Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Animal and Aquaculture Sciences

Master Thesis 2014
30 credit

THE ACCURACY OF GENOMIC PREDICTION FOR VIRAL NERVOUS NECROSIS AND VIBRIOSIS DISEASE RESISTANCE IN ATLANTIC COD

Gebreyohans Tesfaye

Gebreyohans Tesfaye

NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Department of Animal and Aquaculture Sciences

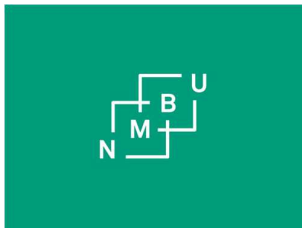
THE ACCURACY OF GENOMIC PREDICTION FOR VIRAL NERVOUS NECROSIS AND VIBRIOSIS DISEASE RESISTANCE IN ATLANTIC COD

GEBREYOHANS TESFAYE

EUROPEAN MASTER IN ANIMAL BREEDING AND GENETICS

THESIS ANIMAL BREEDING AND GENETICS (M30-IHA)

August 2014



Main supervisor:
Prof. Theodorus H.E. Meuwissen

Department of Animal and Aquaculture Sciences (IHA)

Norwegian University of Life Sciences (NMBU)

Co-supervisors:

1. Anna K. Sonesson, PhD
2. Marie Lillehammer, PhD
Nofima AS

Table of contents

List of Tables

List of Figures

Acknowledgments

I would like to express my deepest gratitude to my academic advisors, Prof. Theodorus Meuwissen, Anna K. Sonesson (PhD) and Marie Lillehammer (PhD) for their unreserved support, comments, guidance, reference material provision and encouragement.

I am deeply grateful to Nofima for providing me the data as well as office.

I like to thank Koepon foundation for financial coverage and EMABG secretarial for their nice coordination and guidance.

I am also grateful to Matthew Price for his comments and encouragement.

Finally, I would like to thank my family. They are always supporting me and encouraging me.

Abstract

The objective of this study was to estimate accuracy of genomic prediction for disease resistance to viral nervous necrosis and vibriosis using sparse and genome sequence SNP-data in Atlantic cod. The disease challenge test data of viral nervous necrosis and vibriosis used in this study were obtained from the National Atlantic cod breeding program which is running in Tromsø, Norway and we used disease challenge test data of year-class 2009 for both traits. Disease resistance for both traits was measured as survival at a fixed point in time and assessed as a binary variable. We obtained the result of challenge test data of 707 and 728 individuals for viral nervous necrosis and vibriosis respectively. The individuals came from75 full-sib and half-sib families for both diseases and the number of individuals per family varied from 7 to 20 (average of 9.7) in viral nervous necrosis, and 6 to 10 in vibriosis. On top of pedigree information of 1,743 individuals, three genotype data sets were used in this study, and based on these data sets three different genomic relation matrices were calculated. These were SPARSE8 (genotype data of 283 SNP markers at chromosome 8 of 1,743 individuals), SPARSE GENOME (1,577 individuals' genotype data of 8,658 SNP markers across the entire genome) and DENSE8 (imputed high density genotypes (759,270 SNPs) of chromosome 8 of 1,743 individuals). The genomic relation matrices were used in the GBLUP with polygenic models to estimate the variance components which were explained by the genomic information, and the genomic estimated breeding values using ASReml software. Fivefold within-family cross validations were carried out by randomly masking 20% of phenotypic records within each family in order to evaluate the accuracy of prediction for the viral nervous necrosis disease trait. Each observation was masked once and 141 phenotypes were masked in the first, second and third cross validation tests, whereas 142 phenotypes were masked in the fourth and fifth cross validation tests. Finally, the phenotypic values of the masked individuals were predicted based on the 566 or 565 phenotypic observations of the unmasked individuals. In the case of a between- families cross validation test, the phenotypic values of 20% of the families were masked at a time and their phenotypic values were predicted from the other families' phenotypic values. A total of 15 families were masked in each cross validation and the total masked phenotypes were 140, 142, 136, 137 and 153 in the first, second, third, fourth and fifth cross validations respectively. The accuracy of prediction was calculated based on the correlation between the predicted phenotypic values and observed phenotypic values. The results of analysis showed that for the trait disease resistance to viral nervous necrosis, heritability estimates of the trait using

the traditional BLUP (h^2= 0.359) and GBLUP (SPARSE8) (h^2= 0.355) were almost equal. However, GBLUP (DENSE8) (h^2= 0.335) and SPARSE GENOME (h^2= 0.371) had the lowest and the highest heritability estimates respectively, but these differences were not significant according to their log- likelihood estimates. In the case of vibriosis, our data were not able to distinguish between the genetic variation explained by the genomic information and the pedigree information. The SPARSE GENOME gave a 0.117 heritability estimate by fixing the variance explained by pedigree information at the boundary 0. According to a within-family cross validation test for viral nervous necrosis, the accuracies were estimated as 0.329 in the case of the traditional BLUP and GBLUP (DENSE8) models, but 0.336 in the SPARSE GENOME model. In addition to this, results of between-family cross validation showed that the accuracy of prediction of the DENSE8 (0.15) was less than that of the SPARSE8 (0.16). In our study we found a high heritability of resistance to viral nervous necrosis in Atlantic cod in all models. However, our heritability estimate was lower than the extremely high estimates of other studies in Atlantic cod. The total number of fish, the average number of fish per family, and the model we used in our study could be possible reasons for our relatively lower estimate of heritability for disease resistance to viral nervous necrosis. In our study, the accuracy of prediction of the genomic estimated breeding values using the sparse SNP markers (SPARSE GENOME) did not show a big difference compared with the traditional estimated breeding values, and this could be due to the fact that the phenotypical and genotypical data we used for training was too small to accurately capture the whole fraction of the variance explained by the SNP chip. Moreover, the accuracy of prediction of imputed high density genotypes (DENSE8) of chromosome 8 for disease resistance for viral nervous necrosis was not better than that of SPARSE 8, and this could be because in within- family genomic selection, big segments are inherited together and so the sparse SNPs could be sufficient to detect the chromosome segments. The low heritability estimate of our study to the trait disease resistance for vibriosis is consistent across all studies. However, the accuracy of genomic prediction could not be assessed by cross-validation, since we were not able to distinguish the genetic variance explained by the genomic and pedigree information. In conclusion, for both traits more phenotypic and genotypic data are required in order to properly evaluate the accuracy of prediction of the genomic information.

Keywords: Accuracy of prediction /Atlantic cod/disease resistance/SNP/ viral nervous necrosis

1. Introduction

Fish diseases are major limiting factors facing the Norwegian fish farming industry today (Nakai *et al*., 2009, Woo *et al.*, 2011).Vaccination, drug therapy, hygiene and eradication of infected population are the current strategies to control infectious diseases (Woo *et al*., 2011). Therefore, the role of on-going research and implementation of appropriate disease control strategies are critically important to the fish farming industry (Woo *et al*., 2011). A variety bacterial, viral and parasitic caused diseases are known on farmed and wild cod (Woo *et al.*, 2011). Vibriosis is the most important bacterial disease in farmed Atlantic cod in Norway and it is causing singinificant mortalities and economic losses (Johansen *et al.,* 2011). Viral nervous necrosis (VNN) is a viral disease caused by Nodavirus (Nakai *et al.,* 2009). The virus infects more than 30 fish species and is causing mass mortality in aquaculture hatcheries worldwide (Korsnes, 2008, Yang *et al.,* 2012). Fish which are infected by the virus show abnormal swimming, dark coloration and anorexia as clinical signs, and in severe outbreaks the mortality rate can exceed 95% (Ransangan and Manin, 2010, Yang *et al.*, 2012). The virus also infects Atlantic cod (Gadus morhua) and causes economic losses. Disease outbreaks due to VNN have been reported in juvenile and adult farmed Atlantic cod in Norway (Hellberg *et al.*, 2010).

