Norwegian University
of Life Sciences

Master's Thesis 2018 60 ECTS
Department for Animal and Aquacultural Science (IHA)

Genome-wide association study of resistance to cardiomyopathy syndrome (CMS) in Atlantic salmon (*Salmon salar L.*)

Jingwen Ding
Aquaculture

ACKNOWLEDGMENTS

This master thesis was accomplished at Nofima As, with the support of Norwegian University of Life Sciences. Data for analysis was provided by industrial company Marine Harvest As.

This project was challenging, interesting and exciting. I have learnt a lot from undertaking this thesis in the field of animal breeding and genetics. And I would not have accomplished it without the help from these professional, patient and lovely people, I want to present my appreciation to them:

I would first like to thank my main supervisor Bjarne Gjerde (NMBU/Nofima AS). Thank you, Dr. Gjerde, for your patience and encouragement. Also, thanks for answering my email efficiently, reading my draft carefully and giving good suggestions.

I would like to show my deepest gratitude to my co-supervisor Dr. Solomon Antwi Boison. I could not have finished my thesis without you. Thank you for teaching me how to write R and basic knowledge in genetics. I am really appreciated that you answer my simple questions patiently and read my thesis clearly.

I shall extend my thanks to: Dr. Muhammad Luqman Aslam, thank you for sharing your knowledge on bioinformatics. Dr. Hooi Ling Khaw, thank you for encouraging me and giving me confidence when I doubted myself.

My sincere appreciation also goes to NMBU student advisor Stine Telneset, who gave me help when I was frustrated. Also, I want to say thank you to NMBU academic writing center and my classmate Erica Ann Curles for helping me in English writing.

Last but not least, I would like to say thanks to my parents who support me patiently to finish my master degree and always love me!

ABSTRACT

Cardiomyopathy syndrome (CMS) is considered as a serious viral disease in Norwegian farmed Atlantic salmon. Outbreaks of the disease occur without any observable signs and CMS primarily affects adult salmon. In 2011, piscine myocarditis virus (PMCV) was detected as the causative agent of CMS and the attempts to develop effective vaccines have not been successful so far. In this study, we estimated genetic variation for resistance to CMS and performed genome-wide association study (GWAS) to uncover quantitative trait loci (QTL) for resistance to CMS in a Marine Harvest population. A total of 620 commercial Atlantic salmon from a population that were infected by CMS in a field outbreak were genotyped with a 57K single-nucleotide polymorphism (SNP) Affymetrix array and were used in this study. The heritability of resistance to CMS was moderate ($h^2 \sim 0.39$). Two significant QTL regions were identified, one on chromosome 12 and other on chromosome 27 that explained ~12.4% and ~12.7% of the total additive genetic variance, respectively. Candidate genes, which were identified in the QTL region on chromosome 12, showed strong relationship with energy production in the cells (SLC25A26, MAPKAPK3), while some of them were related to immune responses (RAB7A) and the process of heart formation (ADAMTS9). Candidate genes in the region of the QTL on chromosome 27 included immune related genes (PRUNE, MLLT11) and genes (PVRL2L) that are expressed upon virus infection. The results of this study provide useful information for increasing CMS resistance in Atlantic salmon.

Sammendrag

Kardiomyopati syndrom (CMS) er en alvorlig virussykdom i norsk oppdrettslaks. Utbrudd av sykdommen oppstår uten noen observerbare tegn og CMS påvirker primært voksen laks. Siden 2011 er piscint myokarditt virus (PMCV) ansett som årsaken til CMS, men forsøk på å utvikle effektive vaksiner har så langt ikke vært vellykket. I denne studien estimerte vi genetisk variasjon for resistens mot CMS hos laks fra Marine Harvest og utførte en genom-vid assosiasjons studie (GWAS - genome-wide association study) for å påvise og lokalisere kromosom regioner med relative store effekt (QTL) på resistens mot CMS hos laks fra Marine Harvest.

Totalt 620 laks fra en populasjon infisert med CMS ble genotypet med en 57K enkelt-nukleotid polymorfi (SNP - single-nucleotide polymorphism) Affymetrix matrise. Arvbarheten for resistens mot CMS var moderat ($h^2 \sim 0,39$). To signifikante QTL-regioner ble identifisert, en på kromosom 12 og en på kromosom 27 og som forklarte henholdsvis ~ 12,4% og ~ 12,7% av den totale additiv genetiske variasjonen.

Kandidatgener i QTL-regionen på kromosom 12 er funnet å være sterkt koplet til energiproduksjonen i cellene (SLC25A26, MAPKAPK3), andre til immunrespons (RAB7A) og oppbygging av hjertet (ADAMTS9). Kandidatgener i QTL-regionen på kromosom 27 er immunrelaterte gener (PRUNE, MLLT11) og gener som styrer virus reaksjoner (PVRL2L). Resultatene av denne studien gir nyttig kunnskap for å kunne øke laksens resistens mot CMS.

CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Atlantic salmon is one of the largest farmed marine aquaculture species in the world. Global production was about 2.4 million tons in 2018 and Norway had the largest production with about 1.3 million tons. In the Norwegian Atlantic salmon farming industry, salmon louse (*Lepeophtheirus salmonis*) treatment, infectious viral diseases and bacterial diseases are the main causes of fish mortality. Although salmon louse treatment results in major economic losses, viral diseases also induce large mortalities and reduce welfare of the fish. These diseases (pancreas disease (PD), infectious salmon anemia (ISA), infectious pancreatic necrosis (IPN), heart and skeletal muscle inflammation (HSMI), and cardiomyopathy syndrome (CMS)) are the most prevalent viral diseases in Norway. Among these viral diseases, CMS is one of the major reasons for limiting the expansion of the Norwegian aquaculture industry. In 2017, 100 out of 1000 farms along the coast of Norway were infected by CMS and over the last decade, an average of 82 farms per year have been infected (Hjeltnes 2018). CMS has also been reported in wild Atlantic salmon populations and is considered as a potential threat to wild populations (Poppe & Seierstad 2003).

CMS is a serious cardiac disease and was first detected in Norway in the mid-1980s (Ferguson *et al.* 1990). In most cases, the disease occurs suddenly without any warnings. Infected fish are mostly close to harvest weight and therefore, outbreaks of CMS result in large economic losses. Although, the infected fish may sometimes show signs of nonspecific abnormal swimming behavior. Histopathological diagnoses show severe pathological changes and inflammation on the heart (Ferguson *et al.* 1990). The causative etiology of CMS was unknown until 2011, a double-stranded RNA virus which belongs to the family *Totiviridae*, named piscine myocarditis virus (PMCV), was detected in infected CMS salmon (Haugland *et al.* 2011). However, the virus's biophysical characteristics are unclear and therefore biosecurity strategies of controlling CMS are still challenging. In addition, no vaccines have been

successfully developed against CMS. Although some management strategies to reduce the mortalities and outbreaks of CMS have been successful, selective breeding could be an additional strategy to reduce outbreaks by increasing resistance of Atlantic salmon to CMS. In 2015, AquaGen AS reported that there was genetic variation among families to CMS resistance being reported, genetic selection of resistance PMCV individuals could be the next step of controlling CMS (AquaGen AS, 2015).

Selective breeding programs of Atlantic salmon in Norway started in the early 1970s after the first collection of salmon eggs in different Norwegian rivers (Gjedrem *et al.* 1991). Family-based (sib information for several traits) and within-family (individual information on desirable traits) breeding programs are used for several decades, but recently both breeding programs use molecular marker information to speed up the selection of desirable individuals as parents for the next generation. Compared to classical pedigree-based family breeding programs, the use of marker information for selection increases accuracy and accelerates genetic gains.

With the emergence of new sequencing technologies, thousands of single nucleotide polymorphism (SNP) markers can be obtained in high through-put. The availability of the dense SNP information accelerated the mapping of quantitative trait loci (QTLs) of desired trait. In Atlantic salmon, marker information has been used successfully to map QTLs for infectious pancreatic necrosis (IPN) virus (Moen *et al.* 2009), pancreas disease (Gonen *et al.* 2015) and also to estimate breeding values for selection of breeding candidates (Ødegård *et al.* 2014; Tsai *et al.* 2015; Bangera *et al.* 2017; Correa *et al.* 2017).

The objective of this study is to: 1) estimate heritability of resistance to CMS in a population of Atlantic salmon which were infected by a natural CMS outbreak and 2) identify QTLs for resistance to CMS.

LITERATURE REVIEW

Distribution and occurrence of CMS

Cardiomyopathy syndrome (CMS), also known as congestive cardiomyopathy, was first recorded in the southern-western part of Norway in 1985. There were extensive and unheralded fish deaths, attributed to an unknown heart disease (Amin & Trasti 1988). In 1997, 60% of adult Atlantic salmon were dead within 5 weeks, with a symptom of endomyocarditis and pathological lesions in one Scottish fish farm (Rodger & Turnbull 2000). CMS was firstly diagnosed in Scotland and was subsequently detected in Canada (Brocklebank & Raverty 2002) and Ireland (Rodger *et al.* 2014) (Figure 1). In Norway, CMS outbreaks occur along most of the coast. However, middle-Norway (including counties Nord- and Sør-Trøndelag and Møre & Romsdal) remained the dominant region of CMS outbreak in last decade. Northern-Norway (including counties Finnmark, Troms and Nordland) became another CMS outbreak hotspot in 2006, where 38% of all diagnosed cases of CMS occurred in that year (Hjeltnes 2017) (Figure 2 and Figure 3).

CMS is not a notifiable disease, hence not all the CMS cases are registered and diagnosed. In Norway, CMS analyses are based on Norwegian Veterinary Institute (NVI) records. Official CMS statistics in NVI date back to 2004 and have been reported from all salmon producing areas in Norway since then (Figure 4). According to the NVI fish health reports, Atlantic salmon (*Salmo salar L.*) is the main identified reservoir of CMS. However, in 2017, Piscine Myocarditis Virus (PMCV), the etiological agent of CMS (Haugland *et al.* 2011), was detected in corkwing and Ballan wrasse in Ireland (Scholz *et al.* 2018). This indicates CMS also could infect other fish besides Atlantic salmon. On average, CMS primarily affects adult Atlantic salmon which are close to marketing size. However, Fritsvold reported a successful transmission of CMS in unvaccinated 33-week-old fish (Fritsvold *et al.* 2009), and in another study, CMS was detected in as early as 5 months salmon after being transferred to sea cages. The median time from initial CMS diagnosis to outbreak usually lasts 5 months (Bang *et al.*

2013). Detection CMS cases in one study showed that most cases occurred in autumn and spring (Brun *et al.* 2003).

Figure 1. World map with red ovals which indicate the CMS distribution in the world.

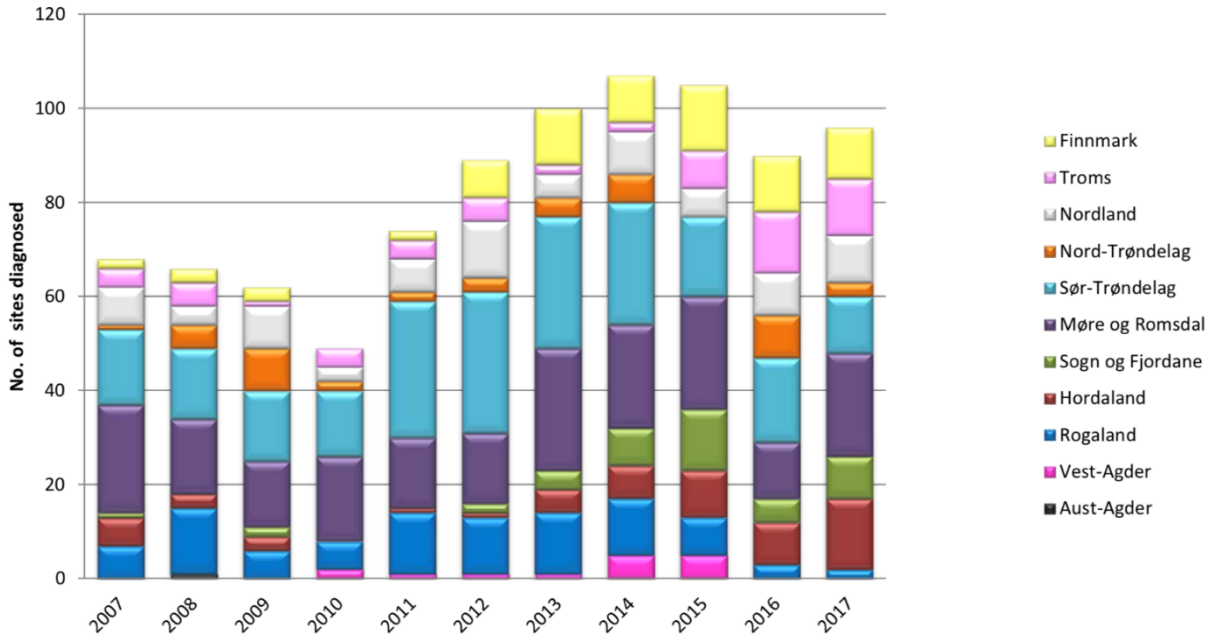


Figure 2. Summary of farms number where CMS had been registered from 2007 – 2017, distributed by year and region. Source: Norwegian Veterinary Institute (https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2018/fish-health-report-2017).

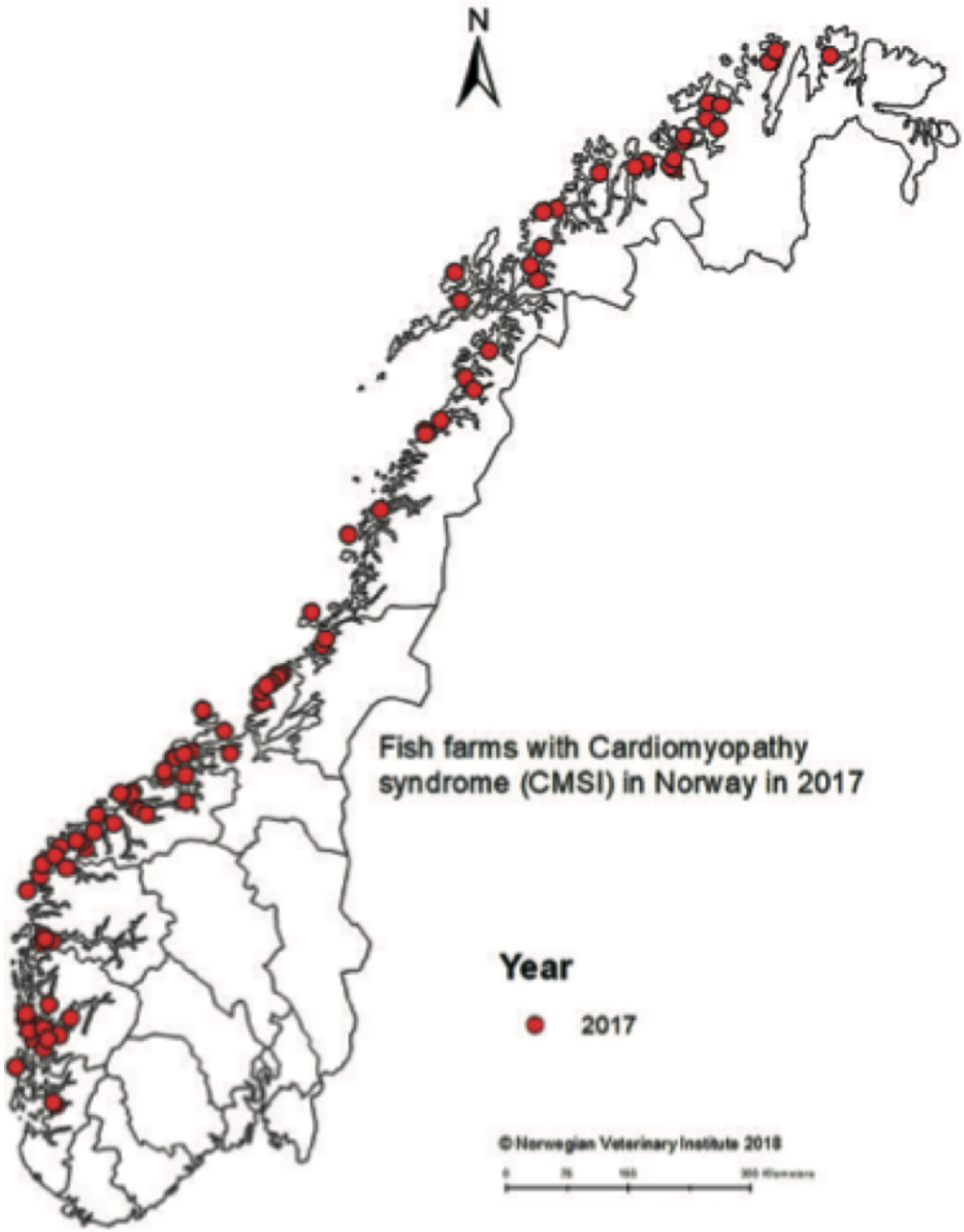


Figure 3. Map shows farms in Norway affected by CMS in 2017. The red dots represent affected fish farms. Source: 2017 Norwegian Veterinary Institute (https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2018/fish-health-report-2017).

