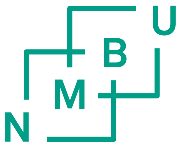


N
M
B
U

ACKNOWLEDGMENTS

This study was financed by Norwegian University of Life Sciences, with agreement to accomplish it at Nofima AS. Data for analysis was provided by industrial company SalMar AS.

First and foremost I would like to express most sincere thanks to my supervisor Dr. Bjarne Gjerde for always finding time to answer any question, efficiency in reading drafts of the thesis and positive criticism, not forgetting his sense of humor with which he was lightening the days for me and colleagues in the department.

Big thanks to both of my co-supervisors, Dr. Matthew Baranski for his help whenever I needed it and for his warm attitude to everybody and Dr. Hooman Khaleghi Moghadam for providing additional data for analysis and his non-vanishing good mood.

Special thanks to Dr. Solomon Antwi Boison for helping me managing the data, as well as to Dr. Celeste Jacq and Dr. Luqman Mohammad Aslam for their patience in explaining things which were new to me and for their constructive advices.

I express my sincere gratitude to NMBU student advisor Stine Telneset, who was helping me all the way during studies. Also I want to say thank you to PhD student Siri Storteig Horn and senior engineer Katrine Hånes Kirste for their friendship and for teaching me Norwegian language and habits. I express my gratitude for all my colleagues at Nofima Marine division, for cozy and friendly working environment.

Very big and warm thanks goes to my family and kids for being extremely patient, supportive and helpful during the whole studying period.

ABSTRACT

Pancreas disease (PD) is an emergent disease which causes large economical losses to Atlantic salmon industry and severely affects welfare of fish. Two different subtypes of the PD virus, SAV2 and SAV3, affect Atlantic salmon in Norway. Even subtype SAV2 don't produce high mortalities, affected individuals reduce their growth and are left with lesions which affect carcass quality. Currently only between-family selection is applied to this trait, which reduces accuracy and selection intensity for resistance to PD. Methods of marker assisted selection (MAS) would improve results of genetic selection of resistance to PD as within family selection would be possible. The aim of this study was to detect linkage of single nucleotide polymorphism (SNPs) with QTLs related to resistance to PD. Fry of Atlantic salmon, from crosses of three unrelated populations, were challenged with SAV2 in fresh water and mortalities were recorded during the test period. In addition, tissue samples were collected from a subset of individuals in the test and genotyped using a custom 57K SNP array. Fry in the test were assigned to possible family groups using estimates of genomic identity by descent (IBD) relationship calculated from SNP data. Heritability of survival to PD, analyzed as a binary trait, was estimated from genotyped individuals using a genomic relationship matrix constructed by two methods. Both methods resulted in similar values, revealing a moderate heritability ($h^2 \sim 0.2$). Genome wide association analysis (GWAS) from genotyped individuals revealed twenty-six putative QTLs. Four of the putative QTLs explained on average 10.9% of the genetic variance (~ 2.5% of the phenotypic variance). Genes related to immune response, metabolism and brain damage (in humans) were found in the area of chromosome twenty-one where the QTL was found. Twenty-two of the significant SNPs were classified as "lonely significant markers" as no SNPs in the neighbor region show association with the trait. In addition, because of the large number of duplicated areas in the Atlantic salmon genome, the exact location in the genome of "lonely significant markers" was put on doubt, leading us to not consider them suitable for MAS. Because of the limited amount of variance explained by the putative QTLs and the uncertainty of the value of the twenty-two "lonely significant markers", the benefit of MAS may be limited. Genomic selection may be more effective tool to increase accuracy and genetic gain in resistance to PD.

CONTENTS

ACKNOWLEDGMENTS I
ABSTRACT II
CONTENTS III
LIST OF TABLES AND FIGURES V
INTRODUCTION 1
LITERATURE REVIEW 5
IMPORTANCE AND GENERAL DESCRIPTION OF PANCREAS DISEASE 5
PREVALENCE, OUTBREAKS AND CONTAMINATION PATHWAYS 5
DISCOVERY AND GEOGRAPHICAL DISTRIBUTION 5
TRANSMISSION 8
VIRAL GENOME STRUCTURE AND RESEMBLANCE BETWEEN SUBTYPES 9
CLINICAL SIGNS AND HISTOLOGY OF AFFECTED ORGANS 9
PANCREAS DISEASE INFLUENCE ON FARM ECONOMY 11
RECOMMENDED PREVENTION METHODS 12
GENERAL MANAGEMENT 13
VACCINATION 13
GENETIC IMPROVEMENT 15
SELECTIVE BREEDING 16
ESTIMATION OF GENETIC PARAMETERS IN SELECTIVE BREEDING 17
SELECTIVE BREEDING PROGRAMS IN AQUACULTURE 18
MAS 19
MATERIALS AND METHODS 25
FISH MATERIAL 25
CHALLENGE TEST 25
SNP GENOTYPING 26
QUALITY CONTROL AND SNP FILTERING 26
PARENTAGE ASSIGNMENT AND FAMILY GROUPS 27
PRINCIPAL COMPONENT ANALYSIS 27
VARIANCE COMPONENTS AND HERITABILITY ESTIMATION 28
GENOME WIDE ASSOCIATION ANALYSIS USING LINEAR MIXED ANIMAL MODEL 29
QUANTILE-QUANTILE (Q-Q PLOT) PLOT AND INFLATION FACTOR (LAMBDA) 31
VARIANCE EXPLAINED BY SNP 31

BIOINFORMATICS AND CANDIDATE GENES 32

RESULTS 33

CHALLENGE TEST OUTCOME, PHENOTYPES 33

GENOTYPING QUALITY CHECK 36

PRINCIPAL COMPONENT ANALYSIS 37

ESTIMATED VARIANCE COMPONENTS AND HERITABILITIES 38

GENOME-WIDE ASSOCIATION RESULTS 38

QUANTITATIVE TRAIT LOCI (QTL) 41

QUANTILE-QUANTILE (Q-Q) PLOT 42

SUMMARY STATISTICS OF SIGNIFICANT SNPS ON CHROMOSOME 21 42

BIOINFORMATICS AND CANDIDATE GENES 46

DISCUSSION 49

CONCLUSIONS 55

CITED LITERATURE 57

APPENDIX 63

LIST OF TABLES AND FIGURES

Table 1. QTLmapped for disease resistance traits in aquaculture species (Gjerde et al., 2011). 22
Table 2. Variance component and heritability estimate for pancreatic disease (PD) using genomic relationship matrices computed based on VanRaden (2008) and Yang et al., (2010). 38
Table 3. Marker name and position, allele frequency (p), allele substation effect (a) and the percentage of variance captured by four significant SNPs on chromosome 21. 43
Table 4. Summary information on genotype counts and allele frequency for the dead (MAF-D) and survivors (MAF-S). Allele count of the minor allele (left), heterozygotes (middle) and major allele (right) for the four significant SNPs on chromosome 21. 43
Table 5. Candidate genes in the region 34.5-35.9 Mb on chromosome 21. Only markers in the protein coding genes are presented. 47

Figure 1. Map showing the current distribution of the different subtypes of Salmonid alphavirus (SAV). 6
Figure 2. Number of PD outbreaks in Norway and their location.. 7
Figure 3. Map of suspected and confirmed cases of pancreas disease (PD) in Norway in 2012.. 8
Figure 4. Pancreas disease in Atlantic salmon infected by a Norwegian subtype 2 related salmonid alphavirus, light microscopy. 10
Figure 5. Healthy salmon compared with a PD affected fish of same age.. 11
Figure 6. Fillets of farmed Atlantic salmon diagnosed with pancreas disease (PD) at slaughter.. 12
Figure 7. Vaccination side effects on salmon welfare and fillet quality.. 14
Figure 8. Severe side effects of vaccination, expressed in the PD vaccinated Atlantic salmon. 15
Figure 9. Example of accumulation of genetic gain in aquaculture. 16
Figure 10. Organogram showing the main elements in a fish breeding program. 19
Figure 11. Mortality profiles of challenge test.. 33
Figure 12. Percentage mortality per family. 34
Figure 13. Heatmap of identity by descent genomic relationship matrix (GIBD) obtained with PLINKv1.9. 35
Figure 14. Plot of within groups sum of squares and number of K- clusters.. 36
Figure 15. Principal component analysis (PCA) for the three distinct populations. 37
Figure 16. Manhattan plot of genome-wide –log10 p-values for resistance to PD –SAV2. 40
Figure 17. Plot of genome-wide –log10 p-values of SNPs on chromosome 21. 41
Figure 18. Quantile-quantile (Q-Q) plot for the observed and expected genome wide –log10 p-values. 42
Figure 19. Plot of genome wide $-\log_{10}$ p-values and linkage disequilibrium (LD) R^2 values on the 30-40 Mb region on chromosome 21. 45
Figure 20. Genes from region 34.5Mb -35.9Mb on chromosome 21. 46

INTRODUCTION

Aquaculture is one of the biggest industries in Norway, which produces about 1.4 million tons of various fish species per year. Almost ninety-five percent of Norway's aquatic production is Atlantic salmon, which corresponds to a first-hand value of 44.4 million NOK (Statistics Norway, 2016). Reared salmon production has increased from 0.5 million tons in 2003, to 1.38 million tons in 2015. Increased production comes from an increasing number of farming sites and there is a trend to have higher production densities per farming site. The combination of both factors is likely to cause higher susceptibility to diseases. Therefore, diseases became one of the most limiting factors and biggest risks for optimal husbandry.

There are several common diseases for Atlantic salmon in Norway, which cause large economical losses. One of them is pancreas disease (PD) which remains one of the major problems in the Norwegian Salmon industry. Pancreas disease is a viral disease which is caused by *Salmonid alphavirus* (SAV) (Weston et al., 2002) (also known as *Salmon pancreas disease virus* (SPDV) (Nelson et al., 1995) and *Sleeping disease virus* (SDV) (Castric et al., 1997), and belongs to the *Togaviridae* family. PD first occurred in Scotland in 1976 (Munro A.L.S., Ellis A.E., McVicar A.H., 1984)). The virus itself was discovered, isolated and characterized about 20 years ago by Nelson et al. (1995). The virus which causes PD is now divided into 6 subtypes according to the differences in viral genotypes (Fringuelli et al., 2008). The different subtypes are found in different geographical locations where the disease occurs and affects the salmon with varying degrees of severity.

For many years Atlantic salmon in Norway were affected by only the SAV3 subtype, and in the south region it was declared endemic (Stormoen et al., 2013). From 2011 outbreaks from SAV2 started in the central part of the country (Hjortaas et al., 2013), but is constantly spreading and is becoming prevalent in central and even northern Norway. Outbreaks of PD occur each year in endemic zones and result in millions of kroner of losses to the industry.

While the number of PD outbreaks in Norway is rising and most of the southern regions are endemic (with SAV3), a few vaccines are now available and used in commercial farming to avoid disease outbreaks and keep the fish healthy (Sommerset et al., 2005). However, the PD vaccines have little effect compared to vaccines against bacterial diseases (Gudding et al., 2014). Furthermore vaccination has negative side effects, such as adhesions of internal organs or melanin spots on bellies (Drangsholt et al., 2011) and in the filet (Larsen et al., 2014). Sometimes these effects are lethal (Poppe and Knudsen, 2005).

Due to the high cost and limited effectiveness of the vaccines against PD, selective breeding for disease resistance has become very popular among the salmon producing companies. For many years classical methods has been applied for selection and breeding of Atlantic salmon which led to significant increase in production. Although successful, classical methods of selection had a number of limitations. Selection using only phenotypic records limit the genetic progress for the traits which are difficult to measure and/or can only be recorded late in life (e.g. fertility, feed efficiency, longevity), require animals to be sacrificed (e.g. meat quality), or challenged for pathogens. Applying conventional methods of family selection for disease resistance also limits the selection that could only be performed between the families and the variation within families could not be explored, reducing the selection potential through a reduced selection differential. As individual genetic values can't be estimated in family breeding programs, estimations of genetic merit would be less accurate compared to state of the art available advanced selection method, - genomic selection (GS). Genomic selection (GS) is an advanced methodology by which breeding values of individuals for complex traits are predicted by combining statistical methods with genome-wide distributed genetic markers (Meuwissen et al., 2001).

Studies utilizing genomic tools such as SNP arrays and genome resequencing have led to the identification of chromosomal regions and genes affecting important commercial traits, among them pathogen/disease resistance. The most high-profile example of this is the discovery of the gene underlying a major quantitative trait loci (QTL) that explains nearly all genetic variation for infectious pancreatic necrosis (IPN) virus resistance in salmon, where a combination of high-density genotyping and whole-genome resequencing were among the methods used to identify the likely causative gene and mutations (Moen et al., 2009). Gene discoveries are of high value to the aquaculture industry for characterizing the precise genetic mechanisms that cause variation in a trait(s) and also to improve our understanding on both innate and adaptive immunology.

Due to an increased knowledge of the salmon genome, high-density SNP arrays containing large number of SNPs are now widely available in Atlantic salmon (both as commercial products and as customer developed resources). This has facilitated advances in applying GS approach. However, the implementation of GS in family-based salmon breeding programs is in its infancy when compared to terrestrial livestock species and require advancements in resources and methods for efficient utilization in breeding schemes. Resistance against PD is reported to be moderately heritable trait (Norris et al., 2008), and selective breeding, using genomic information, would be a powerful tool for creating Atlantic

salmon populations, highly resistant to PD. Identifying trait associated markers and implementing marker assisted selection in salmon breeding allow estimation of breeding values for all the individuals (within and between a families) with relatively higher accuracy and ultimately greater response to selection.

The objective of this study was to identify quantitative trait loci (QTL) for PD by running genome wide association analysis (GWAS), and possibly to use the information in marker assisted selection (MAS).

LITERATURE REVIEW

Importance and general description of pancreas disease

Pancreas disease (PD) is a severe infectious disease making big impact on salmonid aquaculture in Norway (Houston et al., 2010). This disease affects big range of farmed Atlantic salmon *Salmo salar* L. in sea water (pancreas disease, PD) and rainbow trout *Oncorhynchus mykiss* (Walbaum) in fresh water (sleeping disease, SD) in the country. Pancreas disease is caused by Salmon Pancreas Disease Virus (SPDV) (Nelson et al., 1995) and Sleeping disease virus (SDV) (Castric et al., 1997), which were treated as different viruses until Graham et al. (2014) studies have shown that these two viruses serologically are very closely related. Nowadays, when more advanced methods for analyzing genomes and genomic relationships are available, viruses are often named based on genomes and genetic relationships, and therefore the new name Salmonid alphavirus (SAV) (Weston et al., 2002) was launched (T.Taksdal, personal communication, 15 June 2016) and used in scientific publications since. Based on genetic differences, at present SAV is divided into six different subtypes named from 1 to 6 (Fringuelli et al., 2008).

Pancreas disease affects first year Atlantic salmon smolts (usually from May to September) and causes significant economic losses in fish farming, due to high morbidity and mortality in outbreak sites. Dependent on the SAV subtype, mortality levels due to PD show great diversity between farms (Stormoen et al., 2013) and varies from 0.1% to over 60% (Menzies et al., 1996; Desvignes et al., 2002; McLoughlin et al., 2003; Rodger and Mitchell, 2007; Fringuelli et al., 2008).

Prevalence, outbreaks and contamination pathways

Discovery and geographical distribution

Pancreas disease (PD) in farmed Atlantic salmon was first recorded and described in Scotland in 1976 by Munro A.L.S., Ellis A.E., McVicar A.H. (1984). Later on it has been diagnosed in North America (Kent and Elston, 1987), Norway (Poppe et al., 1989) and Ireland (Murphy et al., 1992). Some outbreaks were also reported in France, Spain and Italy (Raynard, R.;Houghton, G.;Munro, 1992; Graham et al., 2007). Disease is strictly distributed

over geographical locations, according to the SAV subtype (Figure 1). Subtypes do not overlap within the farming sites.

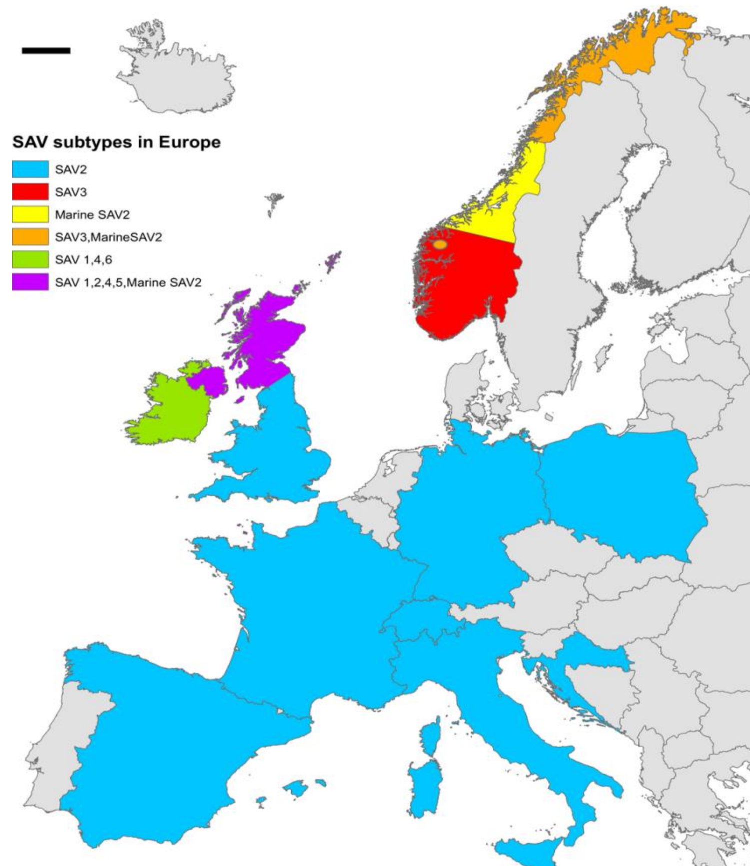


Figure 1. Map showing the current distribution of the different subtypes of Salmonid alphavirus (SAV) (Jansen et al., 2016).

Subtypes 1, 4, 5 and 6 have been discovered only in Scotland or Ireland (Fringuelli et al., 2008; Graham et al., 2012). Subtype SAV3 has been found only in Norway (Hodneland et al. 2005; Fringuelli et al. 2008; M. Karlsen et al. 2006; McLoughlin and Graham 2007; J. H. Weston et al. 2005; Jansen et al. 2010). SAV subtype 2 has been reported in France, England and Germany (Graham et al., 2003; Bergman et al., 2005) and was first recognized as a rainbow trout fresh water disease (Villoing et al., 2000; Castric et al., 1997). Marine variant of SAV2 was recorded in Scotland in reared Atlantic salmon (Fringuelli et al., 2008) and later on in Norway (Hjortaas et al. 2013).

Until recently there has been only subtype three (SAV3) recognized in Norway, in farmed Atlantic salmon (Hjortaas et al., 2013). First appearance of pancreas disease in Norway was observed in the 1980s (Poppe et al., 1989), but since 1990s outbreaks were recorded every year in salmonids (Kristoffersen et al., 2009). Pancreas disease was spreading

markedly through the farming sites since 2002 (Kristoffersen et al., 2009) and became a serious problem for fish farming industry due to large economical losses. Therefore, in 2007 PD was included into list of notifiable diseases by the Norwegian Food Safety Authority (Skjelstad et al., 2007). To limit the rapid spread of SAV, in 2008 Norway was divided into endemic zone south and a non-endemic zone north (Stormoen et al., 2013). Hustadvika was set as border line between those two regions (Jensen et al., 2012) (Figure 2).

