Genetic Resistance to Infectious Pancreatic Necrosis virus in Pedigreed Atlantic salmon (*Salmo salar*)

Derrick R Guy

B.Sc. (London, UK)

M.Sc. (Birmingham, UK)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Edinburgh

2011

Contents

Declaration of contribution to the work

I declare that all chapters of this thesis were composed by myself and are entirely the result of my own work, under the guidance of my supervisors. Interpretations and conclusions are entirely my own.

Chapters two and three are already published as journal papers. Supervisors supervised appropriately as co-authors.

Initiation of this work attracted further research collaboration in parallel to this specific study, and with which I was directly involved. I have made it explicit in the Introduction (1.1) what my contribution was to those projects since they are extensively discussed in this thesis and I was co-author of several of the papers cited, as listed below.

Unless I am cited as first author of any referenced work it should be assumed it is primarily the work of others, and not myself, even if I am co-author. Such work will only be related to contextual discussions.

The animals and data upon which this study depends are derived from a breeding program with which I have shared technical responsibility, and involves the collaborative use of facilities.
My contribution to the provision of such material is :

- I initially set up and managed the breeding program, the genotyping laboratory and the program of data collection.
- Data collection was a collaboration between company staff, owners of field sites and experimental facilities, and friends and family coerced into participating.
- I have personally taken a hand at some stage in collecting all forms of data and traits analysed here.
- I remain responsible for data management and analysis.
- I remain responsible for nominating families and fish material going to field trials and collaborative projects.

The work reported here has not been submitted to any other degree or professional qualification.

Signed : .....

Dedication

There are times in my life when I have found myself climbing out from the bottom of a very deep hole - literally - with the force of all the world's rivers seemingly to conspire to send me back in the opposite direction - fear and exhilaration merged into one. This was one of those times.

" ***When in doubt, do not look forward .....When in regret, do not look back .....But look deep inside, and ask yourself :- Is there anything, I've yet to unpack ?"***

- An Englishman in Dunblane (with apologies to Quentin Crisp)

This thesis is dedicated to my family, who have had to patiently endure what must seem like an adolescent dream that I should have got out of my system over thirty years ago.

To my Dad, who even now won't quite understand fully how those trips to the allotment round the back of a Sheffield coal tip, on the back of his bike at the age of five, was to start me off on a lifetime's ambition - just a "young kid with a crazy dream" (L. Cohen, 2009) - of becoming a 'naturalist' when I grew up. Well, I hope by this to have finally achieved the first. The second will probably be forever out of reach.

To my girls, Jane Laura and Rebecca (an Edinburgh Univeristy alumni herself), dragged along in the slipstream, or more correctly, the other way round :-if this looks like one last desperate attempt to keep up with the combined academic achievements of wife, daughters and nieces, well, you would be absolutely right.

Most of all, this is dedicated to Oliver, first born of the next generation. There will come a time when all I will have left you is a random sample of one quarter of my genes (or is it an eighth? I never could work it out) and hopefully this - lasting proof that you are never too old, and it is never too late, to follow your dreams.

Acknowledgments

At the time of writing, of course it is still not absolutely guaranteed that this thesis passes muster. That quite rightly is for others to decide. My only hope at this stage is that this work does justice to the outstanding support that I, individually, and we, collectively, have received during the carrying out of this work.

To my supervisors, Prof John Woolliams, Prof Steve Bishop and Dr Sue Brotherstone : I well remember a warm august night in Montpellier 2002, post the most lavish post-conference dinner I have yet to surpass, post the 'Bishop's Course' on genetic epidemiology, when I became vaguely aware (through the heady mix of wine and jazz) that the seeds of an offer to support our work in this way were being sown, there and then. I'm sure I said 'do you really, really, know what you are letting yourselves in for?' but probably not loud enough. If this work has scaled any heights at all, then it is entirely by standing on the shoulders of these giants of quantitative genetics.

To Genesis Faraday (now KTN) in particular Chris Warkup, my former boss from a former life - for reaching out across various burnt, smouldering bridges and helping to provide me with a valuable taste of proper post-grad life and training

To the LNS executive and shareholders, particularly Hugh Currie and Professor Ron Roberts for firstly, paying the bills for far longer than I'm sure was anticipated, but far more importantly, for keeping the faith in the whole enterprise, and in myself in particular, during some of the worse trading conditions a company has ever had to face.

To LNS Colleagues, past and present - in the beginning there was just Ron and myself so I just can't begin to describe what it is like to be surrounded by such a lively and challenging bunch of misfits (ok, that's just me then), every one of which has had to go the extra mile (or two) to realise this project described herein.

To Landcatch colleagues, past and present -out on the fish farm looking after the fish - I am only too aware that almost any request from me has resulted in that extra mile or two being in the horizontal sleet of a cold Scottish winter's night, way beyond the call of duty. And I know that me saying 'been there, done that' doesn't really compensate. Thanks anyway.

To the guys and lasses at the **CEFAS freshwater facility Weymouth** - I know it reads in the accounts like a purely business arrangement, but Marianne and I appreciate our first visit all those years back as the genuine extended hand of friendship that it was, and it transformed the fortunes of this project overnight.

To the many **owners of the seawater sentinel sites, processing plants, friends at the IoA, Stirling, friends and friends family members** who all got coerced into helping out with the collection of field data over the years - it may seem a small, fleeting contribution in the grand scheme of things, but it is remembered and appreciated every time the data passes across my desk. This project would not have even got off the ground without your efforts.

Finally, thanks again to **my family,** who have had to pay the real price for showing patience, love and understanding way beyond the call of duty. Your (not inconsiderable) demands, were always a legitimate, welcome and essential distraction, and decidedly the only thing keeping me grounded during this somewhat selfish adventure.

Whatever truths may be attested and challenged herein, the following is unequivocal : On approaching the end of their career, no father has ever said: 'I wish I hadn't spent so much time with my family, and had devoted more time to my career'.

If that is the only message picked up by what I imagine to be the young, enthusiastic, career-building readers of this thesis, I will consider my time well spent.

List of Tables

Tables are labelled in order within chapter :

List of Figures

List of Publications

List of papers and presentations submitted or published by DRG as principal author relating to this thesis, while undergoing postgraduate training, by date.

As principal author constituting chapters of this thesis :

Guy D.R., Bishop S.C., Brotherstone S., Hamilton A., Roberts R.J., McAndrew B.J., Woolliams J.A. (2006)
Analysis of the incidence of infectious pancreatic necrosis mortality in pedigreed Atlantic salmon, *Salmo salar* L. populations
Journal of Fish Diseases 29 (2006) 637-647

[Chapter 2]

Guy D.R., Bishop S.C., Woolliams J., Brotherstone S., (2009)
Genetic parameters for resistance to Infectious Pancreatic Necrosis in pedigreed Atlantic salmon (*Salmo salar*) post-smolts using a Reduced Animal Model
Aquaculture 290 (2009) 229-235

[Chapter 3]

Conference presentations, as principal author :

Guy D.R., Hamilton A., Pearson M.D. (2005)
Genetic resistance to infectious pancreatic necrosis (IPN) in Landcatch strains of Atlantic salmon, Salmo salar
Aquaculture 247 p18 (2005) presented at
Genetics in Aquaculture VIII, 9-15 November 2003, Puerto Varas, Chile

Guy D.R., Hamilton A. (2005)
Genetic parameters for Harvest and Processing Traits in Landcatch Strains of Atlantic Salmon Salmo salar
Aquaculture vol 247 p17 (2005) presented at :
Genetics in Aquaculture VIII, 9-15 November 2003, Puerto Varas, Chile

Other relevant papers discussed having DRG as co-author are listed in the bibliography with **Guy D.R.** indicated in bold.

Thesis Abstract

Infectious Pancreatic Necrosis (IPN), due to infection with the IPN virus (IPNv), continues to cause heavy mortalities and is endemic across the major Atlantic salmon farming regions of the world. Prevalances of *0.3-0.8* or more at the freshwater stage and *0.05* to *0.3* in the seawater phase of production are typical. Partially effective injectable vaccines are available against seawater IPN but biosecurity measures remain the main methods of control. To explore the feasibility of selecting salmon for resistance to IPN, a selective breeding program was initiated in 1996, including a series of field and experimental trials challenging known full-sib families with IPNv. A total of *404,723* fish faced IPNv challenge (*376,541* seawater and *28,182* freshwater) covering 14 years and 17 separate locations across 7 sites.

Mortalities and survivors following IPN challenge were counted by full-sib family and analysed as binomial data (alive / dead). Initial heritabilities were obtained from expressions based on the variance and covariance of full-sib family means for the 2001 year-group, indicating heritabilities (h^2) of *0.16*, range *0.08* to *0.24*, and genetic correlations (r_g) between replicate families of *0.71* to *0.78*. These results were then confirmed by residual maximum likelihood across all seawater challenged data (year-groups 1997-2003), indicating a h^2 of *0.43* (*s.e.0.02*) across all sites, range *0.06* to *0.40* for individual sites, and a range of r_g between replicates of *0.70* to *0.87* (*s.e. approx 0.05*). To accommodate datasets and pedigrees approaching half a million individually identified fish, an implementation of the Reduced Animal model (RAM) was used to obtain these estimates. A similar level of genetic variation

for resistance to freshwater IPN (year-groups 2005-2009) was confirmed with a h^2 of *0.49*, (*s.e. 0.03*), *range 0.31* to *0.59*, and r_g between replicates *(0.80 to 0.95,* s.e. approx *0.05*), using an Individual Animal Model. When all the data were analysed together, assuming seawater and freshwater survival to be the same trait, the heritability increased to *0.67*, (s.e. *0.02*). On testing this assumption, the genetic correlation between freshwater and seawater survival was found to be 0.68 s.e. 0.09. Both these pooled estimates account better than those for the individual site estimates, for the known selection of superior families that was incorporated at the earliest opportunity (2001) into the selective breeding program. To further investigate if there were favourable or antagonistic relationships operating between traits under active selection, genetic correlations between IPN mortality and a range of performance and harvest traits were obtained. When restricting the harvest data to year-groups where the harvested fish had not experienced an IPN event (2003 for seawater IPN, 2005 for freshwater IPN) : fish length and flesh colour just reached significance with seawater IPN (*0.27* to *0.53. s.e. 0.14*), while only harvest weight *(0.30 s.e 0.11*) attained significance with freshwater IPN mortality. All these correlations were antagonistic. When all the data were combined, (ie both IPN and harvest events taken from all yeargroups) these became non-significant. Taken as a whole, these results indicate that selecting salmon for resistance to both seawater and freshwater IPN challenge certainly is feasible, and that adverse effects on selection for other important production traits is not expected. How these medium to high heritabilities relate to the discovery of a major QTL for IPN resistance segregating in these populations, reported in a parallel scheme of work but based on a sub-set of the same families, is discussed.

Chapter One

General Introduction

1.1 Aim of the study

This thesis details the results of a part-time post-graduate study carried out by DRG over the period 2003 to 2010 whilst employed full-time as a geneticist at Landcatch Natural Selection Ltd, an Atlantic salmon breeding company in Scotland, UK. Through a series of field and experimental challenge trials, the study investigates mortalities from a viral disease of Atlantic salmon (Infectious Pancreatic Necrosis, IPN) in terms of the quantitative genetics of host mortality. Results are discussed in relation to breeding Atlantic salmon for genetic resistance to IPN and in relation to a parallel scheme work which confirmed that a segregating QTL is a major determinant of host resistance to IPN virus infection in these populations.

1.2 Background to the study

Landcatch Natural Selection Ltd (LNS) sells salmon eggs to the world market from its sister company hatchery and rearing facilities at Landcatch Ltd located on the west coast of Argyll, Scotland. Five years prior to this post-graduate study (1998), DRG was brought into Landcatch to establish and manage a selective breeding program based on full pedigree and trait recording for all breeding stock. For 25 years prior to that, the Landcatch

breeding program had successfully relied on mass selection to remain competitive. The selective breeding program was initiated entirely in private company facilities in response to rising epidemics caused by Infectious Pancreatic Necrosis virus (IPNv) across the Norwegian, Scottish and Chilean industries. These epidemics resulted in significant mortalities on commercial farms shortly after the delivery of eggs and juvenile fish to customers. There was no treatment or vaccination available at the time apart from preventative biosecurity measures, consequently the virus has become endemic in the North Sea (Eastern Atlantic seaboard), between Scotland and Norway.

Microsatellite DNA markers for parentage assignment had recently become available (Doyle et al., 1994) which motivated the setting up of a genotyping laboratory at the Landcatch hatchery to support a program of field trials backed up by parentage assignment using the novel markers. Each year thereafter, pedigreed breeding program families were exposed to seawater IPN epidemics on customer's commercial sites. DNA from IPN mortalities was collected and the parentage recovered through in-house genotyping. The strategy was to identify the high survival families resistant to the IPN challenge on the field trials and then to breed from the same families identified in breeding program elite stock (broodstock) held in biosecure facilities at the hatchery. Such field trials can be referred to as 'sentinel' trials and the tested populations as 'sentinels' derived from the concept of animals being specifically located to providing early warning of environmental hazards. This approach has some precedence in being attributed to the success of the acclaimed 'green revolution' in which genetically improved

yields of rice brought Asia out of impending starvation (Borlaug N.E., 1989; Borlaug N.E. and Dowswell C.R., 1995). The sentinel testing was built onto the general framework of breeding program good practice : the sentinel trial populations were sampled for processing traits when harvested, with parentage similarly recovered by genotyping, and performance traits were taken on individual broodstock, identified to family by electronic passive integrated transponders (PIT tags). These were inserted while the fish were reared in separate family tanks in the hatchery 'family unit', built specifically for that purpose in 1998. Over all these considerations, the structure of the breeding program was designed to generate data and pedigrees specifically to allow Individual Animal Model (IAM-BLUP) genetic evaluations, as had by now become best practice across all terrestrial livestock breeding.

The genotyping laboratory and technical genetic staff were moved off-site to Alloa, Scotland, to become a commercially (client) based, spin-off sister company, Landcatch Natural Selection Ltd (LNS) in 2001. By 2003 there were four years of sentinel IPN mortality data with completed parentage genotyping available (year-groups 1997 to 2000). This was sufficient material for this Roslin Institute and Edinburgh University supervised PhD to commence study on the quantitative genetics of resistance to IPN infection. Data continued to be gathered and in 2005 the offer of facilities at the CEFAS government research establishment in Weymouth allowed the focus of the field trials to transfer to experimental freshwater challenge. Six year-groups have currently gone through the freshwater facilities (2005 to 2010) and the first three have passed through the Alloa lab genotyping schedules. This

thesis is primarily concerned therefore with the quantitative genetic analysis of the seawater and freshwater data arising from the breeding program sentinel and associated broodstock performance testing.

Being supervised by the Roslin Institute, further BBSRC funding became available to use and enhance the data and DNA resources generated by the breeding program and this study in particular. A parallel collaborative study investigating genomic markers for QTLs associated with IPN resistance was initiated and remains ongoing. That study has revealed critical insights into the genomic determinants underlying the data described here to the extent that no discussion in this thesis is complete without reference to it. I remain a member of the collaborative team with primary responsibility for choice and supply of family and animal material along with associated data, and a co-author contributor to the papers (Houston et. al.) arising from it. I do not, however, claim any analytical input to that work.

This work has similarly attracted a second collaborative (training) project, completed over the past three years to establish Marker Assisted Selection protocols within the breeding program with part direct government funding and also part supervision from Roslin. I was a supervisory member of that collaboration with some responsibility for data management and in helping embed the analytical techniques into company practice. Whilst I provided a similar co-author contribution to the papers arising from that work (Gheyas et al), I make no claim to any direct analytical input. Work from both these projects are however referenced in this thesis where appropriate to the discussion.

1.3 Salmon life cycle and farming

Without some exposure to it, salmon farming remains unfamiliar to many of those versed in terrestrial livestock production and breeding. The following is an outline of salient points :

Salmon farming is based on stock no more than eight generations from the wild and perforce must follow the natural life cycle of wild salmon. Atlantic salmon are 'anadromous' in that they spend half their life in freshwater, after which they go to sea (as 'smolts'), grow in seawater until sexual maturity, whereupon they return to their natal river to spawn. These changes are determined by internal and external triggers but largely require the fish to be sufficiently well grown to withstand the physiological changes required. The normal cycle is to spawn in October - November, the fry hatch in spring, stay in freshwater for a year until the following spring, whereupon they smolt and 'go to sea' at approximately 80 gm.. They will stay at sea for two winters by which stage they may have attained 10 to 20 kg or more after which they return to freshwater and mate. On returning both males and females stop feeding as they make their way upstream. The females can produce from five to fifteen thousand eggs. In the wild, fish can delay the move to sea for one or two years if growth rate is not sufficient and the adults can return after one, two or three winters in the sea. Farmed fish must follow the same cycle although it is possible to retain smolts to adulthood in freshwater if required. After mating ('stripping' of milt and eggs with external fertilisation) farmed salmon are usually culled, partly for diagnostic

disease testing by tissue biopsy and partly because their condition has naturally deteriorated and they become a focus for disease. Commercial salmon are generated from eggs sold into commercial hatcheries then reared to the smolt stage, or purchased as smolts direct from breeding program suppliers and placed straight into the sea cages, at 60,000 to 100,000 per cage. The fish are harvested after 12-14 months in the sea at about 3-6 kg. Salmon are carnivorous with farmed diets obtained from oceanic trawl fisheries derived from oily fish lower down in the food chain. It is very difficult however to provide the crustaceans that provide the necessary pink pigmentation in sufficient quantities and the right balance and so nature-identical astaxanthin is manufactured and added to the feed, constituting a considerable portion of the rearing costs. Salmon are able to synthesise omega-3 fatty acids from dietary precursors as well as retain them from the diet, which makes it a unique and valuable natural source of long-chain, polyunsaturated fatty acids (LC-PUFA, the correct term for omega-3 fatty acids) in human nutrition.

Salmon farming took off in Norway and Scotland during the 1970's when hatchery and breeding programs were started from wild stock collected from a range of rivers. This closed the life cycle and took advantage of very high survival rates (80-90%) once eggs were incubated in a protected environment. Because individual identification of very large numbers of fish was unfeasible, improvement programs for over twenty years were based on mass selection with limited opportunity to gather performance data other than just prior to stripping. Norway had a government supported national breeding

program until privatised in the 1990s and was able to develop family identification on the basis of fin-clipping (for visual identification) and freeze branding marks, aiding the construction of family pedigrees from an early stage. Atlantic salmon are currently farmed in Norway, Scotland, Ireland, Chile, Canada and Tasmania, being restricted in their range to cold waters to maintain reproductive competence and proper development of gametes.

In common with newly domesticated species, the major challenge after closing the life cycle and providing sufficient nutrition, was the emergence of diseases: pathogens taking advantage of much greater concentrations of hosts in the farmed environment than are found in the wild. The bacterial pathogens which emerged initially were mostly amenable to control, although veterinary treatment of hundreds of thousands of fish in sea cages remains a challenge. Vaccination against the most important bacterial pathogen (Furunculosis) before transfer to sea is an established routine. IPN was the first of several viral pathogens which have proved much harder to control, the main difficulty being to set up controlled challenges on which vaccination efficacy may be quantified. On the other hand by the mid 1990's IPN was sufficiently endemic in the North Sea that some sea-cage sites around Shetland could guarantee an outbreak with significant mortalities within a couple of months of placing smolts to sea. The idea of field testing was born.

1.4 Structure of the thesis

Prior to the commencement of this study it was not known whether genetic resistance to IPN in terms of any family variance in mortality existed in these Atlantic salmon populations. The field and experimental challenge testing was designed to enable any genetic variation that was found to be quantified and compared to the null hypothesis of no genetic variation. Replication of families in the data was used to explore and quantify if possible the extent of genetic x environment interactions, in particular between seawater and freshwater life stages, and whether there was any significant maternal influence on mortality to IPN. Finally, genetic correlations between IPN resistance and other performance and harvest traits undergoing selection were investigated to inform predictions of the consequences of selection for IPN resistance on these other traits.

Section 1.5 describes all the data sets now available and analysed and describes the basic procedures for data collection from the field trials, since this is otherwise fragmented over the chapters that follow. The four main chapters of the thesis, chapters two to five, investigate separate aspects of the data mostly related to the order in which the data became available :

Chapter two investigates the most informative year-group of data to become available at the start of the period of study (the year-group born in 2000 and challenged with IPN in 2002). This data was spread over three seawater sites and was used to explore the genetic and statistical nature of observed family differences in IPN mortality as revealed by the sentinel field

testing. This was written up and published as a paper (Guy et al., 2006) Chapter two comprises the published paper omitting the bibliography which is included in those at the end of this thesis.

Chapter three details the investigation of all the seawater field trials separately and pooled together in the context of 'linear mixed model' statistical analysis suitable for routine genetic evaluation as required by breeding programs. This covers year groups 1998 to 2003. The study was written up and published as a paper (Guy et al., 2009). Chapter three therefore also comprises the published paper.

Chapter four investigates the results and inferences from the first three freshwater experimental challenges (year-groups 2005 to 2007) and explores how they relate to what was learned from the seawater challenges of Chapter three. Genotyping to establish parental assignments has only recently been completed, and the three remaining yeargroups (2008, 2009 and 2010) await their turn in the LNS genotyping laboratory commercial work schedules. Over the period of this section of the study, the existence of QTLs segregating in the families investigated here emerged from the parallel investigations and those results are related to these quantitative genetic results in the discussion. It is intended to publish Chapter four as a journal paper in the near future.

Chapter five completes the analytical investigations by exploring the relationship between IPN resistance and the other performance and harvest traits undergoing routine evaluation by the breeding program. This allows

inferences to be made about including IPN with the other traits in the annual selections of the associated breeding program.

Chapter six reviews the objectives of this thesis and summarises the conclusions that could be drawn from the main body of work detailed in Chapters two to five. The chapter then addresses the commercial application of these results in breeding Atlantic salmon that are genetically resistant to IPN in the context of emerging technologies.

Chapter seven deals with questions left unanswered by the thesis and thus points to further research that may be done.

1.5 Description of test populations

Data consisted of mortality records on families comprising 13 year-groups from 1997 to 2009, represented schematically by Figure 1.1 at the end of this chapter (section 1.8) and listed in Table 1.1.

Year-groups represent year of stripping of the parents (October to December) and fertilisation of families. IPN mortalities were seen in ten year-groups of which eight (1998 to 2008) are investigated here. The challenge populations were constructed at the Landcatch hatchery as mixed families of eggs, parr or smolts totalling from 5,000 to 60,000 fish and placed either on commercial seawater cage sites (year-groups 1997 to 2003) or in experimental freshwater containment facilities (year-groups 2005 to 2009) where they underwent a specific challenge with IPNv. Each population and each cage or tank over which the population was replicated comprised all 200 selected full-sib families designated as breeding program broodstock each year, minus any families previously deselected from the program. Sires were repeated over families so that half-sib groups consisting of one to three full-sib families were formed. All parents were culled on first stripping to allow statutory biopsy sampling for IPN and other pathogens. A combination of pit tagging, farm mating records and genotyping enabled pedigrees to be traced back directly for four generations to 1991-born parents used at the 1996 stripping season, at which parentage assignment by genotyping became available. A further two generations were traced as genetic groups derived from three discrete imports of eggs from Norway in 1983 to 1985 which,

when added to the local Scottish strain already on site in 1982, constituted the four founding genetic groups of the Landcatch breeding program. Selected broodstock generally became parents of the families created 4 years later. From 10 to 15 of the 200 families each year were however created by one or both parents maturing at either three years or five years of age (figure 1.2, section 1.8). Therefore a limited number of families could be represented by brothers and sisters (but not re-bred individuals) becoming parents in adjacent year-groups, allowing some genetic connectedness in the pedigrees across years.

Tail finclips of all parents of broodstock families have been collected since the 1996 stripping season. Approximately 360 selected broodstock parents each year were genotyped in preparation for parentage assignment of any subsequent un-tagged challenged test populations. Where fish were individually PIT tagged and recorded to family as parr, parentage was recovered through the hatchery records as verified by genotyping samples of two to five fry taken from each family whilst reared in separate family tanks in the Landcatch Family Unit freshwater facility. Where family full-sib groups were mixed (e.g. as eggs) before they could be tagged, fin clips were taken during the recording phase of each investigation. Microsatellite DNA analysis using a 10-marker multiplex panel (Chapter 2) was then used to recover parentage and form the pedigrees used in genetic evaluation.

The data consisted of 2,822 families, approximately 200 per year, between 1996 and 2009, descended from four separate importations into the Landcatch hatchery over the period 1982 to 1985. These formed four genetic groups of 'founders' i.e. animals with unknown parentage and assumed to be unrelated at the base of the pedigrees. 119,000 smolts (juvenile salmon transferred to sea) with seawater IPN records and 16,300 fry with freshwater IPN records, comprising mortalities plus survivors were included in the pedigrees. Table 1.1 summarises all the populations for which data was available. Freshwater challenges (2005 to 2009 year-groups) replaced seawater sentinel challenges following four years of low or zero data recovery from seawater sites. At the time of analysis, the 2008 and 2009-strip year-groups tested in freshwater at site *s7* had yet to be genotyped and assigned to families.

Table 1.1 : Prevalence of IPN and number of fish challenged by year group.

Year-group stripped	site	cages (c) or tanks (t)	Year of IPN event	% prevalence by year-group	Numbers stocked
seawater					
1997	s0		1999	22.7	17,006
1998	s1	c1, c2	2000	16.4	54,844
1999	s1	c3	2001	30.0	53,179
2000	s2, s3, s4	c4, c5, c6	2002	10.8	76,786
2001	s4		2003	0	5,000
2002	s2		2004	0	55,000
2003	s5	c7	2005	3.2	54,726
2004	s6		2006	0	60,000
freshwater					
2005	s7	t8, t9	2006	60.4	6,215
2006	s7	t10, t11	2007	40.4	4,846
2007	s7	t12, t13	2008	28.6	5,247
2008	s7	t14, t15	2009	20.7	5,874
2009	s7	t16, t17	2010	21.9	5,450

1.6 General description of seawater challenges

Seawater challenges (1997 to 2004 year-groups) were conducted on post-smolts placed on commercial customer sea sites as is described in detail in Chapters 2 and 3. The following is a general description to allow comparison with experimental testing in freshwater which replaced the seawater challenges from the 2005 year-group onwards. Each seawater site was chosen with a high expectation of significant IPN mortalities (typically 10 % overall between 8 and 12 weeks post transfer) based on previous commercial experience. The aim was to place a known number of smolts from each of the breeding program families with commercial customers. Adipose finclips were collected and counted from mortalities until the natural end of the IPN epidemic. Under or over-stocking of cages was avoided due to the adverse effect on fish performance and importantly to ensure epidemiological relevance. At the termination of the trial, fish were finally graded and absorbed into the production routines of the site.

Up to 60,000 smolts covering all breeding program families from each year-group were provided, depending on the sizes of the commercial cages made available by the customer. Post-transfer mortalities unrelated to IPN in the first couple of weeks typically reduced this number by up to 5 % before mortalities attributable to IPN started to appear at around week eight. Because it could not be assumed that all families were affected equally, mortalities immediately following transfer were also genotyped to adjust the total fish available to be subsequently challenged by IPN in each family, by

subtracting pre-IPN mortalities. By week 12 mortalities had typically returned to background levels without further problems, having reached prevalences of between 3 to 30 percent. Further opportunities for data acquisition were taken in the 2000 and 2001 year-groups, sites *s3* and *s4* (Table 1.1) : 5,000 ungraded smolts that were destined to be recorded for harvest traits had been similarly exposed to an IPN epidemic following transfer to sea, and mortalities, where they occurred, collected. Parentage was initially based on a four-marker genotyping system and the first challenge test to be piloted (1997 year-group, site *s0*) was excluded from subsequent genetic analysis due to insufficient recovery of data and parentage. The 1998 year-group had all families replicated into two separate cages on the same site. Families from the 2000 year-group were replicated over three widely separated sites.

While it was practical to collect and later genotype up to 6,000 mortalities each season, only under certain circumstances was it possible to sample survivors of the IPN epidemic, collect DNA and assign to family. The procedure was as follows: a random selection of five thousand survivors of the disease challenge placed in seawater in each of the consecutive years at site *s1* (1998 and 1999 year-groups) were recorded the following year for harvest traits for which family assignment by genotyping was necessary. Similarly, 5,000 smolts placed on sites *s3* and *s4* (2000 year-group) were fully pit tagged and traced to family through the Landcatch family unit parentage records. Of these, only site *s3* had fin clips additionally available.

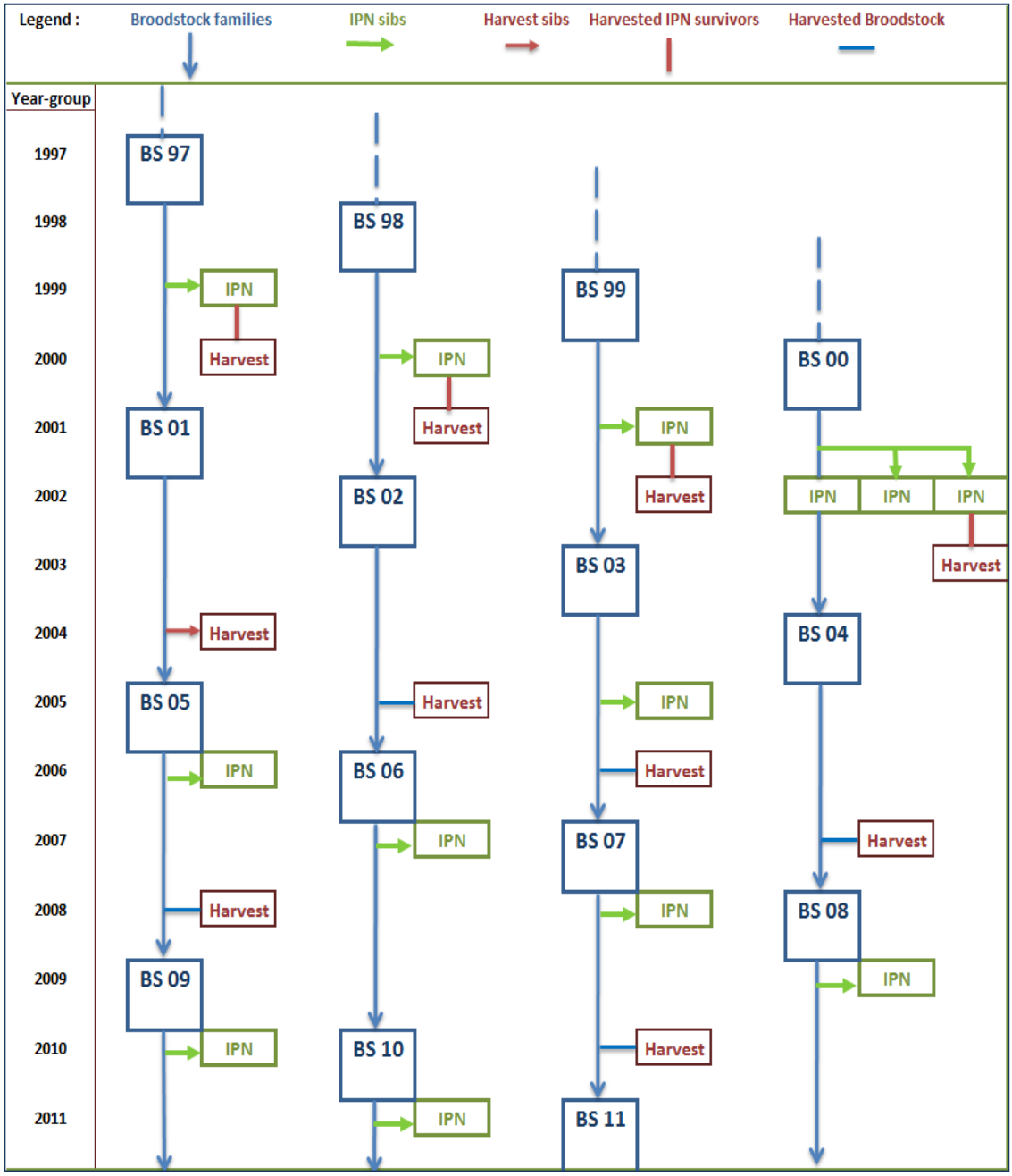
1.7 General description of freshwater challenges

Each year from the 2005 year-group onwards, 30 eggs per family plus allowances for losses at hatching were reserved from the 200 selected broodstock incubators as soon as the final selection had been determined. These were made into two batches of 3,000 'eyed' eggs each (ie viable embryos identified by the presence of black eyes and at a suitable stage to be transported) with all families represented in each batch at a minimum of 15 eggs each. Batches were transported to the CEFAS experimental containment facility at Weymouth. Experimental challenges with IPN virus were conducted simultaneously in two replicate tanks over a period of 6 weeks following hatching. All mortalities and all survivors were individually collected with a record of date and time and tissue samples return to the LNS laboratory for genotyping.