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
ISA	12	8	16	11	4	7	17	10	7	1	2	10	10	15	12	14
PD	14	22	43	45	58	98	108	75	88	89	137	99	142	137	138	176
HSMI			54	83	94	162	144	139	131	162	142	134	181	135	101	93
IPN	174	178	172	208	207	165	158	223	198	154	119	56	48	30	27	23
CMS			88	71	80	68	66	62	49	74	89	100	107	105	90	100

Figure 4. Prevalence of various viral diseases in Norwegian farmed salmonids during the period 2001-2017. Source: Norwegian Veterinary Institute (https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2018/fish-health-report-2017).

Piscine myocarditis virus (PMCV)

Finding and the transmission of PMCV

In late 1980's, virus was proposed as the etiology of CMS, because of the occasional intranuclear eosinophilic inclusion bodies presence (Amin & Trasti 1988). Virus became the most debated cause of CMS in next decades. However, there were no viral particles found by transmission electron microscopy (TEM) in the following years (Ferguson *et al.* 1990). The isolation of virus from the heart, spleen, and kidney of diseased fish also could not prove this indication (Rodger & Turnbull 2000). Fortunately, a nodavirus-like agent was diagnosed in one study, which likely was associated with the cardiac lesions of CMS. And the endothelial cells may be the primary target cells of the virus in the heart (Grotmol 1997). In 2003, one study demonstrated that there was a significant correlation between outbreaks of infectious pancreatic necrosis (IPN) and CMS. The risk of experiencing an outbreak of CMS on the farm ranged from 1.6 to 10 times higher when IPN had been diagnosed in the early seawater phase, which indicated that the causative factors of CMS and IPN were similar (Brun *et al.* 2003). An IPN outbreaks may even be included in the causal pathway of the CMS pathogenesis (Brun *et al.* 2003). Another disease, heart and skeletal muscle inflammation (HSMI), was suspected to have an association with CMS, but could not be entirely included (Kongtorp *et al.* 2006). Bacteria or parasites, nutrition, autoimmunity and environment conditions were all considered

as the risk factors of CMS, but no significant correlation between them and CMS has been found (Kongtorp 2005).

In 2009, it was indicated that CMS had an infectious etiology and should be treated as a potentially contagious disease. Since after injection of the tissue homogenate (heart and kidney) from infected CMS salmon which were in an outbreak grow-out farm, histopathological lesions were observed and consistent with CMS in postsmolt (Fritsvold *et al.* 2009). In 2011, a double-stranded RNA virus was detected as the causative agent of CMS. Determining the causative agent was a crucial step toward better disease control and develop vaccines against CMS. In the experimental challenge, real-time reverse transcription (RT)-PCR and in situ hybridization, both methods detected a viral genome in infected salmon. The amount of viral genome was detected at 2 weeks post challenge (p.c.) and peaked 6 weeks post challenge. According to the RNA-dependent RNA polymerase (RdRp) analysis, the virus was clustered into the family of *Totiviridae* and named piscine myocarditis virus (PMCV) (Haugland *et al.* 2011). PMCV is the first virus in the *Totiviridae* family which can infect vertebrate host. Viruses of this family usually asymptomatically and persistently infect fungi and protozoans (Haugland *et al.* 2011).

Haugland also described that PMCV could transmit horizontally through releasing virus particles from infected hosts or cells. In an *in vivo* experiment, they found that PMCV transferred to healthy cohabitating fish from the fish which were injected PMCV intramuscularly (Haugland *et al.* 2011). Infected cohabitating fish shared the similar histopathological changes with the CMS in clinical cases. Mucosal surfaces (gill, stomach, gut, or even skin) were pointed out as the pathway of PMCV spreading to cohabiting fish (Haugland *et al.* 2011). Wiik-Nielsen suggested that vertical transfer was not the primary transmission pathway of PMCV, as they did not detect any virus RNA in both eggs and newly hatched fry after stripping hatchlings, but diagnosed PMCV at early first feeding (Wiik-Nielsen *et al.* 2012).

PMCV genome structure, composition and replication

The buoyant density of the purified virus from infected heart tissue is 1.3842 g/ mL which corresponds well to that of the *Totiviridae* family. The shape of the virus is spherical, approximately 50nm in diameter with surface structural symmetry (Figure 5) (Haugland *et al.* 2011).

PMCV has a double-stranded (ds) RNA genome, and the complete nucleotide sequence is 6,688 base pairs (bps), including three large open reading frames (*ORF 1, 2, 3*) on the positive-sense strand (Figure 6). *ORF1* is located at the nucleotide position from 445 to 3, 027 bps, and it encodes a protein that contains 861 amino acids (aa). The ORF1 protein has a predicted molecular mass of 91.8 kDa and an isolation point (pI) of 5.4. The molecular mass and pI of ORF1 in PMVC is the only part which is similar to the infectious myonecrosis virus (IMNV) in the *Totiviridae* family. This suggests that *ORF1* of PMCV may code for the coat protein, like in IMNV. *ORF2* is at nucleotide positions from 3,114 to 5,291 bps, and the encoded protein includes 726 amino acids with a predicted molecular mass of 83.1 kDa and a pI of 9.42. In a homology search, the predicted amino acid sequence of *ORF2* matches that of the RNA-dependent RNA polymerase (RdRp) in the family *Totiviridae*. PMCV has the closest relationship with the gal-pol fusion protein of *Giardia lamblia virus (GLV)* and *Giardia canis* virus. However, an overlap region between *ORF1* and *ORF2* suggests that *ORF1* and *ORF2* may work together, ORF2 works as a fusion protein with ORF1 due to ribosomal -1 frameshifting. *ORF3* as a unique and separate sequence only exists in PMCV and there is a 250-nucleotide gap region between *ORF2* and *ORF3*. The nucleotide position of *ORF3* starts from 5,542 to 6,447 bps and encodes a 302 amino acid long protein, the predicted molecular mass of which is 33.4 kDa and a pI of 6.99. The function of ORF3 has a weak homology (E value =0.011) with C-X-C motif chemokine 11, the chemokine superfamily motif at the N-terminal end of the *ORF3* gene product. In human, the immunodeficiency virus (HIV) uses chemokine receptors when entering a cell. To what extent PMCV using specific chemokine

receptors to enter the cell is still uncertain. *ORF3* encoded proteins may also have the function of enhancing or modulating the observed inflammation in CMS, which could be very prominent in chronic cases of the disease (Haugland *et al.* 2011).

Most viruses in family *Totiviridae* transmit into new cells during cell division, sporogenesis and cell fusion (Haugland *et al.* 2011). However, GLV could be shed extracellularly and be highly infectious in its purified form (Wang *et al.* 1993). Unfortunately, the PMCV replication method in the cell is still unknown.

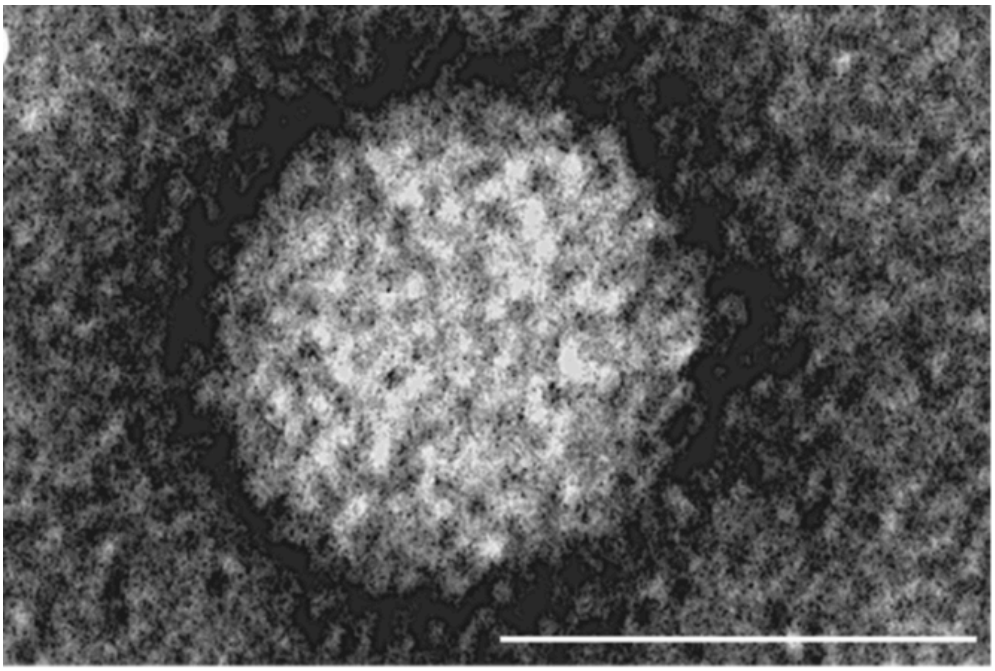


Figure 5. Negative staining of electron micrographs for PMCV. From Haugland et al. 2011.

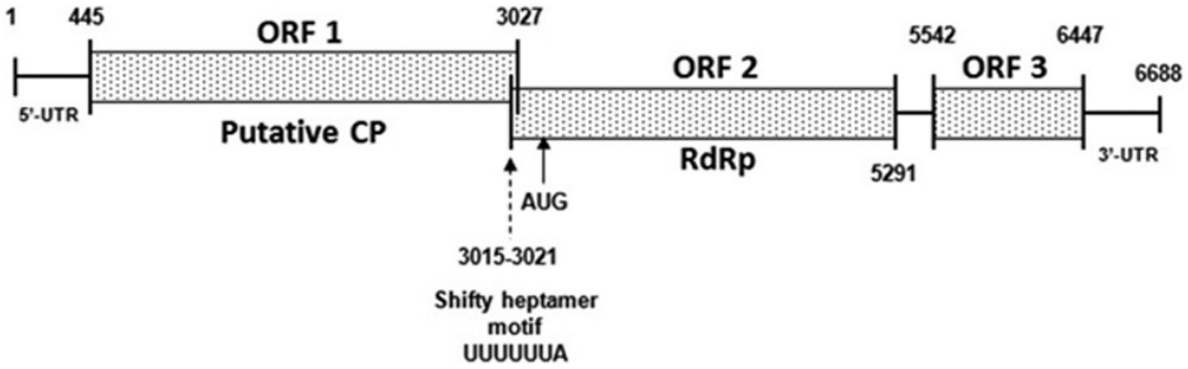


Figure 6. Genome organization of PMCV. The genome includes three large ORFs. From Haugland et al. 2011.

Clinical manifestations and histology of affected organs

CMS is a severe cardiac disease with acute outbreaks or chronic presentation, affecting adult Atlantic salmon primarily during the second year in seawater when fish is close to harvest (Ferguson *et al.* 1990). Acute outbreaks are characterized by relatively high mortality, few observations, fish following the normal weight-gain curve, having normal body condition and eating normally before sudden death. Chronic presentation lasts for a longer period and has moderate mortality (Brun *et al.* 2003).

CMS affects heart, liver, kidney, and spleen. All affected organs have lesions to a certain degree, like alternating dark and light mottled lesions in the liver. This kind of "nutmeg" appearance represents focal degeneration and normal cells, indicating chronic passive congestion. The spleen showed a mild or moderate degree of congestion (Amin & Trasti 1988). Widespread scale-pocket oedema and hemorrhages in the ventral skin are typical external macroscopic findings (Bruno 2013) and fish scales raise because of that (Poppe & Seierstad 2003) (Figure 7). Typical findings when doing necropsy are an enlarged atrium, a hidden ventricle in the pericardial cavity with hemopericardium and/or blood clots, a mottled liver with a fibrinous coat, ascites and general congestion (Bruno 2013) (Figure 8).

Diagnosing CMS should be based on the results of histopathological examination. Macroscopic findings are not specific for CMS symptoms because they also have been detected

in other diseases. Lesions in a fibrinous perihepatitis and enlarged or ruptured atriums with consequent hemopericardium are the most significant histological observations in CMS fish, indicating that the cause of hemopericardium and sudden death are that the wall of the atrium or sinus venosus weakens or ruptures (Ferguson *et al.* 1990). In infected fish, lesions have different stages of reaction and are restricted in the spongy myocardium of atrium and ventricle, while the compact layer of ventricular is intact (Ferguson *et al.* 1990). Lesions start from the outer compact layer of the atrium as a focal pattern, and they spread to the ventricular endocardium spongy layer, developing into multifocal or diffuse and extensive patterns. In the end, with the lysis of contractile fibers, loss of eosinophilia and striation in affected myocardial cells, intact cells are rare in the atrium and spongy layer of the ventricle (Figure 9, Figure 10 and Figure 11) (Timmerhaus 2011a). Inflammatory infiltrates are mainly observed in mononuclear cells, partly in lymphocytes and macrophages (Ferguson *et al.* 1990).

Figure 7. The external macroscopic finding of CMS fish. Picture from Per Anton Sæther, MarinHelse AS (Garseth et al. 2018).

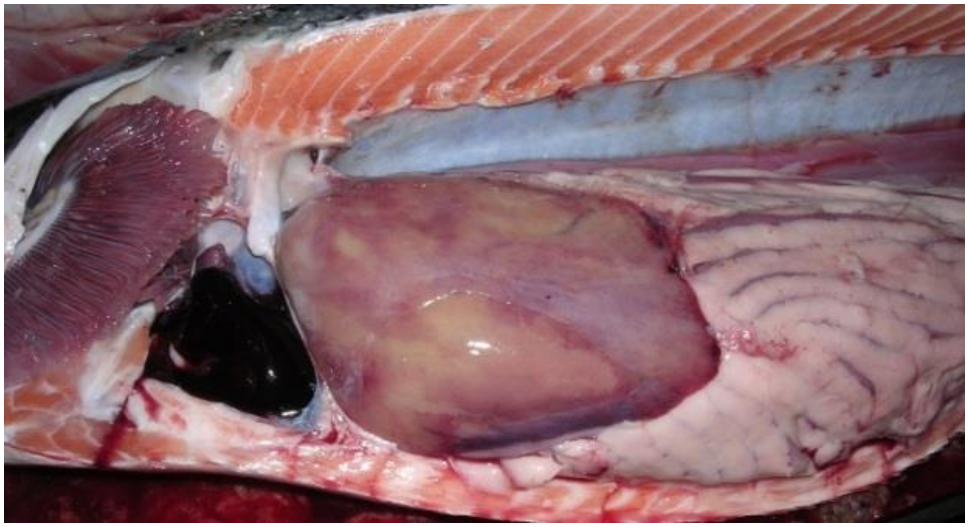


Figure 8. The internal macroscopic finding of CMS fish. Picture form Brit Tørud, Norwegian Veterinary Institute (Garseth et al. 2018).