Disease control or management using host genetic resistance is increasingly recognized as a key component of effective disease control, complementing or sometimes replacing existing strategies (Daetwyler *et al.,* 2008). This is further supported by the significant genetic variation in disease resistance reported in different fish species (Chevassus and Dorson, 1990). The traditional breeding methods have been effective in selecting for animals with easy-to-measure production traits in the absence of molecular knowledge of the genes acting on quantitative trait loci (Goddard and Hayes, 2009b). However, the efficiency of these traditional methods decrease when traits are measured in one sex, after death or late in life, or if measuring the trait is expensive and difficult (Goddard and Hayes, 2009b, Eggen, 2012). Disease resistance is a difficult trait to improve using the traditional selection method and often shows a dichotomous distribution of phenotypes (diseased or non- diseased). Resistance against specific diseases affecting aquaculture species show moderate to high heritability when tested in artificial challenge tests (Ødegård et al., 2011, Fjalestad et al., 1993) . In artificial challenge test, fish are exposed to one specific pathogen at a time in a controlled

environment and mortalities are recorded on a daily basis or in some cases more frequently. The test is continued until at least 50% of the fish have died or until mortality of the fish stops rising (Fjalestad *et al*., 1993) . Although the method improves the heritability estimate to disease resistance, fish that have been challenge-tested for disease resistance cannot be used as breeding stock. As a result of this, selection for diseases resistance is based on a performance test of siblings of the candidates. However, in case of the traditional aquaculture breeding schemes, with a sib test, only 50% of the total genetic variance of the candidates is exploited, perhaps less (Nirea *et al*., 2012). In conclusion, the traditional breeding methods with a sib test using artificial challenge test do not capture the majority of the genetic variance to disease resistance and are therefore not very efficient to genetically improve disease resistance.

In order to exploit the full benefit of the genotypic information from whole genomes in the genetic evaluation of animals, Meuwissen *et al*. (2001) proposed the genomic selection (GS) method. It is a selection method, which simultaneously uses high density markers that cover the whole genome to predict the breeding values of selection (Meuwissen *et al*., 2001). The method assumes that every QTL (large or small) affecting the trait of interest can be explained by several nearby markers, which together explain all the genetic variation caused by QTL. Implementation of genomic selection involves estimation of the effect of chromosome segments/SNP markers in a reference population and prediction of genomic estimated breeding values for selection candidates not in the reference population. So, in order to implement genomic selection, a phenotyped and genotyped reference population for large sets of markers should be available (Goddard and Hayes, 2009b). The reference population is used to develop the prediction equation for the estimation of genomic breeding values in non-phenotyped individuals (selection candidates) based on their genotype alone. Therefore, genomic selection could overcome the shortcoming of the traditional selection methods and marker assisted selection and it can predict breeding values more accurately (Meuwissen *et al.*, 2001, Zhang *et al*., 2011).

Genomic selection in aquaculture industry can be used to solve the main drawbacks of traditional selection which are lack of pedigree recording and the large increase in inbreeding (Gjedrem and Gjøen, 1995, Nielsen *et al.*, 2009). It has been studied using computer simulations and all the simulation studies showed that genomic selection can be used in aquaculture to use both within- and between-family variation during selection for both continuous and dichotomous (disease resistance) traits to increase the accuracy of selection

(Nielsen *et al.*, 2009, Sonesson and Meuwissen, 2009). Simulation studies with the assumptions of high density genotyping and a large number of fish showed that genomic selection could increase accuracy of selection and genetic gain in aquaculture species particularly for traits measured on sibs of the selection candidates (Nielsen *et al.*, 2009). Another simulation study to extend the Bayes B method of genome wide evaluation (Meuwissen *et al.*, 2001) to include dichotomous traits indicated that the accuracy of genome wide evaluation for disease resistance in aquaculture sib based programs is increased by 16% compared with the linear model, in the case of low heritability and disease prevalence (Villanueva *et al*., 2011). Despite of the simulation findings, the actual applications of genomic selection in fish are few. Lack of dense marker maps and high- throughput SNP arrays are the constraint (Nielsen *et al.*, 2009, Sonesson and Meuwissen, 2009). Therefore, the contribution of currently developed high–density (~130 K) SNP genotyping array of Atlantic salmon has significant contribution to the practical application of genomic selection in aquaculture breeding programs (Houston et al., 2014).

In aquaculture breeding, full implementation of genomic selection is more expensive. This is due to the large number of selection candidates and test individuals required to genotype, and low economic value of each selected individual (Ibañez-Escriche and Gonzalez-Recio, 2011). Then, the cost of genotyping is one of the limitations for the implementation of genomic selection in aquaculture breeding programs. Within- family genomic selection is one option for low marker density genomic selection strategies to keep the cost of genotyping low (Lillehammer *et al.*, 2013). The method is a combination of genomic within-family breeding values, based on low density genotyping and conventional BLUP between family breeding values. Another alternative using the central idea of low density SNP panels is proposed by Goddard and Hayes (2009a), where the key ancestors are genotyped with dense panels or with whole-genome re-sequencing and the selection candidates are genotyped with standard low-density panels. Then, the chromosome segments in the selection candidates are traced back to the key ancestors and their genotypes are inferred at all markers assayed on the key ancestors. This method is known as imputation (Mulder *et al*., 2012). It is a method of exploiting linkage disequilibrium and/or linkage analysis by deducing a higher density genotype from a lower density genotype and which results in reduced genotyping costs. In Holstein-Friesian dairy cattle, an average accuracy of imputation from 2,909 SNPs to 54,001 SNPs of 0.98 was reported (VanRaden *et al.*, 2011). So, lower density genotyping together

with imputation can be used to reduce the cost of genomic selection (van Marle-Köster *et al.*, 2013).

The objective of this study was to estimate accuracy of genomic prediction for disease resistance to viral nervous necrosis and vibriosis using sparse and genome sequence SNP-data in Atlantic cod.

2. Literature review

2.1. The principle of genomic selection

Genomic selection is defined as a selection decision based on genomic breeding values. The method simultaneously uses high density markers that cover the whole genome to predict the breeding values of selection candidates (Meuwissen *et al.*, 2001). It assumes that every QTL (large or small) affecting the trait of interest can be explained by several nearby markers, which together explain all the genetic variation caused by QTL due to at least one of the markers being in linkage disequilibrium with each gene or QTL affecting the trait of interest. As a result of this, theoretically, it is possible to capture all genetic variance by markers if the marker density is high enough (Zhang *et al.*, 2011, Meuwissen *et al*., 2001).

2.2. Steps in genomic selection

Getting a large group of animals with accurate phenotypes for the trait(s) of interest is the first step in the genomic selection process. This large group of animals is called a reference population or training animals (Meuwissen, 2007, Goddard and Hayes, 2009b). The reference population also needs to be genotyped. Using both the genotypic and phenotypic information of the reference population, a prediction equation will be developed that predicts the breeding value from the SNP genotypes (Goddard and Hayes, 2009b). Then, the genomic breeding value of candidate animals can be calculated using the prediction equation, based on their genotypes from the SNP array without any accurate phenotypes for these animals. Finally, the candidate animals are ranked on the estimated genomic breeding values, and the best candidate animals are selected to breed the next generation (Goddard and Hayes, 2009b, Eggen, 2012).

2.3. Methodologies for genomic selection

Estimating the effect of each SNP on the trait is the main challenge in genomic prediction, since, the number of SNPs is much larger than the number of phenotypes available (Goddard and Hayes, 2009b). Several statistical methodologies have been developed to address this problem by using prior information about the distribution of the SNP effects. The methodologies can be grouped in two based on their assumptions about the SNP effects. GBLUP, BayesA and Bayesian LASSO assume all SNPs have effects but with different prior information. However, the second group such as BayesB, and BayesC take the assumption that many SNPs have no effect (de los Campos *et al.*, 2013). On top of the assumption on the SNPs effect, they also have an assumption on the distribution of SNPs effects. GBLUP

assumes the SNP effects are normally distributed (Meuwissen *et al*., 2001). BayesA and BayesB both assume SNP effects have a student`s t-distribution but some SNPs are assumed to have no effect in the case of BayesB (de los Campos *et al.*, 2013, Meuwissen *et al.*, 2001). Bayesian LASSO assumes that SNP effects have an exponential distribution. (Hayes *et al.*, 2013, Park and Casella, 2008).