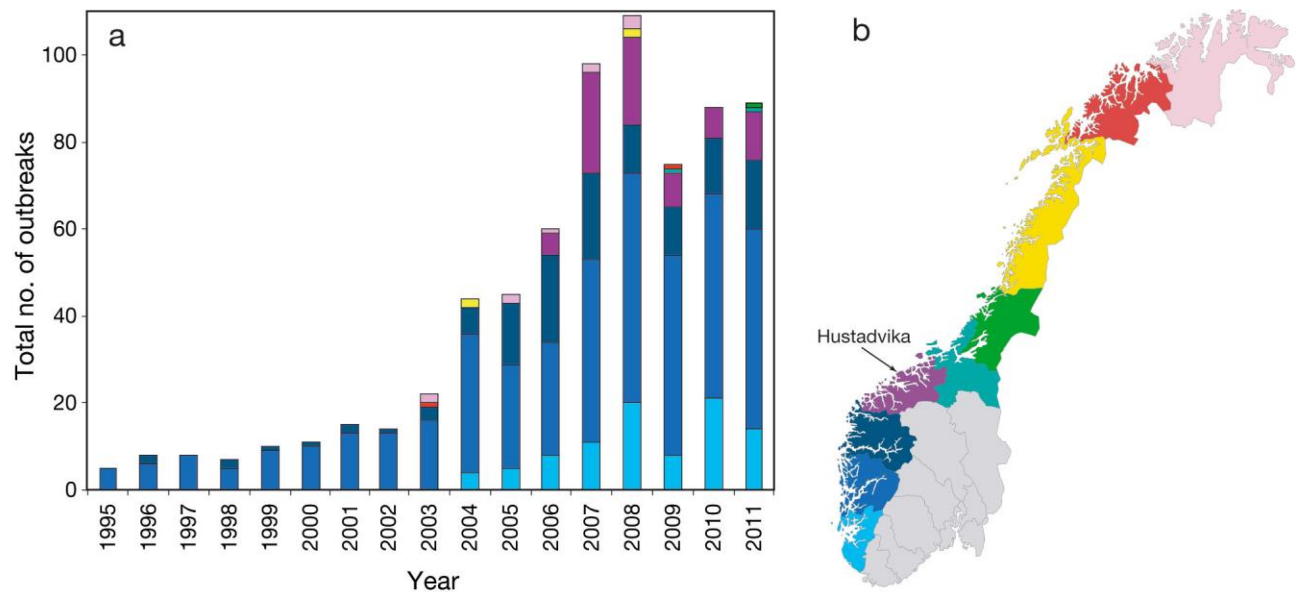


Figure 2. Number of PD outbreaks in Norway and their location. (a) Annual number of all outbreaks of pancreas disease between 1995 and 2011, divided by counties.(b) Location of each county, with colors corresponding to the graph (data collected by the Norwegian Veterinary Institute). No pancreas disease has been reported in counties shown in grey (Jensen et al., 2012).

However, this border between SAV diseased and disease-free zones have been crossed together with SAV2 coming to Norway in 2009 (V. Lund, personal communication, 17 November 2015) and spreading in mid-Norway in late 2011 (Hjortaas et al., 2013). Therefore another endemic area was added for SAV2 in mid-Norway since 2010 (Jansen et al., 2015; Hjortaas et al., 2016) (Figure 3).

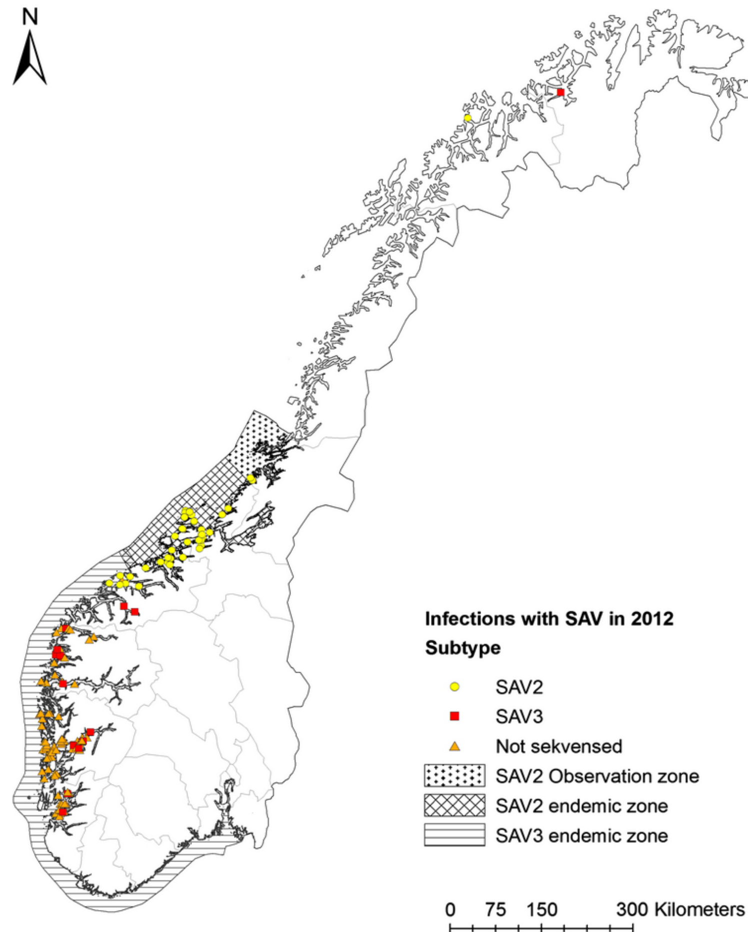


Figure 3. Map of suspected and confirmed cases of pancreas disease (PD) in Norway in 2012. The SAV subtypes, together with the defined endemic, and observation zones are shown (Jansen et al., 2015).

Transmission

PD virus appears to be spreading by horizontal transmission via sea water (Taksdal et al., 2015; Raynard and Houghton, 1993; Fringuelli et al., 2008). However, the true contamination pathways are not completely clear (T. Taksdal, personal communication, 10 June 2016). However there is not strongly supported that SAV3 appears more often when there are outbreaks in the neighbor farm sites (Rodger and Mitchell, 2007; McLoughlin et al., 2003). Also there is no clear knowledge of why SAV2 and SAV3 do not overlap within the same farming sites. There was suggested that SAV can be transmitted through transport of infected smolt (Bratland and Nylund, 2009; Karlsen et al., 2006) or via sea lice (La Linn et al., 2001). There is a lack of evidence for the vertical transmission of the virus (Jansen et al., 2010; Bratland and Nylund, 2009; Kongtorp et al., 2010). Also has been shown that virus

without host does not survive long (up to 6 days at 10 ^{o}C) in the sea water (Fringuelli et al., 2008).

Viral genome structure and resemblance between subtypes

PD is caused by *salmonid alphavirus* (SAV) which belongs to family *Togaviridae* (Nelson et al., 1995; Weston et al., 2002). The SAV genome consists of a 11-12 kb length single-stranded RNA, with two large open reading frames, where first one encodes for four non-structural proteins nsP1 to nsP4 and the second encodes for the structural proteins E1 to E3, TF and 6K (Hjortaas et al., 2016; Weston et al., 2005; Karlsen et al., 2009).

All 6 SAV subtypes are immunologically similar but slightly different in the viral genomes. Genes E2 and nsP3 were sequenced (Hjortaas et al., 2013) and assignation of virus to each subtype is based on phylogenetic analysis of the E2 gene (Fringuelli et al., 2008). Main differences between viral genomes appear in regions within the E1, nsP4 and nsP3 genes, (Weston et al., 2005).

By comparing nucleotide E2 and nsP3 sequences in different subtypes, there were found 0 to 4.8% differences within SAV2 group and 7.8 to 19.7 % difference in viral sequences between SAV2 and other five SAV subtypes (Fringuelli et al., 2008), where SAV2 differed from SAV3 by 7.1% in nucleotide sequence (Karlsen et al., 2009). Marine SAV2 subtype genome sequence showed very high resemblance to English and Scottish marine SAV2 variant which suggests very close relationship between mentioned viral strains and possibly the same origin source (Hodneland et al., 2005; Fringuelli et al., 2008; Hjortaas et al., 2016). As SAV2 from Scotland and Norwegian SAV2 nucleotide sequences were almost identical, this suggests that the virus could be brought from Scotland to Norway(Graham et al., 2012; Hjortaas et al., 2016).

Clinical signs and histology of affected organs

PD outbreaks affects smolts after 3-10 months in the sea (Hodneland et al., 2005; Munro A.L.S., Ellis A.E., McVicar A.H., 1984). Clinical signs of disease can be noticed after 4-6 weeks of smolt transfer to the sea water from the reduced movement (lethargy) or impaired swimming performance (cannot maintain stable position in water), increased faecal casts, sudden loss of appetite (for 5-6 weeks) and mortality (McLoughlin and Graham, 2007; Graham et al., 2007; Poppe et al., 1989).

SAV virus is causing severe histopathological changes in pancreas (pancreatic acinar cell loss), heart (cardiac degeneration and inflammation), kidney and skeletal muscle (degeneration and fibrosis) of infected fish (Herath et al., 2012; Graham et al., 2011, 2007; McLoughlin and Graham, 2007; Taksdal et al., 2007; Hodneland et al., 2005; Hjortaas et al., 2016). Tissue leisions are shown in Figure 4. Pancreas tissue (Taksdal et al., 2015) as well as muscle (Lerfall et al., 2012) can fully recover in up to 80% of salmon in the population (Munro A.L.S., Ellis A.E., McVicar A.H., 1984).

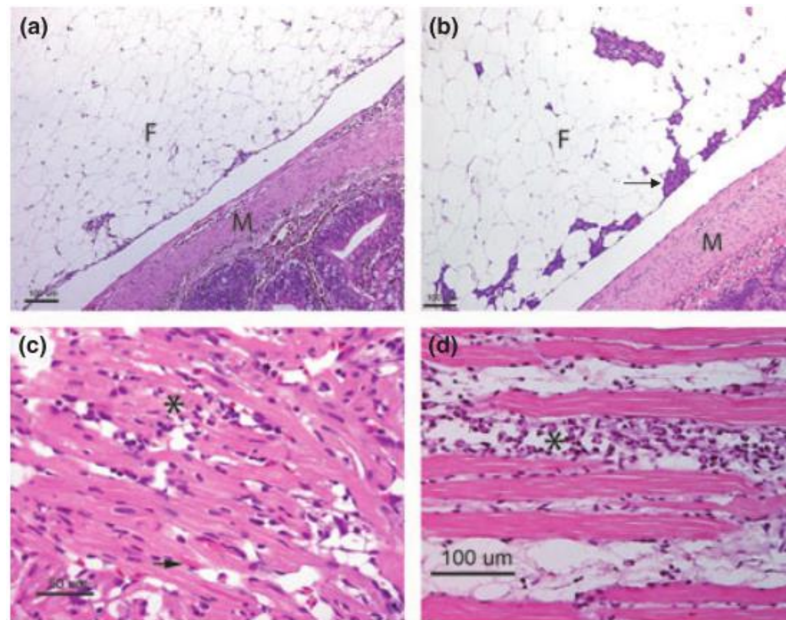


Figure 4. Pancreas disease in Atlantic salmon infected by a Norwegian subtype 2 related salmonid alphavirus, light microscopy. (a) Loss of pancreatic tissues, F, fat tissue; M, muscularis of a pyloric caecum. (b) Normal exocrine pancreatic tissues (arrow) in perivisceral fat tissue in a non-affected fish. (c) Heart ventricle, spongious part: mild inflammation (*), necrotic myocyte (arrow). d) Inflammation (*) in red skeletal muscle. Bars a, b, d: 100 lm, bar c: 50 lm.(Hjortaas et al., 2013).

Clinical signs and histopathology of SAV2 and SAV3 are similar (T. Taksdal, personal communication, 10 June 2016), however due to higher virulence of SAV3, reduction in appetite is more severe than in salmon affected with SAV2 and induces longer duration of inappetance period (Jansen et al., 2015; Fringuelli et al., 2008). However, mortality levels during outbreaks are variable.

Pancreas disease influence on farm economy

During and after PD outbreak fish experience increased susceptibility to other diseases and parasites (sea lice, gill problems), also when the fish stop eating, they stop growing and their shape is changing, as well as fillet quality is reduced (white spots in the flesh). Recovery period is long (up to 8 months) and fish from acute form can get into chronic (according to Munro A.L.S., Ellis A.E., McVicar A.H. (1984) around 20-30%) where they have low immunity and reduced growth during the whole time in the sea before slaughter. Moreover, salmon which has been affected by PD cannot be sold in the market (V. Lund, personal communication, 15 November 2015), as growth reduction (Figure 5) together with damage in the red muscle, are reflecting on the fillet quality (Figure 6).

Figure 5. Healthy salmon compared with a PD affected fish of same age. (Photo: Trygve Poppe) (Stene, 2013).

Figure 6. Fillets of farmed Atlantic salmon diagnosed with pancreas disease (PD) at slaughter. From top to bottom: example fillet from group A=SAV negative and PD negative, B=SAV positive and PD negative, and C=SAV positive and PD positive (Larsson et al., 2012).

In most cases economic losses for SAV2 infected sites occur mainly because of reduced growth rates (Christie et al., 2007; Jansen et al., 2015) and feed conversion efficiencies (Graham et al., 2007; Taksdal et al., 2007), as for SAV3 mortalities are the main cause of loss in profits (Graham et al., 2007). Overall culling rate from PD sick and not recovered salmon usually makes up to 15% of the whole population (Munro A.L.S., Ellis A.E., McVicar A.H., 1984). Thus, the disease causes significant economic problems (Aunsmo et al., 2010; Hjortaas et al., 2016) to farmed Atlantic salmon industry in many countries. The amounts of these losses may vary within different years and farming sites (Munro A.L.S., Ellis A.E., McVicar A.H., 1984), depending of number of outbreaks and severity level of the disease and can reach to 1.8 million euros for one farming site in Norway (Aunsmo et al., 2010). Losses of 35 and 12 million euros were estimated in Ireland during 2003 and 2004 years respectively (Fringuelli et al., 2008) due to high mortality from PD.

Recommended prevention methods

Growth reduction and mortality gives big impact to the fish economy (Taksdal et al., 2015) and there is no treatment against PD, therefore prevention from this viral disease plays major role in avoiding the outbreaks. There are several methods to provide biosecurity, such

as general management, vaccination and selective breeding towards higher disease resistance. However, vaccination is also questioned in concerns of animal welfare.

General management

The PD risk factors have not been fully recognized and described (Kristoffersen et al., 2009), however, there are several factors which could be taken into consideration to reach higher level of prevention against the infection. As PD is transmitted horizontally, it may be spread via sea water directly infecting one fish from another, by transporting infected fish with the well boats (Rodger and Mitchell, 2007), or even via farm employees if they have not been keeping good hygiene and farm safety rules. Therefore good disinfection of well boats, farm equipment and top-up water, as well as hygiene or the staff in the farm is crucial. To avoid direct infection between farmed salmon, fish densities should not exceed the recommended figures and sea cages has to be distributed in adequate distance from each other. Proper quarantine and health check of the new-coming fish to the farm has to be ensured. Sea water currents and location of neighboring farming sites also might be a risk factor for higher probability of infection, especially if neighboring sites has outbreaks. Maintenance of water temperatures is a significant effect to PD outbreaks (Stormoen et al., 2013), as infection mostly occurs at higher temperatures (around 12°C, within April-September).

Vaccination

The number of outbreaks from PD has increased dramatically in 2007-2008, therefore the national vaccination program has been approved in 2008 (Ødegård et al., 2011) and the vaccine against PD (Norvax® Com-pact PD, Intervet International B.V.) was approved for commercial use (Jensen et al., 2012).

During cross-neutralization studies there were only small differences found in 6 SAV subtypes, as all of them belongs to the same virus species indicate that these subtypes are serologic closely related members of the same virus species (Graham et al., 2014). Therefore just one type of vaccine was developed against PD, which was tested on subtype SAV3 and showed good results in mortality reduction (Gudding et al., 2014).

Vaccination can provide immunity to farmed Atlantic salmon at least for 9 months but it is not clear if recovered fish carry virus inside for the rest of their lives or if they get constant immunity. To obtain long lasting effect of vaccination, mostly oil-adjuvant vaccines

are used. Unfortunately these vaccines have many side effects, which are affecting salmon welfare and carcass quality and sometimes so severe that causing mortality of fish (Figure 7, B). Most commonly internal organs are “glued” (Figure 7, A) and melanin spots appear in the final product (Figure 7 C, D). In addition, fish inoculated with oil-adjuvant vaccines reduce their appetite and growth (Drangsholt et al., 2011).

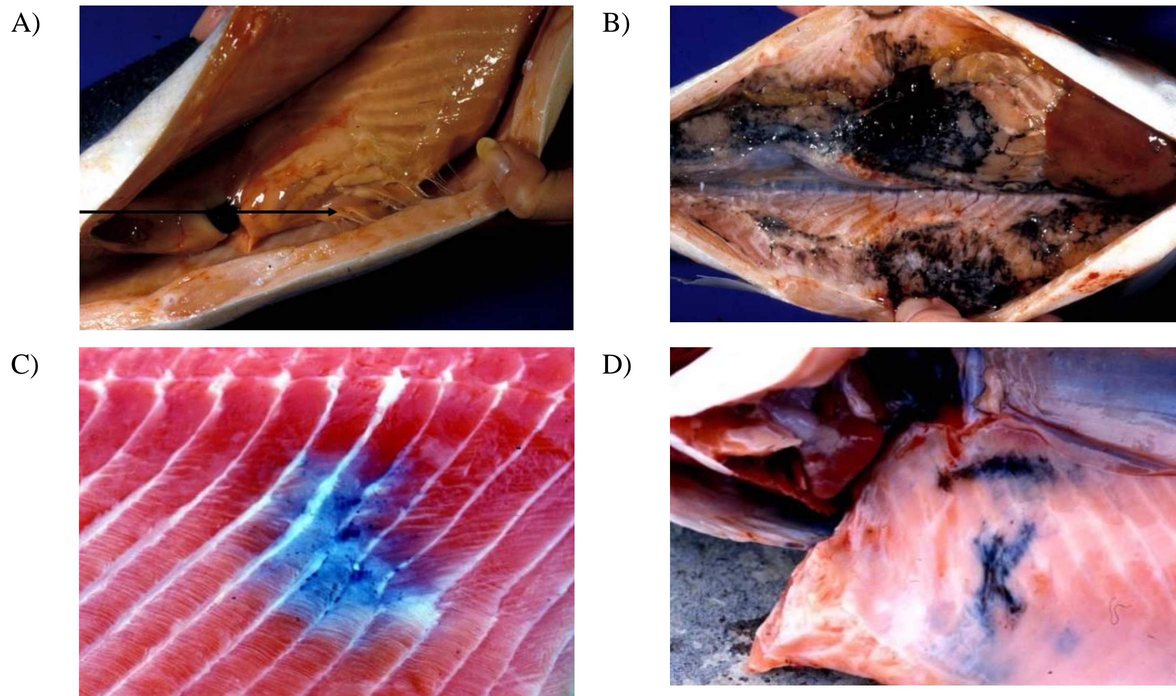


Figure 7. Vaccination side effects on salmon welfare and fillet quality. (A) Lesions within acceptable limits when it does not affect fish welfare and fillet can be sold in the market. (B) Severe lesions which have been occurring as a consequence of vaccination. (C), (D) Heavy melanization of the fillet, as a vaccination outcome. (Photo: Trygve Poppe) (Poppe and Knudsen, 2005).

Figure 8. Severe side effects of vaccination, expressed in the PD vaccinated Atlantic salmon.

Vaccination can provide partial immunity to farmed Atlantic salmon for 9 months but it is not clear if recovered fish are carriers of the virus for the rest of their lives or if they life lasting immunity.

As growth reduction in SAV2 is economically more important than mortality, and causes probably bigger economical losses, vaccination from SAV2 would not have the same economical effect as vaccination against SAV3 where mortalities are higher. Moreover, most of currently available vaccines produce side effects, which might be more severe than those caused by infection for PD. Therefore, other different methods to provide protection against PD must to be explored, including selective breeding programs for the increase of disease resistance against PD, by simultaneously improving general management.

Genetic improvement

Selective breeding is based on improvement of the genetic merit in animal. Genetic superior animals are identified and chosen as breeders in order to transmit their superiority to the next generations. In aquaculture, even if management and vaccination are good preventive methods, more sustainable and least polluting solution is to use genetically improved individuals which are bred towards higher resistance against disease. Selection improvement accumulates over generations (Figure 9), and together with other management methods, provides a more healthy and fast growing fish.