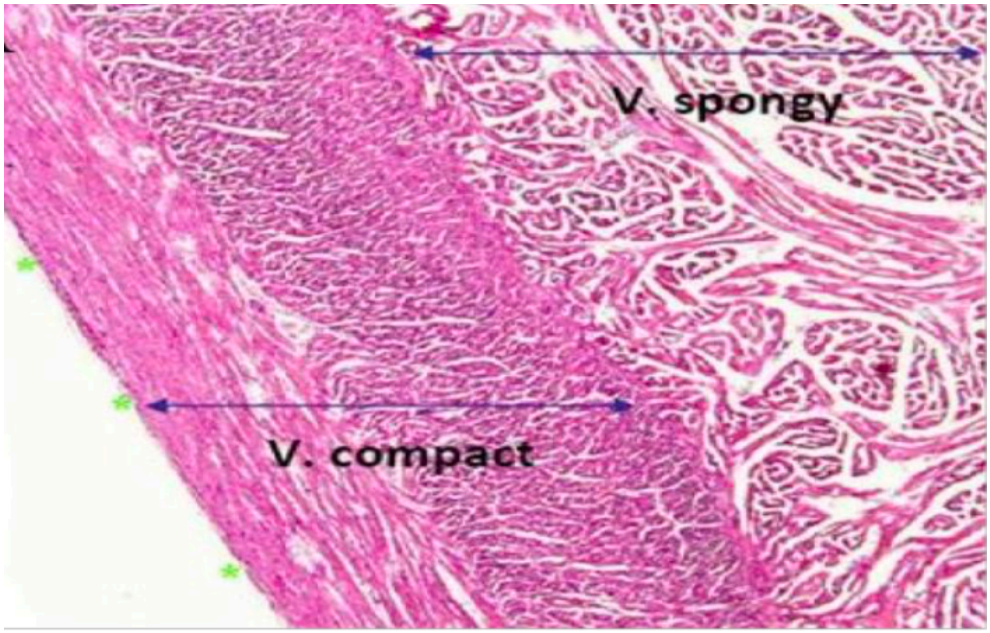


Figure 9. The intact ventricle with three layers, spongy, compact and epicardium (green stars marked) (Timmerhaus 2011a).

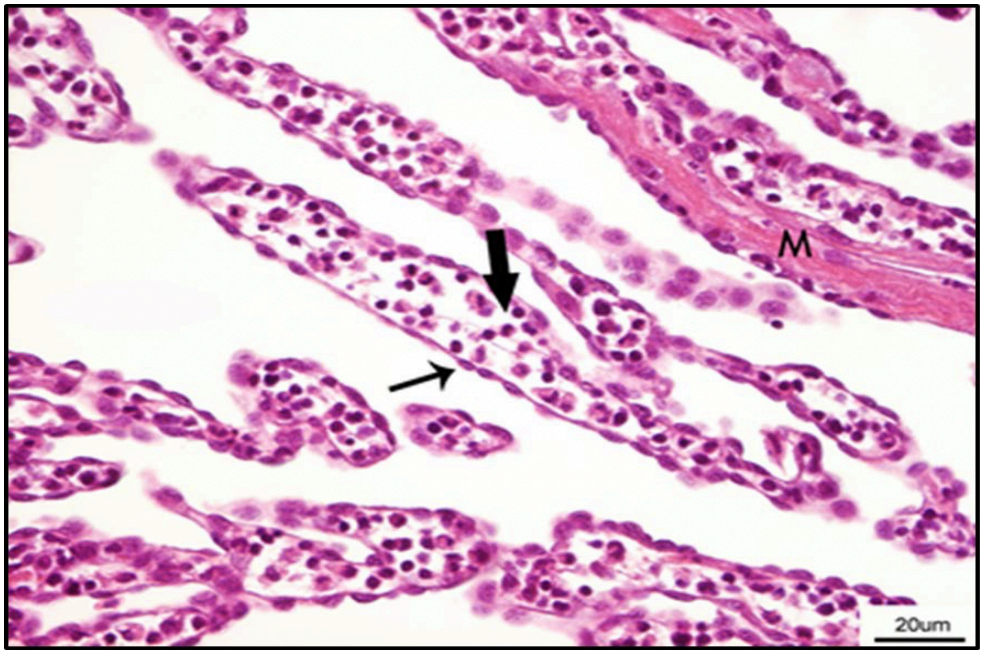


Figure 10. Severe inflammatory sponge layers of atrial, the muscle (M) has almost disappeared and is replaced by inflammatory cells (big arrow), endocardial cells are hypertrophic (small arrow) (Timmerhaus 2011a).

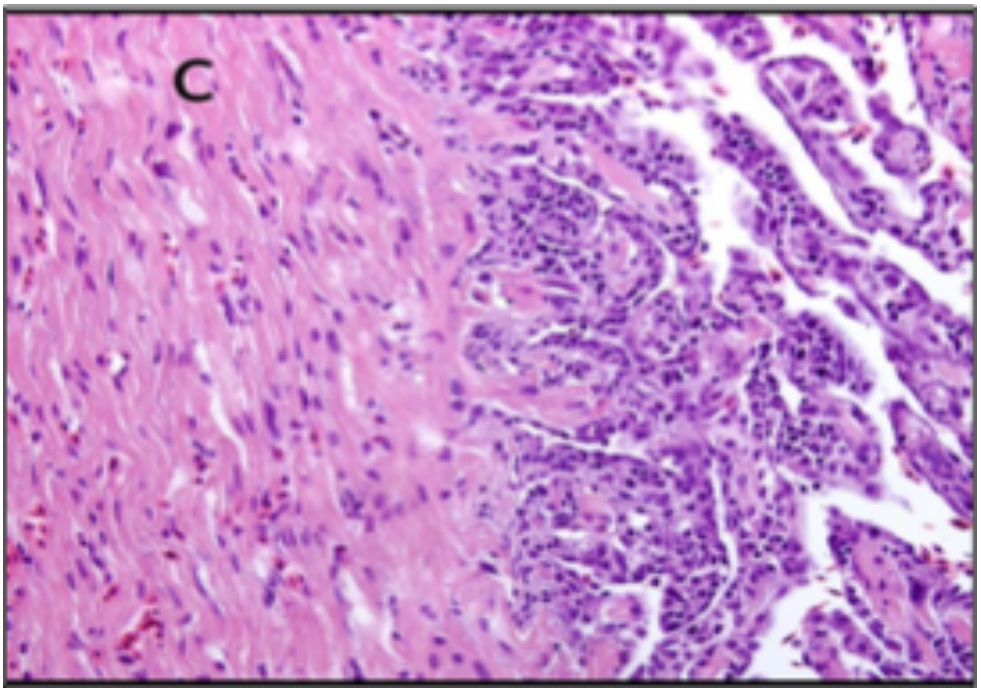


Figure 11. Intact compact muscle layer (left) compared to severe inflammatory sponge layer (Right) of cardiac ventricle with a distinct border, Photo comes from: Trygve Poppe Pharmaq Analytic AS / Torunn Taksdal, Norwegian Veterinary Institute (Timmerhaus 2011a).

Financial losses of CMS

CMS causes huge financial losses in farmed Atlantic salmon as mortality primarily affects adult salmon which are close to harvest size. In addition, CMS outbreaks usually occur suddenly without any signals, therefore, taking precautions is difficult. Once CMS breaks out, preventive prescheduled slaughtering of salmon is the main performance in most farms, which is costly. Brun *et al.* (2003), estimated the total annual direct economic loss (solely on biological loss) in Norway from € 4.6 to 8.3 million when the occurrence of CMS is 14% in the target population, based on data from 2002. The annualized number of CMS cases show a waving trend in the recent 10-year period, but the overall trend of case numbers has increased. The price of salmon has also increased in recent years, and these factors result in higher financial losses. Accordingly, developing an efficient approach to prevent CMS is an urgent task.

Prevention and control strategies

Measures to control and prevent CMS outbreaks will play crucial part in improving fish welfare and decreasing financial losses. Since CMS is a transmissible disease (Haugland *et al.* 2011), the crux of prevention is to cut off the introduction of the virus to the fish farm and reduce risk factors of CMS. Several multifaceted methods have been used, including general management, vaccination, clinical nutrition and selective breeding.

General management

In general, management practices to reduce CMS outbreaks focuses on reducing stress during the production period. Although the fish might be infected with the PMCV virus, large mortalities occur when the fish undergoes stressful activities, therefore, stress should be kept to a minimum until slaughter (Garseth *et al.* 2018). The probability of acquiring CMS is related to the infection pressure, cohort size, the length of time fish stay in the sea and the history of

CMS/HSMI outbreaks in previous cohorts (Bang *et al.* 2013). To decrease pressure, we should prevent the fish density from exceeding the recommended number and choosing proper farm locations where there is no history of the disease. Currently most salmon farms employ the "all in-all out" principle combined with fallowing (Garseth *et al.* 2018). In 2013, stamping-out of infected farms or pre-slaughter of infected fish to reduce overall infection pressure were established as the most efficient methods to prevent the spread of CMS among the farms (Bang *et al.* 2013). Nevertheless, pre-slaughter brings large financial losses.

Vaccination

Vaccination is the one of most efficient and popular methods to prevent Atlantic salmon from bacterial and viral diseases. However, due to the lack of knowledge on the replication processes of PMCV within cells and the exact function of the different genomic regions of PMCV, the development of vaccines against PMCV is hampered but still in progress.

Clinical nutrition

Owing to the lack of efficient commercial vaccines against PMCV, clinical nutrition remains a valuable approach to decrease morbidity from PMCV. Functional feeds, as the main clinical nutrition prevention, contain altered ratios of nutritional compositions or additional new ingredients, those aid in resisting disease and mitigating clinical disease symptoms (Martinez-Rubio *et al.* 2014). In a study, researchers proved that increased dietary levels of EPA (from 3.6% to 14%) and reduced dietary levels of lipid (from 31% to 18%), compared to a reference diet, could enhance the immune response which is embodied in a lower pathology and viral load in heart tissue after an injection of PMCV. Another study concluded that lipid content and composition may have an immunomodulatory effect (Garseth *et al.* 2018). These results are applied to develop commercial functional feeds to prevent heart diseases like CMS and IHSM.

Selective breeding

Selection for increased disease resistance is expected to decrease the risk of infection and increase resistance to diseases. The great advantage of selective breeding programs is that the genetic gain per generation is cumulative, meaning that the resistance to a disease is increased over generations. In Atlantic salmon, several authors have reported that there are genetic variations for resistance to CMS. Vassgård (2017) reported that heritability (liability scale) of CMS resistance in an Atlantic salmon population was 0.25 ± 0.05, Boison *et al.* (2019, under-review) and Hillestad & Moghadam (2018) reported that heritability of CMS resistance was 0.46±0.05 and 0.41 ± 0.05, respectively. These estimates of heritability suggest that there is potential for increasing CMS resistance in Atlantic salmon populations through selective breeding. Both traditional (pedigree-based) and current (genomic selection) approaches of selective breeding can be used to increase resistance to CMS, however, traditional (pedigree-based) genetic selection for increased resistance to CMS does not allow for selection within families (family-based breeding programs), because the trait cannot be recorded on the live breeding candidate. Although family-based breeding programs for disease traits (traits not recorded on live candidate) are less efficient, genetic gain for genetic selection of disease traits (e.g. infectious pancreatic necrosis, vibriosis, pancreas disease, etc.) in Atlantic have been moderate to high per generation. For example, it was shown that, Atlantic salmon with low resistance to infectious pancreatic virus (IPNV) had 79.0% mortality whereas high-resistant salmon had 32.0% mortality (Storset *et al.* 2007).

With the development of genome-wide SNPs and statistical methods that combine phenotypic information and genetic markers information to estimate the genetic merit of breeding candidates, selection within families (even when trait is not recorded on candidate) is possible which leads to higher genetic gain. In Atlantic salmon, selection for disease resistance using phenotypic and genomic information has resulted in genetic improvement of population

to resist several disease traits. Therefore, selective breeding for resistance to CMS will lead to reduction in CMS outbreak.

Quantitative trait loci (QTL) mapping

With the development of technology to detect genetic markers, such as Restriction fragment length polymorphisms (RFLPs), Microsatellites and Single nucleotide polymorphisms (SNPs), selection of breeding candidates for the next generation could be based on marker information. Microsatellites and SNPs are the most described and used genetic markers. Compared to microsatellites, SNP are the most abundant polymorphism in the genome and high-throughput methods have been developed to genotype several individuals at a relatively cheaper cost.

Marker information has been used successfully to map quantitative traits (measurable traits which is characterized by cumulative action of many genes and the environment), but also for monogenic traits (controlled by few genes). Quantitative trait loci or QTLs are loci which are associated with the desirable traits. The assumption of QTL mapping is that marker information (genomic variants) are in linkage disequilibrium (LD) with the causative loci.

Linkage disequilibrium (LD) describes the degree to which one allele of SNP is inherited or correlated with the other allele of another SNP within a population. (Bush & Moore 2012). It is the nonrandom association of alleles at different loci in a given population, and it also describe changes in genetic variation within a population over time (Slatkin 2008). D′ and r^2 are the most common used statistics symbol for describing LD, they are related to the deviations between the observed frequency of co-occurrence for two alleles (i.e. a two-marker haplotype) and the frequency expected if the two markers are independent. D' is a population genetics measurement which is related to recombination events between markers. It is scaled between 0 and 1. r^2 is a statistical measure of correlation between makers. There are dependencies between these two statistics (Bush & Moore 2012).

Identifying causative mutations using marker information have been very successful for selecting some complex traits in aquaculture populations. In several aquaculture populations, QTL mapping has made progress in identifying genetic variants which increase resistance to several diseases, like resistance INP and PD in Atlantic salmon (Moen *et al.* 2009; Gonen *et al.* 2015) and also increases growth related traits in Atlantic salmon (Tsai *et al,* 2015). The uncovering of these genetic markers in QTL mapping studies has allowed for their implementation in breeding programs through marker assisted selection (MAS).

Marker-assisted selection (MAS) and Genome-wide association studies (GWAS)

Marker-assisted selection (MAS) has worked successfully for some quantitative traits (Ribaut & Ragot 2007). The approach of MAS was that used microsatellite and few SNP markers were used to map QTLs for the desirable traits, using linkage mapping techniques (Gutierrez *et al.* 2015). Once a QTL was identified, the favorable individuals could be selected using these few genetic variants (microsatellite or SNPs). In livestock populations, it had been shown that, MAS could increase the rate of genetic response in animal breeding programs by of 5% to 64% (Hayes & Goddard 2003). The main drawback of MAS was that QTLs spanned a large chromosomal region with large confidence intervals because of the small number of genetic markers that were used in the QTL mapping step (Goddard & Hayes 2009). Furthermore, these markers only explained a small proportion of the total genetic variance. The implication of this drawback was that the LD between the QTL and the marker was weak, therefore selection of individual, based on these markers, did not always lead to the desirable gain in genetic response. In addition to that, the cost of genotyping individuals for these markers was prohibitive. Interestingly, MAS for selecting for IPN resistance in Atlantic salmon had been very successful, when a single locus explained a large proportion of the genetic variance (80% - 90%) (Houston *et al.* 2010; Moen *et al.* 2009).