1.8 Schematic of breeding program families, IPN sentinel and harvest sentinel sib groups, by year.

Figure 1 represents a summary schematic of how families are represented by the various yeargroups and how sib groups are created for the IPN and harvest sentinel trials. **Blue** boxes and lines indicate the creation of broodstock (200 families per year), **green** boxes and lines represent IPN challenged sibs and **red** boxes and lines represent harvested sibs.

Figure 1.1 : Breeding program families and creation of IPN and harvest sentinels by year -group.



BS 97 ***: Broodstock (BS) families created at stripping in Nov 1997***

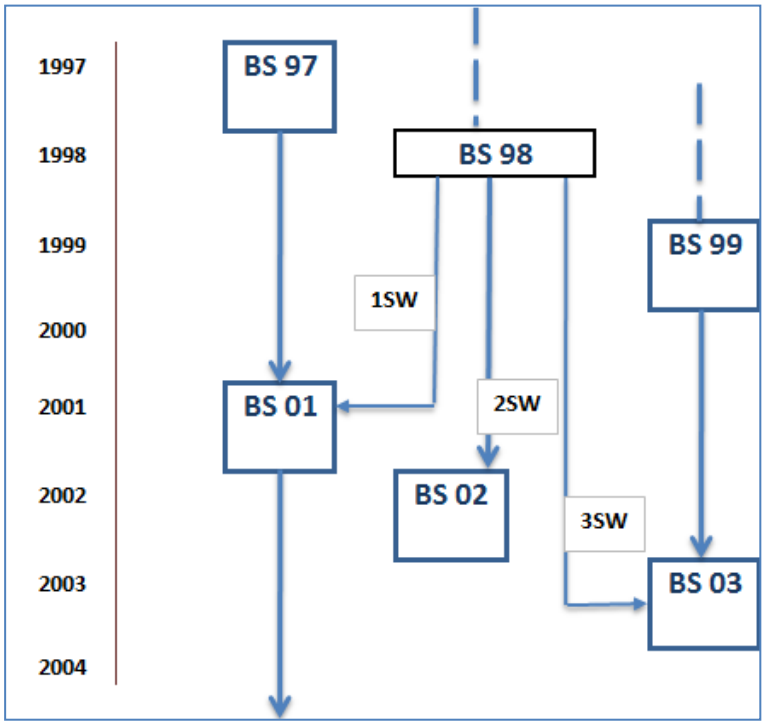
IPN ***: Family sibs challenged with IPN (in seawater up to 2005, freshwater from 2006) in the year indicated by the row label***

Harvest ***: Family sibs harvested in the year indicated.***

Horizontal arrows denote the separate creation of sib groups from broodstock families while lines without arrows denote the same fish are involved (eg survivors of the IPN events), while blue boxes joined by vertical blue arrows denote the creation of the next generation of families from selected parents, forming the 'spine' of the pedigrees.

The broodstock constituting the pedigree spine have four separate tracks, with only a limited number of families being jointly represented in adjacent year groups. These pedigree connections are created by the use of broodstock parents which mature either one year earlier (one sea-winter 1SW), or one year later (3SW) than the typical 2SW 4-year cycle. Figure 1.2 uses the 1998 broodstock year-group (BS98) as an example.

Figure 1.2 : Creation of pedigree connections by using sibs that mature and are mated as parents at 3, 4 or 5 years old.



Chapter Two

Genetic Analysis of Family IPN Mortality

2.1 Abstract to the published paper

A total of 77,124 Atlantic salmon post-smolts, representing 197 full-sib families produced by 149 males and 197 females, experienced a field challenge from Infectious Pancreatic Necrosis Virus (IPNv), following seawater transfer to three separate seawater sites. The first IPN mortality was observed 45 days after transfer, and the duration of the epidemic varied between 37 and 92 days among sites. Mortalities were traced to their parental families by PIT tag records and DNA genotyping. Full-sib family mean prevalence of mortality was calculated for each family on each site. Heritabilities were estimated based on the heterogeneity of $\chi 2$ using prevalence within half-sib families and the variance in prevalence among full-sib families, both on the observed and underlying liability scale. The observed correlation among families across sites was used to estimate genetic correlations. The overall mortality rate was 10.8%, with only small differences between sites, ranging from 10.3 to 11.9%. Heritabilities on the liability scale were found to be moderate to strong, and ranged between 0.24 and 0.81, with a pooled estimate of 0.43, greater than is typically associated with disease traits. Genetic correlations among sites were all substantial,

between 0.71 to 0.78, and indicated that a substantial component of the genetic variation displayed within sites was common to all. The results show that field challenges can yield very good genetic information on family differences in resistance, especially when replicated over sites, which may then be developed for use in selection for breeding strains of Atlantic salmon with greater resistance to IPN.

2.2 Introduction

Infectious pancreatic necrosis virus (IPNv) increasingly causes serious mortality and economic loss in both freshwater and seawater stages of farmed salmonid production particularly in Atlantic salmon, Salmo salar L. (Jarp 1999; Murray, Busby, & Bruno 2003). In Europe, it has become endemic in the North Sea. In freshwater, juveniles are most susceptible before the immune system is fully developed, and hatchery mortality losses in fry around first-feeding can reach 70 % or more (Roberts & Pearson 2005). In seawater, a clearly defined window of susceptibility coincides with the stress of smolting and seawater transfer (Roberts et al 2005), at approximately 15 months post hatching. Mortalities in seawater currently average 10 %, rising annually, but can range from 5-40 % or more depending on site (Roberts et al., 2005). Survivors of an outbreak at any stage acquire a high degree of immunity to clinical disease but may continue to excrete the virus in their faeces for the rest of their lives (Bootland, Dobos, & Stevenson 1991). To protect against the possibility of vertical transmission from infected parents (McLoughlin & Weigall 2002) some regulatory authorities impose costly

biosecurity procedures in commercial hatcheries, to help prevent infection becoming established (McLoughlin et al., 2002). Comprehensive reviews of the disease exist in both scientific (Reno 1999) and industry (FRS 2003; Kjell 2003) literature.

IPNv is a widespread aquatic birnavirus with different serotypes of IPNv broadly coinciding with sampling locations from Europe, the USA or Canada. European serotypes appear to be particularly virulent, with more virulent strains reported to be emerging (Smail, McFarlane, Bruno, & McVicar 1995). Recombinant vaccines against IPNv had been developed (Mikalsen, Torgersen, Alestrom, Hellemann, Koppang, & Rimstad 2004) and these claimed to halve the losses experienced among unvaccinated fish. (McLoughlin et al., 2002). Increased production costs and the lack of a reproducible challenge model (a quantifiable, predictable outcome is required to demonstrate efficacy) however hampered further vaccine development (McLoughlin et al., 2002), although attempts continue (Bowden, Smail, & Ellis 2002; Bowden, Lockhart, Smail, & Ellis 2003). Despite considerable effort, the disease has proved difficult to control either by vaccination, husbandry, or by voluntary and statutory controls (Bruno 2004) over biosecurity and fish movements.

Evidence is emerging of genetic differences in resistance to IPNv within salmon breeding stocks (Midtlyng, Storset, Michel, Slierendrecht, & Okamoto 2002). Differences in mortality among Atlantic salmon fry can vary as much as 80% between families with the greatest and least incidence (Midtlyng et al., 2002), suggestive of a genetic component. Stronger evidence

comes from another member of the salmonid family where a strain of rainbow trout (RT201) appeared resistant to infection after several years of natural selection in the face of regular field outbreaks (Okamoto, Tayama, Kawanobe, Fujiki, Yasuda, & Sano 1993). Subsequent work identified two putative QTLs (Quantitative Trait Loci, chromosomal regions containing genes with large effects) involved with resistance in this population (Ozaki, Sakamoto, Khoo, Nakamura, Coimbra, Akutsu, & Okamoto 2001). Therefore, if lessons from rainbow trout can be extrapolated to salmon, breeding for increased resistance to IPNv has the potential to provide effective and sustainable control of this disease.

A technical challenge in breeding for genetic resistance on the basis of observed mortality, is that, what might be a quantitative trait (e.g. 'resistance') is only scored as two categories (alive or dead) defined by a single threshold determining prevalence. A wide range of non-genetic factors that influence the prevalence of the disease may then affect the position of this threshold. By averaging mortality rates over a family, however, individual non-genetic effects can cancel out. Therefore, large family differences in mortality rates from experimentally induced or natural field outbreaks can indicate a substantial genetic component of disease resistance. (Gjoen, Refstie, Ulla, & Gjerde 1997) This chapter explores this principle, using data obtained by a sentinel test of post-smolts from full-sib Atlantic salmon (Salmo salar) families, replicated over three seawater sites, facing natural outbreaks of IPN.

2.3 Materials and Methods

One hundred and ninety seven full-sib family groups were generated from farmed Atlantic salmon parents spawned in November 2000 and subsequently incubated and reared in freshwater, on the west coast of Scotland. The full sib families derived from 197 females and 149 males, with three males fertilising three females each, 42 males fertilising two females each and 104 males fertilising a single female each. All fish were vaccinated with monovalent furunculosis vaccine (*Furogen II, Novartis*) three months prior to becoming tolerant of seawater (referred to as 'smolting'). Prior to transfer to seawater, non-smolting parr were culled and all freshwater mortalities and culls were collected. Three replicate groups of all 197 families were transferred to three separate seawater sites: two located off the west coast of Scotland, and one in Shetland. Mortalities were collected from each site twice weekly for at least 12 weeks following sea-water transfer. Seawater mortalities due to IPN were identified, dated and counted, and assigned to family by either genotyping *(Site 1)* or recording PIT tags *(Sites 2* and *3)*.

2.3.1 Origin and management of test populations

The three replicate full-sib groups destined for the three seawater sites (*Sites 1, 2 and 3 respectively*), were sampled from a single breeding program population at different times, either from egg incubators, or later as juvenile parr from 197 individual 1-metre diameter tanks, each holding a single full-sib family, located in a 'family unit' in the hatchery. Fish destined for *Site 1* were randomly sampled from these tanks as three separate groups (A, B and

C) of 100 individuals per family at each of the egg (group A), fry (group B), and parr (late juvenile, group C) stages, then combined into a single group (A+B+C) of circa 55,000 individuals (197 x 280 per family) just prior to transfer to sea, in April 2002. The fish were counted on transfer. They were then reared in a single seawater cage, *Site 1,* located in Shetland, until harvested 14 months later. Fish destined for *Sites 2* and *3* were randomly sampled from the individual family tanks in September 2001 at an average weight of 50g and PIT tags were inserted into the abdominal cavity, with the tank and family recorded against each tag. The numbers sampled were 90 and 30 per family for *Sites 2* and *3* respectively. After subsequent rearing in 3.6 metre diameter tanks as mixed family groups, and following vaccination and smolting, fish for *Sites 2* and *3* were transferred to seawater 25 days earlier and 26 days later, respectively, than the fish for *Site 1*. *Sites 2* and *3,* 50 kilometres apart, were located off the west coast of Scotland, 700 km distant from *Site 1*

2.3.2 Procedure for establishing mortality and cause of death

For fish destined for *Site 1,* mortalities were collected, dated and counted, continuously from the point of sampling to the observed end of the IPNv challenge at the seawater site. Early mortalities from groups A and B were counted in total, but assumed random with respect to families given the impracticality of genotyping them. Parr mortalities (group C) and culls from the freshwater tanks of the hatchery were individually counted, and fin clips were taken for DNA parentage analysis to allow for unequal losses from

families prior to the occurrence of IPN mortalities. None of the freshwater mortalities were associated with IPN. On day 38 post transfer, a routine veterinary inspection confirmed absence of IPN mortality. The start of the IPN mortality window was identified by an observable rise in reported mortalities, at approximately 7 weeks post transfer. Diagnosis was based upon finding the external pathology associated with IPN and confirmed by histology. No other causes of mortality were observed. Fin clips were collected from the point of transfer and stopped when mortality prevalence dropped to insignificant levels, six weeks later. During the post-transfer period, approximately 25% of reported mortalities were sampled. DNA was extracted and genotyped to assign those mortalities sampled to families. The fraction sampled varied between weeks, both before and after the onset of IPN.

For fish destined for *Sites 2 and 3*, individual mortalities were dated, from the time of PIT-tagging to several weeks past the end of the seawater IPN mortality window, assigned a cause of death by individual inspection, and their PIT-tags recorded and recovered. The window of IPN mortality was confirmed by veterinary inspection and histology sampling. Almost all mortalities were traced to their full-sib family groups by matching PIT tags to the recorded family of origin. After this each population was reared with very little further mortality.

2.3.3 Genotyping to establish pedigree on Site 1

Assignment to their parental origins by genotyping was done using a single PCR and gel-electrophoresis multiplex of 10 microsatellite markers, run on a 96-lane Applied Biosystems ABI-377 genotyper. The markers were developed from those originally reported by (O'Reilly, Herbinger, & Wright 1998), and (Cairney, Taggart, & Hoyheim 2000; Paterson, Piertney, Knox, Gilbey, & Verspoor 2004), together with an additional five developed in-house. Parentage assignment was carried out using FAP (see Taggart J B, 2007), This achieved greater than 99.8% assignment with fresh DNA samples. Therefore in this study, genotyped but unassigned mortalities can most probably be ascribed to degradation of the DNA prior to sampling.

2.4 Data Analysis

2.4.1 Definition and calculation of traits

Data analysis aimed to describe the pattern of IPN mortality across families, in terms of reported date of mortality *(days to die)* and family percent mortality *(prevalence).* For *Sites 2* and *3*, information on family prevalence was complete and directly available using the pit tag records. Non-IPN mortalities were subtracted from the total mortality for each family to calculate summary prevalence, or were treated as censored in the survival analysis of days to die described below. For *Site 1* however, it was necessary to account for the sampling that was done prior to genotyping. We will be

analysing prevalence (died / total challenged) where the numbers we can count as died are only a sub-sample of those that actually died. It is important to maintain the prevalence collated over families to be the same as that observed, because this affects the heritabilities obtained. Consequently, the family representation of IPN mortality was assumed to be as observed by the sampling, but was treated as if it were the total mortality that had occurred in a smaller subpopulation. This approach better represented the sampling uncertainties of the observations. The overall prevalence in the sub-population was assumed to have been the same as the full population in *Site 1*. For example, if the total overall prevalence was $\boldsymbol{p}$ in a population of size $\boldsymbol{N}$, and a fraction $\boldsymbol{q}^*$ of the mortality was sampled, then the size of the sub-population assumed was given by $\boldsymbol{N}^* = \boldsymbol{q}^*\boldsymbol{N}$. This value of $\boldsymbol{N}^*$ was used in the heritability calculations for *Site 1* described below.

2.4.2 Within and between site sampling properties

The observed pattern of days to die was described by non-parametric Kaplan-Meier survival curves, which express the rate of mortality over time as a proportion of the number surviving to each time point. The survival curves were compared across sites and across families.

If mortality was random with respect to families, then the family prevalence should follow a binomial distribution. Conversely, the extent of departure from a binomial distribution indicates the degree of genetic control, in the absence of family-related common environment effects. The full-sib family prevalence data for each site was compared with its binomial sampling

properties, as follows: if there is genetic variation for survival, relatives will perform more alike than non-relatives and the covariance between full-sibs ***i*** and ***j*, $\mathbf{cov}_{ij}$** will be greater than zero by half the additive variance (in the absence of dominance). The variance within families will be observed to decrease, and variance between families increase, compared to random sampling from a binomial distribution.

Heritability ($\boldsymbol{h^2}$) was inferred from the variance and covariance of family means in two different ways:

(i) by equating the 'heterogeneity chi-squared' of sire (half-sib) family means to its expectation, following (Robertson & Lerner 1949)**;**

(ii) directly describing the expectation of the full-sib family variance in terms of the additive genetic variance.

The between–site properties were explored using genetic correlations, derived from the correlation of full-sib family means across paired sites. Full details of the formulae and their derivations are given in the appendix (2.8) for reference. The methods estimate heritabilities $\boldsymbol{h^2_p}$ on the (0,1) observed binary scale, with 0 and 1 indicating survival and death from IPN. These are sensitive to the observed prevalence and so were converted to an assumed underlying liability scale, again following Robertson and Lerner (1949). This involves calculating $\boldsymbol{h^2_u = h^2_p (pq/z^2)}$ where $\boldsymbol{p}$ is the observed prevalence, $\boldsymbol{q = 1-p}$, and $\boldsymbol{z}$ is the ordinate of the Normal distribution truncated at $\boldsymbol{p}$.

2.5 Results

2.5.1 Description of response to IPNv challenge

During the period 1996-2000, differential Enzyme-linked immunosorbent assay (ELISA) serotyping of freshwater and seawater isolates from the Scottish mainland and islands, confirmed the virulent strain of IPNv to be *Sp A2* (Smail, Bain, Bruno, King, Thompson, Pendrey, Morrice, & Cunningham 2006). Table 2.1 compares the time course of the IPN epidemic over the three sites. There was a 51 day variation in the calendar dates of sea transfer, but the range of onset following sea transfer was only 39 days, although the ranking of dates remained the same. The time from transfer to the commencement of the epidemic varied by just 17 days, ranging from 45 to 62 days, but the duration of the epidemics showed much greater variation from 37 to 92 days.

Table 2.1: The timing and duration of IPN epidemics according to site.

Location	Date of transfer to sea	Date of IPN onset	Period (days) from transfer to:	
			Initial IPN	Last IPN
Site 1	12 Apr 02	27 May 02	45	102
Site 2	20 Mar 02	21 May 02	62	154
Site 3	10 May 02	29 Jun 02	50	87

A total of 77,124 fish were transferred to sea, and Table 2.2 shows the size of the population at each site. *Site 1* had approximately three-fold more fish exposed to IPNv than *Site 2* and there was a similar difference between *Sites*

2 and *3*. Despite the wide geographical separation and difference in the size of population exposed, the prevalence of IPN was very similar on all three sites with an overall IPN prevalence of 10.8%. Overall, 4.7% fish died prior to the onset of IPN, with the greatest proportional loss occurring at *Site 1* (6.4% of those transferred). Of the IPN mortalities, all were allocated to families at *Sites 2* and *3*, whilst only 22.6% (1216 fish) were allocated to families at *Site 1*. Consequently more IPN deaths were allocated to families at *Site 2* than at *Site 1*, despite its smaller population size.

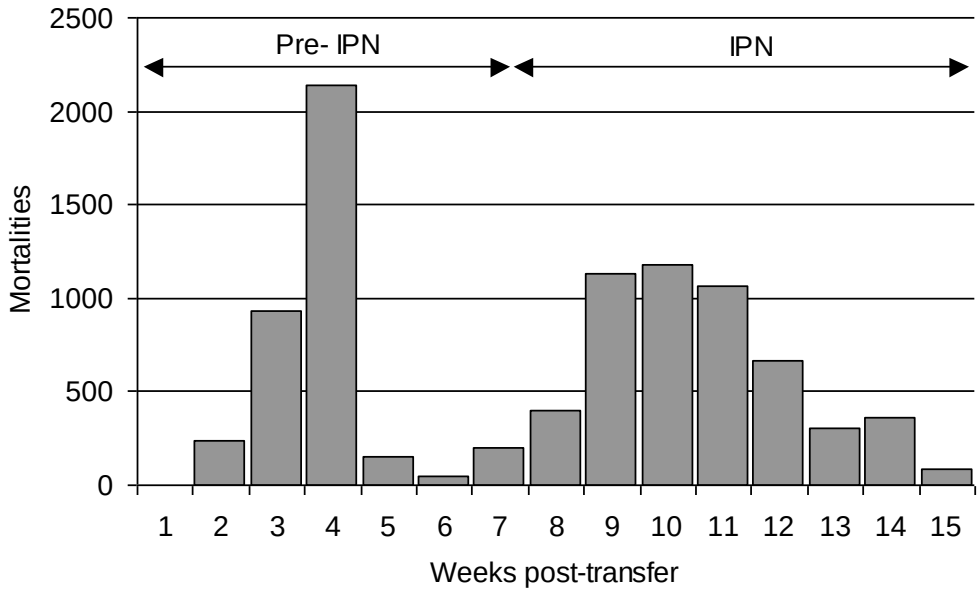
Table 2.2 Population size and reported IPN mortalities for each site and over all sites

	Transferred to sea	Pre-IPN mortality	Alive at start of IPN onset	IPN Mortality	Prevalence %
Site 1	55,007	3,507	51,500	5,384	10.5
Site 2	16,207	109	16,098	1,918	11.9
Site 3	5,910	43	5,867	605	10.3
Total	**77,124**	**3,659**	**73,465**	**7,907**	**10.8**

The weekly pattern of mortality is shown in figure 2.1 for *Site 1*. This shows an initial peak in mortalities unrelated to IPN, typically consisting of undersized, incompletely smolted parr and those otherwise stressed by the transfer process. This continued for four weeks post-transfer and was followed by a period up to week 7 when very little mortality was observed.

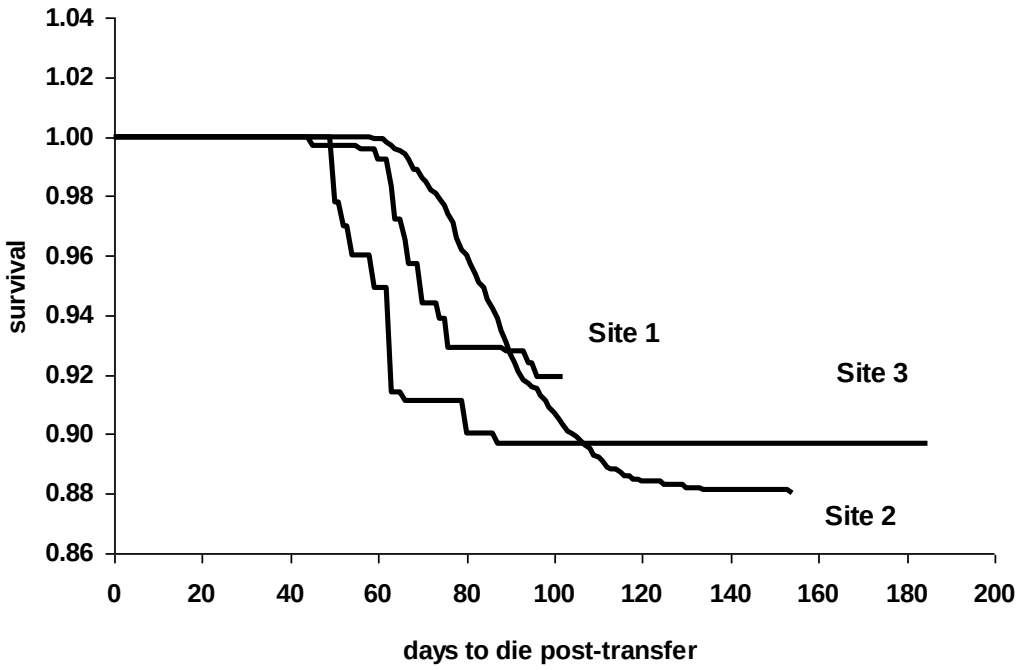
The mortalities up to week 7 make up the pre-IPN mortality reported in Table 2.2. The onset of IPN was signalled by rising mortalities following week 7, reaching a peak at week 10 and falling to background levels close to zero by week 15. This pattern is similar to those in *Sites 2* and *3* and is very typical of seawater IPN outbreaks in post-smolt Atlantic salmon.

Figure 2.1 Weekly reported mortalities from Site 1.



The Kaplan-Meier survival curves (Figure 2.2) show the pattern of IPN survival in terms of *days to die* for each of the three sites. These curves contrast the sudden start and end to the IPN epidemic experienced at *Sites 1* and *3*, with the more protracted event seen at *Site 2*.

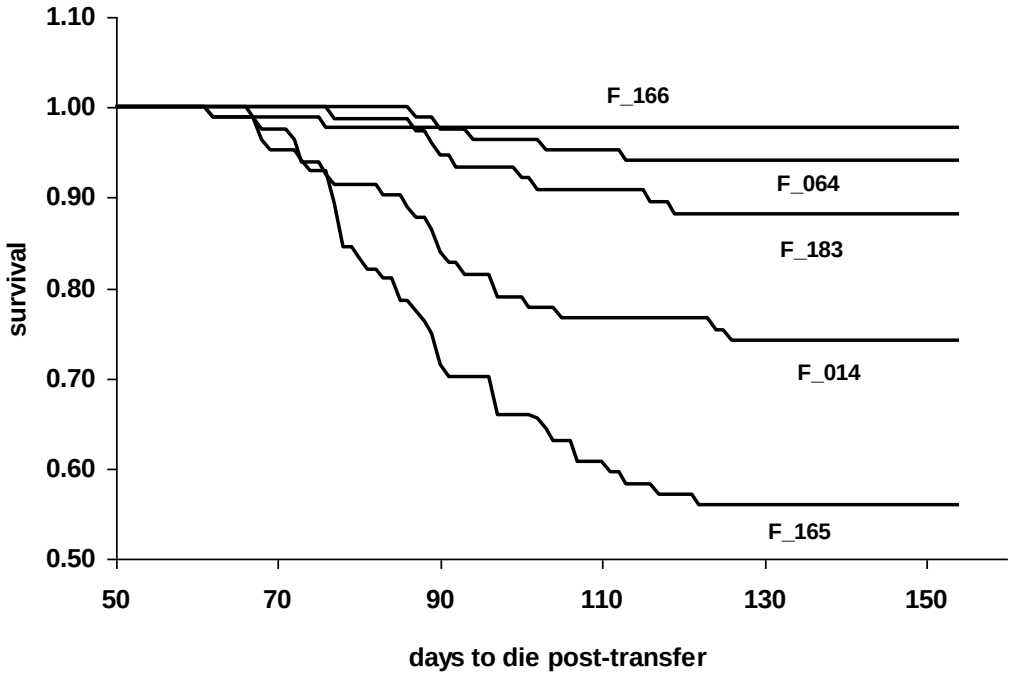
Figure 2.2 Kaplan-Meier survival curves for IPN infection by site.



2.5.2 Family variation in IPN mortality

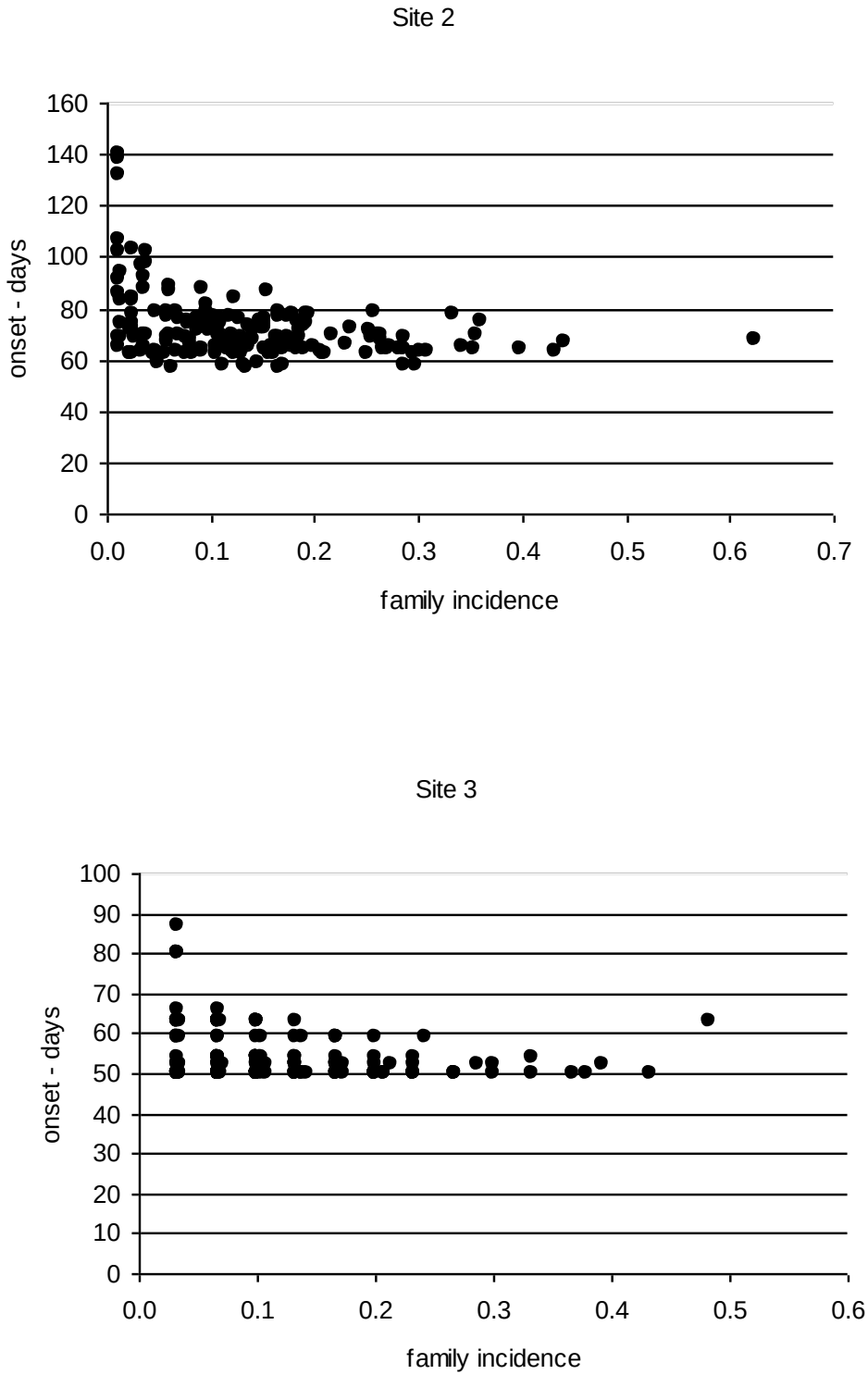
Five families representing the range of family prevalences from *Site 2* are plotted as Kaplan-Meier survival curves in Fig. 2.3. There were clear differences in the time of onset (first observation of IPN in that family) and mortality rate for IPN over the period of the epidemic.

Figure 2.3 ***Kaplan-Meier survival curves from IPN infection for five representative families in Site 2, demonstrating the extent of observed variation.***



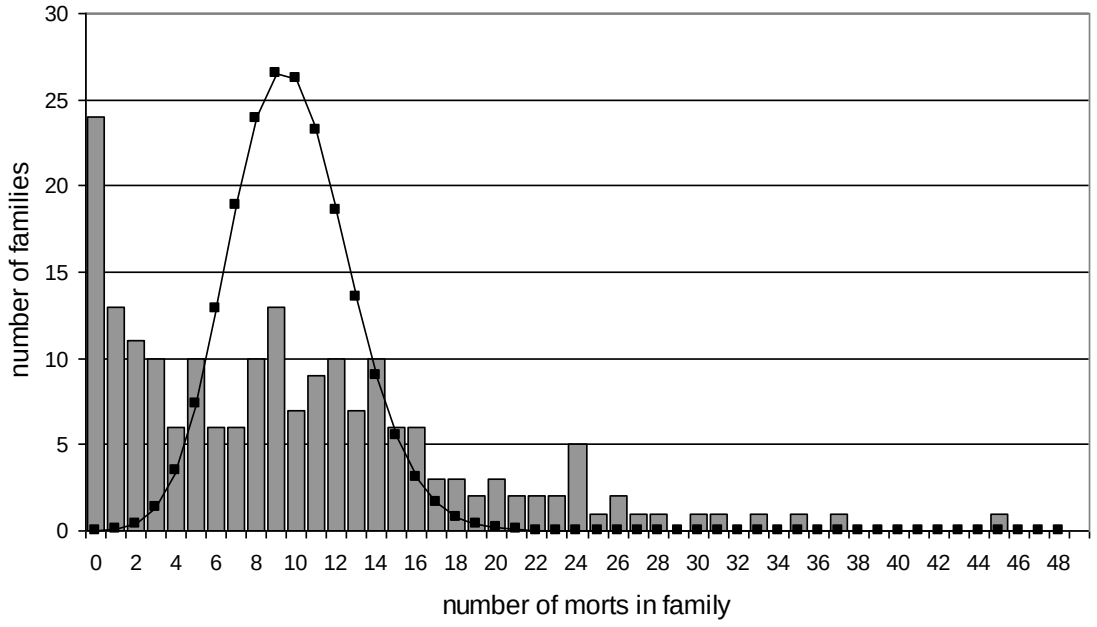
The time of onset (first observed IPN mortality) was only weakly related to the overall prevalence displayed by each family (Figure 2.4). The correlation ***r*** is -0.22 and -0.36 for *Sites 2* and *3* respectively ($P<0.05$, 195 d.f.). *Site 1* was consistent with the other sites (***r*** = -0.22) but is not shown because of the low and variable proportion of mortalities that were sampled and traced to families each week. Although families with relatively late onset had lower overall prevalence, the converse was not true. Whilst families with low prevalence can show both early and late onset, there were no examples of families with high prevalence and late onset.