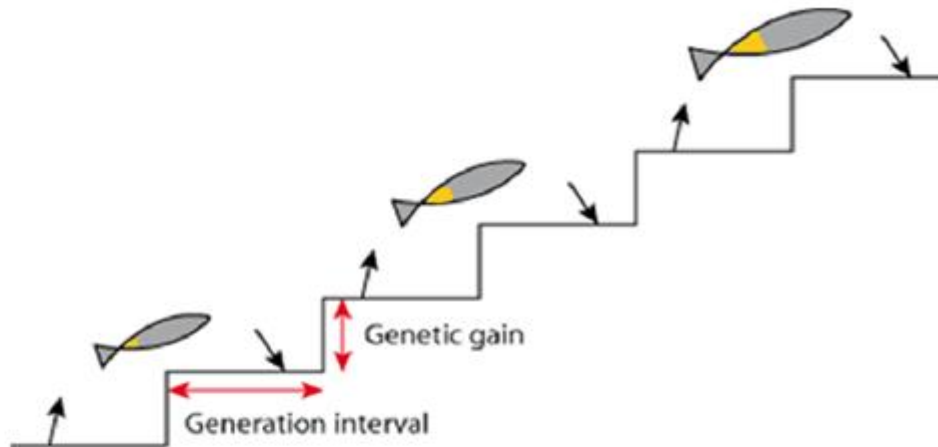


Figure 9. Example of accumulation of genetic gain in aquaculture.

Atlantic salmon in Norway, have been selected for faster growing for few generations and for many years was the only trait included into breeding programs. Later on, according to the market and industry demand, more traits have been included into breeding programs, including disease resistance. Due to high fecundity of Atlantic salmon and together with large variation in many economical important traits, family breeding programs have been proved to be effective in improving the perform of fish in farming conditions.

Selective breeding

As mentioned before, selective breeding is an effective method for improving disease resistance in farmed Atlantic salmon, which is crucial to economy, animal welfare and sustainability of the industry. Genetic selection is often based in the assumption that most of the traits under selection are polygenic, i.e. influenced by many genes, which in combination with environment determine the phenotypes. Genetic progress accumulates over generations improving the population under selection constantly. Selective breeding is possible thanks to advanced statistical methods that allow to identify genetic those individuals which performance is superior because genetic effects. Higher genetic gains can be achieved when strong selection intensity (i.e. the selection of a small number of individuals in proportion to the total available breeders) and high selection differential (large differences between selected individuals to the average performance of the population) and when the targeted trait has continuous range of variation. Genetic improvement in farmed Atlantic salmon population, is largely attributed to selective breeding programs that rely on highly heritable phenotypic traits, such as growth rate and disease resistance.

However, many of economically important traits in Atlantic salmon cannot be measured on the breeding candidates (e.g. carcass quality traits, disease resistance), but only on their relatives (full-sibs or half-sibs). When the genetic merit of an individual is measured only through the performance of relatives only half of the total genetic variation can be used as it makes impossible to differentiate between relatives from the same family for that trait, and at the same time reducing the selection differential, factors that ultimately would reduce selection intensity in the population. In addition those traits are expensive to measure increasing the cost of the programs.

Estimation of genetic parameters in selective breeding

In genetics, a quantitative trait is defined as a phenotype which is influenced by the cumulative effect of many genes and environment, A quantitative trait varies among individuals providing them statistical properties from a normal standard distribution. By applying specific statistical models, environmental effects and other genetic effects different to cumulative ones can be separated making possible to estimate the genetic merit of an individual. A fundamental measurement in selective breeding is heritability, which provides information about how much of phenotypic variation in the population is due to the additive (cumulative) genetic effects and is expressed as a proportion of additive genetic variance (part transmitted to the offspring, while epistatic and dominance effects are not taken into account) to the total phenotypic variance (sum of total genetic and environmental variances):

$$h^2 = \frac{\mathrm{Var}(A)}{\mathrm{Var}(P)}$$

Heritability is a population measure (has to be calculated for each population) and may vary from 0 to 1 and provides the basic bases for any breeding program. Heritability, as well as genetic and phenotypic variances, is used when predicting the response to selection and ultimately for estimate the genetic merit of an individual also known as estimated breeding values. High or moderate heritability of the trait shows good opportunities for selection as the higher estimate is, the more of phenotypic variance is explained by underlying genes and therefore lower environmental effect is expected. Low heritability indicates that the additive genetic contribution to a trait is small resulting in a limited genetic gain.

Studies in Atlantic salmon shown moderate to high heritabilities for bacterial and viral diseases, when estimated based on challenge test and field data (Ødegård et al., 2011).

Hence due to high heritabilities and biological aspects of Atlantic salmon, i.e. high fecundity and possibility to use smolts in breeding (which reducing generation interval), selection intensity is expected to be high.

Another important parameter in animal breeding is correlation, which can be calculated only for related traits and describes the relation between different trait. Correlation is very useful when performing indirect selection and in salmon breeding programs is used for indirect selection of survival (Gjedrem, 2004), as it has been found in many cases to be positively correlated with growth (increasing growth increases the survival). This kind of selection is desirable as can be performed on the easy measurable trait and can be recoreded on the selection candidate.

Selective breeding programs in aquaculture

To design an effectively working breeding strategy, traits for selection must be chosen carefully (according to market needs and genetic parameters) and best contributing individuals should be selected to become parents for the next generation. Farmed Atlantic salmon breeding is based on family (between-family) selection and individual (within-family) selection.

By performing family selection, estimated breeding values (EBV) are calculated to be identical for all individuals in the family. In this way many individuals from the same family have uniform breeding values for the certain traits and only the best performing families are selected for further breeding. Limitation of this selection method consists that individuals are assumed to have equal EBVs for each selection trait, which in reality is not true, as differences among individuals from the same family are expected. Families with low overall breeding values might contain valuable breeding candidates, which in family breeding programs are culled out and their contribution to the genetic pool for the following generation is eliminated.

Furthermore, in cases where the perform of individuals against a disease is tasted, survivors from such test would may made good selection candidates as a phenotypic record is available allowing to perform within family selection, but these individuals are usually excluded because the health risk that they represent.

However, advances in molecular biology and increased availability of genomic information, may provide family breeding programs with necessary tools to perform within family selection even when the trait cannot be measured in the candidates. Mainly, the use of

marker assisted selection through identification of quantitative trait loci and/or the use of genomic selection, may provide the necessary accuracy to select individuals within families given that affordable prices for the required genotyping can be achieved.

Typical breeding program in Atlantic salmon breeding is shown in Figure 10 and is used no matter of selection type.

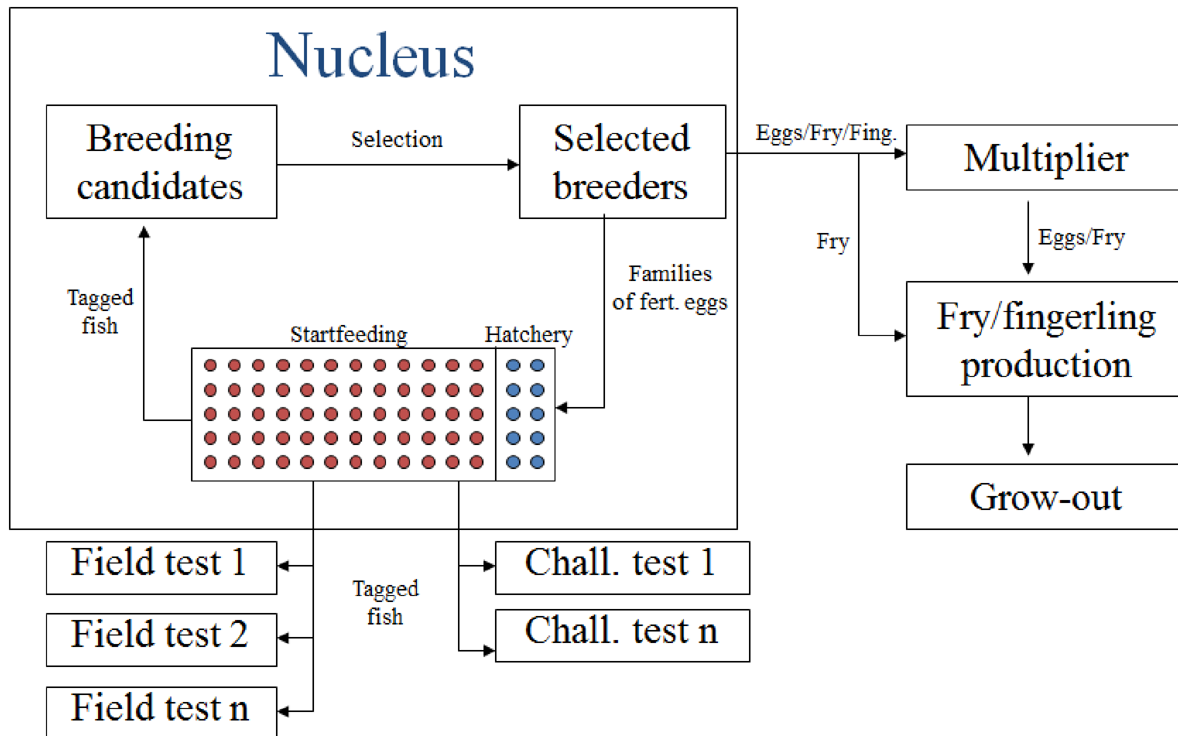


Figure 10. Organogram showing the main elements in a fish breeding program (Gjerde, 2004).

Marker assisted selection (MAS)

By using marker assisted selection (MAS) it is possible to differentiate the genetic merit of individuals from the same family, even if the trait cannot be measured in the same individual. When MAS was proposed, the main goal was to identify variations with large effects on the trait under interest. Lately Meuwissen et al. (2001) proposed the use of a methodology known as genomic selection (GS) this method proposed that when information of dense genetic markers across the genome is available, and a trait is influenced by many genes across the genome, at least one of the markers might be close enough to one of these genes making possible to estimate its effect, and by adding these small effects estimate the

genetic merit of an individual with higher accuracy and making possible to differentiate among individuals from the same family even when records of their phenotypes are not available. With advances in genotyping technology, information of thousands of genetic markers become possible at lower prices in many farming species including Atlantic salmon, because of that making feasible the use of GS in selective breeding programs (Hayes et al., 2006).

Genetic markers

Genetic markers may be defined are detectable variations of the genome, which possibly emerged due to mutation or alteration in the genomic loci. A genetic marker may be an alteration of a single nucleotide in DNA sequence (single nucleotide polymorphism, SNP) that occurs at a specific position in the genome or multiple bases such as variation in short or variable number tandem repeats and including phenotypic differences related to known polymorphism.

Nowadays the most commonly used genetic markers in aquaculture species are single nucleotide polymorphism (SNP), based on single base variations, and microsatellites, consisting of one to six base-pair repeats, sequencing nuclear and mitochondrial DNA. Other useful DNA markers are restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) and random amplification of polymorphic DNA (RAPD) (Aslam, 2012; Lien et al., 2016; Gjedrem and Baranski, 2009).

SNP based genotyping using microarrays has become the preferred method for genotyping because of their high number and distribution across the entire genome. Microarrays are relatively inexpensive, quick and easy to automatized and uses limited human intervention which reduces errors due manual work, for the contrary, microsatellite based genotyping is time consuming and usually result expensive, as the relative number of markers that can be genotyped is low compared with SNPs array. In addition, high density maps of microsatellites are not available for Atlantic salmon (Moen et al., 2004). However, SNPs are less informative than microsatellites due limited number of alleles, therefore bigger numbers are needed to obtain the same information (Gjedrem and Baranski, 2009).

Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) is a powerful tool for annotating phenotypic effects or mapping QTL on the genome by using single nucleotide polymorphism

(SNP) microarrays in gene-based selection (Cole et al., 2011). Genome-wide association analysis is a relatively new way to identify genes and QTLs involved in different traits of importance, including diseases. By GWAS is possible to check hundreds or thousands of SNPs simultaneously and identify associations between specific outcomes (i.e. diseased, dead or healthy) in individuals and in this way possibly to identify genes or genomic regions, involved in development of the certain diseases. A positive association arises when there is a greater frequency in the presence of a genetic variant in diseased individuals than in unaffected ones. As disease resistance is a complex trait and most likely affected by many genes, thus several QTLs usually are found across the whole genome and each of those genetic variants might provide different contribution to the trait.

Quantitative trait loci (QTL)

Quantitative trait loci (QTL) are regions in the genome that affects variation in the quantitative trait and genes in that genomic region control phenotypic expression of the trait. QTLs are usually mapped by applying GWAS with dense SNP arrays, searching for markers which correlate with an observed trait and in this way linking phenotypic data (trait measurements) and genotypic data (usually molecular markers). For that purpose phenotypic records of the trait on a sample population are necessary and a linkage map is essential for the mapping of QTL. If good association of QTL with the genetic marker is found and identified genomic region is explaining big part of genetic variance of the polygenic trait, identified QTL is preferably to be used in MAS as can increase selection response in animal breeding programs, especially for traits that are difficult to improve by traditional selection, such as disease resistance.

Mapping of QTL for disease resistance can be one of the approaches for getting better understanding about effects of the effects of the genes influencing the trait and providing more information on the location of those genes. Mapping design of QTL for disease resistance is based on information from challenge tests in which fish groups from different families are infected with pathogen, where survival is recorded and used in GWAS as a binary trait (Moen et al., 2007). Analysis of such data may be analyzed using various methods (Moen et al., 2004), but in all cases requires large number of families (Massault et al., 2008; Hayes et al., 2006) as only a small number of individuals per each family have phenotypic records for disease resistance.

QTL can be identified using a linkage map or by fine mapping (Gjedrem and Baranski, 2009), but in both cases thousands of markers and samples will be prerequisite to obtain sufficient power for detection as polygenic traits are usually controlled by several causative variants (and environment) and thus the detection process is complex.

Mapping of QTL for disease resistance traits has been studied in aquaculture species. However, only a small number of QTLs with higher importance for selective traits have been identified in several aquatic species, and in Atlantic salmon (Massault et al., 2008). Mainly QTLs for disease resistance have been identified, and all of these QTLs might be or are already used in MAS (Sonesson, 2007).

The biggest success was obtained with Atlantic salmon disease resistance against infectious pancreas necrosis (IPN), where major QTL was found by two independent studies (Houston et al., 2008) and (Moen et al., 2009) in Scottish and Norwegian Atlantic salmon populations. This QTL for IPN explains 80-98% of the genetic variance for disease resistance.

However for other diseases more QTLs with smaller effects seem to appear, for example (Moen et al., 2007) found a QTL for infectious salmon anemia (ISA), where identified genomic region explained only 6% of variation. (Table 1).

Table 1. QTLmapped for disease resistance traits in aquaculture species (Gjerde et al., 2011).

Species	Pathogen/parasite	QTLs found (PVE)*	References
Atlantic salmon (*Salmo salar*)	IPNV	2 (24.6%, 18.2%)† 1 (50.9%)	Houston *et al.* (2008) Houston *et al.* (2009)
Atlantic salmon (*Salmo salar*)	IPNV	2 (23.4%, 0.9%)‡	Moen *et al.* (2009)
Atlantic salmon (*Salmo salar*)	ISAV	1 (6%)	Moen *et al.* (2007)
Atlantic salmon (*Salmo salar*)	*Gyrodactylus salaris* parasite	10 (27.3% total)§	Gilbey, Verspoor, Mo, Sterud, Olstad, Hytterod, Jones and Noble (2006)
Eastern oyster (*Crassostrea virginica*)	*Perkinsus marinus* parasite¶ and *Haplosporidium nelsoni* parasite‖	12 (NP)**	Yu and Guo (2006)
European flat oyster (*Ostrea edulis*)	*Bonamia ostreae* parasite	4 (NP)††	Lallias, Gomez-Raya, Haley, Arzul, Heurtebise, Beaumont, Boudry and Lapègue (2009)
Japanese flounder (*Paralichthys olivaceus*)	*Lymphocystis disease* virus	1 (50%)	Fuji *et al.* (2006)

*Number of significant QTL (percentage of phenotypic variance explained, PVE).
†Genome-wide significant QTL, one additional suggestive QTL found with 8.9% PVE.
‡Genome-wide significant QTL, 13 additional suggestive QTL found, each with < 1% PVE.
§ 10 individual marker trait linkages for parasite counts at different days post infection. Largest PVE is given (marker-QTL linkages on day 30).
¶Causing the disease MSX.
‖Causing the disease Dermo.
**No estimate of the amount of phenotypic variance explained by the QTL provided (NP).
††Chromosome-wide significant QTL based on regression interval mapping.
ISAV, infectious salmon anaemia virus; IPNV, infectious pancreatic necrosis virus; QTL, quantitative trait loci.

Genomic selection

Nonetheless, by using marker assisted selection (MAS) it is possible to differentiate the genetic merit of individuals from the same family, even if the trait cannot be measured in the same individual. When MAS was proposed, the main goal was to identify variations with large effects on the trait under interest. Lately Meuwissen et al. (2001) proposed the use of a methodology known as genomic selection (GS) this method proposed that when information of dense genetic markers across the genome is available, and a trait is influenced by many genes across the genome, at least one of the markers might be close enough to one of these genes making possible to estimate its effect, and by adding these small effects estimate the genetic merit of an individual with higher accuracy and making possible to differentiate among individuals from the same family even when records of their phenotypes are not available. With advances in genotyping technology, information of thousands of genetic markers become possible at lower prices in many farming species including Atlantic salmon, because of that making feasible the use of GS in selective breeding programs.

MATERIALS AND METHODS

Fish material

The Atlantic salmon (*Salmo salar*) used in this study, belonged to the 2015-year class of the Rauma-strain (Rauma Eik origin) from SalMar Farming AS, Norway. In breeding nucleus each dam was mated to 3-4 sires, i.e. eggs from one female were divided into four batches and one male was used to fertilize each batch. In cases when females did not have enough eggs for four groups, eggs were divided in three batches for fertilization. Sires with highest estimated breeding values (EBVs) (for growth and IPN) were mated to more females than those with lower EBVs values. Thirty eyed eggs were collected from each family for the PD challenge test (about 6000 eggs from approximately 200 full-sib families). Eggs of each full-sib family were produced and incubated in separated hatching trays at Rauma Stamfisk AS, Reistad. Around one week prior to hatching, eyed eggs were mixed and transported to the hatchery at Rauma Eik AS, Vestrefjord where hatchlings were kept for a few weeks until the yolk sack was consumed. Ready to feed fry (n=5974) were shipped to VESO Vikan. Breeders from three different year classes were used to produce these fry with the aim to form one breeding nucleus population.

Challenge test

Challenge test was performed at VESO Vikan; 5817 fry (157 fry died before the test) were put in a single tank with fresh water where they were fed by automatic feeders over an acclimatization period of approximately three weeks. On 1 April 2015 fry were exposed to a salmon pancreas disease virus (SPDV) subtype SAV2 by a modified cohabitant model as follows: 130 Atlantic salmon parr (“shedders”; average weight ~ 38 g) were infected with SAV2, through intraperitoneal injection, and kept in a tank with no water exchange apart from daily addition of top-up water and no additional oxygenation. Infected parr were allowed to shed virus into the tank for one week and then effluent water from the parr tank was passed into the fry challenge tank of 1.0 m size (250 l in volume) as the sole water source during the first 21-days of the challenge test period. Water temperature in the fry challenge tank was maintained at 12 °C and minimum 70% O_2 saturation in effluent water. Water quality parameters, fish density and other test environment conditions were standardized and equalized as much as possible during whole challenge test period. The challenge test was

carried out in accordance with guidelines from the Norwegian Food Safety Authority (Mattilsynet | Statens tilsyn for planter, fisk, dyr og næringsmidler, 2016).

Challenge test lasted for 64 days during which dead fry were collected twice a day (morning and afternoon). Fry that died within the first 23 days after exposure to the virus were considered to be dead from reasons other than PD infection and therefore omitted from the data by requirement of SalMar Farming AS. Thirty dead fry that were collected between 24th of April and the 11th May were tested to verify the presence of PD virus by polymerase chain reaction (PCR). After PCR confirmation of SAV in the dead fish, all mortalities during the challenge were assigned to PD and tissue samples were collected for genotyping.

Mortality recording at the challenge test was terminated when the mortalities leveled off (at 51 days post-challenge, as a standardized recording duration). All survivors were euthanized at the end of the challenge test (days 62 -64). Almost equal number of dead fry (n=694) and survivors (n=693) were tissue sampled and sent for genotyping.

SNP genotyping

Tissue samples for genotyping were taken from the tail of each fry. Genomic DNA was extracted at IdentiGEN in Ireland, using a magnetic bead based method. In total four plates of samples, containing 384 samples per plate, were genotyped with the NOFSAL02 Affymetrix axiom 57K SNP array (number of markers was 58,184) at AROS Applied Biotechnology A/S in Denmark. Genotypes of 1418 individuals (138 possible parents and 1280 offspring, of which n=657 mortalities and n=623 survivors) passed Affymetrix quality control (QC) and were retained for subsequent analyses.