After the completion of the human International HapMap Project in 2001, ability to genotype individuals with dense SNP markers became possible, affordable and could be done in high-throughput. Therefore, large scale QTL mapping with high number of markers which spread across the genome could be undertaken (Genome-wide association study). Genome-wide association study (GWAS) evaluates the association between single-nucleotide polymorphisms (SNPs) and the desired traits, relying on the possible high levels of LD between each SNP and the causative mutation (Gutierrez *et al.* 2015). GWAS has worked successfully to identify QTLs that are used in the selection of individuals with resistance to several diseases. Boison *et al.* (2019, under-review) detected QTLs for resistance to CMS on chromosome 12 and 27, using GWAS method. GWAS is based on the genetic merit of animals, compared to early MAS, it is expected to obtain higher genetic improvement, since higher selection accuracies and selection intensity can be obtained.

However, in GWAS, the detected markers are not usually the causal markers of the desired trait, therefore, subsequent analyses to find or narrow down the region of interest are undertaken (fine mapping techniques). It is important to note that, most economically important traits are polygenic and individual genes usually do not explain all the genetic variation of the trait, therefore, statistical methodologies are developed which use all genome wide markers to estimate the genetic merit of breeding candidates in the population. The method is referred to as genomic selection.

Genomic selection (GS)

Genomic selection as proposed by Meuwissen *et al.* (2001), uses all available dense genome wide SNP markers for prediction of genetic values of breeding candidates. The method assumed several markers contribute to the underlying additive genetic variation. Although the approach is similar to MAS (selection is according to the genotype), GS requires dense, uniformly distributed DNA markers throughout the whole genome, as this leads to QTLs in

high LD with at least one genetic marker and almost all genetic variance could be explained by these genetic markers. The recent (one decade) advancement in whole genome sequencing and SNP genotyping technology has allowed for rapid adoption of genomic selection in several animal breeding programs (Goddard & Hayes 2007). Genomic selection has been used in aquaculture population for selection disease-resistant and growth-related traits. Higher accuracies of selection have been obtained compared to using pedigree (traditional) based selection approach.

MATERIAL AND METHOD

Population

The Atlantic salmon (*Salmo salar*) used in this study was obtained from a commercial grow-out farm, owned by Marine Harvest AS, in the Trøndelag region, Norway. The fish were start-fed in March 2016 and were transferred to sea in October 2016. The PMCV virus outbreak occurred in a single cage with 190,000 fish. The virus was confirmed to be PMCV based on clinical signs (veterinarians checked for signs of oedema and fibrin in the liver and spleen as well as congestion; i.e. ascites, hemorrhage of the atrium and discolored liver with fibrinous casts) and quantitative polymerase chain reaction (qPCR) analysis. The field outbreaks started in January 2017 and lasted for 6 weeks. We collected tissue samples from 320 fish that died during the outbreak period. After the cessation of mortalities, tissue samples of a random 300 survivors sample were collected in April 2017. A binary phenotype was analyzed, with dead fish coded as 0 and PMCV virus free fish coded as 1.

Genotyping batch

Tissue sample for DNA extraction were taken from the adipose fin-clip of each fish. DNA extraction and genotyping of the population were undertaken by IdentiGen, Ireland (https://identigen.com/). A total of 620 samples were genotyped with a non-commercial Affymetrix axiom 57K SNP array (NOFSAL03, 55735 markers) which was developed by Nofima in collaboration with Marine Harvest and SalmoBreed. After Affymetrix quality control, 571 fish (of which 253 survivors and 318 dead) were left for the subsequent analyses.

Quality control analyses

The purpose of genotype quality checks is to remove samples and SNPs which are of poor quality. Quality control analyses were undertaken using PLINKv1.9

(http://zzz.bwh.harvard.edu/plink/). Individuals and SNPs with call rate < 98% were removed, SNPs with a very low minor allele frequency (MAF) of < 1% were discarded (Supplementary figure 1). Samples with too low (< 0.32) or too high (> 0.43) heterozygosity were discarded (Supplementary figure 2). Furthermore, SNPs that deviated too largely (p-values from Fishers exact test < 10^{-10}) from Hardy-Weinberg equilibrium (HWE) were also discarded. When SNPs deviate too largely from HWE, this might be as a result of genotype calling errors. At the end of the genotype quality checks, 53,258 SNPs and 551 (304 dead and 247 alive) samples remained.

Population stratification (structure)

Principal component analysis (PCA) was performed with PLINKv1.9 to assess stratification in our study population. In most cases, eigenvectors from the PCA analysis are used as covariate to account for population structure in genome wide association analysis to help reduce the potential obtaining false positive results (Martin & Eskin 2017). The first five eigenvectors were included as covariates in our association analysis. Firstly, the k-means algorithm (R Core Team (2017)) was used to obtain the number of possible family clusters in our population (Figure 12). Furthermore, to visualize stratification in our population, we plotted the first two eigenvectors and computed the proportion of the total genomic variation that explained by the first two eigenvectors, using their eigenvalues. Three clusters were obtained and were presented in Figure 13.

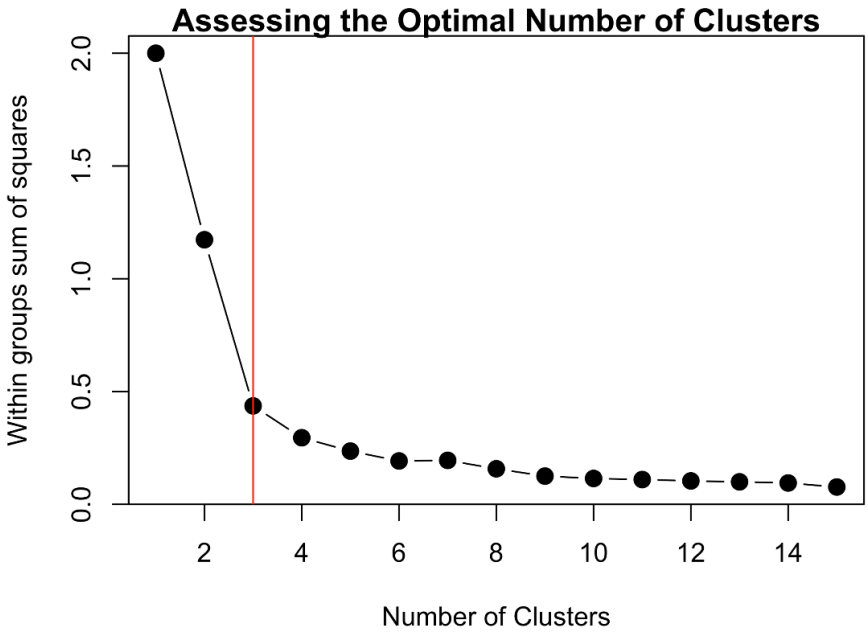


Figure 12. K-Means method of assessing the optimal clusters of families. Turning point indicates the optimal number of clusters. Solid line indicates the turning point.

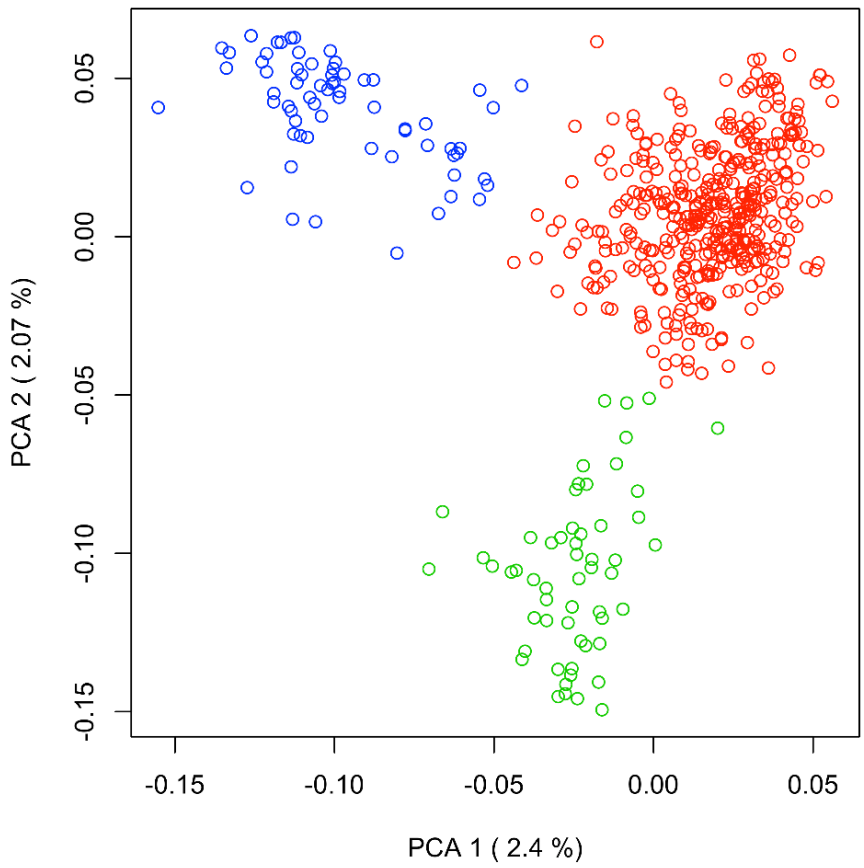


Figure 13. Sample family structure was identified by PCA with the first two principal components. Three different color represented separated family according to k-means algorithm result.

Estimating the variance components

Because these fish were from a commercial population, the pedigree of these fish was unknown. Therefore, variance components estimation was based on the genomic relationships matrix (G) between fish, using marker information. A linear mixed animal model [Model 1] was applied to estimate genetic variance components in the statistical software package GCTA (Yang *et al.* 2011a), using a restricted maximum likelihood (REML) approach. The following model was applied:

$$y = u + \sum_{j=1}^{N=5} pca_j + Zg + e \qquad \text{[Model 1]}$$

where: y is a vector of binary phenotypes (dead = 0, alive = 1), u is the overall mean, pca_j is the first 5 eigenvectors (N = 5), 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, $g \sim N\ (0, G\sigma_g^2)$, where G is the genomic relationship matrix and σ_g^2 is the genetic variance, e is the vector of random residual effects, $e \sim N\ (0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance.

The G matrix used in this study was computed according to the method of VanRaden (VanRaden 2008). Genomic relationships were calculated as:

$$G = \frac{ZZ'}{[2\sum_i^l p_i\,(1 - p_i)]}$$

Where Z = M – P, 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 ith column as $2p_i$, and p_i is the allelic frequency for SNP i. Allele frequencies were estimated from the observed genotypic data.

Heritability on the observed scale (h_{obs}^2) was calculated as:

$$h_{obs}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

We also estimated variance component without fitting the first 5 eigenvectors as covariate in the model. This allowed us to estimate how much of the genetic variance was reduced by accounting for population stratification.

Genome-wide association analysis

A mixed linear model (MLM) analysis was used to estimate marker effect. The analysis was undertaken with GCTA (Yang *et al*., 2011a) using the following model:

$$y = u + \sum_{j=1}^{N=5} pca_j + w_{ij}a_i + Zg + e \qquad \text{[Model 2]}$$

Where y, u, pca_j, Z, g and e have been described earlier. The column vector w_{ij} is the incidence matrix for SNP i containing marker genotypes codes as 0 = *AA*, 1 =*AB*|*BA*, 2= *BB*. The allele substitution effect (a_i) was computed as $a_i = \left(w_{ij}{}'V^{-1}y\right) \times (w_{ij}{}'V^{-1}w_{ij})^{-1}$, where $V = \mathrm{G}\sigma_g^2 + \mathrm{I}\sigma_e^2$.

We used the "--mlma-loco" approach of GCTA to estimate the allele substitution effect. This approach ensured that the allele substitution effect of SNP i was conditioned on a genomic relationship matrix computed from markers of all chromosomes except the chromosome on which the candidate SNP was located (Yang *et al.* 2011a; Yang *et al.* 2014). The results of the genome-wide association analysis were presented in following tables and figures. The P-values from the association analysis were converted to $-\log_{10}P$ and were displayed with a Manhattan plots. Markers were considered significantly associated with resistance to CMS when they exceeded Bonferroni threshold. Bonferroni threshold was computed as 0.05 divided by the total number of SNPs (53,258). The Bonferroni threshold used in this study was $P < 9.38 \times 10^{-7}$ with an equivalent $-\log_{10}P$= 6.02. Quantile-quantile (Q-Q) plots were also used to visualize the deviation of the observed P values from the expected P values (based on the theoretical χ2-distribution (Ehret 2010)). The parameter *Lambda* (λ, an inflation factor), was used to indicate the deviation between observed and expected *p*-values. Lambda was calculated as

$\frac{median\left((\alpha^2/s.e^2)^2\right)}{0.456}$, a was the allele substitution effect and $s.e$ was the standard error of the allele substation effect. If the observed values correspond to the expected values, *Lambda* (λ) should be close to 1.0. According to (Yang *et al.* 2011b), a lambda value below 1.1 is considered acceptable.

Variance explained by SNPs

Variance explained by each significant SNP was calculated from the estimated allele substitution effects (a_i) and the allele frequencies as $\sigma^2_{SNP} = 2p_i(1-p_i)a_i^2$. The proportion of the total genetic variance ($\frac{\sigma^2_{SNP}}{\sigma^2_g}$) and the proportion of phenotypic variance ($\frac{\sigma^2_{SNP}}{\sigma^2_p}$) explained were also computed.

Linkage disequilibrium in the QTL region

We estimated and visualized the linkage disequilibrium (LD) within the QTL regions. Linkage disequilibrium was computed as the correlation between genotypes and SNP in the QTL region by using PLINK 1.9. Heatmaps of the LD in the QTL regions were presented and the heatmaps were made with "LD. plot" function of the "gaston package" of R statistical software (Claire 2018).

Annotation of candidate genes

The genes within the genomic region, harboring the significant SNPs which were associated with CMS resistance, were mapped by GBrowse in SalmoBase (https://salmobase.org). The reference map in SalmonBase is version 2.0 of Atlantic salmon genome assembly (https://www.ncbi.nlm.nih.gov/genome/369?genome_assembly_id=248466). Gene functional

annotation and cluster were undertaken with Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/home.jsp).

RESULTS

Quality check of genomic data

We deleted 611, 614 and 15 markers that did not pass the genotype call rate, minor allele frequency and HWE equilibrium test, respectively. We also removed 10 and 20 samples that failed the sample heterozygosity test and individuals call rate threshold, respectively. At the end of the genotype quality check, 53,258 markers information and 551 samples (304 survivors and 247 dead) remained for further analysis.

Heritability estimates

Table 1 shows the variance components estimated by genomic information with and without accounting for population stratification, through including or excluding 5 eigenvectors as covariates. The two estimated heritability were very similar.

Table 1. Variance components and heritability of CMS.

Model	$\sigma_g^2 \pm$ SE	$\sigma_e^2 \pm$ SE	$\sigma_p^2 \pm$ SE	$h^2_{obs} \pm$ SE
With PCA[1]	0.102 ± 0.024	0.156 ± 0.017	0.258 ± 0.018	0.395 ± 0.077
Without PCA	0.095 ± 0.023	0.159 ± 0.017	0.254 ± 0.017	0.374 ± 0.074

[1] *Accounting for population stratification.*

CMS resistance association analysis

Figure 14 and Figure 15 demonstrate the Manhattan plot and quantile-quantile (Q-Q) plot respectively, according to the association analysis results. One region on chromosome 12 and one region on 27 exceeded the Bonferroni threshold ($-\log_{10}P$= 6.02). In total 8 SNPs (2

SNPs on chromosome 12 and 6 SNPs on chromosome 27) were significant and the most significant marker ($-\log_{10}P$ =7.31) was found on chromosome 27, while the second most significant marker ($-\log_{10}P$ = 6.76) was found on chromosome 12. Because more than one marker was significant in each of these QTL regions, we checked whether there were multiple QTLs in these regions. When the most significant marker on chromosome 12 and 27 were included as covariate, after re-running the association analysis, the QTL peak disappeared, thus suggesting that there was only a single QTL on each of the two chromosomes (Figure 14(B)). An inflation factor (λ) of 1.14 was obtained in this study.