2.4. The accuracy of genomic estimated breeding value

The correlation between the GEBVs and true breeding values is called accuracy (r) of genomic estimated breeding value (Meuwissen *et al.*, 2001). Unlike simulation studies, true breeding values are unknown in a real population. Daetwyler *et al*. (2008) derived the first formula that predict the accuracy of genomic estimated breeding values and it was derived as

$$r = \sqrt{(N_p h^2/(N_p h^2 + N_{qtl}))} \quad (1)$$

where *Np* is the total number of phenotypic records in the reference population, h^2 is the heritability of the trait investigated, and N_{qtl} is the number of independent QTLs affecting the trait. A different form of the formula was derived by Goddard (2009) as.

$$r = \sqrt{1 - \lambda/(2Np\sqrt{a})}\ \log\left(\frac{1 + a + 2\sqrt{a}}{1 + a - 2\sqrt{a}}\right) \quad (2)$$

where $a = 1+2\boldsymbol{\lambda}/Np$, $\boldsymbol{\lambda}= (1-h^2)\ M_e/\ (h^2\log\ (2N_e))$ in which *Ne* is the effective population size in historical population, and M_e is the effective number of chromosome segments estimated from $M_e = (2N_eL)/\log\ (4N_eL)$, where L is the length of the genome. After derivation of the above formula (2) Daetwyler *et al.* (2010) included M_e into the previous formula (1). They deduced that N_{qtl} has an influence on the BayesB method but not on GBLUP, so based on this conclusion they derived two formulas.

In the case of GBLUP,

$$r = \sqrt{N_p\ h^2/(N_p\ h^2 + M_e)} \quad (3)$$

Whereas in the case of BayesB method,

$$r = \sqrt{N_p\ h^2/(N_p\ h^2 + \min(M_{e,}\ N_{qtl}))} \quad (4)$$

Deviations between the observed and predicted accuracies can be due to the violation of the different assumptions made while the formulas were derived. In general both predicted and

observed accuracies respond in similar ways to change in the relevant parameters (Zhang *et al*., 2011).

2.5. Statistical analysis of disease- related traits

Cross–sectional and longitudinal models are the most common models which are used to analyse disease related traits (Noordhuizen et al., 2001). Cross-sectional models are the simplest statistical models used for analyzing challenge test data and in these models, disease resistance is measured as a single record (e.g., alive or dead) at a specific point in time (Ødegård *et al.*, 2011). The cross-sectional studies therefore take a snapshot of the situation at a specific moment but time of death is not taken into account (Thrusfield, 2013) . In the case of binary outcome variables, the residuals do not have normal distribution, which results in the violation of the assumption of linear regression (Noordhuizen *et al.*, 2001, Thrusfield, 2013). Therefore, generalized linear mixed models which account for the binary nature of the data are the appropriate type of model to analyse the nature of such binary data, using the *probit* link function (Threshold model) (Thrusfield, 2013) . However, as Ødegård *et al* (2011) reviewed most of studies found limited differences between the linear models and threshold models with respect to accuracy of predicted breeding value. Further, they added that, if the heritability and the family size increases, the relative difference is expected to decrease. They also noticed that studies which used a plain binary expression of alive or dead did not capture the full extent of the disease resistance as well as those studies which used a longitudinal measure of time until death (for example, the first fish to die are the least resistance ones). In such a follow up study, the exact survival time is only known for those individuals that show the event (dead) while only the minimum survival time is known for the survivors (Ødegård *et al.*, 2011). Moreover, the effect of environmental factors may not be constant over time and this usually causes the survival time to not be normally distributed. Proportional hazard models or survival score, which approximates proportional hazard models have been suggested for analysis of life time data (Ødegård *et al*., 2011). So, the types of model we use could vary depending on the data we have and the situation we want to study.

2.6. Nervous necrosis virus

Nodavirus or nervous necrosis virus is an emerging pathogen, which belongs to the viral family Nodaviridae (Samuelsen *et al.*, 2006). The Nodavirus causes the diseases commonly known as viral nervous necrosis (VNN) or viral encephalopathy and retinopathy (VER) in

marine fish farming worldwide (Korsnes, 2008, Samuelsen *et al.,* 2006). Although these viruses can affect older market size fish, it mainly affects a younger stage of fish and damages the central nervous system in susceptible fish species. Infected fish show abnormal swimming behavior, abnormal posture and muscle tremors (Chi *et al*., 2005).

2.6.1. Transmission of the virus

Horizontal and vertical transmission of VNN has been reported in several research findings (Korsnes, 2008, Hellberg *et al*., 2010). Vertical transmission can occur as transovum transmission where the pathogen is present on the egg surface, and the other way if the pathogen only is present inside the egg (transovarian) (Korsnes, 2008, Fraser, 1986). The horizontal transmission occurs through influent and rearing water, and via utensils, vehicles, and human activity (Nakai *et al*., 2009)

2.6.2. Diagnosis of the virus

Establishment of specific and sensitive methods for nodavirus detection are important, both as diagnostic tools and for scientific studies of the virus (Dalla Valle *et al.,* 2000). Tentative diagnosis of VNN is based on the appearance of vacuoles in the brain, spinal cord, and/or retina as seen by light microscopy (Hellberg *et al.*, 2010). However, individual fish showing only a few vacuoles in the nervous tissues pose a difficulty in diagnosis (Nakai *et al*., 2009). In general, the most commonly confirmatory diagnosis methods used could be divided into molecular, immunological and cell culture (Korsnes, 2008, Dalla Valle *et al.*, 2000). The immunologically based assays could be subdivided into direct and indirect detection of virus particles, and included enzyme-linked immunosorbent assay (ELISA), indirect fluorescent antibody technique (IFAT) and later immunohistochemistry (IHC) and neutralization tests (Korsnes, 2008). The most widely used immunological methods for diagnosing VNN have been ELISA and IHC (Breuil and Romestand, 1999). RT-PCR and realtime RT-PCR assays are molecular methods for the detection of VNN and have played an even more important role in detecting and characterizing viruses (Muroga, 1994).

2.6.3. Control of VNN

Korsnes (2008) suggested strategies for controlling VNN in farming of fish, and the strategies involve screening of fish for the disease and development of a vaccine. This resulted in the reduction of the risk of introduction and spread the virus in farmed populations of fish by

blocking both vertical and horizontal transmission routes (Nakai *et al.*, 2009). Screening and monitoring of health status of fish in order to identify infected (carrier) fish and keep the brood stock in a sheltered aquaculture system to prevent horizontal transmission of VNN are strategies which help to have VNN-free brood (Muroga, 1994, Hellberg *et al.*, 2010). Avoiding immunosuppression and subsequent increased chance of infection with the virus are key control strategies in an open system such as marine growth site, where fish are kept in cages at higher density and sources of virus might be other farmed fish at the site or wild stocks of fish in the vicinity (Fraser, 1986). Ødegård *et al.* (2010) have reported high heritability (0.75 in case of threshold mixed model and 0.43 using linear mixed model) for disease resistance to VNN using challenge test in Atlantic cod. So, in addition to the above control strategies, selective breeding of fish for VNN resistance is one alternative to control and prevent the disease in fish farming industry.

2.7. Vibriosis

Vibriosis is one of the most prevalent fish disease caused by a gram negative bacteria belonging to the genus Vibrio and it is commonly found in the aquatic environment, the majority of which are non-pathogenic (Actis *et al.*, 2011). *Vibrio anguillarum*, *Vibrio salmonicida* and *Vibrio ordalii* are the species which cause the most economically serious disease (Gratacap, 2008).

2.7.1. Transmission of Vibriosis

The interaction of the host-pathogen-environment trigger the occurrence of vibriosis outbreak, and the main predisposing factor is the stress level of the fish (Gratacap, 2008). Increase in temperatures has a known effect on the stress level of fish which result in disease outbreaks (Buller, 2004). Transport and transfer from freshwater to seawater and pollution can also increase the rate of vibriosis, on top this, sometimes vaccination of Atlantic cod can be followed by outbreaks (Gratacap, 2008). The mode of transmission of the bacterium is through water, with fish carriers or contaminated fish farm equipment and it can enter through skin, fins, gills and anus (Buller, 2004). The clinical sign of the disease are haemorrhage to intestines, spleen, muscle, body cavity and darkened coloration to skin and fins. In addition to this, changes to the eyes in size as well as color and grey or white lesion on intestines and spleen are also clinical sign of the disease (Actis *et al.*, 2011, Beacham and Evelyn, 1992).