Quality control and SNP filtering

Genotyped samples were quality checked with PLINKv1.9 using the following procedure: samples and SNPs with call rate <95% were discarded. Furthermore, SNPs with Hardy Weinberg P-value (Fishers exact test) $< 10^{-15}$ and those with minor allele frequency <2% were removed. One sample (corresponding to a dead fry) failed heterozygosity test (which is based on observed versus expected numbers of homozygous genotypes) at a specified significance threshold (five standard deviations from the mean) and was therefore removed from the data set. After the quality checks, the final data consisted 48395 SNPs and 1417 samples.

Parentage assignment and family groups

Around ~1300 SNP markers with high polymorphism (MAF>0.40) were extracted from the parental and offspring populations and CERVUS software was used to perform parentage assignment. However, more than 80% of the offspring could not be assigned to any of their parents. Additional parent-offspring assignment using a highly-informative microsatellite panel specifically designed for assignment testing in Norwegian Atlantic salmon (Baranski et al., 2014), identified correct relationships in less than half of the samples. Instrumental problems were possibly the cause of the low assignment rate using microsatellite data. Due to time constraint and relatively high re-genotyping cost of parents (with either SNP or micro-satellites), no further attempt were made to assign offspring to their parents. Therefore 138 genotypes of possible parents were omitted from further analysis.

However, the offspring were assigned to possible family groups (cluster groups), using estimates of genomic identity by descent (IBD) relationships (G_{IBD}) obtained from PLINKv1.9 software. Briefly, PLINKv1.9 estimates G_{IBD} by detecting the extended chromosomal segmental IBD sharing between pairs of related individuals by use of a identity by state (IBS), in which the underlying hidden IBD state is estimated given the observed identity by state (IBS), see Purcell et al., (2007) for detail description of the method. A clustering analysis was applied on the G_{IBD} using the "k-means" algorithm implemented in R software (R Development Core Team, 2013). As parents were not known, family mortalities were based on "k-means" clustered families. "K-means" are calculated as the average of the cluster groups of many rounds of iterations by randomly chosen SNP groups, used for calculating variation levels in analyzed data set. Individuals were clustered into 150 possible full-sib family groups.

Principal component analysis

Analysis of the genomic data for possible population stratification (structure) was undertaken with PLINK v1.9 (Chang et al., 2015). Population stratification was performed using principal components analysis (PCA) to identify and adjust for ancestry differences among individuals. Eigenvectors obtained from PCA analysis can be viewed as axes of variation that reflect genetic variation due to ancestry effect in the samples. Eigenvector decomposition, based on the singular value decomposition method of the genomic

relationship matrix (G) was used. The G matrix was constructed based on the method of (Yang et al., 2010).

Variance components and heritability estimation

Estimates of variance components for the recorded phenotypes (dead or alive) could not be obtained using pedigree relationships as parents of the tested fry were not genotyped. Due to those limitations, variance components and heritability estimations were based not on pedigree relationships, but on genomic relationships. Genomic relationship matrixes were constructed using two different methods: VanRaden (2008) and Yang et al (2010).

The estimates of genetic, residual and phenotypic variances were obtained using the ASREML v4 software (Gilmour et al., 2009). The following linear mixed animal model (Model 1) was applied to estimate variance components:

$$y = u + \sum_{j=1}^{N=10} pca_j + Zg + e \quad (1)$$

where:

y is a vector of binary phenotypes (dead = 1, alive = 0),

u is the overall mean,

pca_j is the first 10 eigenvectors (N = 10), computed from the genomic relationship matrix,

Z is the incidence matrix of genotyped individuals (linking animal to phenotype),

g is the vector of genomic breeding values and

e is the vector of random residual effects.

It was assumed that $g \sim N(0, G\sigma_g^2)$, and $e \sim N(0, I\sigma_e^2)$, where σ_g^2 and σ_e^2 are the genetic and residual variances respectively, estimated with restricted maximum likelihood (REML) and I is an identity matrix.

The genomic relationship matrix G was constructed using R software following (VanRaden, 2008) as:

$$G = \frac{WW'}{[2\sum_{i=1}^{l} p_i(1-p_i)]}$$

where:

$W = M - P$,

and M is a $N \times l$ matrix of genotypes coded as 0, 1, 2, with N number of genotyped animals by l number of SNP-markers,

P is a matrix with all elements in the i-th column as $2p_i$,

where p_i is the allelic frequency for SNP i. Allele frequencies were estimated from the observed genotypic data.

Heritability on the observed scale (h^2_{obs}) was computed as:

$$h^2_{obs} = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}$$

Heritability on the observed scale was transformed to the underlying (liability) scale (h^2_{lib}) following (Dempster and Lerner, 1950) as:

$$h^2_{lib} = \frac{h^2_{obs} * k(1-k)}{s^2}$$

where k is the proportion of survivors , with $(1-k)$ as the proportion of fish that died. The proportion of survivors k was used to compute the height (s) of the normal standard curve at the threshold corresponding to that proportion k.

Reduced model without pca_j was also fitted to estimate variance components. When there is population stratification (similar to using multi-breed data), not accounting for pca_j effect might lead to inflated variance components.

Genome wide association analysis using linear mixed animal model

Genome wide association study (GWAS) was performed using a linear mixed animal model approach. A linear mixed model was preferred over the simple linear regression model, to account for random polygenic effect , That is important when familial information (covariance between relatives) is used in performing GWAS .Following GWAS model was applied and analysis undertaken with the GCTA software (Yang et al., 2011a):

$$y = u + \sum_{j=1}^{N=10} pca_j + M_i\alpha_i + Zg + e \qquad (2)$$

where:

y is a vector of binary phenotypes (dead = 1, alive = 0),

u is the overall mean,

pca_j is the first 10 eigenvectors (N = 10), computed from the genomic relationship matrix,

Z is the incidence matrix of genotyped individuals (linking animal to phenotype),

g is the vector of genomic breeding values and

e is the vector of random residual effects.

M_i is the incidence matrix for SNP i containing marker genotypes coded as $0 = AA, 1 = AB|BA, 2 = BB$,

α_i is the allele substitution effect of SNP i.

The genomic relationship matrix used here was based on the method of Yang et al. (2010) and was calculated as:

$$G = WDW'/l$$

where:

W and l are as defined before, and

D is a diagonal element, calculated as $\frac{1}{\sqrt{2p_iq_i}}$.

Allele substitution effect was calculated as:

$$\alpha_i = \ (M_i'y) \times (M_i'V^{-1}M_i)^{-1}$$

with variance structure: $V = G\sigma_g^2 \ + \ I\sigma_e^2$.

This approach ensures that when the effect of SNP i. is estimated, it is accounted for the variance due to all markers (σ_g^2) on the other chromosomes.

Test statistics of each SNP effect was calculated as $\chi^2 = {\alpha_i}^2/var(\alpha_i)$ and SNPs were considered to be significant when they exceed the Bonferroni threshold for multiple testing (alpha =0.05) of $0.05/l$, with l = 48395 SNPs. The Bonferroni threshold used in this study was $P < 1.03 \times 10^{-6}$ with an equivalent $-log_{10}(P) = 5.99$. The estimate of the allele substitution effect for SNP i was obtained by excluding the entire number of markers on the chromosome where SNP i is located, and only computing the genomic relationship with

markers from all the other chromosomes. For example, assuming that there are 29 autosomal chromosomes in the genome and the goal is to estimate the allele substitution effect of SNP i, located on chromosome 1. Then genomic relationship matrix would be based on markers located on chromosomes 2 to 29, with chromosome 1 excluded from the calculations.

The estimate of heritability calculated with the G matrix of (Yang et al., 2010) under the GWAS model (Model 2) was compared to that obtained from Model 1. Both models run with and without pca_j.

To identify multiple QTLs in a significant region, the most significant SNP for that region was subsequently included as a fixed effect in the GWAS model and a second GWAS analysis was performed. Additionally, the linkage disequilibrium (calculated as the squared correlations between markers) structure of the significant region was computed using PLINKv1.90.

Quantile-quantile (Q-Q plot) plot and inflation factor (lambda)

Quantile-quantile plots were calculated from the observed p-values (based on our test statistics) and expected theoretical p-values. The magnitude of deviation of observed and expected p-values (i.e. inflation/deflation of p-values - lambda) was calculated as:

$$lambda(\lambda) = \frac{median(\chi^2)}{0.456};$$

If $\lambda < 1.1$ the inflation was considered acceptable (Yang et al., 2011b).

Variance explained by SNP

Variance explained by each significant SNP was calculated from the estimated allele substitution effects (α) based on allele frequencies as:

$$var_{SNP_i} = 2p_i q_i \alpha_i^2$$

The proportion of the total genetic variance ($\%varG_{SNP}$) and the proportion of phenotypic variance ($\%varP_{SNP}$), explained by each SNP, were calculated as:

$$\%varG_{SNP_i} = \frac{var_{SNP_i}}{\sigma_g^2}$$

$$\%varP_{SNP_i} = \frac{var_{SNP_i}}{\sigma_p^2}$$

where:

var_{SNP_i} is the variance explained by SNP i,

p_i is the allele frequency of the major allele for SNP i,

$q_i = 1 - p$, α_i is the allele substitution effect of SNP i,

σ_g^2 and σ_p^2 are the genetic and phenotypic variances on the observed scale from Model 1.

Bioinformatics and Candidate Genes

The map Viewer tool SalmoBase (http://salmonbase.org/) was used for identification of genes. The SalmoBase map viewer uses version 2.0 of the Atlantic salmon genome assembly as the reference map (http://www.ncbi.nlm.nih.gov/genome/369?genome_assembly_id=248466). The genes within the significant regions were identified and subsequently a manual search to know the functions of the genes was performed.

RESULTS

Challenge test outcome, phenotypes

During the entire 64 day of challenge test period (from 1 April 2015 to 3 June 2015) a total of 998 mortalities were recorded and 4819 fry were recorded as alive at the end of the test. The overall level of mortality in the experiment was 17.2 %, calculated from all dead fry in the challenge test, from the day of infection until termination of the challenge test when where mortalities leveled off (less than five dead fry per day for three consecutive days). Figure 11 (A and B) presents mortalities of the fry that were sampled for genotyping (n=693). Mortality curves for the entire challenged tested population (n=5817) are also presented in Figure 11 (C and D, n=998).

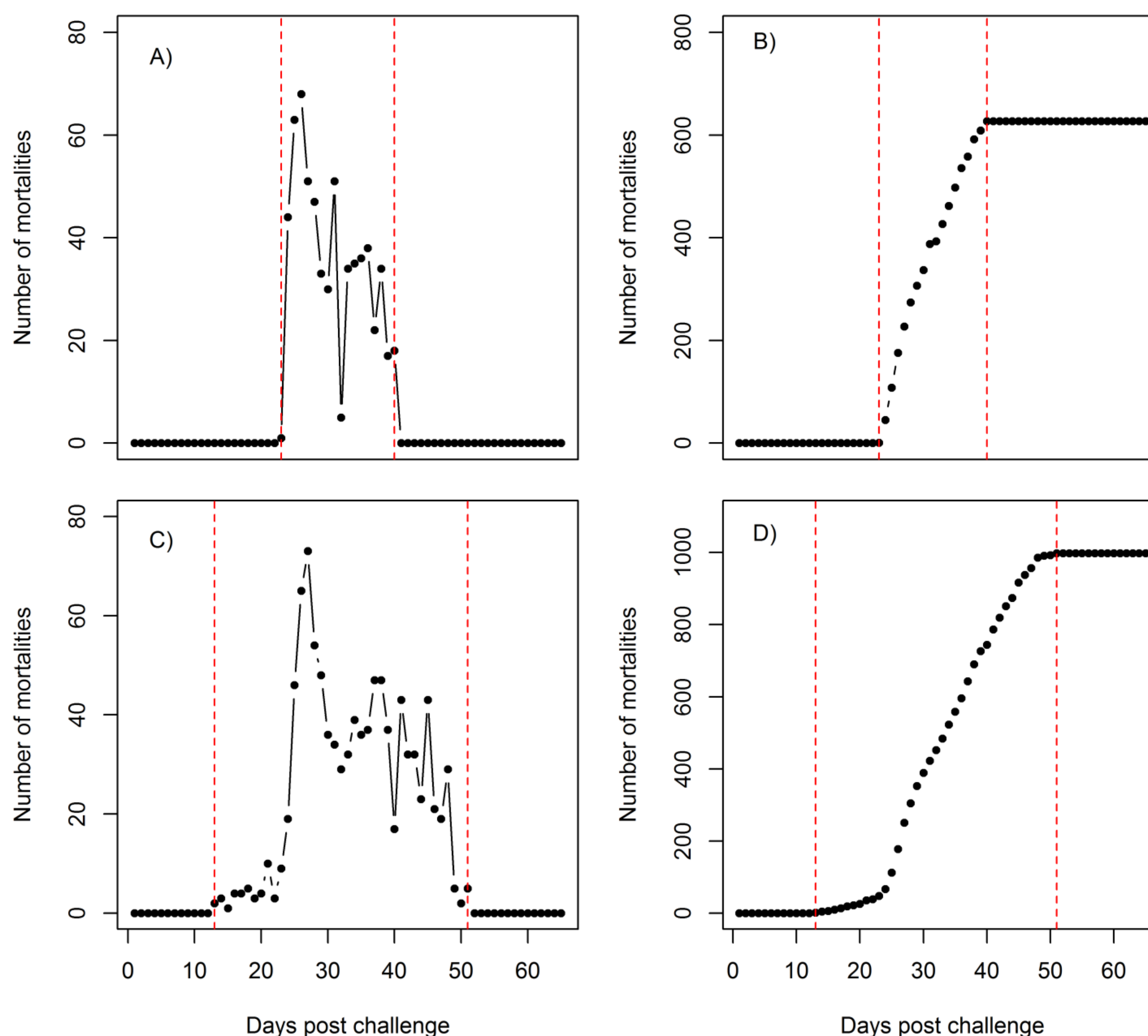


Figure 11. Mortality profiles of challenge test. Daily mortality curve for the genotyped fry (A) and cumulative mortalities (B) during the period 24 April to 11 May, 2015. Daily mortality

(C) and cumulative mortality curve (D) for the entire challenge test period (1 April to 3 June, 2015). Time spans are marked as red dots.

Mortalities in the challenged tank increased at day 13th post-infection, peak in mortalities was reached at day 27th post-challenge (Figure 11, C). At 51 days post-challenge, the recording of mortalities was terminated, as mortalities leveled off. To maximize power for the association study, ~50% (n=694) dead fry and ~50% (n=693) survivors were sampled for genotyping.

As parents were not known, family mortalities were based on “k-means” clustered families. The average percentage of mortalities for the clustered families is shown in Figure 12 where family mortality ranged from 0 to 100%, with an average of 43.3% (SD = 23.4%).

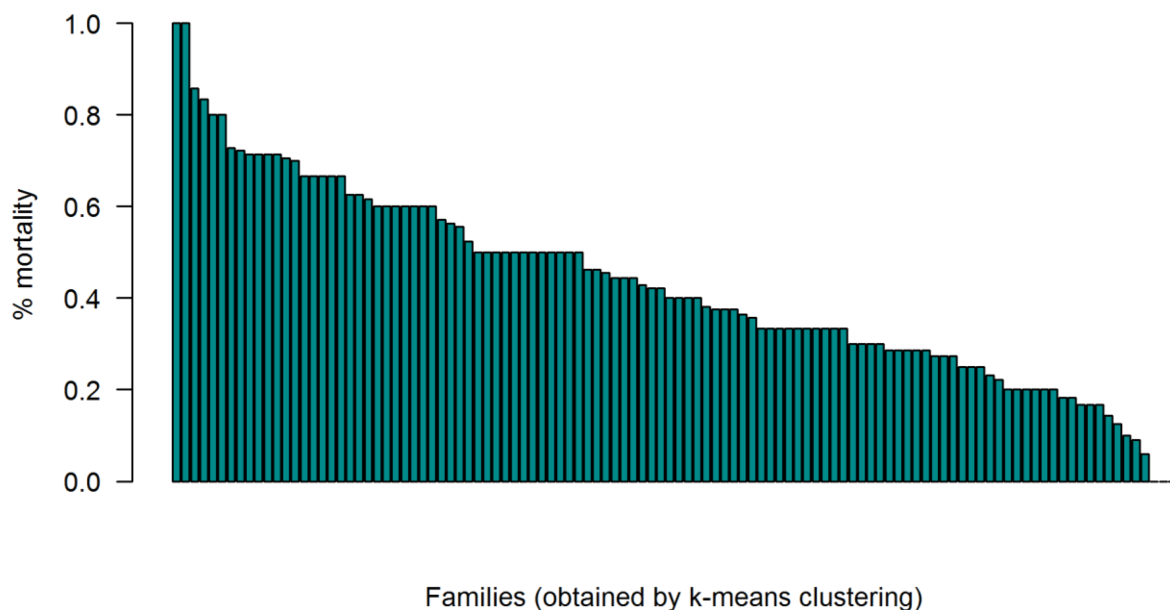


Figure 12. Percentage mortality per family. About 150 full-sib families were obtained by K-means clustering of identity by descent genomic relationship matrix obtained with PLINKv1.9. Only families with ≥ 5 offspring were used (n=110 family groups).

A heatmap of the G_{IBD} is presented in Figure 13, where mortalities within full-sib families (cluster groups) containing more than five individuals are plotted.

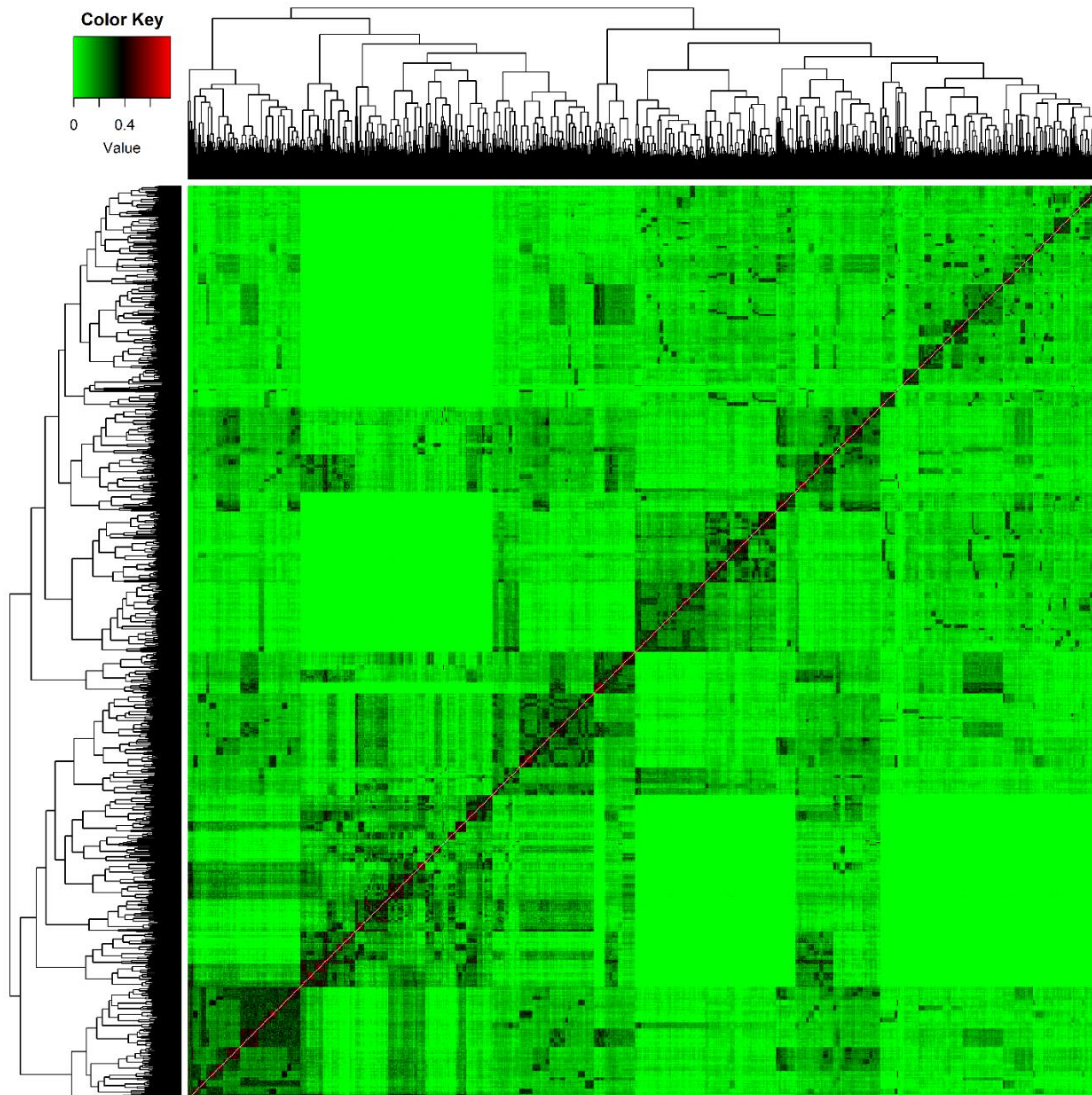


Figure 13. Heatmap of identity by descent genomic relationship matrix (GIBD) obtained with PLINKv1.9.