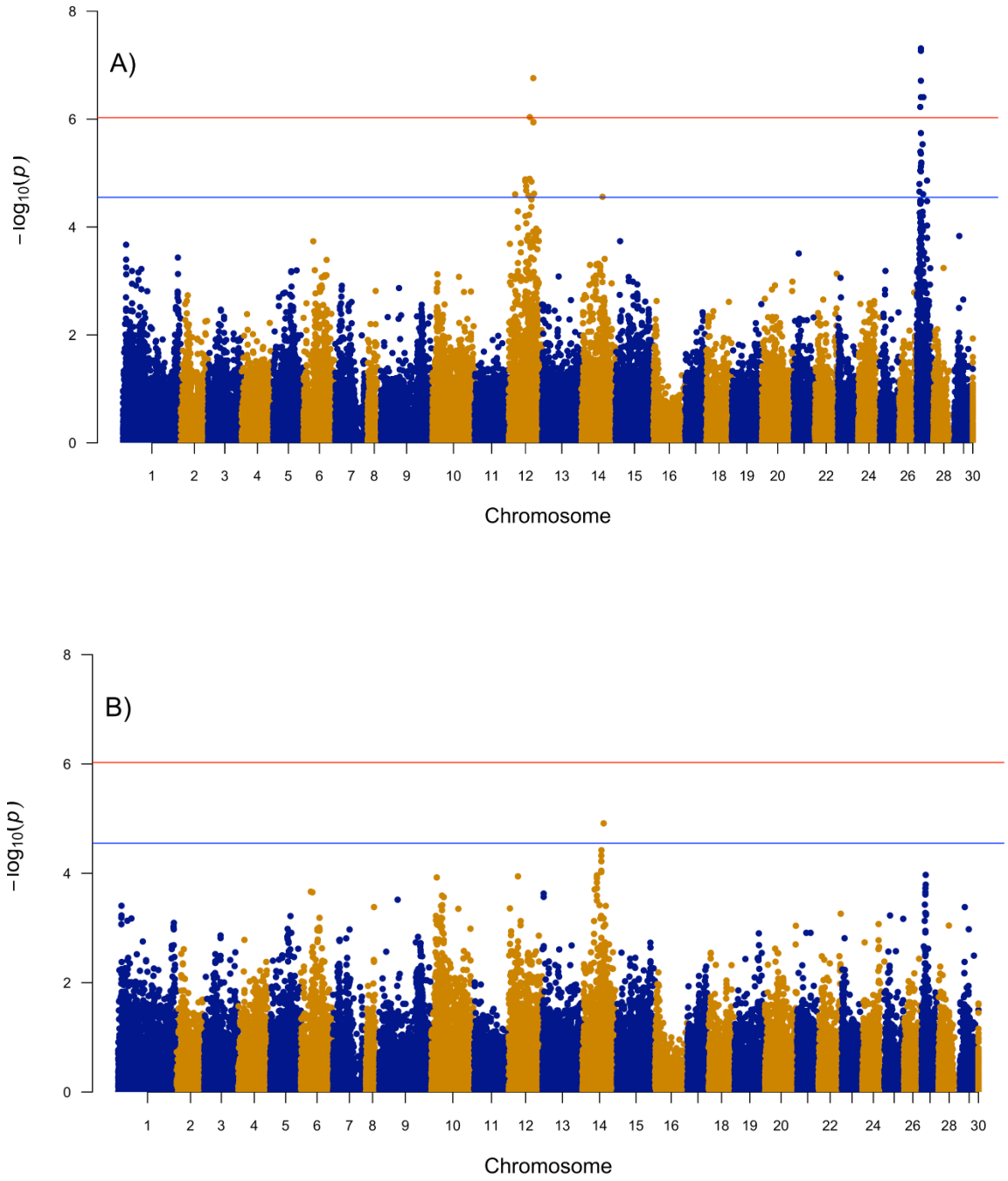


Figure 14. Manhattan plot of genome wide $-log_{10}$P-values for CMS resistance. Plot (A) includes all SNPs; Plot (B) when the most significant SNP on chromosome 12 and 27 were included as covariates. The horizontal red line is the Bonferroni significance threshold ($-log_{10}$P-value = 6.02) and horizontal blue line is the chromosome-wide significant threshold ($-log_{10}$P-value =4.55).

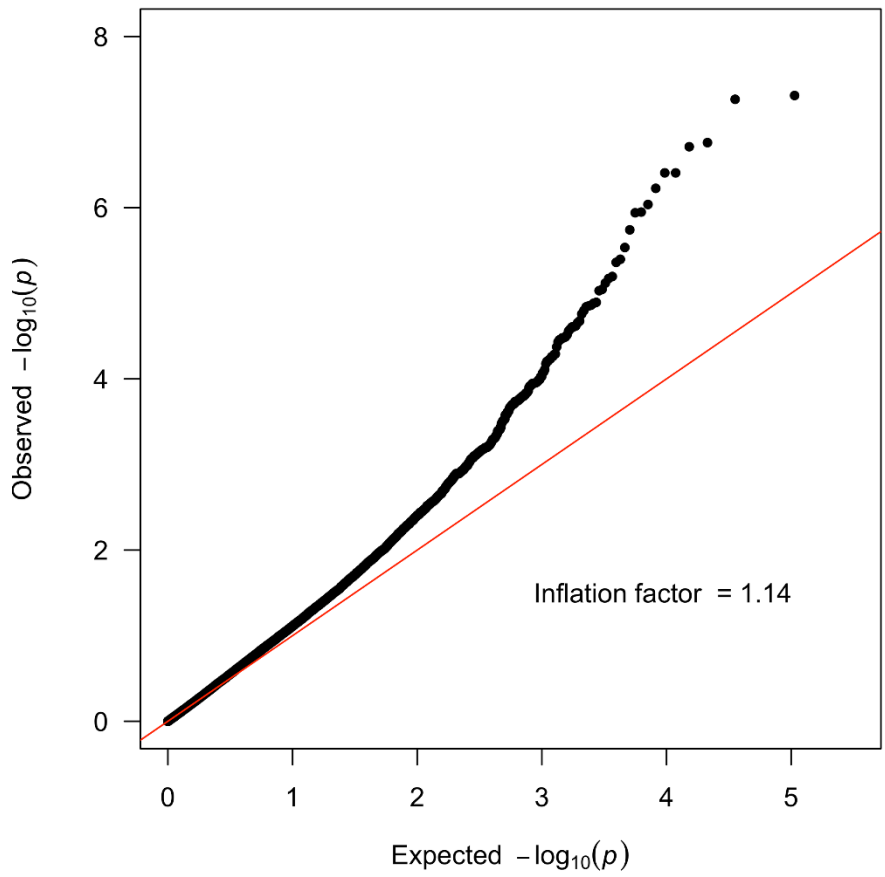


Figure 15. Quantile-quantile (Q-Q) plot of the observed and expected genome wide $-log_{10}$P-value. Red line is when the observed $-log_{10}$P-value equals the expected $-log_{10}$P-value

Quantitative trait loci on chromosome 12

Two SNPs (AX-88118298 and AX-96382208) on chromosome 12 were significant ($-\log_{10}P > 6.02$). And as stated earlier, after fitting the marker AX-96382208 as covariate, any indication of an additional QTL peak was removed, chromosome 12 only had one QTL (Figure 16). The most significant marker (AX-96382208) was located at 67,351,339 bps and marker AX-88118298 was located at 57,843,208 bps and thus 9.51 Mbps apart between these two markers. Interestingly, linkage disequilibrium between the top significant markers (markers with $-\log_{10}P > 4.55$) on chromosome 12 was weak. The LD between the two markers that exceeded $-\log_{10}P > 6.02$ was 0.31 (Figure 17 and Supplementary figure 3).

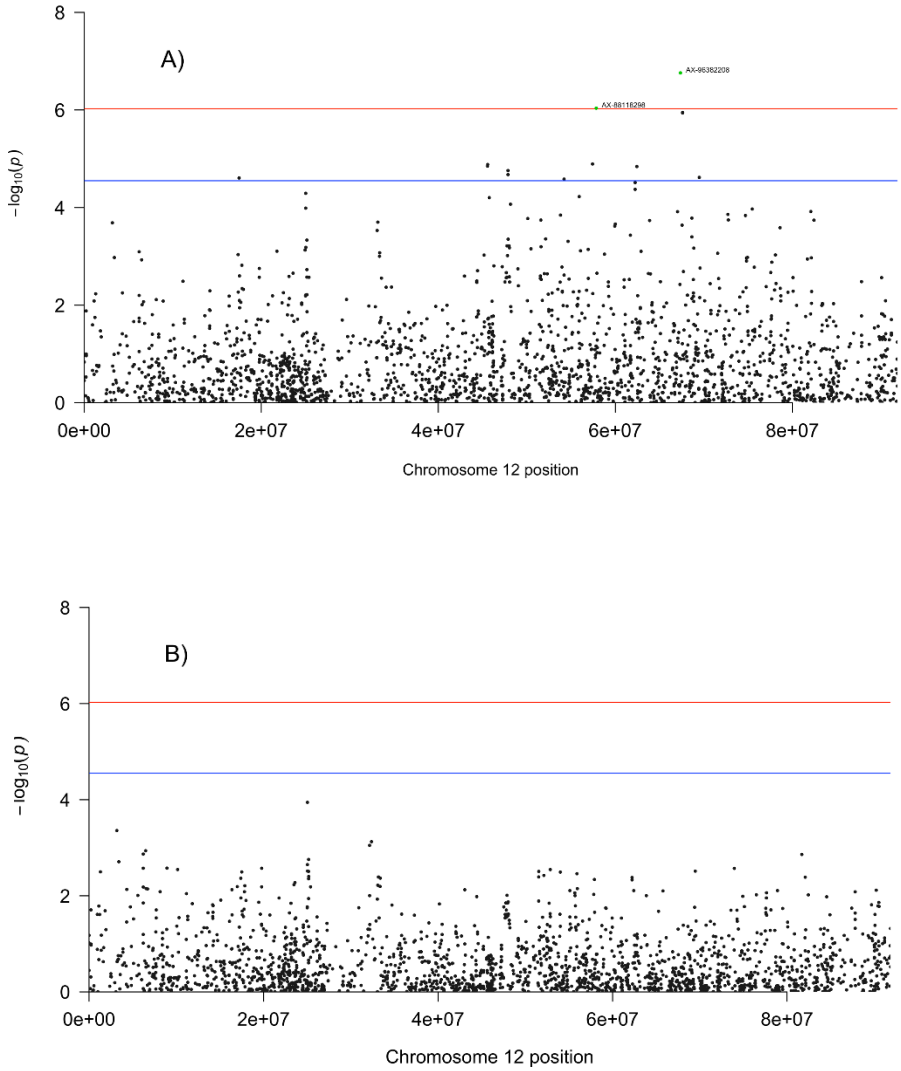


Figure 16. Manhattan plot of chromosome 12 (A) with all SNPs and (B) after accounting for the top significant SNP on chromosome 12. Red line is the Bonferroni significance threshold ($-log_{10}$P-value = 6.02) and blue line is the chromosome-wide significant threshold ($-log_{10}$P-value = 4.55).

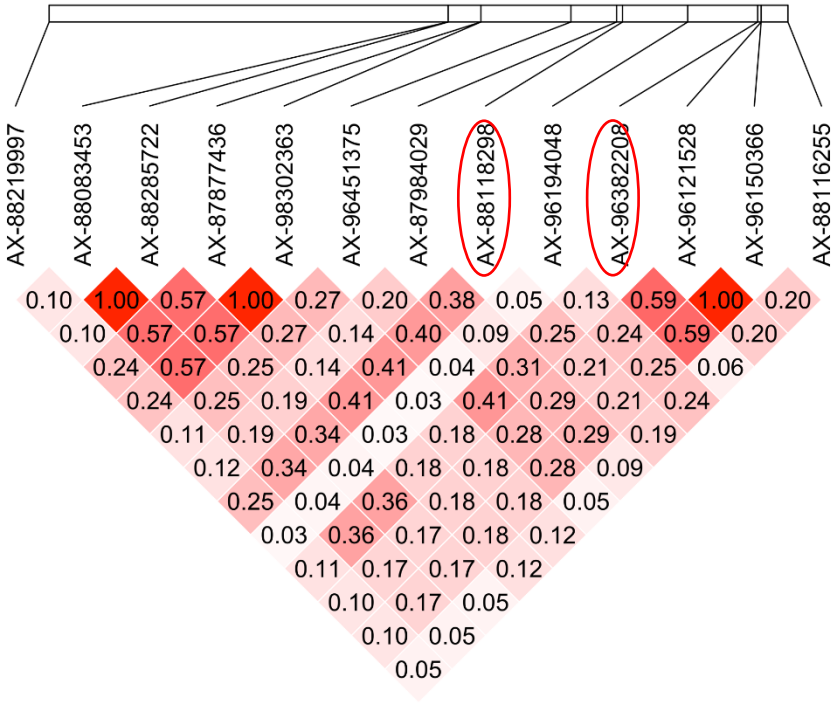


Figure 17. Heatmap of LD structure between SNPs which were chromosome-wide significant ($-log_{10}$P-value is over 4.55) on chromosome 12. The two genome-wide significant SNPs are circled in red.

Quantitative trait loci on chromosome 27

Six markers exceeded the Bonferroni significant threshold ($-\log_{10}P$= 6.02) on chromosome 27. The most significant marker (AX-97886034) was located at 11,723,738 bps. Four out of the six markers were 21.45 Mbps apart (11,723,738 to 11,938,274 bps). The other two markers were located at 10,023,660 bps and 18,612,000 bps. Interestingly, the allele frequency, the allele substitution effect of the marker located at 18,612,000 bps (AX-87482245) was the same as the significant marker at 11,938,274 bps (AX-87110213). Linkage disequilibrium between these two markers was 1.00. Therefore, we hypothesis that the marker located at 18,612,000 bps was misplaced possibly due to genome assembly errors.

When we fitted the top significant markers as covariate, no additional QTL peak was observed (Figure 18). A heatmap of all markers that exceeded the chromosomal significant

threshold is presented in Figure 19. Linkage disequilibrium between the 5 significant markers (11,723,738 to 11,938,274 and 18,612,000 bps) was moderate to high with ranged from 0.50 to 1.00. However, the LD of these 5 markers with the other significant marker which located at 10,023,660 bps was low < 0.20 (Figure 19 and Supplementary figure 4).

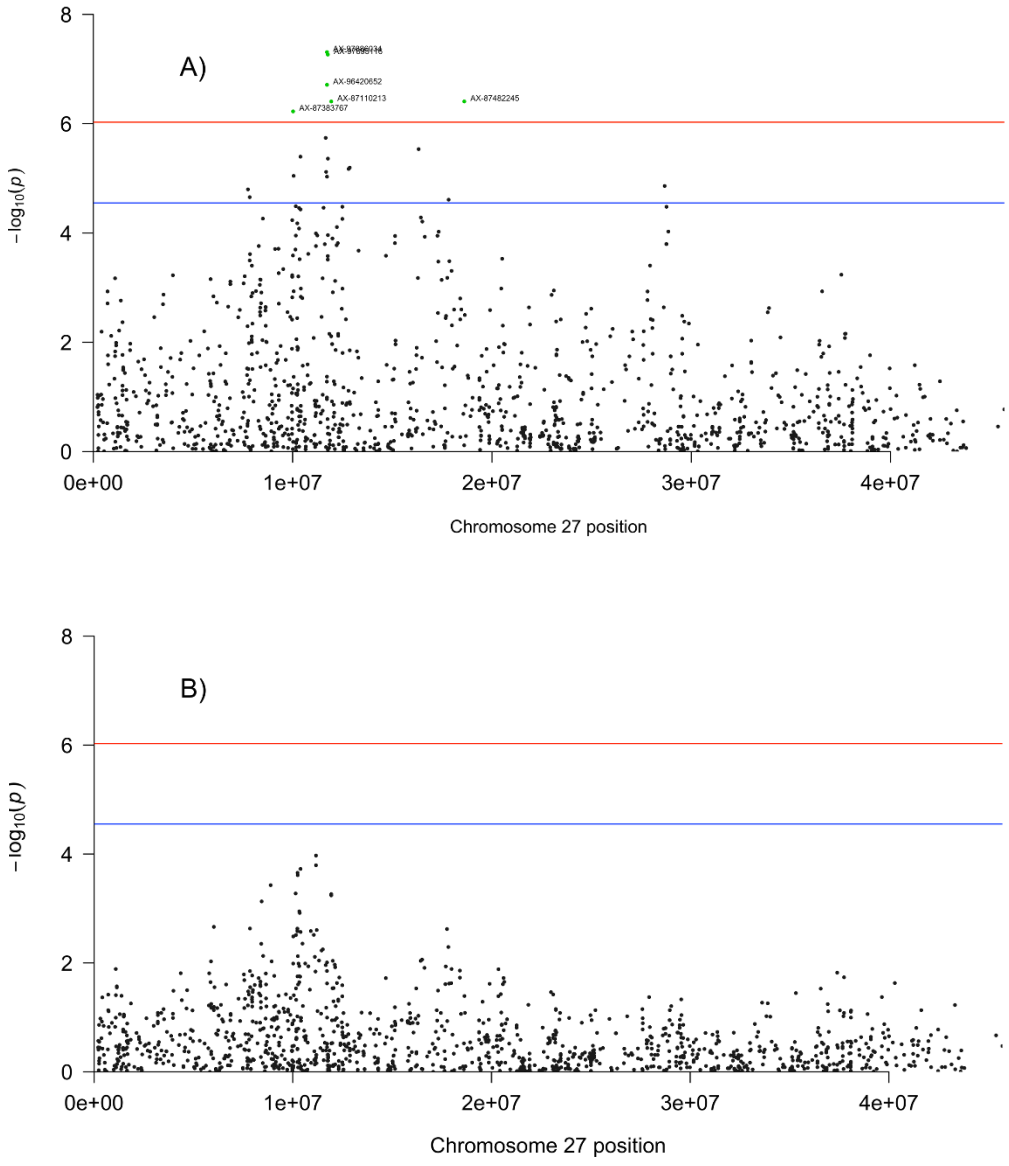


Figure 18. Manhattan plot of chromosome 27 (A) with all SNPs and (B) after accounting for the top significant SNP on chromosome 27. Red line is the Bonferroni significance threshold ($-log_{10}$P-value = 6.02) and blue line is the chromosome-wide significant threshold ($-log_{10}$P-value = 4.55).

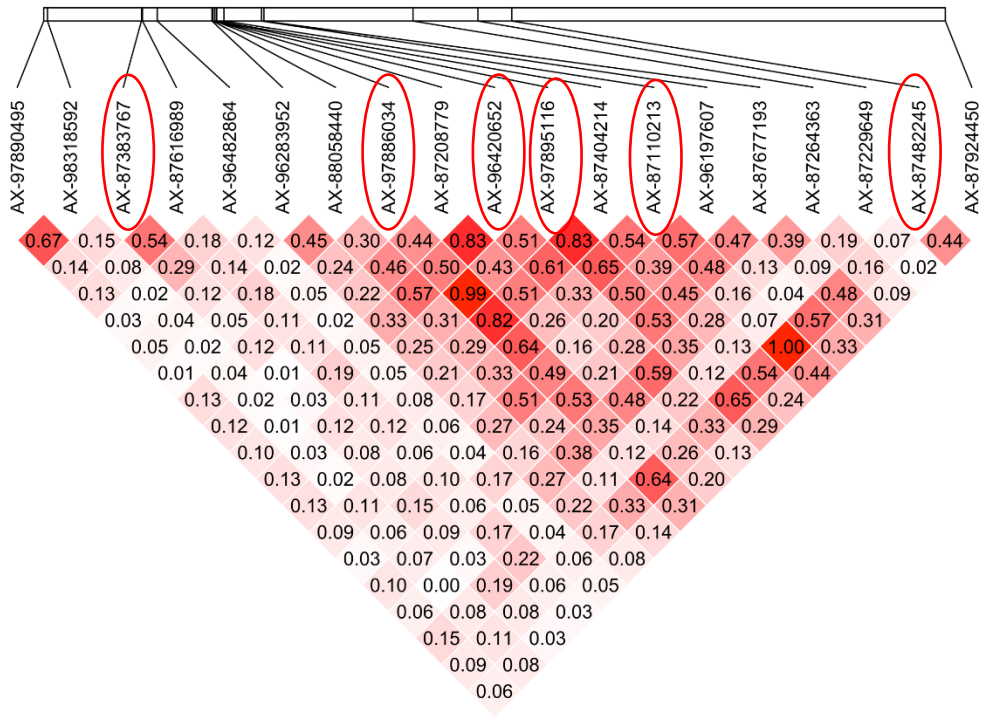


Figure 19. Heatmap of LD structure between SNPs which were chromosome-wide significant (-log$_{10}$P-value is over 4.55) on chromosome 27. The six genome-wide significant SNPs are circled in red.

Summary statistics of significant SNPs on chromosome 12 and 27

Additional summary statistics of the eight significant SNPs are presented in Table 2. Of the total genetic and phenotypic variance in resistance to CMS, the QTL region on chromosome 12 explained 12.4% and 4.9%, respectively. While the QTL region on chromosome 27 explained 12.7% and 5.0 % of the total genetic and phenotypic variance, respectively. These values were computed as the average of the 2 and 6 significant markers on chromosome 12 and 27, respectively. The reason for averaging not accumulating the variance explained by significant markers was only one marker could tag the QTL in this region, which we had shown that before. The dataset was analyzed as survival to CMS outbreak (dead=0 and alive=1), a positive allele substitution effect means increase possibility in survival, only 2 markers had negative allele substitution effect (AX-96382208 and AX-87383767). The allele

frequency of the favorable allele on chromosome 12 ranged from 0.48 to 0.52, and on chromosome 27 from 0.09 to 0.50 (Table 2 and Supplementary table 1).

Table 2 Statistics information of significant SNPs on chromosome 12 and 27.

Chr.	Variant name (SNP ID)	Position (BPs)	Minor allele	MAF (p)[1]	ASE (α)[2]	$-\log_{10}$ *P- value*	%Gsnp[3]	%Psnp[4]
12	AX-88118298	57,843,208	B	0.478	0.154	6.037	11.703	4.615
12	AX-96382208	67,351,339	B	0.471	-0.163	6.760	13.055	5.148
27	AX-87383767	10,023,660	A	0.497	-0.165	6.226	13.395	5.282
27	AX-97886034	11,723,738	A	0.130	0.239	7.310	12.723	5.017
27	AX-96420652	11,724,500	B	0.250	0.187	6.712	12.964	5.112
27	AX-97895116	11,764,773	B	0.130	0.241	7.266	12.949	5.106
27	AX-87110213	11,938,274	A	0.089	0.275	6.406	12.018	4.739
27	AX-87482245	18,612,000	A	0.089	0.275	6.406	12.018	4.739

[1]MAF(p): The frequency of the minor allele.
[2]ASE (α): Allele substitution effect.
[3]%Gsnp: The proportion of the total genetic variance ($\frac{2p_i(1-p_i)a_i^2}{\sigma_g^2 = 0.102}$) which was captured by each SNP.
[4]%Psnp: The proportion of the total phenotypic variance ($\frac{2p_i(1-p_i)a_i^2}{\sigma_p^2 = 0.258}$) which was captured by each SNP.

Bioinformatics and candidate genes

Genes surrounding the top significant markers were investigated and results are shown in Table 3 and Supplementary table 2. For chromosome 12, the two most significant markers were about 9.5 Mbps apart, and therefore these two regions were investigated separately. The first candidate region on chromosome 12 span from 67,101,339 to 67,601,339 bps (significant marker - AX-96382208 was located at 67,351,339 bps) and second from 57,593,208 to 58,093,208 bps (significant marker - AX-88118298 was located at 57,843,208 bps). The two candidate regions contained 20 genes, including 17 protein coding genes. The most significant

SNP (AX-96382208) was located within the gene MAGI1 (ssa12: 67,249,118 to 67,428,971 bps). This gene encoded protein participates in the assembly of multiprotein complexes on the inner surface of the plasma membrane at regions of cell-cell contact (Hata *et al.* 1998). The gene SLC25A26 (ssa12: 67,501,032 to 67,577,876 bps) and ADAMTS9 (ssa12: 67,122,389 to 67,164,932 bps) were on the downstream and upstream of the significant SNP AX-96382208, respectively. SLC25A26 belongs to the mitochondrial carrier family and is related to transport of important small molecules across the mitochondrial inner membrane (Kishita *et al.* 2015). Members of the ADAMTS family have been implicated in controlling organ shape during the development and the inhibition of angiogenesis, it affects heart valve morphogenesis and ventricular cardiac muscle tissue development (Longpré *et al.* 2009). The rest of the protein-coding genes (fourteen) were around the significant SNP AX-88118298 which located at 57,843,208 bps. Among them, GATA2A (ssa12: 57,806,423 to 57,826,331 bps) was located closely to the SNP AX-88118298, whose function may involve in DNA binding and RNA polymerase II transcription factor activity and affect atrioventricular valve formation. The candidate gene RAB7A (ssa12: 57,989,273 to 58,029,764 bps) has direct functional properties with relevance to immune response processes, like antigen processing and presentation of exogenous peptide antigen via MHC class II and positive regulation of viral processes.

DAVID functional annotation cluster divided candidate genes on chromosome12 into three different group, the most significant cluster was related to the energy production in the cells, mainly in ATP binding, while the last two clusters' function were related to cytosol, cytoplasm, nucleus and nucleoplasm.

Three candidate gene regions were defined on chromosome 27. The regions were from 97,75,337 to 10,275,337 bps (significant marker - AX-87383767 was located at 10,023,660 bps), 11,598,578 to 11,848,578 bps (four markers were located within this region) and from 18,360,340 to 18,860,339 bps (significant marker - AX-87482245 was located at 18,612,000 bps). Forty-five different genes were located on these three candidate gene regions. The two

most significant SNPs AX-97886034 and AX-96420652 were located within the PRUNE gene (ssa27: 11,721,333 to 11,744,694 bps). PRUNE encoding-protein has a role in cell proliferation, migration and differentiation. Machida *et al.* (2006) reported that, PRUNE plays an important role in regulating differentiation, survival and aggressiveness of the tumor cells in human. Another gene upstream of the AX-97886034 SNP was MLLT11 (ssa27: 11,671,014 to 11,674,843 bps). MLLT11 encoding-protein regulates lymphoid development by driving multipotent hematopoietic progenitor cells towards a T cell fate (Parcelier *et al.* 2011). The FAM63A gene located at 11,748,704 to 11,766,583 bps, is involved in regulation of transcription by RNA polymerase II. The SNP AX-87383767 was located within the gene TPSN (ssa27: 10,022,846 to 10,029,227 bps). TPSN is involved in immune system response, antigen processing and presentation (Li *et al.* 2000). In addition to these genes, several other putative genes within the region were related to immune responses. For the QTL region between 9,775,337 to 10,275,337 bps, the TUBB5 (ssa27: 9,960,561 to 9,967,866 bps), RFX5 (ssa27: 11,821,440 to 11,840,571 bps), PSMB11 (ssa27: 10,154,707 to 10,163,995 bps) and PSMB8 (ssa27: 10,158,088 to 10,168,546 bps) genes were located within this region. The TUBB5 gene is related to GTP binding, which affects MHC class I protein binding. The RFX5 gene is related to immune response and activate class II MHC promoters. The last two genes, PSMB11 and PSMB8, are the gene for coding of proteasome. Proteasome is related to proteolysis in cellular protein catabolic processes, antigen processing and presentation of exogenous peptide antigen via MHC class I and TAP-dependent, etc.

Twenty-nine candidate genes were annotated in DAVID data base for chromosome 27 and six different clusters were obtained. The most significant cluster was related to innate immune responses. The second significant cluster was about the viral process in host-virus interaction, especially in nucleoplasm and extracellular exosome; the last clusters were related to transcription regulation, alternative splicing and protein and metal binding (include the zinc iron involves).

Table 3. Candidate genes tagged by the significant SNPs directly on chromosome 12 and chromosome 27.

Chr.	SNP	Gene code	Gene function
12	AX-96382208	MAGI1	The protein encoded by this gene plays a role as scaffolding protein at cell-cell junctions.
27	AX-87383767	TPSN	The protein encoded by this gene involves in the association of MHC class I with transporter associated with antigen processing (TAP) and in the assembly of MHC class I with peptide (peptide loading).
27	AX-97886034 AX-96420652	PRUNE	The protein encoded by this gene acts function as both a nucleotide phosphodiesterase and an exopolyphosphatase. This protein is believed to stimulate cancer progression and metastases through the induction of cell motility.
27	AX-97895116	FAM63A	The protein encoded by this gene has exodeubiquitinase activity, may play a regulatory role at the level of protein turnover.

Discussion

The main aim of this study was to estimate heritability and to detect QTLs for CMS resistance from field outbreak information (survival to CMS) of a commercial population. We obtained moderate heritability of 0.395 ± 0.077 and identified one putative QTL on chromosome 12 and one on chromosome 27. Candidate genes in these QTL regions were both directly or indirectly involved in immune functions.

Heritability estimate

There was a significant additive genetic (genomic) variation for resistance to CMS in this study. Vassgård (2017) reported that the estimated heritability (liability scale) of CMS resistance from a field outbreak in a pedigreed population of Marine Harvest was 0.25 ± 0.05. Heritability estimates for resistance to CMS from challenge tests have been reported from two different studies. Boison *et al.* (2019, under-review), reported a heritability of 0.46 using histology information from a pedigree population of the Marine Harvest breeding nucleus. Hillestad & Moghadam (2018) reported a heritability of 0.41 ± 0.05 for resistance to CMS using a quantitative real-time PCR (qRT-qPCR) information obtained from a challenge test of the SalmoBreed breeding nucleus population. Similar heritability estimates have also been reported for other diseases in Atlantic salmon (Ødegård *et al*. 2014). The moderate heritability estimates indicate that there is potential for increasing resistance to PMCV through selective breeding.

The heritability estimates obtained from the model with and without population stratification correction were very similar. This suggests that, although using the top eigenvectors as covariate can reduce the risk of detecting false positive marker associations, the effect of this on the estimated heritability was minimal in our study.

Without genomic information, we could not have estimated heritability in this study. Because the studied fish population was produced from a few sires and dams (~1-2 sires and ~10-15 dams), the estimated heritability was not very reliable as seen by their relatively large standard errors. However, with genomic information, it is possible to estimate heritability from non-pedigreed populations as well as from populations with very few families and large number of offspring per family (Ødegård & Meuwissen 2012).