Figure 2.4 Relationship between onset of IPN mortalities and family prevalence, plotted for each family, for Sites 2 and 3.



The distribution of observed IPN mortality by family is shown as a histogram in Fig. 2.5 for *Site 2*, overlaid by a line denoting that expected from a binomial distribution with the same prevalence and family size prior to exposure but assuming no family influences. The existence of families with prevalences outside the range of values expected from binomial sampling can be readily observed. The over-dispersion seen in both extended upper and lower tails is typical over all sites and in the absence of family-specific environmental effects, is direct evidence for genetic differences between families.

Figure 2.5 ***A comparison of profiles for family prevalence of IPN mortality for Site 2. The bars show the observed distribution as a histogram, whilst the line shows that expected from binomial sampling assuming no variation in mortality prevalence among families.***



2.5.3 Inferences on heritabilities and genetic correlations

For calculations of heritabilities in *Site 1*, to account for the incomplete genotyping (see Materials and Methods for details) the genotyped mortalities were assumed to come from a sub-population of size 11,581 with the same prevalence (10.5%) that was observed in the full population, with 58.8 offspring per full-sib family.

The appendix (2.8) to this chapter applies the heterogeneity chi-square formula of (Robertson et al 1949), to data for *Site 2* as an example :

With s = 149 sire groups, Ns = 108.04 offspring per sire, Nd = 81.72 offspring per dam, and an average prevalence of 11.9 %, the estimated heritability on the observed (0,1) scale is h^2_p = 0.159. Adjustment to the underlying normally distributed liability scale using pq / z^2 = 2.65, leads to h^2_u = 0.42 on the liability scale. The same calculation for *Sites 1* and *3* gives h^2_u = 0.81 and 0.24 respectively.

Table 2.3 compares results for each site, calculated by the 'heterogeneity chi-squared' method (appendix *2.8.1*), and the full-sib family variance V_f method (appendix *2.8.2*). The h^2_u estimates from the different methods are in close agreement for each site. The estimates vary, however, from 0.24 to 0.81 among the sites. Because the sites are similar in prevalence, it was possible to derive an approximate pooled estimate. This was achieved by concatenating all observed data across the three sites before calculating the observed prevalence for each family. Using Appendix *2.8.2*, the

heritability was 0.16 on the observed (0,1) scale, and 0.43 when transformed to the underlying liability scale.

Table 2.3 ***Estimates of heritability for IPN mortality on the observed and underlying liability scale, from using heterogeneity χ^2 among half-sib families (appendix 2.8.1) and variance of full sib families (appendix 2.8.2).***

	Site 1	*Site 2*	*Site 3*
Average full-sib family size	58.8[#]	81.7	29.8
Prevalence %	10.5	11.9	10.3
Family variance, V_f	0.0128	0.0109	0.0079
h^2_p het. χ^2,observed	0.28	0.16	0.08
h^2_p var FS families, observed	0.24	0.19	0.11
h^2_u het. χ^2, liability	0.81	0.42	0.24
h^2_u var FS families, liability	0.69	0.49	0.31

[#] For *site 1* the average full sib-family size was reduced to account for the incomplete sampling of mortality as described in Materials and Methods.

The appropriateness of the pooling across sites will depend on both the similarity of prevalence and the genetic correlations among the sites. Correlations of family prevalence between sites, r_{12}, and the derived genetic correlations from using appendix 2.8.3, rg_{12}, are listed in Table 2.4.

Table 2.4 ***Observed across-sites correlations r_{ij} of family percent IPN mortality, upper right triangle with associated genetic correlations r_{gij} ,lower left triangle, calculated as described in appendix 2.8.3).***

	Site 1	***Site 2***	***Site 3***
Site 1	-	0.62	0.53
Site 2	0.71	-	0.59
Site 3	0.73	0.78	-

The correlations demonstrate a consistent ranking of families over the three sites, with the genetic correlations being remarkably similar, ranging from 0.71 to 0.78. Note that the genetic correlations are always higher than the observed correlations between families and are a function of the within-site heritabilities (appendix 2.8.3).

2.6 Discussion

As far as we are aware, this is the first report of quantified mortality differences between families following seawater exposure to IPNv, and the first study to show the correlations of such observations across sites. The results demonstrate a mortality response pattern to IPNv challenge comparable to those currently being experienced in the industry (Roberts et al

2005). There was no evidence of exposure to IPNv in the freshwater phase of rearing, and it can be assumed that exposure was immediate and continuous on transfer to seawater at each of the three sites. The prevalence of mortality from IPN across three replicates with the same genetic background was consistent at approximately 10 %, with a period from first exposure to first mortality ranging from 45 to 62 days, a difference of just 17 days across sites. This period between presumed exposure and death contrasts sharply with freshwater salmonid IPNv infection, where susceptible fish can start to shed infective virus within hours of exposure and mortalities can start after just ten days incubation (Bebak, McAllister, & Smith 1998), and where final mortalities can reach 90 % or more; (Bebak et al., 1998; Roberts et al., 2005) A defined window, before and after which no pathology from IPN could be observed, was also consistent over the three replicate seawater populations. These similarities are evident despite the large differences in biomass during exposure, and the wide geographic and temporal separation of the replicate sites, all aspects that could reasonably be expected to affect the epidemiology and progression of the disease.

The Kaplan-Meier curves contrast the sudden start and end to the IPN epidemic experienced in the sea-water cages (*Sites 1* and *3*), with the more protracted event seen in the on-shore seawater tanks of *Site 2*. This may be a reflection of the more protected conditions, increased management and more precise recording applied to the on-shore site, allowing the non-feeding fish afflicted with IPN to survive longer. No direct linear relationship could be seen between the period of onset and overall prevalence for each family.

Therefore sampling families only from the initial stages of the epidemic may give a poor indication of prevalence over the full period of mortality. Late onset families were however observed to have reduced total mortalities. While this could be simply related to the finite period over which the epidemic develops and ends, it does suggest selection from late onset families may be a strategy that assists in reducing overall prevalence. The results here suggest that specific factors are operating to prevent mortality resulting from exposure prior to and following the mortality window in addition to the life-time natural immunity possessed by all which survive the IPNv challenge at any stage of their lifecycle. One aspect that is co-incident with the falling mortalities is the rise in ambient seawater temperature. (McLoughlin et al., 2002). Whilst there is little opportunity to manipulate the timing of seawater transfer given the narrow window available to the fish for smolting, this aspect may be worthy of further investigation.

Previous literature reports are anecdotal but the range of percent mortality from IPN across families found in those studies is comparable to these reported here (Midtlyng et al., 2002). All analyses in this study show the heritability for IPN mortality to be moderate to strong. Heritability estimates derived from the chi-square and variance of family mean prevalence closely agree because they use the observed variation between families in similar ways. Chi-square uses the ratio of between family sums of squares to that expected if there were no family differences, (appendix 2.8.1), while the variance of family means uses the ratios of family and phenotypic variances directly, (appendix 2.8.2). Epidemiological differences between the

sites (eg cage or tank, or geographical location) are expected to influence the host response to pathogen challenge and subsequent development of the epidemic. These may appear as a scale effect on the observed family variances and heritabilities derived from them, despite the overall pattern of disease progression being very similar. The pattern of genetic correlations between sites is more consistent than the pattern of heritabilities observed. Whilst these correlations independently suggest a genetic component to resistance, they also suggest that the genetic response in the host to the pathogen challenge is broadly similar on each site. In general, the family variance estimates are sensitive to anything that directly affects the calculated family dispersion, including the efficiency of sampling and family assignment from genotyping.

An important conclusion from this work is that it is possible to use epidemiological data for drawing genetic inferences. Where the genetic structure of the host population is replicated across sites, interpretation of the family variance in mean prevalence can lead to deeper insight into the epidemiology of the pathogen. For example, the higher heritability observed from the cages of *Site 1* may reflect a real difference in the expression of host genetic resistance under severe epidemiological and environmental challenge, compared to more controlled environments experienced at *Sites 2* and *3*.

While the mechanisms of pathogen virulence and host resistance are expected to be complex, these results indicate that selection for resistance is likely to be successful even if the underlying mechanisms of resistance are

unknown. The testing on which selection is based aims to allow all relevant mechanisms of resistance to be expressed so that the selection process can target a balanced range of those most effective. In freshwater, natural field outbreaks of IPN are invariably managed to minimise their impact, leading to uncertain and differential patterns of challenge within the facility, although opportunities for balanced testing of families can be taken if they arise. Effective protocols exist for experimental IPNv challenge of early fry stages (Midtlyng et al., 2002). These, however, demand expensive containment facilities in order to be run on the regular and repeatable basis required by breeding schemes. Estimates of correlations as high as 0.95 have been reported (Gjoen et al 1997) between experimental and field mortalities for furunculosis (a bacterial disease), while (Midtlyng et al., 2002) reported 0.56 for freshwater IPN (although source data were unpublished). This value is very similar to that between replicates over seawater sites reported here, and confirms the value of replication where it is at all possible.

The relevance of freshwater challenge to seawater challenge in post-smolts, and vice-versa remains to be demonstrated in relation to the epidemiological differences and host resistance mechanisms discussed above. In contrast, a large but epidemiologically relevant field test under commercial conditions (a 'sentinel' test) is potentially more informative than a smaller scale, but more closely controlled, experimental challenge, if the extra demands of recovering data and pedigrees can be met. Breeding from survivors of a commercial challenge is generally regarded as an unacceptable risk due to their likely carrier status and the consequent risk of introducing

IPNv into the freshwater facilities of a hatchery. If, however, pedigreed populations of fish are exposed to the field challenge, with unchallenged full-sibs remaining in biosecure conditions, it is possible to make informed selection decisions for breeding broodstock that themselves remain unchallenged. Finally, an epidemiologically relevant field test can be expected to target the widest range of genetic resistance mechanisms available in the population. This may lead to the development of strains resistant to a wider range of pathogens when compared to more direct experimental and molecular approaches.

2.8 *Appendix. Derivation of heritability from variance and covariance of family mean prevalence.*

A fish is coded 0 or 1, according to whether it survives the IPNv challenge or not, respectively. Define the following terms as :

$\boldsymbol{p}$ = overall prevalence, probability of mortality, with $\boldsymbol{q = 1-p}$

$\boldsymbol{V_g}$ = additive genetic variance on the observed (0,1) scale

$\boldsymbol{V_p}$ = phenotypic variance of individual fish on the observed (0,1) scale = $\boldsymbol{pq}$

$\boldsymbol{V_f}$ = variance of family mean prevalence, each of size n

$\boldsymbol{h^2_p} = \boldsymbol{V_g / (pq)}$ = heritability on (0,1) scale,

$\boldsymbol{h^2_u}$ = heritability transformed to the underlying liability scale.

$\boldsymbol{r_{ij}}$ = observed correlation between full-sib family means on sites ***i*** and ***j***

$\boldsymbol{r_{gij}}$ = genetic correlation between full-sib family means on sites ***i*** and ***j***

2.8.1 Expected value of heterogeneity χ^2 for half sib families

Genetic (or environmental) differences among families creates an excess of observed variation in family prevalence over that expected from the random binomial expectation, and this excess can be tested using the 'heterogeneity' χ^2-distribution (Robertson et al 1949). The test compares a random variable $\boldsymbol{Z}$ distributed as χ^2 , with the associated degrees of freedom, since if $\boldsymbol{Z}$ is distributed χ^2_r then $\boldsymbol{E[Z] = r}$. Following (Snedecor & Cochran

1989) , when there is no genetic or environmental variation then the random variable ***Z*** given by :

$Z = \Sigma_i \; n_i \, (f_i - p)^2 / (pq)$ is asymptotically distributed as χ^2_{s-1}, where n_i and f_i are the family size and proportion of mortalities respectively for the offspring of sire ***i***, with summation over all ***s*** sire families. ***Z*** is the ratio of the observed between-family sums of squares to the expected variance within families assuming no family variance.

The heritability (ratio of additive genetic variance to the phenotypic variance) on the observed (0,1) prevalence scale was derived following (Robertson et al., 1949), page 399 by:

$$h^2_p = (Z - (s-1)) / (rn_0)$$

where ***s*** = number of sire groups, ***r*** = average genetic relationship within sire families, and n_0 = an adjusted total to account for variation in sire group size. For hierarchical crossing schemes yielding mixtures of full and half sibs (Robertson et al., 1949) can be extended to yield:

$$r = 0.25 \, (1 + (N_d - 1) / (N_s - 1))$$

where N_d = average number of offspring per dam , N_s = average number of offspring per sire and

$$n_0 = N - (\Sigma_i \, n_i^2 / (N - (s-1))) ,$$

an adjusted N to account for variable family size, and where

$\Sigma_i \, n_i^2$ = squared number of offspring for sire ***i*** summed over all ***s*** sires, and ***N*** = total population size.

Taking data for *Site 2* as an example, with s =149 sire groups, N_s =108.04 offspring per sire, and N_d = 81.72 offspring per dam, gives Z =1249 for 148 d.f., $\Sigma_i n_i$ =16,098 post-smolts challenged, p =0.1191, z = 0.199, n_0 =15,827 and r =0.439. The estimated heritability on the observed (0,1) scale is

$$h^2_p = [Z - (s - 1)] / rn_0 = 0.159.$$

Adjustment to the underlying normally distributed liability scale using pq / z^2 =2.65, leads to h^2_u = 0.42 on the liability scale. The same calculation for *Sites 1* and *3* gives h^2_u = 0.81 and 0.24 respectively.

2.8.2 Heritability from variance of full-sib family means

Let f_i be the mean prevalence of full sib family i:

$$f_i = (X_{i1} + X_{i2} + \dots X_{in}) / n$$

where $X_{ij} = 0$ or 1 according to whether individual j from family i survived or died.

$$V_f = cov[(X_{i1} + X_{i2} + \dots X_{in})/ n, (X_{i1} + X_{i2} + \dots X_{in}) / n]$$

$$= [n.cov(X_{ij}, X_{ij}) + (n^2 - n) cov(X_{ij}, X_{jk})_{j \neq k}] / n^2$$

For a full sib family structure $cov (X_{ij}, X_{ik}) = \frac{1}{2}V_g$ and $cov (X_{ij}, X_{ij}) = V_p$ then, after dividing top and bottom by n

$$V_f = V_p [1+0.5h^2(n-1)] /n$$

Thus, V_f approaches 0.5 V_g for large n.

Therefore, after substituting $V_p = pq$ and rearranging terms,

$$h^2_p = 2[nV_f - pq] / [(n-1)pq]$$

2.8.3 Genetic correlation between family mean prevalence over two sites.

The correlation between family means across sites, say f_1 and f_2, is

$$r_{12} = cov(f_1, f_2) / (V_{f1} . V_{f2})^{1/2}.$$

V_{f1} and V_{f2} are estimated, as described in Appendix 2.8.2.

With full sibs on two different sites with no environmental covariance, $cov(f_1,f_2)$ is ½ the genetic covariance, or $½ r_{g12} (h^2_1 h^2_2 V_{p1} V_{p2})^{1/2}$ so that

$$r_{12} = ½ r_{g12} (h^2_1 h^2_2 V_{p1} V_{p2})^{1/2} / (V_{f1} V_{f2})^{1/2}.$$

All terms have been calculated above for each site so r_{g12} can be calculated directly.

Alternatively squaring both sides and substituting

$V_{f1}/V_{p1} = [1+0.5h^2_1(n_1-1)] / n_1$ as derived in Appendix 2.8.2 for each site we obtain:

$$rg_{12}^2 = 4 r_{12}^2 [1+0.5h^2_1(n_1-1)] [1+0.5h^2_2(n_2-1)] / (n_1 n_2 h^2_1 h^2_2)$$

and taking the square root of this yields the genetic correlation.

Chapter Three

Heritabilities and Genetic Correlations from a Reduced Animal Model

3.1 Abstract to the published paper

Infectious Pancreatic Necrosis Virus (IPNv) is an important cause of mortality and economic loss across all species of commercially farmed salmonids, and genetic variation in survival to IPN challenge has been previously demonstrated. In order to exploit this variation in the development of resistant strains, robust procedures are required to quantify the extent of genetic variation and to provide estimated breeding values used to select candidates for breeding. This paper applies a recently developed implementation of the Reduced-Animal Mixed-model procedure (RAM) to field data describing percent mortality following IPN epidemics in Scottish farmed Atlantic salmon, covering 1369 full-sib family groups distributed over four years and a total of seven sites. Pedigrees were established through a combination of electronic (PIT) tagging and parentage assignment using microsatellite DNA analysis. Heritabilities between 0.07 and 0.56 (standard errors below 0.04) were obtained, genetic correlations between sites sharing the same families were uniformly high, 0.70 to 0.85, (standard errors below 0.06) and low levels of full sib family effect due to common environment

(proportion of phenotypic variance 0.04, s.e. 0.002) were observed. These results confirmed that exploitable genetic variation exists for mortality to IPNv over a range of epidemiological conditions inherent in field data, that can be used to select strains of salmon with increased resistance to IPNv.

3.2 Introduction

Mortality from Infectious Pancreatic Necrosis (IPN) is an important cause of economic loss across all species of commercially farmed salmonids. In anadromous species mortalities can occur at both the juvenile post-hatch freshwater stage (typically 30-80 %) and the early post-smolt seawater stage (5-30 %), both stages being when the fish are particularly vulnerable, both immunologically and physiologically (Roberts et al., 2005). A significant genetic component to mortality from the IPN virus (IPNv) has been clearly demonstrated from experimental and field challenges of pedigreed Atlantic salmon *Salmo salar* L. populations, in both freshwater stage fry (Wetten et al., 2007; Kjoglum et al., 2008) and in seawater post-smolts (Chapter 2; Guy et al., 2006). Heritabilities of 0.11 to 0.24 for seawater mortality to IPN on the observed binary scale, and 0.31 to 0.69 on the underlying liability scale, were demonstrated (Chapter 2) by Guy et al., (2006) using simple expressions derived from the variance and covariance of family mean prevalence.

Given a significant heritability, and a population measured on one site, a ranking of family means is sufficient to apply a selection differential and to generate genetic gain for decreased mortality from disease. Breeding

from survivors is, however, undesirable where biosecurity may be compromised through a risk of viral transmission, as is the case with IPN. In commercial breeding programs, therefore, the information on disease is more likely to come from one or more test sites, possibly with differing levels of infection and testing sibs of candidates with only a subset of the families in common. These data then require to be linked back to the breeding population, which are maintained separately with the intention of remaining unexposed. Additionally, information on disease requires to be properly weighted in combination with other traits. This process is usually addressed through application of multi-trait mixed models, in particular parameterised using the Individual Animal Model, (IAM; (see Henderson, 1986, and Mrode, 2005) to estimate variance components and predict breeding values (EBVs). The IAM provides genetic evaluations for any animal that is linked via the pedigree to any other animal having a data record. This is especially appropriate when breeding for disease resistance where the animals providing data from disease challenge (in aquaculture, these are usually mortalities) are different to candidates for breeding.

Mixed model methodology allows for complex genetic and environmental modelling, for example: estimates of direct additive genetic effects, dam maternal genetic, dam environmental effects, and effects due to common environment where family members are reared together. While maternal effects may seem less relevant to species which do not gestate or rear their young and which invest little compared to mammals in parenting their offspring, the female fish parent does invest resources in the egg that

may affect survival and initial growth, which may then persist to affect performance measurements taken later in life. Moreover, broodstock families are often reared in separate 'family' tanks until the fry are large enough to have electronic tags inserted at around 20 g weight and up to six months of age. Rearing families separately may induce a common environment effect on growth records (e.g. Winkelman et al., 1994) which can inflate estimates of genetic, or maternal effects if the data does not enable them to be independently estimated.

With binary mortality data, recorded on the scale of observation as zero (alive) and one (dead), family information based on data from individual animals can be summarised without loss of information by the full-sib sample mean prevalence p within year and location. Summarised family information includes the sample variance, equal to $p(1-p)$ for binary traits, allowing suitable variance partitioning, the basis of mixed models. An implementation of mixed models and the IAM in particular, based on family mean prevalence may therefore be more computationally efficient in some cases and has attracted some attention in the past (Simianer et al. 1991). One such implementation has recently been described by White et al., (2006) based on a revival of the Reduced Animal Model (RAM) of Quaas et al., (1980). RAM separates the data into two layers, parents and non-parents. Only the parental layer is represented in the pedigree relationship matrix used for analysis and data on non-parents is accumulated into the parental layer, from which the variance components and associated estimated breeding values are derived, without directly processing the non-parents. If required, estimates for non-

parents can be obtained indirectly using simple linear functions (backsolving). The 'absorbing' of non-parental data also affords the possibility of presenting that data as family mean prevalence weighted by sample size rather than by lists of individual animals, since the variance structures are known.

We have applied the RAM method of White et al., (2006), to mortality data obtained from natural seawater field challenges of Atlantic salmon post-smolt families to IPNv. This paper formalises the earlier results of Chapter 2, (Guy et al., 2006) that were based on simpler algebraic functions of family means applied to a single year group, and extends them by including three further year groups. Additionally, it explores the importance of common environment effects and estimates genetic correlations between sites.

3.3 Materials and Methods

3.3.1 Description of test populations

Data consisted of mortality to IPN following seawater transfer, recorded as alive or dead, and covered seven sites, two recorded in year 2000, one in year 2001, three in year 2002 and one in year 2005. Different sites in the same year were replicates of the same families. The populations were coded as *year_site*. For example *2001_3* refers to the population of full-sib families derived from parents stripped (i.e. mated and eggs fertilised) in November 1999 and transferred to seawater at *site 3* in April 2001. Populations *2002_4, 2002_5* and *2002_6* correspond to *sites 1, 2* and *3*

respectively described in Guy et al. (2006), which gives further details of population construction and mortality sampling. For brevity we will refer to a family represented on a particular site at any one time as a 'full-sib group'. Note that populations *2001_3* and *2005_7* were the only populations tested in those years, so in those cases only, families were not replicated across sites.

Table 3.1 ***Numbers of parents and family structure for each year_site and the age of the parents at mating.***

Sites	cohort born	Parents born	Age of parents at mating	Number of female parents	Number of male parents	Numbers of males with differing mating ratios (females to males)			
						1	2	3	4+
2000_1 and **2000_2**	**1998**	1993	5	113	71	45	24	1	1
		1994	4	79	51	30	10	4	7
2001_3	**1999**	1994	5	4	20	11	4	2	3
		1995	4	197	110	80	23	3	4
		1996	3	0	8	7	1	0	0
2002_4, 2002_5 and **2002_6**	**2000**	1996	4	197	149	104	42	3	0
2005_7	**2003**	1999	4	186	83	31	21	22	9
		2000	3	7	8	7	1		0
All families				**783**	**500**	**315**	**126**	**35**	**24**

Table 3.1 shows the mating scheme giving rise to these families. In total 783 dams and 500 sires were crossed, with each dam being crossed to a single sire, so forming progeny groups of paternal half-sib and full-sib families. With replication of families over sites within year, these formed 1369 full-sib groups. 315 sires were single-pair mated (i.e. one mate only), 126 sires produced two full-sib families each, and 59 sires produced three to seven full-sib families each. The overall level of single pair mating (40 % of all dams mated) was a consequence of post-mortem diagnostic testing for the presence of IPNv in asymptomatic parents of eggs destined for export (Roberts et al., 2005) since single pair matings minimised the potential cull rate as a consequence of a male being detected as infected. Post-mortem disease testing also necessitated that males and females contributed offspring to only a single year. Between 0 and 58 % of matings each year involved parents from adjacent year-classes (one-sea-winter and three-sea-winter broodstock respectively). Although identification of individuals through DNA microsatellite analysis commenced with the parents used in the 1996 matings, identification of families common to adjacent year-groups could not start to be determined until four years later, when the offspring of family assigned individuals became parents themselves. Therefore this mortality data was collected when there was only weak pedigree connectedness between the early year-groups. Note, however, that parents of the 2005 year group, mated in 2003, were themselves offspring of parents of the 2001 year group, mated in 1999. In total the pedigree included 1980 further ancestors in addition to the 239,535 offspring recorded as being challenged with IPN.

Each female produced a single spawning of 10,000-20,000 eggs, which were split and eventually distributed as batches of full sib groups of various size to the appropriate site. At smolting in April of each year at 15 months of age (when the fish physiologically adapt to the seawater environment) batches containing equal numbers of each family were transferred as juveniles to the various seawater sites. Adjacent cages used in year 2000 were treated as separate sites, (*2000_1* and *2000_2*). The other sites were single cages in proximity to commercial stocks, apart from *2002_5*, representing a typical IPN event in seawater tanks of mixed broodstock families reared for egg production. The disease challenges occurred June to August, when fish were approximately 18 months of age.

Early mortalities within four weeks of transfer to sea occurred on each site and were assumed to be related to failure in the smolting process. Exposure to low levels of IPN virus in the aquatic environment was assumed to be immediate on transfer, but clinical signs specific to infection by the IPN virus and consequential mortality typically did not present until five to six weeks following transfer. This latency period is characteristic of the disease in seawater (Roberts et al., 2005). Once clinical signs were observed, a rapid rise in the prevalence of mortality indicated the presence of an IPN epidemic. A peak in mortality was typically reached 8 weeks post-transfer, after which mortalities declined to insignificant levels over the following two to four weeks. A more detailed description was given in Chapter 2, (Guy et al., 2006). For each site, veterinary inspection defined a period of time in which observed mortalities could be reliably attributed to IPN infection. The length

of the period varied between sites but typically fell within a window of 5-12 weeks, a pattern also characteristic of the disease. The analysis is restricted to those mortalities from IPN for which parentage could be assigned, either by electronic tagging (sites *2002_5* and *2002_6* only) or DNA analysis of fin-clips taken at mortality sampling. Efforts were taken to ensure all mortalities were monitored, recorded and fin clips collected for DNA parentage assignment (Chapter 2 and Guy et. al., 2006), from seawater transfer to the defined end of the epidemic on each site. Where possible, the small number of mortalities attributed to causes other than IPN during this period were censored from the analysis. It was not, however, practically possible in all cases to collect samples from all mortalities that were observed, and it was not possible to assign to family all those sampled. To maintain the statistical sampling properties of the population, in particular the overall prevalence, the sample size for each family (total alive and dead) was reduced from those initially placed on the site, by the percent of mortality subsampling that had taken place, as described in Guy et al., (2006). Parameters derived from the binary data therefore preserved the observed percent mortality in the analysis but at a reduced population size to allow for the greater uncertainty when only a proportion of mortalities can be sampled and assigned to parents.

Table 3.2 shows summary data for the number of families, number stocked, observed mortality and number of mortalities assigned to family.

1 ***Table 3.2: Prevalence (percent mortality) and its components observed for each site***

Year group _ site	Number of families	Total number stocked	Number of mortalities reported	Number of mortalities assigned	Proportion of mortalities assigned	Effective Number stocked	Average effective number stocked per family per site	Percent mortality
2000_1	192	27,433	5,103	1,534	0.30	8,230	43	18.6
2000_2	192	27,411	3,865	2,047	0.53	14,528	76	14.1
2001_3	201	53,179	15,847	2,594	0.16	8,705	43	29.8
2002_4	197	54,821	5,262	1,216	0.23	12,609	64	9.6
2002_5	197	16,098	1,918	1,918	1.00	16,098	82	11.9
2002_6	197	5,867	605	605	1.00	5,867	30	10.3
2005_7	193	54,726	1,751	1,700	0.97	52,857	274	3.2
All sites	**783**	**239,535**	**23,231**	**11,614**	**0.50**	**118,894**	**87**	**9.8**

Exactly 783 families were sufficiently traceable to be available for analysis, compared to the 800 intended. The Effective Number Stocked (table 3.2, column 7) both total and by family, was calculated as the Total Number Stocked reduced by the same proportion as the ratio of assigned to observed mortalities for reasons given above, and used as denominator in the calculation of family prevalences. This degree of sub-sampling of mortalities varied from 100 % on sites *2002_5* and *2002_6* where all mortalities were assigned, down to 16 % for site *2001_3*, where it was not possible to collect good quality DNA samples from all the IPN mortalities.

3.3.2 Data Analysis

Assuming each observation is recorded as either a survival ($y = 0$) or mortality $y = 1$), then factors contributing additively to the observation can be identified (borrowing terminology from Mrode 2005, section 3.4.1) as :

$$y_{ijk} = p_i + (½\ a_s + ½\ a_d + m_{ijk}) + f_j + e_{ijk} \qquad \text{Equation [1]}$$

where y_{ijk} is the observation on animal k within full-sib group j (j includes 1 to g replicates within the combination of sire s with dam d) located on site i. Note that any one site has only one of the g full-sib groups representing a family and replication occurs where a particular sire x dam combination is distributed over multiple sites within the same year, as described in section *3.3.1* above. Site and year are combined (p_i in equation 1) and represent a fixed effect in the model to account for site differences in mean prevalence across year by location combinations. All other terms in the model are

assumed to be random and normally distributed: a_s and a_d represent breeding values of the sire and dam respectively, and m_{ijk} is the mendelian sampling term particular to animal k (representing genetic differences between full sibs within family). Collectively, the terms in brackets make up the breeding value of the animal, $a_k = (½\ a_s + ½\ a_d + m_{ijk})$. The term f_j represents any common environment experienced by members of a full-sib group, which may in concept be due to a combination of sharing a common dam (dam maternal genetic and dam maternal environment), separate rearing of families prior to tagging, and non-additive genetic effects common to full sibs such as dominance and epistasis (Odegard et.al., 2007b). The term e_{ijk} represents a random environmental effect particular to each individual observation, and is the residual of the IAM model.

The IAM can process the individual $\mathbf{y}_{ijk}$ observations directly and yield variances and evaluations for fixed effects and breeding values where the contrasts in the data used for estimation do not coincide exactly (i.e. where the components are not completely confounded). Ancestral pedigrees on the dams partially remove the confounding of dam additive genetic with full-sib group effects. Dam maternal environment and all remaining full-sib group effects remain however largely confounded unless dams are replicated across sires (by splitting buckets of eggs and fertilising by different sires) which was not the case. In this field data the mating and site allocation design as described with the existence of ancestral pedigrees allowed some separation of full-sib group effects from dam additive genetic effects. The dam maternal effect and common environment (replicate) effect both rely on

the full-sib grouping data and so can be fitted separately but not together since they are completely confounded.

The Reduced Animal Model takes advantage of the fact that the breeding values of non-parents can be expressed in terms of the breeding values of their parents, by combining the terms m_{ijk} and e_{ijk} of individual *k* in equation [1] into a single 'RAM residual'

$$e^*_{ijk} = m_{ijk} + e_{ijk}.$$ equation [2]

This RAM residual variance (i.e. pooled within full-sib group variance) is fixed to $\sum [N_k P_k (1-P_k)] / (\sum(N_k) - r)$ where for each site, N_k is the number of full-sibs in family k, P_k is the observed family prevalence of mortality, and summation is over r families and all relevant sites. The RAM as implemented here is an extension of simple sire plus dam models, the differences being that the sire and dam variances are constrained to be equal, with full pedigree information used in the analysis. ASReml (Gilmour et al., 2006) provided the required flexibility and was used for all analyses. For clarity an example of the ASReml model is given in the appendix section 3.7.

Variance components for mortality were estimated for each of three models :

Model 1 : $y_{ijk} = ½ a_s + ½ a_d + e^*_{ij}$

was a separate univariate analysis for each site fitting only sire and dam (overlaid but typically reported as 'sire' in mixed model output) and RAM residual terms from equation [2] so excluding terms for full-sib group replicates of families (f_j) across sites. The additive genetic variance σ^2_a was estimated as $4\sigma^2_s$, the phenotypic variance σ^2_p was estimated as the RAM

residual plus twice the sire variance (computationally equivalent to $\sigma^2_a + \sigma^2_e$), and heritability was estimated as σ^2_a / σ^2_p. This provided a comparison with the same model calculated according to Chapter 2, (Guy et al., 2006).

Model 2 : $y_{ijk} = ½\, a_s + ½\, a_d + e^*_{ij}$

was a series of bivariate models treating each site as a separate trait (with its own column in the data) each according to Model 1. Genetic covariances were estimated between those sites that were genetically linked. The across year-group bivariate covariances that involved sites insufficiently linked to achieve convergence were constrained to zero. There was no environmental covariance between sites since individuals were reared on one site only and therefore this term was also constrained to zero in the iterations. This model regarded differences between full-sib group replicates as a genotype by environment interaction. Once good starting values were obtained for the iterations, a more convenient but equivalent model was available by specifying the covariance as a *Sire x Site* interaction.

Model 3 : $y_{ijk} = p_i + (½\, a_s + ½\, a_d) + f_j + e^*_{ij}$

was the full univariate model applied across all seven sites combined, including the full-sib group replicate (f_j) effect. Note that site (p_i) is fitted independently as a fixed effect to account for differences in mean prevalence across all families reared on the same site. The full-sib group effect can be coded either as an unreplicated common environment or alternatively as a general maternal effect (with replication within dam) that may have both genetic and environmental components. The resulting 'dam maternal' or

'common environment' effects respectively were expressed as a proportion of the phenotypic variance and calculated along with the heritability from the same model. It was not possible to estimate these together due to confounding caused by the high level of single pair mating. Standard errors of the variance component estimates and ratios (heritabilities and correlations) derived from them were provided by ASReml .

3.4 Results

Table 3.2 shows that the total number of fish transferred to sea across the seven sites was almost 240,000 smolts from a total of 783 families. The number stocked into a single cage on each site and available to be challenged by IPNv (after deduction of non-IPN losses) varied ten-fold from 5,867 on site *2002_6* to nearly 55,000 on site *2002_4* and reflected wide epidemiological differences in regional location, cage type, size and fish stocking density. Percent mortality differed between sites and years, averaging between 14 and 30 percent in 2000 and 2001 respectively, dropping to 10 percent in 2002, and just over three percent in 2005. The average number of fish challenged per family varied between sites from 30 to 240. There was no clear relationship between the percent mortality and the number of fish stocked as might possibly have been expected.

3.4.1 Univariate analysis of separate sites (Model 1)

The simple model fitting *sire*, *dam* and *residual* estimated the genetic variance as four times the 'pooled' estimate of sire and dam variance, achieved through constraining the dam and sire variances to be equal. Results are given in Table 3.3. Significant heritabilities were obtained on the observed (0,1) scale, varying from 0.07 on site *2005_7*, through 0.11 on site *2002_6* (the smallest dataset) to 0.56 on site *2001_3*, with standard errors approximately 10% of the heritability value. Across all sites, the estimated heritability was 0.38 as shown in Tables 3.3 and 3.5. The low heritability of mortality on site *2005_7* (0.07) was partly related to the low prevalence of mortality seen on that site and when transformed to the underlying liability scale (Robertson et al., 1949) the heritability ($\boldsymbol{h^2}_x$ = 0.42) was comparable to the others when similarly transformed.

As expected with binary traits, (Robertson et al., 1949), the heritability on the binary scale increased as prevalence increased from 0.03 to 0.30 (correlation r = 0.78 over the 7 sites). The last column of Table 3.3 gives heritabilities on the observed scale calculated according to Chapter 2 and Guy et al. (2006) and is consistently very close to the RAM estimates. For example the largest difference (Year group_site 2002_4) is 0.41 vs 0.38. This similarity is not surprising since the two methods use the variance of full-sib family means (V_f) in much the same way. The two replicates of sites *2000_1* and *2000_2* (two adjacent but separate cages in the same location) show significantly different heritabilities, although both had broadly similar prevalences, with a genetic correlation between them of 0.85 (see below).

Table 3.3 ***Heritabilities (h^2) and standard errors (se), for IPN mortality from RAM model 1, compared with heritabilities calculated from V_f, the variance of full-sib family means (h^2_p), and RAM estimates on the underlying liability scale (h^2_x)***

Year group _ site	Prevalence	Variance of family means	RAM			V_f
			h^2	se	h^2_x	h^2_p
2000_1	0.186	0.0234	0.28	0.029	0.59	0.27
2000_2	0.140	0.0114	0.16	0.018	0.39	0.17
2001_3	0.300	0.0618	0.56	0.043	0.97	0.56
2002_4	0.096	0.0178	0.41	0.036	1.22	0.38
2002_5	0.119	0.0109	0.20	0.021	0.53	0.19
2002_6	0.103	0.0079	0.11	0.017	0.32	0.11
2005_7	0.032	0.0009	0.07	0.008	0.42	0.05
All sites	**0.108**		**0.38**	**0.017**	**1.06**	

3.4.2 Multivariate analysis (Model 2)

The genetic correlations between family replicates on different sites are presented in Table 3.4. The multivariate heritabilities listed on the diagonals were consistent with the univariate single-site estimates, in size and standard error. Genetic correlations obtained between sites within the same year-group, were high and consistently between 0.70 and 0.85, with approximate standard errors below 0.06 in all cases. Blank cells in Table 4 indicate insufficient pedigree connectedness linking the sites associated with IPN data for the reliable estimation of genetic correlations. There was a high and significant genetic correlation of 0.72 (s.e. 0.13) between sites *2001_3* and *2005_7*. This is particularly notable since the parents of the *2005_7* population were themselves offspring of the parents of the *2001_3* year-group.

Table 3.4 : Heritabilities (diagonals), genetic correlations (lower triangular) and standard errors of correlations (upper triangular) between replicate families on different sites for prevalence of IPN mortality (model 2).

Year group _ site	***2000_1***	***2000_2***	***2001_3***	***2002_4***	***2002_5***	***2002_6***	***2005_7***
2000_1	**0.28**	0.03[#]					
2000_2	**0.85**	**0.16**					
2001_3			**0.56**				0.13
2002_4				**0.41**	0.05	0.06	
2002_5				**0.70**	**0.20**	0.05	
2002_6				**0.72**	**0.81**	**0.11**	
2005_7			**0.72**				**0.07**

.# for example rg (2000_1,2000_2) = 0.85 +/- 0.03

3.4.3 Univariate analysis over sites fitting common environment effects (Model 3)

The data description of equation [1] was applied to determine the proportion of the variance that could be explained by replication of fullsib families over sites, assuming a genetic correlation of one across sites. Table 3.5 compares the univariate model with and without these common environment effects.

Table 3.5 : Univariate variance components including common environment effect expressed as ratios of total phenotypic variance (model 3).

Model	*Additive Genetic h^2*	*Dam maternal*	*Fullsib group*	*Error*	*Pheno-typic variance*
genetic	0.378	-	-	0.622	0.097
+ fullsib group	0.291	-	0.039	0.670	0.096
+ dam maternal	0.329	0.016	-	0.655	0.096
Approx s.e.[#]	0.02	0.02	0.002	0.03	-

[#] *The components had approximately the same standard errors across each model*

Table 3.5 indicates a significant full-sib family component of 4%. While the phenotypic variance remained equal to ***p(1-p)*** over all models, the additional component reduced the proportion of additive genetic variance from 0.38 to 0.29, suggesting an upward bias to the heritabilities if these additional effects are not accounted for, although within range of twice the standard error of the estimates. The error term $\boldsymbol{V_e}$ was also seen to increase, as expected, since the RAM residual term is fixed howsoever the other components are partitioned. A Likelihood Ratio test of the addition of a random full-sib family effect to the simple additive genetic model yielded a highly significant chi-square of 373 for one degree of freedom (P <0.001). When modelled as a dam maternal effect, a non significant estimate of 0.016, less than its standard error, was obtained. Combinations of variance components not reported, including dam genetic effects, were not estimable from this data.

3.5 Discussion

3.5.1 Variance Components

Using mixed model techniques that utilise all phenotypic and pedigree information, the significant genetic variation in survival to seawater IPN challenge reported in Chapter 2 (Guy et. al., 2006) is confirmed for those populations and three additional year groups. The estimates of heritability and genetic correlations derived from simpler algebraic functions in Chapter 2 (Guy et. al., 2006) are also confirmed by the mixed model approach given

here, with the estimates in very close agreement where the models are comparable. Values greater than one were obtained for some of the heritabilities on the transformed scale, however the transformation employed ($\boldsymbol{pq / z^2}$) is an approximation. Heritabilities obtained for each site alone (Table 3.3) were essentially derived from the variance between families compared to the expected within family variance. Such estimates are highly sensitive to any aspects of husbandry or sampling which result in the observed variance between families being inflated. These problems are a challenge in field testing under commercial conditions and can only be reliably overcome routinely by using defined populations at the outset followed by rigorous sampling of mortalities. Where families are assigned by DNA microsatellite analysis, it becomes critical to use marker panels with the highest assignment rates over all the families likely to be encountered. In this work a 10-marker panel was found to be necessary, and it was sufficient to achieve 98% assignment of individuals to families using good DNA template.

The effect of replication of full-sib family groups over sites was found to be low at just under 4% (s.e. 0.003) but highly significant, if we make the assumption that the genetic correlation between sites is one. Possible sources could include small common environment effects from early rearing, lack of fit of the model, and genotype by environment interaction. Large common environment effects are not expected in this study since they would have to act at the very early egg and fry stages prior to mixing of the families although two populations (*2002_5* and *2002_6*) consisted of families that were not mixed until they were PIT tagged at approximately 50g. While the

effect of common rearing environment on growth traits is well known (Winkleman et. al. 1994), a quantifiable effect on survival persisting through to the seawater phase would not be expected. By contrast, in freshwater IPN challenge tests on fry, families are typically held for convenience in separate tanks for the duration of the challenge. Through replication of families, Kjoglum et al., (2005) was able to show a significant tank (i.e. replicate) effect of 6.3% amounting to half the estimated additive genetic effect present in three families isoallelic for MHC genes. A similar level was reported from routine challenge testing for IPN in freshwater fry (Wetten et al., 2007), although Kjoglum et al., (2008) reported a higher value of 0.23 when one of the year-groups was re-analysed with a threshold liability model. Partial factorial mating designs with sufficient replication (Winkelman et al., 1994) are required for a complete partitioning of common environment and dam maternal genetic and maternal environment effects. In this study, a degree of genotype by environment interaction is a more likely reason for both the full sib family replicate effects and the genetic correlations between sites being less than one. The epidemiological conditions prevailing on each site are expected to differ, particularly as some were geographically dispersed, so differential genetic response in the host to local epidemiological conditions may be expected. One such condition is the 'force' of infection as reflected in the levels of mortality observed at each site. For example it is interesting to note that the low mortality of 3.2 % on site *2005_7* was reported despite observing heavy mortalities from IPN (over 25 %) in adjacent cages of commercial fish stocks. The significant difference in mortality between

adjacent cages shown by the 2000 year group (containing the same families) is also characteristic of the epidemiology of IPN as seen in commercial practice. Different genetic mechanisms of resistance may be exposed and targeted as epidemiological conditions change across field sites. At the same time, while significant re-ranking of families can occur, the fact that the genetic correlations between sites are nevertheless moderate to high leads to some confidence that selection of resistant families based on such field data has the potential to result in strains of salmon more resistant to IPN over a wide range of epidemiological conditions.

3.5.2 Inheritance of resistance

It is interesting to note the high and significant genetic correlation of 0.72 (s.e. 0.13) between sites *2001_3* and *2005_7*. Although entirely different full-sib families are involved, there is strong pedigree connectedness (parent-offspring) between these two sites as described in Materials and Methods. The genetic correlation suggests a realised heritable transmission of resistance from one generation to the next (Odegard et. al., 2007b), whereas heritabilities restricted to within year-group patterns of variation merely imply the possibility of such as suggested in Chapter 2, (Guy et. al., 2006). This was the first opportunity in the ongoing sib-testing program to re-test a population that had passed through one generation of selection for IPN. It is shown elsewhere (Houston et al., 2008) that there is strong evidence for one or more QTL with major effects on IPN resistance segregating in these two

particular year-groups. If selection has successfully captured the favourable QTL alleles, then this would account for both the reduced prevalence and reduced genetic variance in the second generation. These results therefore lend support to the finding that QTL of large effect are segregating in these populations.

Midtlyng et al., (2002) reported a study which showed that MHC class 1 and 2 alleles were strongly associated with resistance or susceptibility to bacterial and viral diseases, including IPN. This led the authors to suggest that standardising for MHC alleles may be essential for reproducible challenge trials and a possible reason for variable results. Involvement of MHC genes may explain the low genetic correlations found between resistance to different viral diseases and the sometimes unfavourable genetic correlations between resistance to bacterial and viral diseases (Gjoen et al., 1997; but neutral in Henryon et al, 2005, and positive in Odegard et al., 2007). Using three families that were iso-allelic for MHC genes, Kjoglum et al., (2005), however, were able to show significant non-MHC associated genetic variation for IPN resistance, which, estimated at 12.7%, was the highest out of the three diseases (IPN, ISA and furunculosis) studied. As selection is directed more precisely at these diseases it will become more important to know exactly which genomic mechanisms are being targeted.

3.5.3 Confirmation of heritable resistance to IPNv

There is now a growing literature derived from industry-based breeding schemes confirming the presence of considerable genetic variation

in the resistance of Atlantic salmon to mortality caused by the IPN virus : in fry (Wetten et al., 2007; Kjoglum et al., 2008) and in post-smolts (this chapter and chapter 2 (Guy et al., 2006). Wetten et al., (2007) reported a linear mixed model analysis of freshwater fry challenges in Norway covering eight years using experimental facilities which yielded heritabilities on the untransformed observed scale of 0.16 to 0.39, (pooled 0.31) relating to prevalences between 23 to 63% mortality. These prevalences were subject to the constraint, described in Kjoglum et al., (2008) that the induced challenges were terminated once an overall mortality of approx 50% had been achieved. This is in contrast to the naturally occurring seawater challenges and mortality reported here which were allowed to run to the end of the epidemic. Although operationally convenient and known to maximise the phenotypic variance, the utility of a 50 % prevalence for maximising the heritability depends on the nature of the disease and the disease outbreak. In addition the definition of the trait is effectively modified to include an element of 'time to death' which may include different genetic mechanisms from survival *per se*. Previous work has shown that clear family differences (Guy et al., 2006) and heritable variation (Henryon et al., 2002; Odegard et. al., 2006; Odegard et. al., 2007b) can be demonstrated from salmonid survival analysis data collected as a time series.

Kjoglum et. al., (2008) also reported a heritability of survival to freshwater IPN challenge in fry of 0.55 in a re-analysis of one of the year-classes using a liability threshold model with Gibbs sampling. Two other year-classes from the Wetten et. al., (2007) study also experienced field

outbreaks from the same experimentally challenged families – one in seawater (1997, mortality 18.5%) and one in fresh water (2002, mortality 10%). While the heritability from both outbreaks was low at 0.10, Wetten et al., (2007) were able to show a genetic correlation of 0.78 (s.e. 0.16) between the previous freshwater challenge and the seawater field outbreak. Similarly an estimated r_g of 0.83 (s.e 0.07) was obtained between the freshwater challenge and the following freshwater field outbreak, suggesting that experimental fry challenge was a suitable and efficient predictor of field performance. Further confirmation came from the offspring of the 2001 year-group (Storset et al., 2007), where selected High and Low resistant families were subject to both freshwater and seawater IPN challenge. Although designed as a commercial demonstration, they were able to show that low resistant families had more than double the mortality of the high resistant families.

3.6 Conclusions

These results suggest that selection on the basis of EBVs derived from mixed models could lead to the development of resistant families and strains of Atlantic salmon, on time scales similar to that for other traits recorded on a family basis, such as harvest and processing traits. Midtlyng et al., (2002) suggest that because of the high reproductive rate and the opportunity to score large families in challenge tests, fish have higher potential for improving resistance to infectious diseases than most other food producing animals, to which these data add support (discussed further in this thesis in

section 6.4.3). Furthermore, these analytical models open up the possibility of estimating the relationship between disease and other binary traits and other quantitative performance traits, leading to a full integration of selection objectives in commercial breeding programs. All these results further confirm the existence of a broadly similar genetic response of the host to the viral challenge over wide epidemiological conditions. That would suggest that selection for increased survival based on these field challenges would lead to improved host resistance across a wide range of environments including those experienced from year to year.

Pathogens are, however, expected to evolve in response to increasing selection pressure on the various resistance mechanisms possessed by the host (Bishop et al., 2003; Gandon, S., 2001). Host resistance mechanisms which prolong the time to death without, for example, blocking the infectivity of the virus may reduce the overall impact of a disease in the short term, but have other long term consequences. For example, partially effective vaccines (or equivalently incomplete genetic resistance) which prolong the period in which fish shed active virus into the environment before succumbing themselves has consequences for population wide immunity and may promote more rapid evolution of virulence in the pathogen (Gandon, S., 2001). How different host mechanisms are targeted by different species of pathogen, vaccination regimes, challenge and data collection protocols may all be important in determining how the pathogen evolves and at what rate. That in turn will determine the long term sustainability of selective breeding as a strategy for controlling disease.

3.7 Appendix : Example ASReml coding for RAM model

For data in the following format :

Fixed	Family	Sire	Dam	morts	total	morts %	residual
site1	family1	sire1	dam1	2	43	0.0465	1.9065
site1	family2	sire2	dam2	11	43	0.2558	8.1857
site2	family1	sire1	dam1	3	76	0.0395	10.7787
site2	family2	sire2	dam2	13	76	0.1711	2.8834

Ram residual : site 1 = 0.120146, site 2 = 0.184623, over all sites = 0.100654

ram.as

```
        site   !A  family  !A    sire     !A    dam !A
        morts   total  mortspc
pedfile.prn
datafile.csv     DOPART 1       # or DOPART 2

!PART 1                          # univariate heritability

mortspc   !wt   total ~   mu   site   !r   sire   and(dam)
1 1 0
0 0 0    !S2==0.100654

VPREDICT !DEFINE            # calculation of heritabilities
F Vs           # 1
F Ve           # 2
F Vg  1 * 4   # 3                 (ie Vs * 4)
F Vf   1 * 2   # 4                (ie Vs * 2)
F Ve  2         # 5                (ie Ram Residual)
F Vp  2 + 4    # 6                (ie Ram residual + (2 * Vs)
H h1  3   6

!PART 2                          # correlation between sites

mortspc  !wt  total  ~  site   !r  site.sire   -site.dam  and(site.dam)
2 1 1                            # 2 sites (traits) to correlate, environmental
2 0 0  !S2== 0.12015             # 2 families on site 1, fixed ram residual
2 0 0  !S2== 0.184623            # 2 families on site 2
site.sire 2                      # genetic, 2 dimensions
                                  # dimension 1, variance / covariance
2 0 US  !GPPP                    # unstructured, keep positive definite
0.0104                           # genotypic starting values, lower triangular
0.01     0.0049
sire                             second dimension, sire
```

Chapter Four

Freshwater challenge and genetic correlation with seawater challenge.

4.0 Abstract to the chapter

Mortalities from Infectious Pancreatic Necrosis (IPN) due to infection with the IPN virus (IPNv) remains a major problem across the major Atlantic salmon farming regions of the world. The virus causes mortalities at both stages of the life-cycle, and prevalences of 0.3-0.8 or more (freshwater) and 0.05-0.3 (seawater) are typical. Available injectable vaccines are only partially effective in seawater, and not applicable in freshwater since IPN attacks at the early fry stage before a competent immune system has developed. Host genetic resistance has been previously demonstrated in Landcatch smolts transferred to seawater for yeargroups 1998 to 2003 with heritabilities up to 0.56 being observed. Freshwater challenges were conducted annually on mixed families of broodstock sibs from year-groups 2005 to 2009 in experimental disease containment facilities. A similar level of genetic variation for resistance to freshwater IPN as was found with seawater IPN with a heritability of 0.49, (s.e. 0.03), range 0.31 to 0.59, and genetic correlation between family replicates of 0.80 to 0.95, (s.e. approx 0.05), using an Individual Animal Model. When seawater and freshwater

mortality data were analysed together, assuming seawater and freshwater survival to be the same trait, the heritability increased to 0.67, (s.e. 0.02). On testing this assumption, the genetic correlation between freshwater and seawater mortality was found to be 0.68 (s.e. 0.09). Being significantly different to one, the possibility remains open of some functional differences in genetic resistance between the two life-stages of infection. Further results from elsewhere have confirmed the existence of one or more QTLs for IPN resistance segregating in these and other populations, and are discussed in relation to these results.

4.1 Introduction

Promoting genetic resistance in the host has become widely recognised as both achievable and sustainable (Gjedrem, 2005). The routine creation of very large families of full-sibs, in excess of 10,000 in salmon and other aquaculture species, facilitates large scale field and experimental exposure trials where family differences in mortality rates can be readily observed. In Chapter 2 it has been shown that this is sufficient to demonstrate genetic variation, where it exists, that can be exploited by selection. For selection to be sustained to the point that resistant strains can be made available to the commercial industry requires the framework of a pedigree-based breeding program. Two such programs have addressed the genetic improvement of resistance to IPN mortality : the Norwegian national program, now privately operated (AquaGen AS, 2010) has run experimental

trials based on freshwater bath exposure of fry since 1997 (Wetten et al., 2007). The Scottish program, now operated by Landcatch Natural Selection Ltd (LNS, 2010), ran seawater field trials from 1999 until 2006 (Chapter 3) since when they were replaced by immersion exposure in experimental freshwater facilities. Both programs have included IPN resistance as a selection trait in their respective breeding goals since the 2001 breeding season. Moderate mean heritabilities (0.31) were reported for the annual Norwegian freshwater, fry stage tests (Wetten et al, 2007) and Scottish seawater field smolt stage tests (0.38, Chapter 3). Wetten et al., (2007) further reported high genetic correlations (0.8) between the freshwater fry tests and seawater field outbreaks of IPN. Storset et al., (2007) reported two commercial demonstrations of IPN resistance in the Norwegian program involving a single generation of divergent selection of parents from the year-group stripped in 2000, based on the fry test in 2001. In mixed families of fry from the 2004 strip challenged in freshwater in 2005, 29% and 67% mortality was seen in the high and low resistance groups, respectively. In mixed families of advanced smolts ('s-zero' category smolts, go to sea before their first winter) also tested in 2005, 32% and 79% mortalities were seen in the two groups, respectively. These trials therefore gave confidence that selection should be effective in reducing mortalities from IPN.

This Chapter four reports heritabilities and genetic correlations between replicated families obtained from the experimental freshwater challenges carried out in the Scottish program since 2005. These are discussed in relation to the seawater testing carried out previously. In the

Scottish program there has been no opportunity for testing any single year-group in both freshwater and seawater. It was possible however to derive estimates of genetic correlations between year-groups including those between freshwater and seawater sites where sufficient pedigree connectedness existed in the accumulated pedigrees. Three questions were investigated: (i) is there significant host genetic variation in IPN resistance based on freshwater challenge tests?, (ii) is IPN resistance consistent between family replicates? and (iii) is freshwater IPN mortality consistent with seawater IPN mortality.

4.2 Materials and Methods

4.2.1 Freshwater challenges

4.2.1.1 Choice of fish

Freshwater challenges (2005 to 2009 year-groups) were conducted in experimental disease containment facilities at the Centre for Environment, Fisheries and Aquaculture Science (CEFAS) Weymouth, UK : sentinels were created as mixed populations of eggs after final selection of the 200 reserved broodstock families in February each year at the Landcatch hatchery.

Degree-day development (the sum of daily water temperatures, upon which fish development primarily depends) was synchronised to a common hatching window of two weeks using mixtures of heated, ambient and cooled fresh water for incubation. At the eyed stage, a sample of 15 eggs was taken

from each of the 200 families to form a single batch of 3,000 eggs. The process was repeated to form a second batch after which both batches were transferred to a single tank each at CEFAS, site *s7*. Additionally, approx. 700 eggs of mixed families were provided for pre-challenge testing to establish appropriate viral titres for the main trial. In the first freshwater trial at CEFAS (2005 year-group) a control population of 3,000 fry from the same families was additionally provided. With IPNv having been confirmed in the challenged mortalities and confirmed absent in the controls, along with any other defined causes of mortality, the controls were not analysed and in subsequent years were dispensed with on the basis that within-tank family comparisons were sufficient for quantitative genetic analysis.

4.2.1.2 Virus source and preparation for freshwater challenge

All year-groups were challenged with the same virus isolate V0512-1 serotype Sp A2. This was obtained from infective tissue of hatchery-reared Atlantic salmon fry experiencing a high mortality IPN event in May 2005, the year previous to the first freshwater challenge described here. The *Sp* virus strain is responsible for all seawater IPN mortality events in Atlantic salmon farms monitored to date (Smail et al., 2006). Infected tissue was provided by the Institute of Aquaculture, Stirling and virus isolated at CEFAS Weymouth according to the process previously reported (Houston et al., 2010). Chinook salmon embryo (CHSE-214) cells are typically used for reactivating virus stocks and diagnosis of IPNv strains. Repeated passaging of IPNv in cell culture using CHSE-214 has been shown by Song et al., (2005) to promote

specific mutations in the viral RNA genome leading to reduction or loss of virulence. Therefore, the isolate was cultured instead on rainbow trout gonad (RTG-2) cells at 15 degrees C, and stored at -20 degrees C in dilute (1:1) glycerol to prevent a drop in titre through repeated freeze-thaw. On reactivation for each challenge, the virus was passaged only twice in CHSE-214 to build up stocks before challenge. Cell culture harvests were tested by Enzyme-linked immunosorbent assay (ELISA, TestLine Clinical Diagnostics Ltd) for confirmation of virus. Virus concentrations of the stock were confirmed by titration on 96-well CHSE-214 plates as 1.5×10^7 Tissue Culture Infective Dose (TCID) per ml, re-confirmed on the day of challenge. It was important to confirm that the virus was infective in the stock to be challenged before the main trial since there would be no further opportunity to repeat the trial in that year group. Also, with all-or-nothing (0,1) traits, variance is maximised at a prevalence of 50 %. Therefore, a pre-test was performed each year on additional batches of approx 700 fry aiming to establish the virus concentration required to achieve an approximate 50 % mortality in the main test. Fry development for the pre-test was advanced over that of the main batches by water temperature control. This was to ensure that fry for the main challenge did not enter the development phase during which they were less susceptible to the challenge, typically some 4 months after hatching.

4.2.1.3 Description of freshwater immersion challenge

On arrival at CEFAS, the batches remained separate and families hatched as previously synchronised within the first two weeks of March. First feeding followed 4 weeks later and swim-up 4 weeks after that. Each batch of fry was transferred to one of two 150 litre replicate tanks to acclimatise 7 days prior to challenge at the end of May The control used in the 2005 year group experienced all procedures except exposure to the virus. Dechlorinated water was supplied to each tank at a flow rate of 3 to 5 litres per minute at 10 degrees C and 80 % minimum oxygen saturation at all times. Light was provided at approx 200 lux at the water surface on a 12 hour day/night cycle with a 30 minute dusk/dawn period. During the experimental phase feed was offered manually three times per day and both tanks starved overnight prior to the challenge. Feeding was observed to return to normal the following day.

On the day of challenge, all fry in each of the challenge groups were transferred by hand net into a bucket containing 40 L aerated water inoculated with 40 ml of IPNv at 1.5 x 10^7 TCID per ml, giving a final challenge dose of 1.5 x 10^4 TCID per ml. After four hours the buckets were removed and fry returned to the appropriate challenge tank by hand net. The procedure was repeated for each replicate. Following the immersion challenge, physical tank and water parameters were monitored continuously, and mortalities were removed and individually recorded twice daily. Records included date and time of mortality, observed clinical signs, and general changes of behaviour and appetite. The trial was terminated at the latest by 42

days post challenge when mortalities had fallen to background levels. All mortalities and survivors at the end of each trial were collected. The target mortality of 50% was only achieved in the first 2005 year-group, falling from 40% to 20% in subsequent years.

4.2.1.4 Confirmation of IPN infectivity in freshwater

All the mortalities from each tank were frozen whole at -70 ^{0}C. At the end of the study, pools of 5 mortalities from day 14 and day 21 post challenge from each tank were tested for the presence of IPNv in cell culture : 24 hour old confluent monolayers of CHSE-214 cells in 96 well cell culture trays were inoculated with filtered tissue homogenate diluted 10^{-3} and 10^{-4} in cell culture medium, incubated at 15 ^{0}C, and monitored daily for cytopathic effect (CPE, destruction of the cell layer) characteristic of IPNv. Positive cell culture supernatant was then tested by ELISA to confirm IPNv. On termination of the study, one pool of five survivors from each tank were similarly tested which proved positive for cytopathic effect, confirmed by ELISA. When titrated, many cell cultures were found to be below the sensitivity of the titration assay (5 x 10^{3} $TCID_{50}$ per ml) suggesting there may have been a persistent infection established at a very low level or alternatively was in the process of being cleared out of the fish when the trial was ended.

4.2.2 Seawater challenges

Freshwater data will be analysed in relation to seawater data where there are pedigree connections through the pedigrees. The seawater data available were listed and described in the introductory Chapter 1, section 1.5. Although also extensively discussed in Chapter 3, the year-groups listed there were labelled by the year when the challenge took place, rather than the year when fertilised, as in this Chapter and the general description of Chapter 1. Table 1.1 is reproduced as Table 4.1 below since it will be extensively referred to.

Even though it can be seen some year-groups failed to elicit useable seawater data (1997, 2001, 2002 and 2004), they are retained in Table 4.1 to allow the tracking back of the 4-year parent to offspring breeding lines, connecting the seawater to the freshwater data. For example, the 2007 year-group families challenged in freshwater in 2008 had: (i) 2003 year-group parents whose families were challenged in seawater in 2005, and (ii) 1999 year-group grandparents whose families were challenged in 2001. Alternatively, the 2005 year-group in the first freshwater challenge in 2006 had no useable linked seawater challenge data at either the parental (2001) or grandparental (1997) year-group levels. The 1997 year group was challenged but IPN mortality collection and subsequent assignment to family was sufficiently problematic to exclude this data set from statistical analysis. Because of the low mortality rate in the 2003 year-group (3.2 %), mortality recovery and assignment to family was nevertheless good and this population formed an important link between the seawater and freshwater data.

Note, however, that there was no year-group that itself experienced both a seawater and freshwater challenge.

Table 4.1 : Prevalence of IPN and number of fish challenged by year group.

Year-group stripped	site	cages (c) or tanks (t)	Year of IPN event	% prevalence by year-group	Numbers stocked
seawater					
1997	s0		1999	22.7	17,006
1998	s1	c1, c2	2000	16.4	54,844
1999	s1	c3	2001	30.0	53,179
2000	s2, s3, s4	c4, c5, c6	2002	10.8	76,786
2001	s4		2003	0	5,000
2002	s2		2004	0	55,000
2003	s5	c7	2005	3.2	54,726
2004	s6		2006	0	60,000
freshwater					
2005	s7	t8, t9	2006	60.4	6,215
2006	s7	t10, t11	2007	40.4	4,846
2007	s7	t12, t13	2008	28.6	5,247
2008	s7	t14, t15	2009	20.7	5,874
2009	s7	t16, t17	2010	21.9	5,450

Genetic parameters from mixed-model analyses of seawater IPN mortality on the remaining four year groups (1998, 1999, 2000 and 2003) were previously reported (Chapter 3). Results are confirmed here in this study and extended in relation to freshwater challenge data obtained subsequently. Note again that in Chapter 3, for the convenience of publication, year groups

were labelled by the year when the fish were challenged, rather than here in Table 4.1, by the year when parents were stripped, in order to provide consistency with the freshwater challenge trials. Note also that column 4 of Table 4.1 gives the equivalent year of the IPN event.

4.2.3 Data Analysis

From the individual mortality data, average percent mortality was calculated by site, cage or tank and family. This is equivalent to prevalence since challenges were allowed to proceed to their natural end. Mortalities not unequivocally caused by IPN and those that were unable to be assigned to family were removed (censored) from analysis. All uncensored fish were assigned a binary variable $\boldsymbol{Y_i}$, *with* $\boldsymbol{Y_i = 1}$ if the fish died and $\boldsymbol{Y_i = 0}$ if the fish survived.

Data from IPN challenged cages and tanks from all year groups were analysed fitting an individual animal mixed-model, (IAM, Mrode 2005, section 3.4.1) using all known ancestors of the challenged fish for four generations. Note that all the seawater data sets in RAM format (analysing prevalence) used for Chapter 3 were converted to individual animal model data sets (individual fish 0,1 observations) for these Chapter 4 analyses since the freshwater data was already in that format when collected. Genetic groups based on the year-group of the missing parent were fitted where appropriate. The analysis was implemented using ASReml version 3.0 (Gilmour et al., 2009) with the additive genetic variance calculated using the model:

$$Y_{ijk} = g_i + c_j + a_{ijk} + e_{ijk}$$

where Y_{ijk} is the observation on the k^{th} fish, g_i is the fixed effect of the i^{th} genetic group, c_j is the fixed effect of the j^{th} cage or tank, a_{ijk} is the random (additive genetic) effect of the k^{th} fish and e_{ijk} is the random environmental effect particular to each individual observation, the residual term of the model. Sex cannot be determined until the fish start to mature at 3 years old.

Cages and tanks were uniquely labelled (Table 4.1) and so fitting cage and tank also removed the effects of year, site and the replication of families over cages or tanks within year-group, where that occurred. Family replication can be fitted either as a dam effect (assuming a genetic correlation of unity between sites as explored in Chapter 3) or as a genetic correlation between family effects in different cages or tanks. Since an aim of this study was to derive genetic correlations within and between year-groups, dam effects were not fitted. The heritability (h^2) for mortality (equivalently, survival) to IPN was estimated as $h^2 = \sigma^2_a / (\sigma^2_a + \sigma^2_e)$ where σ^2_a and σ^2_e are the estimated additive genetic variance and residual variance respectively. To answer objective (i), whether there was significant host genetic variation in resistance to IPN in freshwater, univariate heritabilities were obtained first within tank or cage location (omitting the cage component c_j from the model), and then across locations. The variance components from these then served as starting values for between cage and between site bivariate covariance component analyses to determine whether any genetic resistance is consistent between replicates and between seawater and freshwater life-stages of the host objectives (ii) and (iii).

4.3 Results

4.3.1 Freshwater challenge

4.3.1.1 Freshwater prevalence

The five year-groups exposed to experimental freshwater challenge are listed in the lower half of Table 4.1. Only in the first year-group (2005) was the target mortality of 50 percent exceeded, averaging 60% by the end of the epidemic. In subsequent years, mortalities had returned to background levels of less than two or three per day after reaching approximately 40, 30 and 20% respectively. These figures therefore represent prevalences thus avoiding any artificial censoring of individual families due to the duration of observations, which was a maximum of 42 days in all trials. 2005 was also the only year-group to show a significant difference in prevalence between replicates, 0.47 and 0.74 respectively. In all other year-groups, of both freshwater trials (Table 4.2 column 3) and seawater trials (Table 4.3 column 3), replicates were consistent, over a range of sites and conditions. In particular, in the 2009 year-group, mortalities in the two replicate tanks with prevalence 0.218 and 0.220 respectively, matched each other almost exactly for the whole 42 days duration of the trial.

Figure 4.1 - Cumulative mortalities, 2005-strip

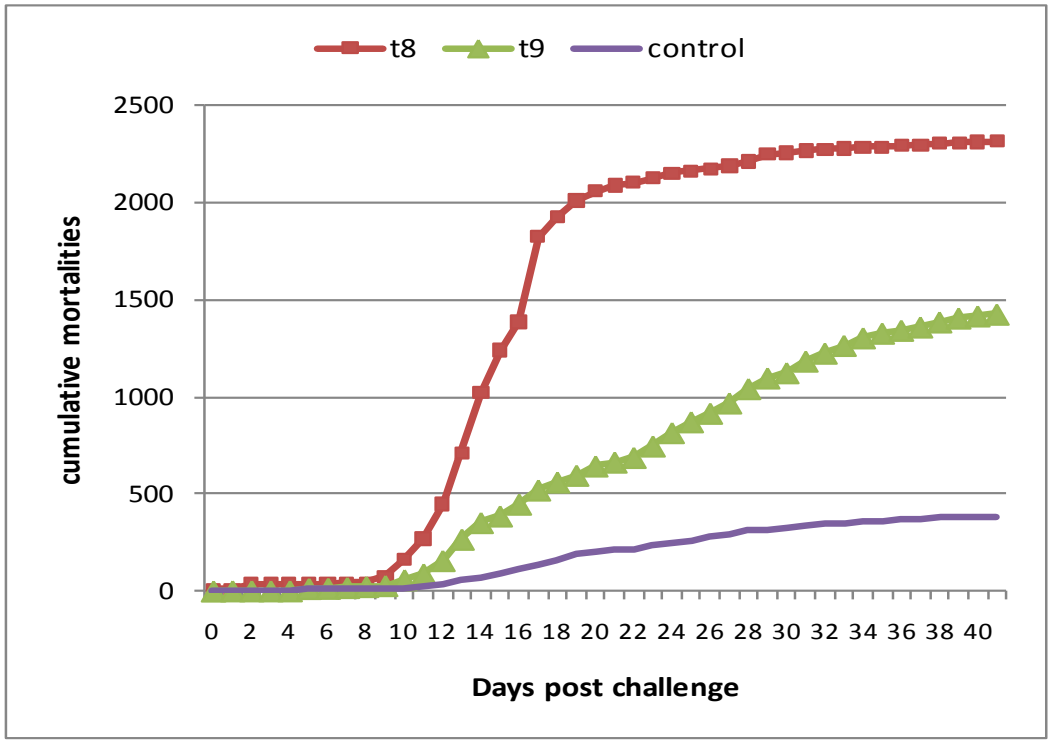
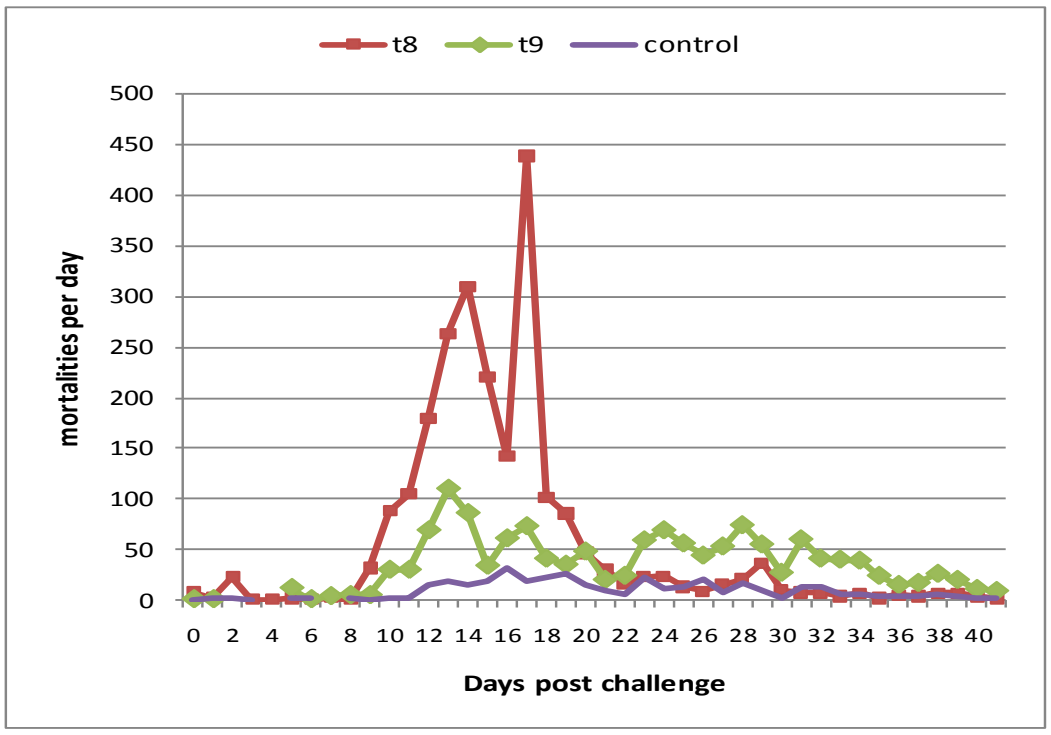
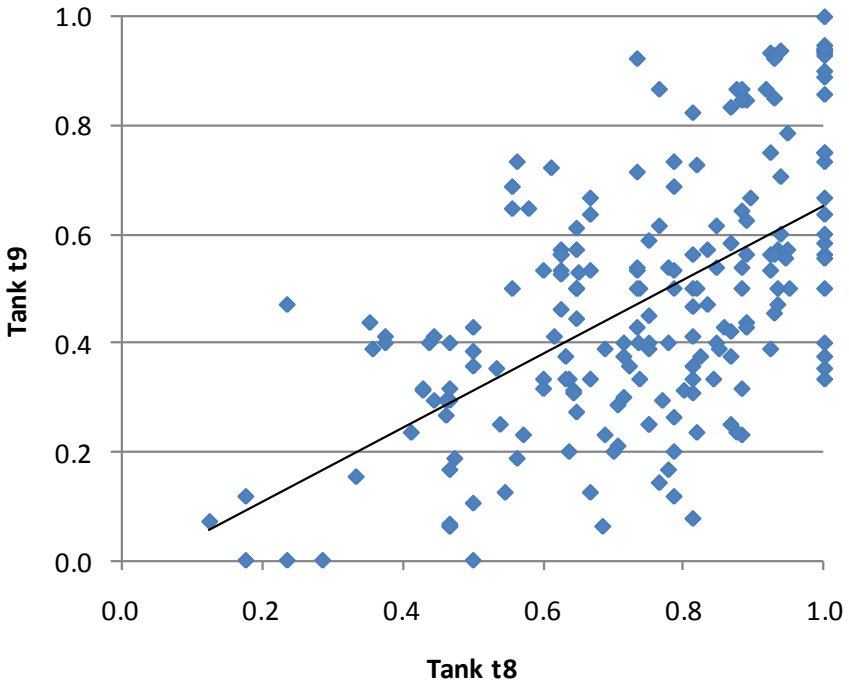


Figure 4.2 - Incidence of mortality by day, 2005-strip



The cumulative mortality curves for the three replicates of the 2005 year-group are shown in Figure 4.1. Apart from the controls almost all the mortalities were due to IPN. The development of the epidemic is clearly different between the two challenge tanks, with a sudden rise and rapid return to background levels in tank *t8* while tank *t9* showed a steadier rate of mortalities and a less clear end to the epidemic. This is illustrated by Figure 4.2 which plots the daily mortality (ie incidence) clearly showing the spike of mortality in tank *t8* that is absent from *t9*. Given the standardised families, environmental and challenge conditions, it was not possible to attribute the different pattern of mortalities over the two tanks to any particular cause, and it was not seen between replicates in subsequent years. In total, 385 mortalities unrelated to IPN were seen in the controls, which is a mortality level not unexpected when handling eggs and fry. The overall mortality in controls was about 0.13, but family assignment was not performed and within-family prevalence was not estimated. Despite the difference in the trajectory of the epidemic in the challenge tanks, the phenotypic correlation of family prevalence between the two challenge tanks was positive $\boldsymbol{r_p} = 0.51$, (plotted in Figure 4.3) and higher than this in the other years. This, itself, is indicative (Chapter 2) of high genetic variance. In tank ***t8***, 24 families had 100 % mortality, while all families had at least one mortality. In tank ***t9***, just one family had 100 % mortality, and 4 families had no mortalities at all.

Figure 4.3 : Correlation between family prevalence in the two challenge tanks, 2005-strip



x axis *:family prevalence observed in Tank t8 ;* ***y axis*** *: family prevalence observed in Tank t9;* ***each point*** *: one family, replicated in both tanks ;* ***fitted line*** *: least squares best fit (regression of y on x)*

4.3.1.2 Freshwater - univariate heritabilities

Univariate heritabilities were calculated separately by tank for all three year-groups (2005, 2006 and 2007-strip) where family assignment had been completed, and are presented in Table 4.2.

Table 4.2: Freshwater - prevalence, univariate variance components and heritabilities obtained by year and cage site within year.

Yeargroup Cage/ tank	Number challenged	Average IPN mortality	Genetic variance V_A	Environ-mental variance V_e	Herit-ability h^2	s.e. h^2
2005_t8	3,125	0.737	0.061	0.136	**0.311**	0.042
2005_t9	3,090	0.470	0.082	0.169	**0.325**	0.043
2006_t10	2,135	0.415	0.128	0.104	**0.550**	0.058
2006_t11	2,711	0.395	0.142	0.098	**0.592**	0.057
2007_t12	2,609	0.283	0.098	0.101	**0.491**	0.053
2007_t13	2,639	0.289	0.117	0.088	**0.570**	0.055
2008_t14	2,991	0.199	#			
2008_t15	2,883	0.214				
2009_t16	2,730	0.218				
2009_t17	2,720	0.220				
All freshwater	**27,633**	**0.348**				
Assigned freshwater	**16,309**	**0.442**	**0.109**	**0.116**	**0.485**	**0.026**

[#] *blank cells denote family assignments not yet available (yeargroups 2008 and 2009)*

The lowest heritabilities were obtained for the first year of freshwater testing, 2005 strip, and were consistent at 0.31 and 0.33 for the two tanks respectively, despite the differences in mortality. Heritabilities were higher in subsequent year-groups, the highest being 0.59 achieved in the 2006 year-group and maintained for the year-group following, despite average mortalities falling. These heritabilities are on the observed scale (probabilities) and all other things being equal are expected to decline as the mean mortality deviates from a value of 0.5 (Chapter 2). The standard errors were between 0.042 and 0.058. These were higher than those obtained for the seawater mortality data, (Chapter 3), reflecting the family size here of 15 sibs per tank compared to families of 25 to 250 or more in the seawater data. Over all three year-groups (fitting tank as a fixed effect) the pooled heritability obtained from the freshwater trials to date was 0.49 (s.e. 0.03).

4.3.2 Seawater challenge - univariate heritabilities

Table 4.1 shows a range of prevalence from a high of 30% at site *s1* to an unexpected zero at site *s6*. Significant mortality to IPN was only seen in the first four years (1997 to 2000 year-groups) of seawater testing, with prevalence in the final four years being very low or zero. As well as the 1997-strip year-group, three further sites failed to elicit useable mortality data: the 5,000 smolts (2001 year-group) placed on site *s4* had encountered a freshwater IPN event in the hatchery that was rapidly contained by biosecurity measures. This appears to have mitigated against an IPN event

when they went to sea (Chapter 2; Roberts and Pearson, 2005). IPN was reported on site but just 48 morts (0.11 %) were attributed, between weeks 4 to 9. Site *s2*, 2002 year-group, experienced high levels of general mortality not associated with IPN. Therefore collection and processing of mortalities was not carried out. The 2004 year-group (60,000 smolts) appeared not to produce any mortalities to IPN, although IPN was reported in smolts from other sources in adjacent cages.

Table 4.3 gives effective number challenged (corrected for the proportion of fish assigned to family) and average mortality to IPN for year-groups providing seawater data. Overall, seawater mortality on sites with confirmed IPN was 10.8 %**.** Heritabilities ranged from 0.06 for cage *c7* to 0.40 in cage *c3*, consistent with those previously reported in Chapter 3. The standard errors (0.01 to 0.04) are slightly improved over those in Chapter 3, due to a small amount of pedigree verification through re-genotyping. Overall seawater populations, fitting cage as a fixed effect, the heritability on the observed scale was 0.43 (s.e. 0.02). This was higher than any of the estimates for individual cages and a little lower than the pooled freshwater heritability (0.49, s.e. 0.03) presented previously in Table 4.2. When the seawater and freshwater data was combined, on the, as yet untested, assumption that the same trait is involved, the heritability rose to the highest value obtained so far, 0.67 (s.e. 0.02).

Table 4.3 : Seawater IPN - Prevalence, univariate variance components and heritabilities obtained by year and cage site within year.

Yeargroup Cage / tank	Effective Number challenged	Average IPN mortality	Genetic variance	Environ -mental Variance	Heritability		Estimates reported in Guy et al (2009)	
					h^2	s.e. h^2	h^2	s.e. h^2
1998_c1	8,256	0.186	0.041	0.111	**0.271**	0.028	0.28	0.029
1998_c2	14,592	0.140	0.018	0.102	**0.153**	0.017	0.16	0.018
1999_c3	15,075	0.300	0.085	0.126	**0.403**	0.034	0.56	0.043
2000_c4	12,609	0.096	0.036	0.055	**0.398**	0.036	0.41	0.036
2000_c5	16,098	0.119	0.020	0.086	**0.189**	0.020	0.20	0.021
2000_c6	5,867	0.103	0.009	0.084	**0.093**	0.016	0.11	0.017
2003_c7	52,857	0.032	0.002	0.029	**0.060**	0.007	0.07	0.008
All seawater	**125,354**	**0.108**	**0.043**	**0.057**	**0.431**	**0.019**	**0.38**	**0.017**
All data	**141,663**	**0.147**	**0.090**	**0.044**	**0.670**	**0.019**		

4.3.3 Genetic correlations between replicate families

Table 4.4 lists the bivariate heritabilities and genetic correlations (r_g) for cases where sufficient pedigree links between cohorts of fish ensured that correlations were estimable. In the seawater data, family replicates existed for the 1998 and 2000 year-groups and are presented in the first four rows of Table 4.4 The genetic correlations were uniformly high, from 0.70 to 0.87 with standard errors sufficiently low (around 0.05) to interpret them as significantly different from unity. These results confirm those reported previously (Chapter 3). In the freshwater data, all year-groups were replicated across two tanks and r_g ranged from 0.80 in the 2007 year-group to 0.95 in the 2006 year-group.

4.3.4 Genetic correlations between year-groups

Genetic correlations between year-groups were successfully obtained where there was direct pedigree connectedness across generations, i.e. where families were direct offspring of families created four years earlier, both having associated data. It can be seen from Table 4.1 that the only unbroken line of data involved the 1999 (i.e. grandparents), 2003 (parents) and 2007 (offspring) year-groups. The two bivariate genetic correlations from this line successfully converged and are shown at the bottom of Table 4.4. A very high r_g of 0.89 (s.e. 0.06) was obtained between the two single cages of year-groups 1999 and 2003, comparable in size and accuracy to that obtained between freshwater tank

replicates and not significantly different to unity. This was achieved despite the low prevalence and heritability obtained for cage *c7*. The estimate was improved over that obtained in Chapter 3, (0.72 s.e. 0.13) as a result of further genotyping leading to updating of pedigrees.

Table 4.4 : ***Genetic correlations between families replicated within year-group, and between test populations across year-groups where sufficient genetic links exist.***

	yeargroup cage / tank 1	yeargroup cage / tank 2	h^2_1	h^2_2	rg_{12}	s.e. rg_{12}	Guy (2009) rg_{12}
seawater	**1998_c1**	**1998_c2**	0.262	0.153	**0.869**	0.030	0.85
	2000_c4	**2000_c5**	0.390	0.187	**0.704**	0.045	0.70
	2000_c4	**2000_c6**	0.399	0.100	**0.754**	0.056	0.72
	2000_c5	**2000_c6**	0.190	0.101	**0.811**	0.048	0.81
freshwater	**2005_t8**	**2005_t9**	0.312	0.333	**0.824**	0.053	
	2006_t10	**2006_t11**	0.555	0.565	**0.952**	0.031	
	2007_t12	**2007_t13**	0.556	0.630	**0.796**	0.068	
between year groups	**1999_c3**	**2003_c7**	0.424	0.064	**0.893**	0.057	0.72
	2003_c7	**2007**	0.063	0.582	**0.545**	0.110	
	seawater	**freshwater**	0.429	0.511	**0.678**	0.093	

The genetic correlation between the single cage of the 2003 year-group and the 2007 replicate freshwater tanks (replicates were fitted by including tank as a fixed effect) was 0.55. Using all the available data, the final row of Table 4.4 shows the most accurate estimate we have from this data of the genetic correlation between freshwater and seawater IPN mortality. This was the result of fitting cage or tank as a fixed effect in a bivariate analysis between all seawater and freshwater data respectively, taking into account the extra data and genetic links that exist across the pedigrees. This yielded a genetic correlation of 0.68 (s.e. 0.09). Note that this is higher (0.678 vs 0.545), and with an improved standard error (0.093 vs 0.110) over the bivariate genetic correlation between the two sites upon which the seawater-freshwater comparison is based (2003_c7 and 2007). Although the standard errors are highest of all correlation estimates derived so far, the correlation is positive, high and significantly different from either unity or zero.

4.4 Discussion

4.4.1 Genetic variation for resistance

All the results presented here confirm a high and exploitable level of genetic variation for both seawater and freshwater resistance to IPNv mortality. Heritabilities obtained from seawater field testing on the observed scale in this study were variable but reached a high of 0.43 for the pooled data. Controlled experimental testing in freshwater revealed heritabilities up to 0.59 for individual sites, and the pooled estimate over all available freshwater data was 0.49. These are higher than those reported by Wetton et al., (2007) covering eight years of freshwater fry challenge (overall 0.31). The Norwegian data, however, fitted individual family tank as a random effect, included in and therefore inflating the denominator variance. This may have resulted in more conservative heritabilities compared to those reported here from mixed family tanks. The 2000 year-group in Wetton et al., (2007) consisted of mixed tanks with families recovered by genotyping and so was more comparable to the freshwater trials reported here. That year-group reported one of the highest heritabilities (0.39), consistent with those reported here.

When combining the seawater and freshwater data, we report a pooled heritability of 0.67 (s.e. 0.02), higher than any estimate obtained from any of the constituent data sets. This assumes seawater and freshwater IPN mortality can be analysed as the same trait, and accounts for the selection of the parents of the 2003 year-group, based on the 1999 year-group challenged families, and selection of 2007 year-group parents, based on 2003 year-group challenged

families. If the IPN challenge data on which parents were themselves selected is not included, such as with the individual year-group analyses, the possible reduction in the sire and dam variances (but not the mendelian segregation variance) would depress heritability estimates. The higher heritability obtained when freshwater and seawater data are combined therefore supports the suggestion that selection based on seawater sib challenge is realised in family performance of descendents tested in freshwater. This is the complementary scenario to that reported by Storset et al., (2007) who showed significant differentiation in seawater survival to IPN based on parents selected on the basis of freshwater IPN challenge.

Genetic correlations between family replicates in this study approached unity, even in datasets from field trials with low prevalence resulting in low heritabilities. This suggests that genetic resistance for survival to IPN is robust to a wide range of epidemiological challenge conditions. As mentioned above, two comparisons were available where families were genetically connected as parental and offspring generations : where both were seawater sites, the genetic correlation was effectively unity; where the sites were seawater and freshwater respectively the genetic correlation was still relatively high at 0.55 (s.e. 0.11). This represents the genetic correlation between the two life stages of infectivity, fry and smolt, freshwater and seawater. Being predominantly based on the comparison between the 2003 and 2007 year-groups, this estimate also accounts for the selection of parents as discussed above. Using all available data and links in the pedigree, the genetic correlation between IPN mortality regarded as a seawater trait and as a freshwater trait respectively, was 0.68 (s.e. 0.09). Wetton

et al., (2007) reported genetic correlations for two of the eight year-groups where IPN outbreaks in the field were subsequently experienced : that between fry challenge and IPN mortalities in freshwater parr following tagging was 0.83, between fry challenge and vaccinated smolts following post seawater transfer was 0.78. The heritability of IPN mortality in the field outbreaks was 0.1 in both cases which supports the view introduced in Chapter 2 that replication of families can reveal genetic variation otherwise hidden by that based on simple comparisons of family means. Storset et al., (2007) was also able to show that families observed as high or low average mortality from the 2000 year-group performed consistently when offspring were tested as fry (67 and 29 percent mortality, respectively), and as advanced smolts (79 and 32 percent mortality).

These figures support the view that resistance in freshwater and seawater may be largely a reflection of the same genetic mechanisms. Testing in freshwater, which is more timely (15 months earlier) and convenient (in experimental containment facilities), would therefore be expected to generate improvements in seawater performance. In the dataset reported here, it did not prove possible to challenge any one year-group in both seawater and freshwater. By accumulating pedigrees through a sustained breeding program and appropriate statistical analysis, however, the genetic correlation of 0.68 between freshwater and seawater resistance compares well with those obtained elsewhere and supports the view of common mechanisms operating in the two environments despite very large differences in disease development, epidemiology and the physiological status of the host.

4.4.2 Field and experimental testing for disease resistance

The intention of the seawater field trials was to expose the full range of (as yet unknown) resistance mechanisms that may be operating under commercial conditions and then target the resistant families in unchallenged broodstock when making selections. That principle still holds even though field exposure by definition can be difficult to control and sustain over the time periods required. Experimental trials afford greater control (Kjoglum et al., 2008) but perforce target a particular subset of resistance mechanisms determined by the challenge protocols adopted, particularly the virus's mode of entry to the host, and may miss some relevant mechanisms that are only expressed under field conditions. For example, as noted by Wetten et al., (2007), intra-peritoneal injection of IPN virus to generate infective cohabitants often resulted in very few appearing as mortalities. Polygenic heritabilities and genetic correlations as high as these, however, are not usually anticipated for an animal livestock disease (Bishop et al., 2010), and these values help explain how clear results have been so readily obtained in a variety of situations. IPN was a model trait for field testing because it is endemic, acts quickly at two separate stages of development and has a clear method of observational scoring. Other diseases tend not to have these characteristics and may not be quite so amenable to analysis.

4.4.3 Statistical approaches

All the results presented here are from analyses on the observed (0,1) scale, using linear mixed models and assuming normality. Thompson (1979) and Gianola (1980) point out that in theory, methodology for the continuous case is not suitable for the discrete case. Certainly recording any continuous trait using only two values would be expected to lose information compared to when all values across the range are available, and is functionally equivalent to recoding the two values as zero and one. Lower apparent heritabilities then result due to additive genetic variance being progressively transferred to the non-additive fraction, and hence the residual term, as the frequency of one value increases over the other (Robertson and Lerner, 1949). A transformation to an underlying normal distribution is naturally suggested where the frequency of each value across a population represents the area under a normal curve truncated by a threshold measured along the x-axis. Threshold models, first proposed by Wright (1934) and placed in the mixed model context by Gianola and Foulley (1983) therefore map the observed (0,1) discrete data onto an underlying, unseen but normally distributed 'liability' scale. It was shown in Dempster and Lerner (1950, quoting Robertson) that genetic parameters obtained on the observed scale can be approximately transformed to those obtained on the liability scale by multiplying by $\boldsymbol{p(1-p)/z^2}$ or equivalently by $\boldsymbol{(1-p)/i^2p}$ where $\boldsymbol{z}$ is the height of the ordinate and $\boldsymbol{i}$ is the corresponding mean of the proportion $\boldsymbol{p}$ cut off by the threshold. The lowest value of the multiplicand occurs at $\boldsymbol{p}$=0.5 and is 1.571, i.e. always greater than one ensuring that liability scale heritabilities are greater than on the observed scale. We are reminded however by Gianola 1982 that genetic

parameters transformed from the observed scale to the underlying scale are not entirely free of the bias that is due to non-additivity.

Matos et al., (1997a), in a comparison of non-linear models applied to sheep reproductive data, hypothesised that threshold models seem appropriate for discrete data and may capture a higher proportion of the additive genetic variance than is possible with linear models (ie using the observed scale) and that selection criteria based on threshold models may lead to better predicted genetic gain. While it could be shown that the advantage of threshold over linear models increased as heritabilities decreased, they concluded that in terms of goodness of fit and predictive ability (Matos et.al., 1997b) there was no advantage of threshold over linear models and that the failure of threshold models to demonstrate superiority over linear models was observed in several other studies.

Non-linear models, including the logit and probit transformations to the underlying liability scale, are implemented in ASReml 3.0 (Gilmour et.al., 2009) by a class of Generalised Linear Models (GLMs) which include random effects according to Schall (1991) and more appropriately called Generalised Linear Mixed Models. GLMs use a transformation of the vector of observations $\boldsymbol{y}$ to an adjusted dependent variable $\boldsymbol{z}$, by applying a 'link function' $\boldsymbol{g(.)}$ to the mean $\boldsymbol{\mu}$, where

$$\boldsymbol{z = g(y) = g(\mu) + (y - \mu)\, g'(\mu)}$$

The linear random effects model for z is then

$$\boldsymbol{z = X\beta + Ub + eg'(\mu)}$$

which ASReml can solve by likelihood iteration using algorithms already programmed. While the method is perfectly general with respect to the link function and number of dispersion (variance) components (Schall, 1991), as pointed out by Sawalha et al., (2007) the likelihoods must be approximated since there is a lack of explicit expression for the likelihood of these models except as an approximate integral. When applied to the data throughout the work described here, GLMMs using the logit link only rarely, and with the probit link, never, iterated successfully to converged values. The estimation of heritability of all the combined IPN data was repeated using a GLMM with logit link, successfully converging to a value 0.252, considerably less than that obtained on the observed scale (0.67, Table 4.3) and unlikely to be correct. Applying Robertson's transformation to the observed scale estimate takes the heritability to approx 1.6. The conclusion is that while the underlying liability scale is very desirable for several good reasons, the behaviour of the likelihood models at these high levels of heritability is generally too unstable and numerically problematic to obtain reliable estimates. Odegard et al., (2010b) explored the inherent bias to heritabilities in threshold models due to having only one binary observation per animal, and due to the possibility of fixed categories containing only a single category of response variable, leading to individuals which can be regarded as 'non-informative'. Odegard proposes a novel algorithm based on Gibbs sampling is proposed which seems to overcome these problems.

Typically, quantitative genetic investigations on livestock diseases are based on field data or alternatively experimental data based on restricted numbers, and estimated heritabilities are consistently low even though the

underlying mechanisms such as immune parameters may display strong genetic control. Bishop et al., (2010) showed there is a downward bias to heritabilities by failure to account for quantifiable epidemiological determinants such as exposure rates to the pathogen or misclassification of response all of which add to the environmental variance. They show in a re-interpretation of the seawater IPN data presented here (figure 5 of Bishop et al., (2010)), that the strong linear dependence of observed-scale heritabilities on prevalence (as observed here in Table 4.3) was not completely removed by transformation to the liability scale (i.e. $\boldsymbol{p(1-p)/z^2}$). This dependence disappeared, however, when differences in relative exposure probabilities were hypothesised and the induced biases were removed. The resulting heritabilities were consistent and greater than 0.70 and indicative of the major QTL (section 4.4.5) now known to be segregating in these populations.

4.4.4 Evidence for realised selection

There were three clear phases to the challenge testing - (i) four years, 1999 to 2002 (year of event), where seawater epidemics were well established and mortalities recovered in high numbers, (ii) four years (2003 to 2006) where seawater epidemics either failed to establish or mortalities were recovered in low numbers and (iii) five years (2006 to 2010) where freshwater epidemics were well established but with gradually declining mortality rates for reasons that are as yet unclear. The first pilot field trial in 1998 (year-group 1996, not shown) and the first full scale trial in 1999 (shown in Table 4.1) were not included in these analyses although extreme families were readily identified and used in selection.

Therefore, significant seawater mortalities, sufficient to inform selection decisions, only came from the first four years of challenge testing. Families created in the final four years of seawater testing (2001 to 2004 year-groups, with unexpectedly low or zero prevalence) were therefore each subject to a single round of selection, the parents being selected sibs of IPN challenged populations. Alternatively, with freshwater data, only the 2007 year-group had parents with associated IPN data, (2003 year-group with seawater challenge). For the other freshwater-tested year-groups, only seawater data from the grandparental level was available. The extent to which selection may be the cause of the low observed mortality is difficult to separate from the other factors that were identified (section 4.3.2) and is subject to further investigation. Houston et al., (2010) in a parallel study, repeated the 2007 year-group challenge on a selected subset of 20 families reared separately in freshwater tanks. Four susceptible, twelve intermediate and four resistant families were selected based on family EBVs derived from the 1999 and 2003 year-group seawater IPN challenges. Average mortalities after 44 days post-challenge were 15.5, 8.6 and 4.5 percent (average 10 percent) respectively. Nine of the families showed negligible mortality and seemed to be completely resistant to the challenge. Storset et al., (2007) similarly showed a response in seawater to one generation of divergent selection of high and low resistant families, based on freshwater data (32 and 79 percent respectively). All the results discussed show a far greater response to selection than expected with a purely polygenic trait, and strongly suggest the plausibility of a major quantitative trait loci (QTL)

segregating in families of both the Scottish (Landcatch) and Norwegian (AquaGen) breeding programs, and that is indeed the case.

4.4.5 Evidence for segregation of QTLs

Use of polygenic (parental average) EBVs for the selection of candidate sibs that were themselves unchallenged, exploits only between-family variation, which is usually approximately half of the genetic variation available, since all members of a family receive the same parental average EBV. Discrimination between individuals within full-sib families requires either a correlated polygenic trait or the identification and tracking of segregating genomic markers linked to Quantitative Trait Loci (QTLs) associated with resistance.

The opportunistic availability of DNA from both mortalities and survivors, along with the judicious selection of families, has allowed a search for genomic markers to progress in parallel with the investigation of polygenic inheritance described so far. Houston et al., (2008a), Houston et al., (2008b) investigated 19 families of intermediate resistance from the 1999 year-group challenged with the field outbreak of IPN in seawater reported here. The DNA used was originally collected for parentage assignment on fish mortalities and harvested survivors. A whole genome scan using 72 microsatellite markers revealed for the first time the involvement of a major QTL on linkage group 21 which explained ~21 % of the within family phenotypic variation in IPN resistance. Gheyas et al., (2010a) confirmed these findings in 20 families with intermediate prevalence from the unrelated 2005 year-group freshwater IPN data

reported in Table 4.3. Houston et al., (2010) confirmed the involvement of the same QTL in 10 families from the 2007 year-group (i.e. 2 generation descendents of the 1999 year-group) in a separate challenge experiment with freshwater IPN to that reported in Table 4.3. In a mixed model partitioning the polygenic and QTL effects, polygenic component was estimated to be negligible. Essentially all the genetic variation in these 10 families was due to the QTL, with some evidence of dominance for the resistant allele, not seen in the seawater challenge or 2005 year-group data. Moen et al., (2009) confirmed the segregation of what unequivocally appears to be same QTL in 10 families from a Norwegian breeding program, first in post-smolts, then confirmed in fry samples from the same year-group. The QTL appeared to explain ~28 % of the phenotypic variance and 83 % of the genetic variance. Haplotypes of the 4 markers closest to the QTL are reported to have population wide correlation with the QTL, suggesting that it could be used between as well as within families to identify selection candidates with the favourable QTL allele.

4.5 Conclusions

Three questions were posed at the beginning of this work :

(i) is there significant host genetic variation in IPN resistance based on freshwater challenge tests

(ii) is IPN resistance consistent between family replicates and

(iii) is freshwater IPN mortality consistent with seawater IPN mortality to the extent they may be the same trait ?

These trials were initiated in the expectation that there was exploitable genetic variation in host resistance to IPN. That has clearly been confirmed by these results and those reported elsewhere. The estimate of heritability pooled over all year-groups was 0.67. This is unusually high for a disease trait and the involvement of a QTL that explains most if not all of the genetic variation has been subsequently confirmed in both this and one other strain of farmed Atlantic salmon.

Genetic correlations between family replicates in this study approached unity, even in datasets from field trials with low prevalence resulting in low heritabilities. This suggests that genetic resistance for survival to IPN is robust to a wide range of epidemiological challenge conditions. The genetic correlation between the disease expressed in seawater and freshwater based on all replicate and linked populations was 0.68. This was achieved without the availability of a direct, within year-group comparison of freshwater and seawater IPN on the same full-sib families. It is therefore sufficiently high to suggest that resistance to seawater and freshwater IPN may be regarded as the same trait at least for selection purposes. Similar effect in both seawater and freshwater challenge has subsequently been confirmed for the QTL effect (Gheyas et al., 2010a)

There is no evidence yet that the IPN resistance QTL refers to either a single gene, or a tightly linked gene complex or indeed what its function might be. It is confirmed however to be functional at both the seawater, smolt stage and freshwater fry stage. The smolt stage is a period when the immune system is compromised as the fish adapt to seawater. The fry stage of challenge is before a functional immune system has been established (Roberts and Pearson 2005).

Mortalities then cease and the fish effectively become carriers at the time the immune system starts to become established. This strongly suggests, as pointed out by Moen et al. (2009), that the function of the QTL may well turn out to be related to the innate rather than the adaptive immune system. While this would explain the lack of association found so far between IPN susceptibility and immune function, and partly explain why IPN vaccines and genetic resistance appear to be additive (Ramstad et al., 2008), it would not rule out mechanisms of resistance associated directly or indirectly with physiological aspects of fish that are growing while exposed to the farmed environment. That possibility is explored in the final investigation of this study, described in Chapter five.

Chapter Five

Genetic correlations between IPN mortality and performance traits at harvest.

5.0 Abstract to the chapter

Despite the availability of vaccines, the Infectious Pancreatic Necrosis virus (IPNv) continues to cause high mortalities in both freshwater and seawater stages of Atlantic salmon. Within the Landcatch breeding program, a high level of host genetic resistance to mortality from the virus has been previously reported with heritabilities of 0.49 (s.e. 0.03) in freshwater, 0.43 (s.e 0.02) in seawater, and 0.67 (s.e. 0.02) for data combining the two stages and ten year-groups. Consequently, IPN has been under active selection in the Landcatch breeding program for the past 10 years. To investigate if there were favourable or antagonistic relationships operating between the traits under active selection, genetic correlations between IPN mortality and a range of performance and harvest traits, collected on sibs to those tested for IPN, were investigated. Only 7 out of the 44 genetic correlations that were investigated approached significance at the 5 % level compared with their standard errors. Two colour traits, hue (r_g = 0.37 s.e. 0.16) and intensity (r_g = 0.54 s.e.0.14) had significant and favourable genetic correlations with seawater IPN mortality, but these and the others mentioned were not confirmed by any strong consistency over the various

datasets. Only one antagonistic correlation with harvest weight (0.30 s.e 0.11) attained significance with freshwater IPN mortality. When the IPN and harvest data from all year-groups were combined, these became non-significant. Taken as a whole, these results indicate that selecting salmon for resistance to both seawater and freshwater IPN challenge certainly is feasible, and that adverse effects on selection for other important production traits is not expected. How these medium to high heritabilities relate to the discovery of a major QTL for IPN resistance segregating in these populations, is discussed.

5.1 Introduction

The heritability and genetic correlation parameters estimated for IPN resistance in the previous chapter have informed selection in the associated LNS breeding program based on mixed model EBVs since the trials began. Likewise, markers for IPN QTLs have formed the basis for Marker Assisted Selection (MAS) evaluations since their discovery in these populations in 2006. Given the devastating nature of the disease and the size of the QTL effect, there may well be pressure on breeding programs to bring the resistance allele of the QTL to, or near to, fixation as rapidly as possible (Moen et al., 2010) since this would allow a breeding company to supply fully resistant stock without the expense of further testing. This raises the question as to why the unfavourable (susceptible) QTL appears to have been maintained in farmed populations prior to active selection against it (Moen et al., 2009). One possible reason may be that IPN susceptibility could be negatively correlated with one or more traits relevant to aquaculture, and yet remain neutral in the wild. For that reason genetic progress

in IPN resistance needs to be carefully managed in relation to maintaining genetic variation and selection targets for other traits. As a pre-requisite it becomes important to show how IPN resistance may be genetically correlated with other traits under selection.

The evidence for substantial genetic variation in growth and flesh quality (Guy et al.,2005; Quinton et al., 2005; Powell et al., 2008), lipids (Leaver et al., 2010; Morais et al., 2010; Bell et al., 2010), skeletal deformity (Sullivan et al., 2007) and early maturity (Gjedrem, 2005) is well established, and the moderate to high heritabilities are utilised to various extents in current selective breeding schemes (Gjedrem, 2005). Powell et al., (2008) presented results for a series of bivariate analyses of performance traits taken on broodstock and harvest sibs from the LNS 2001 year-group, although not including skeletal deformities, early maturity or disease traits. Evidence for the involvement of one or more QTLs affecting performance traits in farmed populations is also accumulating (Reid et al., 2005; Boulding et al., 2008; Houston et al., 2009b; Baranski et al., 2010).

An unfavourable genetic correlation between IPN resistance and one or more of these other traits targeted for selection could have at least two consequences : (i) the choice of candidates favourable for both traits would be much reduced, restricting the rate of genetic gain that can be made in either, and (ii) the prediction of genetic gain in correlated traits will be inaccurate and biased if the correlation is ignored. The amount of restriction experienced by each trait will depend on the size of the genetic correlation and the relative importance (economic weighting) of each trait. A genetic correlation of

whatever sign, favourable or unfavourable can be used to advantage where one trait is not measured directly on candidates while the other trait is. For a trait measured on relatives only, candidates for selection will be allocated the family mean EBV and there will be no discrimination within family. If evaluated as a genetically correlated bivariate or multivariate trait, candidates will receive their own evaluations on a trait for which they have not been measured, and an additional degree of within family selection becomes possible. Marker assisted selection is a special case of this general principle.

Therefore the objective of the final scheme of work for this PhD study was to screen the available year groups for suitably balanced combinations of IPN mortality and harvest sentinel data, then to conduct bivariate mixed model analysis to identify genetic correlations that may be operating between IPN mortality and performance traits.

5.2 Materials and Methods

5.2.1 Description of test populations

Data consisted of mortality records and harvest performance on families comprising all the 10 year-groups from 1998 to 2007 (Table 4.1, reiterated in Table 5.1). Year-group denotes year of stripping of the parents (October to December) and subsequent fertilisation of the families. The thesis Introduction described (i) how 'sentinel' test populations were constructed for each year-group and placed on customer sites for disease challenge and possible later

sampling at harvest, and (ii) how each sentinel population comprised all 200 breeding program families reared under commercial 'field' conditions.

Field data being dependent to a large extent on opportunity, there was variation to the intended plan as was shown in Table 4.1 and again in Table 5.1. Particularly we have seen there were three year groups (2001, 2002 and 2004) which failed to elicit any accountable IPN mortalities, and consequently, from year-group 2005 onwards, the IPN field challenge was replaced by experimental challenge in closed facilities. Harvest data were available for each year-group except 2006 and the numbers sampled each year varied between 1,000 and 5,000, depending on resources available for processing the samples. Table 5.1 also shows how the IPN data related to the harvest data for each year-group. The IPN data was as described in Chapter 3 except 'number stocked' is the total placed on site and not the effective number challenged of Table 3.2. The harvest sentinels in the first four year-groups, (1997-2000), were sampled from the survivors of the IPN event which occurred the previous year in the same cages. The 2001 year-group were similarly sampled from the same cage, although the site did not experience an IPN event at any stage. In all subsequent year-groups (2002-2007) fish sampled for harvest traits were randomly extracted in equal numbers of each family from the broodstock population as identified by pit tag (noted as Ex-BP in the last column of Table 5.1)

In summary, (i) IPN sentinels and harvest sentinels were the same families if in the same year-group; (ii) harvest sentinels were survivors of the IPN event if they were the same cage, and (iii) harvest sentinels were surplus broodstock sibs if denoted 'ex-BP'

1 ***Table 5.1 Populations contributing IPN and harvest data***

2

	IPN sentinels					Harvest sentinels			
Year-group	**Year of event**	**site**	**cages / tanks**	**Reported % IPN mortality**	**Number stocked**	**Year of harvest**	**Harvest cage**	**Number harvested**	**status of harvest**
seawater									
1997	1999	S0	c0	22.7	17,006	**2000**	c0	1,014	survivors
1998	2000	S1	c1, c2	16.4	54,844	**2001**	c1, c2	4,974	survivors
1999	2001	S1	c3	30.0	53,179	**2002**	c3	5,000	survivors
2000	2002	S2	c4	14.8	54,821				
2000	2002	S3	c5	11.9	16,098				
2000	2002	S4	c6	10.3	5,867	**2003**	c6	4,992	survivors
2001	2003	S4	c20	0	5,000	**2004**	c20	3,171	sentinels
2002	2004	S2		0	55,000	**2005**	c21	1,837	ex-BP
2003	2005	S5	c7	3.2	52,857	**2006**	c22	1,372	ex-BP
2004	2006	S6		0	60,000	**2007**	c23	1,118	ex-BP
freshwater									
2005	2006	S7	t8, t9	0.604	6,215	**2008**	c24	1,935	ex-BP
2006	2007	S7	t10, t11	0.404	4,846	**2009**			
2007	2008	S7	t12, t13	0.286	5,248	**2010**	c25	2,000	ex-BP
2008	2009	S7	t14, t15	0.206	5,874				
2009	2010	S7	t16, t17	0.219	5,450				

5.2.2 Description of data

5.2.2.1 IPN mortality

Mortality and survival to each IPN challenge was recorded on each individual animal in the disease sentinels as 1 and 0 respectively. IPN in freshwater (***IPNfw***) and seawater (***IPNsw***) were first considered separately and then together. Mean prevalence of mortality by year-group and site combination was summarised in Table 5.1 above. Mean prevalence of mortality by year-group and cage combination was as given by Table 4.2 (freshwater) and Table 4.3 (seawater) of Chapter 4.

5.2.2.2 Harvest traits

Table 5.4 lists in the lower section all twelve traits recorded on the sentinel population at the time of harvesting. Powell et al, (2008) described in detail trait collection from the 2001 year-group which applied generally across all year-groups. Methods of collection were subject to technical development over the period, but related to a core list of component weights, yields (relative weights), instrumental lipid assessment, and fillet colour all taken at the same harvest event. Weights and lengths (***harvwt, harvlen***) collected as harvest sentinels were processed, at around two and a half years of age, after approx 14 months rearing in the sea. Fat (***fatpc***) was scored as estimated percent of the gutted fish using the mean of 8 readings taken at various points along the body using a Distel Fatmeter (Kent, M., 1990). The fatmeter used microwave attenuation to read water content, which is inversely

proportional to fat percent of the tissue and is displayed on the instrument following application of a species-specific standard curve calibrating water against fat. In the 2003 year-group harvest, a Fatmeter was not available and a Near Infra-Red Reflectometer, based on the same principle of assessing water content and standardising against a calibration curve was employed (***nirfats***). As harvested fish were gutted, the component parts (***guttedwt, filletwt***) and yield as a percent of the pre-processed fish weight giving ***guttedyld*** (*gutted wt/harvwt*)**,** ***filletyld*** **(***filletwt / harvwt***)**, and ***guttedfilletyld*** (*filletwt / guttedwt*) were determined.

Routine broodstock recording was carried out over the period of this study, where weights, deformities, and maturity rates also became available to be included in an overall analysis (*section 5.2.3.2, analysis [B]* below). Performance traits measured on live broodstock included weight (*wt*), length (*len*) and condition factor *(cf = weight / length^3)* taken as growing fish were transferred to larger tanks, once a year January to March. Measurements were taken at the time smolts were transferred to sea, 12 months post-hatching (***suwt, sulen*** and ***sucf***, respectively), then after one winter in seawater, approx 12 months post-transfer (***sw1wt, sw1len, sw1cf***). In recent years fat percent was also taken on live broodstock at the one-seawinter moves (***sw1fat***). Broodstock were also visually scored for ***deformity*** as a binary trait (0 and 1 for unaffected or affected respectively), at the one-seawinter moves, and early maturity **(*grilse*)** as a binary trait when seen to mature early at three years old (1), or later (0).

5.2.2.3 Colour measurement

A uniform deep pink flesh colour in salmon and trout is taken as an indicator of quality and is critical for consumer acceptability of fresh or smoked products. The colour derives from the carotenoids particularly astaxanthin obtained from crustaceans in the natural diet of wild fish or nature-identical manufactured pigments added to feed given to farmed fish. The pigment functions as an anti-oxidant, stored in the flesh as fish grow and transferred to eggs as females sexually mature. Poor flesh colour at harvest is therefore sometimes indicative of poor feeding or precocious sexual maturity and leads to downgrading of products. Genetic variation with medium heritability (0.1 to 0.2) has been observed for the visual assessment of colour related to the absorption and retention of colour pigments (Gjedrem T, 2005) and is a component objective for selection in the LNS breeding program.

Fillet colour was taken on harvested fish, using either visual score related to a colour chart or instrumental readings taken from either a Minolta Chroma Meter (www.konicaminolta.eu) or digital camera images. Visual scores ranged from 20-34, yellow to red respectively, compared against the industry standard Roche colour chart (***salmofan***). Instrumental readings and digital images were recorded on the HSI model as ***hue*** (1 to 90^{o}***,*** red to yellow respectively, lower is better) ***saturation*** 0 to 255 (higher is better) and ***intensity*** (0 to 255, lower is better). The eye mainly perceives differences in pigment concentration as saturation when assessing fillet quality which tends to white in salmon lacking pigment, (eg those reared under an organic regime). Image-Pro Express software (MediaCybernetics

Inc, Bethesda USA, www.mediacy.com) was used to read images and output HSI parameters. How colour was measured and transformed to the HSI model is given in Appendix 5.1 below since there appears to be no accessible written description in the aquaculture literature.

5.2.3 Selection of datasets

Harvest data from the first four year-groups (1997-2000, Table 5.1) was derived from survivors of the earlier IPN event. In effect, when correlating disease and harvest traits, we are missing harvest data on the IPN mortalities due to the timing of the two events. This introduces some uncertainty when interpreting what correlations one might expect from such data. For example, it is not clear how the stress of the IPN challenge on surviving post-smolts may have impacted on the regulation of harvest traits, and surviving fish may have allocated energy towards immune response at the expense of growth (Houston et al., 2009). Therefore this difficulty was avoided by first focussing on two subsets of data, IPN in seawater [A], and freshwater [B], that were consistent and more likely to be unbiased, before combining all year-groups [C] as follows :

5.2.3.1 Seawater IPN (1999 and 2003 year-groups) versus Harvest traits (2003): dataset [A]

The strongest and most consistent IPN dataset in terms of data collection, linked pedigrees and subsequent analysis is represented by the grandparent -

parent - offspring (1999-2003-2007 year groups respectively). The harvest data was taken from 1372 random culls (av 6.7 per family) from the Ormsary broodstock in 2006 (year-group 2003), therefore not subject to an IPN event, but was the same year-group as the last of the seawater IPN challenges. This combination was chosen to investigate the genetic correlation of harvest traits with seawater IPN.

5.2.3.2 Freshwater IPN (2005 + 2007) versus Harvest traits (2005), dataset [B]

The only harvest data available for analysis at the time of writing that is associated with full-sib families challenged with IPN in freshwater, is the 2005 year group, with 1935 sibs, approx 10 per family. Freshwater IPN challenge data from the 2007 year-group was added since it represents the only existing link with seawater IPN data, completing the connection to the 1999-2003-2007 cohort mentioned in analysis [A].

5.2.3.3 All IPN year groups (1998 - 2007) versus all harvest year groups (1997-2005) : dataset [C]

IPN and harvest data from all available year-groups was used, which conforms with the routine annual analyses from which broodstock selections are made. Since the sampling of fish and families is not ideal in all datasets, it is of interest to see if results are consistent with datasets A and B.

5.2.4 Analysis of data

All datasets were analysed for genetic parameters by applying linear mixed models to mortality data on the observed (0,1) scale and a range of harvest and performance traits. An Individual Animal Model (IAM, see Mrode, 2005) was implemented using restricted maximum likelihood (Patterson H D, and Thompson R, 1971) based on the Average Information algorithm (Gilmour et al., 1995), as implemented in ASReml 3.0 (Gilmour et al., 2009). For each dataset, significant fixed effects, univariate variance components and heritabilities were first determined for each trait investigated. These models were then implemented for each separate trait in a bivariate model containing IPN mortality as the first trait and a harvest trait as the second, using the univariate variance components obtained previously as starting values. Where convergence became problematic or updates kept going out of bounds, ASReml allowed the starting variances to be fixed at their univariate values, iterating the genetic covariances alone. In such cases, convergence was confirmed by applying the iterated values as starting values in an unrestricted re-run of the models.

Bivariate animal models were used to estimate the trait variances and covariances for each pair of traits, where trait 1 was always IPN mortality and trait 2 a harvest trait. Asreml allows trait specific fixed effects and can accommodate missing data for each trait. The model fitted for IPN mortality was :

$y_{ijk} = \mu + G_i + TY_j + a_{ijk} + e_{ijk}$ where :

y_{ijk} = observation of mortality (1) or survival (0) on animal ***k***,

μ = trait mean

G_i = the fixed effect of genetic group ***i,***

TY_j = the fixed effect of tank (or cage) and year-group combination ***j***

a_{ijk} = the random additive genetic effect of animal ***k,*** and

e_{ijk} = the random residual error for animal ***k***

The model fitted for performance and harvest traits was :

$$y_{ijkl} = \mu + G_i + TY_j + (TY.S)_{jl} + a_{ijkl} + e_{ijkl}$$

where additionally, ***S*** = the effect of sex ***l***, within tank and year-group, when recorded.

Performance traits ***grilse*** and ***deformity*** were also measured as 1 (affected) and 0 (not affected). For IPN mortality, fixed effect categories consisted of tank (or cage), which also took account of year-group and seawater / freshwater differences since they were uniquely labelled (Table 5.1). For harvest and broodstock performance traits, fixed effects were fitted as an interaction of year-group and tank (cage), and also sex where known and appropriate. A single pedigree file covering all year-groups and all known ancestors was used for all analyses. Genetic groups (families with unknown parentage categorised by year group) were fitted to the pedigrees to account for separate origins of the four consecutive year-groups that constitute one generation, and to account for animals that had data but missing parentage appearing in each year-group.

5.3 Results

Results for each of the three analyses conducted *(A-C)* are presented in Tables 5.2, 5.3 and 5.4 respectively, giving a description of the traits (number of records and mean trait values,) heritabilities and genetic correlations obtained, with associated standard errors. Heritabilities for the IPN trait are presented just once, consistent with the univariate estimates used as starting values. Significant deviations of the bivariate heritabilities from their univariate estimates were taken as supportive evidence of poor convergence. Estimates deemed not to have converged were replaced in the tables with 'nc' since further interpretation would be unjustified. An estimate was interpreted as being statistically significant at alpha=0.05 if the ratio of the estimate to its standard error was greater than two. This is only an approximation since the standard error estimated by ASReml is itself an approximation and also the true confidence intervals around the parameter estimate may actually be asymmetric.

5.3.1 Seawater IPN and Harvest traits [A] :

The seawater IPN data consisted of the two genetically connected (parent-offspring) year-groups and, as before, the heritabilities were high (Table 5.2). Heritabilities for the weight-based harvest traits (***harvwt, harvlen, guttedwt, and filletwt***) were high, approaching 0.5 and are consistent with those found generally in the routine breeding program evaluations using all data. Genetic correlations with seawater IPN mortality

were effectively zero, given the high standard errors. An unfavourable correlation with ***harvlen*** (0.27) just attains significance, although being unsupported by ***harvwt***, is unlikely to be important.

At the 2003 harvest, fat percent was measured by NIR rather than by fatmeter. Despite the heritability of ***nirfat*** being close to zero, (0.014) the genetic correlation with IPN mortality converged to a moderate and favourable 0.45, albeit very imprecisely estimated (s.e. 0.48). There was no ***salmofan*** data but HSI model estimates of colour showed moderate and significant heritabilities (0.16 to 0.23), revealing a significant and favourable genetic correlation of IPN mortality with ***hue*** (0.37, se. 0.16) and with ***intensity*** (0.54, se 0.14). This is not confirmed by the unfavourable (but not significant) correlation with ***saturation*** (0.14 se 0.16) however, suggesting there may be some numerical issues remaining with these particular traits. All three yield traits failed to show any heritability, in common with all earlier findings and as discussed by Powell et al., (2008).

Table 5.2 : Seawater IPN (1999 and 2003 year-groups) versus Harvest traits (2003) [A]

Trait	units	records	mean	h^2	s.e. h^2	r_g	s.e. r_g
Seawater IPN #	0 - 1	67,932	0.092	0.767	0.037		
Harvwt	kg	1,339	4.00	0.481	0.069	-0.071	0.134
Harvlen	cm	1,362	66.58	0.434	0.064	**0.270**	0.128
guttedwt	kg	1,323	3.57	0.466	0.068	-0.026	0.`136
Filletwt	kg	1,256	2.92	0.526	0.070	-0.015	0.138
Nirfat	%	1,228	13.99	0.014	0.021	0.448	0.477
Hue	0 - 90^o	1,233	10.20	0.157	0.044	**0.367**	0.160
saturation	0 - 256	1,233	178.53	0.232	0.054	0.143	0.164
intensity	0 - 256	1,233	82.16	0.175	0.045	**0.543**	0.139
guttedyld	kg / kg	1,299	0.88		nc		nc
Filletyld	kg / kg	1,233	0.70		nc		nc
guttedfilletyld	kg / kg	1,251	0.80		nc		nc

 # Traits as defined in section 5.2.2.2 ; nc = not converged, estimates blank.

5.3.2 Freshwater IPN and Harvest traits [B]

The two year-groups with combined freshwater IPN mortality data were not directly linked by pedigrees and the heritability was reduced somewhat compared to the seawater IPN situation, although still high and significant (Table 5.3). Heritabilities for the growth related traits from the 2005 year-group were comparable to those seen in the 2003 year-group but also a little lower. Only one correlation exceeded twice its standard error (***harvwt**,)* but was unfavourable (*0.30, s.e. 0.11).* Fat percent (***fatpc***) using the fatmeter achieved a moderately high heritability, and the genetic correlation with IPN mortality was very close to zero. Unlike the results for IPN in seawater, ***hue, saturation*** and ***intensity*** failed to converge, with genetic variation tending to zero. ***Salmofan*** data existed for this year-group, but was observed and scored from displayed digital images, and the mixed model failed to converge, with the genetic variance tending to zero. As expected, yields also failed to display any genetic variation, although the heritability of fillet weight as a percent of gutted weight (***guttedfilletyld***) did become marginally significant at 0.08 s.e. 0.04. The heritabilities tending to zero were sufficient to prevent the genetic correlations from converging.

Table 5.3 : Freshwater IPN (2005 + 2007) versus Harvest traits (2005) [B]

	Units	records	mean	h2	s.e. h2	rg	s.e. rg
Freshwater IPN	0 -1	11,462	0.458	0.424	0.030		
Harvwt	kg	913	3.43	0.424	0.030	**0.301**	0.109
Harvlen	cm	1,832	68.73	0.422	0.058	0.159	0.106
guttedwt	kg	1,825	3.06	0.440	0.059	0.190	0.103
Filletwt	kg	1,221	2.06	0.387	0.058	0.119	0.118
Fatpc	%	1,928	11.12	0.398	0.055	-0.020	0.107
salmofan	20-34	1,002	31.00		nc		
Hue	0 - 90^{o}	1,005	9.66		nc		
saturation	0 - 256	689	152.69		nc		
intensity	0 - 256	689	116.90		nc		
guttedyld	kg / kg	803	0.89	0.002	0.045		nc
Filletyld	kg / kg	814	0.60	0.000	0.027		nc
guttedfilletyld	kg / kg	1115	0.67	0.085	0.039		nc

 # Traits as defined in section 5.2.2.2 ; nc = not converged, (estimates are blank).

5.3.3 All IPN year groups and all harvest year groups [C]

Heritabilities and genetic correlations from combining all IPN data sets and all harvest data sets are presented in Table 5.4. Trait definitions are as given in section 5.2.2.2.The parameters used as starting values for IPN mortality, combining the seawater and freshwater data, were those established in *Chapter 4.*

For performance traits there were approx 120,000 records available over the 10 years of recording in the broodstock population, and approx 20,000 records collected on harvest traits. Fat percent on live broodstock taken at the one-sea-winter moves (***sw1fat***) has only recently become available (7,449 records, year-groups 2004 to 2007). Since ***Grilse*** were identified as maturing fish (coded 1) following the stock reductions at one-sea-winter, the total records, including those going on to become mature at two-sea-winter were only half those seen at one-sea-winter. Heritabilities were medium to high for all broodstock recorded traits and also for harvest traits related to weigh and growth. Of these traits, only weight at the smolt moves (***suwt***, transfer to sea) had a genetic correlation with IPN mortality approaching significance. Although numerically small it was favourable, families with smaller fish being associated with higher mortality (-0.09 s.e. 0.03). Useful moderate heritabilities were also obtained for ***sw1fat***, (0.39) ***deformities*** (0.18) and ***grilse*** (0.21). Of these, only ***deformities*** had a favourable genetic correlation approaching significance (-0.09 s.e. 0.04) with lower deformities associated with lower IPN mortality.

Table 5.4 : Combined all years, bivariate analysis of IPN mortality with a range of broodstock performance and harvest traits.

	units	records	mean	h^2	s.e. h^2	r_g	s.e. r_g
IPN all years	**0 -1**	**141,662**	**0.147**	**0.670**	**0.019**		
broodstock[#] :							
Suwt	gm	154,046	123.47	0.563	0.010	-0.030	0.030
sulen	cm	119,150	21.21	0.620	0.013	**-0.085**	0.033
Sucf	gm/cm^3	119,146	1.20	0.117	0.006	-0.013	0.044
sw1wt	kg	117,574	2.68	0.522	0.013	-0.045	0.034
sw1len	cm	117,568	57.89	0.576	0.014	-0.041	0.035
sw1cf	gm/cm^3	117,562	1.35	0.370	0.013	-0.033	0.038
sw1fat	%	7,449	12.53	0.390	0.030	-0.016	0.061
deformity	0 - 1	117,640	0.10	0.180	0.008	**-0.085**	0.040
Grilse	0 - 1	49,649	0.09	0.205	0.011	0.100	0.054
Harvest							
Harvwt	kg	22,705	2.60	0.539	0.021	0.004	0.043
Harvlen	cm	18,740	61.20	0.436	0.021	-0.056	0.054
guttedwt	kg	22,690	2.38	0.521	0.020	-0.015	0.043
Filletwt	kg	20,454	1.88	0.526	0.022	0.002	0.044
Fatpc	%	19,765	11.04	0.165	0.013	-0.044	0.056
Nirfats	%	1,430	13.60	0.032	0.027	0.340	0.316
salmofan	20-34	6,735	29.22	0.736	0.042		nc
Hue	$0 - 90^o$	6,589	14.19	0.104	0.013	-0.097	0.102
saturation	0 - 256	4,708	161.05	0.209	0.029	-0.151	0.113
guttedyld	kg / kg	19,901	0.87	0.034	0.007		nc
Filletyld	kg / kg	18,442	0.68	0.063	0.009		nc
guttedfilletyld	kg / kg	19,259	0.78	0.079	0.009	**0.121**	0.068

 # as defined in section 5.2.2.2; nc = not converged (estimates are blank)

Hue and ***saturation*** showed similar moderate but useful heritabilities, while ***salmofan,*** (heritability 0.74, s.e. 0.04) is suggestive of structural issues in the data, given that the genetic correlation with IPN mortality also failed to converge. The three yield traits (***guttedyld***, ***filletyld*** and ***guttedfilletyld***) all achieved low but significant heritabilities (0.03, 0.06 and 0.08 respectively) in this larger data set. The genetic correlations of these yield traits with IPN mortality failed to converge as expected except ***guttedfilletyld*** which was unfavourable but not quite attaining significance (0.12 s.e. 0.068). Apart from those highlighted already, all other genetic correlations with IPN were non-significant, low and approximately centred on zero, but nevertheless in the favourable direction.

5.4 Discussion

5.4.1 Polygenic (co)variance parameters

Thirty-one out of forty-four bivariate trait combinations converged successfully and hence yielded genetic correlations between IPN resistance and performance traits. Those failing to converge were mostly due to very low genetic variances seen in performance traits. Mortality to IPN, although recorded as a binary trait, was modelled on the observed (0,1) scale assuming normality, justified to some extent by the intermediate levels of mean prevalence found, as discussed in Chapter 4. ASReml does allow bivariate analyses applying a Generalised Linear Mixed Model (GLMM) to the first trait (only) using typically a logit or probit link to analyse the true binary data

on the underlying liability scale, and always assuming the second trait is normally distributed. Where these models were tried in this study, successful convergence was not achieved in a substantial number of cases, for reasons, discussed in Chapter 4, that remain unclear, and the observed scale was adopted for all analyses.

Overall this survey of the available datasets has revealed 7 out of 44 genetic correlations between IPN mortality and performance traits approaching significance at the 5 % level. Of these, four were biometric traits (***sulen, harvwt, harvlen, guttedfilletyld***) and three carcase quality traits (***hue, intensity and deformity)***. The strongest correlations were with seawater IPN looking at the single harvest in the 2003 year group. The two highest and most significant correlations were with flesh colour (***hue,*** 0.37 and ***intensity,*** 0.54) in a data set with reasonably well behaved and significant heritabilities for all traits except yield. These results could not be confirmed in the freshwater IPN data set (2005 year-group harvest) likely due to poorer data structure as there were few samples per family, i.e. an average of less than four for the colour traits. In the freshwater IPN dataset, only harvest weight (***harvwt***, 0.30) attained a significant correlation, contrasting with harvest length (***harvlen***, 0.27) in the seawater IPN data. Generally with biometric traits in salmon, effects observed for weight are unlikely to be important if not supported by a similar observation for length, since they are very highly correlated.

When all the available data was analysed together (analysis C), significant heritabilities were achieved for most traits, consistent with results published previously looking in detail at the 2001 year-group (Powell et al., 2008). Generally, biometric traits were found to have high heritabilities, (0.4 to 0.6), quality traits such as fat, colour, deformity and grilsing were intermediate (0.1 to 0.4) with yields low (below 0.1) but in the combined datasets significantly greater than zero. Heritability estimates reported elsewhere where traits are comparable (Rye and Gjerde, 1996; Gjedrem 2005; Quinton 2005) tend to be lower than those reported here. A notable exception is a heritability for colour obtained using the Minolta Chroma Meter of 0.46 reported in Gjedrem 2005. ***Salmofan*** and ***nirfats*** (heritabilities obtained here of 0.74 and 0.03 respectively) remain problematic and these estimates should be regarded as unreliable. Instrumental readings of both fat (Torry fat meter) and colour (Minolta chroma meter) in commercial breeding programs have tended to rely on affordable and portable equipment designed for quality control screening of commercially harvested stock. The only report we are aware of in the literature of an instrument that achieves high correlations with chemical analysis of both fat and astaxanthin content in the flesh is the Perton DA7000 (Solberg C, 2000). This uses Near Infrared Reflectance spectroscopy at a range of wavelengths to achieve correlations of 0.94 with a standard error of prediction (s.e.p) of 0.82% (fat percent) and 0.92 s.e.p. 0.33 mg kg^{-1} flesh (astaxanthin concentration) with chemical assays.

In the larger dataset, most of the genetic correlations were close to, and not significantly different from zero with only ***sulen, deformity*** and

guttedfilletyld achieving marginal significance. If the favourable ***sulen*** and other weight-based biometric trait correlations were confirmed elsewhere, then a functional association might be speculated, however there is no strong evidence from anywhere in this work that families of smaller fish are more predisposed to IPN mortality. It was also not unexpected that the high and significant correlations observed between ***hue*** and ***intensity*** with seawater IPN would be lost when pooled with all data from other years, for reason discussed in Materials and Methods. The general lack of observed genetic variation for colour may well be real and a reflection of the fact that in all these populations, commercial levels of pigment in the feed were offered to ensure adequate flesh colour saturation, since fish were commercially harvested and processed for sale. Despite this, however, heritabilities and genetic correlations are strikingly high and favourable in the seawater IPN dataset (0.37 and 0.54 for hue and intensity respectively) and so could be investigated further to identify which particular families may be implicated, and whether the result is confirmed elsewhere. If this result is reflected in the EBVs allocated to candidate families, then selection against IPN mortality would be expected to support any selection for better fillet colour.

5.4.2 QTL effects

One of the motivations for this chapter was the finding discussed in Chapter 3 and 4, that a QTL allele for IPN resistance explained a large proportion of the genetic variance for this trait, (Houston et al., 2008; Moen et al., 2009; Houston et al., 2010). Subsequently, the QTL susceptibility allele

appeared to be maintained at an unexpectedly high frequency in populations studied in both Scotland and Norway (Moen et al., 2009). One possible reason for this could be that the susceptible allele may have a positive effect on another trait undergoing active selection in farmed populations. Moen et al., 2009, however, pointed out that so far, such negative genetic correlations have not been observed. In the Norwegian breeding program only one of 14 recorded traits was found to be negatively and significantly correlated with IPN resistance. (AquaGen, unpublished, quoted by Moen et al., 2009). The trait was not disclosed, was only weakly associated ($r = 0.17$) and was not one of those under strong selection. This study adds seven more traits, only two of which are other than weakly associated and in a favourable direction.

Evidence is growing for the involvement of several QTLs affecting performance traits, for example growth traits (Reid et al., 2005; Boulding et al., 2008; Baranski et al., 2010), fat percent (Derayat et.al., 2007) and colour (Baranski et al., 2010). Genome-wide significant QTL affecting harvest length and head weight were detected on linkage group LG 1 in a sub-set of 10 families from the 1999 year-group by Houston et al., (2009b), and a significant QTL also affecting head weight was detected on LG 5. Suggestive levels were also detected for over 25 QTLs on 11 of the 29 linkage groups, although all but one (head weight) became non-significant after Bonferroni correction for multiple traits was made. Correction for body weight as a covariate also reduced the number of significant QTLs. Sex is known to have a measurable effect on weight traits taken at one sea-winter, prior to the usual time of harvest (Powell et al., 2008) and the sex determining locus is known

to map to LG 1 (Artieri et al., 2007). Therefore spurious associations with LG 1 may arise where sex is not properly accounted for in the analyses of weight related traits (Houston et al., 2009b).

In a study of 188 families from the 2003 year-group Gheyas et al., (2010) found no significant effect of the IPN resistance QTL (previously discovered on LG 21, Houston et al., 2008) on harvest performance including weights, lengths, yields (percentage weight following processing) and fat percent. More importantly, although finding suggestive QTLs for colour on LG 16, 18 and 23, Houston et al., (2009b) did not find LG 21 implicated in any of the harvest trait QTLs suggested in the 10 families studied from the 1999 year-group. This may partly be explained by the fact that those families were survivors of the IPN event which gave rise to the discovery of the IPN QTL. The skewed allele frequencies may possibly therefore have reduced the number of families informative of any association with LG21.

It is known that fat metabolism is implicated in immune function (Petropoulos et al., 2009). Morais et al., (2010) in a transcriptomic study of four 2006 year-group families from the LNS breeding program, observed a strong involvement of immune response genes in families contrasting for total lipid (26 % of all genes implicated) and omega-3 fatty acids, (38 %) respectively. Taken together these results remain suggestive but cannot be taken as confirmation of an association between IPN resistance and performance traits.

Interestingly, Baranski et al., (2010) found a genome-wide significant QTL for flesh colour (salmofan), mapping to chromosome 26, at 44 cM. This

is linkage group 21 (Philips et al., 2009), the same chromosome as the IPN QTL discovered previously at location 25cM (Moen et al., 2009) and 69 cM (Houston et al., 2008a). The QTL mapping population was an F2 cross between a wild, landlocked, isolated ice age relic population and commercial Norwegian salmon selected for fast growth and high colour. This generated extremes of the salmofan distribution (less than the lowest category of 20 and averages of 25) on which selective genotyping was performed. A large correlation between colour and biometric traits (weight and length, $r_p = 0.75$) seen in the mapping population was fitted as a covariate in the QTL analysis. Further, in other unreported studies, (quoting T.Moen pers. comm.) some genotypes at the IPN QTL appeared to be positively correlated to flesh colour. Baranski et al., (2010) hypothesise that non-lethal infection by IPN may lead to differential processing or depositing of pigment, depending on the IPN QTL genotype, but do not say whether the association would be favourable or unfavourable.

5.5 Conclusions

In this chapter we investigated whether the high genetic variation for IPN mortality seen in previous chapters was associated in any way with other performance traits, particularly harvest, collected over the same period. If significant and important associations were to be found, this would affect the balance of selection for traits targeted as breeding program objectives, the prediction of genetic change in both targeted and non-targeted traits, and

affect the desirability or not of rapidly bringing the underlying IPN QTL resistance allele to a high frequency in the broodstock population.

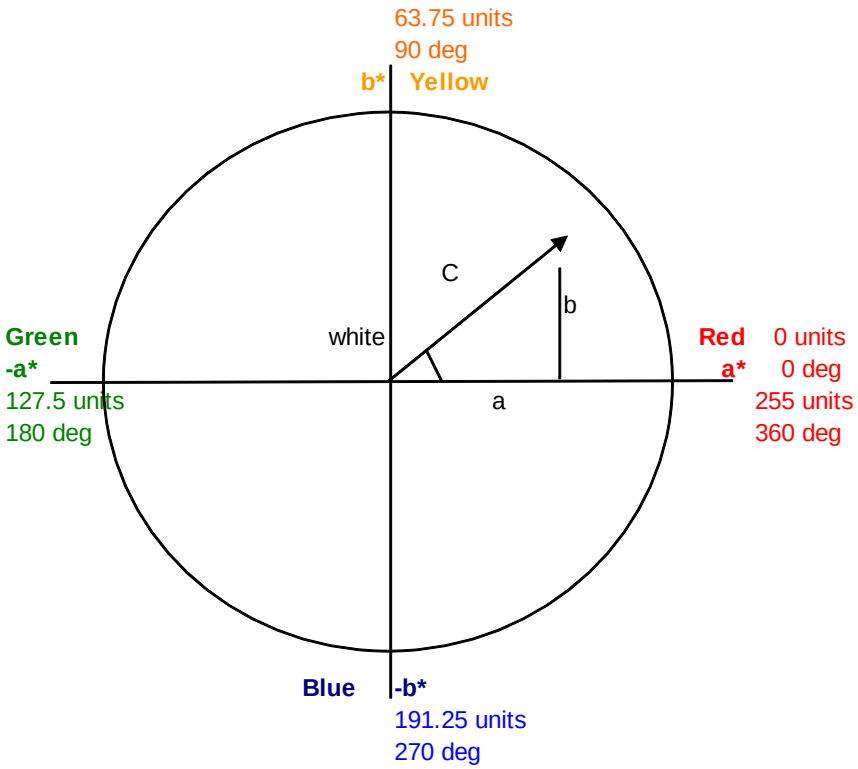
Out of 44 genetic correlations investigated, two appeared to be highly significant and a further five marginally so. This is more than those reported in similar studies elsewhere. The two significant traits (both favourable in direction) were the colour parameters ***hue*** and ***intensity*** related to seawater IPN mortality***.*** These colour traits have less functional significance than the third parameter ***saturation***, since the hue (redness) is largely a characteristic of the pigment astaxanthin, and intensity is a grey-scale measurement confounded to some extent but not completely, with saturation. The correlations were not confirmed by saturation, nor the freshwater and overall IPN mortality results. Of the biometric traits, ***sulen, harvwt and harvlen*** were all individually significant but there was no consistency over the three datasets. Unless confirmed in other datasets and given the lack of consistency, no firm recommendations can be given to utilise these genetic correlations in selection. This work generally supports that from elsewhere in suggesting IPN resistance is independent of other, non-disease traits being recorded by breeding programs. In view of the study by Baranski et al., (2010), however, it might be premature to dismiss the positive results reported in this chapter associating seawater IPN with colour too readily, and a need for further work in this area is indicated.

5.6 Appendix

Recording colour from images of salmon fillets :

Instrumental measures of colour are captured along three axes of colour space : red to green (axis ***a***); blue to yellow (axis ***b***); and light to dark (axis ***L***) represented respectively by the three ***x, y*** and ***z*** coordinates of Cartesian space. The range in colour variation of salmon and trout fillets is however from red to yellow, tending to white at very low pigment concentrations. Therefore analyses on the ***Lab*** or ***xyz*** models invariably give very inconsistent results since the visual colour response is the result of an interaction between the two parameters, ***a*** and ***b.***

Figure 5 : Lab and HSI colour space



Fortunately the ***HSI*** colour model is available which transforms the ***Lab*** axes into Polar coordinates in which the full range of colours are distributed around a circular disc lying on the ***ab*** plane. The ***HSI*** model describes colour (***hue***) as the angle around the circle ranging anticlockwise from red (0^o, extreme right of circle) to yellow (at 90^o) then through green and blue, back to red. Depth of colour (***saturation*** or equivalently *'chromaticity'*) is measured from the centre of the disc (a washed-out white) out towards the circumference. ***Intensity*** is equivalent to Lightness, represented in grey-scale as a cone rising vertically out of the plane (black) to white.

Greyscale of image files is 0 - 255 units (8 bit) for each parameter.

Hue = 'colour' = angle of hue :

= arctan (b/a) for a > 0 and b> 0

= 180 + arctan (b/a) for a < 0

= 360 - arctan (b/a) for a > 0 and b < 0

Instruments will then usually rescale the 0 - 360 degree scale to 0 - 255 eight-bit scale to conform with saturation and intensity.

Saturation = chroma = saturation of colour

= sqrt (a2 + b2) = line C in the above diagram

Intensity = Lightness

Saturation tends to white at zero (centre of circle)
Intensity tends to black at zero lying on the AB plane and rising to white as a cone above the plane)
Intensity and saturation are inversely related.

Applied to salmon fillets, ***hue*** is scaled from a deep red at 1^{o} to yellow at 90^{o}, therefore lower values are more desirable. ***Saturation*** is recorded on an 8-bit scale from 0 (washed out white) to 255 (deeply coloured) with higher values more desirable, ***intensity*** (equivalent to Lightness) is from degree of black (0) to white (255), irrespective of colour, hence lower values are more desirable. There is little information in intensity that is not present in saturation and the HSI model essentially resolves to two independent parameters.

The Minolta Chroma Meter, based on reflectance at specific wavelengths, can provide results transformed to either colour model. It is widely suspected, however, that the natural wetness of fillet flesh can lead to false reflections and readings in an unpredictable way and reduce the numerical accuracy of output.

Digital camera images usually provide image files based on the RGB (red, green, blue) colour model denoting the degree of each by an 8-bit scale 0-255. Natural variation in colour and the presence of artefacts will exist across fillets and so care is required to be consistent in selecting 'areas of interest' in digital images from which to take readings.

Chapter Six

General Discussion

6.1 Objectives overview

At the start of the series of field trials, before this study commenced, it was not known whether field exposure to pathogens in general and IPNv in particular would result in significant family i.e. genetic differences. The default expectation (null hypothesis) from experience with livestock was that any genetic component was likely to be polygenic, with a low heritability and a high level of environmental noise inherent in the difficulty of capturing data in a commercial environment. The most optimistic expectation was that there might be low but sufficient measurable genetic variation to allow the sensible application of Individual Animal Model (IAM) BLUP to establish a selection differential resulting in genetic gain over a number of generations. Detectable family differences were apparent from the first genotypes to be obtained from the first set of field trials. The initial application of mixed models to quantify the variation were not well behaved, showing 'unfeasibly high' heritabilities where convergence actually occurred.

As soon as this PhD study commenced, the actuality of high levels of genetic variation was rapidly established. This promoted a whole new round of ongoing investment funding centered on the BBSRC which aims to drill-

down to eventually discover the functional determinant or determinants of genetic resistance to IPN in Atlantic salmon.

The objectives of this PhD were restricted to the following questions:

(i) what is the extent of polygenic (co)variation in these datasets ?,

(ii) how does the genetic variation interact with those aspects of epidemiology, environment and inheritance that could be revealed by the format of the field and experimental trials, ? and

(iii) can an analytical framework be established in order to select for resistance to IPN within the routines of the breeding program ?

This chapter six initially aims to summarise the main findings of each of the four studies (chapters two to five) carried out for this PhD (*section 6.2*) , along with comment comparing the use of field and experimental trials to yield data on genetic resistance (*6.3*). Those results are then placed in the context of existing commercial procedures (*6.4.1 to 6.4.3*), and challenges raised by them are identified. Subsequent sections (*6.4.4 to 6.4.6*) will indicate how technology is emerging to better address these problems.

6.2 Key findings by chapter

Chapter two

Chapter two took a single year-group of data with full-sib families replicated over three sites for an in-depth exploration of what could be learned genetically and epidemiologically from the distribution of observed family differences. Building on the seminal work of Roberson and Lerner (1949), and Dempster and Lerner (1950) it was realised that relatively simple but novel expressions could be constructed inferring genetic heritability and correlation parameters directly from the family differences observed. Moreover this showed that genetic information could be directly gleaned from epidemiological observations.

Chapter three

The expressions derived in Chapter two, however, did not directly yield Estimated Breeding Values, for example required for routine selection schemes. Whilst it was possible to substitute the parameters derived from such expressions for the 'real' parameters required by BLUP evaluations, parameter estimates derived from REML are more appropriate, and preferred. Chapter three therefore investigated the totality of all the seawater data (no more became available after 2006), on the basis of REML linear mixed models. An efficient algorithm permitting scaling-up to very large datasets was implemented (the Reduced Animal Model) which had previously become available. The variance components thus obtained confirmed those derived earlier (Chapter 2) and further indicated high heritabilities and genetic

correlations within and between year-groups. Dam effects were explored but were considered to be largely confounded with family replication over sites (a GxE interaction) given these data structures.

Chapter four

Parentage assignment by genotyping had been completed for the first three year-groups of freshwater challenge trials, by October 2009. This permitted a comparison of (co)variance components with those obtained from seawater challenge and combined results looking over all the combined datasets. The highest heritability for a single site was obtained (0.59). When combined, the seawater and freshwater data yielded an even higher heritability (0.67) implying the possibility that the two data sets may be treated as the same trait. This was supported by the high, positive, genetic correlation found between freshwater and seawater resistance (0.68), although being significantly different to one, still leaves open the possibility of some functional differences in genetic resistance between the two life-stages of infection. During this period of study further results from elsewhere confirmed the existence of one or more QTLs for IPN resistance segregating in these and other populations, and these are related to this work in discussion.

Chapter five

Chapter five completed the analytical investigations by exploring the relationship between IPN resistance and the other performance and harvest traits undergoing routine evaluation by the breeding program. Only 7 out of

the 44 genetic correlations that were investigated approached significance at the 5 % level as assessed (approximately) by comparison with their standard errors. Two colour traits, hue (r_g = 0.37) and intensity (r_g = 0.54) had highly significant genetic correlations with seawater IPN resistance, but these and the others mentioned were not confirmed by any strong consistency over the various datasets. The field-based data structures, however, were not ideal and further specifically targeted investigations may provide deeper insights.

6.3 Field versus experimental testing

Field trials under commercial conditions may be difficult to sustain but, as we have shown, can produce extremely valuable data and can become the foundation of more targeted genomic investigations. Experimental facilities are expected to produce more precise and accurate data but usually at a higher cost and often on a minimally acceptable number of animals. For that reason experimental data may not be any more or less sustainable than field data. Fortunately, the choice appears to be a purely practical one : it was entirely fortunate that IPN was as amenable to genetic investigation as we have found, and under genetic control by a single major determinant that seems to operate predictably across all life-stages and under all environments (seawater and freshwater) and circumstances investigated so far. That situation certainly was not anticipated at the start of this study. Over the period that IPN has come under control, the negative impact of other viral diseases such as Infectious Salmon Anaemia (ISA), Pancreas Disease (PD) and parasitic diseases such as sealice infestation has increased to critical

levels, significantly impacting on the growth of the commercial salmon farming industry. For example in Chile, the need for regulatory control over ISA outbreaks has halved the size of the Atlantic salmon farming industry over the last three years, leading currently to a shortage on the world market. While the potential for discovering genetic and molecular determinants of resistance to these other diseases remains high, obtaining good data from either controlled experimental or field exposure trials appears to be considerably more problematic than is the case with IPN.

6.4 Selection for increased disease resistance

6.4.1 Polygenic EBV selection.

The main findings of this study were integrated into the commercial practice of the LNS breeding program immediately they arose. The implementation of the RAM model (Chapter 3) was a useful addition to the techniques available to solve linear mixed models for genetic parameters and BLUP evaluations. Its main utility was in allowing full animal model solutions in situations where it was impractical to list all individuals providing data, which was the case at the start of this study. The other obvious aquaculture application is in evaluating egg survival - the LNS breeding program for instance collects and evaluates data on up to 50 million eggs each year from breeding program families. That would give memory allocation or time problems to most mixed model solvers implementing an individual animal model across several years of data on modern desktop

workstations. RAM condenses this to 3,000 families where the contribution of both male and female to egg survival is still properly accounted for. It has some similarity to Sire and Dam models and is not so straight forward to implement in multi-trait as it is for univariate models. Analytical facilities, however, were sufficiently upgraded to allow the IPN data to be formatted directly for IAM (one row per animal) which was routinely used for chapters 4 and 5 and subsequent breeding program evaluations.

We have seen that it has proved possible to routinely collect field and experimental challenge mortality data on IPN infection, sufficient to support an ongoing selective breeding program. The inclusion of IPN resistance in the breeding goal, even without directly targeting the QTL and in the absence of genetic correlations with other traits, nevertheless impacts on the amount of progress that can be made simultaneously in other traits. Polygenic EBVs based on sib-testing are unable to discriminate between candidate full-sibs for resistance, each sib being allocated the family EBV (numerically equivalent to the mean EBV of the parents). Selection for IPN resistance then essentially resolves to selection among a limited number of families, where the trade-off between genetic progress made and genetic variation lost is exacerbated compared to traits where selection candidates have their own measurements and hence their own individual EBVs. If such an individual animal trait exists that is genetically correlated with the family-based disease trait, then additional within family selection becomes available for the disease trait, since the correlation will modify the family mean EBV for the disease trait allocated to each animal within the family. As well as identifying eligible

selection candidates within family, candidates from families that would have otherwise been excluded may now be selected. The benefit is a more optimal balance, leading to increased selection intensity at a decreased rate of loss of genetic variation (Sonesson A.K., 2007) as discussed further in section *6.4.4*. An unfavourable genetic correlation, however, at least with a trait also undergoing selection, would have the opposite, restricting effect. It is important therefore to quantify these possibilities, using parameters obtained from the data on which selection is to be based, and this becomes a potential topic for future work. There is however little evidence from the work presented here of the existence of unfavourable or favourable genetic correlations that could be used either way to enhance selection for IPN resistance.

Because of critical biosecurity constraints, selection for disease resistance in aquaculture species will mostly rely on destructive challenge tests on full-sibs to selection candidates. While full sib groups may be much larger than is common with other livestock, particularly mammalian, species and allow resistant families to be identified with high precision, without further information sib testing cannot differentiate between resistant and susceptible, but at the same time, untested, individuals within families (Odegard J., 2010a) and a great deal of the advantage of scale can be lost as selection differentials fall and the trait becomes more difficult to integrate across a balanced set of objective traits and inbreeding constraints. Correlated traits and targeting of QTLs by marker assisted selection can provide the extra information required, but presupposes they exist to be found. The

availability of what appears to be a single QTL determinant of resistance with such a strong effect is certainly not expected to be the norm with other important disease traits waiting to be targeted.

6.4.2 QTL EBV Selection

Other IPN resistance mechanisms particular to each life-stage may await to be discovered, however confirmation that a single QTL has a similar large effect at both life stages is perhaps surprising and suggests that alternative resistance mechanisms, if they exist, may be harder to detect while the QTL is segregating. The proportion of the polygenic heritability of IPN resistance that is attributable to the QTL is sufficiently high to suggest that selection on the basis of polygenic EBVs will itself have successfully increased the frequency of the resistant QTL allele. Tracking the frequency of QTL alleles over generations requires unequivocal marker genotyping results, leading to identified haplotypes where the phase can be calculated accurately within family. This is usually only possible where parents are heterozygous and where full-sib family sizes of the challenged population are greater than say five. Without markers closely linked to the QTL, many families and individuals become non-informative for the QTL.

Markers that are very close to the QTL (or, in fact, are the QTL itself) can be used across families without having to establish the association (phase) of each marker allele with each QTL allele within each family, since this will be constant over all families ('population wide') and so only has to

be established once.. If a point mutation giving rise to the marker occurred at the same time as the mutation in the gene giving rise to the QTL, then each allele of the marker will directly indicate a different allele (eg resistant or susceptible) of the QTL. Eventually, even this linkage-disequilibrium (LD) will break down by the occurrence of a rare recombination event between the QTL mutation and the mutation that gave rise to the marker. The favourable marker allele will then be associated with the unfavourable QTL allele in a way that will remain undetected. Selection by mistake of any offspring where this has occurred will degrade the selection for the QTL that can be achieved by using the marker.

Even if LD is maintained, some families will remain uninformative unless the marker is the actual functional mutation. The reason is because it is improbable that the marker mutation and the QTL mutation will have occurred simultaneously if they are separate events. Consider a pre-existing QTL mutation that has an intermediate frequency in the population. A marker mutation (eg nucleotide base **A** changes to **T**) occurring later as a single event very near to a resistant allele, will be diagnostic for that allele. If the marker frequency remains low in the population, however, there will be many more resistant alleles segregating with the original **A** marker allele haplotype. These will be missed, and selecting on the marker will remain inefficient. If the marker mutated first and rises to intermediate levels, the situation becomes a mirror image, but the inefficiency remains. The situation can be improved by constructing a panel of closely linked markers where a range of marker allele combinations (haplotypes) can be identified which together

indicate a particular QTL allele (Moen et al., 2009; Moen T, 2010) that applies across families. The incentive remains, however, to find the functional mutation.

Kinghorn (1997) outlined a probabilistic procedure to infer the allocation of QTL alleles to individuals (and hence allele frequencies in populations) based on their EBVs. Mostly this is not essential, since markers can be incorporated into mixed models in a framework familiar to ASReml. For example Pong-Wong et al., (2001) described the construction of an IBD (identical by descent) matrix giving the probability of inheritance of the marker alleles recorded to each individual from each of its parents. When provided with the best available estimate of the distance of the markers from the known QTL, a separate genotypic relationship matrix (***G***-matrix) can be provided (in its inverse form), to the mixed model which tracks the probability of an individual inheriting either allele of the QTL from each of its parents and forms the random IBD structure used to calculate the QTL variance and estimated breeding values. This is analogous to the provision of the ***A*** relationship matrix, which in the absence of marker data, just assumes a probability of 50 % of inheriting a calculated genetic effect (EBV) from each parent, as used to estimate the polygenic variance. When both matrices are supplied together, the total variance for the trait (recorded as (0,1) alive or dead by individual) is partitioned appropriately between the polygenic and QTL components. Once the mixed models have converged the EBVs pertaining to each component are provided from the BLUP solutions as polygenic EBVs and QTL EBVs respectively. The QTL EBVs can then used

as an additional trait in selection. Interpreted as probabilities of mortality or survival (if measured on the 0,1 observed scale), QTL EBVs take full account of the patterns of uncertainty in the inheritance of the QTL in relation to the available data and are therefore preferable to using assigned QTL scores (0, 1 or 2 resistant alleles etc) in a selection framework involving other traits. When such QTL analysis was applied to each of the three freshwater datasets (year-groups 2005, 2006 and 2007) the variance of the polygenic component reduced to effectively zero in the presence of the QTL variance, confirming the findings from previous experimental work that the QTL is explaining almost, if not all, the genetic variance for this trait.

6.4.3 Applying Marker Assisted Selection (MAS)

The direct selection of candidate breeders on the basis of markers linked to QTLs is called Marker Assisted Selection (Sonessen A K, 2005; Sonessen A K, 2007; Odegard et al., 2010a). It is a technology which can be contemplated whenever a QTL is discovered to be linked with markers located on the same chromosome, assuming they are not at opposite ends.

MAS is a long anticipated technology that has potential in two complementary areas. Firstly, to increase the precision with which future progeny performance can be predicted, for example in allocating parents to generate commercial stock for sale or multiplication. This can have a bigger and more immediate impact on an industry than the longer term aim of creating genetic gain in the breeding program families. Secondly, MAS can help control the rate of change in frequency of the favourable QTL in the

breeding program elite families. The benefit of MAS arises from the ability to discriminate between individual members of a full sib family without requiring them to be themselves exposed to the disease. This allows an increase in the selection differential obtained compared to where full-sib candidates would otherwise all receive the same parental evaluations as discussed above. Drawing eligible candidates from a wider set of families then allows better control of genetic variance in general and inbreeding in particular, for the same target rate of genetic gain.

There is a further (commercial) advantage to MAS related to the extremely large full-sib family sizes found in species cultivated for aquaculture : for Atlantic salmon the number of eggs produced by a four year-old female can be 15,000 eggs or more, almost all of which may be expected to be fertilised under hatchery conditions. For species such as cod or catfish, there may be greater than a million eggs per female initially fertilised. A superficial assessment would suggest very large selection differentials are possible. The operational costs of a breeding program are, however, largely determined by how many families are maintained and requiring evaluation and not the number of offspring generated. On that assessment Atlantic salmon breeding programs are comparable in size to mammalian livestock breeding units with 200-400 elite breeding families only. Over the four year cycle the number of families accumulate to 800-1600 respectively, but because parents die or are culled at fertilisation, year-groups tend to remain discrete. Selection differentials cannot then be easily increased by retaining the highest evaluated parents for rebreeding the following year, although

developments in sperm cryo-preservation (Salte et al., 2004) are starting to have an impact as they have with other livestock species.

The limitations detailed above however, do not exist for stock destined for sale. Each year the 200 pedigreed parental families of the LNS breeding program have generated, upon mating and selection of parents, over 3,000 progeny families totalling approximately 50 million eggs for sale for commercial rearing to harvest. Depending on the initial frequency of the favourable (resistant) QTL allele, and the genotyping investment required to identify sufficient numbers of homozygous or heterozygous resistant parental stock, the opportunity arises to use MAS to attain very high selection differentials in stock destined for commercial harvest. A lift in genetic merit at the commercial industry level may therefore by achieved with immediate effect and because of the numbers involved, without the customary genetic lag (Guy et al., 1981) imposed by multiplication layers in the industry breeding structures. Breeding programs experience constraints in balancing a range of traits and managing genetic variation for long term sustainability. It becomes possible therefore to create negative genetic lag, ie the genetic merit of stock for a particular trait passing through the commercial tier being permanently higher than that concurrently going through the breeding program stocks, a situation believed to been seen elsewhere only in very sophisticated dairy cattle breeding dissemination structures. While this would apply to any trait with heritabilities as high as those reported here, MAS gives a much increased degree of confidence in predicted commercial performance, which could be expected to influence uptake of the technology by industry.

6.4.4 Balancing Traits, Selection and Relatedness

Moving on to future practice, the QTL effect indicated here is sufficiently large to expect that attempts by properly resourced breeding programs to fix the favourable allele within one to three generations would be successful, even without locating the gene itself. Whether a cautious, slower approach is advised or an attempt made to achieve the clearance of all susceptible alleles from the breeding population in a single generation is a choice that can be made, and will be impacted to some extent by the position in the egg and smolt supply market that the breeding program occupies or into which it sells. Large, national, well resourced programs can afford, and be expected, to be cautious. In Atlantic salmon two generations takes a minimum of eight years as each generation involves four consecutive year-groups. That approach requires continued diagnostic testing of each year-group so that susceptible alleles passing into the stock for commercial sale can be identified, the customer informed and allowed to make a choice based on price. Smaller competing programs may not have that luxury. Prolonged expensive testing may be less affordable and other diseases will emerge that assume greater importance and demand on resources. The use of markers very close to the causative QTL can dramatically reduce testing and diagnostic costs and help the screening out of susceptible alleles from stock destined for commercial sale. The only way to avoid testing altogether, however, and still offer a competitive product in the market is to fix the resistant allele in all broodstock as fast as possible. Homozygous resistant stock can then be offered to customers without the expense of continued

testing, either incurred by the program, or passed on to the customer. Other positions within these two extremes can be adopted, but in all commercial business, the opportunity exists to gain advantage from accepting the risks involved.

There is a duty on technical managers, therefore to investigate and quantify the associated risks of any course of action. This raises the critical need to ensure that desired genetic progress in other traits under selection is properly managed along with measures to minimise any loss of genetic variation and the monitoring of any unexpected and undesired effects on important traits not directly under selection.

Good analytical techniques now exist to deal with these issues. For example, given genetic (co)variance parameters obtained from mixed model analyses, classical selection indexes (Cameron N D., 1997) can be used to predict the optimum combination of the measured trait EBVs to achieve desired rates of progress in each of the traits under selection. A practical implementation for use in a commercial context is the 'desired gains' software utility of Kinghorn (2010). Efficient software utilities to establish the optimum balance between selection and maintaining genetic variation, suitable for aquaculture applications exist (OCSelect), based on the 'optimum contributions' method (Hinrichs et al., 2006). A practical implementation based on a different optimisation algorithm to that of OCSelect can be found in the 'mate selection' routine within the Pedigree Viewer software utility of Kinghorn (2010) described in Carvalheiro (2010) and Kremer et al., (2010).

Both these optimisations require the setting of constraints to future increases in inbreeding, and provide the maximum genetic gain under the defined circumstances, and can be used for both simulation and as operational decision-making tools. The mate allocation strategy of Kinghorn additionally takes advantage of any dominance effects inherent in the trait or associated QTL by maximising the expression of heterosis in the offspring of suggested matings. Without dominance, however, there is little advantage to the nominated specific matings above that inherent in the nominated parents except a finer control to avoid matings between more distantly related individuals.

By use of the above utilities, risks based on current knowledge can be quantified and informed decisions made. The greatest risk, however, from a deliberate loss of genetic variation in order to fix a specific disease resistant allele is the possible genetic association with a currently unimportant or unknown disease that emerges to give critical levels of problems at some stage in the future. That scenario includes the possibility of viral evolution to more virulent and pathogenic forms, potentially by-passing the host resistance previously fixed. Such risks are dangerous because they are difficult to imagine. IPN is a very good example. When it was first characterised, it was an unimportant viral disease of salmonids in freshwater, with most strains avirulent and non-pathogenic. The rise in virulence and incidence of the currently most pathogenic strain (Sp) coincided with the growth of salmon farming in the 1970's. The increasing problems from IPN and ISA impacting the industry in Chile similarly coincides with the very

rapid expansion of commercial salmon farming in the 1990's despite a tightening of regulation.

6.4.5 Identifying the gene

Uncovering the actual mechanism and function of the QTLs will require further work. Next-generation genotyping technologies based on SNPs are becoming available which will lead to dense saturation of markers across the genome and finer localisation of QTLs (Odegard et al., 2010a; Houston et al., 2010b). A benefit of using SNP markers to investigate QTLs is that an extremely closely linked marker may indicate the actual causal mutation. A microsatellite marker will never be the causal mutation. Even if one is found in complete linkage disequilibrium with a QTL (i.e. no recombinants observed over a very large number of offspring), the prospect of sequencing what may be several million base-pairs along the chromosome (chromosome walking) to a putative coding region is currently hampered by the lack of a public domain salmon genome sequence with which to align any specific observed sequence. The ultimate aim therefore is to identify the causal mutation and the application of gene-assisted selection (GAS) using a rapid and cheap diagnostic DNA test for each allele of the mutation. Transcriptomic analyses (Houston et al., 2010b) use an entirely different approach to search for candidate genes by taking families which contrast in their high or low susceptibility to a disease and quantifying the up or down-regulation of messenger RNA that is produced as families respond to an actual challenge. Very high throughput screening of the cDNA

(complementary DNA transcribed from the collected RNA) for tens of thousands of known genes simultaneously can help to elucidate biosynthetic and metabolic pathways which may in turn point towards possible modes of action of the QTL. This approach is capable of producing a very large amount of information in a very short time and doesn't rely on the archived banks of collected DNA necessary for QTL analyses, at least in the discovery phase. The bioinformatic challenge of deciding which signals are relevant against the background noise is orders of magnitude more challenging than for classical QTL searches although this area is currently seeing heavy research investment and very rapid development. By whatever means, if candidate genes can be suggested, then a search for marker sequences in the chromosomal DNA within or near to the gene will turn the effort back to establishing rapid and cost effective DNA diagnostic marker tests for the favourable alleles.

6.4.6 Genomic selection

Recent QTL searches (e.g. Baranski et al., 2010) using increasingly dense marker panels are tending to identify multiple QTLs with varying levels of effect on a particular performance trait, spread across the whole genome, in effect moving back towards a more polygenic mode of inheritance. At the same time there is very strong imperative from medical genetic-based diagnostics to develop ever more efficient and cheaper very high throughput sequencing and genotyping technologies. These are starting

to become available and affordable to aquaculture and livestock applications. Being based on SNP chips, these are expected to promote both the construction of very highly dense genome maps along with more rapid development of genotyping technologies that allow markers across the whole genome to be screened on each of a large number of individuals. The culmination of this technology is genomic selection (Meuwissen et al., 2001) where phenotype recording establishes the trait-marker associations assigned to each segment of chromosome identified by very closely spaced markers. Subsequently, only the inheritance of each segment needs to be tracked from parents to offspring. Selection is then based purely on the sum total breeding value of the inherited segments received by an individual. Prior knowledge of the effect and location of specific QTLs then becomes unimportant, effectively returning the process back to using polygenic mixed models. Inheritance through the pedigree is replaced by the actual observed inheritance of each chromosomal segment although Solberg et al., (2009a) suggest that retaining polygenic evaluations in the models can have some advantage as selections are repeated over generations.

There are still considerable hurdles to overcome before this technique can become a practical reality. In extensive simulation studies, Solberg et al., (2008) suggest that for a 30-morgan genome (i.e., comparable in size to that of Atlantic salmon), perhaps up to 24k (24,000) SNP markers would be needed. (Recent estimates update this to 70-100k or more while the currently available SNP chip is only 17k). Analytically that leads immediately to the possibility of having orders of magnitude more parameters than data points,

indicating a requirement to reduce dimensionality to make the models tractable, (Solberg et al., 2009b). Efficient analytical methods suitable for practical application are, however, now available (Meuwissen et al., 2009). Given that it is widely agreed (Tier, B., 2010) that this is the direction in which the technology is leading, work is already in hand to enhance the genomic and phenotypic databases supporting the LNS breeding program to accommodate this level of scalability.

Chapter Seven

Unanswered Questions and Concluding Remarks

7.1 Other species - trout QTLs

This study was restricted to investigating the extent of polygenic variation for IPN resistance based on recorded mortality. When the Landcatch Breeding program was initiated (1998) to facilitate the pedigree construction, data collection and selective breeding leading to this study, the only specific indication that there may be genetic determination of resistance to IPN was a report from Japan (Okamoto et al., 1993) that simple selection of survivors appeared to have very quickly fixed genetic resistance in the strain of rainbow trout under investigation. By 2001 (Ozaki et al., 2001) two QTLs had been identified on rainbow trout linkage groups RT LG3 and RT LG22 respectively, each explaining approximately 17 % of the phenotypic variance in mortality, in what was the first reported discovery of a QTL in any fish species. The markers closest to the rainbow trout QTLs were tested in our Atlantic salmon populations but found uninformative (Houston et al., 2008a). Another rainbow trout marker, however, *OmyRGT44TUF*, maps to both AS LG21 and RT LG22 at opposite ends to where the QTLs were respectively located. Further evidence of homology between RT LG22 and

AS LG21 is being investigated. Fortunately IPN has yet to emerge as a major problem in freshwater trout production possibly because the cyclic exposure to a seawater phase is avoided.

7.2 Analytical methods

The initial results and data collected from this study naturally motivated a parallel QTL investigation based on the same material and data. This has been equally successful in elucidating the QTL determinants underlying the genetic basis of resistance to IPN in Atlantic salmon. All these investigations have however been based on the binary response of mortality or survival. While the analytical models developed were sufficient to yield the insights we required for this study, the likelihood based analyses did not always produce clean results in terms of failure to converge or logit models yielding heritability estimates lower than those from observed models. The combination of binary data with a major QTL determinant of the variance structures seems to provide a sufficiently parsimonious explanation for this, but leaves open the possibility that other approaches may give different insights. Bayesian solutions to the mixed models (Odegard et al., 2010b) may behave better under such conditions, although they are not expected to improve on models that poorly fit the data in the first place. As another example, it is clear from Chapter 2 that there is an element of time to death involved. All the freshwater and most of the seawater datasets on which this study is based have time to death recorded and it is expected that analysis by survival models for example (not considered in this thesis) may lead to

additional information (Odegard et al., 2010a) about variation not currently explained by the known QTL. This may become more important as the QTL is progressively brought under control in breeding populations.

The opportunity for rapid genetic change in breeding program stocks makes it imperative that selection is targeted appropriately, both in balance with other traits and in maintaining sufficient genetic variation for future goals and challenges including those related to other diseases. In the datasets reported here there appear to be sufficient pedigree connectedness over years to observe what are favourable rates of genetic change in the traits targeted for selection. The favourable result, however, is consequential purely on heritabilities and applied selection differentials being significantly greater than zero. There are no alternative methods, such as random bred control populations, available for estimating genetic gain in these datasets. More detailed work is required and is reserved for a time outside this study where it can be investigated in appropriate depth. Of particular interest from the ongoing genotyping for MAS will be the revealing of marker and QTL allele frequencies for more year-groups and it should be possible to track how these change over generation and year-groups. In particular, there has not been a sufficient number of generations (4-6) since the original imports (1982-85) to produce a homogeneous mix of founder groups in current year-groups, so it would seem possible to investigate connectedness and evaluate the comparative merit of the founder strains for a range of traits. These results would then inform future decisions on drawing the balance between targeted genetic gain and the need to maintain genetic variation.

7.3 Other diseases

The domestication of Atlantic salmon is regarded as still being in its infancy. Emerging diseases capable of having a high economic impact in particular are very difficult if not impossible to anticipate over the medium term. For example the scale and continued development of the salmon industry world-wide is currently (2010) under severe restriction through the impact of biosecurity measures both within and across trading borders. These security measures are in place to control the devastating effect on national industries of what are otherwise uncontrolled emerging viral diseases. For this reason, the viral pathogen causing Infectious Salmon Anaemia (ISA) has emerged to overtake IPN virus in its impact on the industry, over the time course of this study. Likewise, as IPN has been brought under control, Pancreas Disease (PD) has become predominant in its economic impact on the industry. PD is altogether more challenging to investigate than with IPN : the alpha virus causing the disease has only relatively recently been identified, and suitable challenge protocols only very recently established. The pathology is not ideal for investigation in that not all infected animals die, while performance is usually compromised. Scoring simply on mortality observation therefore misses a great deal of information (Odegard J., 2010a) Attempts to set up sentinel field testing for PD in the LNS breeding program along the lines of those established successfully for IPN have so far failed to generate the required data. This is partly because it has been difficult to establish significant and countable mortality events in breeding program families despite heavy mortalities in adjacent stock from other sources, and

partly because mortalities occur over a much longer time span than the rather narrow window seen with IPN. Sealice infestation is also an emerging problem as pathogens develop resistance to chemical treatments currently used. Suitable recording protocols for field trials on sealice are similarly difficult to establish. For all these reasons, the current thrust of research is in establishing well defined experimental challenge protocols and moving straight to genomic investigation. While these techniques are being brought to bear on these diseases, their pathology and epidemiology is not as easily recorded or addressed as was found with IPN and they are equally unlikely to reveal their secrets as easily.

7.4 Sustainability

It remains to be seen whether selection on the QTL leads to a sustainable long term solution for the industry, or whether unforeseen aspects of host-pathogen interaction or continued pathogen evolution shortens the time over which benefit can be extracted. This could be important both as the known QTL leads to fixation and as functional genomics begin to shed light on the mechanisms involved. In particular it is not currently known whether the QTL facilitates survival by preventing the virus from entering and/or establishing replication within the host, or the alternative of enabling the infected fish to survive the carrier, 'persistently infected' condition, and to continue, or not, to shed virus into the environment. Clearly this latter option is less acceptable given the close relationship between the environments shared between farmed and wild salmon, and the wide species range of the

virus. Vaccination gives some protection (and currently in the seawater phase only) but being based on immune response, this seems to imply an independent and additive mode of action to the QTL effect. If the combination of QTL and vaccination is not completely protective then the possibility of a rapid evolutionary response from the virus (Gandon et al., 2001) becomes real, given continued very heavy concentrations of hosts in the farmed environment. Given suitable insights into these epidemiological parameters, the approach of Bishop et al., (2010) would become required and critical work to properly inform on future directions to take. By far the best outcome would be if vaccination and genetic resistance combine to ensure virus is not replicated and shed into the aquatic environment by the host and the virulent strain of IPNv ceases to be endemic in all farmed and wild environments. Whether that is the actual outcome, remains to be seen.

List of References

AquaGen AS (2010)
url : http://www.aquagen.no

Artieri C.G., Mitchell L.A., Ng S.H., Parisotto S.E., Danzmann R.G., Hoyheim B., Phillips R.B., Morasch M., Koop B.F. & Davidson W.S. (2007)
Identification of the sex-determining locus of Atlantic salmon (Salmo salar) on chromosome 2.
Cytogenetic and Genome Research 112, 152–9.

Baranski M., Moen T., Vage D I (2010)
Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (Salmo salar)
Genetics Selection Evolution 42:17

Bebak J., McAllister P.E., & Smith G. (1998)
Infectious pancreatic necrosis virus: Transmission from infectious to susceptible rainbow trout fry.
Journal of Aquatic Animal Health 10, (3) 287-293.

Bell J. G., Pratoomyot J, Strachan F, Henderson R, Fontanillas A, Hebard A, **Guy D R**, Hunter D, Tocher D R. (2010)
Growth, flesh adiposity and fatty acid composition of Atlantic salmon (Salmo salar) families with contrasting flesh adiposity : effects of replacement of dietary fish oil with vegetable oils
Aquaculture 306 (2010) 225-232

Bishop, S.C., Mackenzie, K.A., 2003.
Genetic management strategies for controlling infectious diseases in livestock populations.
Genetics Selection Evolution 35, S3-S17.

Bishop S. C., Woolliams J. A. (2010)
On the Genetic Interpretation of Disease Data
PLoS ONE 5(1): e8940. doi: 10.1371/journal.pone.0008940

Bootland L.M., Dobos P., & Stevenson R.M.W. (1991)
The IPNv Carrier State and Demonstration of Vertical Transmission in Experimentally Infected Brook Trout.
Diseases of Aquatic Organisms 10, (1) 13-21.

Borlaug N.E. (1989)
Address to the 1988 World Food Conference
1988 World Food COnference Proceedings Vol 1 : Policy Addresses pp79-83. Published 1989 ISBN0-8138-0897-9

Borlaug N.E. and Dowswell C.R.(1995)
Mobilising Science and Technology to get agriculture moving in Africa.
Development Policy Review 1995 vol 13 (2) : pp115-129

Boulding E.G., Culling M., Glebe B., Berg P.R., Lien S. & Moen T.(2008)
Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses.
Heredity 101, 381–91.

Bowden T.J., Lockhart K., Smail D.A., & Ellis A.E. (2003)
Experimental challenge of post-smolts with IPNv: mortalities do not depend on population density.
Journal of Fish Diseases 26, (5) 309-312.

Bowden T.J., Smail D.A., & Ellis A.E. (2002)
Development of a reproducible infectious pancreatic necrosis virus challenge model for Atlantic salmon, Salmo salar L.
Journal of Fish Diseases 25, (9) 555-563.

Bruno D.W. (2004)
Changes in prevalence of clinical infectious pancreatic necrosis among farmed Scottish Atlantic salmon, Salmo salar L. between 1990 and 2002.
Aquaculture 235, (1-4) 13-26.

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

Cameron N. D., (1997)
Selection Indices and Prediction of Genetic Merit in Animal Breeding
CAB International, Wallingford UK. ISBN 0 85199 169 6

Carvalheiro, R., Kinghorn B.P. and Queiroz S.A. 2010.
Mate Selection accounting for connectedness.
Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, (2010)

Dempster E. R., and Lerner L. M., (1950)
Heritability of threshold characters
Genetics 35 : 212-236

Derayat A., Houston R. D, **Guy D.R**., Hamilton A,. Ralph J, .Spreckley N., Taggart J. B., McAndrew, B. J., Haley, C.S., Bishop, S.C. (2007)
Mapping QTL affecting body lipid percentage in Atlantic Salmon (Salmo Salar).
9th International Symposium on Genetics in Aquaculture, Montpellier, France 26-30th June 2006
Aquaculture 272, (2007) doi: 10.1016/j.aquaculture.2007.07.046

Doyle R.W. and Herbinger C. (1994)
The use of DNA fingerprinting for high-intensity, within-family selection in fish breeding
Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, 7-12 Aug 1994, vol 19, pp 364-371

Falconer D.S., Mackay T.F.C. (1996).
Introduction to Quantitative Genetics.
Fourth edition, Longman & Co, London.

FAO (2008).
Fishery and Aquaculture Statistics 2006.
Food and Agriculture Organisation Yearbook 2008 , Rome.

FRS , 2003.
Final Report of the Aquaculture Health Joint Working Group sub-group on Infectious Pancreatic Necrosis in Scotland.
Fisheries Research Services Aberdeen. 90 pp. ISBN : 0-95 464 90 0-0-1

Gandon S., (2001).
Imperfect vaccines and the evolution of pathogen virulence.
Nature 414, 751–756.

Gheyas A. A., Houston R. D., Mota-Velasco J. C., **Guy D. R.**, Tinch A. E., Haley C. S., Woolliams J. A. (2010a)
Segregation of IPN resistance QTL in the early life cycle of Atlantic salmon (*Salmo salar*).
Animal Genetics: doi:10.1111/j.1365-2052.2010.02032.x

Gheyas A. A., Haley C. S., **Guy D. R.**, Hamilton A., Tinch A. E., Mota-Velasco J. C., Woolliams J. A. (2010b)
Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon.
Animal Genetics: doi:10.1111/j.1365-2052.2010.02051.x

Gianola D., (1980)
A method of sire evaluation for dichotomies
J. Anim. Sci. 51 : 1266

Gianola D., (1982)
Theory and analysis of threshold characters.
J. Anim. Sci 54: 1079

Gianola D., and Foulley J. L. (1983)
Sire evaluation for ordered categorical data with a threshold model.
Genet. Sel. Evol 15 : 201

Gilmour, A. R., Thompson, R. and Cullis, B. R. (1995).
AI, an efficient algorithm for REML estimation in linear mixed models
Biometrics 51: 1440–1450.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2006.
ASReml User Guide, Release 2.0
VSN International Ltd.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R., (2009).
ASReml User Guide, Release 3.0
VSN International Ltd.

Gjedrem, Trygve, (2005)
Selection and Breeding Programs in Aquaculture
AKVAFORSK, Institute of Aquaculture Research AS, Norway
ISBN-13 978-1-4020-3341-4 Springer, Dordrecht, The Netherlands

Gjoen H.M., Refstie T., Ulla O., & Gjerde B. (1997**)**
Genetic correlations between survival of Atlantic salmon in challenge and field tests.
Aquaculture 158, (3-4) 277-288.

Guy D. R. and Smith C. (1981).
Derivation of Improvement Lags in a Livestock Industry.
Animal Production 32 : pp 333 - 336.

Guy D. R., Hamilton A., (2005)
Genetic parameters for Harvest and Processing Traits in Landcatch Strains of Atlantic Salmon Salmo salar
Aquaculture vol 247 p17 (2005) presented at Genetics in Aquaculture VIII, 9-15 November 2003, Puerto Varas, Chile

Guy D.R., Bishop S.C., Brotherstone S., Hamilton A., Roberts R.J., McAndrew B.J., Woolliams J.A. (2006)
Analysis of the incidence of infectious pancreatic necrosis mortality in pedigreed Atlantic salmon, Salmo salar L. populations
Journal of Fish Diseases 29 (2006) 637-647

Guy D.R., Bishop S.C., Woolliams J., Brotherstone S., (2009)
Genetic parameters for resistance to Infectious Pancreatic Necrosis in pedigreed Atlantic salmon (Salmo salar) post-smolts using a Reduced Animal Model
Aquaculture 290 (2009) 229-235

Henderson, C.R., 1986.
Estimation of variances in animal-model and reduced animal-model for single traits and single records.
Journal of Dairy Science 69 : 1394-1402.

Henryon, M., Jokumsen, A., Berg, P., Lund, I., Pedersen, P.B., Olesen, N.J., Slierendrecht, W.J., 2002.
Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within a farmed population of rainbow trout.
Aquaculture 209, 59-76.

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

Hinrichs D., Wetten M., Meuwissen T. H. E. (2006)
An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates
J Anim Sci 84 : 3212-3218 doi: 10.2527/jas.2006-145

Houston R.D., Haley C.S., Hamilton A., **Guy D.R.**, Tinch A.E., Taggart J.B., McAndrew B.J., Bishop S.C. (2008a)
Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar)
Genetics 178 (2008) 1109-1115

Houston R.D., Gheyas A., Hamilton A., **Guy D.R.**, Tinch A.E., Taggart J.B., McAndrew B.J., Haley C.S., Bishop S.C. (2008b)
Detection and Confirmation of a Major QTL Affecting Resistance to Infectious Pancreatic Necrosis (IPN) in Atlantic salmon (Salmo salar)
International Symposium on Animal Genomics for Animal Health, Oct 25-27 2007, Paris
Animal Genomics for Animal Health 132 (2008), 199-204

Houston R.D., Bishop S.C., Hamilton A., **Guy D.R.**, Tinch A.E., Taggart J. B., Derayat A., McAndrew B. J., Haley C. S. (2009b)
Detection of QTL affecting harvest traits in a commercial Atlantic salmon population.
Animal Genetics 40 (2009) 753-755

Houston R.D., Haley C.S., Hamilton A., **Guy D.R.**, Mota-Velasco J.C., Gheyas A.A., Tinch A.E., Taggart J.B., Bron J.E., Starkey W.G., McAndrew B.J., Verner-Jeffreys D.W., Paley R.K., Rimmer G.S.E., Tew I.J., Bishop S.C. (2010)
The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL
Heredity (2010) doi: 10.1038/hdy.2009.171

Houston R.D., Haley C.S., Hamilton A., **Guy D.R.**, Mota-Velasco J.C., Gheyas A.A., Tinch A.E., Taggart J.B., Bron J.E., Starkey W.G., McAndrew B.J., Verner-Jeffreys D.W., Paley R.K., Rimmer G.S.E., Tew I.J., Bishop S.C. (2010b)
Functional characterisation and fine mapping of a major disease resistance QTL in Atlantic salmon
32nd Conference of the International Society of Animal Genetics (ISAG 2010), July 26th -30th 2010, Edinburgh, Scotland, U.K

Jarp J. (1999)
Epidemiological aspects of viral diseases in the Norwegian farmed Atlantic salmon (Salmo salar L.).
Bulletin of the European Association of Fish Pathologists 19, (6) 240-244.

Kent M. (1990)
Hand-Held instrument for fat/water determination in whole fish.
Food Control 1 : 47-53 available at :
http://www.distell.com/products/prd-fish-fatmeter/technical-data/ffm-research-paper

Kinghorn B. J. (2010)
homepage :
url : http://www-personal.une.edu.au/~bkinghor/

Kinghorn, B.P. and Clarke B. E. (1997).
Genetic evaluation at individual QTL.
Animal Biotechnology, 8:63-68

Kjell, M. 2003**.**
IPN in salmonids - a review.
Editors : Norwegian Seafood Federation-Aquaculture (FHL havbruk). VESO Trondheim. 115 pp. ISBN : 82-91743-16-9 www.veso.no

Kjoglum, 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.

Kjoglum S, Henryon M, Aasmundstad T, Korsgaard I (2008)
Selective breeding can increase resistance of Atlantic salmon to furunculosis, infectious salmon anaemia and infectious pancreatic necrosis.
Aquaculture Research 39: 498–505.

Kremer, V.D., Newman S., Wilson E.R. and Kinghorn B.P.. 2010.
Mate Selection for sustained genetic improvement in small populations.
Proceedings of the 9th World Congress on Genetics Applied to Livestock Production. Leipzig (2010)

LNS (2010) Landcatch Natural Selection Ltd,
url : http://www.landcatch.co.uk/lns/

Leaver M.J., Taggart J.B., Villeneuve L., Bron J.E., **Guy D.R.**, Bishop S.C., Houston R.D., Matika O., Tocher D.R. (2010)
Heritability and mechanisms of n-3 long chain polyunsaturated fatty acid deposition in the flesh of Atlantic salmon
Comparative Biochemistry and Physiology D, doi:10.1016/j.cbd.2010.04.002

Matos C.A. , Thomas D.L., Gianola D., Templeman R.J. and Young L.D. (1997a)
Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models: I Estimation of genetic parameters.
J Anim Sci 75 : 76-87

Matos C.A., Thomas D.L., Gianola D., Perez-Enciso M. and Young L.D. (1997b)
Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models: II Goodness of fit and predictive ability.
J Anim Sci 75 : 88-94

Matos C.A.P., Thomas D.L., Young L.D., Gianola D. (2000)
Genetic analyses of lamb survival in Rambouillet and Finnsheep flocks by linear and threshold models.
Animal Science 71 : 227-234

McLoughlin M. & Weigall B. (2002**)**
Infectious Pancreatic Necrosis (IPN) of Atlantic salmon : Summary of FVS seminar June 2000.
Fish Veterinary Journal, (6) 73-82.

Meuwissen T.H., Hayes B.J., Goddard M.E., (2001)
Prediction of total genetic value using genome-wide dense marker maps
Genetics, 157, 1819-1829

Meuwissen T.H.E., Solberg, T.R., Shepherd, R., Woolliams, J.A. (2009)
A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value.
Genetics Selection Evolution 41:2

Midtlyng P.J., Storset A., Michel C., Slierendrecht W.J., & Okamoto N. (2002)
Breeding for disease resistance in fish.
Bulletin of the European Association of Fish Pathologists 22, (2) 166-172.

Mikalsen A.B., Torgersen J., Alestrom P., Hellemann A.L., Koppang E.O., & Rimstad E. (2004)
Protection of Atlantic salmon Salmo salar against infectious pancreatic necrosis after DNA vaccination.
Diseases of Aquatic Organisms 60, (1) 11-20.

Moen T., Baranski M., Sonesson A., Kjoglum S. (2009)
Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait
BMC Genomics (2009) 10:368 doi:10.1186/1471-2164-10-368

Moen T., (2010)
Response to Selection At A Single Locus : Large Change In Allele Frequency Between Two Consecutive Generations At A Major QTL For IPN Resistance In Atlantic Salmon (Salmo salar)
Poster PP4-122, World Congress on Genetics Applied to Livestock Production IX, Leipzig, Germany August 1-6, 2010.

Morais S., Taggart J.B., **Guy D.R.,** Bell J.G., Tocher D.R. (2010)
Possible association between flesh lipid and IPN resistance in Atlantic salmon (Salmo salar)
32nd Conference of the International Society of Animal Genetics (ISAG 2010), July 26th -30th 2010, Edinburgh, Scotland, U.K.

Mrode, R. A., (2005).
Linear Models for the Prediction of Animal Breeding Values.
2nd Edition, CAB International, Oxon UK.

Murray A.G., Busby C.D., & Bruno D.W. (2003**)**
Infectious pancreatic necrosis virus in Scottish Atlantic salmon farms, 1996-2001.
Emerging Infectious Diseases 9, (4) 455-460.

Odegard, J., Olesen, I., Gjerde, B., Klemetsdal, G. 2006.
Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (Salmo salar) : Prediction of field survival.
Aquaculture 259, 116-123.

Odegard, J., Olesen, I., Gjerde, B., Klemetsdal, G., 2007.
Positive genetic correlation between resistance to bacterial (furunculosis) and viral (infectious salmon anaemia) diseases in farmed Atlantic salmon (Salmo salar).
Aquaculture 271, 173-177.

Odegard, J., Olesen, I., Gjerde, B., Klemetsdal, G. 2007b.
Evaluation of statistical models for genetic analysis of challenge test data on ISA resistance in Atlantic salmon (Salmo salar) : Prediction of progeny survival.
Aquaculture 266, 70-76.

Odegard J., Baranski M, Gjerde B, Gjedrem T,(2010a)
Genetic evaluation of disease resistance in aquaculture species – challenges and future prospects
Scottish Aquaculture Research Forum (SARF), 'Scottish Aquaculture : A Sustainable Future', Edinburgh, April 2009
Aquaculture Research 2010; 41 (suppl.1): in press 2010)

Odegard J., Meuwissen THE., Heringstad B., Per Madsen (2010b)
A simple algorithm to estimate genetic variance in an animal threshold model using Bayesian inference
Genetics Selection Evolution 42:29

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

Okamoto N., Tayama T., Kawanobe M., Fujiki N., Yasuda Y., & Sano T. (1993)
Resistance of a Rainbow-Trout Strain to Infectious Pancreatic Necrosis.
Aquaculture 117, (1-2) 71-76.

Ozaki A., Sakamoto T., Khoo S., Nakamura K., Coimbra M.R.M., Akutsu T., & Okamoto N. (2001**)**
Quantitative trait loci (QTLs) associated with resistance / susceptibility to infectious pancreatic necrosis virus (IPNv) in rainbow trout (Oncorhynchus mykiss).
Molecular Genetics and Genomics 265, (1) 23-31.

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

Patterson, H. D. and Thompson, R. (1971).
Recovery of interblock information when block sizes are unequal.
Biometrika 58: 545–54.

Petropoulos I.K., Thompson K.D., Morgan A., Dick J.R., Tocher D.R., Bell J.G. (2009)
Effects of substitution of dietary fish oil with a blend of vegetable oils on liver and peripheral blood leukocyte fatty acid composition, plasma prostaglandin E2 and immune parameters in three strains of Atlantic salmon (*Salmo salar*)
Aquaculture Nutrition 15: 596-607

Philips R.B., Keatley K.A., Morasch M.R., Ventura A.B., Lubieniecki K.P., Koop B.F., Danzmann R.G. Davidson W.S. (2009)
Assignment of Atlantic salmon (salmo salar) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (Oncorhynchus mykiss)
BMC Genetics 10:46
doi:10.1186/1471-2156-10-46

Pong-Wong R., George A.W., Woolliams J.A., Haley C.S. (2001)
A simple and rapid method for calculating identity-by-descent matrices using multiple markers
Genet. Sel. Evol 33 : 453-471

Powell J., White I., **Guy D.**, Brotherstone S. (2008)
Genetic parameters of production traits in Atlantic salmon (Salmo salar)
Aquaculture 274 (2008) 225 - 231

Quaas, R.L., Pollak, E.J., (1980).
Mixed Model Methodology for Farm and Ranch Beef-Cattle Testing Programs.
Journal of Animal Science 51, 1277-1287.

Quinton C. D., McMillan I., Glebe B. D., (2005)
Development of an Atlantic salmon (*Salmo salar*) genetic improvement program: Genetic parameters of harvest body weight and carcass quality traits estimated with animal models
Aquaculture 247 (2005) 211-217

Ramstad A., Midtlyng P.J. (2008).
Strong genetic influence on IPN vaccination-and-challenge trials in Atlantic salmon Salmo salar L.
J Fish Dis. 31: 567-578.

Reid D.P., Szanto A., Glebe B., Danzmann R.G. & Ferguson M.M.(2005)
QTL for body weight and condition factor in Atlantic salmon (Salmo salar): comparative analysis with rainbow trout (Onchorhynchus mykiss) and Arctic charr (Salvelinus alpinus).
Heredity 94, 166–72.

Reno, P.W (1999**)**
Infectious Pancreatic Necrosis and Associated Aquatic Birnaviruses., in *Viral, Bacterial and Fungal Infections, pages 1-55, Series : Fish Diseases and Disorders (Eds Woo, P. T. K and Bruno, D. W.), CAB International.*

Roberts R.J. & Pearson M.D. (2005)
Infectious pancreatic necrosis in Atlantic salmon, Salmo salar L.
Journal of Fish Diseases 28, 383–390.

Robertson A. & Lerner I.M. (1949)
The heritability of all-or-none traits - viability of poultry
Genetics 34, 395-411

Rye, M., Gjerde, B., (1996).
Phenotypic and genetic parameters of body composition traits and flesh colour in Atlantic salmon, *Salmo salar.*
Aquaculture Research 27, 121–133.

Salte R., Galli A., Falaschi U., Fjalestad K.T., and Aleandri R., (2004)
A protocol for the on-site use of frozen milt from rainbow trout (Oncorhynchus mykiss Walbaum) applied to the production of progeny groups: comparing males from different populations.
Aquaculture, 231:337-345

Sawalha R., Conington J., Brotherstone S., Villanueva B. (2007)
Analysis of lamb survival of Scottish Blackface sheep
Animal 1 : 151-157

Schall R., (1991)
Estimation in Generalised Linear Models with Random Effects
Biometrika 78 : 719-727

Simianer, H., Gjerde, B., (1991).
Estimating variance-components from fullsib group means.
Journal of Animal Breeding and Genetics 108, 270 – 279.

Smail D.A., McFarlane L., Bruno D.W., & McVicar A.H. (1995)
The pathology of an IPN-Sp sub-type (Sh) in farmed Atlantic salmon, Salmo salar L, post-smolts in the Shetland Isles, Scotland.
Journal of Fish Diseases 18, (6) 631-638.

Smail D.A., Bain N., Bruno D.W., King J.A., Thompson F., Pendrey D.J., Morrice S. & Cunningham C.O. (2006)
Infectious pancreatic necrosis virus in Atlantic salmon, Salmo salar L., post-smolts in the Shetland Isles, Scotland: virus identification, histopathology, immunohistochemistry and genetic comparison with Scottish mainland isolates.
Journal of Fish Diseases 29, 31–41.

Snedecor, G. W. and Cochran, W. G. (1989)
Statistical Methods,
Edition 8, page 204, Equation 11.7.2, Iowa State University, Ames.

Solberg, C. (2000)
Diode-Array NIR analysis of whole and filleted salmon
in : Near Infrared Spectroscopy, Davies A.M.C., and Giangiacomo R., eds NIR Publications, pp 889-891

Solberg, T.R., Sonnesson, A.K., Woolliams, J.A., Meuwissen, T.H.E.,(2008)
Genomic selection using different marker types and densities.
Journal of Animal Science, 86, 10, 2447-2454

Solberg, T.R., Sonnesson, A.K., Woolliams, J.A., Meuwissen, T.H.E.,(2009a)
Reducing dimensionality for prediction of genome-wide breeding values.
Genetics Selection Evolution, 41:29

Solberg, T.R., Sonnesson, A.K., Woolliams, J.A., Odegard, J., Meuwissen, T.H.E.,(2009b)
Persistence of accuracy of genomic-wide breeding values over generations when including a polygenic effect.
Genetics Selection Evolution, 41:56

Sonesson A.K., (2005)
Possibilities for marker-assisted selection in aquaculture breeding schemes. *In : Marker-Assisted Selection: Current Status and Future Application in Crops, Livestock, Forestry and Fish, pp 309–328 edited by E. P. Guimaraes, J. Ruane, B. D. Scherf, A. Sonnino and J. D. Dargie. FAO, Rome.*

Sonesson A K, (2007)
Within-family marker-assisted selection for aquaculture species
Genet Sel Evol, 39, 301-317

Song H., Santi N., Evensen O., Vikram N. V. (2005)
Molecular determinants of Infectious Pancreatic Necrosis Virus Virulence and Cell Culture Adaptation.
Journal of Virology 49 (2005) 10289 - 10299

Storset, A., Strand, C.,Wetten, M., Kjoglum, S., Ramstad, A., (2007).
Response to selection for resistance against infectious pancreatic necrosis in Atlantic salmon (Salmo salar L.).
Aquaculture 272 (Supplement 1), S62–S68.

Sullivan M. S., **Guy D. R.**, Roberts R. J., Manchester N. J. (2007)
The Aetiology of Spinal Deformity in Atlantic Salmon Salmo salar L. : Influence of genetic factors on the frequency and severity of the condition in freshwater stages.
Journal of Fish Diseases (2007) vol 30, 753-758

Taggart, J.B., (2007)
FAP: an exclusion-based parental assignment program with enhanced predictive functions.
Mol. Ecol. Notes 7: 412–415.

Tier B., (2010)
Genomic selection: promises and propriety
J. Anim Breed. Genet 127, 169-170

Thompson, R. (1979)
Sire evaluation.
Biometrics 35:339

Wetten T., Aasmundstad T., Kjøglum S., Storset A. (2007).
Genetic analysis of resistance to infectious pancreatic necrosis (*Salmo salar L.*).
Aquaculture 272: 111–117.

White, I.M.S., Roehe, R., Knap, P.W., Brotherstone, S., (2006).
Variance components for survival of piglets at farrowing using a reduced animal model.
Genetics Selection Evolution, 3,8 359-370.

Winkelman, A.M., Peterson, R.G., (1994).
Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in chinook salmon.
Aquaculture 125, 17-30.

Wright S., (1934)
An analysis of variability in number of digits in an inbred strain of guinea pigs
Genetics 19 : 506-536

_____ *end of thesis* _____