The number of groups was determined from a plot (see Figure 14) of the within group sum of squares by number of clusters. The average (± standard deviation) G_{IBD} within a cluster was 0.42 ± 0.08 and the number of animals within a cluster ranged from 2 to 34 (mean=8.5; SD=5.0).

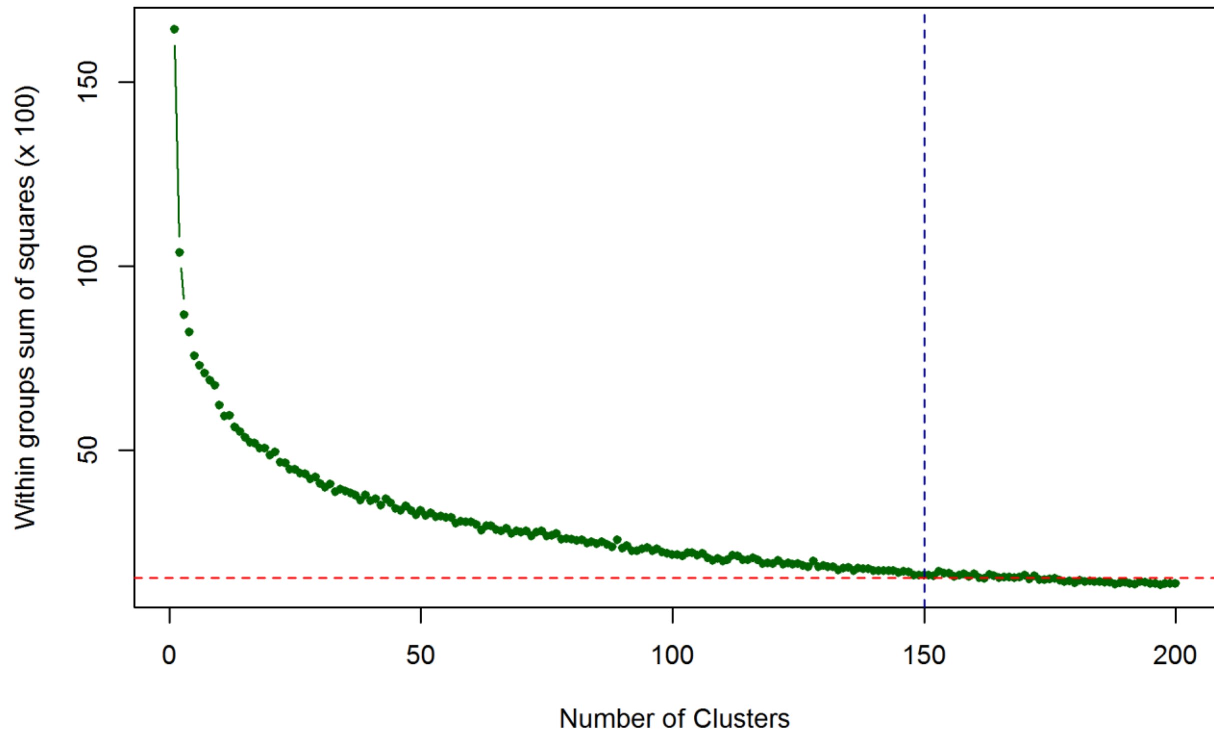


Figure 14. Plot of within groups sum of squares and number of K- clusters. The point at which the blue line meets the red line was chosen as the optimal number of clusters.

Genotyping quality check

Salmon genome contains 29 chromosomes. Each of the SNPs is assigned to one of the chromosomes. However, for some SNPs location was not defined and thus set to be on the chromosome zero (here after referred as chromosome 30).

Around 330 out of the 1387 samples failed Affymetrix quality check because of poor genotyping quality and therefore have been excluded from further analysis. During quality control, genotyping call rates QC scores under value of 0.90 were discarded. In total, 57 184 SNPs were used for genotyping and out of these, 5445 SNPs were withdrawn from further analysis (on the base of missing genotype rate, in our case SNPs that were non-informative in more than 5% of samples), 874 SNPs failed the Hardy Weinberg test, 2470 markers excluded as they were below MAF (minor allele frequency) threshold (2%). Minor allele frequency refers to the frequency at which the least common allele occurs in a given population. One animal failed heterozygosity test (which is based on observed versus expected numbers of homozygous genotypes) at a specified significance threshold and therefore was removed from the data set. In total 8789 SNPs and one individual (dead fry) were removed from the dataset, thus after filtering of genotyping data, 48 395 SNP markers and 1279 genotypes were left for the QTL (GWAS) analysis.

Principal component analysis

Principal component analysis on the genomic data revealed genetic stratification among the individuals used for the study. Figure 15 demonstrates the level of population stratification in the data. The first two eigenvectors captured 6.87% (PCA1) and 2.88% (PCA2) of the variation in the genomic data. To reduce the potential confounding effect of population stratification on the GWAS analysis, 10 eigenvectors (PCA1-10) that captured about 23% of the variation in the genomic data and was added as a covariate to the GWAS model.

Three distinct clusters were identified for the three different year classes from which the parents of the offspring originated. According to SalMar Farming AS (owner of the breeding nucleus), in the breeding program all three year classes were combined into one single breeding nucleus, thus the entire dataset was analyzed as a single population.

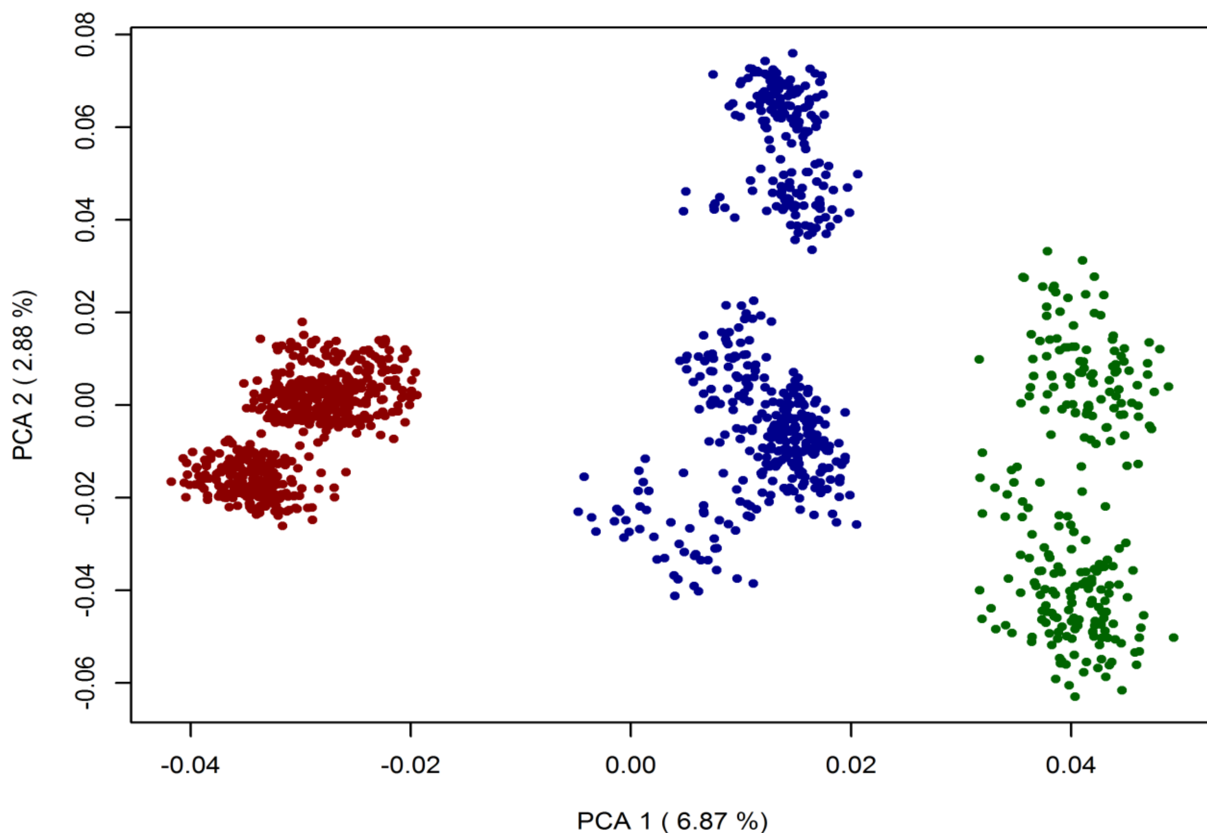


Figure 15. Principal component analysis (PCA) for the three distinct populations. The percentage of variance explained by PC 1 and 2 is in brackets and the colors represent each distinct year class of the parental population.

Estimated variance components and heritabilities

Heritabilities, calculated using two methods, ranged from 0.185 to 0.227 (Table 2) on the observed binary scale and both methods gave similar heritability estimate (h^2 = 0.196±0.04 with VanRaden (2008) and 0.225±0.05 with Yang et al. (2010)). When heritabilities were transformed to the underlying liability scale with a prevalence level of 17.2% (the percentage mortality in the entire challenge test), h^2 values ranged from 0.34 to 0.38. Accounting for the observed population structure by using the 10 largest eigenvectors resulted in a reduction of heritability by about 9.5% with respect to heritability when no PCA where included, although not significant. Since full-sib families were not known, common environmental effect was not accounted for in the model. However, when the "k-means" clustered families were fitted in the model but not significant, and thus omitted from further analysis.

Table 2. Variance component and heritability estimate for pancreatic disease (PD) using genomic relationship matrices computed based on VanRaden (2008) and Yang et al., (2010).

Method	**Model1**	σ_g^2 **± SE**	σ_e^2 **± SE**	σ_p^2 **± SE**	h_{obs}^2 **± SE**
VanRaden	With PCA	0.043 ± 0.011	0.191 ± 0.010	0.235 ± 0.010	0.185 ± 0.043
(2008)	Without PCA	0.051 ± 0.012	0.194 ± 0.010	0.244 ± 0.011	0.207 ± 0.045
Yang et al.,	With PCA	0.056 ± 0.014	0.191 ± 0.011	0.247 ± 0.012	0.227 ± 0.051
(2010)	Without PCA	0.055 ± 0.013	0.191 ± 0.010	0.246 ± 0.011	0.223 ± 0.046

[1]The statistical model used to estimated variance component included a covariate term of either 10 principal component (with PCA) or not (without PCA) to account for population structure displayed in Figure 15.

Genome-wide association results

GWAS results for survival to PD SAV2 virus are presented in Manhattan plots (Figure 16). In total, 26 SNPs on 14 chromosomes showed significant association (Bonferroni $-log_{10}$ (P) > 5.9) with the trait. On twenty-two SNPs out of 26 SNPs, no neighbor markers associated with the trait were found. These 22 markers are hereafter referred to as "lonely significant markers" (Figure 16, A).

Due complexity of salmon genome, SNPs in the array used for this study were classified in 3 types: type 1 marks just one certain point, type 2 uncertainties in the exact position, meaning approximate coordinates of the SNP were reported but the exact position is

uncertain, and; type 3 are markers which alternative allele usually mapped to a different position in the genome. All twenty-two “lonely significant SNP markers” were not taken into consideration as they were classified as type 3 SNPs, which lead to the possibility of being false positive.

One significant QTL (multiple markers in the same region showing significant associations with the trait) was identified on chromosome 21 (Figure 16, B). Thus the main focus of the present study was on chromosome 21, where four significant markers exceeded the threshold and multiple neighbor markers in the region showed weaker association with the trait.

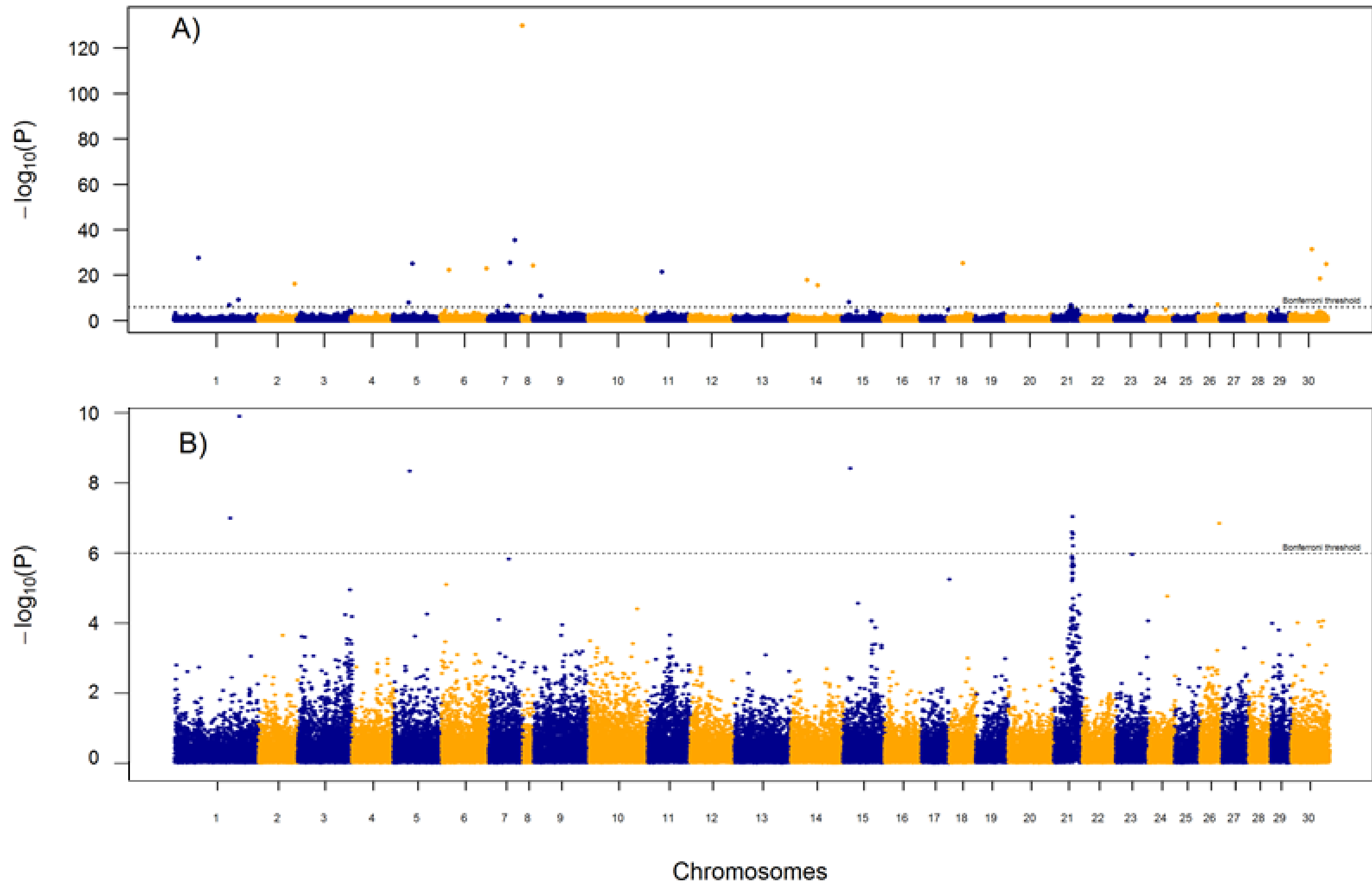


Figure 16. Manhattan plot of genome-wide –log10 p-values for resistance to PD –SAV2. Plot with all SNPs (A) and with SNPs with genome-wide –log10 (P) <10 (B). The horizontal dashed line represents the Bonferroni significance threshold (-log10 (P) = 5.99).

Quantitative trait loci (QTL)

The QTL region on chromosome 21 spans from 34,474,927 to 35,975,753 bp (Figure 17). In this region, 17 SNPs were A) in high LD and B) weakly associated ($-log_{10}$ (P)>5) with the four top significant markers. When the significant SNP (AX-88148479; located on chromosome 21, was included as a fixed effect in the GWAS model, none of the three surrounding SNPs showed associations with the trait.

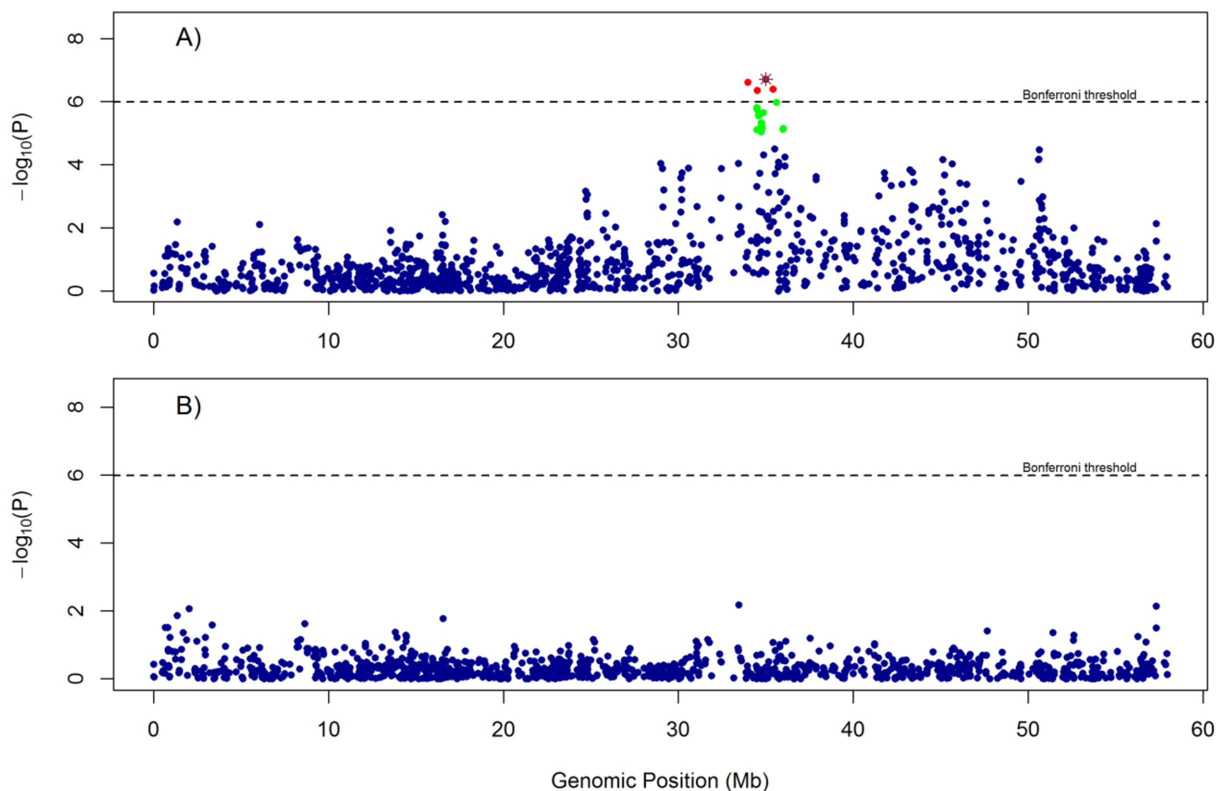


Figure 17. Plot of genome-wide –log10 p-values of SNPs on chromosome 21. The horizontal dashed line represents the Bonferroni significance threshold (-log10 (P) = 5.99). (A) Plot with all markers on chromosome 21 where the most significant markers are highlighted in red and the marker [AX-88148479] with the highest significance is asterisked (*). (**B**) Plot after correcting for the highest significant SNP [AX-88148479].