GWAS results

One QTL was identified on chromosome 12 and the other on chromosome 27. These two chromosomal regions are not duplicates (Lien *et al.* 2016). The two significant SNPs on chromosome 12 were about 9.5 Mbps apart but the most significant markers were located at 67.35 Mbps. Recently, Hillestad & Moghadam (2018) identified a significant QTL on chromosome 12, by using an real-time qRT-qPCR information which obtained from a challenge population. They reported that most of the significant markers on chromosome 12 were located between 61.39 Mbps to 61.70 Mbps, although one SNP was located at 33.47 Mbps (Hillestad & Moghadam 2018). Boison *et al.* (2019, under-review) also reported a significant QTL on chromosome 12 located at 62.82 Mbps using histology information from a Marine Harvest breeding nucleus population.

The six significant SNPs identified on chromosome 27 spanned from 10.02Mbps to 18.61Mbps. However, the marker located at 18.61 Mbps was identical (high LD, same allele frequency and allele substitution effect) to another significant marker located at 11.93 Mbps. Therefore, we conclude that the marker at location 18.61 Mbps was misaligned (possibly due to genome assembly errors) and that its true location was around 11.93 Mbps on chromosome 27. This means that the QTL region on chromosome 27 spanned from 10.02 Mbps to 11.94 Mbps. Meantime, Hillestad & Moghadam (2018) and Boison *et al.* (2019, under-review) identified a significant QTL at 10.39 Mbps and 11.72 Mbps, respectively, on chromosomes 27.

The large difference in the QTL location on chromosome 12 between our study (67.35 Mbps) and that of Hillestad & Moghadam (2018) and Boison *et al.* (2019, under-review) could be explained by the differences in the study populations and the method of infections (challenge test vs. field outbreak). Unlike field outbreaks where mortalities during outbreaks can still be as a result of several unknown causes, data from disease challenged tests are much more cause-specific and much more powerful for GWAS analysis, consequently.

The proportion of genetic variance explained by the two QTL was about 25% (~12.4% for that on chromosome 12 and ~12.7% for that on chromosome 27) in total. Gonen *et al.* (2015) reported that the QTL region on chromosome 3 for PD explained about ~23% of the total genetic variance. Furthermore, Houston *et al.* (2010) and Moen *et al.* (2009) reported that 80-90% of the genetic variance for IPN resistance was captured by a single QTL. When QTL regions explain large proportion of the genetic variance, marker assisted selection can be used to improve the resistance of animals for the traits.

Candidate genes

The heart is the primary organ infected by PMCV in Atlantic salmon. Several putative candidate genes on chromosome 12 showed significant relationship to heart development, like gene ADAMTS9, GATA2A and SFMBT1. The results from the gene cluster analysis on chromosome 12 demonstrated that, the network of candidate genes was mainly involved in ATP binding. A sudden cessation in the production of energy (ATPase) could result in sudden death which might be the major reason for no-visible signs of PMCV infection until a sudden outbreak. Candidate genes on chromosome 12 were also related to immune response, like RAB7A; MHC I protein binding and MHC class II antigen, were consistent to the MHC antigen reported by (Timmerhaus *et al.* 2011b) for gene expression studies of CMS.

Putative candidate genes on chromosome 27 had strong association with immune responses. The immune response genes in the QTL regions of chromosome 27 were consistent

with the results reported by Timmerhaus *et al.* (2011b) after transcriptome profiling of Atlantic salmon infected with the PMCV virus. Both studies suggested that immune responses include antiviral and IFN response, MHC antigen presentation, T cell response and apoptosis in the reaction to CMS resistance. Some of the candidate genes on chromosome 12 and 27 are also related to virus receptor activity which affect the host and virus interaction.

CONCLUSION

In this study, we used data from a field outbreak of CMS in Atlantic salmon to estimate heritability and map QTLs for resistance to CMS. The heritability for CMS resistance was moderate ($h^2 = 0.395 \pm 0.077$). Markers on chromosome 12 and 27 were detected to have strong association with resistance to PMCV, the causative agent of CMS. A QTL on chromosome 12 explained 12.4% and 4.9% of the total genetic and phenotypic variation for CMS resistance, respectively; while a QTL on chromosome 27 explained 12.7% and 5.0% of the total genetic and phenotypic variation. Candidate genes in these two QTL regions were related to immune responses and host reaction to virus (RAB7A, TPSN). We have shown that there was genetic variation for CMS resistance and that field outbreak information can be used to identify QTLs for CMS. We identified QTLs of moderate effect and therefore they could be used for marker assisted selection or genomic selection to increase resistance of Atlantic salmon to the PMCV virus. Further studies to understand the function of associated genes and gene network will be helpful to identify the causative mutation.

REFERENCES

Amin A., & & Trasti J. (1988) Endomyocarditis in Atlantic salmonin Norwegian seafarms. *Bulletin- European Association of Fish Pathologists* **8**, 70.

Bang Jensen B., Brun E., Fineid B., Larssen R.B. & Kristoffersen A.B. (2013) Risk factors for cardiomyopathy syndrome (CMS) in Norwegian salmon farming. *Dis Aquat Organ* **107**, 141-50.

Bangera R., Correa K., Lhorente J.P., Figueroa R. & Yanez J.M. (2017) Genomic predictions can accelerate selection for resistance against Piscirickettsia salmonis in Atlantic salmon (Salmo salar). *BMC Genomics* **18**, 121.

Brocklebank J. & Raverty S. (2002) Sudden mortality caused by cardiac deformities following seining of preharvest farmed Atlantic salmon (Salmo salar) and by cardiomyopathy of postintraperitoneally vaccinated Atlantic salmon parr in British Columbia. *Can Vet J* **43**, 129-30.

Brun E., Poppe T., Skrudland A. & Jarp J. (2003) Cardiomyopathy syndrome in farmed Atlantic salmon Salmo salar: occurrence and direct financial losses for Norwegian aquaculture. *Dis Aquat Organ* **56**, 241-7.

Bruno D.W., Noguera, P. A., & Poppe, T. (2013) *A Colour Atlas of Salmonid Diseases*. Springer Netherlands: Dordrecht.

Bush W.S. & Moore J.H. (2012) Chapter 11: Genome-wide association studies. *PLoS Comput Biol* **8**, e1002822.

Claire Dandine-Roulland (2018). gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. R package version 1.5.3. https://CRAN.R project.org/package=gaston.

Correa K., Bangera R., Figueroa R., Lhorente J.P. & Yanez J.M. (2017) The use of genomic information increases the accuracy of breeding value predictions for sea louse (Caligus rogercresseyi) resistance in Atlantic salmon (Salmo salar). *Genetics Selection Evolution* **49**, 15.

Ehret G.B. (2010) Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Current hypertension reports* **12**, 17-25.

Ferguson H.W., Poppe T. & Speare D.J. (1990) Cardiomyopathy in Farmed Norwegian Salmon. *Diseases of Aquatic Organisms* **8**, 225-31.

Fritsvold C., Kongtorp R.T., Taksdal T., Orpetveit I., Heum M. & Poppe T.T. (2009) Experimental transmission of cardiomyopathy syndrome (CMS) in Atlantic salmon Salmo salar. *Dis Aquat Organ* **87**, 225-34.

Garseth A.H., Fritsvold C., Svendsen J.C., Bang Jensen B. & Mikalsen A.B. (2018) Cardiomyopathy syndrome in Atlantic salmon Salmo salar L.: A review of the current state of knowledge. *J Fish Dis* **41**, 11-26.

Gjedrem T., Gjøen H.M. & Gjerde B. (1991) Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture* **98**, 41-50.

Goddard M.E. & Hayes B.J. (2007) Genomic selection. *J Anim Breed Genet* **124**, 323-30.

Goddard M.E. & Hayes B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**, 381-91.

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

GrotmolS T., KryviH (1997) Detectionofanodavirus-likeagentinheart tissue from reared Atlantic salmon Salmo salar suffering from cardiac myopathy syndrome (CMS). *Diseases of Aquatic Organisms 1997* **29**, 79-84.

Gutierrez A.P., Yáñez J.M., Fukui S., Swift B. & Davidson W.S. (2015) Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar). *PLoS One* **10**, e0119730.

Hata Y., Nakanishi H. & Takai Y. (1998) Synaptic PDZ domain-containing proteins. *Neuroscience research* **32**, 1-7.

Haugland O., Mikalsen A.B., Nilsen P., Lindmo K., Thu B.J., Eliassen T.M., Roos N., Rode M. & Evensen O. (2011) Cardiomyopathy syndrome of atlantic salmon (Salmo salar L.) is caused by a double-stranded RNA virus of the Totiviridae family. *J Virol* **85**, 5275-86.

Hayes B. & Goddard M.E. (2003) Evaluation of marker assisted selection in pig enterprises. *Livestock Production Science* **81**, 197-211.

Hillestad B. & Moghadam H.K. (2018) Genome-wide association study of piscine myocarditis virus (PMCV) robustness in Atlantic salmon (Salmo salar). *bioRxiv*, 458901.

Hjeltnes B B., G, Jansen, M D, Haukaas, A, Walde, C (Eds.) (2017) The Health Situation in Norwegian Aquaculture 2016. Norwegian Veterinary Institute 2017.

Hjeltnes B.-J.B., Bornø G, Haukaas A, & (Ed.) W.C.S. (2018) The Health Situation in Norwegian Aquaculture 2017. Norwegian Veterinary Institute 2018.

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

Khlestkina E.K. (2014) Molecular markers in genetic studies and breeding. *Russian Journal of Genetics: Applied Research* **4**, 236-44.

Kishita Y., Pajak A., Bolar N.A., Marobbio C.M., Maffezzini C., Miniero D.V., Monné M., Kohda M., Stranneheim H. & Murayama K. (2015) Intra-mitochondrial methylation deficiency due to mutations in SLC25A26. *The American Journal of Human Genetics* **97**, 761-8.

Kongtorp R., Taksdal, T., & Lillehaug, A. (2005) Cardiomyopathy syndrome (CMS): a literature review. National Veterinary Institute.

Kongtorp R.T., Halse M., Taksdal T. & Falk K. (2006) Longitudinal study of a natural outbreak of heart and skeletal muscle inflammation in Atlantic salmon, Salmo salar L. *J Fish Dis* **29**, 233-44.

Li S., Paulsson K.M., Chen S., Sjögren H.-O. & Wang P. (2000) Tapasin is required for efficient peptide binding to transporter associated with antigen processing. *Journal of Biological Chemistry* **275**, 1581-6.

Lien S., Koop B.F., Sandve S.R., Miller J.R., Kent M.P., Nome T., Hvidsten T.R., Leong J.S., Minkley D.R. & Zimin A. (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200.

Longpré J.-M., McCulloch D.R., Koo B.-H., Alexander J.P., Apte S.S. & Leduc R. (2009) Characterization of proADAMTS5 processing by proprotein convertases. *The international journal of biochemistry & cell biology* **41**, 1116-26.

Machida T., Fujita T., Ooo M., Ohira M., Isogai E., Mihara M., Hirato J., Tomotsune D., Hirata T. & Fujimori M. (2006) Increased expression of proapoptotic BMCC1, a novel gene with the BNIP2 and Cdc42GAP homology (BCH) domain, is associated with favorable prognosis in human neuroblastomas. *Oncogene* **25**, 1931.

Martin L. & Eskin E. (2017) Review: Population structure in genetic studies: Confounding factors and mixed models. bioRxiv.

Martinez-Rubio L., Evensen O., Krasnov A., Jorgensen S.M., Wadsworth S., Ruohonen K., Vecino J.L.G. & Tocher D.R. (2014) Effects of functional feeds on the lipid composition, transcriptomic responses and pathology in heart of Atlantic salmon (Salmo salar L.) before and after experimental challenge with Piscine Myocarditis Virus (PMCV). *BMC Genomics* **15**, 462.

Meuwissen T.H.E., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-29.

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

Ødegård J. & Meuwissen T.H. (2012) Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genetics Selection Evolution* **44**, 16.

Ødegård J., Moen T., Santi N., Korsvoll S.A., Kjøglum S. & Meuwissen T.H. (2014) Genomic prediction in an admixed population of Atlantic salmon (Salmo salar). *Frontiers in genetics* **5**, 402.

Parcelier A., Maharzi N., Delord M., Robledo-Sarmiento M., Nelson E., Belakhdar-Mekid H., Pla M., Kuranda K., Parietti V. & Goodhardt M. (2011) AF1q/MLLT11 regulates the emergence of human prothymocytes through cooperative interaction with the Notch signaling pathway. *Blood*, blood-2011-01-333179.

Poppe T.T. & Seierstad S.L. (2003) First description of cardiomyopathy syndrome (CMS)-related lesions in wild Atlantic salmon Salmo salar in Norway. *Diseases of Aquatic Organisms* **56**, 87-8.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ribaut J.-M. & Ragot M. (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany* **58**, 351-60.

Rodger H. & Turnbull T. (2000) Cardiomyopathy syndrome in farmed Scottish salmon. *Vet Rec* **146**, 500-1.

Rodger H.D., McCleary S.J. & Ruane N.M. (2014) Clinical cardiomyopathy syndrome in Atlantic salmon, Salmo salar L. *J Fish Dis* **37**, 935-9.

Scholz F., Ruane N.M., Morrissey T., Marcos-Lopez M., Mitchell S., O'Connor I., Mirimin L., MacCarthy E. & Rodger H.D. (2018) Piscine myocarditis virus detected in corkwing wrasse (Symphodus melops) and ballan wrasse (Labrus bergylta). *J Fish Dis* **41**, 147-52.

Slatkin M. (2008) Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-85.

Solomon Boison, Jingwen Ding, Erica Leder, Bjarne Gjerde, Per Helge Bergtun, Ashie Norris, Matthew Baranski & Robinson N. (2019, under-review) QTLs associated with resistance to cardiomyopathy syndrome in Atlantic salmon. Journal of Heredity.

Storset A., Strand C., Wetten M., Sissel K. & Rarnstad A. (2007) Response to selection for resistance against infectious pancreatic necrosis in Atlantic salmon (Salmo salar L.). *Aquaculture* **272**, S62-S8.

Timmerhaus G. (2011a) Cardiomyopathy syndrome (CMS) in Atlantic salmon, Salmo salar L.: functional genomics studies of host-pathogen responses and disease markers. In: *Department of Animal and Aquacultural Sciences Norwegian University of Life Sciences*. Norwegian University of Life Sciences.

Timmerhaus G., Krasnov A., Nilsen P., Alarcon M., Afanasyev S., Rode M., Takle H. & Jorgensen S.M. (2011b) Transcriptome profiling of immune responses to cardiomyopathy syndrome (CMS) in Atlantic salmon. *BMC Genomics* **12**, 459.

Tsai H.Y., Hamilton A., Tinch A.E., Guy D.R., Gharbi K., Stear M.J., Matika O., Bishop S.C. & Houston R.D. (2015) Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**, 969.

VanRaden P.M. (2008) Efficient methods to compute genomic predictions. *Journal of dairy science* **91**, 4414-23.

Vassgård S. (2017) Quantitative genetics of cardiac traits in Atlantic salmon (Salmo salar L.). Norwegian University of Life Sciences, Ås.

Wang A.L., Yang H.M., Shen K.A. & Wang C.C. (1993) Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by a translation frameshift. *Proc Natl Acad Sci U S A* **90**, 8595-9.

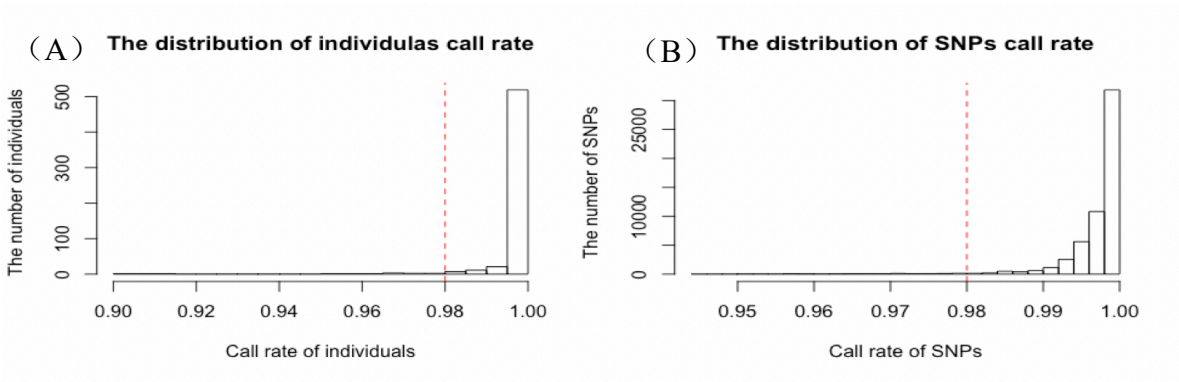
Wiik-Nielsen C.R., Ski P.M., Aunsmo A. & Lovoll M. (2012) Prevalence of viral RNA from piscine reovirus and piscine myocarditis virus in Atlantic salmon, Salmo salar L., broodfish and progeny. *J Fish Dis* **35**, 169-71.

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011a) GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82.

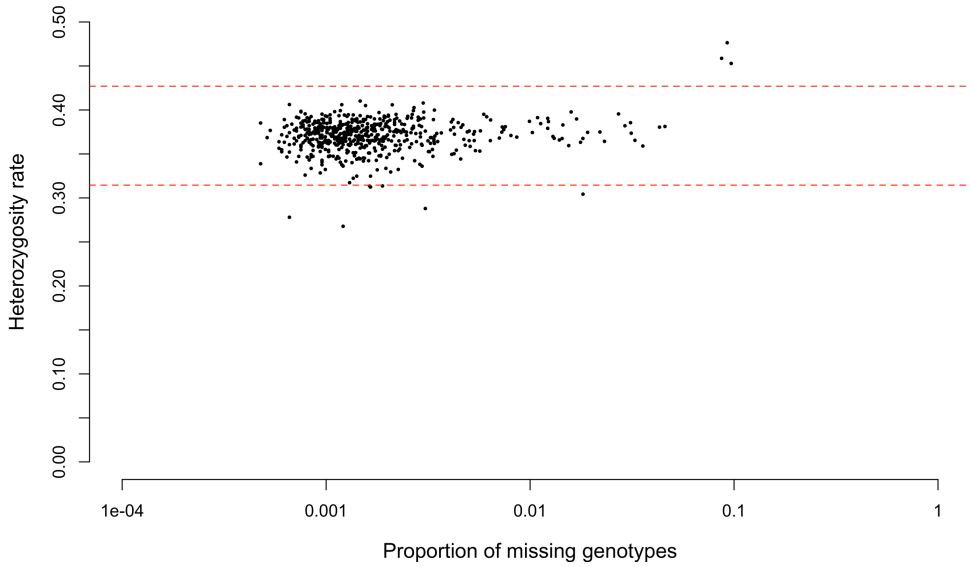
Yang J., Weedon M.N., Purcell S., Lettre G., Estrada K., Willer C.J., Smith A.V., Ingelsson E., O'connell J.R. & Mangino M. (2011b) Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807.

Yang J., Zaitlen N.A., Goddard M.E., Visscher P.M. & Price A.L. (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics* **46**, 100.

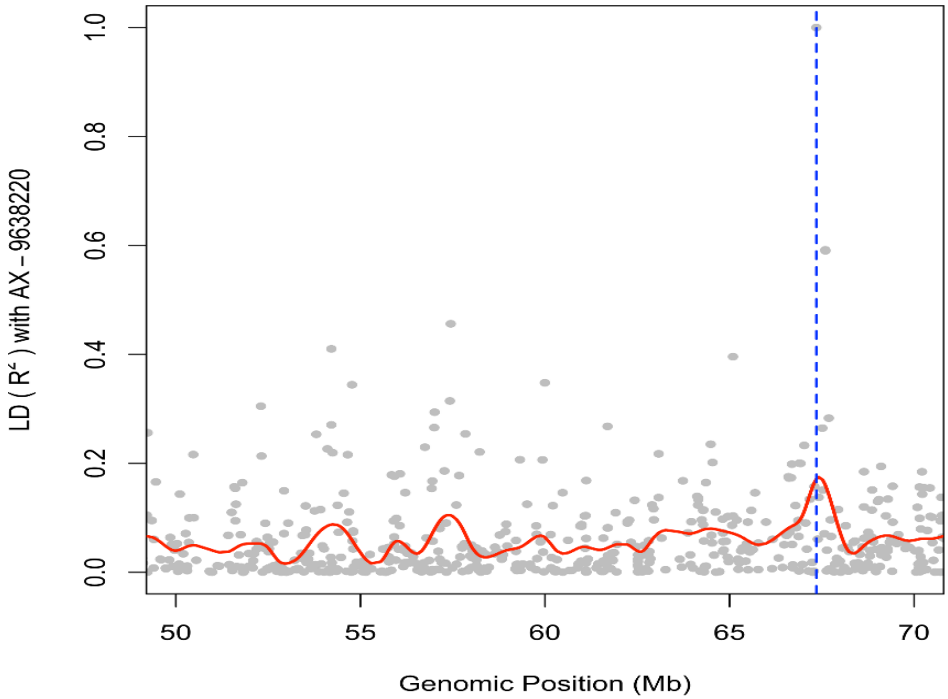
APPENDIX



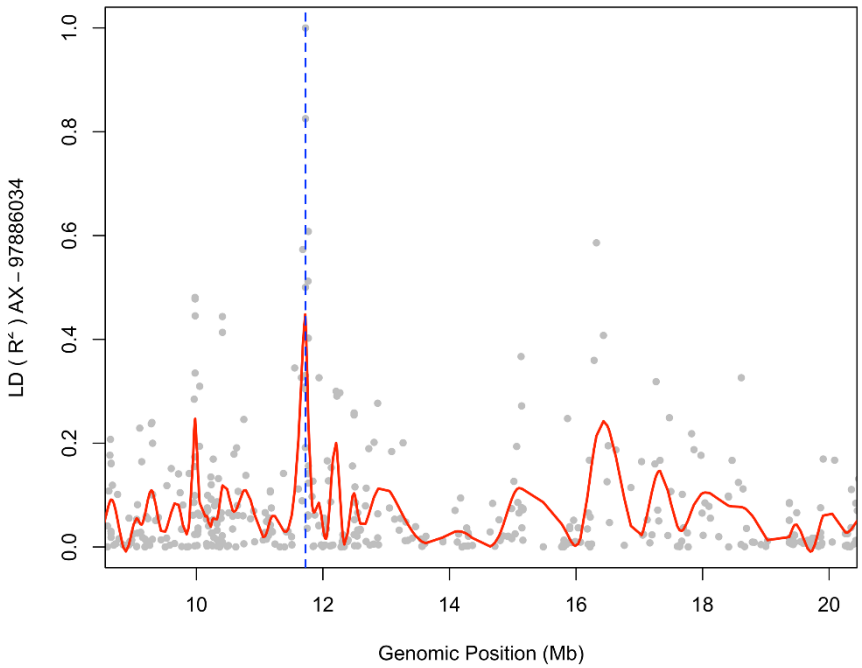
Supplementary figure 1. Call rate of A) individuals and B) SNPs. Quality check threshold for SNP and sample were set to >0.98 (red-dotted-line).



Supplementary figure 2. Heterozygosity rate compared to proportion of missing genotypes of all individuals. Red dash lines were +/- three deviations from mean.



Supplementary figure 3. Degree of LD of all SNPs on the chromosome 17 from 55Mbps to 70 Mbps with the most significant marker (AX-96382208). Red curve is a smoothed spline and the blue dashed line is the position top SNP.



Supplementary figure 4. Degree of LD of all SNPs on the chromosome 27 from 9Mbps to 20 Mbps with the most significant marker (AX-97886034). Red curve is a smoothed spline and the blue dashed line is the position top SNP.

Supplementary table 1. Summary information of genotype distribution and allele frequency in dead and survival fish.

Chr.	Variant name (SNP ID)	[1]GENO-S	[2]GENO-D	[3]MAF-S	[4]MAF-D
12	AX-88118298	60/132/112	72/131/44	0.415	0.557
12	AX-96382208	94/140/70	38/115/94	0.540	0.387
27	AX-87383767	97/148/57	36/132/79	0.566	0.413
27	AX-97886034	6/26/268	13/78/154	0.063	0.212
27	AX-96420652	14/80/209	29/109/109	0.178	0.338
27	AX-97895116	4/31/269	13/78/156	0.064	0.211
27	AX-87110213	2/20/287	5/65/177	0.038	0.152
27	AX-87482245	2/20/287	5/65/177	0.038	0.152

[1]GENO-S: the genotype counts in survival fish, with the allele count of the minor allele (left), heterozygotes (middle) and major allele (right).

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

[3]MAF-S: the frequency of the minor allele in survival fish.

[4]MAF-D: the frequency of the minor allele in dead fish.

Supplementary table 2. The function of genes which locates in the QTL regions.

Chr.	Name[1]	Function[2]
ssa12	ADAMTS9	Protein coding gene. It is implicated in controlling organ shape development and the inhibition of angiogenesis. It affects heart valve morphogenesis and ventricular cardiac muscle tissue development.
ssa12	SLC25A26	Protein coding gene. It is related to ATP binding, which transports important small molecules across the mitochondrial inner membrane.
ssa12	MAGI1	Protein coding gene. This gene participates in the assembly of multiprotein complexes on the inner surface of the plasma membrane at regions of cell-cell contact
ssa12	GAG-POL	Protein coding gene. Its essential events are in virion assembly, including binding the plasma membrane, making the protein-protein interactions necessary to create spherical particles, recruiting the viral proteins, and packaging the genomic RNA via direct interactions with the RNA packaging sequence (Psi).
ssa12	ARPC4	Protein coding gene. It controls actin polymerization in cells and has been conserved throughout eukaryotic evolution, produces fusion protein when combines with the downstream tubulin tyrosine ligase-like family, member 3 (TTLL3).
ssa12	RFT1	Protein coding gene. Its coded protein has virus receptor activity, helps virus entry into host cells.
ssa12	MAPKAPK3	Protein coding gene. It actives MAPK activity and affects ATP binding.
ssa12	CHCHD4	Protein coding gene. It affects protein disulfide oxidoreductase activity; the function is related to protein transportation.
ssa12	ORF2P	Protein coding gene. Its function is related to RNA-directed DNA polymerase and type II site-specific deoxyribonuclease activity, which affects reverse transcription involved in RNA-mediated transposition process and nucleic acid phosphodiester bond hydrolysis respectively.
ssa12	RAB7A	Protein coding gene. It participates in immune response process, like antigen processing and presentation of exogenous peptide antigen via MHC class II, positive regulation of viral process.
ssa12	GATA2A	Protein coding gene. It affects DNA binding and RNA polymerase II transcription factor activity, participating atrioventricular valve formation.
ssa12	SFMBT1	Protein coding gene. Histone-binding protein, which is part of various corepressor complexes. Negative regulation of muscle organ development
ssa12	RUVBL1	Protein coding gene. Possesses single-stranded DNA-stimulated ATPase and ATP-dependent DNA helicase (3' to 5') activity
ssa12	ORF1	Protein coding gene. Nucleic acid-binding protein which is essential for retro transposition of LINE-1 elements in the genome. Functions as a nucleic acid chaperone binding its own transcript and therefore preferentially mobilizing the transcript from which they are encoded.

ssa12	PFKFB4B	Protein coding gene. Participates ATP binding activity.
ssa12	SEC61AB	Protein coding gene. Appears to play a crucial role in the insertion of secretory and membrane polypeptides into the endoplasmic reticulum. It is required for assembly of membrane and secretory proteins. Found to be tightly associated with membrane-bound ribosomes, either directly or through adaptor proteins.
ssa12	PRKCDA	Protein coding gene. It plays contrasting roles in cell death and cell survival by functioning as a pro-apoptotic protein during DNA damage-induced apoptosis but acting as an anti-apoptotic protein during cytokine receptor-initiated cell death, is involved in tumor suppression.
ssa27	TCF19	Protein coding gene. Potential trans-activating factor that could play an important role in the transcription of genes required for the later stages of cell cycle progression.
ssa27	TPSN	Protein coding gene. Involved in the association of MHC class I with transporter associated with antigen processing (TAP) and in the assembly of MHC class I with peptide.
ssa27	PSB9	Protein coding gene. Cleavage of peptide bonds with very broad specificity and participate in immune system process.
ssa27	ZN384	Protein coding gene. This gene is related to DNA binding.
ssa27	DAXX	Protein coding gene. Death domain associated protein resides in multiple locations in the nucleus and in the cytoplasm. In the nucleus, the encoded protein functions as a potent transcription repressor that binds to sumoylated transcription factors. In the cytoplasm, the encoded protein may function to regulate apoptosis. Potentially participates virial process.
ssa27	ZBTB22	Protein coding gene. The proteasome is a multi-catalytic proteinase complex. Molecular functions are hydrolase, protease, threonine protease.
ssa27	HA1F	Protein coding gene. It is involved in the presentation of foreign antigens to the immune system.
ssa27	PSMB11	Protein coding gene. It codes the proteasome, which is a multi-catalytic proteinase complex and characterized by its ability to cleave peptides. Participates antigen processing and presentation and is related to virus process.
ssa27	PSMB8	Protein coding gene. It codes the proteasome, which is a multi-catalytic proteinase complex and characterized by its ability to cleave peptides. Participates antigen processing and presentation and is related to virus process.
ssa27	BRD2A	Protein coding gene. It participates in the regulation of apoptotic process.
ssa27	PVRL2L	Protein coding gene. The coded protein participates in virion attachment to host cell process.
ssa27	FAM73A	Protein coding gene. The coded protein is the regulator of mitochondrial fusion.

ssa27	MLLT11	Protein coding gene. Its encoding-protein is relevant to immune system, because it regulates lymphoid development by driving multipotent hematopoietic progenitor cells towards a T cell fate.
ssa27	PSMB4	Protein coding gene. Proteasome complex has negative regulation of inflammatory response to antigenic stimulus.
ssa27	PI4KB	Protein coding gene. The corresponding protein may regulate Golgi disintegration/reorganization during mitosis, possibly via its phosphorylation.
ssa27	PRUNE	Protein coding gene. Proteasome complex has a role in cell proliferation, migration and differentiation.
ssa27	FAM63A	Protein coding gene. It codes the proteasome, function as hydrolase, protease or thiol protease. It has exodeubiquitinase activity and may have relation with protein turnover.
ssa27	RFX5	Protein coding gene. It activates transcription from class II MHC promoters.
ssa27	POGZ	Protein coding gene. This gene plays a role in mitotic cell cycle progression and is involved in cell division.
ssa27	EDN2	Protein coding gene. The products of this gene are endothelium-derived vasoconstrictor peptides, with positive regulation of cell proliferation and heart rate.
ssa27	STMN1B	Protein coding gene. This gene's function is tubulin binding.
ssa27	TPASE	Protein coding gene. Bifunctional mRNA-capping enzyme, exhibiting RNA 5'-triphosphatase activity in the N-terminal part and mRNA guanylyl transferase activity in the C-terminal part. Participates GTP binding process.

[1]Gene only include the one has certain known function.

[2]Function is refer to https://www.uniprot.org and https://www.ncbi.nlm.nih.gov.

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

Postboks 5003
NO-1432 Ås
Norway