2.7.2. Diagnosis of vibriosis

In the case of vibriosis there is no reliable presumptive diagnosis of the disease because of its similarity to other septicemic diseases caused by gram –negative bacteria (Bullock, 1977, Actis *et al.*, 2011). Diagnosis of the disease carried out using bacteriological examination (Actis *et al.*, 2011).

2.7.3. Control and prevention of the disease

Good management practices such as maintenance of water quality, low stocking densities and good husbandry should be used to control vibriosis (Beacham and Evelyn, 1992, Samuelsen *et al.,* 2006). In diseased fish, a successful treatment using antimicrobial is dependent on a rapid diagnosis and immediate treatment and Sulfamerazine and Oxytetracycline drugs can be used for the treatment of the disease(Actis *et al*., 2011). More over vaccination has proven to be an efficacious method in preventing vibriosis (Samuelsen *et al.*, 2006)

3. Materials and methods

3.1. Fish data

The disease challenge test data of viral nervous necrosis and vibriosis used in this study were obtained from the National Atlantic cod breeding program which is running in Tromsø, Norway. The national Atlantic cod breeding program was started from 2002 and three parallel year-classes were formed as progeny of the wild-caught Atlantic cod. Details of the year- classes and mating design can be found in Bangera *et al*. (2011). In this study, we used disease challenge test data of year-class 2009 for both traits. The 2009 year–class data represent the second generation (F_2) of the progeny of selected fish from year-class 2006 (Bangera *et al.*, 2011).

3.2. Data on disease challenge tests

Disease resistance for viral nervous necrosis and vibriosis traits was measured as survival at a fixed point in time. It was also assessed as a binary variable where the observed value of each individual fish was 1 if the fish died following challenge test and 0 if it survived.

3.2.1. Viral nervous necrosis challenge test data

The challenge test was conducted using injection of a nodavirus suspension. The challenge test period was 35 days and the first mortality was set as a starting point. Preparation of the viral isolate, titration of virus and methods of the challenge test is presented in Ødegård *et al.* (2010). We obtained the result of this challenge test data of 707 individuals from 75 full-sib families and half-sib families (75 dams and 60 sires) of Atlantic cod. The number of individuals per family varied from 7 to 20 (average of 9.7).

3.2.2. Vibriosis challenge test data

728 individual challenge test results of vibriosis were obtained from The National cod breeding program. The duration of challenge test was 27 days like the other trait the first mortality was set as a starting point but it was conducted using bath challenge. More details about the challenge test protocol and fish population can be found in Kettunen *et al.* (2007). The individuals were from 75 full-sib families and half-sib families (75 dams and 57 sires) of Atlantic cod and the number of individuals per family varied from 6 to 10.

3.3. Genotype data

The Illumina Atlantic cod 12K SNP array were developed by the CSC (www.cigene.no) and used to genotype the Atlantic cod. Three genotype data sets were used in this study and based on these data sets three different genomic relation matrices were calculated. The first data set

contained 1,577 genotyped animals and the total pedigree of these animals contained 1,743 entries. Genotype data on chromosome 8 (283 SNP markers) were used and the genotypes of the ungenotyped ancestors were imputed by LDMIP (Meuwissen and Goddard, 2010). We used this genotype data of 283 SNP markers to estimate the genomic relation matrix of the 1,743 individuals at chromosome 8 (which we will call the SPARSE8). The second data set was 1,577 individuals genotyped data across the entire genome with Illumina Low-Density Atlantic cod SNP chip and based on this data set we calculated genomic relation matrix of 1,577 individual using 8,658 SNP markers across the whole genome. This genomic relation matrix is called in this paper SPARSE GENOME. Low density genotypes data of chromosome 8 of 1,743 individuals' were imputed up to sequence density (759,270 SNPs) genotypes the genome sequence (approx. 10X) of 111 sequenced parents of the challenge tested families (imputation by LDMIP).These imputed high density genotypes (759,270 SNPs) were used to calculate the third genomic relation matrix (which we call the DENSE8). The genomic relation matrices were used in the GBLUP model to estimate the variance components which were explained by the genomic information and genomic breeding value. Details of the model and genomic relation matrix are presented in section 3.3.2. On top of the genotypes data, pedigree information of 1,743- individuals' were used in the traditional BLUP and GBLUP with polygenic models.

3.4. Statistical models

Variances explained by the pedigree only, DENSE8 or SPARSE8 or SPARSE GENOME were estimated using ASReml software (Gilmour *et al*., 2009). These variance components estimates were used to calculate the heritabilities of the traits. We used two best linear unbiased prediction (BLUP) models for the estimation and prediction of the traditional breeding value and genomic breeding values. The first model was the traditional BLUP, while the second model was the GBLUP with a polygenic effect (using the pedigree based relationship matrix). Both models are presented below (Henderson, 1975, Meuwissen *et al.*, 2001).

3.4.1. Traditional BLUP

The traditional estimated breeding values and predicted values for both traits were determined using phenotypes and pedigree information based on the following model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Za} + \mathbf{e} \quad (5)$$

Where y is the vector of observed phenotypic values of individuals, μ is the overall mean, **a** is the vector of additive genetic effects of the phenotyped individuals and their parents, Z is the incidence matrix of **a**, and **e** is the vector of residual errors. The variance-covariance matrices of **a** and **e** are $\mathbf{A}\sigma_a^2$ and $\mathbf{I}\sigma_e^2$, respectively, where A is the additive genetic relationship matrix, σ_a^2 is the additive genetic variance, and σ_e^2 is the residual variance (Henderson, 1975).

3.4.2. GBLUP with a polygenic effect

The genomic breeding values using SPARSE8 or DENSE8 or SPARSE GENOME genomic relation matrices were estimated using GBLUP including the polygenic effect (Meuwissen *et al.*, 2013, Meuwissen *et al.*, 2001). In this paper GBLUP (SPARSE8), GBLUP (DENSE8) and GBLUP (SPARSE GENOME) model means a GBLUP model with polygenic effect and used the genomic relation matrix of SPARSE8, DENSE8 and SPARSE GENOME respectively. These models were also used to predict the genomic breeding values of masked individuals in the case of cross validation test (Meuwissen *et al.*, 2013), and details of the cross validation test are presented in section 3.4. The model in matrix notation is presented below and y, $1_{n'}$ and μ have the same definition as in equation 5

$$\mathbf{y} = 1_n u + \mathbf{Za} + \mathbf{Wg} + \mathbf{e} \qquad (6)$$

Z= is the incidence matrix for the random polygenic effect

a= is a vector containing a random polygenic effect for each individual
a is assumed to follow $N(0,A\sigma_{\mathbf{a}}^2)$ where A is the pedigree based relationship matrix

W= is the incidence matrix for the random genomic effect

g= vector of random additive genetic effect using the genomic relationship matrix (G) coming from $N(0, G\sigma_g^2)$

e= vector of random residual errors $N(0, \mathbf{I}\sigma_e^2)$

The genomic relationship matrix (G) was calculated by using SNP marker genotype according to VanRaden (2008)

$$G = XX'/m$$

X= matrix of standardised SNP genotypes X_{ij}

m=number of SNPs

X_{ij} denotes the standardised SNP genotypes of animal *i* for SNP *j*

Genotypes values of the SNP are standardized to a mean of zero and a standard deviation of 1 by subtracting the mean ($2p_j$) and dividing by the standard deviation √H as presented below

For genotypes 0, 1 and 2

X_{ij} :$(0 - 2p_j)/\sqrt{H}$; $(1-2p_j)/\sqrt{H}$; $(2 - 2p_j)/\sqrt{H}$

Where heterozygosity (H) = $2p_j$ $(1 - p_j)$ and p_j is the allele frequency of allele 1 and 0, 1, or 2 are genotypes value for SNP with genotypes "0 0," "0 1," or "1 1," respectively. Thus, the genomic relationships between two animals were calculated and the calculated genomic relationship matrix is implemented in the above equation to calculate GBLUP breeding values and the variance explained by the markers using ASReml (Gilmour *et al*., 2009). In order to test whether SPARSE8 explains more variance of the trait, i.e. whether chromosome 8 explains more variance than other SNPs (disease resistance to viral nervous necrosis) than expected based on the SPARSE GENOME genomic relation matrix, we used the model GBLUP with polygenic effect by incorporating both SPARSE8 and SPARSE GENOME genomic relation matrices simultaneously by considering the covariance between the two genomic relation matrices. The model is presented below,

$$y = 1_n u + Za + Wg + Qv + e \quad (7)$$

Where y, u, Z, **a** **W**, **g** and **e** are similar like equation 6 and **Q** is a vector of genetic effect for a chromosome 8 assumed to be normally distributed with mean 0 and $G\sigma_v^2$.

3.5. Cross validation test

Within- and between-family cross validations were carried out in order to evaluate the accuracy of prediction for the viral nervous necrosis disease trait. The within-family cross validation will predict the accuracy of genomic breeding values under the assumption that the candidates have full-sibs with records. The between family cross validation will give the predicted accuracy under the assumption that phenotypic records come from more distant relatives.

3.5.1. Within family cross validation

Fivefold within-family cross validation tests was carried out by randomly masking 20% of phenotypic records within each family. As a result of this, each observation within family were masked once and 141 phenotypes were masked in the first, second and third cross validation tests whereas 142 phenotypes were masked in the fourth and fifth cross validation tests. The phenotypic values of the masked individuals were predicted based on the 566 or 565 (in the case of 142 individuals masked) phenotypic observations of the unmasked individuals. The accuracy of prediction was calculated based on the correlation between the predicted phenotypic values and observed phenotypic values. This correlation equals no more than the square root of the heritability of the trait, so we scaled the correlation result by the square root of the heritability to obtain an unbiased estimate of the accuracy of prediction from 5-fold cross validation (Meuwissen *et al.*, 2013).

3.5.2. Between family cross validation

Like the above procedure, fivefold between-family cross validation tests was carried out by randomly masking the phenotypic values of 20% of the families at a time, and their phenotypic values were predicted from the other families' phenotypic values (Meuwissen *et al.*, 2013). A total of 15 families were masked in each cross validation and the total masked phenotypes were 140, 142, 136, 137 and 153 in the first, second, third, fourth and fifth cross validation respectively.

4. Results

4.1. The variance components

The variance components explained by the traditional BLUP, GBLUP (SPARSE8), GBLUP (DENSE8) and SPARSE GENOME for both traits are presented in table 1 below. The correlation between the predicted values of the data (Yhat) of the model GBLUP (SPARSE8) and GBLUP (DENSE8) for the trait disease resistance to Viral nervous necrosis was high (figure 1).

4.1.1 Viral nervous necrosis

For the trait disease resistance to viral nervous necrosis, in the traditional BLUP model the pedigree information variance equal to 0.084 and the error variance was 0.15 (Table 1). When the data was analyzed using GBLUP(SPARSE8) or GBLUP(DENSE8) model the variance explained by SPARSE8 was a bit less than that of DENSE8 (Table1). Heritability estimates of the trait using the traditional BLUP and GBLUP (SPARSE8) were almost equal. However, GBLUP (DENSE8) and SPARSE GENOME had the lowest and the highest heritability estimates respectively (Table 1).

When the disease challenge test results were analyzed using equation 7 by incorporating both SPARSE8 and SPARSE GENOME genomic relation matrices, 0.021, 0.047 and 0.017 of the variation were explained by SPARSE GENOME, PEDIGREE and SPARSE8 respectively and the error variance was 0.150. So, SPARSE8 explained 20% of the total genetic variance. The log-likelihood was 190 and not significantly better than same model without SPARSE8.

4.1.2. Vibriosis

The analysis of the disease challenge test result for the trait disease resistance to vibriosis is also presented in table 1.Using our data we were not able to distinguish between the genetic variation explained by the genomic information and the pedigree information. In the GBLUP (SPARSE8) and GBLUP (DENSE8) models the variance explained were estimated by fixing the variance explained by SPARSE8 and DENSE8 at the boundary of 0 respectively, however, in case of the SPARSE GENOME model, the PEDIGREE information was fixed at the boundary of 0. Traditional BLUP, GBLUP (SPARSE8) and GBLUP (DENSE8) models gave similar heritability estimate but SPARSE GENOME heritability estimate was the highest. So, the heritability estimate was highest when the variance explained by the PEDIGREE information was fixed at the boundary of 0. (Table 1)

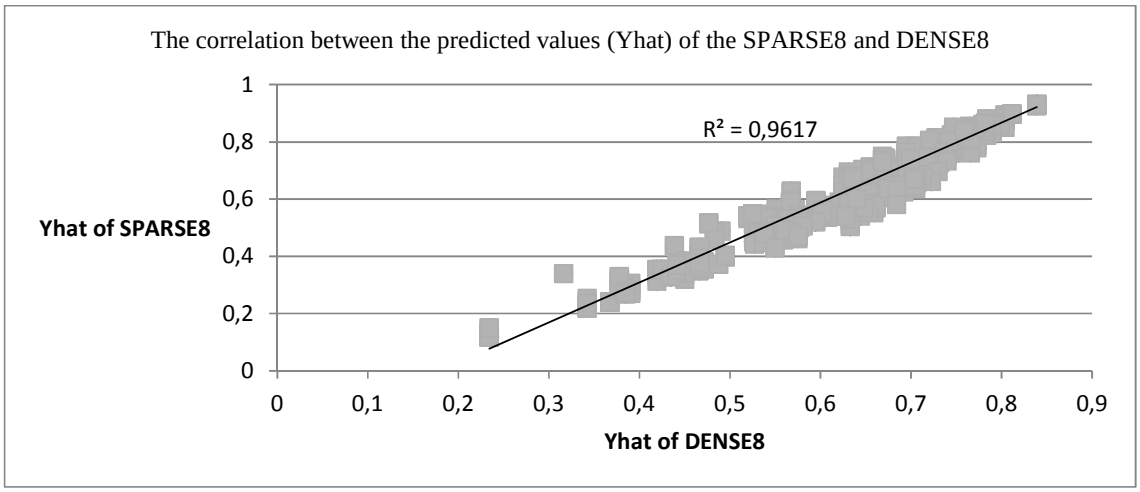


Figure 1. The correlation between the predicted values of the data (Yhat) of the model with SPARSE8 and DENSE8 for the trait viral nervous necrosis

4.2. Evaluation of accuracies of prediction

The correlation between the predicted phenotypic values and observed phenotypic values of the trait disease resistance to viral nervous necrosis for the different models are presented in table 2 and 3. The correlation results were scaled by dividing by the square root of the heritability (h^2).

4.2.1. Within-family cross-validation test

Within–family cross validation test results for the trait disease resistance to viral nervous necrosis are presented in table 2. The traditional BLUP and GBLUP (DENSE8) models gave the same correlation (0.329), however the GBLUP (SPARSE8) gave higher correlation (0.334). The GENOME SPARSE model further improved the correlation (0.336). When the correlation was divided by h, GENOME SPARSE and SPARSE8 had the highest accuracy (0.56) (Table 2).

4.2.2. Between-family cross-validation test

The between-families correlations of the predicted phenotypic values and observed phenotypic values for the trait disease resistance to viral nervous necrosis are presented in table 3. The GBLUP (SPARSE8) had highest correlation (0.16) but the traditional BLUP had the least correlation (0.148). The comparison of within- and between- family accuracies of prediction for the trait disease resistance to viral nervous necrosis is presented in figure 2. In both case the DENSE8 had less accuracy of prediction than the SPARSE8.

Table 1. The summary of the variance components of ASReml results for Viral nervous necrosis and Vibriosis

	Traits							
	Viral nervous necrosis				**Vibriosis**			
Source of variance	Traditional BLUP	GBLUP (SPARSE8)	GBLUP (DENSE8)	GBLUP SPARSE GENOME	Traditional BLUP	GBLUP (SPARSE8)	GBLUP (DENSE8)	GBLUP SPARSE GENOME
PEDIGREE[1]	0.084	0.075	0.06	0.066	0.009	0.009	0.009	0[B]
SPARSE8[1]	N	0.008	N	N	N	0[B]	N	
DENSE8[1]	N	N	0.018	N	N	N	0[B]	N
SPARSE GENOME[1]	N	N	N	0.02	N	N	N	0.019
Error	0.15	0.151	0.155	0.146	0.152	0.152	0.152	0.144
LogL.	189	190	189	190	304	304	304	302
h^2	0.359	0.355	0.335	0.371	0.056	0.056	0.056	0.117

N = not part of the model 1= genetic variance B= fixed at a boundary

Table 2. The accuracies of prediction for the trait disease resistance to viral nervous necrosis for the within-family cross-validation test

Models	**h^2**	**Correlation**	**Accuracy**
Traditional BLUP	0.359	0.329	0.55
GBLUP (SPARSE8)	0.355	0.334	0.56
GBLUP (DENSE8)	0.335	0.329	0.55
GBLUP (SPARSE GENOME)	0.371	0.336	0.56

Table 3. The accuracies of prediction for the trait disease resistance to viral nervous necrosis for the between-family cross-validation test

Models	**h^2**	**Correlation**	**Accuracy**
Traditional BLUP	0.359	0.148	0.25
GBLUP (SPARSE8)	0.355	0.16	0.27
GBLUP (DENSE8)	0.335	0.15	0.25

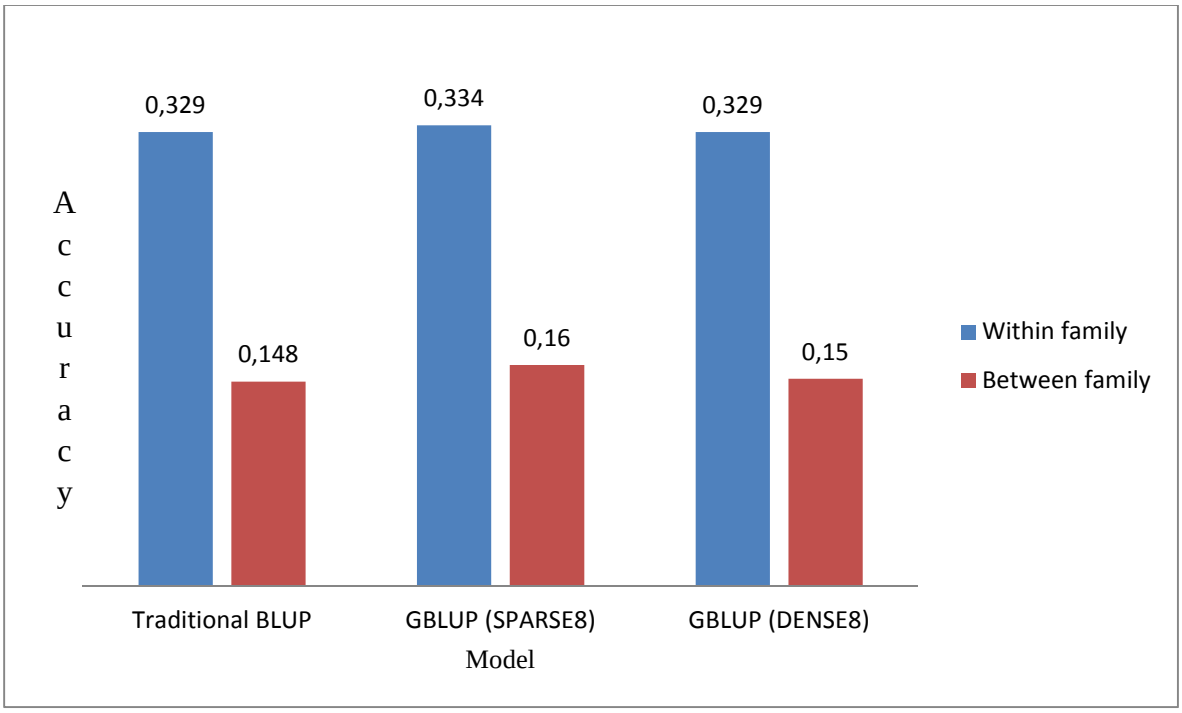


Figure 2. The comparison of within- and between- family accuracies of prediction for the trait disease resistance to viral nervous necrosis

5. Discussion

The main objective of this study was to evaluate the accuracy of genomic prediction to disease resistance for viral nervous necrosis and vibriosis using sparse and genome sequence SNP-data in Atlantic cod. Several computer simulation studies showed that genomic selection can be used in aquaculture to use both within- and between-family variation during selection for both continuous and dichotomous (disease resistance) traits to increase the accuracy of selection (Nielsen *et al.*, 2009, Sonesson and Meuwissen, 2009). However, our results did not show the significant role of genomic information in both traits.

5.1. Viral nervous necrosis

The heritability determines the strength of the relationship between phenotypes and genotypes is one of the most important factors that affects accuracy of selection and the success of genetic improvement of traits (Hedrick, 2011) . In this study the heritability of resistance to viral nervous necrosis in Atlantic cod had a slightly higher estimate in the case of the GBLUP (SPARSE GENOME) model, which was 0.371, than the traditional BLUP model, which was 0.359 (table 1). But, for both models heritability estimates from our study to the trait were lower than the extremely high estimate of other studies in Atlantic cod. Using the 2007 year-class disease challenge test data of the National Atlantic cod breeding program, Ødegård *et al.* (2010) estimated the heritability of survival at the end of test to viral nervous necrosis, and the estimates were 0.75 and 0.43 using threshold mixed model and linear mixed model respectively. Another exceptionally high heritability (0.68) for viral nervous necrosis has been reported by Bangera *et al*. (2011) using 6185 individuals of the 2007 and 2009 year-classes disease challenge test data of the National Atlantic cod breeding program. In aquaculture, high heritability estimates to disease resistance have been reported in different studies, and among these Park and Casella (2008) reported heritability for resistance to three different diseases ranged between 0.42 and 0.57 in rainbow trout, and Hedrick (2011) also reported a high heritability of 0.26 on the observed scale and 0.55 on the underlying scale for infectious pancreatic necrosis (IPN) in Atlantic salmon. The total number of fish, the average number of fish per family, and the model we used in our study could be possible reasons for our relatively lower estimate of heritability for disease resistance to viral nervous necrosis.

In aquaculture, computer simulation studies showed high accuracy of selection based on the genomic information (Nielsen et al., 2009, Sonesson and Meuwissen, 2009). In addition to this, Lillehammer *et al.* (2013) also showed using a simulation study that a combination of genomic within-family breeding values based on low–density genotype and conventional BLUP family breeding values can be a possible low marker density implementation of genomic selection without compromising the effect of genomic selection on genetic gain for species with large full-sib families. However in our study, based on within-family cross-validation tests, the accuracy of prediction of the genomic estimated breeding values using the sparse SNP markers (SPARSE GENOME) did not show a big difference compared with the traditional estimated breeding values. The accuracies of prediction were estimated as 0.336 and 0.329 in the SPARSE GENOME and traditional BLUP respectively (Table 2). A similar result was also reported by Mulder *et al.* (2012) in sheep for the trait concentration of omega 3 fatty acid compounds, and they found that the accuracy of the genomic estimated breeding values were very close to that of traditional estimated breeding values. Despite of a few reports that did not show higher accuracy of genomic estimated breeding values than the traditional estimate breeding values, the accuracy of the genomic estimated breeding value that can be achieved has been evaluated in different livestock species for a range of traits (Meuwissen *et al.*, 2013), and genomic selection could overcome the shortcoming of the traditional selection method, whilst predicting the breeding values more accurately (Zhang *et al*., 2011, Meuwissen *et al.*, 2001). In genomic estimated breeding values, increases in reliability varying between 20% and 29% were reported over those of the traditional selection for milk-yield traits in dairy cattle (Nielsen et al., 2009). In beef cattle the accuracy of genomic selection for growth and carcass traits were evaluated using cross-validation, and up to 0.42 and 0.65 accuracies were reported for growth and carcass traits respectively (Nath *et al.*, 2004). In pigs, for the trait feed conversion ratio, Meuwissen *et al.* (2001) reported that the genomic information gave more accurate genomic estimated breeding values than pedigree only estimated breeding values. In our study, the loss of the contribution of the genomic information to improve the accuracy of prediction over the traditional method could be due to the fact that phenotypical and genotypical data we used for training was too small to accurately capture the whole fraction of the variance explained by the SNP chip. SPARSE GENOME and SPARSE8 had almost similar accuracy of prediction (Table 2) that means the

SPARSE GENOME was not better than SPARSE8 and this could be due to within-family effects are hard to predict because of small family size (approximately 10).

Using within-family cross-validation tests, we checked the accuracy of prediction of imputed high density genotypes (DENSE8) of chromosome 8 for disease resistance for viral nervous necrosis, but its accuracy of prediction was a bit less than that of SPARSE 8 (Figure 2). Generally, high density markers give more accurate prediction than low density markers or at least equal prediction (Solberg *et al.*, 2008, Nielsen *et al.*, 2009). This could be in within- family genomic selection, big segments are inherited together and the sparse SNPs could be sufficient to detect the chromosome segments. Further, we checked the accuracy of prediction of DENSE8 using between-family cross-validation tests, since higher density markers are required for accurate prediction in the case of between-families, than for within-families (Meuwissen, 2009). But, in our result the DENSE8 resulted in lower accuracy of prediction than the SPARSE8. The lower accuracy of prediction of the DENSE8 than the SPARSE8 could be due to imputation errors compensating for the effect of a higher density. The SNP chip explain 24% of the variance from the total genetic variance (Table 1) and chromosome 8 (SPARSE8) explained 20% of genetic variance from the total genetic variance (result of equation 7). But, we did not get a similar finding using chromosome 8 (SPARSE8) and SPARSE GENOME for the trait disease resistance to vibrosis. The possible explanation for this could be due to differences in the genetic architecture and heritability of the traits.

5.2. Vibriosis

In this study using the traditional BLUP model, the heritability of disease resistance for vibriosis (which was measured as survival at a fixed point in time) was estimated to be low (0.059). Even though we were not able to distinguish the variances explained by the genomic information and pedigree information alone, the heritability estimate using GBLUP (SPARSE GENOME) was low but higher (0.117) than the traditional BLUP model. These low heritability estimates to disease resistance to vibriosis are in agreement with the results reported from other studies. Kettunen *et al.* (2007) reported low heritability estimates to disease resistance (measured as time at death) for vibriosis, which ranged from 0.08 to 0.17, depending upon the method used. Similar low heritability (0.16) to disease resistance (measured as survival at the end of test) to vibriosis were reported by Bangera *et al.* (2011). Another low heritability estimate (0.13) was reported for

disease caused by the bacteria Vibrio in other fish species (cold water vibriosis in Atlantic salmon) (Gjedrem and Gjøen, 1995). Moreover, very low heritability estimates of mortality and time of death for vibriosis in Chinook salmon was reported by Beacham and Evelyn (1992). It appears that the finding of low heritabilities for disease resistance for vibrosis is consistent across all studies.

The low heritability estimate of disease resistance for vibriosis (which was measured as survival at a fixed point in time) in our study as well as others studies may not reflect the true additive genetic variation inherent in a population but rather a deficiency in the philosophy underpinning the models that are currently fitted (Lipschutz-Powell *et al*., 2012). Genetic analyses of resistance to infectious disease from disease data, which focus on individual's susceptibility to infectious disease by ignoring the effect of its group members to the disease status of the individual tends to underestimate heritabilities estimate (Nath *et al*., 2004). Nath and his colleagues identified the impact that individuals have on each other as critical parameters for the risk and severity of infectious disease, and by taking into account the variation an individual's impact on its group mates to severity of infectious disease; they improved the low heritabilities estimate (Nath *et al.*, 2004). Moreover, the indirect genetic effects (IGE) model which considers the indirect genetic effect captures the inherent genetic variation an individual's impact on its group mates to severity of infectious disease (Lipschutz-Powell *et al.*, 2012). So, the low heritability estimate of disease resistance to vibriosis can be improved using the indirect genetic effects model and genomic information of large phenotypes and genotypes data (Ibañez-Escriche and Gonzalez-Recio, 2011, Lipschutz-Powell *et al.*, 2012).

A large number of phenotypes and genotypes are required for lower heritability traits than for higher heritability traits in order to get acceptable genomic accuracy (Ibañez-Escriche and Gonzalez-Recio, 2011). Due to this fact (and may be also due to the quality of the SNP chip), the genomic information (SPARSE GENOME) did not capture considerable amounts of the genetic variance. Since, we were not able to distinguish the genetic variance explained by the genomic and pedigree information, the accuracy of genomic prediction could not be assessed by cross-validation.

6. Conclusions

In aquaculture industry genomic selection can be used to increase the accuracy of selection and to solve the lack of pedigree recording. Moreover, it has also a significant role to control inbreeding (Nielsen *et al.*, 2009, Ibañez-Escriche and Gonzalez-Recio, 2011, Sonesson and Meuwissen, 2009). However, in our study, the accuracy of prediction using the genomic information (sparse markers) for disease resistance did not show a big difference compared to using only pedigree information for any of the traits and the traits showed different genetic architecture. In conclusion for both traits more phenotypic and genotypic data are required in order to properly evaluate the accuracy of prediction of the genomic information. But, disease resistance to vibriosis needs more phenotypic and genotypic data than viral nervous necrosis and this is due to the low heritability estimate of vibriosis (Ibañez-Escriche and Gonzalez-Recio, 2011).

7. References

ACTIS, L., TOLMASKY, M., CROSA, J., WOO, P. & BRUNO, D. 2011. Vibriosis. *Fish diseases and disorders. Volume 3: viral, bacterial and fungal infections***,** 570-605.

BANGERA, R., ØDEGÅRD, J., PRÆBEL, A. K., MORTENSEN, A. & NIELSEN, H. M. 2011. Genetic correlations between growth rate and resistance to vibriosis and viral nervous necrosis in Atlantic cod (Gadus morhua L.). *Aquaculture,* 317**,** 67-73.

BEACHAM, T. & EVELYN, T. 1992. Population and genetic variation in resistance of Chinook salmon to vibriosis, furunculosis, and bacterial kidney disease. *Journal of Aquatic Animal Health,* 4**,** 153-167.

BREUIL, G. & ROMESTAND, B. 1999. A rapid ELISA method for detecting specific antibody level against nodavirus in the serum of the sea bass, Dicentrarchus labrax (L.): application to the screening of spawners in a sea bass hatchery. *Journal of fish diseases,* 22**,** 45-52.

BULLER, N. B. 2004. *Bacteria from fish and other aquatic animals: a practical identification manual*, CABI.

BULLOCK, G. 1977. Vibriosis in fish.

CHEVASSUS, B. & DORSON, M. 1990. Genetics of resistance to disease in fishes. *Aquaculture,* 85**,** 83-107.

CHI, S., WU, Y. & CHENG, T. 2005. Persistent infection of betanodavirus in a novel cell line derived from the brain tissue of barramundi Lates calcarifer. *Diseases of aquatic organisms,* 65**,** 91-98.

DAETWYLER, H. D., PONG-WONG, R., VILLANUEVA, B. & WOOLLIAMS, J. A. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics,* 185**,** 1021-1031.

DAETWYLER, H. D., VILLANUEVA, B. & WOOLLIAMS, J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One,* 3**,** e3395.

DALLA VALLE, L., ZANELLA, L., PATARNELLO, P., PAOLUCCI, L., BELVEDERE, P. & COLOMBO, L. 2000. Development of a sensitive diagnostic assay for fish nervous necrosis virus based on RT-PCR plus nested PCR. *Journal of Fish Diseases,* 23**,** 321-327.

DE LOS CAMPOS, G., HICKEY, J. M., PONG-WONG, R., DAETWYLER, H. D. & CALUS, M. P. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics,* 193**,** 327-345.

EGGEN, A. 2012. The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers,* 2**,** 10-15.

FJALESTAD, K. T., GJEDREM, T. & GJERDE, B. 1993. Genetic improvement of disease resistance in fish: an overview. *Aquaculture,* 111**,** 65-74.

FRASER, C. M. 1986. *The Merck veterinary manual*, Merck & Co.

GILMOUR, A. R., GOGEL, B., CULLIS, B. & THOMPSON, R. 2009. ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK.*

GJEDREM, T. & GJØEN, H. 1995. Genetic variation in susceptibility of Atlantic salmon, Salmo salar L., to furunculosis, BKD and cold water vibriosis. *Aquaculture Research,* 26**,** 129-134.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136**,** 245-257.

GODDARD, M. & HAYES, B. Genomic selection based on dense genotypes inferred from sparse genotypes. Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics, 2009a. 26-29.

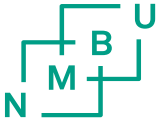
GODDARD, M. E. & HAYES, B. J. 2009b. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics,* 10**,** 381-391.

GRATACAP, R. M. 2008. Characterisation of Vibrio anguillarum for the development of vaccine in cod (Gadus morhua).

HAYES, B. J., LEWIN, H. A. & GODDARD, M. E. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics,* 29**,** 206-214.

HEDRICK, P. 2011. *Genetics of populations*, Jones & Bartlett Learning.
HELLBERG, H., KVELLESTAD, A., DANNEVIG, B., BORNØ, G., MODAHL, I., HALDORSEN, R., VIK-MO, F., OTTESEN, K., SAETRE, E. & SINDRE, H. 2010. Outbreaks of viral nervous necrosis in juvenile and adult farmed Atlantic cod, Gadus morhua L., in Norway. *Journal of fish diseases,* 33**,** 75-81.
HENDERSON, C. 1975. Rapid method for computing the inverse of a relationship matrix. *Journal of Dairy Science,* 58**,** 1727-1730.
HOUSTON, R. D., TAGGART, J. B., CÉZARD, T., BEKAERT, M., LOWE, N. R., DOWNING, A., TALBOT, R., BISHOP, S. C., ARCHIBALD, A. L. & BRON, J. E. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). *BMC genomics,* 15**,** 90.
IBAÑEZ-ESCRICHE, N. & GONZALEZ-RECIO, O. 2011. Review. Promises, pitfalls and challenges of genomic selection in breeding programs. *Spanish Journal of Agricultural Research,* 9**,** 404-413.
JOHANSEN, L.-H., JENSEN, I., MIKKELSEN, H., BJØRN, P.-A., JANSEN, P. & BERGH, Ø. 2011. Disease interaction and pathogens exchange between wild and farmed fish populations with special reference to Norway. *Aquaculture,* 315**,** 167-186.
KETTUNEN, A., SERENIUS, T. & FJALESTAD, K. 2007. Three statistical approaches for genetic analysis of disease resistance to vibriosis in Atlantic cod (Gadus morhua L.). *Journal of animal science,* 85**,** 305-313.
KORSNES, K. 2008. *Nervous necrosis virus (NNV) in farmed Norwegian fish species.* The University of Bergen.
LILLEHAMMER, M., MEUWISSEN, T. H. & SONESSON, A. K. 2013. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genetics Selection Evolution,* 45**,** 39.
LIPSCHUTZ-POWELL, D., WOOLLIAMS, J. A., BIJMA, P. & DOESCHL-WILSON, A. B. 2012. Indirect genetic effects and the spread of infectious disease: are we capturing the full heritable variation underlying disease prevalence? *PloS one,* 7**,** e39551.
MEUWISSEN, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics,* 124**,** 321-322.
MEUWISSEN, T. 2009. Accuracy of breeding values of 'unrelated'individuals predicted by dense SNP genotyping. *Genet Sel Evol,* 41**,** 35.
MEUWISSEN, T., HAYES, B. & GODDARD, M. 2013. Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.,* 1**,** 221-237.
MEUWISSEN, T. H. E., HAYES, B. & GODDARD, M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157**,** 1819-1829.
MULDER, H., CALUS, M., DRUET, T. & SCHROOTEN, C. 2012. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *Journal of dairy science,* 95**,** 876-889.
MUROGA, K. 1994. Polymerase chain reaction (PCR) amplification of RNA of striped jack nervous necrosis virus (SJNNV). *Diseases of Aquatic Organisms,* 18**,** 103-107.
NAKAI, T., SUGAYA, T., NISHIOKA, T., MUSHIAKE, K. & YAMASHITA, H. 2009. Current knowledge on viral nervous necrosis (VNN) and its causative betanodaviruses.
NATH, M., WOOLLIAMS, J. & BISHOP, S. 2004. Identifying critical parameters in the dynamics and control of microparasite infection using a stochastic epidemiological model. *Journal of animal science,* 82**,** 384-396.
NIELSEN, H. M., SONESSON, A. K., YAZDI, H. & MEUWISSEN, T. H. 2009. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture,* 289**,** 259-264.
NIREA, K. G., SONESSON, A. K., WOOLLIAMS, J. A. & MEUWISSEN, T. H. 2012. Strategies for implementing genomic selection in family-based aquaculture breeding schemes: double haploid sib test populations. *Genetics Selection Evolution,* 44**,** 1-9.

NOORDHUIZEN, J. P. T. M., FRANKENA, K., THRUSFIELD, M. & GRAAT, E. 2001. *Application of quantitative methods in veterinary epidemiology*, Wageningen Pers.
ØDEGÅRD, J., BARANSKI, M., GJERDE, B. & GJEDREM, T. 2011. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquaculture research,* 42**,** 103-114.
ØDEGÅRD, J., SOMMER, A.-I. & PRÆBEL, A. K. 2010. Heritability of resistance to viral nervous necrosis in Atlantic cod (Gadus morhua L). *Aquaculture,* 300**,** 59-64.
PARK, T. & CASELLA, G. 2008. The bayesian lasso. *Journal of the American Statistical Association,* 103**,** 681-686.
RANSANGAN, J. & MANIN, B. O. 2010. Mass mortality of hatchery-produced larvae of Asian seabass, Lates calcarifer (Bloch), associated with viral nervous necrosis in Sabah, Malaysia. *Veterinary microbiology,* 145**,** 153-157.
SAMUELSEN, O. B., NERLAND, A. H., JØRGENSEN, T., SCHRØDER, M. B., SVÅSAND, T. & BERGH, Ø. 2006. Viral and bacterial diseases of Atlantic cod Gadus morhua, their prophylaxis and treatment: a review.
SOLBERG, T., SONESSON, A. & WOOLLIAMS, J. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science,* 86**,** 2447-2454.
SONESSON, A. K. & MEUWISSEN, T. 2009. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol,* 41**,** 37.
THRUSFIELD, M. 2013. *Veterinary epidemiology*, John Wiley & Sons.
VAN MARLE-KÖSTER, E., VISSER, C. & BERRY, D. 2013. A review of genomic selection-Implications for the South African beef and dairy cattle industries. *South African Journal of Animal Science,* 43**,** 1-17.
VANRADEN, P. 2008. Efficient methods to compute genomic predictions. *Journal of dairy science,* 91**,** 4414-4423.
VANRADEN, P. M., O'CONNELL, J. R., WIGGANS, G. R. & WEIGEL, K. A. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol,* 43.
VILLANUEVA, B., FERNÁNDEZ, J., GARCÍA-CORTÉS, L., VARONA, L., DAETWYLER, H. & TORO, M. 2011. Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. *Journal of animal science,* 89**,** 3433-3442.
WOO, P. T., LEATHERLAND, J. F. & BRUNO, D. W. 2011. *Fish diseases and disorders*, CABI.
YANG, S. Y., WU, J. L., TSO, C. H., NGOU, F. H., CHOU, H. Y., NAN, F. H., HORNG, H. E. & LU, M. W. 2012. A novel quantitative immunomagnetic reduction assay for Nervous necrosis virus. *Journal of Veterinary Diagnostic Investigation,* 24**,** 911-917.
ZHANG, Z., ZHANG, Q. & DING, X. 2011. Advances in genomic selection in domestic animals. *Chinese Science Bulletin,* 56**,** 2655-2663.



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no