Quantile-quantile (Q-Q) plot

The quantile-quantile (Q-Q) plot from the fitted GWAS model is presented in Figure 18. The inflation factor (lambda) of the fitted GWAS model with all markers was 0.99, which indicates relatively good concordance between observed and assumed distributions of the chi-square test statistics. Markers were declared significant if they exceeded the Bonferroni significance threshold of p = 1.03 × 10-6 equivalent to -log10 (P) = 5.99. When the "lonely significant markers" were excluded from the plot, the inflation factor dropped to 0.98 (Figure 18).

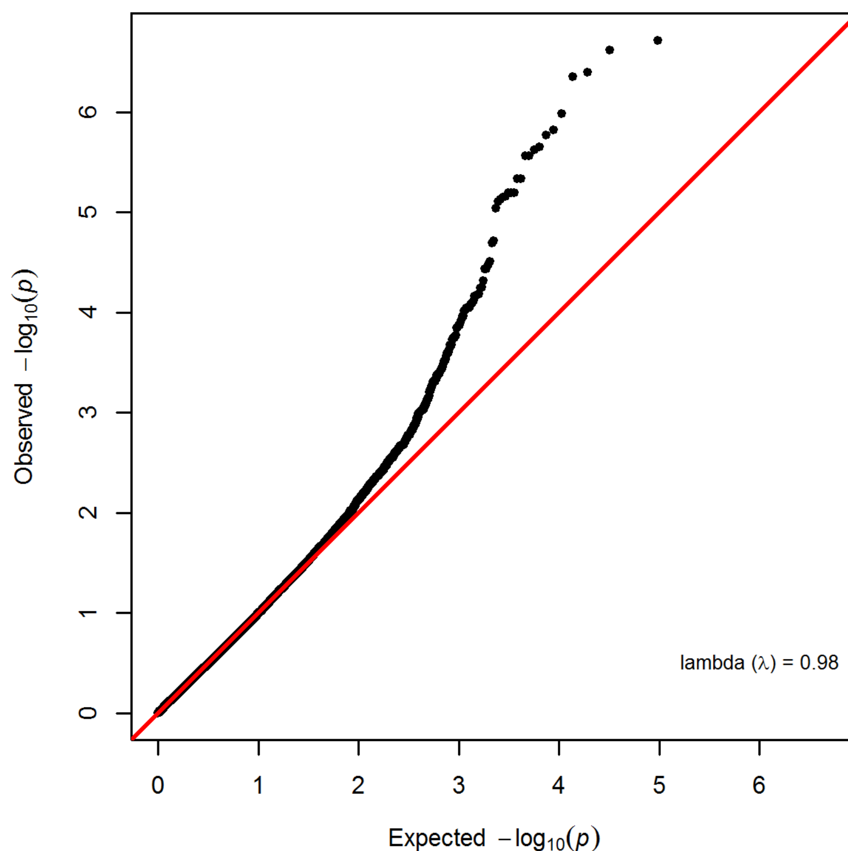


Figure 18. Quantile-quantile (Q-Q) plot for the observed and expected genome wide –log10 p-values.

Summary statistics of significant SNPs on chromosome 21

The frequency of the minor allele, allele substitution effects and proportion of the genetic and phenotypic variances, explained by each of the four significant markers on chromosome 21, are presented in Table 2 and Table 3. The direction of the allele substitution effect of the minor allele was positive for one (AX-87261594) and negative for the remaining

three markers. Minor allele frequency of the four significant markers in the entire population ranged from 0.26-0.50 (Table 3).

On average markers explained 10.9% of the genetic variance and 2.5% of phenotypic variance (Table 3). Because none of the markers showed association after fixing the most significant marker in the model, total variance captured by the QTL region was 10.9% (instead of the sum of the variances).

Frequency of the four significant SNPs in survivors ranged from 0.231 to 0.538 and in the dead fry from 0.303 to 0.473 (Table 4). Hardy Weinberg disequilibrium test showed a moderately significant deviation (Fishers exact p-value >8 × 〖10〗 ^(-4)).

Table 3. Marker name and position, allele frequency (p), allele substation effect (α) and the percentage of variance captured by four significant SNPs on chromosome 21.

Chr.	Variant name (SNP ID)	Position (BP)	Minor allele	Freq $(p)^1$	Beta (α)	-log$_{10}$P	$\%\sigma^2_{G_{SNP}}$ [2]	$\%\sigma^2_{p_{SNP}}$ [3]
21	AX-88148479	35,003,259	A	0.30	-0.126	7.0	11.93	2.72
21	AX-87754063	33,975,046	B	0.28	-0.122	6.6	10.81	2.46
21	AX-87261594	35,414,933	A	0.50	0.109	6.5	10.61	2.44
21	AX-88221678	34,502,311	B	0.26	-0.123	6.4	10.40	2.37

[1]Freq: the frequency of the minor allele.
[2]% $\sigma^2_{G_{SNP}}$: The percentage of the total genetic variance (σ^2_g =0.056) captured by each SNP [$=2pq\alpha^2/\sigma^2_g$].
[3]% $\sigma^2_{p_{SNP}}$: The percentage of the total phenotypic variance (σ^2_p =0.247) captured by each SNP [$=2pq\alpha^2/\sigma^2_p$].

Table 4. Summary information on genotype counts and allele frequency for the dead (MAF-D) and survivors (MAF-S). Allele count of the minor allele (left), heterozygotes (middle) and major allele (right) for the four significant SNPs on chromosome 21.

Chr.	Variant name (SNP ID)	Position (BP)	[1]MAF_D	[2]MAF_S	[3]GENO_D	[4]GENO_S
21	AX-88148479	35,003,259	0,267	0,354	38/335/398	69/214/214
21	AX-87754063	33,975,046	0,255	0,331	41/306/414	55/210/219
21	AX-87261594	35,414,933	0,538	0,436	211/409/152	90/257/154
21	AX-88221678	34,502,311	0,231	0,303	33/274/428	51/197/245

[1]MAF_D: the frequency of the minor allele in affected (dead) fish.
[2]MAF_S: the frequency of the minor allele in unaffected (survivors) fish.
[3]GENO_D: the genotype counts in affected fish, with the allele count of the minor allele (left), heterozygotes (middle) and major allele (right).
[4]GENO_S: the genotype counts in unaffected fish, with the allele count of the minor allele (left), heterozygotes (middle) and major allele (right).

Genome wide $-\log_{10}$ p-values and linkage disequilibrium (LD) R^2 values on the 30-40 Mb region on chromosome 21 are plotted in Figure 19.

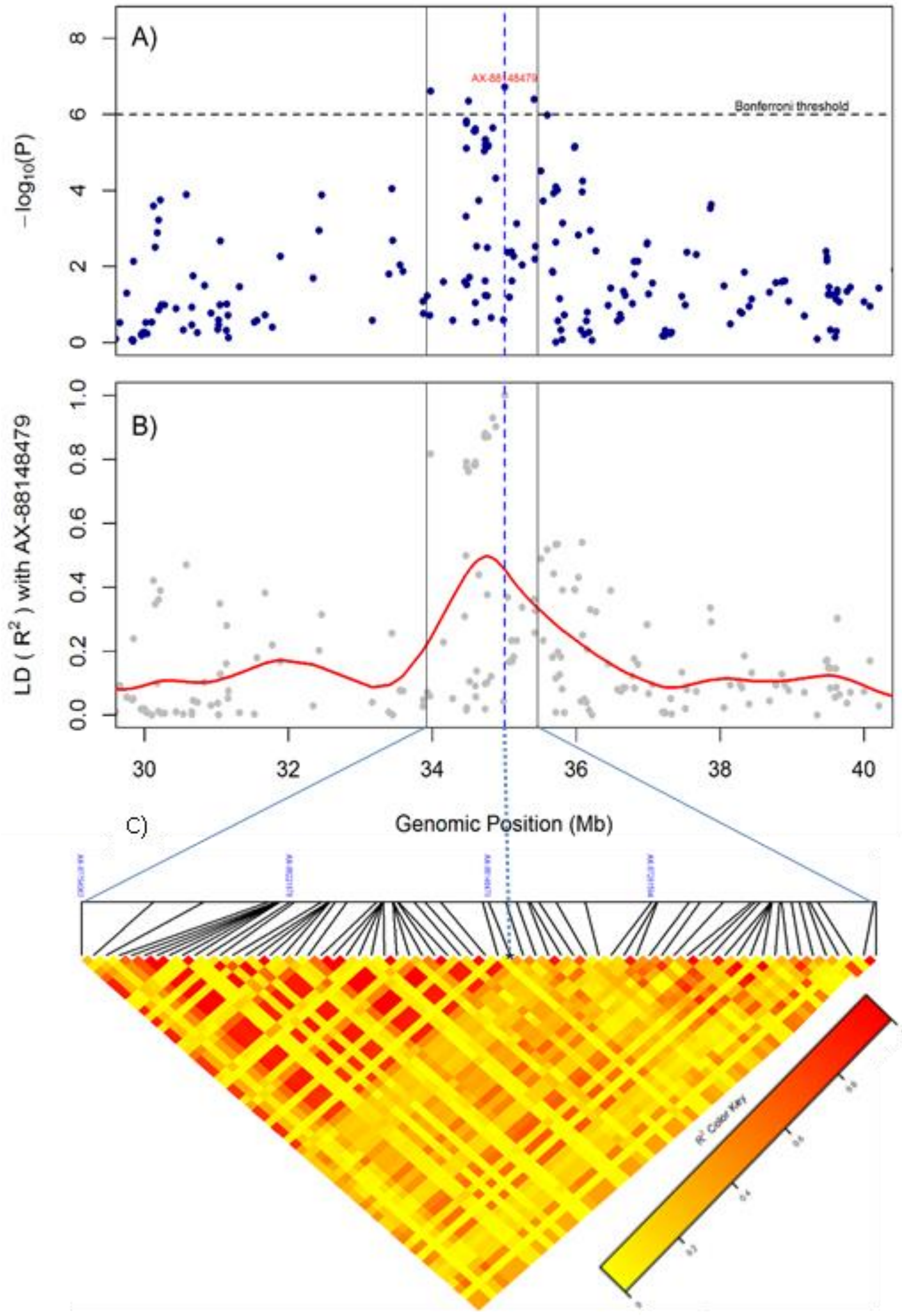


Figure 19. Plot of genome wide $-\log_{10}$ p-values and linkage disequilibrium (LD) R^2 values on the 30-40 Mb region on chromosome 21. (A) Plot of genome wide $-\log_{10}$ p-values of the region, Bonferroni significance threshold (p-value = 1.03×10^{-6}). (B) Degree of LD of all SNPs in the region. Red line shows a smoothed spline of LD values. (C) Heatmap of LD structure between SNPs in the region from 33.90 to 35.99 Mb. Positions of the four significant markers are highlighted in blue text.

Bioinformatics and candidate Genes

Genes from the candidate region on chromosome 21 spanning 34,474,927-35,975,753 bp, where identified using SalmoBase database (The International Cooperation to Sequence the Atlantic Salmon Genome) (Figure 20).

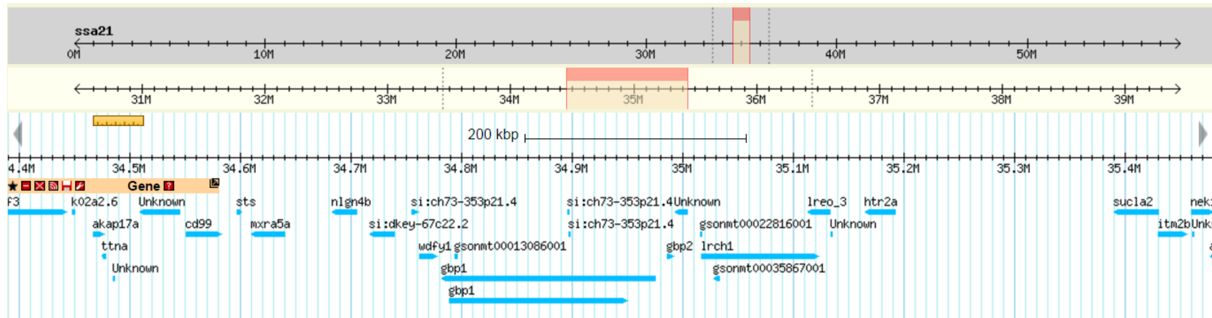


Figure 20. Genes from region 34.5Mb -35.9Mb on chromosome 21.

Two assemblies of the salmon genome were used to identify candidate genes in the QTL region. In total 36 genes within the region were found and grouped into 8 clusters (DAVID data base) by their function. One of the clusters was associated with immune response, one with cancer and several with metabolism processes in the cells (mainly Ca, Zinc ions and elements involved in osmotic balance in fish, as well as energy maintenance in the cell mitochondrias). Eight SNPs from chromosome 21 were targeting genes directly (Table 5). Gene functions were taken from human genome, assuming that they are involved the same or similar functions in Atlantic salmon ortologues.

Table 5. Candidate genes in the region 34.5-35.9 Mb on chromosome 21. Only markers in the protein coding genes are presented.

Chr.	Variant name (SNP ID)	Gene code	Gene function
21	AX-96189480	AKAP17A	Involved in the regulation of alternate splicing in some mRNA precursors.
21	AX-87450144	WDFY1	WD40 domain, found in a number of eukaryotic proteins that cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly. Metal ion binding.
21	AX-96376164	LONRF2	Involved in zinc ion binding (disease induced,-cancer).
21	AX-87267241	PDE1A	Possibly involved into Ca+2 release.
21	AX-87781688	SRRM1	Involved in immune and inflammation responses.
21	AX-86972729	SRRM2	Involved in immune adaptive responses against microbial pathogens, cancer and toxins.
21	AX-88141638	GBP1	Exhibits antiviral activity against influenza virus.
21	AX-87261594	SUCLA2	Succinate-CoA ligase plays a critical role in mitochondria, which converts the energy from food into a form that cells can use.

DISCUSSION

Survival curves for PD SAV2 obtained in our study in fry (infected by cohabitant method) (Figure 11), were found to be very similar to those reported by Gonen et al. (2015) on fry infected with SAV3 using the same method of infection. In this study as in Gonen's et al (2015), peak mortalities were observed within days 26th and 28th post infection and leveling off after fifty one day post-infection. However, results on fry from both studies, differed from those reported by Gonen et al. (2015) in smolt, with mortalities occurring already seven days post infection, reaching a peak at day thirteen and levelling off at day sixteen post-infection. Differences between smolt and fry survival curves may be due to the infection method used in Gonen's et al. (2015) study on smolt, as smolt were infected by intraperitoneal injection while in both studies (ours and Gonen's (2015)), fry were infected by cohabitant method.

Within family mortalities in our study ranged from 0 to 100%, which are in accordance to results obtained by Gonen et al (2015). However, average family mortality for SAV2 in our study when compared to SAV 3 in Gonen's (2015) were smaller, with values of 43.3% in our study compared with mortalities of 61% reported by Gonen et al (2015). Differences observed may be attributed due to different subtypes and virulence levels of the used pathogen, as findings by Jansen, Jensen, & Brun (2015) and Taksdal et al. (2014) suggests that PD outbreaks caused by SAV2 results in lower mortalities and milder clinical signs when compared to SAV3 outbreaks. For example under field conditions Jansen et al. (2015) reported lower levels of mortality due to SAV2 (up to 25 %) when compared to SAV3 mortality (up to 72 %). However, mortality during outbreaks is influenced by environmental conditions and other diseases prior to PD. As given high genetic correlations between disease resistance in field and resistance in challenge tests (r=0.95), as have been reported by (Gjøen et al., 1997) for three bacterial diseases, mortalities under challenge test conditions could be more trustworthy as environment can be better controlled.

Each viral (SAV) subtype has few genetically slight different variants, so called isolates. Mortalities in experimental conditions across SAV2 isolates have been reported to be similar (Taksdal et al., 2014), most likely due to a short period of presence of the virus in Norway and limited geographical distribution in the country For the contrary, SAV3 isolates have shown significant differences in mortalities across isolates, more likely due a longer presence of the virus in Norway and accumulation of mutations (Taksdal et al., 2015).

Heritabilities for PD survival in this study were found to be moderate, around 0.2±0.04 (see Table 1) on the observed scale, which are comparable to those obtained by Norris et al. (2008) and Gonen et al. (2015); h^2=0.21±0.005 and h^2=0.26± 0.07 respectively. As heritability is a population dependent measure, moderate heritability in our case shows good possibilities to select for the trait. Heritability for PD is also similar to those reported on other salmon diseases (Gjedrem and Gjøen, 1995; Ødegård et al., 2011).

As salmon genome contains a large number of duplications (Lien et al., 2016) and SNPs can bind to different regions across the genome. In our study the most significant SNP (AX-88162080), was classified as SNP type 3 as it was binding to two positions in chromosome 8 (at 14.1Mbp and 14.6Mbp). In addition, by analyzing raw Affymetrix data, this particular SNP (AX-88162080) was unclear in its' appearance, therefore unreliable and omitted from further analysis as it was not possible to define if polymorphism corresponded to differences in state of the allele, or due different positions in the genome or due genotyping artifacts. On the same way, the other twenty-one "lonely significant SNP markers" were not taken into consideration due to the possibility of being false positive for similar reasons to those that led to exclusion of the marker AX-88162080. Additional studies are needed to clarify the quality of SNPs and to remove those markers showing uncertainties when allele call is done. While most of the "lonely significant markers" found by running GWAS in our study were SNPs of type 2 or 3, these markers could be on any of the chromosomes, meaning that actually all of them could be on chromosome 21 or segregating together on any other chromosome. It is hard to make clear conclusions based on single SNPs' analyzing polymorphic traits, thus further studies are required.

In addition to unclear SNP positions on the genome, plates with genotyping material for Affymetrix were non-randomized (meaning that unequal number of mortalities and survivors were placed in each of the plates), which made possible to produce artifacts during the genotyping and therefore could result in false polymorphism or wrong allele call. Moreover, allele call was done plate by plate, which means that genotyped samples were analyzed within each plate and not across all analyzed plates, as recommended by Affymetrix in their own manuals for good practices. That might cause mistakes in reading raw Affymetrix data.

Only significant markers on chromosome 21 showed a shape compatible with a true QTL, where several SNPs in the same region were creating a clear peak in the Manhattan plot. Most SNPs in the region of interest were closely linked, opening the possibility to use

fewer markers in the region to identify the different haplotypes. In addition, if the markers are closely linked, they express the phenotypic and genetic variation of only one of those markers.

In our study with SAV2, we did not identify a QTL common to that detected by Gonen et al. (2015) for SAV3. This suggests that either SNP effects are population specific, or that SAV3 induced a different response than SAV2 due to the differences in viral genotypes of SAV2 and SAV3 at section E2 (Fringuelli et al., 2008). The Gonen et al. (2015) study identified a QTL on chromosome three in both fry and smolt life stages, therefore this suggests that the lack of a QTL detected on this chromosome for SAV2 is more likely due to differences in the underlying genetics of resistance rather than a life stage specific effect.

Gonen et al. (2015) found a QTL on chromosome 7 for SAV3 in fry population and QTLs on chromosomes 2 and 14 for SAV3 in smolt. In our study there were “lonely significant markers” on chromosomes 2, 7 and 14, however those were not checked for correspondence to the SNP found by Gonen et al. (2015) on the same chromosomes. More research is needed to confirm if there is any certainty in the match of the results from these above mentioned studies.

Allele frequencies and QTL were calculated for survival as we cannot recognize between individuals which were sick but that recovered during the test period from those who didn’t get infected at all.

All fish heterozygous for SNPs (AX-96351825 (chrom. 5), AX-86994027 (chrom. 15), AX-88095996 (chrom. 30) and AX-88044860 (chrom. 14)) were recorded as dead in the test. On the other hand, all fish heterozygous to SNP AX-87896560 on chromosome 7 survived the test. This could lead to assumption that some alleles are deleterious and some gives complete resistance to disease, or that individuals were affected but survived due to specific (better or worse) immunity response or other unclear factors. However, if the effect for resistance would have been significant, any of the SNPs containing those alleles should explain a high proportion of the genetic variance, which was not in our case. In our study the mentioned above SNPs explained from 12% to 54% of genetic variance and from 2.6% to 12.4% of phenotypic variance.

The second most significant SNP from all analyzed markers was a “lonely significant marker” on chromosome 7 (AX-87896560). In addition, this SNP explained a large proportion of the genetic variance (54%). A candidate gene, targeted by this marker is involved in immune response by inducing mitochondrial stress, while stress is a natural body response to the pathogen. The function of this candidate gene is similar to the functions of

genes found on chromosome 21. This supports the mentioned above hypothesis that the "lonely significant marker" could be located not on chromosome 7 but on chromosome 21. For the contrary, Gonen et al. (2015) reported a QTL on chromosome 7, which if located on the same region as the "lonely significant marker" found in our study, may provide bases to rethink our decision to remove this SNP as a real QTL, together with "lonely significant markers" found in chromosomes 2 and 14.

