not pooled). These individuals were obtained from a laboratory based seawater tank system, and were presumed to represent non-diseased gene expression levels for salmon maintained in seawater. The inclusion of non-infected naïve samples in the qPCR analysis allow as to determine if certain genes were actually differential expressed in the resistant fish or simple a result of the differences in disease/physiological condition compared to the susceptible animals.

Primers were designed from EST sequences using PrimerExpress (Applied Biosystems). Primer sequences for IL-1β, C/EBPβ and the unknown EST, UniGene-

Ssa.25836, were obtained from previous studies (Haugland et al. 2005; Wynne et al. 2008). All primers used for the qPCR are shown in Table 5.1. The housekeeping genes β-actin and elongation factor IA_B (EFA_B) were chosen as the endogenous controls using the primers obtained from Haugland et al. (2005) and Olsvik et al. (2005) respectively.

TRAITS clone / reference	Identity	Primer sequence (5'-3')
CK897380	Immunoglobulin light chain precursor (Ig)	CCCTCCCCACTGACCAG GGTCCTTTAGGGTGGCCTC
EG649443	MHC class II invariant chain-like protein 2 (MH Ic)	GATGAGCAGGGCAACTACCT GCATCCACACACCAGCAA
CK894671	p13 cyclin dependent kinase inhibitor (CDKI)	TCACCAGTGGCGCAGTT CAGTGTGTCGGTCTGCGATA
DW588077	CDC14 cell division cycle 14 homolog a (CDC14)	CGGAACTGTATAATGACGATGAG TCAGGGCTCGCAGTTTATC
AM042284	Serum amyloid A (SAA)	CCAACCGCTACAGACCAAAT GGTCCTTCAGTTTGGCTAGTAAC
AM042158	C1q-like adipose specific protein (C1q)	CCGGCGCCTTAAAGATG CAAGCATGTCTTCACCAACAC
AM049496	C type lectin receptor A (CTL)	TCCATCTGACAAATGCGTTT CCATTCGCTGACCTGGTT
CK889753	Cyclin B1 (CYCB)	TGAACTACACCGAAGACTGCTTA CCTCGTTCACCTTCACAACAT
(Wynne et al. 2008)	CCAAT/enhancer binding protein β (C/EBP)	CGGAGACGCAACACAAAGTG CAGCTGCTCCACGCGTTT
(Wynne et al. 2008)	UniGene-Ssa.25836 (EST)	AATACCATCACTCATACCGCTT CCACTTTGCCACCTGTTATGT
(Haugland et al. 2005)	Interleukin-1β (IL-1B)	GCTGGAGAGTGTGTGGAAGA TGCTTCCCTCCTGCTCGTAG
(Haugland et al. 2005)	β-actin	CCAGTCCTGCTCACTGAGGC GGTCTCAAACATGATCTGGGTCA
(Olsvik et al. 2005)	Elongation factor $1A_B$	TGCCCCTCCAGGATGTCTAC CACGGCCCACAGGTACTG

Table 5.1. List of transcripts chosen for microarray validation using real-time quantitative RT-PCR. The clone sequence ID used to design primers is indicated along with the identity of the most significant BLAST result. In the case where primer sequences were obtained from previous studies, the appropriate reference is cited. PCR and cycling parameters were constant for all transcripts and are described within the text. The sense and anti-sense primer sequences are also presented.

cDNA was generated from total RNA using Superscript III reverse transcriptase (Invitrogen). Briefly, 2 μg of total RNA was quantified and reverse transcribed in a 20 μl reaction containing 4 μl of 5X buffer, 200 nM of oligo $dT_{(17)}$, 0.5 mM dNTPs, 5 mM DTT, 40 units RNaseOUT and 200 units of Superscript III reverse transcriptase. Following the addition of RNA, oligo dT and dNTPs, the reactions were heated to 65°C for 5 min. The buffer, DTT, RNaseOUT and reverse transcriptase were then added and reactions were incubated at 50°C for 1 hour. Two units of RNase H were finally added and reactions were incubated for 20 min at 37°C. The cDNA was then diluted to 4 ng (starting RNA) μl^{-1} with nuclease free water.

All reactions were performed in triplicate in a total volume of 25 μl containing 12.5 μl of 2X SensiMix (Quantace, London, UK), 0.5 μl of SYBR® Green I, 200 nM of forward and reverse primer, 5 μl of cDNA template and 5 μl of water. Reactions were subjected to 95°C for 10 min, followed by 40 cycles of 94°C for 20 sec, 56°C for 20 sec and 72°C for 20 sec using a Rotor-Gene™3000 real-time thermal cycler (Corbett Research, Mortlake, Australia). To verify amplification specificity, melt curves for each gene were examined from 72-95°C with a heating rate of 1°C per 5 sec. In addition, amplicons were examined by agarose gel electrophoresis and also cloned into PGEM-T-easy (Promega, Annandale, Australia) and sequenced using T7 and Sp6 primers.

5.2.9 qPCR data analysis

Raw fluorescence data were obtained and analysed using the Rotor-Gene Analysis Software ver6.1 (Corbett Research). Thresholds were determined manually and C_T

(cycle number at which the signal reaches the thresholds) calculated accordingly. Amplification efficiency (E) for each transcript was determined from a pooled cDNA sample (including the 14 AGD resistant and 14 AGD susceptible samples) serially diluted two-fold five times and calculated as $E=10^{(-1/S)}$ where s is the slope of the standard curve of serial dilutions according to Pfaffl (2001). Differences in gene expression between AGD resistant and susceptible Atlantic salmon were analysed using relative expression, with the software REST-XL (Pfaffl et al. 2002) according to the following equation.

$$\text{Ratio} = (E_{\text{target}})^{\Delta CT[\text{target(control-sample)}]}/(E_{\text{ref}})^{\Delta CT[\text{ref(control-sample)}]}$$

β-actin and EFA_B served as the endogenous reference genes. For the purpose of verifying the microarray results, the susceptible Atlantic salmon were considered the control sample. When the comparisons between naïve and AGD resistant and susceptible were performed, the naïve samples were considered the control. The analysis using different reference genes was performed independently. The REST-XL software calculates the relative expression ratio of a target gene, while adjusting for differences in amplification efficiency. Standard errors represent the variation between the individual fish within each group (i.e. between the 14 resistant fish or 14 susceptible fish).

5.3 Results

5.3.1 AGD histopathology

Histopathology of gill sections from the 14 AGD resistant and 14 AGD susceptible individuals revealed significant differences in the severity of infection (Figure 5.1). Inspection of the susceptible fish revealed substantial pathology including extensive lamella fusion; hyperplasia and hypertrophy of the epithelial cells; oedema; and interlamellar vesicle formation. In contrast, the resistant individuals showed very minor AGD pathology with only a few small lesions (two or three fused lamellae) observed microscopically. Nevertheless, the fact that all AGD resistant fish displayed some degree of pathology suggests resistance is associated with tolerance to the parasite, rather than absolute resistance. A moderate infiltration of inflammatory cells morphologically identified as neutrophils and monocytes/macrophages was observed within the lesions and within the adjacent central venous sinus of susceptible and resistant fish. The presence of amoebae with the symbiont (parasome) was observed within all AGD susceptible individuals and within some AGD resistant fish.

The mean gross gill scores for the 14 fish designated resistant and the 14 designated susceptible for the four different gill scoring events are illustrated in Figure 5.2. Both the resistant and susceptible groups showed considerable signs of AGD pathology during the early stages of the trial. It was not until the third gill scoring event that the

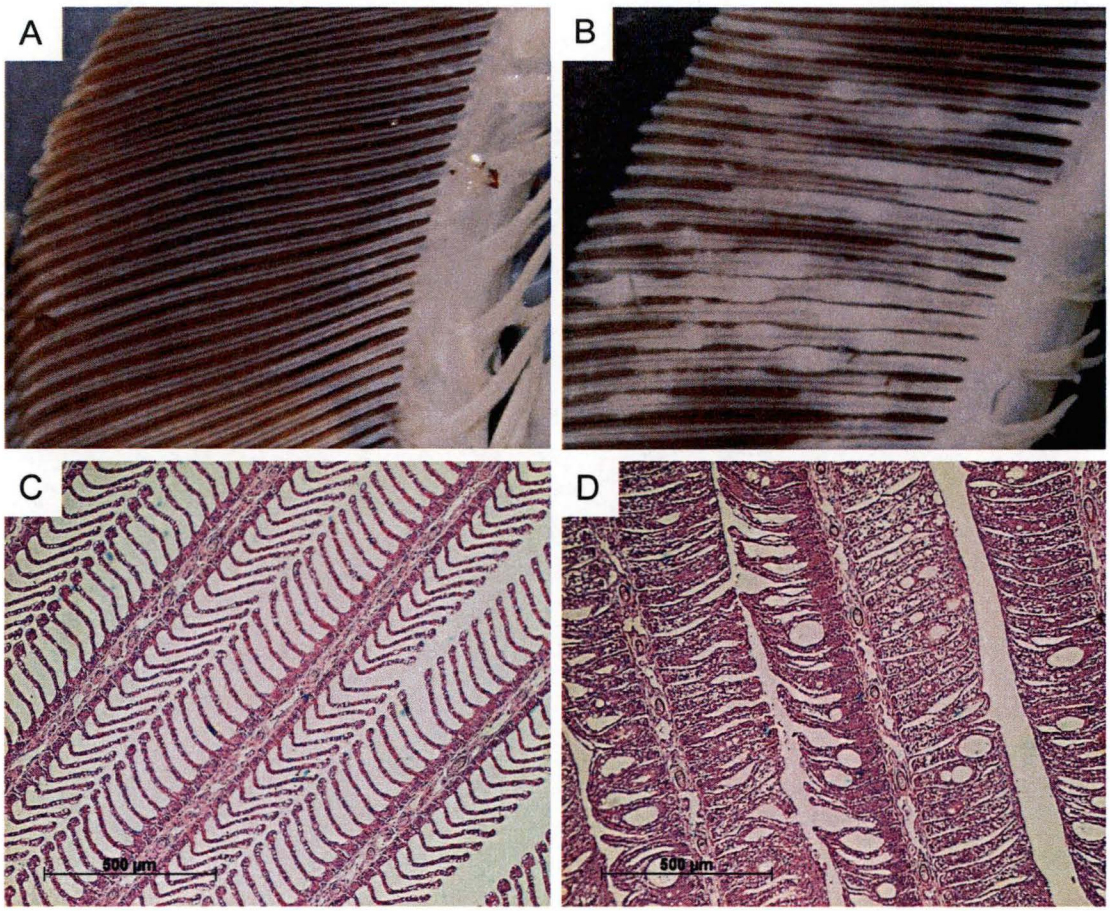


Figure 5.1. Gross pathology and histopathology of AGD resistant and susceptible Atlantic salmon. A: Gross presentation of seawater Davidson's fixed gill section obtained from an Atlantic salmon resistant (gross gill score 0) to AGD. B: Gross presentation of seawater Davidson's fixed gill section obtained from an Atlantic salmon susceptible (gross gill score 5) to AGD. Extensive multifocal mucoid lesions are clearly visible. C: histology from the gill of an AGD resistant Atlantic salmon displaying normal gill morphology. D: histopathology of an Atlantic salmon susceptible to AGD. Extensive hyperplastic lamellar fusion is clearly visible.

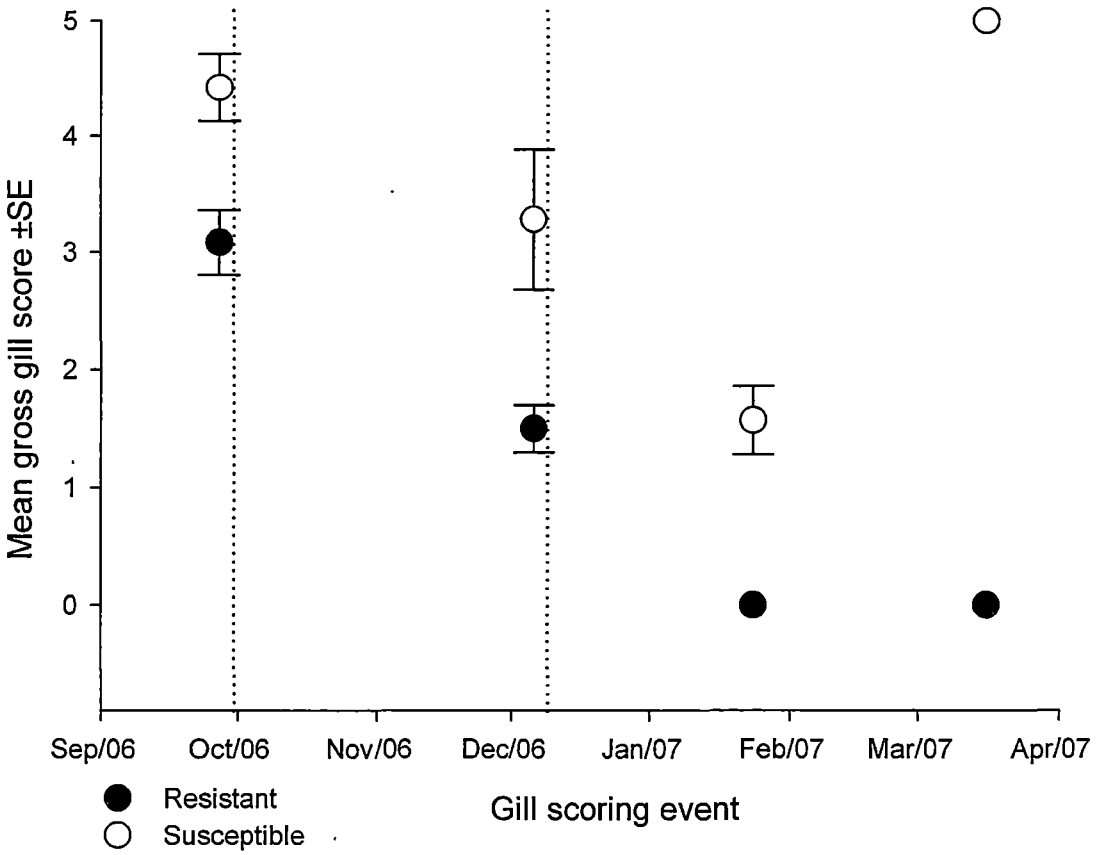


Figure 5.2. Mean gross gill score history ± standard error of the 28 analysed Atlantic salmon designated either AGD resistant (n = 14) or AGD susceptible (n = 14). Gross gill scoring was completed at four time points, with the final time point representing trial termination. *Vertical dashed lines* indicate freshwater bathing events. All fish were assumed to return to a gill score of 0 following bathing.

resistant individuals showed considerable improvement in resistance. The greatest difference in gross gill score between the two groups was observed at trial termination when all resistant fish had a gross gill score of 0, while all the susceptible fish had a gross gill score of 5.

5.3.2 Microarray gene expression profiles, comparison between AGD resistant vs. susceptible Atlantic salmon

In total, 196 transcripts were deemed to be significantly differentially expressed between AGD resistant and AGD susceptible Atlantic salmon (Figure 5.3). One hundred and four transcripts, termed up-regulated, appeared to have significantly higher expression in the AGD resistant individuals compared to AGD susceptible individuals. Ninety-two transcripts, termed down-regulated, had significantly lower expression in the AGD resistant individuals compared to the AGD susceptible individuals (Table 5.2). BLAST analysis of the 104 up-regulated sequences found 70 with positive identities; 34 had no BLAST hit or belonged to a non-annotated UniGene cluster. BLAST analysis of the 92 down-regulated transcript sequences, found 65 positive identities; 27 had either no BLAST hit or represented a non-annotated UniGene cluster. All transcripts with positive identities that were significantly up- or down-regulated in the resistant fish are listed in Table 5.3 and Table 5.4, respectively. Transcripts with either no BLAST hit or homology to a non-annotated UniGene cluster are presented within Appendix 3.1 and 3.2.

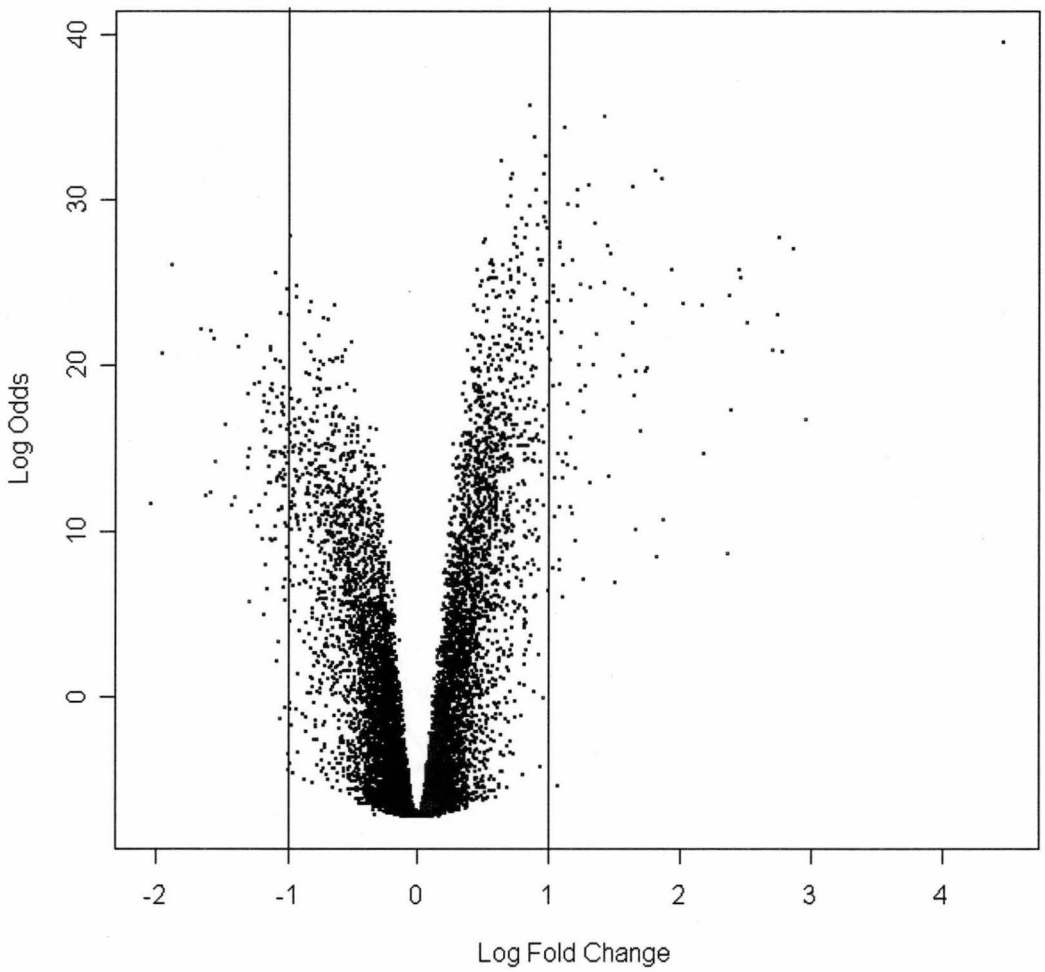


Figure 5.3. Volcano plot of normalised microarray data produced from 28 hybridisations. The mean of each replicate feature is presented. The x-axis represents the $\log_2$ of the fold-change between AGD resistant and susceptible Atlantic salmon. The y-axis corresponds to the log odds which represents the odds (or probability) that a certain gene is differential expressed. A log odds value of 0 represents a 50% chance that the gene is differential expressed. The greater the log odds value the greater the probability the gene is truly differentially expressed and not a false positive. The *vertical* lines represent the two-fold change ($\log_2$) threshold utilised in this study.

	BLAST hit		no BLAST hit	
	BLASTx	BLASTn (UniGene)	non annotated UniGene	unknown
Up-regulated	55	15	26	8
Down-regulated	52	13	23	4

Table 5.2. The number of transcripts significantly up- and down-regulated in the AGD resistant individuals is summarised. Transcripts were identified through either BLASTx of the NCBI non-redundant protein database or BLASTn of the NCBI EST database (annotated UniGene clusters). The number of transcripts which had no BLAST hit or belonged to a non-annotated UniGene clusters is also presented.

Functional class	Clone ID	Identity	Accession	p Value	Fold-change	Biological process GO
Electron transport						
	CO472279	Cytochrome P450 1A	Ssa.31314	1.381E-11	2.79	GO:0006118: electron transport
	AM402919	Cytochrome P450 1A1	Q92110	1.32E-13	2.60	GO:0006118: electron transport
	DW591439	Cytochrome b	YP_001122912	6.30E-09	2.36	GO:0006118: electron transport
Immune and stimulus response						
	CK894662	Guanylate binding protein 4	AAM44075	9.47E-11	2.46	GO:0006955: immune response
	CK897380	Immunoglobulin light chain precursor	AAD38362	5.74E-07	2.45	GO:0006955: immune response
	CK885361	Matrix Gla protein	AAO64980	1.24E-12	2.39	GO:0009607: response to stimulus
	AM041756	Bactericidal permeability-increasing protein	BAB91243	4.70E-13	2.27	GO:0006955: immune response
	AJ425406	Invariant chain INVX	AAL58577	3.95E-12	2.22	GO:0006955: immune response
	EG649443	MHC class II invariant chain-like protein 2	AAL91669	1.91E-12	2.18	GO:0006955: immune response
	CO471779	Bactericidal permeability-increasing protein	BAB91243	3.07E-13	2.14	GO:0006955: immune response
	EG355307	Granzyme k	AAQ54830	5.98E-09	2.13	GO:0006955: immune response
	CK880837	Immunoglobulin light chain precursor	AAG18372	2.05E-05	2.12	GO:0006955: immune response
	CK881973	Immunoglobulin light chain	BAD16718	6.81E-06	2.11	GO:0006955: immune response
	DW588549	Immunoglobulin light chain	AAK97642	2.42E-09	2.11	GO:0006955: immune response
	CK892925	Barrier to autointegration factor 1	NP_446083	7.73E-04	2.09	GO:0009607: response to stimulus
	CK893291	C1q subcomponent, gamma polypeptide	EDL80826	7.82E-12	2.05	GO:0006955: immune response
	var_imm_IGB09_abe_tra_sub_0p	CD8 α	AAW33876	5.11E-09	2.04	GO:0006955: immune response
Cell cycle						
	CK894671	p13 cyclin dependent kinase inhibitor	CAC12808	1.94E-11	2.26	GO:0007049: cell cycle
	DW588093	Septin 5	AAH94347	1.66E-12	2.21	GO:0007049: cell cycle
	ova_oyr_03B05_gal_sal_std_5p	Sjogrens syndrome scleroderma autoantigen 1	NP_001033617	9.78E-11	2.03	GO:0007049: cell cycle
Signal transduction and communication						
	CK884106	CC chemokine SCYA112	AAT52146	9.52E-16	3.66	GO:0007165: signal transduction

Table 5.3. Continued over page

CK877670	Mitogen-activated protein kinase kinase kinase 6 isoform 3	Ssa.709	4.95E-14	3.00	GO:0007165: signal transduction
CK879697	PREDICTED: similar to A-kinase anchoring protein	XP_699814	8.20E-14	2.95	GO·0007165: signal transduction
AJ424476	Arrestin domain containing 2	NP_956233	1.28E-08	2.29	GO:0007165: signal transduction
DW588069	DIX domain containing 1	NP_878304	2.37E-12	2.27	GO·0007154: cell communication
DW589803	Proteolipid protein 2	XP_001507488	2.29E-11	2.19	GO.0007165: signal transduction
CK897432	Mitogen-activated protein kinase kinase kinase 6 isoform 3	Ssa.709	1.30E-13	2.19	GO:0007165: signal transduction
CK897499	Mitogen-activated protein kinase kinase kinase 6 isoform 3	Ssa.709	2.78E-10	2.03	GO:0007165: signal transduction
CK897571	hydroxysteroid 11-beta dehydrogenase	BAC76709	1.45E-10	2.03	GO:0007154: cell communication
AJ424335	Mitogen-activated protein kinase kinase kinase 6 isoform 3	Ssa.709	6.25E-09	2.01	GO:0007165: signal transduction
Cellular process					
CO469753	Spondin 1a	CAM56280	5.27E-11	2.25	GO.0007155: cell adhesion
CK884207	Collagen α-2(I) chain precursor	O93484	8.97E-10	2.24	GO:0001501: skeletal development
CK896922	Growth arrest and DNA-damage-inducible β	NP_056490	6.70E-12	2.23	GO:0008219: cell death
EG649361	Collagen α-2(I) chain	BAB79230	4.49E-10	2.21	GO:0001501: skeletal development
EG649013	Collagen α-3(I)	BAB55662	4.55E-10	2.2	GO:0008283: cell proliferation
DW592221	Collagen α-2(I) chain	BAB79230	1.44E-08	2.17	GO:0001501: skeletal development
CK885060	Collagen precursor (XVI)	Ssa.34037	3.38E-10	2.13	GO:0007155: cell adhesion
CN181319	Collagen precursor (XVI)	Ssa.30921	4.21E-10	2.12	GO:0007155: cell adhesion
CK885437	Echinoderm microtubule associated protein like 2	NP_001025406	3.07E-10	2.1	GO:0016043: cellular organization
DW591808	Collagen α-4(IV)	Ssa.34688	3.50E-10	2 07	GO:0048468: cell development
DW589811	Heat shock 70kda protein 9b (mortalin-2)	AAI02335	1.60E-12	2.06	GO:0016043: cellular organization
CK873772	Collagen α-1(I)	BAA33380	2.06E-10	2 05	GO:0001501: skeletal development
gil_cgi_E4E12_car_tra_sub_0p	Collagen α-2(VIII)	Ssa. 24648	2.38E-07	2.05	GO:0007155: cell adhesion
CK885676	TSC22 domain family, member 3 isoform 2	NP_004080	5.75E-11	2 04	GO:0008219: cell death
EG355157	Collagen α-2(VIII)	Ssa.24648	1.54E-07	2.03	GO:0007155: cell adhesion
AJ424854	Collagen α-3(I)	BAB55662	5.59E-10	2.03	GO:0008283: cell proliferation
EG648912	Collagen α-1(I)	BAA33380	1.89E-11	2.02	GO:0001501: skeletal development
AJ424291	Collagen α-1(XV) isoform 2	Ssa 32837	1.92E-14	2.01	GO:0007155: cell adhesion
CK884893	PREDICTED: similar to collagen α-2(IX)	XP_001498143	2.16E-08	2.01	GO:0001501: skeletal development
Cellular metabolism					

Table 5.3. Continued over page

	CO472136	Sulfotransferase isoform 2	AAO64984	1.02E-09	3.07	GO:0044237: cellular metabolism
	EG647995	Creatine kinase, muscle	EAW57337	1.14E-12	2.33	GO:0044237: cellular metabolism
	DW588077	CDC14 cell division cycle 14 homolog a	NP_956473	9.09E-11	2.19	GO:0044237: cellular metabolism
	EG647833	H1 histone family, member X	NP_001080265	1.31E-12	2.18	GO:0044237: cellular metabolism
	DW588390	Creatine kinase, muscle	EAW57337	7.26E-11	2.18	GO:0044237: cellular metabolism
	CK883781	Ubiquitin	CAG04600	7.01E-12	2.17	GO:0044237: cellular metabolism
	CK885560	protein tyrosine phosphatase	NP_001013058	1.37E-08	2.14	GO:0044237: cellular metabolism
	CK880071	Creatine kinase, brain	NP_775329	4.14E-12	2.07	GO:0044237: cellular metabolism
	CK894889	Dual adaptor of phosphotyrosine and 3-phosphoinositides	EAX06111	4.37E-08	2.01	GO:0044237: cellular metabolism
Other metabolism						
	EG355006	Apolipoprotein D precursor	Ssa.29386	2.18E-12	2.48	GO:0006629: lipid metabolic process
	CK885772	Sulfotransferase 1 isoform 3	ABJ98761	1.13E-11	2.29	GO:0006805: xenobiotic metabolism
Transport						
	CK874017	Epithelial calcium channel	AAP12529	1.85E-09	2.33	GO:0006816: calcium ion transport
	AM042393	Prosaposin isoform 8	Ssa.6260	6.24E-08	2.25	GO:0006810: transport
	CK885326	RAB32, member RAS oncogene family	NP_958489	1.98E-09	2.07	GO:0006810: transport
No GO assigned						
	CK876833	Keratocan	ABG72686	4.03E-15	2.49	N/A
	CK879811	Novel protein containing a SEA domain	CAM14120	6.86E-14	2.49	N/A
	BM414353	FAM83F protein	AAH31099	9.37E-11	2.08	N/A
	CK873807	T-cell immunoglobulin and mucin domain containing 4	Ssa.5164	6.65E-12	2.07	N/A
	CK897241	Neuroblastoma, suppression of tumorigenicity 1	NP_996980	2.05E-09	2.05	N/A
	CK879753	VHSV-induced protein	AAM18475	3.66E-10	2.05	N/A
	CO470982	Anhydrase	Ssa.5502	4.74E-12	2.01	N/A

Table 5.3. Legend over page

Table 5.3. Transcripts with BLAST identities that were significantly up-regulated in AGD resistant compared to AGD susceptible Atlantic salmon are presented. Clone ID (accession number or TRAITS ID) is indicated, along with the identity of the most significant BLAST result, the accession number of that BLAST hit, and the high level Gene Ontology biological process term when available. The fold-change and adjusted p value are also indicated.

Functional class	Clone ID	Identity	Accession	p Value	Fold-change	Biological process GO
Electron transport						
	CK892524	Aldo-keto reductase	CAI20762	2.875E-16	- 2.73	GO:0006118: electron transport
	CK882754	Cytochrome c oxidase subunit IV isoform 2 precursor	AAF79934	6.702E-18	- 2.48	GO.0006118: electron transport
	BI468047	Cytochrome P450, family 2, subfamily J, polypeptide 24	NP_001076518	1.450E-12	- 2.45	GO:0006118: electron transport
	AM412213	Cytochrome c oxidase subunit 1	ABG37914	1.932E-09	- 2.26	GO:0006118: electron transport
Stress response						
	CK890160	Heat shock 60 kD protein 1	NP_851847	8.778E-20	- 2.70	GO:0006950: stress response
	CK885807	Heat shock 60 kD protein 1	NP_851847	3.099E-15	- 2.38	GO:0006950: stress response
	BI468144	NADH dehydrogenase (ubiquinone)	Ssa.5559	1.267E-10	- 2.16	GO:0006950: stress response
	CK880657	Isocitrate dehydrogenase 2 (NADP+)	Ssa.5190	5.643E-14	- 2.14	GO:0006950: stress response
	AJ424875	Heat shock protein 60	ABH88213	4.534E-08	- 2.14	GO:0006950: stress response
	CK878109	Heat shock 60 kD protein 1	NP_851847	3.311E-15	- 2.06	GO:0006950: stress response
Immune/defense response						
	AM042158	C1q-like adipose specific protein	AAM73701	1.85E-13	-6.88	GO:0006955: immune response
	AM042078	C1q-like adipose specific protein	AAM73701	1.75E-13	-6.55	GO:0006955: immune response
	CK896920	C1q-like protein	ABN45966	3.09E-08	-5.17	GO:0006955: immune response
	AM042338	Natterin-like protein	Q5CZR5	9.77E-15	-4.06	GO:0009405: pathogenesis
	AM049678	C type lectin receptor A	Ssa.851	1.28E-15	-3.83	GO:0006955: immune response
	CK889507	Transferrin	BAA84102	4.40E-09	-3.69	GO:0006952: defense response
	AM049496	C type lectin receptor A	AAT77220	4.56E-18	-3.65	GO:0006955: immune response
	gil_cgi_D1E08_car_tra_sub_0p	Differentially regulated trout protein 1	AAG30030	1.04E-14	-3.37	GO:0006955: immune response
	AM041713	Arachidonate 5-lipoxygenase	Ssa.7568	5.55E-13	-3.36	GO:0006952: defense response
	AM049499	Differentially regulated trout protein 1	AAG30030	5.99E-13	-3.18	GO:0006955: immune response
	AM049530	Differentially regulated trout protein 1	AAG30030	2.38E-12	-3.17	GO:0006955: immune response
	AM049476	C type lectin receptor A	AAT77220	7.03E-18	-3.15	GO:0006955: immune response
	AM049895	C type lectin receptor A	AAT77220	5.47E-15	-3.14	GO:0006955: immune response
	AM049483	Differentially regulated trout protein 1	AAG30030	3.89E-15	-3.02	GO:0006955: immune response

Table 5.4. Continued over page

AM049472	C type lectin receptor B	AAT77221	1.79E-07	-2.87	GO:0006955: immune response
AM042284	Serum amyloid A	CAM12347	4.79E-10	-2.5	GO:0006952: defense response
Cell cycle					
CK889994	Karyopherin α2	NP_001002335	9.457E-16	- 2.16	GO:0007049: cell cycle
CK889753	Cyclin B1	ABI54407	2.798E-14	- 2.08	GO:0007049: cell cycle
CK890958	Cyclin B1	ABI54407	5.150E-15	- 2.06	GO:0007049: cell cycle
Transcription factors					
BM414000	General transcription factor IIB	CAM14348	7.352E-09	- 3.20	GO:0045449: regulation of transcription
CK897622	Interferon regulatory factor 6 isoform 1	Ssa.33979	3.315E-14	- 3.13	GO:0045449: regulation of transcription
CK892065	CCAAT/enhancer binding protein β	ABD84407	8.715E-08	- 2.14	GO:0045449: regulation of transcription
Signal transduction					
var_imm_IGA01_abe_tra_sub_0p	Interleukin-1β	AAT36642	2.005E-15	- 5.54	GO:0007165: signal transduction
AY617117	Interleukin-1β	AAT36642	1.203E-15	- 5.49	GO:0007165: signal transduction
AY617117	Interleukin-1β	AAT36642	6.039E-15	- 5.21	GO:0007165: signal transduction
AY617117	Interleukin-1β	AAT36642	1.085E-14	- 4.52	GO:0007165: signal transduction
kid_cki_A2C12_car_tra_sub_0p	Tumor-associated calcium signal transducer	NP_998340	2.840E-15	- 2.70	GO:0007165: signal transduction
EG647772	Mitogen-activated protein kinase kinase kinase 6 isoform 3	Ssa.709	2.860E-09	- 2.28	GO:0007165: signal transduction
CK885687	N-myc and STAT interactor	Ssa.1929	3.178E-16	- 2.12	GO:0007165: signal transduction
Cellular process					
CK885737	Fibroblast growth factor-binding protein	Ssa.25939	3.24E-14	-5.73	GO.0008283: cell proliferation
CK892779	Intestinal mucin-like protein	P98089	5.91E-12	-5.26	GO:0006928: cell migration
BM414013	Myosin, heavy polypeptide 9, non-muscle	NP_990808	4.19E-08	-3.56	GO:0016043: cellular organization
BM414061	Calpain 1, isoform CRA_a	EDM12549	6.52E-12	-2.41	GO:0008283: cell proliferation
CK895032	Desmoglein 1β	NP_859010	4.06E-12	-2.22	GO:0007155: cell adhesion
CK888161	Cystathionase isoform 2	Ssa.31931	1.78E-19	-2.17	GO:0008283: cell proliferation
CK888789	Cystathionase	NP_997769	2.37E-16	-2.13	GO:0008283: cell proliferation
Cell metabolism					

Table 5.4. Continued over page

	CK885733	Cytidine deaminase	NP_991242	1.79E-16	-6.75	GO.0044237: cellular metabolism
	DW589192	Glutamate decarboxylase-like 1	NP_001039075	7.75E-11	-4.54	GO:0044237: cellular metabolism
	CK891048	Ribonucleotide reductase m2 polypeptide	AAH75746	7.51E-17	-2.57	GO.0044237: cellular metabolism
	CK879648	Squalene monooxygenase	Ssa.33937	3 94E-13	-2.54	GO:0044237: cellular metabolism
	BM413697	ribonuclease/angiogenin inhibitor	Ssa.30974	1.87E-12	-2.38	GO:0044237: cellular metabolism
	CK890886	Ribonucleotide reductase m2 polypeptide	NP_571525	2.33E-17	-2.34	GO:0044237: cellular metabolism
	DW591384	Tubulin, α2	AAQ91280	9.18E-18	-2.34	GO:0044237: cellular metabolism
	CK892811	Carbonyl reductase 1	NP_919387	1.23E-12	-2.12	GO:0044237: cellular metabolism
	AM042370	Ribonucleotide reductase m2 polypeptide	AAH75746	8.46E-11	-2.11	GO:0044237: cellular metabolism
Other metabolism						
	CO471940	Isopentenyl-diphosphate delta isomerase 2	NP_001011323	3.67E-15	-2.49	GO:0006629: lipid metabolic process
	EG355381	Aldose reductase	AAA31157	1.38E-13	-2.38	GO:0005975: carbohydrate metabolism
	CK875291	Isopentenyl-diphosphate delta isomerase	EDL86437	4.09E-13	-2.35	GO:0006629: lipid metabolic process
Transport and development						
	DW589322	Type I keratin S8	CAC45059	2.234E-13	- 2.98	GO:0048731: system development
	EG647609	ATP-binding cassette, sub-family F (GCN20), member 2	NP_958472	2.921E-13	- 2.02	GO:0006810: transport
No GO assigned						
	BM414515	Tumor differentially expressed 2	NP_956021	2.620E-18	- 3.54	N/A
	BM414518	ORF1 of novel LINE-like retrotransposon	Ssa.15073	1.847E-10	- 2.31	N/A
	CK884079	ACyLtransferase-like family member	Ssa.22612	7.163E-16	- 2.29	N/A
	CK897877	Cysteine and histidine-rich domain-zinc-binding protein 1	NP_956633	3.002E-11	- 2.25	N/A
	CK885188	Arginine-rich, mutated in early stage tumors	EAW65138	2.233E-17	- 2.21	N/A

Table 5.4. Legend over page

Table 5.4. Transcripts with BLAST identities that were significantly down-regulated in AGD resistant compared to AGD susceptible Atlantic salmon are presented. Clone ID (accession number or TRAITS ID) is indicated, along with the identity of the most significant BLAST result, the accession number of that BLAST hit, and the high level Gene Ontology biological process term when available. The fold-change and adjusted p value are also indicated.

A moderate redundancy was observed within the lists of genes differentially expressed. This provides some degree of internal validation for the microarray results. Furthermore, in most cases, the degree of differential expression between different transcripts representing the same gene was very similar. For instance, the fold-change of four transcripts representing differentially regulated trout protein 1 (DRTP1) ranged from 3.02 to 3.37 down-regulated in the AGD resistant Atlantic salmon. Removing redundancy reduced the number of significantly up- and down-regulated transcripts with known identities to 55 and 49, respectively.

Grouping non-redundant informative transcripts based on high-level biological process gene ontology terms suggested the majority of transcripts differentially expressed were involved in the immune response (GO:0006955) and cellular metabolism (GO:0044237) (Figure 5.4). Some gene ontology terms, such as cell cycle (GO:0007049), signal transduction (GO:0007155) and cellular metabolism (GO:0044237) were almost equally represented in both up- and down-regulated gene lists. However, other gene ontology terms showed large frequency differences between the up- and down-regulated groups. For instance, all transcripts assigned the terms, "stress response" (GO:0006950), "defence response" (GO:0006952) and "regulation of transcription" (GO:0045449) were down-regulated in the resistant individuals, while transcripts assigned the terms "skeletal development" (GO:0001501), "response to stimulus" (GO:0009607) and "cell communication" (GO:0007154) were only up-regulated in the resistant individuals. Transcripts involved in cell adhesion (GO:0007155) and immune response (GO:0006955) were also more frequent in the up-regulated list.

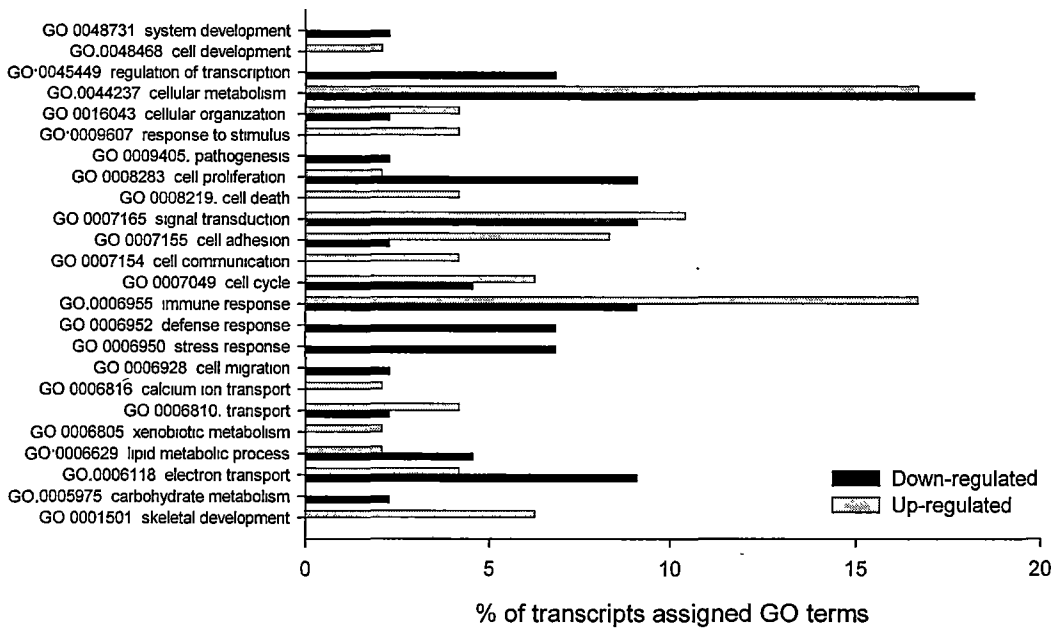


Figure 5.4. The proportion of informative transcripts (redundancy removed) assigned high level biological process Gene Ontology terms that were significantly up- and down-regulated in AGD resistant compared to AGD susceptible Atlantic salmon are presented. A total of 48 (87%) up-regulated informative transcripts and 44 (89%) down-regulated informative transcripts could be assigned biological process Gene Ontology terms.

A number of transcripts with strong homology to genes involved in cell cycle and cellular proliferation were differentially expressed between AGD resistant and AGD susceptible fish. Resistant individuals showed increased expression of the cyclin dependent kinase inhibitor p13, septin 5, growth arrest and DNA-damage-inducible β (GADDβ), CDC14 cell division cycle 14 homolog A and mortalin-2 (70 kD HSP). Resistant individuals showed decreased expression of cyclin B, fibroblast growth factor-binding protein and N-myc STAT interactor. The resistant fish were also found to have an up-regulation of many transcripts resembling various types of collagen. Two transcripts resembling putative tumor suppressors (TSC-22 (TFGβ stimulated clone-22) and neuroblastoma, suppression of tumorigenicity 1) were also up-regulated in the resistant fish, while a transcript with homology to tumor differentially expressed 2 was down-regulated.

The acute phase response genes serum amyloid A (SAA), C-type lectin receptor A, C-type lectin receptor B, C1q complement component, DRTP1 and transferrin were all down-regulated within the gill of AGD resistant fish. Three transcription factors were down-regulated in the AGD resistant fish: interferon regulatory factor 6 isoform 1, general transcription factor IIb, and C/EBPβ. The pro-inflammatory cytokine IL-1β was significantly down-regulated in the resistant fish. The resistant fish demonstrated lower expression of stress related genes, namely transcripts encoding the heat shock 60kD protein.

In contrast, the AGD resistant individuals appeared to have higher expression of transcripts involved in adaptive immunity: immunoglobulin light chain (Ig), MHC class II invariant chain (MH Ic), CD8 and a transcript known as T-cell immunoglobulin and

mucin domain containing 4. Two transcripts known as bactericidal permeability-increasing protein were also up-regulated in the resistant fish.

5.3.3 qPCR of AGD resistant vs. susceptible Atlantic salmon

Eleven transcripts identified as significantly up- or down-regulated within the microarray analysis were further examined with qPCR. Melt curve and sequencing analysis of amplicons indicated the single product of interest was amplified in all cases. The expression of the reference genes β-actin or EFA_B did not appear to be influenced by AGD.

Typically, the fold-change of transcripts examined with qPCR was in agreement with the fold-change determined through the microarray analysis (Pearson correlation = 0.96, $p<0.001$; Figure 5.5). Some minor discrepancies were observed in the magnitude of the differential expression, with the microarray slightly underestimating the differential expression compared to qPCR.

In addition to verifying the microarray results, the genes selected for qPCR allowed us to confirm the expression of genes involved in two distinct processes: the acute phase response and the cell cycle response. The only gene not involved in these pathways (at least to our knowledge) but analysed with qPCR, was the unknown EST, UniGene-Ssa.25836. At present, the function of this gene has not been determined, however, our previous research has found it to be significantly up-regulated during AGD infection

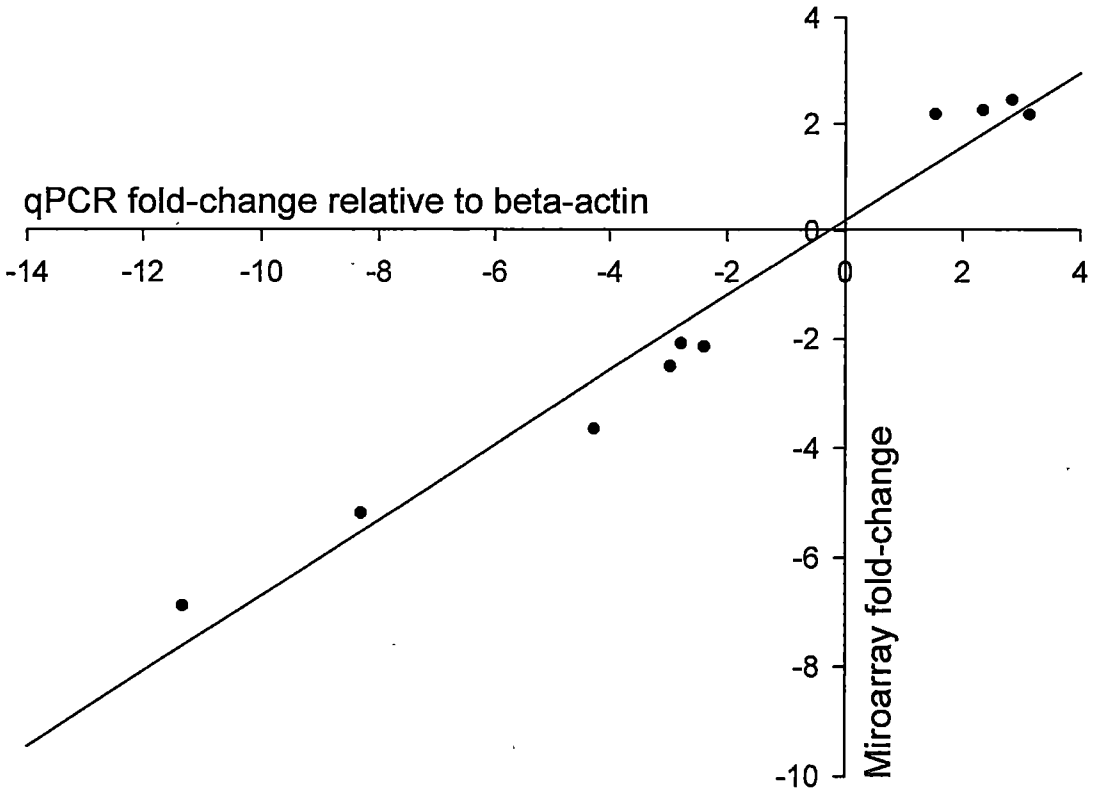


Figure 5.5. Correlation between the fold-change determined through microarray and real-time quantitative RT-PCR (qPCR) analysis of 10 transcripts is presented. Due to showing such a large differential expression according to both the microarray and qPCR analysis the unknown EST UniGene Ssa.25836 was not included within this figure.

(Wynne et al. 2008). The fold-change of each transcript examined with qPCR is shown in Table 5.5.

5.3.4 qPCR of AGD resistant and susceptible vs. naïve Atlantic salmon

Inclusion of samples from naïve seawater-maintained fish in the qPCR analysis allowed us to examine what could be considered the "normal" expression levels of these 11 genes. For some transcripts, the expression in both the AGD resistant and susceptible Atlantic salmon was higher compared to the naïve individuals (Figure 5.6). An example of such was the expression of the immune gene Ig, which was up-regulated by approximately 5.1 and 2.4 fold in the AGD resistant and susceptible individuals compared to naïve, respectively. Another adaptive immune related gene, MH Ic, showed a similar pattern.

Most transcripts involved in the acute phase response showed significantly higher expression in the AGD susceptible individuals compared to the naïve controls, with the exception of SAA. This transcript demonstrated decreased expression in both the AGD resistant and susceptible individuals relative to naïve. The resistant individuals generally demonstrated a small decrease in expression of the acute phase proteins C1q, C-type lectin receptor A (CTL) and IL-1β relative to the naïve samples. As reported previously (Wynne et al. 2008) the expression of C/EBPβ, and the unknown EST, UniGene-Ssa.25836, were both up-regulated in AGD resistant and susceptible fish compared to the naïve fish.

Clone ID	Identity	Fold-change normalised to β-actin (± SE)	Fold-change normalised to EFA_B (± SE)
CK897380	Immunoglobulin light chain precursor	2.83 (2.43)	2.16 (3.09)
EG649443	MHC class II invariant chain-like protein 2	3.13 (3.41)	2.94 (0.33)
CK894671	p13 cyclin dependent kinase inhibitor	2.34 (0.60)	2.04 (0.20)
DW588077	CDC14 cell division cycle 14 homolog a	1.53 (1.25)	1.12 (0.24)
CK889753	Cyclin B1	- 2.79 (0.71)	- 2.61 (1.07)
N/A	Interleukin-1β	- 8.31 (3.01)	- 8.50 (2.48)
CK892065	CCAAT/enhancer binding protein β	- 2.40 (0.78)	- 2.19 (0.77)
AM042284	Serum amyloid A	- 2.97 (2.76)	- 3.16 (2.23)
AM042158	C1q-like adipose specific protein	- 11.33 (18.80)	- 10.07 (16.48)
AM049496	C type lectin	- 4.28 (1.67)	- 3.54 (1.65)
CK880278	UniGene-Ssa.25836	- 70.05 (517.76)	- 61.81 (134.26)

Table 5.5. The fold-changes of a sub-sample of transcripts determined by real-time quantitative RT-PCR (qPCR) normalised to either beta actin or elongation factor $1A_B$ are presented. Negative fold-changes represent transcripts down-regulated in AGD resistant compared to AGD susceptible Atlantic salmon, while positive fold-changes represent transcripts up-regulated in AGD resistant compared to AGD susceptible Atlantic salmon. The standard errors (SE) represent the variation between individuals within the resistant group compared to individuals within the susceptible group.

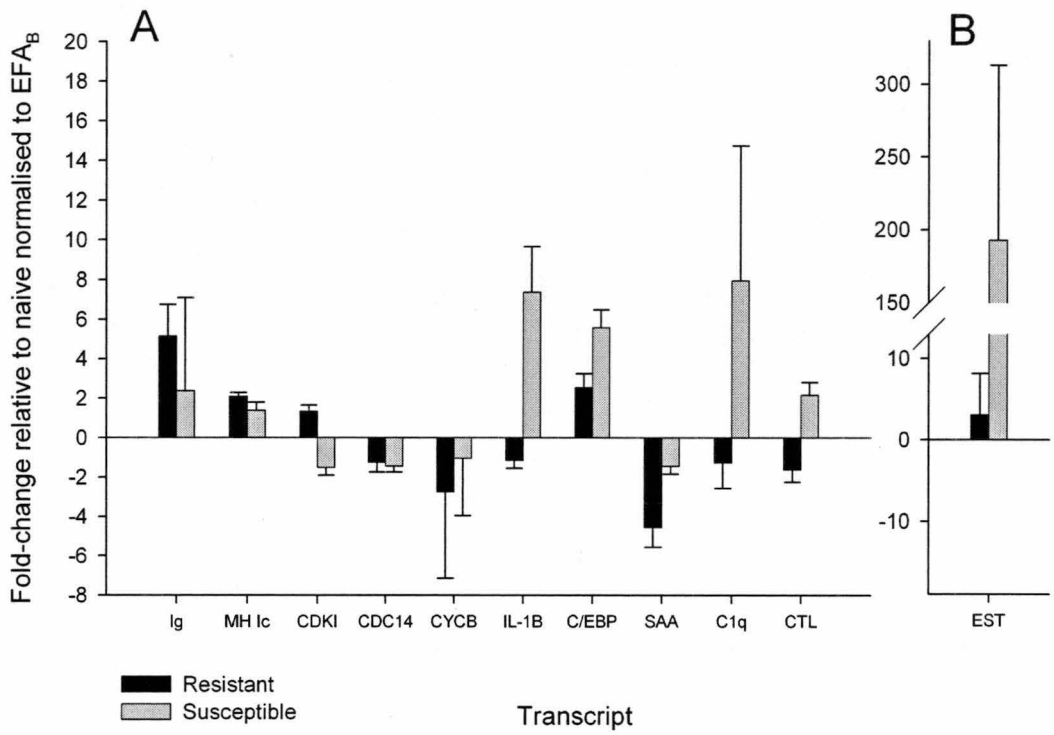


Figure 5.6. The fold-change of a sub-sample of transcripts determined by real-time quantitative RT-PCR (qPCR) of AGD resistant (n=14) and susceptible (n=14) Atlantic salmon compared to naïve Atlantic salmon and normalised to elongation factor $1A_B$. See Table 1 for acronym descriptions for transcripts. Standard errors represent the variation within the resistant group and within the susceptible group compared to naïve Atlantic salmon smolt. Note, scale change on y-axis in part A and part B.

The transcripts involved in cell cycle control also showed differential expression in the AGD affected individuals relative to naïve samples, albeit to a lesser degree than the acute phase proteins. Cyclin B showed decreased expression in both AGD resistant and susceptible Atlantic salmon relative to the naïve individuals. In contrast, compared to naïve fish, the cyclin dependent kinase inhibitor p13 (CDKI) showed a small up-regulation in the resistant individuals and a down-regulation in the susceptible individuals.

5.4 Discussion

Colonisation of the gill by *Neoparamoeba perurans* causes extensive morphological alterations of the gill structure (Adams and Nowak 2001). Although the mechanisms by which the parasite induces such a cellular response remain unclear, some individuals demonstrate a remarkable ability to resist these mechanisms (Taylor et al. 2007). The current study suggests that resistance to AGD is acquired following multiple infection events rather than being innately present. The molecular mechanisms influencing the ability of certain animals to acquire this resistance is a logical avenue worth exploring.

The ability of Atlantic salmon to develop an adaptive immune response against AGD has been the subject of multiple studies (Findlay and Munday 1998; Gross et al. 2004b; Vincent et al. 2006). While the ability of Atlantic salmon to develop anti-*Neoparamoeba* spp. antibodies appears unequivocal (Gross et al. 2004a), their functional relevance to AGD resistance has only recently been demonstrated (Vincent et al. 2006). The results obtained in our study – namely the up-regulation of certain adaptive immune related genes - support the hypothesis that AGD resistance may be associated with an adaptive immune response. Indeed, the up-regulation of transcripts resembling genes such as MH class II invariant chain-like protein and immunoglobulin light chain in the AGD resistant individuals may be an indicator of enhanced antigen presentation and immunoglobulin production within these animals.

Acting as an MH class II chaperone, the class II associated invariant chain stabilises and promotes the exit of MH class II molecules from the endoplasmic reticulum to the endosomal pathway where non-self peptides can be bound (Lotteau et al. 1990). Mice

deficient in the expression of MHC class II invariant chain generally demonstrate lower cell surface expression of MHC class II (Viville et al. 1993). In addition to its role in antigen presentation, MH class II invariant chain can also stimulate B cell maturation (Matza et al. 2003). In our study both the AGD resistant and susceptible Atlantic salmon showed higher expression of the MH class II invariant chain-like protein compared to naïve animals, although the expression in the susceptible individuals was only slightly greater than that of the naïve individuals. As a result of increased invariant chain expression, resistant individuals may have a superior ability to present *Neoparamoeba* derived antigens and stimulate B cell maturation, which would be expected to contribute to an enhanced immune response.

Relative to both susceptible and non-infected naïve animals, immunoglobulin light chain was up-regulated in the resistant fish. Furthermore, expression in susceptible fish was also slightly higher than that in naïve animals. This result is in contrast to the expression of various transcripts resembling immunoglobulin heavy chains, which upon first infection are down-regulated in AGD infected salmon (Young et al. 2008a). The increased expression of immunoglobulin in the AGD resistant animals could at least be partly a function of enhanced B cell maturation and subsequent immunoglobulin production. The up-regulation of CD8 and granzyme K in the resistant fish may also suggest a movement of cytotoxic T lymphocytes into the gills of AGD resistant fish. Therefore, it is possible that the resistant fish are displaying both a humoral and cellular adaptive immune response to AGD. Further research concentrating on comparing the different cell types between the gills of AGD resistant and susceptible Atlantic salmon is required.

The acute phase response (APR) is central to the innate immune system. In response to tissue injury or infection a dramatic increase or decrease of certain plasma proteins – known as the acute phase proteins (APP) – will occur (Gabay and Kushner 1999). The APR is initially stimulated by pro-inflammatory cytokines (IL-1, IL-6 and tumor necrosis factor alpha) which lead to the translocation of the transcription factors NF-κβ and C/EBPβ. Following sequential phosphorylation, the transcription factor C/EBPβ is capable of binding to the promoter and in turn inducing the expression of many APP (Tang et al. 2005; Kim et al. 2007). Previous research in rainbow trout cells has demonstrated that stimulation with recombinant IL-1β will increase the expression of C/EBP and certain other APP (Martin et al. 2007b). The current study demonstrated higher expression of IL-1β and C/EBPβ within those animals susceptible to AGD. Over-expression of the APP serum amyloid A, C1q complement component, DRTP1, transferrin and C-type lectin receptors A and B was also observed in the AGD susceptible animals compared to the resistant animals.

In an earlier study we have reported that, upon first infection, Atlantic salmon demonstrate little or no APR to AGD relative to naïve individuals (Wynne et al. 2008). The over-expression of genes encoding APP within the susceptible animals was therefore unexpected. It seems likely that the up-regulation of APR genes in the susceptible individuals may, to some extent, provide increased protection to AGD which allowed the individuals to survive until the trial termination, despite displaying significant pathology. The ability of an AGD affected animal to progress from this APR to an adaptive or acquired immune response, like that seen in the resistant individuals, may ultimately determine whether that individual will develop resistance to AGD.

Although AGD is a parasitic disease, it also has many hallmarks of a proliferative condition. Fish displaying advanced stages of AGD demonstrate considerable cellular proliferation within the gill. In addition to the dysregulation of immune related genes, the microarray analysis revealed a range of transcripts involved in the regulation of the cell cycle were also differentially expressed between AGD resistant and AGD susceptible Atlantic salmon. The commitment of cells to re-enter the cell cycle is controlled by a range of positive and negative factors. Positive factors, such as cyclins and their specific cyclin-dependent kinases (CDK), stimulate cell cycle progression (Malumbres and Barbacid 2001). Negative factors, such as cyclin-dependent kinase inhibitors (CDKI) and tumor suppressor loci, act to induce cell cycle arrest. The microarray study demonstrated over-expression of two transcripts with high homology to cyclin B in the individuals susceptible to AGD compared to the resistant fish. However, closer examination with qPCR revealed that, when compared to non-infected naïve animals, cyclin B expression was actually down-regulated in the resistant animals and furthermore showed little difference in expression in the susceptible fish. Although somewhat speculative, this down-regulation of cyclin B may in fact be mediated by the CDC14 homolog A phosphatase and the growth arrest and DNA-damage (GADD) proteins, which are both important antagonists of the cyclin B-CDK1 complex (Kipreos 2004; Gao et al. 2005) and were found to be more highly expressed in the AGD resistant Atlantic salmon. GADD has previously been reported to be down-regulated in AGD affected Atlantic salmon (Morrison et al. 2006a). The suppression of cyclin B, through CDC14 and GADD, may be an important mechanism by which the resistant animals can limit the cellular proliferation associated with AGD.

Cell cycle arrest can also be induced through the up-regulation of tumor suppressor loci. Previous research has demonstrated that the tumor suppressor p53 is down-regulated in AGD affected Atlantic salmon (Morrison et al. 2006a). The current study demonstrates that three putative tumor suppressor loci, p13 CDKI (also known as CDKN2X), TSC-22 (TFGβ stimulated clone-22) and neuroblastoma, suppression of tumorigenicity 1, were up-regulated in the AGD resistant Atlantic salmon compared to susceptible fish. Relative to non-infected naïve Atlantic salmon, p13 CDKI was up-regulated in the resistant fish while it was down-regulated in susceptible fish. The p13 gene shares strong similarity with other members of the mammalian CDKN2 family of tumor suppressor loci, namely, CDKN2A (p16) and CDKN2B (p15) (Kazianis et al. 1999) and regulates the cell cycle progression by binding and inactivating CDK4 and CDK6. The higher expression of tumor suppressor loci in the AGD resistant animals is a likely mechanism responsible for limiting the cellular proliferation in response to the infection.

By definition the resistant animals generally displayed few or no AGD lesions on the gill surface. In contrast, the susceptible individuals had a large number of lesions covering almost the entire gill surface. This means comparing resistant with susceptible animals was in fact also comparing lesioned with non-lesioned gill material. Lesion restricted expression of certain genes has been reported previously for AGD affected Atlantic salmon (Morrison et al. 2006a; Morrison et al. 2007; Young et al. 2008a). We therefore expected that some of the differentially expressed transcripts were simply caused by the different cell types present within a lesion rather than resistance/susceptibility per se. To further investigate this situation we included naïve individuals in the qPCR analysis. The expression of IL-1β and C1q in the resistant fish

was almost identical to that of the naïve. In contrast, the susceptible fish had considerably higher expression of IL-1β and C1q compared to the naïve individuals. This result suggests the up-regulation of IL-1β and C1q is probably a consequence of infection, rather than a down-regulation in the resistant animals. Conversely, the expression of immunoglobulin light chain was highest in the resistant animals compared to naïve, suggesting the expression of this transcript is indeed associated with resistance.

In conclusion, we report that the ability of Atlantic salmon to resist the effects of AGD is associated with dysregulation of genes involved in the cell cycle and the adaptive immune response. This study provides preliminary insights into what appears to be acquisition of resistance to AGD. The challenge remains to fully elucidate the role that the adaptive immune response plays in this. Finally, the future identification of quantitative trait loci associated with these pathways may provide valuable insight into the host genetic influences controlling AGD resistance. This will facilitate a more effective exploitation of AGD resistance as a commercial selective breeding trait.

Chapter 6: Further investigation of the unknown transcript CK880278

Abstract

Two previous transcriptome profiling experiments have identified one unknown transcript (accession number CK880278) which shows particularly high differential expression between naïve and amoebic gill disease (AGD) affected Atlantic salmon. This chapter describes a study conducted to further investigate this transcript in an attempt to elucidate its possible function. Briefly, 3' and 5' rapid amplification of cDNA ends was performed to obtain the full length 675bp mRNA sequence. Next, by means of semi-quantitative reverse transcription PCR, I examined the expression distribution in different organs of a naïve Atlantic salmon smolt. Expression was observed in all organs examined, although highest in the anterior kidney and lowest in the muscle and skin. *In situ* hybridisation of digoxigenin labelled RNA probe on gill sections from naïve Atlantic salmon demonstrated that expression of this transcript is isolated to epithelial pavement cells within the gill. In contrast, *in situ* hybridisation on gill sections derived from AGD affected Atlantic salmon demonstrated high expression of this transcript in almost all proliferating cells within the AGD lesion. Despite obtaining the full length sequence the identity of this transcript could not be determined by the use of the basic local alignment search tool (BLAST). Three open reading frames were predicted, which ranged in size between 41 amino acids (aa) and 111aa. In an attempt to identify the correct open reading frame, recombinant proteins for two of the three hypothetical protein sequences were generated. These proteins will be used in a later study to develop polyclonal antibodies for western blot analysis of Atlantic salmon crude lysate. In conclusion, this study describes the full length mRNA sequence and expression characteristics of a transcript previously shown to be highly induced in AGD affected Atlantic salmon. Further research is required to determine which open reading

frame encodes the actual protein and what functional relevance this protein has on AGD. Completing this additional research was beyond the time constraints of my candidature.

6.1 Introduction

AGD is a parasite mediated proliferative disease capable of affecting a range of cultured teleosts (Munday et al. 2001). Caused by the protozoan *Neoparamoeba perurans* (Young et al. 2007; Young et al. 2008c), individuals affected by AGD display extensive hyperplasia of the gill epithelium (Adams and Nowak 2001; Young et al. 2008c). The cellular proliferation in the gill is associated with a down-regulation of genes involved in apoptosis (Chapter 4), cell cycle inhibitors (Chapter 5) and tumor suppressor loci (Morrison et al. 2006a). Genes involved in the major histocompatibility (MH) class I and class II pathways are also down-regulated within the lesions of AGD affected Atlantic salmon (Young et al. 2008a). The mechanism by which the amoebae are capable of mediating the cellular alterations associated with AGD remains unclear.

Previous transcriptome profiling experiments have identified one transcript that displays significant up-regulation in response to AGD in both laboratory based challenge trials (Chapter 4) and natural AGD infections (Chapter 5). Compared to naïve Atlantic salmon, individuals infected by AGD demonstrated an up-regulation of this transcript by over 180-fold. In addition, individuals with higher disease severity display higher expression of this transcript compared to individuals with low disease severity (Chapter 4). Due to its considerable up-regulation, I hypothesise that this gene may have an important role in AGD pathogenesis and is thus worthy of further investigation.

Unfortunately the identity of this transcript remains unknown. Interrogation of nucleotide collections such as databases maintained by the National Centre of

Biotechnology and Information (NCBI) reveal no homology to any known sequence. The clone was originally derived from an anterior kidney cDNA library produced by the Salmon Genome Project (SGP) (Adzhubei et al. 2007). It was subsequently included on the newly developed Transcriptome Analysis of Commercially Important Traits (TRAITS)/SGP cDNA microarray (Martin et al. 2007a).

Determining the identity of a nucleotide sequence with no sequence homology to any previously reported sequence is difficult. Some of the major steps to achieve this objective are illustrated in Figure 6.1. The first step is to obtain the full length mRNA sequence, often undertaken using 3' and 5' rapid amplification of cDNA ends (RACE). BLAST may then be used to search for regions of local similarity between the queries sequences within various databases. If no sequence homology can be detected using BLAST, the next most important step is to determine if the nucleotide sequence actually encodes a functional protein.

To determine if a particular sequence encodes a functional protein the open reading frame/s must first be identified through computational prediction. The hypothetical protein sequences are then used to develop recombinant proteins in either prokaryotic (most commonly in *Escherichia coli*) or eukaryotic cell lines. Despite certain post transcriptional modifications (especially in the prokaryote system) and engineered fusion tags, these recombinant proteins should imitate the actual functional protein as closely as possible. The recombinant proteins can then be used to immunise rabbits for the production of polyclonal antibody sera. These antibodies are subsequently used for western blot analysis against cell lysate derived from an organ or cell culture that is known to expresses this protein.

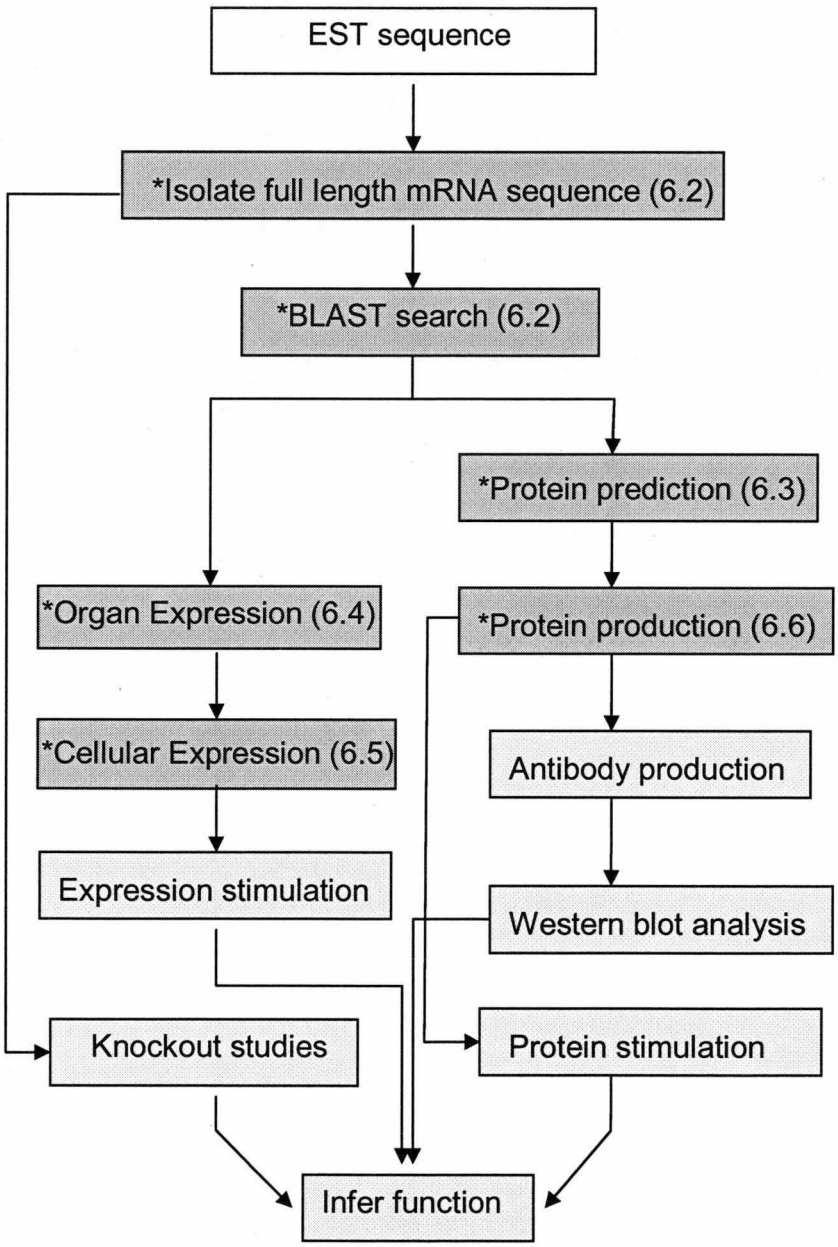


Figure 6.1. Flow diagram of steps involved in the characterisation of the unknown EST CK880278. * represents steps completed in this study. The chapter sections describing each stage are indicated in *parenthesis*.

Although determining if the sequence encodes an actual protein is important, it provides little information regarding the possible function of the gene. For this purpose it may be appropriate to examine the mRNA expression distribution in organs and cells and how this is affected by different stimuli (e.g. lipopolysaccharide). This may help identify possible roles of the gene.

To further elucidate the potential role functional studies are required. Silencing the gene expression through RNA interference (RNAi) is one such approach. Following the knockout or at least knockdown of expression the biological and transcriptional responses are assessed. The dysregulation of biological and transcriptional responses may provide some indication of the role of the gene. Another strategy is to use the recombinant protein to stimulate cell lines derived from the species under investigation. When combined with transcriptome profiling this strategy may prove useful to identify dysregulation of genes and pathways in response to the uncharacterised protein. This approach has been previously used to study the transcriptome response in salmonid cells to stimulation with both interleukin-1β (IL-1β) (Martin et al. 2007b) and interferons (Martin et al. 2007a). Despite the use of these strategies described above, it must be conceded that in many cases the role of a protein can simply not be determined with current molecular procedures.

The present study describes the first steps taken to characterise the unknown transcript (accession number CK880278) previously identified through microarray experiments. More specifically, I first obtained the full length sequence and then examined its expression in the different tissues, including AGD affected gill tissue. Next,

recombinant proteins were generated against the predicted open reading frames and will be used to produce polyclonal antibodies for western blot analysis. Unfortunately due to time restraints, the full series of experiments required to assign a possible function to this gene could not be completed. This chapter is presented in a step-by-step structure where each experiment is presented and discussed in sequential order. The direction of future research is considered at the conclusion of this sequence.

6.2 Isolation of full length mRNA sequence

6.2.1 Materials and Methods

The full length mRNA sequence was obtained using 3' and 5' RACE. All gene specific primer sequences used in RACE were designed based on the EST sequence CK880278. To achieve 3' RACE 1 μg of total RNA from the gill of a single AGD affected Atlantic salmon was quantified with a Qubit fluorometer (Invitrogen, Mount Waverly, Australia) and reverse transcribed using superscript III reverse transcriptase (Invitrogen) and a custom oligo dT. This oligo, designated Race-dT, contains an adapter sequence on the 5' end (Table 6.1). Reverse transcription was performed following the manufacturer's protocol (Invitrogen) with the additional RNase H digestion for 20 min at 37°C. The first round amplification was performed in a total volume of 50 μl containing 5 μl of 10X buffer, 2.5 mM of $MgCl_2$, 500 μM of dNTPs, 400 μM of the gene specific forward primer (EST03G12-20F), 400 μM of adapter reverse primer (Adapt-dT, Table 6.1), 2 units of AmpliTaq-Gold (Applied Biosystems, Foster City, USA) and approximately 50 ng (starting RNA) of template. Reactions were subjected to 10 min at 95°C followed by 37 touchdown cycles of 95°C for 1 min, 60°C decreasing in 1°C increments every 2 cycles for 1 min with an additional 5 cycles at 44°C, 72°C for 3 min and a final elongation step of 72°C for 7 min. Cycling was performed in an Eppendorf thermal cycler (Eppendorf, Hamburg, Germany). Resulting product was diluted 1:20 in water and subjected to a second round amplification using the gene specific primer, EST03G12-50F (Table 6.1), and the Adapt-dT primer. Reactions were performed in a total volume of 50 μl as above, however standard cycling of 95°C for 10 min, 35 cycles

Primer name	Primer sequence 5'- 3'
Race-dT	CTGGAGATCGATGCGGCCGCTTTTTTTTTTTTTTTT
Adapt-dT	CTGGAGATCGATGCGGCCGC
EST03G12-9F	AGAACACCTGATTACTACAACCTTCA
EST03G12-20F	CCAGGACAATACCATCACTCATAC
EST03G12-50F	ACAAACAACTCCCAGAAAAACTG
EST03G12-374R	CCTACATTGATTACATTGTTGAATTCG
EST03G12-324R	GTAGCACTTTTTCACTTATCTGAGATG
EST03G12-615R	ACCTACATTGATTACATTGTTGAATTC
ORF111-F	CACCGAACACCTGATTACTACAACCTTCAA
ORF111-R	CTATGGTGATGCTGGGTGC
ORF71-F	CACCGATTGTGCTTCAACTGTCGTC
ORF71-R	TTAAGCGGTATGAGTGATGGTATT
ORF41-F	CACCGCTCTCTGTGACCTCATCGAA
ORF41-R	TCACTTTCTACTTGGAAGCCAGT
BACT-F	TCTCTGGAGAAGAGCTAC
BACT-R	TTAGAAGCATTTGCGGAG

Table 6.1. Primers utilised in this study. The primer name and sequence is indicated. Nucleotides *underlined* represent bases required for Gateway directional cloning (Invitrogen).

of 95°C for 1 min, 55°C for 1 min and 72°C for 3 min and a final elongation step of 72°C for 7 min was used. Resulting PCR product was visualised on agarose gel and then directly sequenced using the primer EST03G12-50F (Table 6.1) and Adapt-dT with the Applied Biosystems BigDye Terminator Kit (Applied Biosystems) as per the manufacturer's protocol. Sequences were obtained using the ABI 3100 capillary sequencer (Applied Biosystems).

To achieve 5' RACE the GeneRacer kit (Invitrogen) was used. Following quantification (Qubit fluorometer; Invitrogen), 2 μg of total RNA was dephosphorylated, decapped and 5'oligos ligated exactly as stipulated by the manufacturer's protocol. RNA was then reverse transcribed using superscript III (Invitrogen) as per the manufacturer's protocol. The first round amplification conditions were identical to the 3' RACE first round conditions (including the touchdown cycling). The gene specific reverse EST03G12-374R primer (Table 6.1) and the commercial forward GeneRacer 5' primer were used. Resulting PCR product was diluted 1:20 in water and a second round nested amplification was performed using the same PCR conditions as the first round. The gene specific reverse primer, EST03G12-324R (Table 6.1), and the commercial forward GeneRacer Nested 5' primer was used for the second round amplification. PCR product was cloned into pGEM-T easy (Promega, Annandale, Australia) following the manufacturer's protocol and six clones were sequenced using the T7 and Sp6 primers with the Applied Biosystems BigDye Terminator Kit (Applied Biosystems) and ABI 3100 capillary sequencer (Applied Biosystems).

Full length sequences were amplified using the primers EST03G12-9F and EST03G12-615R. PCR was performed in a 50μl reaction containing 5μl of 10X buffer, 2 mM of

$MgCl_2$, 500 μM of dNTPs, 400 μM of the gene specific forward primer (EST-9F), 400 μM of reverse primer (EST-615R), 1μl of PfuUltra II Fusion HS DNA polymerase (Stratagene, La Jolla, USA). Reactions were first subjected to 2 min at 95°C and then 35 cycles of 1 min at 95°C, 1 min at 56°C and 2 min at 72°C. A final extension step at 72°C for 3 min was then used. PCR product was purified using the Qiagen MinElute system (Qiagen) as per the manufacturer's protocol. Six microliters of PCR product was A tailed to facilitate TA cloning. Brefily 6μl of PCR product was incubated with 1μl of 10X buffer, 10 μM of dATP, and 5 units of *Taq* polymerase (Finnzyme, Espoo, Finland) at 70°C for 30 min. The PCR product was then cloned and 19 clones sequenced as described above.

2.2 Results and Discussion

The 3' RACE resulted in only the addition of poly (A) tail. However, the 5' RACE extended the sequence by 245bp to produce a maximum full length sequence of 675bp (not including the poly A) (Figure 6.2). Sequences derived from the six RACE clones showed some minor nucleotide variation. This variation was for the most part concentrated to base pairs 7-10 and a string of A bases at position 135-146. Considering a non-proof reading *Taq* polymerase was used for this RACE reaction it is likely some of these nucleotide variations were caused by *Taq* errors, especially at position 135-146. Primers to amplify the full length sequence were designed based on the conserved regions of these RACE clones.

Amplification, cloning and sequencing of 19 full length clones demonstrated that two sequence variants (not including single nucleotide polymorphisms) were obtained when

```
Clone1   ACG CGG G-- --A GAA CAC CTG ATT ACT ACA ACC TTC AAG CGC GCC ATC GCA TCG  [ 54]
Clone2   ... ... .-- --. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 54]
Clone3   ... ... .G- --. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 54]
Clone4   ... ... .G- --. ... ... ... ... ... ... ... ... ... ... ... ... A.. ...  [ 54]
Clone5   ... ... .-A CT. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 54]
Clone6   ... ... .GA CT. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 54]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [ 54]

Clone1   AGG AAG AAC ACA CAG ACA CAG CAA TGG ATT GTG CTT CAA CTG TCG TCA CTT CAC  [108]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [108]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [108]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [108]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [108]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [108]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [108]

Clone1   ATA CCA ACA GCA AAA ACG ACT GCA ATA AAA AAA AA- -TG TGT TGC CCC ATG TCC  [162]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ..- -.. ... ... ... ... ...  [162]
Clone3   ... ... ... ... ... ... ... ... ... ... ... .-- -.. ... ... ... ... ...  [162]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ..A -.. ... ... ... ... ...  [162]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ..A A.. ... ... ... ... ...  [162]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ..- -.. ... ... ... ... ...  [162]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [162]

Clone1   TAC TGA ATC AGG CAA AAA CAT ATC AGG TCA AAA CAT ATC AGG CCA AAA CAT ATC  [216]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [216]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [216]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [216]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [216]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [216]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [216]

Clone1   AGG CCA AAA CAT ATC AGG CAG TGT GGG CGG AAG TAT TTC CAT TGG CAA CCA GGA  [270]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [270]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [270]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [270]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [270]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [270]
CK880278 --- --- --- --- --- --- --- --- --- --. ... ... ... ... ... ... ... ...  [270]

Clone1   CAA TAC CAT CAC TCA TAC CGC TTA ACA AAC GAC TCC CAG AAA AAC TGT CGC ACC  [324]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [324]
Clone3   ... ... .G. ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...  [324]
Clone4   ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...  [324]
Clone5   ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...  [324]
Clone6   ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...  [324]
CK880278 ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...  [324]

Clone1   CAG CAT CAC CAT AGC CAA CAT AAC AGG TGG CAA AGT GGG TCT CTA CTA CAA GGC  [378]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]
CK880278 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [378]

Clone1   AGA GAA GAT GGC TCT CTG TGA CCT CAT CGA AGG AAG TCA ACT CAG GAA GGG CC [431]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
CK880278 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .. [431]
```

Figure 6.2 Sequence alignment of six clones produced from 5'RACE of the EST CK880278. The 5' region of the EST CK880278 is also shown. *Dots* indicate sequence identities to Clone1 and *dashes* indicate gaps within the alignment.

using these primers. Seventeen of these clones were almost identical to the EST CK880278 at their 3' region (Figure 6.3). Two clones showed considerable nucleotide divergence to both the other 17 clones and the EST CK880278. Considering the sequences derived from the 17 clones were almost identical to the 5' region of the EST CK880278 it seems most likely these clones represent the actual full length transcript of this EST. Given the sequence divergence of clones 7 and 19, these clones probably reflect non-specific amplification rather than a variant allele. Future research using southern blots will be required to confirm the presence of a single locus.

The mRNA sequence contained a polyadenylation site and some putative AU-rich translation inhibitor elements. BLASTx analysis of the full length sequences against the non-redundant (nr) protein database maintained by the NCBI revealed only weak sequence homology. The most significant BLASTx hit was to the hypothetical protein BURPS1710b_1103 from the bacteria *Burkholderia pseudomallei* ($E = 0.007$). BLASTn against the NCBI nr nucleotide collection revealed no homology to any sequence. However BLASTn against the NCBI EST database revealed significant homology ($E = 0$) to a number of Atlantic salmon EST sequences, almost all of which were assigned to the non-annotated UniGene cluster Ssa.25836. Interrogation of the tissue expression distribution of the 71 sequences comprising the UniGene cluster Ssa.25836 revealed this transcript is most highly expressed in the thymus, thyroid and anterior kidney of Atlantic salmon. The lack of significant BLAST results may suggest this transcript is either unique to Atlantic salmon, or more likely, shows extensive divergence to its ancestral ortholog.

Despite the lack of significant BLAST hits it is interesting to note that there is an EST sequence from *Salvalinus fontinalis* that is 94% similar and larger in size than CK880278 and when this sequence is analysed with BLAST it has a hit to a growth factor independent zinc finger protein (Gfi-1). The significance of this alignment is relatively low $E = 5x10^{-8}$ compared to the threshold employed in the previous chapters and therefore makes the annotation more uncertain.

```
Clone1   AGA ACA CCT GAT TAC TAC AAC CTT CAA GCG CGC CAT CGC ATC GAG GAA GAA CAC ACA GAC ACA GCA ATG GAT TGT GCT TCA ACT  [ 84]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone9   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone10  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone11  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone12  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone13  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone14  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone15  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone16  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone18  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone17  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 84]
Clone19  ... ... ... ... ... ... ... ... ..G T.. ... ..C T.. .C. ... ... ..G ... ... ... ... .-- --- -TA ATG .GA G.G GT.  [ 84]
Clone7   ... ... ... ... ... ... ... ... ..G T.. ... ..C T.. .C. ... ... ..G ... ... ... ... .-- -CA .TA ATG .GA G.G GT.  [ 84]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [ 84]

Clone1   GTC GTC ACT TCA CAT ACC AAC AGC AAA AAC GAC TGC AAT AAA AAA AAA TGT GTT GCC CCA TGT CCT ACT GAA TCA GGC AAA AAC  [168]
Clone2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone9   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone10  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone11  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone12  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone13  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone14  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone15  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone16  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone18  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone17  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [168]
Clone19  .CT TCA CA. GAT GG. G.T .CA .AT .GT GGT ACT .CA ... .CC ..C .GT CA. AAC .A. AAC C.. GT. GA. A.C AGT CAT ..G GTA  [168]
Clone7   .CT TCA CA. GAT GG. G.T .CA .AT .GT GGT ACT .CA ... .CC ..C .GT CA. AAC .A. AAC C.. GT. GA. A.C AGT CAT ..G GTA  [168]
CK880278 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [168]
```

Figure 6.3. Continued over page

```
Clone1    ATA TCA GGT CAA AAC ATA TCA GGC CAA AAC ATA TCA GGC CAA AAC ATA TCA GGC AGT GTG GGC GGA AGT ATT TCC ATT GGC AAC  [252]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone17   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [252]
Clone19   TC. GG. CAA A.C C.A .AC AT. TCA GGC C.A .GC A.G AT. AGT GG. GG. GTT ACA TT. .GC AA. ATG .CC CAG A.. .CC ACT C.G [252]
Clone7    TC. GG. CAA A.C C.A .AC AT. TCA GGC C.A .GC A.G AT. AGT GG. GG. GTT ACA TT. .GC AA. ATG .CC CAG A.. .CC ACT C.G [252]
CK880278  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -.. ... ... ... ... ... ... [252]

Clone1    CAG GAC AAT ACC ATC ACT CAT ACC GCT TAA CAA ACA ACT CCC AGA AAA ACT GTC GCA CCC AGC ATC ACC -AT AGC CAA CAT AAC  [336]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone17   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... -.. ... ... ... ... [336]
Clone19   ACC .CT C.A C.. GGA ... .CC CAA AAA ACT GTC T.. C.C AG. .TC .CC .TA .C. AAC GTT .CA GGT GG. A.A GTG .GT .TC T.. [336]
Clone7    ACC .CT C.A C.. GGA ... .CC CAA AAA ACT GTC T.. C.C AG. .TC .CC .TA .C. AAC GTT .CA GGT GG. A.A GTG .GT .TC T.. [336]
CK880278  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ?.. ... ... ... ... [336]
```

Figure 6.3. Continued over page

```
Clone1    AGG TGG CAA AGT GGG TCT CTA CTA CAA GGC AGA GAA GAT GGC TCT CTG TGA CCT CAT CGA GGG AAG TCA ACT CAG GAA GGG CCA  [420]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone17   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [420]
Clone19   TAC AA. GT. GAG AA. AT. TCT ..C G.T .A. CTC ATT ..G ..G AG. .AA CTC AAG A.G G.C .AA TCC ATG CTC ATC CTG .AA GGC  [420]
Clone7    TAC AA. GT. GAG AA. AT. TCT ..C G.T .A. CTC ATT ..G ..G AG. .AA CTC AAG A.G G.C .AA TCC ATG CTC ATC CTG .AA GGC  [420]
CK880278  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... [420]

Clone1    CTC CAT TCT CAT CCT GGA AAA CAC CTT TAA GAA GAC CCC TCT TCC TCC TCT ATG TCA CTG GCT TCC AAG TAG AAA GTG ACA ATA  [504]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
Clone17   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .A. [504]
Clone19   AC. TT. AAG A.G A.C CCC CTC .CT .AG AGC CTC T.T GT. A.. GG. .T. CAA G.A GA. TGT AGC AAT .CA C.T .GC TGC C.T GAG  [504]
Clone7    AC. TT. AAG A.G A.C CCC CTC .CT .AG AGC CTC T.T GT. A.. GG. .T. CAA G.A GA. TGT AGC AAT .CA C.T .GC TGC C.T GAG  [504]
CK880278  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [504]
```

Figure 6.3. Continued over page

```
Clone1    CAT ACA GAT TAA ACT ACA TAC AAA TTA TTG ATT AAA AAT AAA GAG AGA ATA CAT CTC AGA TAA GTG AAA AAG TGC TAC AAG ATA  [588]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]
Clone17   ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ...  [588]
Clone19   G.A T.C CCA CTG .A. GGG CTG G.C AAC C.C C.A GC. G.G T.T .CT GAG GGT ..G .C. .AG .CC TAC .TC CTC ACT GTT G.C .A.  [588]
Clone7    G.A T.C CCA CTG .A. GGG CTG G.C AAC C.C C.A GC. G.G T.T .CT GAG GGT ..G .C. .AG .CC TAC .TC CTC ACT GTT G.C .A.  [588]
CK880278  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [588]

Clone1    TAC TTT GAT TGT ATG GTC GAA TTC AAC AAT GTA ATC AAT GTA GGT --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone6    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone9    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone10   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone11   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone12   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone13   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone14   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone15   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone16   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone18   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone4    ... ... ... ... ... T.. A.. ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone17   ... ... ... ... ... T.. A.. ... ... ... ... ... ... ... ... --- --- --- --- --- --- --- --- --- --- --- --- ---  [672]
Clone19   G.T G.G A.G CTA GCC AAG A.. GGA G.G .TC TGG G.G G.. .AG .AC GGG GTG GAT CAC AAC CCC AAA AAG TAA CGC CAC CAG GAA  [672]
Clone7    G.T G.G A.G CTA GCC AAG A.. GGA G.G .TC TGG G.G G.. .AG .AC GGG GTG GAT CAC AAC CCC AAA AAG TAA CGC CAC CAG GAA  [672]
CK880278  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... TTT TGA CAA AAA AAT AAA TGG TAC CAA ACT G-- --- ---  [672]
```

Figure 6.3. Continued over page

```
Clone1    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone3    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone5    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone6    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone8    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone9    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone10   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone11   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone12   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone13   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone14   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone15   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone16   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone18   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone4    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone17   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]
Clone19   ACC TCC TGT ATC TTA ACA CCC TTT ATT AAT CAC AGA TTA AAC TAC ATA CAA ATT ATT GAT TAA AAA TAA AGA TAA GTG AAA TAG [756]
Clone7    ACC TCC TGT ATC TTA ACA CCC TTT ATT AAT CAC AGA TTA AAC TAC ATA CAA ATT ATT GAT TAA AAA TAA AGA TAA GTG AAA TAG [756]
CK880278  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- [756]

Clone1    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone3    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone5    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone6    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone8    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone9    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone10   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone11   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone12   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone13   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone14   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone15   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone16   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone18   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone4    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone17   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
Clone19   TGC TTC AAG ATA CAC TGA GTG TAT GGT AGA ATT CAA CAA TGT AAT CAA TGT AGG T- [812]
Clone7    TGC TTC AAG ATA CAC TGA GTG TAT GGT AGA ATT CAA CAA TGT AAT CAA TGT AGG T- [812]
CK880278  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- [812]
```

Figure 6.3. Sequence alignment of 19 full length clones of the EST CK880278. The 5' region of the EST CK880278 is also shown. *Dots* indicate sequence identities to Clone1 and *dashes* indicate gaps within the alignment.

6.3 Protein prediction

6.3.1 Materials and Methods

ORF Finder (www.ncbi.nlm.nih.gov/gorf) was used to predict the open reading frame/s of the full length transcript. Primers were then designed – based on the full length sequence – to amplify each of the predicted hypothetical proteins (Figure 6.4). The PROSITE

```
  1 ACGCGGGGAC TAGAACACCT GATTACTACA ACCTTCAAGC GCGCCATCGC ATCGAGGAAG
                    ORF111-F
 61 AACACACAGA CACAGCAATG GATTGTGCTT CAACTGTCGT CACTTCACAT ACCAACAGCA
                            ORF71-F
121 AAAACGACTG CAATAAAAAA AAATGTGTTG CCCCATGTCC TACTGAATCA GGCAAAAACA

181 TATCAGGTCA AAACATATCA GGCCAAAACA TATCAGGCCA AAACATATCA GGCAGTGTGG

241 GCGGAAGTAT TTCCATTGGC AACCAGGACA ATACCATCAC TCATACCGCT TAACAAACGA
                                    ORF71-R
301 CTCCCAGAAA AACTGTCGCA CCCAGCATCA CCATAGCCAA CATAACAGGT GGCAAAGTGG
                         ORF111-R
361 GTCTCTACTA CAAGGCAGAG AAGATGGCTC TCTGTGACCT CATCGAAGGA AGTCAACTCA
                                  ORF41-F
421 GGAAGGGCCA CTCCATTCTC ATCCTGGAAA ACACCTTTAA GAAGACCCCT CTTCCTCCTC

481 TATGTCACTG GCTTCCAAGT AGAAAGTGAC AATACATACA GATTAAACTA CATACAAATT
          ORF41-R
541 ATTGATTAAA AATAAAGAGA GAATACATCT CAGATAAGTG AAAAAGTGCT ACAAGATATA

601 CTTTGATTGT ATGGTCGAAT TCAACAATGT AATCAATGTA GGTTTTTGAC AAAAAAATAA

661 ATGGTACCAA ACTGAAAAAA AAAAAAAAAA A
```

Figure 6.4. Full length mRNA nucleotide sequence of the unknown EST CK880278. Sequence obtained by 5' RACE is *underlined.* Primer sequences used to amplify the predicted open reading frames are indicated by *grey shading.*

database (Falquet et al. 2002) was used to search for motifs and patterns within each of the hypothetical protein sequences.

6.3.2 Results and Discussion

Three potential open reading frames were predicted in the forward orientation (Figure 6.5). These predicted protein sequences were 111aa, 71aa and 41aa in length. The 111aa sequence contained no start codon, while both the 71 and 41aa sequences contained start and stop codons. BLASTp analysis of each open reading frame against the NCBI nr protein database revealed no significant homology, likewise a conserved domain (CD) BLAST search also suggested no CDs were present within the predicted protein sequences.

The short length of all predicted hypothetical protein sequences – with one containing no start codon – raised some concern regarding whether the actual full length mRNA sequence had been obtained through RACE. To address this concern, I repeated the 5' RACE, this time using the SMART™ RACE kit (Clontech Laboratories, Mountain View, USA) as per the manufacturer's protocol. The full length sequence obtained from this kit (data not shown) was identical to the sequence obtain using the GeneRacer kit (Figure 6.2). I therefore concluded that the full length mRNA is indeed a 675bp fragment.

A number of phosphorylation and glycosylation sites were predicted across the hypothetical protein sequences (Figure 6.5). The 71aa hypothetical protein also

Figure 6.5 A

```
ACG CGG GGA CTA GAA CAC CTG ATT ACT ACA ACC TTC AAG CGC GCC ATC GCA TCG AGG AAG    60
 T   R   G   L   E   H   L   I   T   T   T   F   K   R   A   I   A   S   R   K    20

AAC ACA CAG ACA CAG CAA TGG ATT GTG CTT CAA CTG TCG TCA CTT CAC ATA CCA ACA GCA   120
 N   T   Q   T   Q   Q   W   I   V   L   Q   L   S   S   L   H   I   P   T   A    40

AAA ACG ACT GCA ATA AAA AAA AAT GTG TTG CCC CAT GTC CTA CTG AAT CAG GCA AAA ACA   180
 K   T   T   A   I   K   K   N   V   L   P   H   V   L   L   N   Q   A   K   T    60

TAT CAG GTC AAA ACA TAT CAG GCC AAA ACA TAT CAG GCC AAA ACA TAT CAG GCA GTG TGG   240
 Y   Q   V   K   T   Y   Q   A   K   T   Y   Q   A   K   T   Y   Q   A   V   W    80

GCG GAA GTA TTT CCA TTG GCA ACC AGG ACA ATA CCA TCA CTC ATA CCG CTT AAC AAA CGA   300
 A   E   V   F   P   L   A   T   R   T   I   P   S   L   I   P   L   N   K   R   100

CTC CCA GAA AAA CTG TCG CAC CCA GCA TCA CCA TAG                                  336
 L   P   E   K   L   S   H   P   A   S   P   *                                  111

CCAACATAACAGGTGGCAAAGTGGGTCTCTACTACAAGGCAGAGAAGATGGCTCTCTGTGACCTCATCGAAGGAAGTCA  415
ACTCAGGAAGGGCCACTCCATTCTCATCCTGGAAAACACCTTTAAGAAGACCCCTCTTCCTCCTCTATGTCACTGGCTT  494
CCAAGTAGAAAGTGACAATACATACAGATTAAACTACATACAAATTATTGATTAAAAATAAAGAGAGAATACATCTCAG  573
ATAAGTGAAAAAGTGCTACAAGATATACTTTGATTGTATGGTCGAATTCAACAATGTAATCAATGTAGGTTTTTGACAA  652
AAAAATAAATGGTACCAAACTG(A)n                                                       675
```

Figure 6.5 B

```
ACGCGGGGACTAGAACACCTGATTACTACAACCTTCAAGCGCGCCATCGCATCGAGGAAGAACACACAGACACAGCA     77

ATG GAT TGT GCT TCA ACT GTC GTC ACT TCA CAT ACC AAC AGC AAA AAC GAC TGC AAT AAA   137
 M   D   C   A   S   T   V   V   T   S   H   T   N   S   K   N   D   C   N   K    20

AAA AAA TGT GTT GCC CCA TGT CCT ACT GAA TCA GGC AAA AAC ATA TCA GGT CAA AAC ATA   197
 K   K   C   V   A   P   C   P   T   E   S   G   K   N   I   S   G   Q   N   I    40

TCA GGC CAA AAC ATA TCA GGC CAA AAC ATA TCA GGC AGT GTG GGC GGA AGT ATT TCC ATT   257
 S   G   Q   N   I   S   G   Q   N   I   S   G   S   V   G   G   S   I   S   I    60

GGC AAC CAG GAC AAT ACC ATC ACT CAT ACC GCT TAA                                  293
 G   N   Q   D   N   T   I   T   H   T   A   *                                  71

CAAACGACTCCCAGAAAAACTGTCGCACCCAGCATCACCATAGCCAACATAACAGGTGGCAAAGTGGGTCTCTACTACA  372
AGGCAGAGAAGATGGCTCTCTGTGACCTCATCGAAGGAAGTCAACTCAGGAAGGGCCACTCCATTCTCATCCTGGAAAA  451
CACCTTTAAGAAGACCCCTCTTCCTCCTCTATGTCACTGGCTTCCAAGTAGAAAGTGACAATACATACAGATTAAACTA  530
CATACAAATTATTGATTAAAAATAAAGAGAGAATACATCTCAGATAAGTGAAAAAGTGCTACAAGATATACTTTGATTG  609
TATGGTCGAATTCAACAATGTAATCAATGTAGGTTTTTGACAAAAAAATAAATGGTACCAAACTG(A)n           675
```

Figure 6.5. Continued over page

Figure 6.5 C

```
ACGCGGGGACTAGAACACCTGATTACTACAACCTTCAAGCGCGCCATCGCATCGAGGAAGAACACACAGACACAGCAAT    79
GGATTGTGCTTCAACTGTCGTCACTTCACATACCAACAGCAAAAACGACTGCAATAAAAAAAAATGTGTTGCCCCATGT   158
CCTACTGAATCAGGCAAAAACATATCAGGTCAAAACATATCAGGCCAAAACATATCAGGCCAAAACATATCAGGCAGTG   237
TGGGCGGAAGTATTTCCATTGGCAACCAGGACAATACCATCACTCATACCGCTTAACAAACGACTCCCAGAAAAACTGT   316
CGCACCCAGCATCACCATAGCCAACATAACAGGTGGCAAAGTGGGTCTCTACTACAAGGCAGAGAAG               383

ATG GCT CTC TGT GAC CTC ATC GAA GGA AGT CAA CTC AGG AAG GGC CAC TCC ATT CTC ATC    443
 M   A   L   C   D   L   I   E   G   S   Q   L   R   K   G   H   S   I   L   I    20

CTG GAA AAC ACC TTT AAG AAG ACC CCT CTT CCT CCT CTA TGT CAC TGG CTT CCA AGT AGA    503
 L   E   N   T   F   K   K   T   P   L   P   P   L   C   H   W   L   P   S   R    40

AAG TGA                                                                           509
 K   *                                                                            41

CAATACATACAGATTAAACTACATACAAATTATTGATTAAAAATAAAGAGAGAATACATCTCAGATAAGTGAAAAAGTG   588
CTACAAGATATACTTTGATTGTATGGTCGAATTCAACAATGTAATCAATGTAGGTTTTTGACAAAAA AATAAA TGGTAC   667
CAAACTG(A)n
```

Figure 6.5. Full length cDNA sequence of the unknown transcript. The predicted open reading frames representing the 111aa (A), 71aa (B) and 41aa (C) sequences are indicated. Predicted phosphorylation sites are represented by *boxes*. Predicted glycosylation sites are *underlined*. Six consecutive N-myristoylation sites are shown in *shaded grey* text. The polyadenylation site is indicated by *bold* text.

contained six consecutive N-myristoylation sites. These sites were isolated to a pentapeptide repeat region located at the c-terminus of the sequence. Despite these predictions, the motifs are generally too non-specific to provide any clue regarding the function of these hypothetical proteins.

6.4 Organ expression distribution

6.4.1 Materials and Methods

Semi-quantitative reverse transcription PCR (RT-PCR) was conducted to examine the expression of the transcript in different organs. Organ samples were dissected from the gill, brain, skin, anterior kidney, muscle, liver, heart, intestine and spleen of a single naïve Atlantic salmon smolt. Samples were stored in RNA*later* (Ambion, Austin, USA) at -80°C following dissection. For each sample approximately 100 mg of tissue was homogenised in TRI-reagent (Ambion) using a Quicklyse tissue homogeniser (Qiagen, Doncaster, Australia). Total RNA was then isolated using the RiboPure™ kit according to the manufacturer's protocol (Ambion). Following isolation, RNA was treated with rDNase I from the DNA-*free* ™ kit (Ambion). The quantity and quality of total RNA was estimated by examining the 260:280 nm absorbance ratio (Nanodrop Technologies, Delaware, USA).

A total of 500 ng of RNA was reverse transcribed using superscript III reverse transcriptase in a 20 µl reaction as per the manufacturer's protocol (Invitrogen). The

resulting cDNA was diluted to 10 ng (starting RNA) μl^{-1} with nuclease free water. The unknown transcript was amplified using the primers ORF111-F and ORF111-R (Table 6.1). The housekeeping gene β-actin was also amplified using the primers BACT-F and BACT-R (Table 6.1). PCR parameters were identical for both β-actin and the unknown transcript. Reactions were performed in 50 µl containing 5 µl of 10X buffer, 2.5 mM $MgCl_2$, 500 µM of dNTPs, 400 nM of forward and reverse primer, 2 units of AmpliTaq-Gold (Applied Biosystems) and 50 ng of cDNA template. Amplification was conducted in an Eppendorf thermal cycler (Eppendorf) using the following cycling profile: 10 min at 95°C, 30 cycles of 1 min at 94°C, 1 min at 56°C and 2 min at 72°C. Amplicons were electrophoresed on 2% TAE (Tris-acetate-EDTA) agarose gel containing 1X SYBR Safe™ DNA gel stain (Invitrogen) and visualised on a Safe Imager™ Blue-Light transilluminator (Invitrogen).

6.4.2 Results and Discussion

Semi-quantitative RT-PCR was conducted to compare the expression of the unknown transcript in different organs of naïve Atlantic salmon smolt. This analysis revealed that the expression was evident although variable between all organs. The highest expression was observed in the anterior kidney (Figure 6.6), however moderate expression was also observed in the spleen, gill and heart. The lowest expression was observed in the muscle and skin.

Since this transcript was originally derived from an anterior kidney cDNA library (Adzhubei et al. 2007), it was not surprising it should demonstrate high expression in

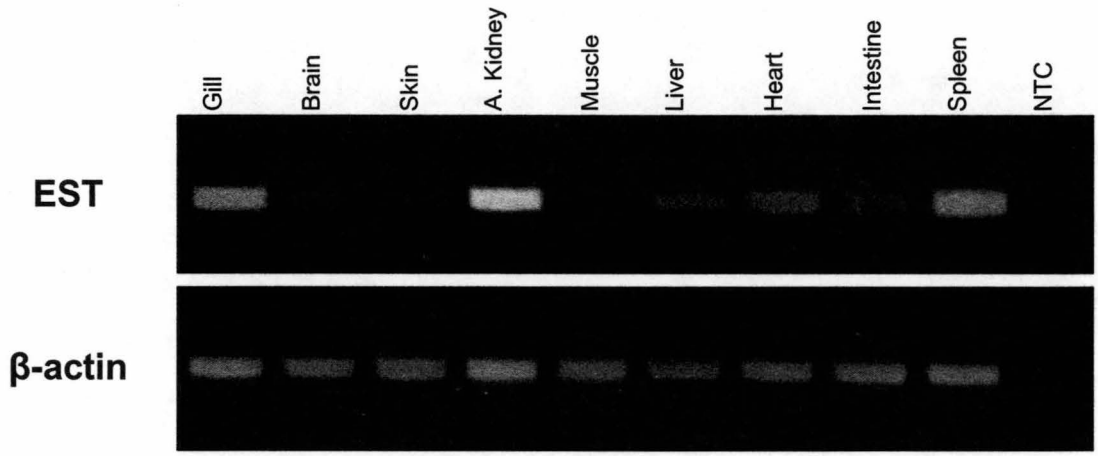


Figure 6.6. Expression of the unknown transcript (EST) assessed by semi-quantitative RT-PCR in various tissues derived from a single naïve Atlantic salmon smolt. The expression of β-actin is shown to confirm equal loading. No template controls (NTC) are also presented.

this organ. It is however interesting that while expression is highest in the anterior kidney in naïve animals, no differential expression is observed in this organ when individuals are infected by AGD (Chapter 4). Differential expression is restricted to the site of infection, the gill (Chapter 4). The expression profile of the 74 EST sequences comprising the UniGene cluster Ssa.25836 demonstrated that expression is highest in the thymus, thyroid and anterior kidney. This result is in agreement with my experimental results described above.

6.5 Cellular expression distribution

6.5.1 Materials and Methods

In situ hybridisation was used to examine the expression of the unknown EST in the gills of naïve and AGD affected Atlantic salmon. RNA probes were generated using *in vitro* transcription with digoxigenin (DIG) labelled UTP (Roche Applied Science, Penzberg, Germany). Briefly, a 216bp fragment was amplified using the primers ORF71-F and ORF71-R (Table 6.1) and parameters described above. PCR products were cloned into pGEM-T easy (Promega) and plasmid was isolated from a single colony – grown overnight – using the Quick Lyse Miniprep kit (Qiagen). Orientation was confirmed by sequencing with T7 and Sp6 primers and the Applied Biosystems BigDye Terminator Kit (Applied Biosystems). Plasmid was quantified (Nanodrop Technologies) and 3.15 µg were linearised with either 50 units of *NcoI* or *SpeI* (New England Biolabs, Ipswich, USA) at 37°C for 17 hours. Linearised plasmid was purified

using phenol:chloroform extraction and ethanol precipitation. "Run off" transcripts were produced using the DIG RNA labelling *in vitro* transcription kit (Roche Applied Science) exactly as per the manufacturer's protocol. Anti-sense probe was generated using the *NcoI* digested template and the Sp6 RNA polymerase. Sense probe (negative control) was generated using the *SpeI* digested template and T7 RNA polymerase. RNA probes were quantified using the Qubit fluorometer (Invitrogen) and diluted to 5 ng μl^{-1} in nuclease free water.

Two AGD affected and two naïve Atlantic salmon smolts were obtained from laboratory tank systems at the University of Tasmania. The second left anterior hemibranch was dissected and fixed for 15 hours at 4°C in phosphate buffered saline (PBS; pH 7.4) containing 4% paraformaldehyde. Tissues were dehydrated with graded ethanol, cleared in xylene and embedded in paraffin wax. Seven micrometer sections were cut and mounted on Poly-Prep™ slides (Sigma-Aldrich, St Louis, USA). Sections were de-waxed in xylene, rehydrated in graded ethanol (100%-70%) and washed twice in diethylpyrocarbonate (DEPC)-treated water for 5 min. Slides were incubated twice for 10 min in DEPC-treated PBS (pH 7.4) and then twice for 10 min in DEPC-treated PBS (pH 7.4) containing 100 mM glycine. Sections were incubated for 15 min in DEPC-treated PBS (pH 7.4) containing 0.1% Triton X-100 (Sigma-Aldrich). Sections were washed twice for 5 min each in DEPC-treated PBS (pH 7.4) and then permeabilised with 5 $\mu g\ ml^{-1}$ Proteinase K (New England Biolabs) in DEPC-treated TE (10 mM Tris, 1 mM EDTA, pH 8.0) for 30 min at 37°C. Sections were post-fixed for 5 min at 4°C with DEPC-treated PBS (pH 7.4) containing 4% paraformaldehyde. Slides were again washed twice (5 min each) with DEPC-treated PBS (pH 7.4) and then

acetylated twice with 0.1 M triethanolamine (Sigma-Aldrich) containing 0.25% acetate anhydride (Sigma-Aldrich) for 5 min each.

Next, sections were pre-hybridised with a solution of 50% HiDi (deionised formamide; Applied Biosystems) and 4X SSC at 37°C for approximately 30 min. Pre-hybridisation solution was removed and sections were overlayed with 60 µl of denatured (80°C for 10 min) hybridisation solution containing 40% HiDi (Applied Biosystems), 4X SSC, 1X Denhardt's solution (Sigma-Aldrich), 10 mg ml^{-1} yeast tRNA (Sigma-Aldrich), 10 mM DTT (Invitrogen), 10% dextran sulfate (Sigma-Aldrich), 10 mg ml^{-1} fish sperm DNA (Roche Applied Science) and 25 ng of sense (control) or anti-sense probe. Hybridisations were performed at 42°C for 15 hours. Coverslips were removed and sections were washed twice in 2X SSC for 15 min at room temperature, twice in 1X SSC for 15 min at 42°C and finally twice in 0.5X SSC for 15 min at 52°C.

Unbound RNA probe was digested with 20 µg ml^{-1} RNase A (New England Biolabs) in NTE buffer (500 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 8.0) for 30 min at 37°C. Following digestion, sections were washed twice in 0.1X SSC for 30 min each. Sections were washed twice in buffer 1 (100 mM Tris, 150 mM NaCl, pH 7.5) for 10 min each and then blocked in buffer 1 containing 0.1% Triton X-100 (Sigma-Aldrich) and 2% sheep serum (Sigma-Aldrich) for 50 min. Blocking solution was removed and sections were incubated for 2 hours in buffer 1 containing 0.1% Triton X-100 (Sigma-Aldrich), 1% sheep serum (Sigma-Aldrich) and a 1:1000 dilution of anti-DIG alkaline phosphatase antibody (Roche Applied Science). Sections were washed twice for 10 min in buffer 1 and then incubated in buffer 2 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5) for 10 min. Sections were incubated in BCIP/NBT solution (Sigma-Aldrich) for

15 hours. Colour development was stopped by rinsing slides in TE (pH 8.0). Sections were counter-stained for 5 min with nuclear fast red (Sigma-Aldrich) and washed twice in water. Sections were dehydrated through graded ethanol (70-100%), cleared in xylene and mounted.

6.5.2 Results and Discussion

In situ hybridisation of the gill of naïve Atlantic salmon suggested expression is restricted to the epithelial pavement cells within the gill (Figure 6.7A). In contrast, the gill of AGD affected Atlantic salmon displayed highest expression within the proliferating cells of the lesion (Figure 6.7B, 6.7D). Hybridisation with sense control probe revealed little or no staining (Figure 6.7C).

Staining of the epithelial pavement cells in the naïve animals resembled the results obtained from previous studies which examined the location of chloride cells in naïve Atlantic salmon gills (Adams and Nowak 2003). However, unlike the latter study, the number of cells expressing this unknown transcript increased considerably in the AGD lesion. In contrast, large hyperplastic lesions showed a relatively low number of chloride positive cells (Adams and Nowak 2003). Previous research has shown that almost all cells in a hyperplastic AGD lesion express proliferating cell nuclear antigen (PCNA) (Morrison et al. 2006a). Likewise, my study found all proliferating cells with the AGD lesion were expressing the unknown transcript. The number of cells expressing immune genes such as MH class II (Morrison et al. 2006b) and immunoglobulin (Gross 2007) are variable in AGD lesions and are not expressed in all

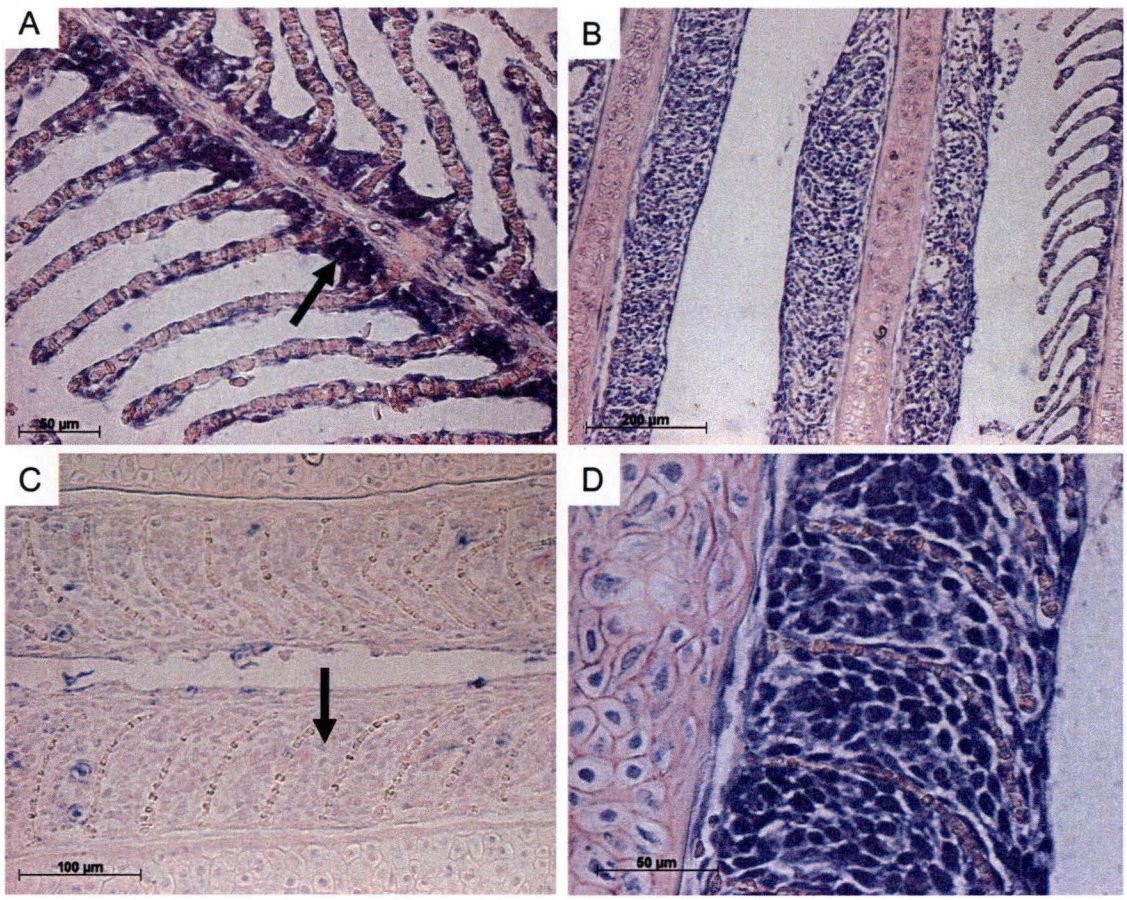


Figure 6.7. *In situ* hybridisation of naïve and AGD affected gill tissue. In naïve Atlantic salmon staining appears to be isolated to the epithelial pavement cells of primary lamellae (A), indicated by *arrow*. In the AGD affected gill tissue extensive staining was observed throughout the lesions (B). Little or no staining was observed when the sense control probe was hybridised to AGD affected gill tissue (C) indicated by *arrow*. Almost all proliferating cells within the AGD lesion appeared to express this transcript (D).

proliferating cells. Although somewhat speculative, the fact that the cellular expression of the unknown transcript more closely resembling that of PCNA may suggest this transcript is involved in cell proliferation pathways.

Given the *in situ* hybridisation, results it also seems plausible that the significant up-regulation of this transcript may be an artefact caused by the cell proliferation within the AGD lesion. For instance, if this transcript is constitutively expressed in naïve epithelial pavement cells which then undergo proliferation upon AGD infection, the expression would appear to increase simply because there are more cells expressing this transcript in the infected compared to non-infected gill material. Under this hypothesis the transcript may represent a housekeeping gene which is specific to epithelial pavement cells.

6.6 Recombinant protein expression

6.6.1 Materials and Methods

Recombinant proteins were produced for the three predicted open reading frames (open reading frame 41, 71 and 111). From a single Atlantic salmon the three open reading frames were amplified using the primers ORF111-F and ORF111-R for the 111aa open reading frame, ORF71-F and ORF71-R for the 71aa open reading frame, and ORF41-F and ORF-41-R for the 41aa open reading frame (Table 6.1). PCR parameters are described within section *6.4 Organ expression distribution.* Amplicons were purified

with the QIA Quick PCR Purification kit (Qiagen) and cloned into pENTR-D-TOPO (Invitrogen) as per the manufacturer's instructions. Plasmid was isolated from a single colony – grown overnight – using the Quick Lyse Miniprep kit (Qiagen). Constructs were sequenced with M13F and M13R primers and the Applied Biosystems BigDye Terminator Kit (Applied Biosystems) to confirm orientation. An LR Clonase™ recombination reaction was then performed between the pENTR-D-TOPO construct and the expression vector pDEST-17 (Invitrogen) as per the manufacturer's instructions. Expression plasmids were isolated from successful recombinants using the Quick Lyse Miniprep kit (Qiagen) and sequenced with T7 and gene specific reverse (ORF111-R, ORF71-R, ORF41-R) primers.

Protein expression was optimised by the Protein Production Unit at Monash University. This optimisation consisted of testing a range of different *E. coli* strains for protein expression. All constructs were transformed into the appropriate cell line and transformants were selected on Luria-Bertani (LB) agar containing 100 $\mu g\ ml^{-1}$ ampicillin (Sigma-Aldrich). Single colonies were picked and grown in autoinduction media (Novagen, Darmstadt) with ampicillin for 24 hours at 20°C. Cultures were then lysed with Popculture (Novagen) and lysonase (Novagen). The lysed cells were applied to a vacuum filter membrane which separates soluble (flow-through) with insoluble (bound to membrane) proteins. The membrane bound insoluble fraction was then solubilised through the addition of 4% SDS and eluted under vacuum.

Proteins were separated on a 4-15% BIS-Tris gel (Invitrogen) and transferred to a nitrocellulose membrane (Bio-Rad, Hercules, USA) overnight. The membrane was washed twice for 5 min each with TBS (Tris buffered saline) containing 0.1% tween

(Sigma-Aldrich). Western blot analysis was then performed using the Westernbreeze chromogenic immunodetection kit and 1:1000 dilution of monoclonal anti-penta HIS antibody (Qiagen) exactly as per the manufacture's protocol (Invitrogen).

6.6.2 Results and Discussion

The expression of recombinant proteins representing the three predicted open reading frames was conducted using *E. coli* (Figure 6.8). Successful expression was achieved for the 111aa and 71aa proteins. No expression of the 41aa could be detected with any cell line tested. Interestingly, western blot analysis of the cells transformed with the 41aa construct did suggest a HIS tag protein of approximately 90 kDa in size was expressed. This however was significantly larger than the predicted 7 kDa size. Further optimisation will be required to express the 41aa recombinant protein. Both the 111aa and 71aa recombinant proteins were successfully detected using western blot analysis with an anti-HIS antibody.

The 71aa recombinant proteins were expressed highest in the Rosetta-gami B (DE3) cell line while the 111aa recombinant protein was most highly expressed in the BL21 (DE3) RP cell line. The 111aa protein was only expressed in the insoluble cellular fraction. This was possibly caused by a failure of the protein to reach native conformation which can lead to the formation of insoluble inclusion bodies and/or degradation via multiple proteases. In contrast, the 71aa protein showed high expression in the soluble fraction.

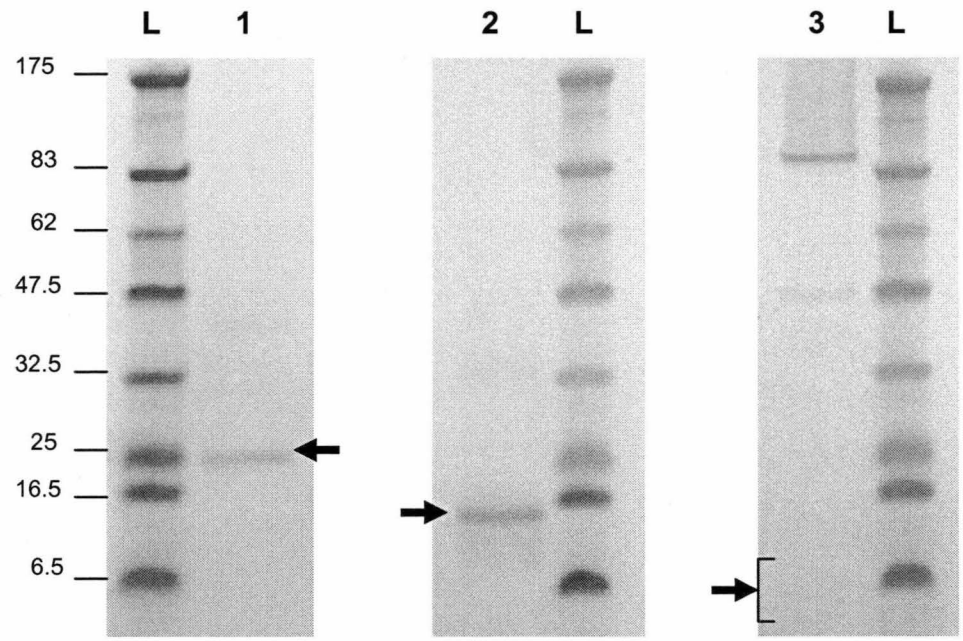


Figure 6.8. Western blot of recombinant HIS fusion tagged proteins. Lanes L: protein standard (protein sizes, kDa, are indicated to the right). Lane 1: 111aa recombinant protein indicated by *arrow*. Lane 2: 71aa recombinant protein indicated by *arrow*. Lane 3: 41aa recombinant protein not detected by western blot analysis *arrow* indicates predicted protein size.

This result suggests that the 71aa open reading frame can indeed encode a peptide capable of correct native confirmation. This provides some evidence to suggest the 71aa open reading frame – rather than the 111aa or 41aa open reading frames – encodes the functional protein from this unknown gene. It must however be conceded that post-translational modifications in *E. coli* are different to those of a eukaryote such as Atlantic salmon and further experimentation is required.

6.7 Future directions

Although this study has performed the first steps in characterising this transcript, considerable research is still required. The most pressing aspect is to determine which open reading frame, if any, encodes the functional protein. To achieve this large scale production of the two recombinant proteins and further optimisation of the 41aa protein is currently being undertaken. I plan to use these recombinant proteins to produce three polyclonal antibodies in rabbits, which will be used to perform western blot analysis on Atlantic salmon crude lysate. This will allow me to determine which open reading frame encodes the functional protein. If the correct protein sequence can be determined, I will then use recombinant proteins to stimulate the Atlantic salmon SHK-1 cell line and a gill derived cell line, and profile the response using microarrays. This approach has been used successfully to study the role of type I and II interferons (Martin et al. 2007a) and the proinflammatory cytokine IL-1β (Martin et al. 2007b) and may provide some insight into the function of this protein. The use of RNAi to silence this transcript is also planned as future research. Considering this transcript is so highly induced during AGD, I propose conducting an experiment in which I silence this transcript *in*

vivo in Atlantic salmon and then challenging the animals (with appropriate controls) with AGD to assess the survival response. Gene expression profiling may also be used to assess the transcriptome response both before and after challenge.

To conclude, this study has taken the first steps to further characterise an unknown transcript which displays significant up-regulation in the gill of AGD affected Atlantic salmon. Despite the investigation undertaken within this study considerable research is still required before a putative function for this transcript could be assigned.

Chapter 7: General discussion

Preamble

AGD is the most significant health issue affecting the culture of Atlantic salmon in Tasmania and is of increasing concern in other salmon farming areas (Steinum et al. 2008). Progress towards reducing the impact of AGD is ongoing. The primary research foci in Tasmania are the development of a vaccine and selective breeding for AGD resistance. Understanding the molecular mechanisms of AGD resistance would significantly enhance both programs. For instance, the development and integration of marker assisted selection (MAS) into the Tasmanian Atlantic salmon selective breeding program has the potential to significantly increase the genetic gains obtained from this program. Indeed, simulation of two-generation within-family aquaculture MAS schemes have indicated that increased genetic gains can be achieved through MAS selective breeding compared to conventional breeding (Sonesson 2007b). The enhanced genetic gain in the MAS scheme was caused by an increasing frequency of the positive quantitative trait loci (QTL) allele within the population (Sonesson 2007b).

MAS is particularly suitable as a selection strategy for AGD resistance within the Tasmanian Atlantic salmon selective breeding program. This is because current biosecurity protocols within Tasmania restrict the transfer of potential broodstock from the marine grow-out sites back to the freshwater hatcheries. Thus the majority of broodstock are never actually exposed to the marine environment or to the causative agent of AGD. In this situation selection must be made on sibling information, rather than actual broodstock performance. Considering the within-family variation of many

traits, selection based on genotype, rather than sibling phenotype, would significantly increase the accuracy of selection.

In this discussion I will consider the major findings of each research chapter and their implications for the development of genetic markers linked to AGD resistance. I will conclude by proposing future research directions arising from this research project.

7.1 High MH class II variation contrasts diversity in non-coding loci

The previous reports of reduced genetic variation in the Tasmanian Atlantic salmon population (Elliott and Reilly 2003; Innes and Elliott 2006) raised some concern regarding the amount of diversity that may be present in functional loci such as the immune related major histocompatibility (MH) genes. I therefore considered it would be prudent to examine the amount of MH variation in this population before investigating the relationship between MH polymorphism and AGD resistance. The results reported in Chapter 2 demonstrated that a high level of MH class II variation has been retained in the Australian Atlantic salmon population, despite previous reports of reduced non-coding genetic variation (Elliott and Reilly 2003; Innes and Elliott 2006). These results suggested adequate variation was present at the MH class II genes to warrant the investigation of the association between MH polymorphism and AGD resistance.

7.2 Significance of a weak association between MH class II *Sasa-DAA* polymorphism and AGD resistance.

Chapter 3 describes a weak, yet statistically significant, association between the presence/absence of the MH class II *Sasa-DAA-3UTR* allele 239 and increased resistance to AGD. A significant association was also detected between the *Sasa-DAA-3UTR* genotypes 239-259 and 259-259 and increased resistance to AGD. The need for caution in the interpretation of these associations has been discussed previously, but to reiterate, these associations should be considered as suggestive rather than conclusive until validated by further research. Further research should concentrate on verifying the association in a larger sample size across a range of genetic backgrounds. As a consequence of this uncertainty, these associations cannot be considered for use within a MAS program in their present form. The main limitations surrounding this research are described below.

7.2.1 Limitations of the *Sasa-DAA-3UTR* and *Sasa-UBA-3UTR* loci

A number of strategies have been utilised to assess variation within the MH, including single strand confirmation polymorphism (Palti et al. 2001), denaturing gradient gel electrophoresis (Miller et al. 2004) and restriction fragment length polymorphism (Langefors et al. 2000). However, the identification and characterisation of the micro and minisatellite markers in the 3' UT (untranslated) regions of *Sasa-UBA* (Grimholt et

al. 2002) and *Sasa-DAA* (Grimholt et al. 2000; Stet et al. 2002) genes, offers the most efficient method to assess variation at the MH, with the possible exception of direct DNA sequencing. These markers have been used to successfully examine MH variation and its relationship to disease resistance in Atlantic salmon (Grimholt et al. 2003; Kjøglum et al. 2006; Glover et al. 2007; Kjøglum et al. 2008). The use of these markers within my research was, however, complicated by the relationship between variation in the 3' UT region markers and sequence variation through the peptide binding region (PBR). In contrast to that reported for Norwegian cultured Atlantic salmon (Stet et al. 2002), not every *Sasa-DAA* allele sequence is represented by a unique *Sasa-DAA-3UTR* allele within the Tasmanian Atlantic salmon population. This situation has also been reported at the *Sasa-UBA* locus (Grimholt et al. 2002). Furthermore, the haplotypic variation between *Sasa-DAA* and *Sasa-DAB* was also unexpected and further complicates the relationship to *Sasa-DAA-3UTR*.

Despite the fact the *Sasa-DAA-3UTR* 239 allele was linked to the sequence *Sasa-DAA*1001* in all individuals analysed, the relationship between variation at the 3' UT regions markers and PBR sequence variation makes the associations reported in Chapter 3 more uncertain. The fact that at least three co-segregating *Sasa-DAB* alleles were linked to *Sasa-DAA*1001* is one such example. In this situation it would be advantageous to compare the effects of different composite haplotypes such as *Sasa-DAA*1001/Sasa-DAB*0402* and *Sasa-DAA*1001/Sasa-DAB*1502*. Such an experiment would allow the effects of the *Sasa-DAA* and *Sasa-DAB* to be separated and may help identify which region is actually associated with the resistance.

Through the Tasmanian Atlantic salmon selective breeding program, families of Atlantic salmon with known levels of AGD resistance – quantified as both disease severity and survival time – have been developed. These families may be utilised to examine the association between AGD and MH polymorphism on a significantly larger scale compared to the challenge test described in Chapter 3. Given the uncertainty between variation in 3' UT region markers and sequence variation in the PBR, a combined strategy of direct DNA sequencing, 3' UT region marker genotyping and allele specific PCR would be required to resolve all alleles in both broodstock and offspring. This strategy has proven successful for other studies (Grimholt et al. 2002; Stet et al. 2002; Grimholt et al. 2003) and with the increased power of larger sample sizes could significantly increase the accuracy of this investigation.

7.2.2 Effects of infection dynamics on the MH

The difference in dynamics between the challenge trial and natural infection may influence the results of my investigation of the effects of the MH on AGD resistance. It is possible that the results could be different in a natural infection compared to a challenge system. When a pathogen infects a host an interaction will develop between the pathogen's ability to cause disease and the host's ability to respond and eliminate that threat. The MH plays a pivotal role in this balance and by virtue of its allelic variation will cause differences in disease resistance between individuals. In the situation when the host is overwhelmed by the pathogen (i.e. due to an inundation of pathogens) the host may be unable to respond successfully. Regardless of whether the host carries an allele conferring resistance the individual may simply succumb before it

has the necessary time to stimulate that superior immune response. This is one concern when using a challenge model, like that described in Chapters 3 and 4, compared to a natural infection (i.e. Chapter 5). Although intended to mimic the natural infection the challenge test is considerably more acute and as a consequence fish may not be able to respond to the best of their ability. With this in mind, it may be more appropriate to examine the relationship between MH variation and AGD resistance in a more natural infection. This will be considered in greater depth later in this discussion.

7.3 AGD resistance as a polygenic trait

The broad sense heritability estimate for AGD resistance measured through histopathology upon first infection in a laboratory challenge was 0.30 ± 0.09 (Taylor et al. 2007) and the difference between the most resistant and most susceptible families was 19% (Chapter 3). Considering this, a locus, such as the MH, that accounts for 4-5% of total variation in resistance (Chapter 3) may actually account for a moderate proportion of the genetic variance. Nevertheless, the relatively small contribution of the MH on AGD resistance suggests resistance may be under polygenic control and non-MH genetic effects could also contribute. For example, despite strong associations between variation at *Sasa-DAA/Sasa-UBA* and resistance to infectious salmon anaemia (ISA) in Atlantic salmon (Grimholt et al. 2003; Kjøglum et al. 2006), non-MH effects are known to contribute significantly (Kjøglum et al. 2005). A similar result was also observed for the associations between MH polymorphism and resistance to furunculosis (Grimholt et al. 2003; Kjøglum et al. 2005; Kjøglum et al. 2008). The large number of genes differentially expressed between Atlantic salmon putatively resistant and

susceptible to AGD (Chapter 5) – often involved in different pathways – further suggests resistance is influenced by multiple loci.

The hypothesis that AGD resistance is under polygenic control is probably not unreasonable given that resistance to other infectious diseases is often a polygenic trait (Vallejo et al. 1998). For instance, in many disease resistance QTL mapping studies, a number of QTLs are identified, each typically accounting for only 2-10% of variation in disease resistance (Yonash et al. 1999). A recent QTL mapping study in Atlantic salmon identified 10 genomic regions associated with resistance to the monogenean ectoparasite *Gyrodactylus salaries*, accounting for a total 27.3% of variation (Gilbey et al. 2006). Multiple loci were also found to be associated with resistance to the myxosporean parasite *Ceratomyxa shasta* in rainbow trout (Nichols et al. 2003). Although resistance to each parasite should be tested empirically, these studies do in fact suggest that parasite resistance in salmonids is most likely a polygenic trait.

7.4 Identifying markers associated with polygenic traits

Identifying loci linked to polygenic traits is not a simple task. The contribution of individual loci affecting a polygenic trait diminishes as the number of influencing regions increase. They also become increasingly more difficult to detect because their control is obfuscated. This situation highlights the limitations of employing a candidate gene approach – like that described in Chapter 3 – to examine a polygenic trait. The candidate gene approach can also introduce bias into the study. This is because

candidate genes with only a presumed relevant function are examined, and therefore many unknown genes which may contribute are simply not investigated.

With the development of genetic markers and maps for Atlantic salmon (Gilbey et al. 2004; Moen et al. 2004b; Hayes et al. 2007) it has become possible to conduct genome-wide QTL mapping studies to search for loci that affect disease resistance. Genome-wide scans using hundreds or thousands of markers are particularly appropriate for polygenic traits. One of the best examples is the identification and validation of a QTL associated with ISA resistance in Atlantic salmon (Moen et al. 2004a; Moen et al. 2007). QTL mapping studies have also been conducted to identify markers associated with resistance to infectious pancreatic necrosis virus in rainbow trout (Ozaki et al. 2001) and resistance to infectious hematopoietic necrosis in rainbow/steelhead trout backcrosses (Rodriguez et al. 2004).

Although the individual QTL contribution may be small in a polygenic trait, aquaculture species such as Atlantic salmon have a number of attributes that make them particularly amenable for the identification of small effect QTLs (Sonesson 2007a). The large family size and possibility of repeat crosses (possibly through cryo-preservation of gametes) means the statistical power and flexibility of within-family analysis can be increased. The use of double haploids, produced through chromosomal set manipulations, is considered to be particularly appropriate (Martinez et al. 2002) for QTL mapping in fish when the effect of the QTL is low. This is because the QTL will be in a homozygous form and its contribution to the trait will be greater than in a full- or half-sibling design (Sonesson 2007a). The use of double haploid Atlantic salmon to

investigate loci associated with AGD resistance may prove useful under the hypothesis that resistance is controlled by multiple small effect QTLs.

The recent identification of 2507 Atlantic salmon expressed sequence tag (EST) single nucleotide polymorphisms (SNPs) through a Canadian and Norwegian initiative (Hayes et al. 2007) will also prove invaluable for the identification of loci influencing polygenic traits and the ultimate development of marker assisted selection. Marker assisted selection based on hundreds of SNPs – also known as genome-wide selection – is particularly useful for polygenic traits, in which numerous loci influence the trait. Examining the association between these SNPs and AGD resistance must be considered as a future research direction for identifying loci associated with AGD resistance and the ultimate development of a marker assisted selection program for this trait.

7.5 The localised host response

The gill is the primary and only site of infestation by *Neoparamoeba perurans*. Localised cellular alterations, including epithelial cell proliferation, are only observed at the site of amoeba attachment (Adams and Nowak 2004a). This localised cellular response was reflected in the transcriptome response compared between the gill, liver and anterior kidney. The majority of differential expression was isolated to the site of infection: the gill. This result provides further evidence that the host response to AGD is extremely localised, at least upon initial infection. This has also been reported when the expression of immune-related genes have been examined in the liver and anterior kidney of Atlantic salmon (Bridle et al. 2006a) and rainbow trout (Bridle et al. 2006b).

The localised nature of the host response is further corroborated by reports of lesion restricted expression of certain genes (Morrison et al. 2006a; Morrison et al. 2007; Young et al. 2008a). The gross presentation of AGD lesions progress from discrete focal spots during the early stages of disease to extensive regions of coalescing mucoid patches (Adams et al. 2004). By comparing the transcriptome response between areas of AGD lesion versus non-lesion gill material, Young et al. (2008a) demonstrated that a large proportion of genes are only differentially expressed within the actual lesion. AGD lesion restricted expression has also been reported for the dysregulation of interleukin-1β (Morrison et al. 2007).

Although the evidence presented thus far suggests the host response to AGD remains localised to the gill, the presence of anti-*Neoparamoeba* spp. antibodies in the serum of AGD affected Atlantic salmon has been demonstrated in several studies (Findlay et al. 1995; Findlay and Munday 1998; Gross et al. 2004a; Vincent et al. 2006). While these studies suggest a systemic response to AGD is possible, the response appears to be heavily influenced by infection history (Vincent et al. 2006). For instance, following the first four weeks of infection in a challenge test, Atlantic salmon display no anti-*Neoparamoeba* spp. antibody response (Vincent et al. 2006). However, in contrast, antibodies can be detected within individuals which have undergone multiple infection events (Vincent et al. 2006). This result suggests the host response may be restricted to the gill upon the initial infection, however following treatment and subsequent re-infection a more systemic response may develop. Currently all gene expression analysis studies have concentrated on the host response to AGD upon first infection (Bridle et al. 2006a; Bridle et al. 2006b; Morrison et al. 2006a; Morrison et al. 2007; Young et al. 2008a), with the exception of the study described in Chapter 5.

The generation of anti-*Neoparamoeba* spp. antibodies through multiple re-infection events suggests the host response upon first infection may be considerably different to the response at later infection events. Although the research presented in Chapter 5 did not aim to examine the host response to AGD per se, the study did identify a few notable results concerning the differences in immune response between the initial and subsequent re-infection events. For instance, when compared to naïve Atlantic salmon, individuals infected with AGD for the first time demonstrated a down-regulation of immune genes, including MH class II invariant chain-like protein and immunoglobulin (Young et al. 2008a). In contrast, individuals which have undergone multiple re-infection events have an up-regulation of these genes compared to naïve Atlantic salmon (Chapter 5). It is expected the increased expression of the adaptive immune genes following multiple re-infection events is at least partially responsible for the generation of the systemic anti-*Neoparamoeba* spp. antibody response.

7.6 The difference in host response between initial and subsequent re-infection

Along with a localised host response, individuals infected with AGD for the first time also demonstrated general gene suppression (Chapter 4). Suppression of the Atlantic salmon immune system by *Neoparamoeba* spp. has also been documented both at the cellular level (Gross et al. 2004a; Gross et al. 2005) and at the transcriptome level (Young et al. 2008a). Immunosuppression by *Neoparamoeba* spp. upon initial infection may allow parasites to escape immune surveillance and therefore permit the development of a localised infection.

The general gene suppression observed in Chapter 4 as well as other studies (Young et al. 2008a) may be influenced by the changes in cellular architecture associated with AGD lesion formation. Indeed, despite a moderate infiltration of inflammatory cells, (Adams and Nowak 2004a) the significant proliferation of epithelial cells would cause the proportion of inflammatory cells to be lower in comparison to epithelial cells within AGD affected tissue. When this tissue is compared to naïve or non-lesion tissue in microarray or real-time PCR analysis, certain genes may appear down-regulated simply by the change in cell types/cell ratios. This situation is probably exacerbated when lesion restricted expression is examined (Young et al. 2008a) or when individuals with very severe infection are compared to naïve or less severely infected individuals (i.e. Chapter 4).

Surprisingly, comparing the transcriptome response between Atlantic salmon putatively resistant and susceptible to AGD following the natural infection (Chapter 5) did not reflect the localised host gene suppression reported in Chapter 4 and recently published by Young et al. (2008a). If *Neoparamoeba* spp. was capable of suppressing the immune system – or the gene suppression reported previously was simply an artefact of the changes in cellular architecture – I would have expected to see localised gene suppression in the AGD susceptible Atlantic salmon when compared to the resistant animals. This was not observed. In fact, despite displaying severe AGD pathology and in contrast to that reported previously (Young et al. 2008a), the expression of the immune genes immunoglobulin and MH class II invariant chain-like protein were actually up-regulated in the AGD susceptible Atlantic salmon compared to naïve animals. This result highlights the difference in the host response between the initial and

subsequent re-infections. Considering the evidence presented above the following model can be proposed. Upon initial infection the host demonstrates general localised gene suppression (including genes of the immune system). Following treatment and subsequent re-infection, however, the host develops a more systemic response which includes the increased expression of immune genes and the production of anti-*Neoparamoeba* spp. antibodies.

7.7 Effects of chronic and acute AGD infections

The evidence discussed above suggests the host response upon initial infection and subsequent re-infection will vary. Another complicating factor relates to the difference in infection dynamics, principally caused by the difference in amoeba concentration between a commercial salmon farm and the laboratory based challenge systems. Typically most challenge tests – including that described in Chapters 3 and 4 – use a dose of between 450-500 cells L^{-1} (Morrison et al. 2006a; Young et al. 2008a). This is significantly higher than the 10-50 cells L^{-1} reported in a natural infection in a commercial fish cage (Douglas-Helders et al. 2003). AGD within a commercial salmon farm therefore resembles a more chronic infection than the acute infection represented by the challenge test.

Considering the differences in infection dynamics between chronic and acute AGD exposure, it may be unreasonable to assume that the host response would be identical in the two situations. The general gene suppression observed within the gill following the acute challenge described in Chapter 4 – as well as in other studies (Young et al. 2008a)

– may be one consequence of the high dose used in the challenge system. It seems plausible that when fish are overwhelmed with an infectious agent they may have a restricted ability to mount a successful response against the pathogen. This situation has been reported for other parasites. Infection of mice, for example, with low doses of the parasitic haemoflagellate, *Leishmania major* Yakimoff et Schokhor, leads to parasite containment while higher doses result in progressive disease (Menon and Bretscher 1998). In regard to AGD it is likely that when the host is exposed to a lower concentration of amoeba (such as in a commercial farming operation) it will have a better chance of progressing from the initial suppressive response to a more positive response. In fact, the localised suppressive response may not occur at all at lower amoeba concentrations. However, this is only supposition and it has not been tested.

It has been demonstrated previously that doses as low as 10 cells L^{-1} will elicit AGD within an experimental system (Morrison et al. 2004) and this more closely reflects the amoeba concentration in a natural infection. Further experimentation to analyse the transcriptome response to AGD at lower amoeba concentrations such as 10 cells L^{-1} is required to determine immune response in a chronic infection.

7.8 Innate resistance versus acquired resistance: different traits?

As stated in Chapter 1, AGD resistance can be quantified in a variety of ways. For the purpose of this thesis I have defined resistance as reduced gill pathology, quantified through either histopathology or gross gill scoring. A complicating aspect concerning AGD resistance relates to when the resistance is actually measured. In short, resistance

to AGD can be measured immediately following first infection or later following multiple exposures (often during which time treatment events may occur). Research presented within this thesis has examined both forms of AGD resistance. The genetic correlation between resistance upon first exposure (innate resistance) and resistance following multiple exposures (acquired resistance) – as measured by the Tasmanian Atlantic salmon selective breeding program – has generally been low (P. Kube pers. comm[9]). I hypothesise innate (measured at first infection) and acquired (measured after multiple exposures) resistance to AGD are in fact different traits. This means a family that is selected for innate resistance may not necessarily demonstrate significantly higher resistance during later infections. And vice versa, families that are selected for acquired resistance may not show increased resistance upon initial infection.

It seems logical that if innate and acquired resistance are different traits they are most likely controlled by different loci. The ability of Atlantic salmon to acquire resistance to AGD through multiple infection events was associated with an increased expression of certain genes involved in the adaptive immune response. In contrast, considering variation in disease severity upon first infection (innate resistance) can be observed following only 19 days of acute challenge (Taylor et al. 2007), it seems unlikely that the adaptive immune system is entirely or even partially responsible. Similarly, when reporting the association between MH variation and resistance to furunculosis, Grimholt et al. (2003) questioned whether adaptive immunity was solely responsible, considering that most mortalities occurred within 20 days and it was unlikely that a significant antibody response could develop in this time. The same situation is probably true for innate AGD resistance measured at 19 days post inoculation. This means if marker

[9] Dr Peter Kube, CSIRO Marine and Atmospheric Research.

assisted selection is to be considered for AGD resistance, loci associated with both forms of resistance will need to be identified. Selection will then be performed upon all the informative markers.

7.9 Was the MH a good candidate?

The evidence considered so far suggests AGD resistance is complex and may be divided into innate and acquired resistance. The ability of Atlantic salmon to acquire resistance to AGD was associated with increased expression of genes involved in the adaptive immune system. This raises the question; was the MH a good candidate to be associated with innate resistance? Considering the variation in innate resistance was observed after only 19 days post inoculation (Taylor et al. 2007) it seems unlikely this resistance was caused by a superior adaptive immune response driven by the MH. However, bearing in mind that acquired resistance was associated with increased expression of genes involved in the adaptive immune system, the MH may be a more appropriate candidate to be associated with this form of resistance. Future research regarding the relationship between MH and AGD resistance within families from the Tasmanian Atlantic salmon selective breeding program should concentrate on acquired resistance.

The recent finding that the genes involved in the MH class I and class II pathways are down-regulated within the gill of AGD affected Atlantic salmon (Young et al. 2008a) raises some concern regarding how appropriate these loci may be for AGD resistance. Interestingly, however, my transcriptome profiling experiments reported no down-

regulation of MH pathways in the gill of AGD affected Atlantic salmon (despite being present on the microarray) following either natural infection or challenge trial. In fact, genes involved in the MH class II pathway including two transcripts resembling MH class II invariant chain-like proteins were up-regulated in infected putatively susceptible fish compared to naïve Atlantic salmon. Nevertheless, despite the reported down-regulation of MH pathways (Young et al. 2008a) MH class II^+ cells are still present in the gills of AGD affected Atlantic salmon (Morrison et al. 2006b) and therefore the ability of different MH class II proteins to present anti-*Neoparamoeba* spp. antigens with varying specificities may still occur. The detection of anti-*Neoparamoeba* spp. antibodies in the serum of AGD affected Atlantic salmon indirectly suggests this response can occur (Gross et al. 2004a; Vincent et al. 2006).

7.10 Future directions: progress towards marker assisted selection.

The main objective of this thesis was to identify the molecular mechanisms of AGD resistance to assist the ultimate development of a marker assisted selection program. It has become clear that AGD resistance is not a simple trait. It may take several different forms (both innate and acquired) and is most likely controlled by multiple genes. This suggests genome-wide association studies are probably more appropriate than candidate gene approaches for these traits. Future research on the molecular mechanisms of AGD resistance should now concentrate of on genome scans using SNPs to identify loci associated with resistance. This research has become possible through the development of both Atlantic salmon SNP resources (Hayes et al. 2007) and families with varying

levels of AGD resistance (through the Tasmanian Atlantic salmon selective breeding program).

Further investigation of the unknown transcript, CK880278, is another future research direction arising from this thesis. This transcript was highly induced following AGD infection in the challenge trial and natural infection. Although some progress towards further characterising this transcript is described in Chapter 6, significant research is still required. The main focus needs to be determining which open reading frame encodes the functional protein. If the protein can be identified, future research could concentrate on functional studies including the use of RNA interference to silence the transcript *in vivo* and perform AGD challenge tests and transcriptome analysis. The use of the recombinant protein to stimulate the Atlantic salmon SHK-1 cell line and perform transcriptome analysis is also planned as future research.

7.11 Conclusion

To conclude, this thesis suggests that AGD resistance is a complex polygenic trait, involving both innate and adaptive mechanisms. Careful consideration needs to be taken when examining the molecular mechanisms causing these different types of resistance. The development of families of Atlantic salmon resistant to AGD by the Tasmanian Atlantic salmon selective breeding program will greatly assist in the future research aimed to investigate the molecular mechanisms of this trait. The major research focus should now concentrate on the use of whole genome scanning and/or SNP analysis to identify molecular markers associated with AGD resistance.

Bibliography

Adams MB, Ellard K, Nowak BF (2004) Gross pathology and its relationship with histopathology of amoebic gill disease (AGD) in farmed Atlantic salmon, *Salmo salar* L. J Fish Dis 27:151-161

Adams MB, Nowak BF (2001) Distribution and structure of lesions in the gills of Atlantic salmon, *Salmo salar* L., affected with amoebic gill disease. J Fish Dis 24:535-542

Adams MB, Nowak BF (2003) Amoebic gill disease: sequential pathology in cultured Atlantic salmon, *Salmo salar* L. J Fish Dis 26:601-614

Adams MB, Nowak BF (2004a) Experimental amoebic gill disease of Atlantic salmon, *Salmo salar* L.: further evidence for the primary pathogenic role of *Neoparamoeba* sp. (Page, 1987). J Fish Dis 27:105-113

Adams MB, Nowak BF (2004b) Sequential pathology after initial freshwater bath treatment for amoebic gill disease in cultured Atlantic salmon, *Salmo salar* L. J Fish Dis 27:163-173

Adams MB, Villavedra M, Nowak BF (2008) An opportunistic detection of amoebic gill disease (AGD) in blue warehou (*Seriolella brama*) collected from an Atlantic salmon (*Salmo salar*) production cage in south eastern Tasmania. J Fish Dis *in press*

Adcock GJ, Ramírez JHB, Hauser L, Smith P, Carvalho GR (2000) Screening of DNA polymorphisms in samples of archived scales from New Zealand snapper. J Fish Biol 56:1283-1287

Adzhubei AA, Vlasova AV, Hagen-Larsen H, Ruden TA, Laerdahl JK, Høyheim B (2007) Annotated Expressed Sequence Tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource. BMC Genomics 8:209

Agrawal A, Cha-Molstad H, Samols D, Kushner I (2001) Transactivation of c-reactive protein by IL-6 requires synergistic interaction of CCAAT/enhancer finding protein beta (C/EBP beta) and Rel p50. J Immunol 166:2378-2384

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. P Natl Acad Sci USA 101:3490-3494

Akhlaghi M, Munday BL, Rough K, Whittington RJ (1996) Immunological aspects of amoebic gill disease in salmonids. Dis Aquat Organ 25:23-31

Arkush KD, Giese AR, Mendonca HL, McBride AM, Marty GD, Hedrick PW (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

Arnér ESJ, Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. Eur J Biochem 267:6102-6109

Bacon LD (1987) Influence of the major histocompatability complex on disease resistance and productivity. Poultry Sci 66:802-811

Beaumont AR, Hoare K (2003) *Biotechnology and genetics in fisheries and aquaculture.* Blackwell Science Ltd, Osney Mead, Oxford

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57:289-300

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

Bertini R, Howard OMZ, Dong HF, Oppenheim JJ, Bizzarri C, Sergi R, Caselli G, Pagliei S, Romines B, Wilshire JA, Mengozzi M, Nakamura H, Yodoi J, Pekkari K, Gurunath R, Holmgren A, Herzenberg LA, Herzenberg LA, Ghezzi P (1999) Thioredoxin, a redox enzyme released in infection and inflammation, is a unique chemoattractant for neutrophils, monocytes, and T cells. J Exp Med 189:1783-1789

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nat Genet 29:365-371

Bridle AR, Butler R, Nowak BF (2003) Immunostimulatory CpG oligodeoxynucleotides increase resistance against amoebic gill disease in Atlantic salmon, *Salmo salar* L. J Fish Dis 26:367-371

Bridle AR, Carter CG, Morrison RN, Nowak BF (2005) The effect of beta-glucan administration on macrophage respiratory burst activity and Atlantic salmon, *Salmo salar* L., challenged with amoebic gill disease - evidence of inherent resistance. J Fish Dis 28:347-356

Bridle AR, Morrison RN, Cunningham PMC, Nowak BF (2006a) Quantitation of immune response gene expression and cellular localisation of interleukin-1 beta mRNA in Atlantic salmon, *Salmo salar* L., affected by amoebic gill disease (AGD). Vet Immunol Immunop 114:121-134

Bridle AR, Morrison RN, Nowak BF (2006b) The expression of immune-regulatory genes in rainbow trout, *Oncorhynchus mykiss*, during amoebic gill disease (AGD). Fish Shellfish Immun 20:346-364

Buates S, Matlashewski G (2001) General suppression of macrophage gene expression during *Leishmania donovani* infection. J Immunol 166:3416-3422

Campbell RD, Trowsdale J (1993) Map of the human MHC. Immunol Today 14:349-352

Carington Smith O, Wadley V (2003) Update on the salmon and trout industry in Tasmania. In *The third scientific conference of the Atlantic salmon aquaculture subprogram* (Edited by Battaglene S. C. and Cobcroft J. M.), p. 14, CSIRO, Marine Laboratories, Hobart

Chwieralski CE, Welte T, Bühling F (2006) Cathepsin-regulated apoptosis. Apoptosis 11:143-149

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

Conesa A, Götz S, García-Gómez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676

Consuegra S, Megens HJ, Leon K, Stet RJM, Jordan WC (2005a) Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. Immunogenetics 57:16-24

Consuegra S, Megens HJ, Schaschl H, Leon K, Stet RJM, Jordan WC (2005b) Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Mol Biol Evol 22:1095-1106

FAO Fisheries Department (2006) State of world aquaculture. In *FAO fisheries technical paper*, Rome

Diez-Tascón C, Keane OM, Wilson T, Zadissa A, Hyndman DL, Baird DB, McEwan JC, Crawford AM (2005) Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. Physiol Genomics 21:59-69

Douglas-Helders GM, Dawson DR, Carson J, Nowak BF (2002) Wild fish are not a significant reservoir for *Neoparamoeba pemaquidensis* (Page, 1987). J Fish Dis 25:569-574

Douglas-Helders GM, O' Brien DP, McCorkell BE, Zilberg D, Gross A, Carson J, Nowak BF (2003) Temporal and spatial distribution of paramoebae in the water column - a pilot study. J Fish Dis 26:231-240

Douglas-Helders M, Carson J, Howard T, Nowak B (2001) Development and validation of a new dot blot test for the detection of *Paramoeba pemaquidensis* (Page) in fish. J Fish Dis 24:273-280

Dyková I, Figueras A, Novoa B (1995) Amoebic gill infection of turbot, *Scophthalmus maximus*. Folia Parasit 42:91-96

Dyková I, Figueras A, Novoa B (1999) Epizoic amoebae from the gills of turbot *Scophthalmus maximus*. Dis Aquat Organ 38:33-38

Dyková I, Figueras A, Peric Z (2000) *Neoparamoeba* Page, 1987: light and electron microscopic observations on six strains of different origin. Dis Aquat Organ 43:217-223

Dyková I, Novoa B (2001) Comments on diagnosis of amoebic gill disease (AGD) in turbot, *Scophthalmus maximus*. Bull Eur Assn Fish P 21:40-44

Dyková I, Nowak BF, Crosbie PBB, Fiala I, Pecková H, Adams MB, Macháčková B, Dvořáková H (2005) *Neoparamoeba branchiphila* n. sp., and related species of the genus *Neoparamoeba* Page, 1987: morphological and molecular characterization of selected strains. J Fish Dis 28:49-64

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

Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ, Stet RJM, Waller MJ, Walter L, Marsh SGE (2006) ISAG/IUIS-VIC comparative MHC nomenclature committee report, 2005. Immunogenetics 57:953-958

Ewart KV, Belanger JC, Williams J, Karakach T, Penny S, Tsoi SCM, Richards RC, Douglas SE (2005) Identification of genes differentially expressed in Atlantic salmon (*Salmo salar*) in response to infection by *Aeromonas salmonicida* using cDNA microarray technology. Dev Comp Immunol 29:333-347

Excoffier LGL, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. In *Evolutionary Bioinformatics Online*, Vol. 1, p. 47-50

Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJA, Hofmann K, Bairoch A (2002) The PROSITE database, its status in 2002. Nucleic Acids Res 30:235-238

Findlay VL, Helders M, Munday BL, Gurney R (1995) Demonstration of resistance to reinfection with Paramoeba sp by Atlantic salmon, *Salmo salar* L. J Fish Dis 18:639-642

Findlay VL, Munday BL (1998) Further studies on acquired resistance to amoebic gill disease (AGD) in Atlantic salmon, *Salmo salar* L. J Fish Dis 21:121-125

Fjalestad KT, Moen T, Gomez-Raya L (2003) Prospects for genetic technology in salmon breeding programmes. Aquac Res 34:397-406

Gabay C, Kushner I (1999) Mechanisms of disease: Acute-phase proteins and other systemic responses to inflammation. New Engl J Med 340:448-454

Gao H, Jin SQ, Song YM, Fu M, Wang MR, Liu ZH, Wu M, Zhan QM (2005) B23 regulates GADD45a nuclear translocation and contributes to GADD45a-induced cell cycle G(2)-M arrest. J Biol Chem 280:10988-10996

Gilbey J, Verspoor E, McLay A, Houlihan D (2004) A microsatellite linkage map for Atlantic salmon (*Salmo salar*). Anim Genet 35:98-105

Gilbey J, Verspoor E, Mo TA, Sterud E, Olstad K, Hytterød S, Jones C, Noble L (2006) Identification of genetic markers associated with *Gyrodactylus salaris* resistance in Atlantic salmon *Salmo salar*. Dis Aquat Organ 71:119-129

Gjedrem T (2000) Genetic improvement of cold-water fish species. Aquac Res 31:25-33

Glamann J (1995) Complete coding sequence of rainbow trout MHC II beta-chain. Scand J Immunol 41:365-372

Glover KA, Grimholt U, Bakke HG, Nilsen F, Storset A, Skaala Ø (2007) Major histocompatibility complex (MHC) variation and susceptibilty to the sea louse *Lepeophtheirus salmonis* in Altantic salmon *Salmo salar*. Dis Aquat Organ 76:57-65

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. J Hered 86:485-486

Grimholt U, Drablos F, Jørgensen SM, Høyheim B, Stet RJM (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, Getahun A, Hermsen T, Stet RJM (2000) The major histocompatibility class II alpha chain in salmonid fishes. Dev Comp Immunol 24:751-763

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

Gross K, Carson J, Nowak B (2004a) Presence of anti-*Neoparamoeba* sp antibodies in Tasmanian cultured Atlantic salmon, *Salmo salar* L. J Fish Dis 27:81-88

Gross KA (2007) Interactions between *Neoparamoeba* spp. and Atlantic salmon (*Salmo salar* L.) immune system components. University of Tasmania, Launceston

Gross KA, Morrison RN, Butler R, Nowak BF (2004b) 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

Gross KA, Powell MD, Butler R, Morrison RN, Nowak BF (2005) Changes in the innate immune response of Atlantic salmon, *Salmo salar* L., exposed to experimental infection with *Neoparamoeba* sp. J Fish Dis 28:293-299

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98

Hansen JD, Zapata AG (1998) Lymphocyte development in fish and amphibians. Immunol Rev 166:199-220

Haugland Ø, Torgersen J, Syed M, Evensen Ø (2005) Expression profiles of inflammatory and immune-related genes in Atlantic salmon (*Salmo salar* L.) at early time post vaccination. Vaccine 23:5488-5499

Hauptmann G, Bahram S (2004) Genetics of the central MHC. Curr Opin Immunol 16:668-672

Hayes B, Laerdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Hoyheim B (2007) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. Aquaculture 265:82-90

Hill AVS (1998) The immunogenetics of human infectious diseases. Annu Rev Immunol 16:593-617

Hordvik I, Grimholt U, Fosse VM, Lie O, Endresen C (1993) Cloning and sequence analysis of cDNAs encoding the MHC class II beta chain in Atlantic salmon (*Salmo salar*). Immunogenetics 37:437-441

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci - evidence for overdominant selection. P Natl Acad Sci USA 86:958-962

Innes BH, Elliott NG (2006) Genetic diversity in a Tasmanian hatchery population of Atlantic salmon (*Salmo salar* L.) compared with its Canadian progenitor population. Aquac Res 37:563-569

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? Microbes Infect 2:1335-1341

Jenny MJ, Chapman RW, Mancia A, Chen YA, McKillen DJ, Trent H, Lang P, Escoubas JM, Bachere E, Boulo V, Liu ZJ, Gross PS, Cunningham C, Cupit PM, Tanguy A, Guo X, Moraga D, Boutet I, Huvet A, De Guise S, Almeida JS,

Warr GW (2007) A cDNA microarray for *Crassostrea virginica* and *C. gigas*. Mar Biotechnol 9:577-591

Kaufman J, Milne S, Göbel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925

Kazianis S, Morizot DC, Della Coletta L, Johnston DA, Woolcock B, Vielkind JR, Nairn RS (1999) Comparative structure and characterization of a *CDKN2* gene in a *Xiphophorus* fish melanoma model. Oncogene 18:5088-5099

Kent ML, Sawyer TK, Hedrick RP (1988) *Paramoeba pemaquidensis* (Sarcomastigophora, Paramoebidae) infestation of the gills of coho salmon *Oncorhynchus kisutch* reared in sea water. Dis Aquat Organ 5:163-169

Kim J, Tang QQ, Li X, Lane MD (2007) Effect of phosphorylation and S-S bond-induced dimerization on DNA binding and transcriptional activation by C/EBP beta. P Natl Acad Sci USA 104:1800-1804

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. Mol Ecol 10:807-821

Kipreos ET (2004) Developmental quiescence: Cdc14 moonlighting in G1. Nat Cell Biol 6:693-695

Kjøglum S, Grimholt U, Larsen S (2005) Non-MHC genetic and tank effects influence disease challenge tests in Atlantic salmon (*Salmo salar*). Aquaculture 250:102-109

Kjøglum S, Larsen S, Bakke HG, Grimholt U (2006) How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). Fish Shellfish Immun 21:431-441

Kjøglum S, Larsen S, Bakke HG, Grimholt U (2008) The effect of specific MHC class I and class II combinations on resistance to furunculosis in Atlantic salmon (*Salmo salar*). Scand J Immunol 67:160-168

Kopitar-Jerala N (2006) The role of cystatins in cells of the immune system. Febs Lett 580:6295-6301

Kotsyfakis M, Sá-Nunes A, Francischetti IMB, Mather TN, Andersen JF, Ribeiro JMC (2006) Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick *Ixodes scapularis*. J Biol Chem 281:26298-26307

Krasnov A, Koskinen H, Pehkonen P, Rexroad CE, Afanasyev S, Molsa H (2005) Gene expression in the brain and kidney of rainbow trout in response to handling stress. BMC Genomics 6:3

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150-163

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

Langefors A, Lohm J, Grahn M, Andersen Ø, von Schantz T (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. P Roy Soc Lond B Bio 268:479-485

Langefors A, Lohm J, von Schantz T, 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

Lidie KB, Ryan JC, Barbier M, Van Dolah FM (2005) Gene expression in Florida red tide dinoflagellate *Karenia brevis*: Analysis of an expressed sequence tag library and development of DNA microarray. Mar Biotechnol 7:481-493

Liu HC, Cheng HH, Tirunagaru V, Sofer L, Burnside J (2001) A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. Anim Genet 32:351-359

Lohm J, Grahn M, Langefors A, Andersen Ø, Storset A, von Schantz T (2002) Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. P Roy Soc Lond B Bio 269:2029-2033

Lotteau V, Teyton L, Peleraux A, Nilsson T, Karlsson L, Schmid SL, Quaranta V, Peterson PA (1990) Intracellular transport of class II MHC molecules directed by invariant chain. Nature 348:600-605

Málaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sültmann H, Klein J (1998) Linkage relationships and haplotype polymorphism among cichlid *Mhc* class II *B* loci. Genetics 149:1527-1537

Malumbres M, Barbacid M (2001) To cycle or not to cycle: A critical decision in cancer. Nat Rev Cancer 1:222-231

Marsh SGE, Parham P, Barber LD (2000) *The HLA factsbook.* Academic Press, San Diego

Martin SAM, Taggart J, Seear P, Bron J, Talbot RT, Sweeney G, Teale A, Høyheim B, Houlihan D, Zou J, Tocher D, Secombes C (2007a) Interferon type I and type II responses in a Atlantic salmon (*Salmo salar*) SHK-1 cell line using the salmon TRAITS/SGP microarray. Physiol Genomics 32:33-44

Martin SAM, Zou J, Houlihan DF, Secombes CJ (2007b) Directional response following recombinant cytokine stimulation of rainbow trout (*Oncorhynchus*

mykiss) RTS-11 macrophage cells as revealed by transcriptome profiling. BMC Genomics 8:150

Martinez VA, Hill WG, Knott S (2002) On the use of double haploids for detecting QTL in outbred populations. Heredity 88:423-431

Matza D, Kerem A, Shachar I (2003) Invariant chain, a chain of command. Trends Immunol 24:264-268

Menon JN, Bretscher PA (1998) Parasite dose determines the Th1/Th2 nature of the response to *Leishmania major* independently of infection route and strain of host or parasite. Eur J Immunol 28:4020-4028

Miller KM, Winton JR, Schulze AD, Purcell MK, Ming TJ (2004) Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. Environ Biol Fish 69:307-316

Moen T, Fjalestad KT, Munck H, Gomez-Raya L (2004a) A multistage testing strategy for detection of quantitative trait loci affecting disease resistance in Atlantic salmon. Genetics 167:851-858

Moen T, Høyheim B, Munck H, Gomez-Raya L (2004b) A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. Anim Genet 35:81-92

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

Morrison RN, Cooper GA, Koop BF, Rise ML, Bridle AR, Adams MB, Nowak BF (2006a) Transcriptome profiling the gills of amoebic gill disease (AGD)-

affected Atlantic salmon (*Salmo salar* L.): a role for tumor suppressor p53 in AGD pathogenesis? Physiol Genomics 26:15-34

Morrison RN, Crosbie PBB, Nowak BF (2004) The induction of laboratory-based amoebic gill disease revisited. J Fish Dis 27:445-449

Morrison RN, Koppang EO, Hordvik I, Nowak BF (2006b) MHC class II+ cells in the gills of Atlantic salmon (*Salmo salar* L.) affected by amoebic gill disease. Vet Immunol Immunop 109:297-303

Morrison RN, Zou J, Secombes CJ, Scapigliati G, Adams MB, Nowak BF (2007) Molecular cloning and expression analysis of tumour necrosis factor-alpha in amoebic gill disease (AGD)-affected Atlantic salmon (*Salmo salar* L.). Fish Shellfish Immun 23:1015-1031

Munday BL (1986) Diseases of salmonids. In *Proceedings of the workshop on diseases of Australian fish and shellfish* (Edited by Humphrey J. D. and Langdon J. S.), p. 127-141. Department of Agriculture and Rural Affairs, Benalla, Victoria

Munday BL, Foster CK, Roubal FR, Lester RJG (1990) *Paramoebic gill infection and associated pathology of Atlantic salmon, Salmo salar L. and rainbow trout, Salmo gairdneri, in Tasmania.* Academic Press, San Diego

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

Murray BW, Shintani S, Sültmann H, Klein J (2000) Major histocompatibility complex class II *A* genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. Immunogenetics 51:576-586

Newell WR, Trowsdale J, Beck S (1996) MHCDB: database of the human MHC (release 2). Immunogenetics 45:6-8

Nichols KM, Bartholomew J, Thorgaard GH (2003) Mapping multiple genetic loci associated with *Ceratomyxa shasta* resistance in *Oncorhynchus mykiss*. Dis Aquat Organ 56:145-154

Nonaka M, Matsuo M, Naruse K, Shima A (2001) Comparative genomics of medaka: The major histocompatibility complex (MHC). Mar Biotechnol 3:S141-S144

Nonaka M, Namikawa C, Kato Y, Sasaki M, Salter-Cid L, Flajnik MF (1997) Major histocompatibility complex gene mapping in the amphibian Xenopus implies a primordial organization. P Natl Acad Sci USA 94:5789-5791

O' Brien SJ, Wildt DT, Goldman D, Merril CR, Bush M (1985) The cheetah is depauperate in genetic variation. Science 221:459-462

Olsvik PA, Lie KK, Jordal AEO, Nilsen TO, Hordvik I (2005) Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. BMC Mol Biol 6:21

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

Palti Y, Nichols KM, Waller KI, Parsons JE, Thorgaard GH (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

Parsons H, Nowak B, Fisk D, Powell M (2001) Effectiveness of commercial freshwater bathing as a treatment against amoebic gill disease in Atlantic salmon. Aquaculture 195:205-210

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). P Natl Acad Sci USA 95:3714-3719

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST (c)) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30:e36

Phillips RB, Zimmerman A, Noakes MA, Palti Y, Morasch MRW, Eiben L, Ristow SS, Thorgaard GH, Hansen JD (2003) Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. Immunogenetics 55:561-569

Pol E, Björk I (2001) Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases. Protein Sci 10:1729-1738

Poli V (1998) The role of C/EBP isoforms in the control of inflammatory and native immunity functions. J Biol Chem 273:29279-29282

Powell MD, Fisk D, Nowak BF (2000) Effects of graded hypoxia on Atlantic salmon infected with amoebic gill disease. J Fish Biol 57:1047-1057

Purcell MK, Nichols KM, Winton JR, Kurath G, Thorgaard GH, Wheeler P, Hansen JD, Herwig RP, Park LK (2006) Comprehensive gene expression profiling following DNA vaccination of rainbow trout against infectious hematopoietic necrosis virus. Mol Immunol 43:2089-2106

Rammensee HG (1995) Chemistry of peptides associated with MHC class I and class II molecules. Curr Opin Immunol 7:85-96

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

Reilly A, Elliott NG, Grewe PM, Clabby C, Powell R, Ward RD (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

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223-225

Rise ML, Jones SRM, Brown GD, von Schalburg KR, Davidson WS, Koop BF (2004a) Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. Physiol Genomics 20:21-35

Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SRM, Smailus DE, Jones SJM, Schein JE, Marra MA, Butterfield YSN, Stott JM, Ng SHS, Davidson WS, Koop BF (2004b) Development and application of a salmonid EST database and cDNA microarray: Data mining and interspecific hybridization characteristics. Genome Res 14:478-490

Roberge C, Páez DJ, Rossignol O, Guderley H, Dodson J, Bernatchez L (2007) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. Mol Immunol 44:1374-1383

Roberts SD, Powell MD (2003) Reduced total hardness of fresh water enhances the efficacy of bathing as a treatment for amoebic gill disease in Atlantic salmon, *Salmo salar* L. J Fish Dis 26:591-599

Rodger HD, McArdle JF (1996) An outbreak of amoebic gill disease in Ireland. Vet Rec 139:348-349

Rodriguez MF, LaPatra S, Williams S, Famula T, May B (2004) Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (Oncorhynchus mykiss) backcrosses. Aquaculture 241:93-115

Sato A, Figueroa F, Murray BW, Málaga-Trillo E, Zaleska-Rutczynska Z, Sültmann H, Toyosawa S, Wedekind C, Steck N, Klein J (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51:108-116

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

Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, Stet RJM, Secombes C, Parham P (2001) Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J Immunol 166:3297-3308

Shum BP, Mason PM, Magor KE, Flodin LR, Stet RJM, Parham P (2002) Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 54:193-199

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3

Smyth GK (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Edited by Gentleman R., Carey V., Dudoit S., Irizarry R. and Huber W.), p. 397-420. Springer, New York

Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21:2067-2075

Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods 31:265-273

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2:16

Sonesson AK (2007a) Possibilities for marker-assisted selection in aquaculture breeding schemes. In *Marker-assisted selection; current status and future perspectives in crops, livestock, forestry and fish* (Edited by Guimarães E., Ruane J., Scherf B., Sonnino A. and Dargie J.), p. 310-328. FAO, Rome

Sonesson AK (2007b) Within-family marker-assisted selection for aquaculture species. Genet Sel Evol 39:301-317

Speed T (2003) *Interdisciplinary Statistics: Statistical Analysis of Gene Expression Microarray data*. Chapman & Hall, Boca Raton, USA

Steinum T, Kvellestad A, Rønneberg LB, Nilsen H, Asheim A, Fjell K, Nygård SMR, Olsen AB, Dale OB (2008) First cases of amoebic gill disease (AGD) in

Norwegian seawater farmed Atlantic salmon, *Salmo salar* L., and phylogeny of the causative amoeba using 18S cDNA sequences. J Fish Dis 31:205-214

Stet RJM, de Vries B, Mudde K, Hermsen T, van Heerwaarden J, Shum BP, 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

Storni T, Bachmann MF (2004) Loading of MHC class I and II presentation pathways by exogenous antigens: A quantitative in vivo comparison. J Immunol 172:6129-6135

Streetz KL, Wustefeld T, Klein C, Manns MP, Trautwein C (2001) Mediators of inflammation and acute phase response in the liver. Cell Mol Biol 47:661-673

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967-978

Tang QQ, Grönborg M, Huang HY, Kim JW, Otto TC, Pandey A, Lane MD (2005) Sequential phosphorylation of CCAAT enhancer-binding protein beta by MAPK and glycogen synthase kinase 3 beta is required for adipogenesis. P Natl Acad Sci USA 102:9766-9771

Taylor RS, Wynne JW, Kube PD, Elliott NG (2007) Genetic variation of resistance to amoebic gill disease in Atlantic salmon assessed in a challenge system. Aquaculture 272S1:S94-299

Tidemand-Johnannessen P (1999) Salmonid culture: history and development. In *Manual of salmonid farming* (Edited by Willoughby S.), p. 1-18. Blackwell Science Ltd, Oxford

Vallejo RL, Bacon LD, Liu HC, Witter RL, Groenen MAM, Hillel J, Cheng HH (1998) Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F-2 intercross chickens. Genetics 148:349-360

Vincent BN, Morrison RN, Nowak BF (2006) Amoebic gill disease (AGD)-affected Atlantic salmon, *Salmo salar* L., are resistant to subsequent AGD challenge. J Fish Dis 29:549-559

Viville S, Neefjes J, Lotteau V, Dierich A, Lemeur M, Ploegh H, Benoist C, Mathis D (1993) Mice lacking the MHC class II associated invariant chain. Cell 72:635-648

von Schalburg KR, Rise ML, Brown GD, Davidson WS, Koop BF (2005a) A comprehensive survey of the genes involved in maturation and development of the rainbow trout ovary. Biol Reprod 72:687-699

von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF (2005b) Fish and chips: Various methodologies demonstrate utility of a 16,006-gene salmonid microarray. BMC Genomics 6:126-144

Walsh PS, Metzger DA, Higuchi R (1991) Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506-513

Wang B, Li FH, Dong B, Zhang XJ, Zhang CS, Xiang JH (2006) Discovery of the genes in response to white spot syndrome virus (WSSV) infection in *Fenneropenaeus chinensis* through cDNA microarray. Mar Biotechnol 8:491-500

Ward RD, Grewe PM, Smolenski AJ (1994) A comparison of allozymes and mitochondrial-DNA in Atlantic salmon from tasmania and from the ancestral population in Canada. Aquaculture 126:257-264

Wong FYK, Carson J, Elliott NG (2004) 18S ribosomal DNA-based PCR identification of *Neoparamoeba pemaquidensis*, the agent of amoebic gill disease in sea-farmed salmonids. Dis Aquat Organ 60:65-76

Wynne JW, Cook MT, Holmes BH, Elliott NG (2007a) Allelic and haplotypic diversity at the major histocompatibility class II within domesticated Australian Atlantic salmon (*Salmo salar* L.). J Fish Biol 70:45-59

Wynne JW, Cook MT, Nowak BE, Elliott NG (2007b) Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (*Salmo salar* L.). Fish Shellfish Immun 22:707-717

Wynne JW, O' Sullivan MG, Cook MT, Stone G, Nowak BF, Lovell DR, Elliott NG (2008) Transcriptome analyses of amoebic gill disease affected Atlantic salmon (*Salmo salar*) tissues reveal localised host gene suppression. Mar Biotechnol *in press*

Yang XL, Cox-Foster DL (2005) Impact of an ectoparasite on the immunity and pathology of an invertebrate: Evidence for host immunosuppression and viral amplification. P Natl Acad Sci USA 102:7470-7475

Yonash N, Bacon LD, Witter RL, Cheng HH (1999) High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. Anim Genet 30:126-135

Young ND, Cooper GA, Nowak BF, Koop BF, Morrison RN (2008a) Coordinated down-regulation of the antigen processing machinery in the gills of amoebic gill disease affected Atlantic salmon (*Salmo salar* L.). Mol Immunol *in press*

Young ND, Crosbie PBB, Adams MB, Nowak BF, Morrison RN (2007) *Neoparamoeba perurans* n sp., an agent of amoebic gill disease of Atlantic salmon (*Salmo salar* L.). Int J Parasitol 37:1469-1481

Young ND, Dyková I, Nowak BF, Morrison RN (2008b) Development of a diagnostic PCR to detect *Neoparamoeba perurans*, agent of amoebic gill disease (AGD). J Fish Dis *in press*

Young ND, Dyková I, Snekvik K, Nowak BF, Morrison RN (2008c) *Neoparamoeba perurans* is a cosmopolitan aetiological agent of amoebic gill disease. Dis Aquat Organ *in press*

Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK, Smith NJ (2003) Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. Plant Breeding 122:40-46

Zilberg D, Munday BL (2001) The effect of anti-*Paramoeba* antibodies on *Paramoeba* sp., the causative agent of amoebic gill disease. J Fish Dis 24:345-350

Zilberg D, Nowak B, Carson J, Wagner T (1999) Simple gill smear staining for diagnosis of amoebic gill disease. Bull Eur Assn Fish P 19:186-189

Appendices

Appendix 1.1 Co-authored manuscript relevant to this thesis.

Taylor, R. S., Wynne, J. W., Kube, P. D., Elliott, N. G., 2007. Genetic variation of resistance to amoebic gill disease in Atlantic salmon (Salmo salar) assessed in a challenge system, Aquaculture, 272(S1), S94-S99
see: https://doi.org/10.1016/j.aquaculture.2007.08.007

Appendix 2.1 Frequency of *Sasa-DAA-3UTR* genotypes within the 30 full-sibling families challenge with AGD. The parental column displays the sire (♂) and dam (♀) genotypes. For each family all possible offspring genotype classes are shown. The expected number of offspring within each class is indicated in *parenthesis* and the observed number of offspring indicated to the *right*. Mendelian inheritance was then tested using χ^2 analysis and the χ^2 and p value are shown. *Dashes* represent families that demonstrated perfect mendelian inheritance and were therefore not tested with χ^2 analysis.

Family	Parental	Observed no. offspring in each genotype class				χ^2	p
04003	♀207/229	207/229	229/229	207/207		1.118	0.572
	♂207/229	8 (8.75)	6 (4.25)	3 (4.25)			
04005	♀229/229	229/259				-	-
	♂259/259	18 (18.00)					
04008	♀207/207	207/239	207/259			-	-
	♂239/259	7 (7.00)	7 (7.00)				
04011	♀239/259	239/259	229/259	229/239	259/259	2.571	0.463
	♂229/259	5 (3.50)	4 (3.50)	4 (3.50)	1 (3.50)		
04017	♀259/259	229/259	259/259			0.059	0.808
	♂229/259	8 (8.50)	9 (8.50)				
04020	♀259/259	229/259	250/259			0.250	0.617
	♂229/250	9 (8.00)	7 (8.00)				
04022	♀207/259	207/259	259/259			0.889	0.346
	♂259/259	11 (9.00)	7 (9.00)				
04023	♀229/229	229/259	229/229			1.667	0.197
	♂229/259	5 (7.50)	10 (7.50)				
04027	♀207/279	207/259	207/207	259/279	207/279	3.308	0.347
	♂207/259	2 (3.25)	2 (3.25)	3 (3.25)	6 (3.25)		
04028	♀259/279	207/259	207/279	229/259	229/279	7.706	0.052
	♂207/229	1 (4.25)	8 (4.25)	6 (4.25)	2 (4.25)		
04029	♀229/229	207/229	229/229			2.882	0.090
	♂207/229	12 (8.50)	5 (8.50)				
04030	♀207/259	207/229	229/259			0.250	0.617
	♂229/229	9 (8.00)	7 (8.00)				
04031	♀229/229	229/279	229/229			0.067	0.796
	♂229/279	7 (7.50)	8 (7.50)				
04033	♀207/239	207/229	229/239			-	-
	♂229/229	7 (7.00)	7 (7.00)				
04034	♀239/250	229/239	229/250			0.067	0.796

Appendix 2.1. Continued over page

	♂229/229	7 (7.50)	8 (7.50)				
04037	♀207/279	207/279	207/207			-	-
	♂207/207	8 (8.00)	8 (8.00)				
04038	♀229/279	229/229	229/279	279/279		4.385	0.112
	♂229/279	6 (3.25)	3 (6.50)	4 (3.25)			
04039	♀239/239	239/259	239/298			1.000	0.317
	♂259/298	6 (8.00)	10 (8.00)				
04042	♀259/298	239/259	239/298			0.250	0.617
	♂239/239	7 (8.00)	9 (8.00)				
04043	♀229/229	229/259	229/279			0.067	0.796
	♂259/279	7 (7.50)	8 (7.50)				
04048	♀229/259	229/259	229/279	229/229	259/279	5.000	0.172
	♂229/279	5 (3.75)	0 (3.75)	5 (3.75)	5 (3.75)		
04049	♀250/259	239/250	250/250	239/259	250/259	7.500	0.058
	♂239/250	5 (4.00)	1 (4.00)	2 (4.00)	8 (4.00)		
04050	♀250/259	229/250	229/259	250/259	259/259	0.176	0.981
	♂229/259	4 (4.25)	5 (4.25)	4 (4.25)	4 (4.25)		
04062	♀250/259	250/259	229/250	229/259	259/259	4.500	0.212
	♂229/259	4 (4.00)	7 (4.00)	4 (4.00)	1 (4.00)		
04065	♀229/259	207/229	207/259	229/229	229/259	0.647	0.886
	♂207/229	3 (4.25)	5 (4.25)	5 (4.25)	4 (4.25)		
04066	♀229/229	207/229	229/229			0.059	0.808
	♂207/229	8 (8.50)	9 (8.50)				
04068	♀207/259	207/239	207/259	239/259	259/259	1.500	0.682
	♂239/259	2 (4.00)	5 (4.00)	5 (4.00)	4 (4.00)		
04078	♀229/279	229/259	229/279	259/279	279/279	1.267	0.737
	♂259/279	4 (3.75)	5 (3.75)	4 (3.75)	2 (3.75)		
04129	♀279/298	207/279	207/298	229/279	229/298	1.800	0.615
	♂207/229	5 (3.75)	3 (3.75)	5 (3.75)	2 (3.75)		
04134	♀207/239	207/229	207/259	229/239	239/259	4.222	0.238
	♂229/259	8 (4.50)	4 (4.50)	4 (4.50)	2 (4.50)		

Appendix 2.2 Frequency of *Sasa-UBA-3UTR* genotypes within the 30 full-sibling families challenge with AGD. The parental column displays the sire (♂) and dam (♀) genotypes. For each family all possible offspring genotype classes are shown. The expected number of offspring within each class is indicated in *parenthesis* and the observed number of offspring indicated to the *right*. Mendelian inheritance was then tested using χ^2 analysis and the χ^2 and p value are shown. *Dashes* represent families that demonstrated perfect mendelian inheritance and were therefore not tested with χ^2 analysis.

Family	Parental	Observed no. offspring in each genotype class				χ^2	p
04003	♀313/317	313/327	313/335	317/327	317/335	2.059	0.560
	♂327/335	6 (4.25)	4(4.25)	2(4.25)	5(4.25)		
04005	♀317/317	317/327	317/335			0.889	0.346
	♂327/335	11 (9)	7 (9)				
04008	♀327/333	321/327	321/333	327/335	333/335	1.429	0.699
	♂321/335	3 (3.50)	4 (3.50)	2 (3.50)	5 (3.50)		
04011	♀321/327	321/333	321/335	327/333	327/335	0.286	0.963
	♂333/335	4 (3.50)	3 (3.50)	3 (3.50)	4 (3.50)		
04017	♀327/333	321/327	321/333	327/327	327/333	0.647	0.886
	♂321/327	5 (4.25)	4 (4.25)	3 (4.25)	5 (4.25)		
04020	♀327/327	313/327	327/327			1.000	0.317
	♂313/327	10 (8)	6 (8)				
04022	♀313/327	313/327	313/321	321/327	327/327	2.000	0.572
	♂321/327	4 (4.50)	3 (4.50)	7 (4.50)	4 (4.50)		
04023	♀317/327	313/317	313/327	317/327	327/327	5.500	0.139
	♂313/327	3 (3.75)	3 (3.75)	8 (3.75)	2 (3.75)		
04027	♀313/327	309/313	309/327	313/333	327/333	3.308	0.347
	♂309/333	2 (3.25)	2 (3.25)	3 (3.25)	6 (3.25)		
04028	♀327/333	321/327	321/333	327/335	333/335	2.059	0.506
	♂321/335	2 (4.25)	6 (4.25)	4 (4.25)	5 (4.25)		
04029	♂315/317	313/315	313/317	315/327	317/327	1.118	0.773
	♀313/327	4 (4.25)	6 (4.25)	3 (4.25)	4 (4.25)		
04030	♂313/327	313/317	317/327			2.250	0.134
	♀317/317	11 (8.00)	5 (8.00)				
04031	♂327/327	321/327	327/335			0.067	0.796
	♀321/327	8 (7.50)	7 (7.50)				
04033	♂313/327	313/327	313/335	327/335	327/327	1.429	0.699
	♀327/335	5 (3.50)	4 (3.50)	2 (3.50)	3 (3.50)		
04034	♂327/329	317/327	317/329	327/333	329/333	7.667	0.053

Appendix 2.2. Continued over page

	♀317/333	2 (3.75)	8 (3.75)	4 (3.75)	1 (3.75)		
04037	♂327/327	327/327	327/333			0.250	0.617
	♀327/333	7 (8.00)	9 (8.00)				
04038	♂327/327	327/327				-	-
	♀327/327	13 (13.00)					
04039	♂327/327	321/327	327/333			2.250	0.134
	♀321/333	5(8.00)	11 (8.00)				
04042	♂327/335	309/327	309/335	327/327	327/335	0.500	0.919
	♀309/327	5 (4.00)	4 (4.00)	4 (4.00)	3 (4.00)		
04043	♂317/335	317/321	317/335	321/335	335/335	3.933	0.269
	♀321/335	5 (3.75)	1 (3.75)	6 (3.75)	3 (3.75)		
04048	♂333/333	327/333	333/335			0.067	0.796
	♀327/335	7 (7.50)	8 (7.50)				
04049	♂327/333	327/329	327/333	329/333	327/327	0.500	0.919
	♀327/329	3 (4.00)	4 (4.00)	5 (4.00)	4 (4.00)		
04050	♂317/327	317/317	317/327	327/327		1.588	0.452
	♀317/327	5 (4.25)	6 (8.50)	6 (4.25)			
04062	♂315/315	315/321	315/327			0.250	0.617
	♀321/327	7 (8.00)	9 (8.00)				
04065	♂313/335	313/313	313/335	313/327	327/335	0.647	0.886
	♀313/327	5 (4.25)	5 (4.25)	4 (4.25)	3 (4.25)		
04066	♂317/333	317/321	317/335	321/333	333/335	7.235	0.065
	♀321/335	9 (4.25)	3 (4.25)	3 (4.25)	2 (4.25)		
04068	♂327/333	313/327	313/333	327/327	327/333	2.500	0.475
	♀313/327	2 (4.00)	6 (4.00)	5 (4.00)	3 (4.00)		
04078	♂313/327	313/327	313/333	327/327	327/333	4.467	0.215
	♀327/333	4 (3.75)	2 (3.75)	7 (3.75)	2 (3.75)		
04129	♂323/333	313/323	313/333	323/327	327/333	4.467	0.215
	♀313/327	7 (3.75)	2 (3.75)	4 (3.75)	2 (3.75)		
04134	♂327/333	309/327	309/333	327/333	333/333	2.059	0.560
	♀309/333	4 (4.25)	2 (4.25)	6 (4.25)	5 (4.25)		

Appendix 3.1 Transcripts up-regulated within the AGD resistance Atlantic salmon compared to the AGD susceptible individuals with either no BLAST hit (Unknown) or a BLAST hit (E<1x10^{-1}) to a non-annotated UniGene cluster. The transcript accession number or TRAITS clone ID is shown along with the fold-change and significance (p).

ID	Identity	P.Value	Fold-change
CK895333	UniGene Ssa.24399	1.57E-09	4.11
EG999226	UniGene Ssa.31308	2.052E-13	3.85
gil_cgi_D4H01_car_tra_sub_0p	Unknown	4.727E-14	3.15
CK884295	UniGene Ssa.3495	8.495E-10	3.01
gil_cgi_D4C10_car_tra_sub_0p	Unknown	5.169E-14	2.99
AM041525	UniGene Ssa.31308	1.289E-10	2.92
CO469776	UniGene Ssa.26196	1.776E-09	2.68
CK877715	UniGene Ssa.31308	1.07E-09	2.64
CK884656	UniGene Ssa.22016	1.932E-10	2.48
CO470831	UniGene Ssa.22718	6.313E-11	2.44
CK884239	UniGene Ssa.3495	2.633E-09	2.44
DW589701	Unknown	3.525E-12	2.28
CK878732	Unknown	2.072E-11	2.28
CK880285	UniGene Ssa.6457	1.238E-06	2.28
CK883288	Unknown	2.65E-07	2.24
CK877868	UniGene Ssa.25803	1.818E-10	2.21
AJ424291	UniGene Ssa.20821	1.721E-13	2.21
CK879778	Unknown	1.36E-08	2.20
DW591257	UniGene Ssa.12317	7.584E-11	2.15
EG648661	Unknown	1.578E-15	2.13
AM041468	UniGene Ssa.26454	2.714E-12	2.13
CK887860	Unknown	7.152E-11	2.10
CK885425	UniGene Ssa.7860	2.61E-10	2.09
CK879523	UniGene Ssa.2004	1.49E-11	2.09
CK874820	UniGene Ssa.4597	3.144E-13	2.09
CK877700	UniGene Ssa.24719	1.705E-14	2.08
AJ424229	UniGene Ssa.25476	7.720E-11	2.07
CK879898	UniGene Ssa.2004	2.25E-12	2.06
CO471746	UniGene Ssa.33113	5.494E-10	2.05
CK877148	UniGene Ssa.25762	8.041E-11	2.05
CK877541	UniGene Ssa.25770	4.82E-13	2.05
CK880525	UniGene Ssa.33252	3.583E-04	2.03
CK879035	UniGene Ssa.25809	5.243E-07	2.03
AJ424195	UniGene Ssa.15328	7.338E-11	2.00

Appendix 3.2 Transcripts down-regulated within the AGD resistance Atlantic salmon compared to the AGD susceptible individuals with either no BLAST hit (Unknown) or a BLAST hit (E<1x10^{-1}) to a non-annotated UniGene cluster. The transcript accession number or TRAITS clone ID is shown along with the fold-change and significance (p).

ID	Identity	p.Value	Fold-change
CK880278	UniGene Ssa.25836	6.809E-22	- 22.07
CO472444	UniGene Ssa.26266	1.012E-11	- 7.83
gil_agi_05E10_abe_tra_sub_0p	Unknown	3.473E-16	- 7.29
CO470395	UniGene-Ssa.35702	1.899E-14	- 6.68
CK885324	Unknown	4.613E-13	- 3.39
CK893353	UniGene Ssa.11643	2.052E-11	- 3.26
ova_oyr_04C10_gal_sal_std_5p	UniGene Ssa.15598	7.798E-13	- 2.94
CK876054	UniGene Ssa.25726	4.708E-16	- 2.79
BM414494	UniGene Omy.4010	2.966E-10	- 2.77
bra_bfo_12B08_fou_sal_nrc_5p	Unknown	6.292E-14	- 2.58
CK881171	Unigene Ssa.6661	1.567E-07	- 2.42
CK888199	UniGene Ssa.4452	1.500E-08	- 2.31
CK893980	UniGene Ssa.756	8.427E-15	- 2.25
CK896510	UniGene Ssa.25495	1.630E-10	- 2.22
CN181449	UniGene Ssa.26192	8.333E-11	- 2.21
CK898963	UniGene Ssa.1723	4.539E-07	- 2.17
CK891700	UniGene Ssa.2450	3.284E-10	- 2.17
CK899096	UniGene Ssa.26131	2.040E-09	- 2.14
CK890296	UniGene Ssa.15229	7.812E-15	- 2.12
mus_snm_06H11_osl_tra_nrc_5p	UniGene Ssa.6148	5.649E-02	- 2.12
AM041585	UniGene Ssa.26445	3.395E-09	- 2.08
EG649380	UniGene Ssa.22205	3.532E-10	- 2.08
bra_bfo_16H09_fou_sal_nrp_5p	Unknown	1.366E-12	- 2.06
CK891055	UniGene Ssa.20915	7.919E-08	- 2.06
CK889354	UniGene Ssa.4452	2.953E-07	- 2.01
CO469805	UniGene Ssa.1583	1.456E-13	- 2.01
CK882496	UniGene Ssa.986	4.424E-12	- 2.01

Appendix 4.0 Animal ethics approvals. All animal experimentation described within this thesis has been approved by the University of Tasmania Animal Ethics Committee and conducted under the guidelines of the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes 7th edition 2004.

(www.nhmrc.gov.au/publications/synopses/ea16syn.htm)

Application title	Project No.
Molecular assessment of resistance to amoebic gill disease (AGD) in Atlantic salmon (*Salmo salar*)	A0008088
Performance and genetic characterisation of Atlantic salmon in the Tasmanian selective breeding programme related to amoebic gill disease	A0009111