Thirty-six genes were found in the QTL region of chromosome 21 on Salmon genome assembly. More than one third of the genes found in this region are involved in immune response, including tumor suppressor genes. Several regulate metabolism of Ca and Zinc ions. Some of the candidate genes are connected with brain damage and muscle inervation disfunction (athony), which might cause reduced ability to move (e.g. HTR2a). This may indicate that PD might have a connection with neurological disease which ends up in movement disorder, loss of appetite due to depression and other neurological based dysfunctions. Few genes are specifically involved in smooth contraction of muscles, including heart muscle. However, the loss of appetite and reduced swimming may as well arise due to physiological changes in pancreas, heart and other organs, and physical pain of the fish. This analysis of genes in the region of interest was made from ortologues of human genome, assuming similar functions of the same genes in both salmon and humans.

However, some genes had similarities with the ones found by other studies proceeded in Atlantic salmon. For example, gene TNIK is one of the inflammatory genes, - the TNF signaling gene. TNF genes were also induced in the heart during pathology in fish, infected with SAV3 (Xu et al., 2012). Gene CD99 is a T-cell related gene, thus involved in adaptive immune system cycle. Several of these T cell-related genes (CD3ε, CD4, CD8, TCR-α and MHC-II) were expressed in pancreas and heart at 8th week post infection in Xu et al. (2012) study with SAV3. To be certain of full functions and interactions between the genes, a separate study is required.

Inclusion of marker assisted selection (MAS) in aquaculture breeding programs is usually done with the aim of exploit the within family variance, as most of these programs are family based and only between family variance can be used. But because of the high cost of including this technology in any breeding program, traits should be carefully selected. In our study, because mortalities caused by pancreas disease SAV2 variant are low, survival analysis may be less relevant as the main cause of economic losses for the industry are caused by reduced salmon growth and high culling rate of non-recovered fish.

Thus breeding should aim to select for a salmon, less susceptible to develop PD. In our study, it is not clear if fry survived because they did not develop the disease, or because they were able to recover from the infection. For this reason, survivors also should be checked for PD by, for example, real time PCR before they are euthanized. Also, survived fry should not be killed right after the challenge, but kept longer to ensure mortalities don't continue to occur. Also efforts should be done to identify infected but not affected fish shedding the virus.

In order to obtain clear and valuable results, survived fry should be kept to draw the growth curves for the same period as in commercial production. In such studies, at least two groups of the fry with different phenotypes should be made: one who did not get infected or got infected but not affected and another where fish got infected but recovered. Drawing such growth curves for each of the groups, should help to identify a better selection trait to include in the breeding programs. Moreover, studies to determine if infected but not affected fish is shedding the virus may also provide information to target more economical important traits for selection.

CONCLUSIONS

In this study fry from a single population but a mix of three year classes was challenged with pancreas disease virus subtype SAV2. Because a moderate heritability for resistance against PD was estimated (h^2=0.2) it shows the feasibility of family selection for PD resistance. A single quantitative trait loci (QTL) was identified on chromosome 21 and explained only 2.5% of phenotypic variation and 10.9% of genetic variation for resistance in the populations. Therefore QTL based MAS is more likely not to be the best approach, but because of the large number of “lonely significant markers”, genomic selection alone or in combination with classical selection may be a better tool for selection in resistance to PD. Also identification of underlying polymorphism of the trait needs to be performed in order to obtain best results in further selection if only QTL MAS is implemented.

CITED LITERATURE

2007 McLoughlin Graham SAV review jfd_848.

Aslam, M.L. 2012. Genetic control and variation in turkey : molecular insights in selection. 184 pp.

Aunsmo, A., P.S. Valle, M. Sandberg, P.J. Midtlyng, and T. Bruheim. 2010. Stochastic modelling of direct costs of pancreas disease (PD) in Norwegian farmed Atlantic salmon (Salmo salar L.). *Prev. Vet. Med.* 93:233–241. doi:10.1016/j.prevetmed.2009.10.001.

Baranski, M., C. Jacq, S. Karlsson, B. Hayes, and R. Blonk. 2014. SporLaks – Industry-wide tracing of Norwegian farmed Atlantic salmon. Final report.

Bergman, S.M., J. Castric, M. Bremont, R. Riebe, and D. Fichtner. 2005. Bergman S. *In* Detection of sleeping disease virus (SDV) in Germany. 154.

Bratland, A., and A. Nylund. 2009. Studies on the possibility of vertical transmission of Norwegian Salmonid Alphavirus in production of Atlantic salmon in Norway. *J. Aquat. Anim. Health*. 21:173–178. doi:10.1577/H08-038.1.

Castric, J., F. Baudin Laurencin, M. Brémont, J. Jeffroy, A. le Ven, and M. Bearzotti. 1997. Isolation of the virus responsible for sleeping disease in experimentally infected rainbow trout (Oncorhynchus mykiss). *Bull. Eur. Assoc. Fish Pathol.* 17:27–30.

Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 4:7. doi:10.1186/s13742-015-0047-8.

Christie, K.E., D.A. Graham, M.F. McLoughlin, S. Villoing, D. Todd, and D. Knappskog. 2007. Experimental infection of Atlantic salmon Salmo salar pre-smolts by i.p. injection with new Irish and Norwegian salmonid alphavirus (SAV) isolates: A comparative study. *Dis. Aquat. Organ.* 75:13–22. doi:10.3354/dao075013.

Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor, B.A. Crooker, C.P. Van Tassell, J. Yang, S. Wang, L.K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics*. 12:408. doi:10.1186/1471-2164-12-408.

Dempster, E.R., and I.M. Lerner. 1950. Heritability of Threshold Characters. *Genetics*. 35:212–236.

Desvignes, L., C. Quentel, F. Lamour, and A. LE Ven. 2002. Pathogenesis and immune response in Atlantic salmon (Salmo salar L.) parr experimentally infected with salmon pancreas disease virus (SPDV). *Fish Shellfish Immunol.* 12:77–95. doi:10.1006/fsim.2001.0356.

Drangsholt, T.M.K., B. Gjerde, J. Ødegård, F. Fridell, and H.B. Bentsen. 2011. Quantitative genetics of vaccine-induced side effects in farmed Atlantic salmon (Salmo salar). *Aquaculture*. 318:316–324. doi:10.1016/j.aquaculture.2011.05.044.

Fringuelli, E., H.M. Rowley, J.C. Wilson, R. Hunter, H. Rodger, and D.A. Graham. 2008. Phylogenetic analyses and molecular epidemiology of European salmonid alphaviruses (SAV) based on partial E2 and nsP3 gene nucleotide sequences. *J. Fish Dis.* 31:811–823. doi:10.1111/j.1365-2761.2008.00944.x.

Gilmour, A., B. Gogel, C. B, and R. Thompson. 2009. ASReml user guide release 3.0. *In* VSN International Ltd, Hemel Hempstead, UK.

Gjedrem, T. 2004. Selection and Breeding Programs in Aquaculture. Springer-Verlag, Berlin/Heidelberg. 364 pp.

Gjedrem, T., and M. Baranski. 2009. Selective breeding in aquaculture : an introduction. Springer. 221 pp.

Gjedrem, T., and H.M. Gjøen. 1995. Genetic variation in susceptibility of Atlantic salmon, Salmo salar L., to furunculosis, BKD and cold water vibriosis. *Aquac. Res.* 26:129–134. doi:10.1111/j.1365-2109.1995.tb00892.x.

Gjerde, B., J. Ødegård, and I. Thorland. 2011. Estimates of genetic variation in the susceptibility of Atlantic salmon (Salmo salar) to the salmon louse Lepeophtheirus salmonis. *Aquaculture*. 314:66–72. doi:10.1016/j.aquaculture.2011.01.026.

Gjøen, H.M., T. Refstie, O. Ulla, and B. Gjerde. 1997. Genetic correlations between survival of Atlantic salmon in challenge and field tests. *Aquaculture*. 158:277–288. doi:10.1016/S0044-8486(97)00203-2.

Gonen, S., M. Baranski, I. Thorland, A. Norris, H. Grove, P. Arnesen, H. Bakke, S. Lien, S.C. Bishop, and R.D. Houston. 2015. Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (Salmo salar). *Heredity (Edinb).* 115:405–14. doi:10.1038/hdy.2015.37.

Graham, D.A., E. Fringuelli, H.M. Rowley, D. Cockerill, D.I. Cox, T. Turnbull, H. Rodger, D. Morris, and M.F. Mc Loughlin. 2012. Geographical distribution of salmonid alphavirus subtypes in marine farmed Atlantic salmon, salmo salar L., in Scotland and Ireland. *J. Fish Dis.* 35:755–765. doi:10.1111/j.1365-2761.2012.01401.x.

Graham, D.A., P. Frost, K. Mclaughlin, H.M. Rowley, I. Gabestad, A. Gordon, and M.F. Mcloughlin. 2011. A comparative study of marine salmonid alphavirus subtypes 1-6 using an experimental cohabitation challenge model. *J. Fish Dis.* 34:273–286. doi:10.1111/j.1365-2761.2010.01234.x.

Graham, D.A., H.M. Rowley, E. Fringuelli, G. Bovo, A. Manfrin, M.F. McLoughlin, C. Zarza, M. Khalili, and D. Todd. 2007. First laboratory confirmation of salmonid alphavirus infection in Italy and Spain. *J. Fish Dis.* 30:569–572. doi:10.1111/j.1365-2761.2007.00826.x.

Graham, D.A., H.M. Rowley, I.W. Walker, J.H. Weston, E.J. Branson, and D. Todd. 2003. First isolation of sleeping disease virus from rainbow trout, Oncorhynchus mykiss (Walbaum), in the United Kingdom. *J Fish Dis*. 26:691–694. doi:10.1046/j.1365-2761.2003.00505.x.

Graham, D.A., H.R. Rowley, and P. Frost. 2014. Cross-neutralization studies with salmonid alphavirus subtype 1-6 strains: Results with sera from experimental studies and natural infections. *J. Fish Dis.* 37:683–691. doi:10.1111/jfd.12167.

Gudding, R., A. Lillehaug, and Ø. Evensen eds. . 2014. Fish Vaccination. John Wiley & Sons, Ltd, Chichester, UK.

Hayes, B.J., A. Gjuvsland, and S. Omholt. 2006. Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. *Heredity (Edinb).* 97:19–26. doi:10.1038/sj.hdy.6800827.

Herath, T.K., J.E. Bron, K.D. Thompson, J.B. Taggart, A. Adams, J.H. Ireland, and R.H. Richards. 2012. Transcriptomic analysis of the host response to early stage salmonid alphavirus (SAV-1) infection in Atlantic salmon Salmo salar L. *Fish Shellfish Immunol.* 32:796–807. doi:10.1016/j.fsi.2012.02.001.

Hjortaas, M.J., B. Bang Jensen, T. Taksdal, A.B. Olsen, A. Lillehaug, E. Trettenes, and H. Sindre. 2016. Genetic characterization of salmonid alphavirus in Norway. *J. Fish Dis.* 39:249–257. doi:10.1111/jfd.12353.

Hjortaas, M.J., H.R. Skjelstad, T. Taksdal, A.B. Olsen, R. Johansen, B. Bang-Jensen, I. Ørpetveit, and H. Sindre. 2013. The first detections of subtype 2-related salmonid alphavirus (SAV2) in Atlantic salmon, Salmo salar L., in Norway. *J. Fish Dis.* 36:71–74. doi:10.1111/j.1365-2761.2012.01445.x.

Hodneland, K., A. Bratland, K.E. Christie, C. Endresen, and A. Nylund. 2005. New subtype

of salmonid alphavirus (SAV), Togaviridae, from Atlantic salmon Salmo salar and rainbow trout Oncorhynchus mykiss in Norway. *Dis. Aquat. Organ.* 66:113–120. doi:10.3354/dao066113.

Houston, R.D., A. Gheyas, A. Hamilton, D.R. Guy, A.E. Tinch, J.B. Taggart, B.J. McAndrew, C.S. Haley, and S.C. Bishop. 2008. Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (Salmo salar). *Dev. Biol. (Basel).* 132:199–204.

Houston, R.D., C.S. Haley, A. Hamilton, D.R. Guy, J.C. Mota-Velasco, A.A. Gheyas, A.E. Tinch, J.B. Taggart, J.E. Bron, W.G. Starkey, B.J. Mcandrew, D.W. Verner-Jeffreys, R.K. Paley, G.S.E. Rimmer, I.J. Tew, and S.C. Bishop. 2010. The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity (Edinb).* 105:318–27. doi:10.1038/hdy.2009.171.

Jansen, M.D., B. Bang Jensen, M.F. Mcloughlin, H.D. Rodger, T. Taksdal, H. Sindre, D.A. Graham, and A. Lillehaug. 2016. The epidemiology of pancreas disease in salmonid aquaculture: A summary of the current state of knowledge. *J. Fish Dis.* doi:10.1111/jfd.12478.

Jansen, M.D., B.B. Jensen, and E. Brun. 2015. Clinical manifestations of pancreas disease outbreaks in Norwegian marine salmon farming - variations due to salmonid alphavirus subtype. *J. Fish Dis.* 38:343–353. doi:10.1111/jfd.12238.

Jansen, M.D., M.A. Wasmuth, A.B. Olsen, B. Gjerset, I. Modahl, O. Breck, R.N. Haldorsen, R. Hjelmeland, and T. Taksdal. 2010. Pancreas disease (PD) in sea-reared Atlantic salmon, Salmo salar L., in Norway; a prospective, longitudinal study of disease development and agreement between diagnostic test results. *J. Fish Dis.* 33:723–736. doi:10.1111/j.1365-2761.2010.01176.x.

Jensen, B.B., A.B. Kristoffersen, C. Myr, and E. Brun. 2012. Cohort study of effect of vaccination on pancreas disease in Norwegian salmon aquaculture. *Dis. Aquat. Organ.* 102:23–31. doi:10.3354/dao02529.

Karlsen, M., K. Hodneland, C. Endresen, and A. Nylund. 2006. Genetic stability within the Norwegian subtype of salmonid alphavirus (family Togaviridae). *Arch. Virol.* 151:861–874. doi:10.1007/s00705-005-0687-6.

Karlsen, M., S. Villoing, E. Rimstad, A. Nylund, M. McLoughlin, D. Graham, S. Villoing, M. Bearzotti, S. Chilmonczyk, J. Castric, M. Bremont, J. Weston, M. Welsh, M. McLoughlin, D. Todd, A. Firth, B. Chung, M. Fleeton, J. Atkins, J. Strauss, E. Strauss, L. Kaariainen, T. Ahola, S. Schlesinger, S. Schlesinger, E. Fringuelli, H. Rowley, J. Wilson, R. Hunter, H. Rodger, D. Graham, K. Hodneland, A. Bratland, K. Christie, C. Endresen, A. Nylund, J. Weston, D. Graham, E. Branson, H. Rowley, I. Walker, V. Jewhurst, H. Jewhurst, D. Todd, M. Karlsen, K. Hodneland, C. Endresen, A. Nylund, C. Moriette, M. Leberre, A. Lamoureux, T. Lai, M. Bremont, J. Weston, S. Villoing, M. Bremont, J. Castric, M. Pfeffer, V. Jewhurst, M. McLoughlin, O. Rodseth, K. Christie, J. Koumans, D. Todd, A. Powers, A. Brault, Y. Shirako, E. Strauss, W. Kang, J. Strauss, S. Weaver, M. Zuker, T. Dubensky, D. Driver, J. Polo, B. Belli, E. Latham, C. Ibanez, S. Chada, D. Brumm, T. Banks, S. Mento, D. Jolly, S. Chang, and R. Hardy. 2009. Characterization of untranslated regions of the salmonid alphavirus 3 (SAV3) genome and construction of a SAV3 based replicon. *Virol. J.* 6:173. doi:10.1186/1743-422X-6-173.

Kent, M.L., and R.A. Elston. 1987. Pancreas disease in pen-reared Atlantic salmon in North America. *Bull. Eur. Ass. Fish Pathol*. 7:29–31.

Kongtorp, R.T., A. Stene, P.A. Andreassen, V. Aspehaug, D.A. Graham, T.M. Lyngstad, A.B. Olsen, R.S. Olsen, M. Sandberg, N. Santi, C. Wallace, and O. Breck. 2010. Lack of evidence for vertical transmission of SAV 3 using gametes of Atlantic salmon, Salmo

salar L., exposed by natural and experimental routes. *J. Fish Dis.* 33:879–888. doi:10.1111/j.1365-2761.2010.01190.x.

Kristoffersen, A.B., H. Viljugrein, R.T. Kongtorp, E. Brun, and P.A. Jansen. 2009. Risk factors for pancreas disease (PD) outbreaks in farmed Atlantic salmon and rainbow trout in Norway during 2003-2007. *Prev. Vet. Med.* 90:127–136. doi:10.1016/j.prevetmed.2009.04.003.

Larsen, H.A.S., L. Austbø, A. Nødtvedt, T.W.K. Fraser, E. Rimstad, P.G. Fjelldal, T. Hansen, and E.O. Koppang. 2014. The effect of vaccination, ploidy and smolt production regime on pathological melanin depositions in muscle tissue of Atlantic salmon, Salmo salar L. *J. Fish Dis.* 37:327–340. doi:10.1111/jfd.12106.

Larsson, T., A. Krasnov, J. Lerfall, T. Taksdal, M. Pedersen, and T. Mørkøre. 2012. Fillet quality and gene transcriptome profiling of heart tissue of Atlantic salmon with pancreas disease (PD). *Aquaculture*. 330–333:82–91. doi:10.1016/j.aquaculture.2011.12.016.

Lerfall, J., T. Larsson, S. Birkeland, T. Taksdal, P. Dalgaard, S. Afanasyev, M.T. Bjerke, and T. Mørkøre. 2012. Effect of pancreas disease (PD) on quality attributes of raw and smoked fillets of Atlantic salmon (Salmo salar L.). *Aquaculture*. 324–325:209–217. doi:10.1016/j.aquaculture.2011.11.003.

Lien, S., B.F. Koop, S.R. Sandve, J.R. Miller, M.P. Kent, T. Nome, T.R. Hvidsten, J.S. Leong, D.R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R.A. Hermansen, K. Von Schalburg, E.B. Rondeau, A. Di Genova, J.K.A. Samy, J.O. Vik, M.D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. Inge Våge, P. De Jong, T. Moen, and M. Baranski. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature*. 533. doi:10.1038/nature17164.

La Linn, M., J. Gardner, D. Warrilow, G.A. Darnell, C.R. McMahon, I. Field, A.D. Hyatt, R.W. Slade, and A. Suhrbier. 2001. Arbovirus of marine mammals: a new alphavirus isolated from the elephant seal louse, Lepidophthirus macrorhini. *J. Virol.* 75:4103–9. doi:10.1128/JVI.75.9.4103-4109.2001.

Massault, C., H. Bovenhuis, C. Haley, and D.-J. de Koning. 2008. QTL mapping designs for aquaculture. *Aquaculture*. 285:23–29. doi:10.1016/j.aquaculture.2008.06.040.

Mattilsynet | Statens tilsyn for planter, fisk, dyr og næringsmidler. 2016.

McLoughlin, M.F., and D. a Graham. 2007. Alphavirus infections in salmonids, a review. *J. Fish Dis.* 30:511–531. doi:10.1111/j.1365-2761.2007.00848.x.

McLoughlin, M.F., E. Peeler, K.L. Foyle, H.D. Rodger, D. O'Ceallachain, and F. Geoghegan. 2003. An Epidemiological Investigation of the Re-Emergence of Pancreas Disease in Irish Farmed Atlantic Salmon (Salmo Salar L .) in 2002. *Mar. Environ. Heal. Ser.* 1–41.

Menzies, F.D., M.F. McLoughlin, S.B. Wheatley, and E.A. Goodall. 1996. Development of a computerized information retrieval system for Atlantic salmon, Salmo salar L., production. *Aquac. Res.* 27:183–190. doi:10.1111/j.1365-2109.1996.tb00983.x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157:1819–1829. doi:11290733.

Moen, T., M. Baranski, A.K. Sonesson, and S. Kjøglum. 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*. 10:368. doi:10.1186/1471-2164-10-368.

Moen, T., K.T. Fjalestad, H. Munck, and L. Gomez-Raya. 2004. A multistage testing strategy for detection of quantitative trait loci affecting disease resistance in Atlantic salmon. *Genetics*. 167:851–858. doi:10.1534/genetics.103.013227.

Moen, T., A.K. Sonesson, B. Hayes, S. Lien, H. Munck, and T.H.E. Meuwissen. 2007. Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (Salmo Salar): comparing survival analysis with analysis on

affected/resistant data. *Bmc Genet.* 8:13. doi:53\r10.1186/1471-2156-8-53.
Munro A.L.S., Ellis A.E., McVicar A.H., N.E.A. 1984. An exocrine disease of farmed Atlantic salmon in Scotland. *Helgoländer Meeresunters*. 37:571–586.
Murphy, T.M., H.D. Rodger, E.M. Drinan, F. Gannon, P. Kruse, and W. Korting. 1992. The sequential pathology of pancreas disease in Atlantic salmon farms in Ireland. *J. Fish Dis.* 15:401–408. doi:10.1111/j.1365-2761.1992.tb01239.x.
Nelson, R.T., M.F. Mcloughlin, H.M. Rowley, M.A. Platten, and J.I. Mccormick. 1995. Isolation of a toga-like virus from farmed Atlantic salmon Salmo salar with pancreas disease. 22:25–32.
Norris, A., L. Foyle, and J. Ratcliff. 2008. Heritability of mortality in response to a natural pancreas disease (SPDV) challenge in Atlantic salmon, Salmo salar L., post-smolts on a West of Ireland sea site. *J. Fish Dis.* 31:913–920. doi:10.1111/j.1365-2761.2008.00982.x.
Poppe, T., E. Ri, M. Ad, and B. Hyllseth. 1989. Pancreas disease in Atlantic salmon (Salmo salar) postsmolts infected with infectious pancreatic necrosis virus (IPNV). *Bull. Eur. Ass. Fish Pathol*. 9:83–85.
Poppe, T.T., and G. Knudsen. 2005. Side-effects of vaccination: an example of the conflict between guidelines and real life.
Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. a R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–75. doi:10.1086/519795.
R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria.
Raynard, R.;Houghton, G.;Munro, A.L.S.B. 1992. Pancreas disease of Atlantic salmon: proceedings of a European Commission workshop. 15 pp.
Raynard, R.S., and G. Houghton. 1993. for the transmission of pancreas disease of Atlantic salmon Salmo salar. *Dis. Aquat. Organ.* 15:123–128.
Rodger, H., and S. Mitchell. 2007. Epidemiological observations of pancreas disease of farmed Atlantic salmon, Salmo salar L., in Ireland. *J. Fish Dis.* 30:157–67. doi:10.1111/j.1365-2761.2007.00799.x.
Skjelstad, H.R., G. Bornø, K. Flesja, H. Hansen, H. Nilsen, M.A. Wasmuth, and B. Hjeltnes. 2007. Helsesituasjonen hos oppdrettsfi sk 2007. 1-19 pp.
Sommerset, I., B. Krossøy, E. Biering, and P. Frost. 2005. Vaccines for fish in aquaculture. *Expert Rev. Vaccines*. 4:89–101. doi:10.1586/14760584.4.1.89.
Sonesson, A.K. 2007. Within-family marker-assisted selection for aquaculture species. *Genet. Sel. Evol.* 39:301. doi:10.1186/1297-9686-39-3-301.
Statistics Norway. 2016. Aquaculture - annually, final figures - SSB.
Stene, A. 2013. Salmon: pancreas disease spread : Compute Scotland. *Salmon pancreas Dis. spread*.
Stormoen, M., A.B. Kristoffersen, and P.A. Jansen. 2013. Mortality related to pancreas disease in Norwegian farmed salmonid fish, Salmo salar L. and Oncorhynchus mykiss (Walbaum). *J. Fish Dis.* 36:639–645. doi:10.1111/jfd.12060.
Taksdal, T., B. Bang Jensen, I. Böckerman, M.F. McLoughlin, M.J. Hjortaas, A. Ramstad, and H. Sindre. 2015. Mortality and weight loss of Atlantic salmon, Salmon salar L., experimentally infected with salmonid alphavirus subtype 2 and subtype 3 isolates from Norway. *J. Fish Dis.* 38:1047–1061. doi:10.1111/jfd.12312.
Taksdal, T., a B. Olsen, I. Bjerkas, M.J. Hjortaas, B.H. Dannevig, D. a Graham, and M.F. McLoughlin. 2007. Pancreas disease in farmed Atlantic salmon, *Salmo salar* L., and rainbow trout, *Oncorhynchus mykiss* (Walbaum), in Norway. *J Fish Dis*. 30:545–558.

doi:10.1111/j.1365-2761.2007.00845.x.

The International Cooperation to Sequence the Atlantic Salmon Genome (ICSASG). SalmoBase.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–23. doi:10.3168/jds.2007-0980.

Villoing, S., M. Bearzotti, S. Chilmonczyk, J. Castric, and M. Bremont. 2000. Rainbow Trout Sleeping Disease Virus Is an Atypical Alphavirus. *J. Virol.* 74:173–183. doi:10.1128/JVI.74.1.173-183.2000.

Weston, J., S. Villoing, M. Bremont, J. Castric, M. Pfeffer, V. Jewhurst, M. McLoughlin, O. Rodseth, K.E. Christie, J. Koumans, and D. Todd. 2002. Comparison of Two Aquatic Alphaviruses, Salmon Pancreas Disease Virus and Sleeping Disease Virus, by Using Genome Sequence Analysis, Monoclonal Reactivity, and Cross-Infection. *J. Virol.* 76:6155–6163. doi:10.1128/JVI.76.12.6155-6163.2002.

Weston, J.H., D.A. Graham, E. Branson, H.M. Rowley, I.W. Walker, V.A. Jewhurst, H.L. Jewhurst, and D. Todd. 2005. Nucleotide sequence variation in salmonid alphaviruses from outbreaks of salmon pancreas disease and sleeping disease. *Dis. Aquat. Organ.* 66:105–111. doi:10.3354/dao066105.

Xu, C., T.C. Guo, S. Mutoloki, Y. Haugland, and Y. Evensen. 2012. Gene expression studies of host response to Salmonid alphavirus subtype 3 experimental infections in Atlantic salmon. *Vet. Res.* 43:2–11. doi:10.1186/1297-9716-43-78.

Yang, J., B. Benyamin, B.P. Mcevoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569. doi:10.1038/ng.608.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. 2011a. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88:76–82. doi:10.1016/j.ajhg.2010.11.011.

Yang, J., M.N. Weedon, S. Purcell, G. Lettre, K. Estrada, C.J. Willer, A. V. Smith, E. Ingelsson, J.R. O'Connell, M. Mangino, R. Mägi, P.A. Madden, A.C. Heath, D.R. Nyholt, N.G. Martin, G.W. Montgomery, T.M. Frayling, J.N. Hirschhorn, M.I. McCarthy, M.E. Goddard, P.M. Visscher, and GIANT Consortium. 2011b. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19:807–812. doi:10.1038/ejhg.2011.39.

Ødegård, J., M. Baranski, B. Gjerde, and T. Gjedrem. 2011. Methodology for genetic evaluation of disease resistance in aquaculture species: Challenges and future prospects. *Aquac. Res.* 42:103–114. doi:10.1111/j.1365-2109.2010.02669.x.

APPENDIX

Supplementary table 1. Summary information on genotype counts and allele frequency for the affected (dead) and unaffected (survivors) fish for significant makers on chromosome 21

Chr.	Variant name (SNP ID)	Position (BP)	Minor allele	Freq (p)[1]	Beta (α)	-$\log_{10}$P	$\sigma^2_{G_{SNP}}$ [2]	$\%\sigma^2_{p_{SNP}}$ [3]	MAF_D[4]	MAF_S[5]	GENO_D[6]	GENO_S[7]
1	AX-88000746	38024743	B	0,257	-0,257	26,5	0,025	10,32	0,189	0,365	19/247/488	66/215/194
1	AX-96244562	121330871	B	0,106	-0,198	9,9	0,007	3,04	0,069	0,168	5/96/671	39/78/347
1	AX-87901482	98437358	B	0,057	0,239	7,0	0,006	2,49	0,079	0,023	5/105/614	0/23/472
2	AX-87954298	64061781	B	0,170	-0,228	15,9	0,015	5,99	0,127	0,244	8/180/587	57/106/287
5	AX-87910240	34339204	A	0,467	0,228	26,9	0,026	10,57	0,556	0,330	228/382/144	35/250/200
5	AX-96351825	29910408	B	0,023	0,412	8,3	0,008	3,10	0,039	0,000	0/56/670	0/0/500
6	AX-88306486	83935112	B	0,213	-0,278	23,9	0,026	10,60	0,148	0,318	0/225/533	63/168/231
6	AX-88290240	25078855	B	0,048	-0,448	22,5	0,018	7,50	0,011	0,107	0/17/743	7/88/382
7	AX-87896560	41750449	A	0,033	-0,693	35,1	0,030	12,38	0,000	0,087	0/0/776	0/81/384
7	AX-88315769	35193002	B	0,482	0,213	25,4	0,023	9,31	0,567	0,356	247/348/148	72/211/216
8	AX-88162080	0	B	0,348	0,582	134,5	0,154	62,80	0,523	0,066	117/555/82	0/62/407
8	AX-87197166	25180474	B	0,490	-0,237	26,4	0,028	11,45	0,396	0,638	131/339/288	181/249/49
9	AX-87098832	14940050	B	0,208	-0,180	10,6	0,011	4,38	0,170	0,268	3/257/516	86/97/319
11	AX-98317035	26177739	B	0,288	0,220	22,0	0,020	8,08	0,366	0,170	132/275/330	0/165/321
14	AX-98657908	26178413	B	0,244	0,215	19,1	0,017	6,95	0,315	0,140	115/231/385	3/135/364
14	AX-88044860	44195756	B	0,042	0,423	16,2	0,014	5,84	0,070	0,000	0/103/634	0/0/502

Chr.	Variant name (SNP ID)	Position (BP)	Minor allele	Freq (p)[1]	Beta (α)	-log$_{10}$P	$\sigma^2_{G_{SNP}}$ [2]	$\%\sigma^2_{p_{SNP}}$ [3]	MAF_D[4]	MAF_S[5]	GENO_D[6]	GENO_S[7]
15	AX-86994027	17035303	A	0,024	0,371	8,4	0,007	2,68	0,041	0,000	0/60/666	0/0/503
18	AX-88054471	28850676	B	0,062	-0,434	24,9	0,022	8,91	0,024	0,123	0/37/726	0/115/353
26	AX-98327033	35447258	B	0,189	-0,135	6,8	0,006	2,27	0,157	0,239	24/195/553	30/178/290
30	AX-88101206	381	A	0,454	-0,260	29,6	0,034	13,75	0,388	0,554	99/367/263	153/237/100
30	AX-88147014	17017	B	0,474	0,230	25,6	0,026	10,75	0,557	0,346	252/355/164	21/303/175
30	AX-88095996	4228	B	0,043	0,435	18,5	0,016	6,42	0,073	0,000	0/107/625	0/0/503

[1]Freq: the frequency of the minor allele.

[2]$\sigma^2_{G_{SNP}}$: Genetic variance $=2pq\alpha^2$.

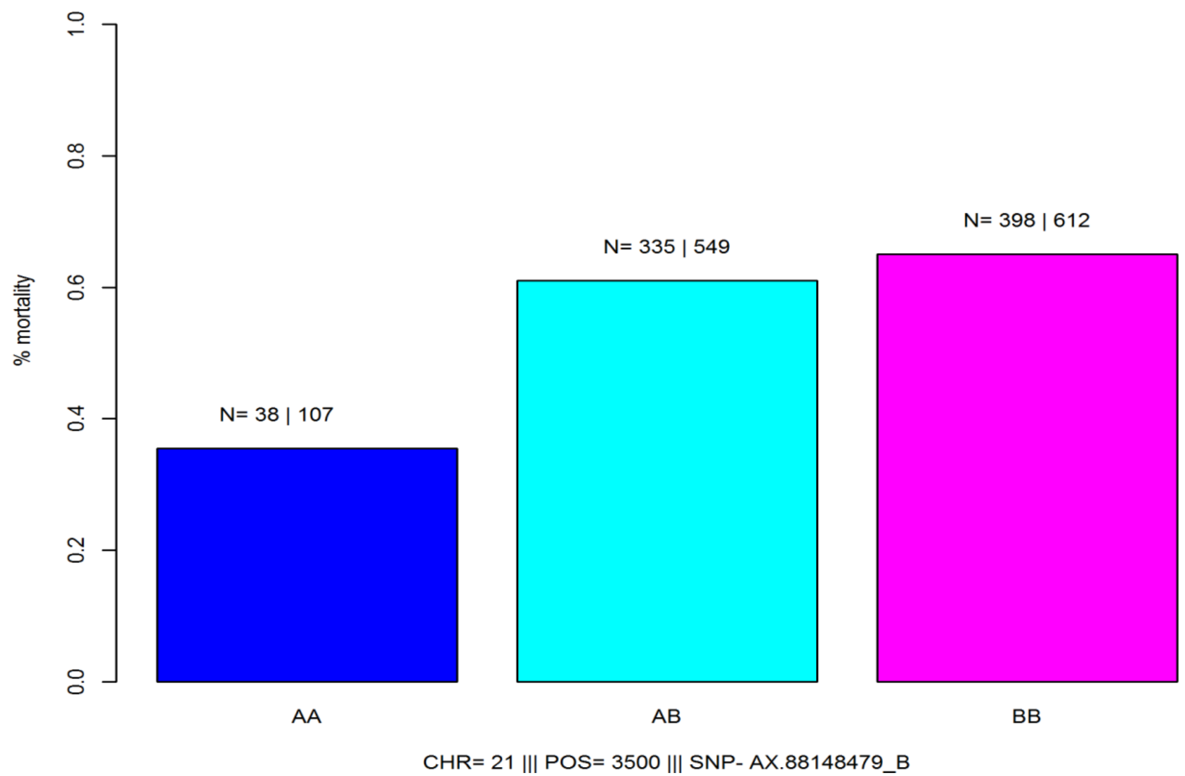
[3]$\%\sigma^2_{p_{SNP}}$: The percentage of the total phenotypic variance (σ^2_p=0.247) captured by each SNP [$=2pq\alpha^2/\sigma^2_p$].

[4]MAF_D: the frequency of the minor allele in affected (dead) fish.

[5]MAF_S: the frequency of the minor allele in unaffected (survivors) fish.

[6]GENO_D: the genotype counts in affected fish, with the allele count of the minor allele (left), heterozygotes (middle) and major allele (right).

[7]GENO_S: the genotype counts in unaffected fish, with the allele count of the minor allele (left), heterozygotes (middle) and major allele (right).



Supplementary figure 1. Counts of mortalities by corresponding genotypes for the most significant SNP (AX-88148479) on chromosome 21, showing that homozygote for favorable allele (first allele) mortality decrease when compared to the heterozygote and homozygote for the second allele.

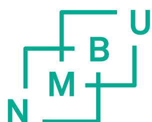
Supplementary table 2. Candidate genes on chromosome 21, region from 34.5 to 35.9 Mb.

Nr.	Gene code	Gene function
1	HTR2A	Protein Coding gene. Affects neural activity, perception, cognition and mood (*loss of apetite*).
2	LRCH1	Protein Coding gene. Diseases associated with LRCH1 include osteoarthritis.
3	GBP2	Protein Coding gene. Among its related pathways are Immune System and Interleukin receptor SHC signaling. Also have *antiviral effects and inhibit tumor cell proliferation.*
4	LREO3	L-amino acid oxidase from Agkistrodon halys pallas (*antibacterial activity*).
5	ITM2B	Protein Coding gene. Function not clear, but possibly associated with retinal dysfunctions and *inhibition of tumor cell proliferation*.
6	NEK1	Protein Coding gene. Associated with *tumor cells proliferation*. In response to injury that includes DNA damage, limits mitochondrial cell death.
7	ALG11	Protein Coding gene. Among its related pathways are Transport to the Golgi and subsequent modification and *Metabolism*.
8	NLGN4B	Protein. Belongs to a family of proteins that affect interactions between neurons.
9	MXRA5A	Protein Coding gene. This gene encodes one of the matrix-remodelling associated proteins. Diseases associated with MXRA5 include lung *cancer*.
10	CD99	Protein Coding gene. Involved in T-cell adhesion processes. Diseases associated with this gene include neuroepithelial *tumor*.
11	TTNA	Protein Coding gene. Connection of this gene with *cardiomyopathy and muscular dystrophy* were found in zebra fish.
12	STS	Protein Coding gene. It belongs to the sulfatase family and hydrolyzes several 3-beta-hydroxysteroid sulfates, which serve as metabolic precursors for estrogens, androgens, and cholesterol. Diseases associated with STS include rear eye and skin diseases.
13	SH3RF3	Paralog SH3 is a protein Coding gene. Among its related pathways are *cardiac conduction and smooth muscle contraction.*
14	CNOT11	Protein Coding gene. Among its related pathways are gene expression and Regulation *of TP53 (tumor suppressor gene)* Activity.
15	LCP1	Protein Coding gene. Among its related pathways are Signaling by GPCR (G protein–coupled receptors, which are involved in many diseases) and Lipoprotein *metabolism*. GO annotations related to this gene include *calcium ion binding and actin binding.*

Nr.	Gene code	Gene function
16	RPL31	Protein Coding gene. Diseases associated with RPL31 include *infectious diseases*, diseases of the blood and blood-forming organs and certain disorders *involving the immune mechanism*. Among its related pathways are Gene Expression and Metabolism. GO annotations related to this gene include poly(A) RNA binding and structural constituent of ribosome.
17	TBC1D8B	Protein Coding gene. Among its related pathways is cellular transport. GO annotations related to this gene include *calcium ion binding* and GTPase activator activity.
18	PDCL3	Protein Coding gene. GO annotations related to this gene include protein binding involved in protein folding and vascular endothelial growth factor receptor 2 binding. Among paralogs` PDCL2 related pathways are *Transport to the Golgi* and subsequent modification
19	TNIK	Protein Coding gene. Among its related pathways are Cellular Senescence and TNF signaling (REACTOME). GO annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity.
20	IL1R1	Protein Coding gene. It is an important mediator involved in many cytokine induced immune and inflammatory responses. Controls many different cellular functions including proliferation, differentiation and cell survival/apoptosis but are also involved in several pathophysiological processes. Among its related pathways are *Immune System and Interleukin-3, 5 a*nd GM-CSF signaling. GO annotations related to this gene include signal transducer activity and protease binding.
21	IL1R2	Protein Coding gene. Diseases associated with IL1R2 include reproductive diseases, as well as *Diseases of the musculoskeletal system and connective tissue.* Function is same as gene il1r1.
22	DNAJC10	Protein Coding gene. Among its related pathways are Protein processing in endoplasmic reticulum. GO annotations related to this gene include chaperone binding and protein disulfide oxidoreductase activity.
23	HMGT	High mobility group (HMG) proteins are ubiquitous nuclear proteins that regulate and facilitate various DNA-related activities such as transcription, replication, recombination and repair. HMGs bind to DNA and chromatin and act as "architectural elements" that induce both short- and long-range changes in the structure of their binding sites. They affect the activities of various regulatory molecules, including hormone receptors, *TP53 (tumor suppressor),* the RAG

Nr.	Gene code	Gene function
		proteins involved in V(D)J recombination, the homeotic protein HOXD9 of HIV integrase, and several transcription factors. The functional motifs of the ubiquitous HMG proteins are widespread and found in the DNA binding domains of numerous regulatory proteins.
24	FRZB	Protein Coding gene. Diseases associated with FRZB include osteoarthritis 1 and osteoarthritis. Among its related pathways are Endochondral Ossification and Adipogenesis. GO annotations related to this gene include G-protein coupled receptor activity and Wnt-protein binding. Diseases associated with paralog FZD10 include *cancer*.
25	NCKAP1	Protein Coding gene. Among its related pathways are Signaling by GPCR and Signaling by Rho GTPases. GO annotations related to this gene include protein complex binding and Rac GTPase binding.
26	DUSP19	Protein Coding gene. Diseases associated with DUSP19 include skin diseases and neurological diseases. GO annotations related to this gene include phosphatase activity and protein tyrosine/serine/threonine phosphatase activity.

*GO- gene ontology



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway