Norges miljø- og biovitenskapelige universitet

Faculty of Veterinary Medicine
Department of Paraclinical Sciences
Virology Unit

Thesis 2022, 40 pt
Project related specialization

Sequence Analysis of PRV-1 Isolates Pre- and Post-emergence of HSMI

Sekvensanalyse av PRV-1-isolater fra tidsrommet da HSMB brøt ut i Norge

Thea Carlsson, Anette Edvardsen & Anneli Yrjönheikki
Class 2017

Supervisors: Espen Rimstad, Øystein Wessel & Nina Askim Vatne

Table of Contents

Preface

We are three veterinary students in our last year of study who have chosen to specialize in scientific research to explore and become more familiar with scientific investigation, methodology and analysis. Due to a common interest in virology and aquaculture we wanted to immerse ourselves in these topics, and fortunately we got the opportunity to work with the virology team at the Department of Paraclinical Sciences at the Faculty of Veterinary Medicine and their ongoing research on *Piscine orthoreovirus*-1 (PRV-1). Today PRV-1 is the most widespread virus in Norwegian Atlantic salmon aquaculture and the causative agent of the important and emerging disease heart and skeletal muscle inflammation (HSMI). Improved knowledge about PRV-1 and HSMI is important for the future productivity and sustainability of the aquaculture sector and working on such a significant project has been incredibly motivating and rewarding as well as an educational process.

Abstract

Title: Sequence Analysis of PRV-1 Isolates Pre- and Post-emergence of HSMI

Authors: Thea Carlsson, Anette Edvardsen and Anneli Yrjönheikki

Supervisors: Espen Rimstad, Øystein Wessel and Nina Askim Vatne, Department of Paraclinical Sciences

Piscine orthoreovirus-1 (PRV-1) is the causative agent of heart and skeletal muscle inflammation (HSMI) in farmed Atlantic salmon, first described in Norway in 1999. Not all infected fish develop HSMI, and the virus has been revived from archived samples dating back to 1988, leading to a hypothesis that viral evolution has generated new genotypes with increased virulence contributing to disease development. The objective was to identify PRV-1 isolates present in Norwegian farmed salmon in the years before and after 1999. Archived samples from Norwegian farmed salmon, comprising 204 samples from 1987-2008, were screened for PRV-1. Coding sequences were obtained for segment S1, S4, M2, L1 and L2, putatively linked to virulence, from six isolates. Phylogenetic analysis and genogrouping was performed. Screening displayed that PRV-1 was common in farmed salmon predating the first described HSMI-outbreak, and sequence analysis demonstrated a mixture of assumed high and low virulent isolates following 1999. Coinfection with isolates belonging to different phylogenetic groups was demonstrated in two fish from 1998 and 2005. Co-infection is a prerequisite for reassortment, linked to the hypothesis of the origin of the high virulent strains through evolution. In the presumed receptor binding protein σ1, a new amino acid change was discovered (E254K).

Sammendrag (Abstract in Norwegian)

Tittel: Sekvensanalyse av PRV-1-isolater fra tidsrommet da HSMI brøt ut i Norge

Forfattere: Thea Carlsson, Anette Edvardsen og Anneli Yrjönheikki

Veiledere: Espen Rimstad, Øystein Wessel og Nina Askim Vatne, Institutt for parakliniske fag

Piscine orthoreovirus-1 (PRV-1) er årsaken til hjerte- og skjelettmuskelbetennelse (HSMB) hos oppdrettslaks, først beskrevet i Norge i 1999. Alle PRV-1-infiserte fisk utvikler imidlertid ikke HSMB og viruset har blitt detektert helt tilbake til 1988, som har ført til en hypotese om at viral evolusjon har generert nye genotyper med økt virulens som har bidratt til sykdomsutvikling. Hovedmålet med studien var å kartlegge genomet til PRV-1-isolater før og etter første beskrevne HSMB-utbrudd. 204 arkiverte prøver fra norsk oppdrettslaks fra 1987-2008 ble screenet for PRV-1. Kodende sekvens ble sekvensert for segment S1, S4, M2, L1 og L2, antatt koblet til virulens, fra seks isolater. Fylogenetisk analyse og genogruppering ble utført. Screeningen viste at PRV-1 var et vanlig forekommende virus hos oppdrettslaks allerede før det første beskrevne HSMB-utbruddet og sekvensanalysen demonstrerte en blanding av antatt høyvirulente og lavvirulente isolater i tidsperioden etter 1999. Studien påviste også koinfeksjon med to isolater tilhørende ulike fylogenetiske grupper hos to fisk, fra 1998 og 2005. Koinfeksjon er en forutsetning for reassortering, som knyttes til hypotesen om at høyvirulente stammer har oppstått gjennom reassortering av ulike lavvirulente stammer. I det antatt reseptorbindende proteinet σ1 ble det oppdaget en ny aminosyreendring (E254K).

Definitions and Abbreviations

Aa	Amino acid
Ambiguous nucleotide	Position with ≥ two peaks in a chromatogram, where the smallest is at least half as tall as the tallest peak
Bp	Base pair
cDNA	Complementary DNA
CMS	Cardiomyopathy syndrome
Codon	Three succeeding nucleotides in a DNA or RNA sequence, together forming a unit encoding a particular amino acid
Co-infection	Virology; simultaneous infection of a single cell by two or more virus particles
Ct	Cycle threshold
ddNTP	Dideoxynucleotide triphosphate
DEPC water	Diethyl pyrocarbonate water
dNTP	Deoxynucleotide triphosphate
DPBS	Dulbecco's phosphate buffered saline
DTT	Dithiothreitol
dsRNA	Double stranded RNA
EF1αB	Elongation factor 1αB of Atlantic salmon
HSMI	Heart and skeletal muscle inflammation
ISA	Infectious salmon anemia
IPN	Infectious pancreatic necrosis
IUPAC nucleotide codes	International letter codes for describing ambiguous nucleotides
mRNA	Messenger RNA
MRV	Mammalian orthoreovirus

NCBI	National Center for Biotechnology Information
NMBU	Norwegian University of Life Sciences
NVI	Norwegian Veterinary Institute
ORF	Open reading frame, a portion of an RNA sequence that, when translated into amino acids, contains no stop codons
PD	Pancreas disease
PCR	Polymerase chain reaction
PRV	*Piscine orthoreovirus*
qPCR	Quantitative polymerase chain reaction
Reading frame	One of three possible ways to translate an RNA strand
RNA	Ribonucleic acid
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
Salmonid species	Taxonomic; Family *Salmonidae*
Secondary peaks	Position with ≥ two peaks in a chromatogram
Silent mutation	A single nucleotide change within a protein-coding portion of a gene that does not affect the amino acid sequence
TAE	Tris-acetate-EDTA
UTR	Untranslated region

Introduction

Atlantic Salmon Aquaculture in Norway

Aquaculture is the fastest growing animal food-production sector globally and Atlantic salmon (*Salmo salar*) is one of the species that is extensively used in fish farming (FAO, 2022). In 2021, the total global production of farmed Atlantic salmon was approximately 2,7 million metric tons (Statista, 2022). Norway is the largest salmon farming nation in the world accounting for more than 50 % of the total global production (The Norwegian Directorate of Fisheries, 2022). The global aquaculture sector is still growing, and is expected to expand further (FAO, 2022).

The Norwegian salmon farming has its origin in the 1970s, when the pioneers Ove and Sivert Grøntvedt slaughtered 20 000 salmon smolt from octagonal fish cages in Laksåvika on the Norwegian west coast. This is considered as the first successful salmon fish farm in the world and has laid the foundation of modern Atlantic salmon aquaculture worldwide. Salmon farming became an industry in Norway during the 1980s, and the sector has over the last decades grown and developed at a rapid pace with a total production of roughly 1,5 million metric tons in 2021 (see figure 1). The rapid expansion and intensification have given rise to challenges that impact mortality, growth rate and slaughter quality (Norwegian Seafood Council, 2020; The Norwegian Directorate of Fisheries, 2022).

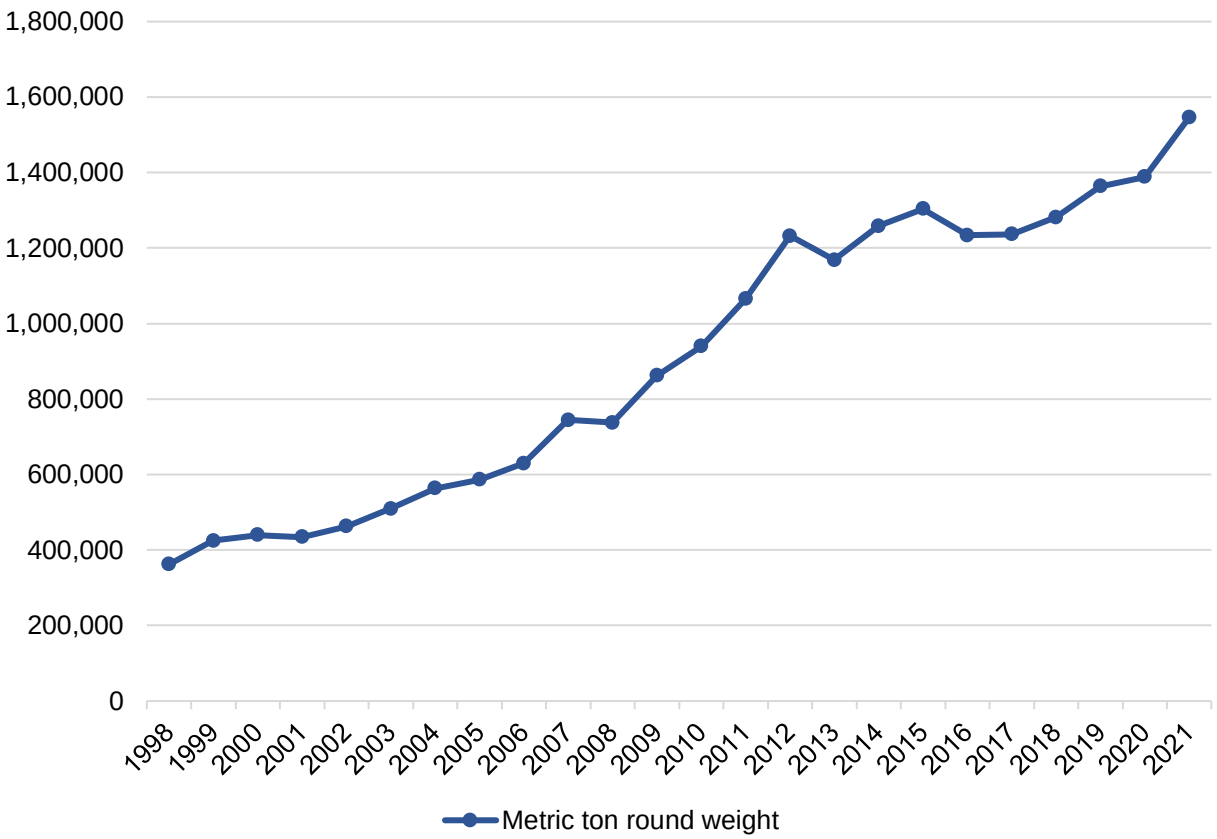


Figure 1. Total slaughtered salmon in Norway from 1998-2021. The Norwegian Directorate of Fisheries. (2022). *Salg 1994-2021*. Available at: https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Laks-regnbueoerret-og-oerret/Matfiskproduksjon (accessed 12.09.2022).

One of the greatest threats to the aquaculture sector is emerging infectious diseases, due to their negative impact on both welfare, environment and profit through loss of biomass and reduced slaughter quality (Aldrin et al., 2009). Large-scale farming of a monoculture at high stocking density as well as transportation of fish and equipment over large geographic distances provides optimal conditions for the spread of infectious diseases. The salmon farming industry has encountered several challenges related to various bacterial, viral and parasite infections since its beginning in 1970, such as furunculosis, infectious salmon anemia (ISA) and salmon lice.

Disease prevention and control is essential to ensure a sustainable farming industry, both ecologically and economically. Immunoprophylaxis has been an important tool in preventing and controlling some of the most common infectious diseases, primarily bacterial (Gudding &

Van Muiswinkel, 2013), which has been crucial for the rapid growth in the industry. Some of the first effective vaccines were introduced against several of the most prevalent bacterial diseases in the 1980s, i.e., classical and cold water vibriosis and furunculosis. The control of furunculosis was pivotal and led to a linear growth in the industry as well as a massive decrease in the use of antibiotics (Gravningen & Berntsen, 2008). Vaccines against viral diseases have not had the same success as for bacterial diseases, and several viral diseases therefore still pose a significant threat to the industry today (Aldrin et al., 2009).

The most important viral diseases in Norway today are heart and skeletal muscle inflammation (HSMI), cardiomyopathy syndrome (CMS), pancreas disease (PD) and infectious salmon anemia (ISA). Out of these, HSMI was the most frequently reported disease in 2021 with a total of 188 diagnoses (see table 1) (Norwegian Veterinary Institute, 2022). To be able to develop methods for preventing and controlling these diseases in the future, in-depth knowledge about the causative agents is crucial.

	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
HSMI	162	142	134	181	135	101	93	104	79	161*	188*
CMS	74	89	100	107	105	90	100	101	82	154*	155*
IPN	154	119	56	48	30	27	23	19	23	22*	20*
ISA	1	2	10	10	15	12	14	13	10	23	25
PD	89	137	99	142	137	138	176	163	152	158	100

Table 1. Number of salmon farming sites in Norway with detected viral diseases from 2011-2021. For ISA and PD, new locations with diagnosed disease are shown. For CMS, IPN and HSMI, annual disease diagnoses are shown. *For the period 2011-2019, the number of diagnoses is based on cases submitted to the Norwegian Veterinary Institute, while data made available from private laboratories is included for 2020 and 2021. Modified table from: Norwegian Veterinary Institute. (2022). *Norwegian Fish Health Report 2021*. Norwegian Veterinary Institute Report, series 2a/2022. Available at: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2022/fish-health-report-2021 (accessed 20.08.2022).

Emergence and Spread of HSMI in Norway

In 1999, the first outbreak of a new and emerging disease in sea-caged Atlantic salmon was reported in Trøndelag, Mid-Norway. The disease was characterized by observation of a unique set of histopathological lesions in heart and skeletal muscle, and the disease was therefore designated "heart and skeletal muscle inflammation" (Kongtorp et al., 2004a). In the following years, the disease spread along the coast and became an increasing problem with a yearly increase in the number of reported outbreaks (see figure 2) (Norwegian Veterinary Institute, 2012; Olsen, 2022).

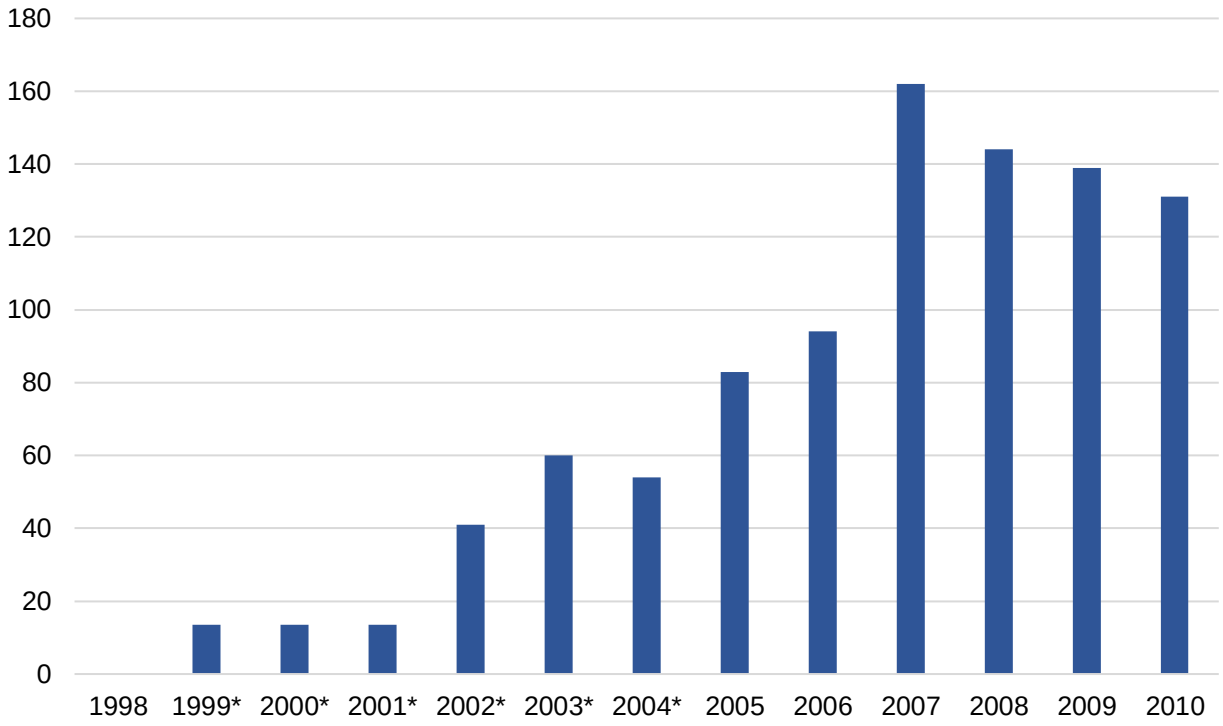


Figure 2. Annual HSMI-diagnosed locations in Norway from 1998-2010. *The numbers are uncertain as the disease probably was underdiagnosed. Norwegian Veterinary Institute. (2012). *The Health Situation in Norwegian Aquaculture 2011*. Norwegian Veterinary Institute Report. Available at: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2012/fish-health-report-2011 (accessed 05.10.2022); Olsen, A. B. (2022). *Number of HSMI Outbreaks 1999-2005* (e-mail to Anne Berit 04.10.2022)

Today numerous outbreaks of HSMI are reported annually in Norway (see table 1). HSMI was classified as a notifiable disease in Norway from 2006 to 2014, with the highest number

reported outbreaks in 2014 (Norwegian Veterinary Institute, 2015). The denotification of HSMI in 2014, i.e. the diagnosis did not have to be verified by NVI, led to increased use of private diagnostic laboratories, resulting in loss of overview of the total number annual HSMI outbreaks and a decline in the number reported outbreaks declined the following years (see table 1). It is assumed that the disease was highly underreported at this time, and the drop in prevalence did therefore not necessarily reflect an actual decrease in disease outbreaks (Norwegian Veterinary Institute, 2018). From 2020, results from private diagnostic laboratories and the Norwegian Veterinary Institute have been merged, which has led to an increase in the total number yearly outbreaks again. In 2021, the currently highest number of annual HSMI diagnoses was reported, with a total of 188 diagnoses. Out of these outbreaks, 75 % were reported in between January and July (Norwegian Veterinary Institute, 2022).

Heart and Skeletal Muscle Inflammation (HSMI)

HSMI is a transmissible viral disease affecting Atlantic salmon, usually occurring 5-9 months after sea transfer during spring and early summer. Clinically affected fish display unspecific symptoms such as anorexia and abnormal swimming behavior. The morbidity during an outbreak can reach 100 %, while the mortality can vary from almost negligible up to 20 %. Environmental and management stress such as delousing can predispose to the development of disease (Kongtorp et al., 2004b).

Gross Pathology and Histopathology

Common autopsy findings indicate signs of circulatory failure and include pale heart, yellow liver, swollen spleen, ascites and petechiae in perivisceral fat (Kongtorp et al., 2004b).

The diagnosis of HSMI is based on histopathological examination (Kongtorp et al., 2004a) and the major histopathological changes are found in the heart and red skeletal muscle. Typical lesions in the heart are epicarditis with different degrees of mononuclear inflammatory cell infiltrations, inflammation in the endocardium and sometimes myocardium, in severe cases resulting in a pancarditis. Together with heart lesions, one often finds degeneration and necrosis of red skeletal muscle. As a result of the heart failure, ischemic necrosis may develop in the liver (Kongtorp et al., 2004b; Bjørgen et al., 2015).

Piscine Orthoreovirus-1 (PRV-1) is the Causative Agent of HSMI

In 1999, at the time HSMI was first discovered, the etiology of the disease was unknown. In 2004, an infectious etiology was demonstrated through transmission of the disease to naive fish by experimental injections with tissue homogenate from previously HSMI-diagnosed fish as well as by cohabitation challenge. The experiment suggested viral etiology as no difference in disease induction was observed between fish inoculated with inoculum containing antibiotics and a non-treated inoculum (Kongtorp et al., 2004a). A later challenge experiment indicated that the causal agent may be a non-enveloped virus through inoculation with tissue homogenate pre-treated with chloroform (Kongtorp & Taksdal, 2009).

In 2010, a novel reovirus, given the name *Piscine orthoreovirus* (PRV), was discovered and found to be associated with HSMI (Palacios et al., 2010). A causal relationship was later confirmed as experimental infection with purified PRV-1 reproduced histopathological lesions consistent with HSMI, which defined PRV-1 as the causative agent of HSMI in Atlantic salmon (Wessel et al., 2017).

PRV-1 can be detected in association with the heart lesions during development of HSMI, underpinning the etiology of the disease (Finstad et al., 2012). Although the virus is mainly known to induce heart pathology, the virus replicates to high titer in its main target cell erythrocytes prior to the onset of cardiac lesions and this seems to be a consistent and important part of the pathogenesis (Finstad et al., 2014). It has also been shown that PRV can infect macrophages (Malik et al., 2019). The heart lesions eventually resolve after an HSMI outbreak. However, the virus is not cleared from the fish but causes a persistent infection (Garver et al., 2016).

Piscine Orthoreovirus-1 (PRV-1)

Piscine orthoreovirus (PRV) belongs to the family *Spinareoviridae* (formerly called *Reoviridae*), genus *Orthoreovirus* (International Committee of Taxonomy of Viruses, s.a.). The type species of genus *Orthoreovirus* is *Mammalian orthoreovirus* (MRV). MRV is associated with rare diseases in young humans, including celiac disease (Bouziat et al., 2017), and has been extensively used as a model virus in virology research. Based on phylogenetic relationship between MRV and PRV and similarities in secondary and tertiary protein structure, MRV has been used as a model to study PRV viral structure, viral replication and function of viral proteins (Schiff et al., 2007; Markussen et al., 2013). Gene segments, PRV proteins and functional properties are in the following described based on studies of MRV, assuming similarity (see table 2).

There are currently three known strains of PRV called PRV-1, PRV-2 and PRV-3, which are associated with different diseases in different salmonid species. PRV-1 is the causative agent of HSMI in Atlantic salmon (*Salmo salar*) (Wessel et al., 2020; Polinski et al., 2020). PRV-2 causes erythrocytic inclusion body syndrome (EIBS), a disease that can induce mass mortality

in Coho salmon (*Oncorhynchus kisutch*) (Takano et al., 2016). PRV- 3 is associated with heart inflammation in Rainbow trout (*Oncorhynchus mykiss*) (Olsen et al., 2015). The present study will only focus on PRV-1.

Virus Particle and Genome

The PRV virion is spherical, non-enveloped with a double-layered icosahedral capsid, like all *Spinareoviruses* (see figure 4) (Joklik, 1981). The genome is linear, segmented dsRNA, consisting of ten segments grouped into three different groups based on nucleotide sequence length: small (S1-S4), medium (M1-M3) and large (L1-L3) (see figure 3) (Palacios et al., 2010). Each segment codes for one protein, except for S1 which is bicistronic, i.e. codes for two proteins. Thus, the genome encodes a total of at least eleven proteins, however S2 and L2 are also potentially bicistronic encoding hypothetical proteins (Markussen et al., 2013).

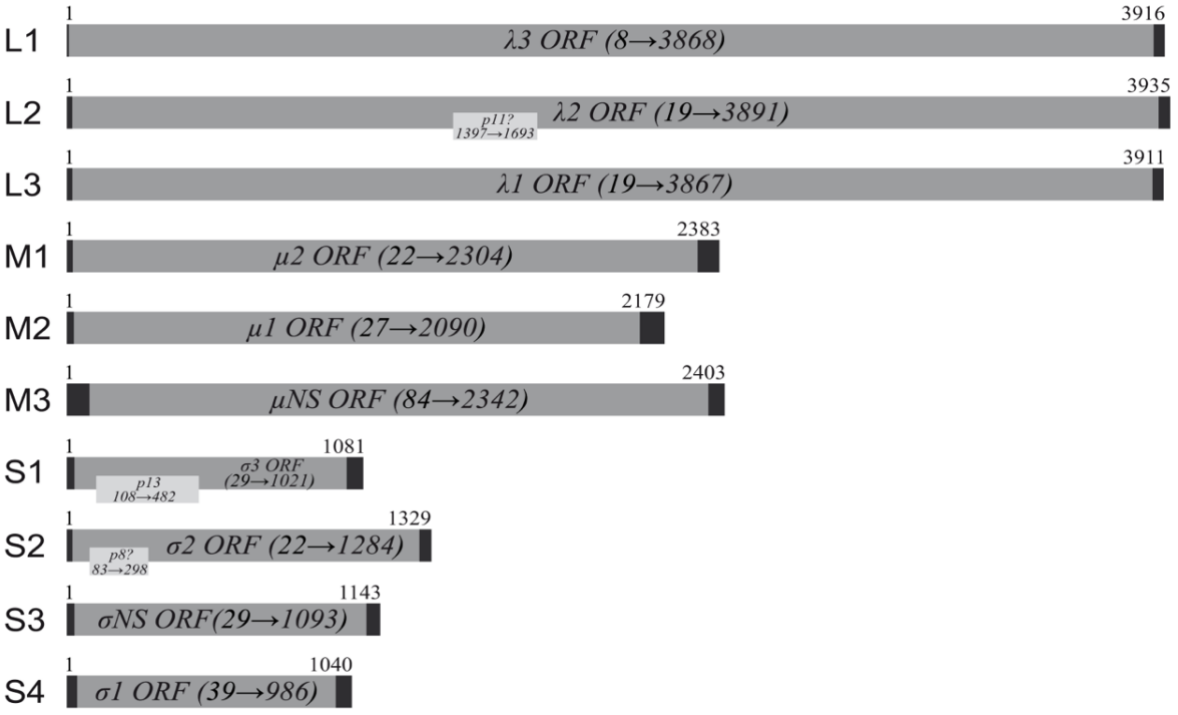


Figure 3. The PRV genome. Gene segments are assigned according to MRV. Open reading frames (ORFs) and putative encoded proteins are indicated by regions in grey, with start and end positions indicated. Non-translated regions (UTR's) at gene segment ends are indicated by regions in black. Gene segment S1 is bicistronic. S2 and L2 are possibly bicistronic. Markussen, T., Dahle, M., K., Tengs, T., Løvoll, M., Finstad, Ø., W., Wiik-Nielsen, C., R, Grove, S., Lauksund, S., Robertsen, B. & Rimstad, E. (2013). Sequence Analysis of the Genome of Piscine Orthoreovirus (PRV) Associated with Heart and Skeletal Muscle Inflammation (HSMI) in Atlantic Salmon (*Salmo Salar*). *PLoS ONE,* 8 (7). doi:10.1371/journal.pone.0070075.

Segment	Protein	Length bp (aa)	Structural/non-structural (predicted location/function)
S1	σ3	1081 (330)	Structural (outer capsid protein, zinc metalloprotein)
	p13	(124)	Non-structural (integral membrane protein, cytotoxic)
S2	σ2	1329 (420)	Structural (inner capsid protein, RNA binding)
	p8		Hypothetical protein
S3	σNS	1143 (354)	Non-structural (virus inclusion formation)
S4	σ1	1040 (315)	Structural (outer capsid protein, cell attachment protein)
M1	μ2	2383 (760)	Structural (NTPase, RNA triphosphatase, RNA binding)
M2	μ1	2179 (687)	Structural (outer capsid protein, membrane penetration)
M3	μNS	2403 (752)	Non-structural (virus inclusion formation)
L1	λ3	3916 (1286)	Structural and non-structural (RNA-dependent RNA polymerase)
L2	λ2	3935 (1290)	Structural (guanylyltransferase, methyltransferase)
	p11		Hypothetical protein
L3	λ1	3911 (1282)	Structural (helicase, NTPase, RNA triphosphatase)

Table 2. List of the PRV encoded proteins and their predicted location and functional properties from comparative studies with MRV. Modified table from: Markussen, T., Dahle, M., K., Tengs, T., Løvoll, M., Finstad, Ø., W., Wiik-Nielsen, C., R, Grove, S., Lauksund, S., Robertsen, B. & Rimstad, E. (2013). Sequence Analysis of the Genome of Piscine Orthoreovirus (PRV) Associated with Heart and Skeletal Muscle Inflammation (HSMI) in Atlantic Salmon *(Salmo Salar*). *PLoS ONE* 8 (7). doi: 10.1371/journal.pone.0070075.

The Outer Capsid

The outer capsid consists of the structural proteins σ1, σ3 and μ1 (see figure 4) (Joklik, 1981). Most of the outer capsid is built up of σ3 and μ1, which together form a heterohexamer structure coating the inner capsid (Joklik, 1981; Dryden et al., 1993). The outer capsid protein σ1 is associated with λ2 turrets in the inner capsid and makes up the spikes of the outer capsid (Dryden et al., 1993). In 1981, Lee et al. (cited in Dryden et al., 1993) characterized these spikes as cell attachment proteins. The cell attachment protein is involved in the virus' ability to bind host cells and facilitate entry (Guglielmi et al., 2006). The outer capsid protein μ1 is believed to play an important role in endosomal membrane penetration of the host cell (Nibert et al., 2005). After entering a host cell, it is demonstrated that PRV-1 σ3 binds dsRNA, which is most likely beneficial for modulating the host cell immune response (Wessel et al., 2015).

The Inner Capsid

The inner capsid, containing the viral genome, consists of the proteins σ2, μ2, λ1, λ2 and λ3 (Joklik, 1981). The two proteins λ1 and σ2 forms most of the inner core, while the other proteins are present in smaller amounts (Schiff et al., 2007). The protein λ3 is the RNA-dependent RNA polymerase, responsible for viral transcription and replication (Markussen et al., 2013). This protein can be categorized as both structural and non-structural, since the protein both have a functional property, but also is a consistent part of the inner capsid (Joklik, 1981; Markussen et al., 2013). The polymerase synthesizes non-polyadenylated mRNA from each dsRNA segment inside the core, which are then translocated to the cell cytoplasm. This way the virus ensures that the dsRNA is never exposed to cytoplasm, preventing activation of an antiviral response by the host cell. The protein μ2 is a minor structural RNA-binding protein, which displays NTPase and RTPase activities and possibly can act as a co-factor for RNA polymerase (Brentano et al., 1998; Markussen et al., 2013). The protein λ2 is a capping enzyme that contains guanylyltransferase and methyltransferase activities, necessary to generate the 5'-terminal cap on virally encoded mRNAs, creating stable, mature mRNA (Markussen et al., 2013). It is arranged in pentamers of twelve copies, forming turrets that project the inner capsid. The capped mRNA exists through this protein (Reinisch et al., 2000).

Non-structural Proteins

The non-structural proteins μNS, σNS and p13 are not a part of the viral particle itself. μNS and σNS are responsible for organizing viral factories in MRV (Becker et al., 2003). For PRV it has been shown that μNS forms viral factory-like structures to where σNS among λ1 and μ2 are recruited (Haatveit et al., 2016). The protein p13 is a cytotoxic integral membrane protein, however, the cytotoxic function has not been determined in vivo but it is likely to be involved in molecular interaction with the host (Key et al., 2013; Dhamotharan et al., 2019).

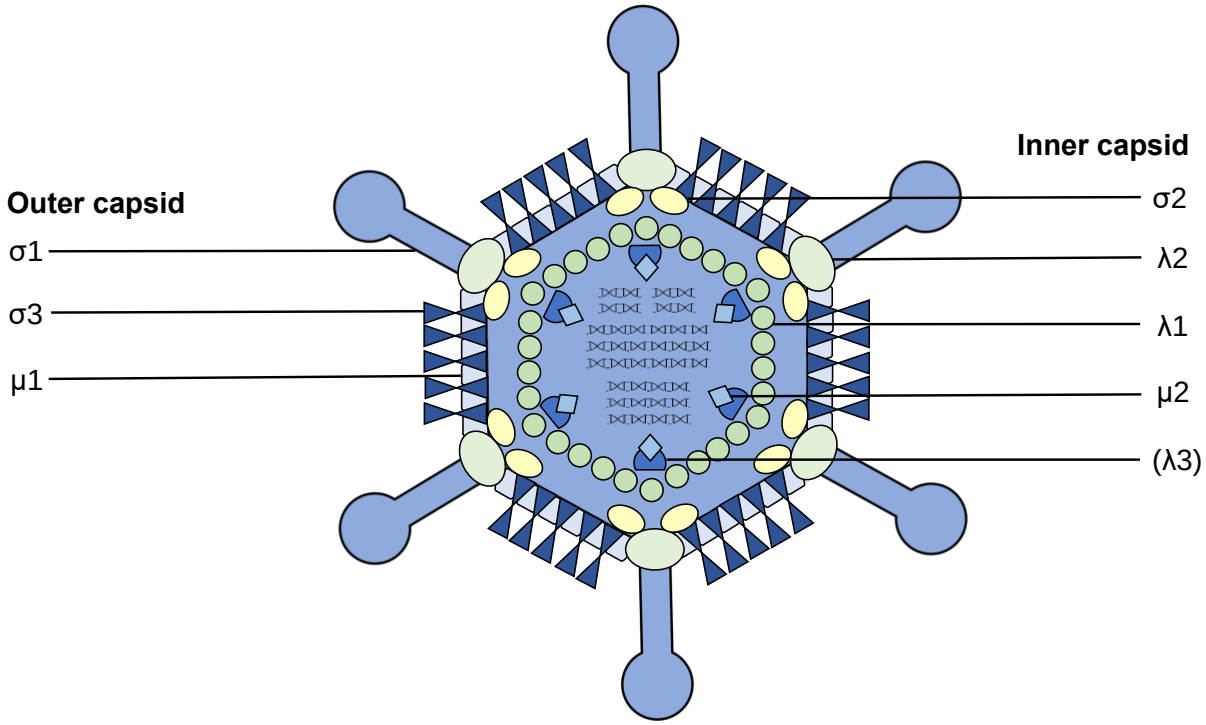


Figure 4. Schematic representation of PRV. Structural proteins are represented. The protein in parenthesis is categorized as both structural and non-structural. Modified figure from Vallejos-Vidal, E., Reyes-López, F. E., Sandino, A. M. & Imarai, M. (2022). Sleeping with the Enemy? The Current Knowledge of Piscine Orthoreovirus (PRV) Immune response Elicited to Counteract Infection. *Frontiers in Immunology*, 13. doi: 10.3389/fimmu.2022.768621.

PRV-1 Virulence Differences

PRV-1 is Ubiquitous while HSMI is not

Today PRV-1 is ubiquitous in Norwegian aquaculture and nearly all farmed Atlantic salmon undergo infection during a production cycle, but not all infected salmon develop HSMI (Wessel et al., 2017). This is reflected in the fact that PRV-1 can be detected in both clinically affected fish during HSMI outbreaks and clinically healthy fish with no reported disease as well as in wild stock (Garseth et al., 2012; Løvoll et al., 2012). PRV-1 is also widespread in other countries with large-scale Atlantic salmon farming, but the prevalence of HSMI appears to differ (Wessel et al., 2020). HSMI has been reported in Scotland, Chile and Canada (Ferguson et al., 2005; Godoy et al., 2016; Di Cicco et al., 2017). However, Canada have only reported a few cases of clinical disease, despite the virus being widespread in the farmed Atlantic salmon

population (Di Cicco et al., 2017, Polinski et al., 2019). Faroe Islands has also reported the virus, but no clinical disease so far (Rimstad, 2022).

Challenge trials performed with Norwegian and Canadian PRV-1 isolates have given different results regarding reproduction of HSMI. Challenge trials using Norwegian PRV-1 isolates have repeatedly reproduced HSMI lesions, whereas trials using Canadian isolates have produced little or no lesions consistent with HSMI (Polinski et al., 2019, Garver et al., 2016). The field observations and challenge trials indicated that unknown viral, host or environmental factors are important for development of HSMI (see figure 5).

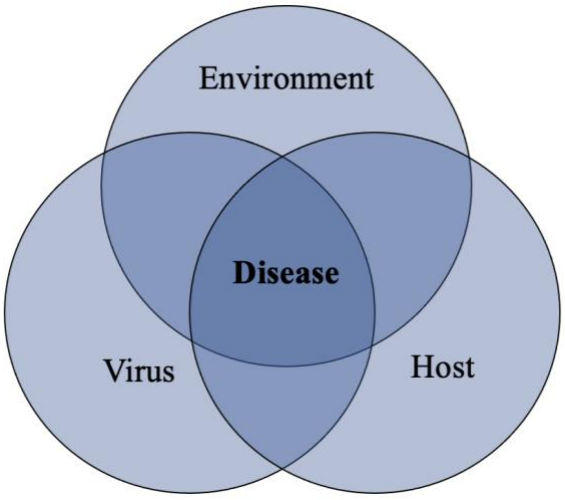


Figure 5. Epidemiologic triangle displaying the interaction of virus, host and environment for development of disease.

PRV-1 Virulence Differences Confirmed in Challenge Trial

Confirmation of the hypothesized difference in virulence was provided by a standardized dose challenge trial, which demonstrated that PRV-1 isolates differ in their ability to induce HSMI. The challenge trial compared three historical Norwegian isolates sampled before the first reported outbreak of HSMI (NOR-1988, NOR-1996 and NOR-1997), two modern Norwegian

field isolates (NOR-2018/NL and NOR-2018/SF) and one Canadian isolate (CAN 16-005ND) (Wessel et al., 2020).

The two Norwegian field isolates from 2018 both induced histopathological lesions in the heart and red skeletal muscle consistent with HSMI, while the three historical Norwegian and the Canadian isolate only resulted in mild lesions. Based on the findings, the isolates were characterized as high virulent (NOR-2018/NL and NOR-2018/SF) or low virulent (NOR-1988, NOR-1996, NOR-1997 and CAN 16-005ND) (Wessel et al., 2020). Segment S1, S4, M2, L1 and L2 were found to differentiate between the high and low virulent isolates, which have putatively linked these segments to virulence.

Mechanisms of Viral Evolution

Virulence differences derive from viral evolution. All viruses can change and evolve through altering of their genomes through point mutations and recombination. Point mutations occur when a single base pair is changed, added or deleted due to errors made by the polymerase. In general, RNA viruses exhibit high mutation rates, but the mutation rate of double stranded RNA viruses is lower than that of single stranded RNA viruses, and the length of the genome has a negative correlation to the mutation rate (Sanjuán & Domingo-Calap, 2016). An exchange of a single base pair alters the codon and can in some cases result in a codon encoding a different amino acid than originally. Other base exchanges do not alter the encoded amino acid and are therefore regarded as silent mutations. It is therefore natural that the nucleotide sequence often varies more than the amino acid sequence. A single base insertion or deletion can on the other hand change the entire reading frame and therefore every subsequent amino acid in the protein.

RNA recombination is a process where the RNA polymerase switches from one template RNA molecule to another during synthesis, resulting in a mRNA with mixed ancestry (Simon-Loriere & Holmes, 2011). Recombination often occurs for non-segmented RNA viruses with positive-sense RNA genome and can happen anywhere in the genome (McDonald et al., 2016). Segmented viruses, like PRV-1 and other *Spinareoviruses*, can also alter their genome through reassortment, where entire segments are exchanged between different viruses. Both recombination and reassortment can only occur during co-infections (see figure 6), where a cell is infected with two or more virus particles of the same virus (Simon-Loriere & Holmes, 2011).

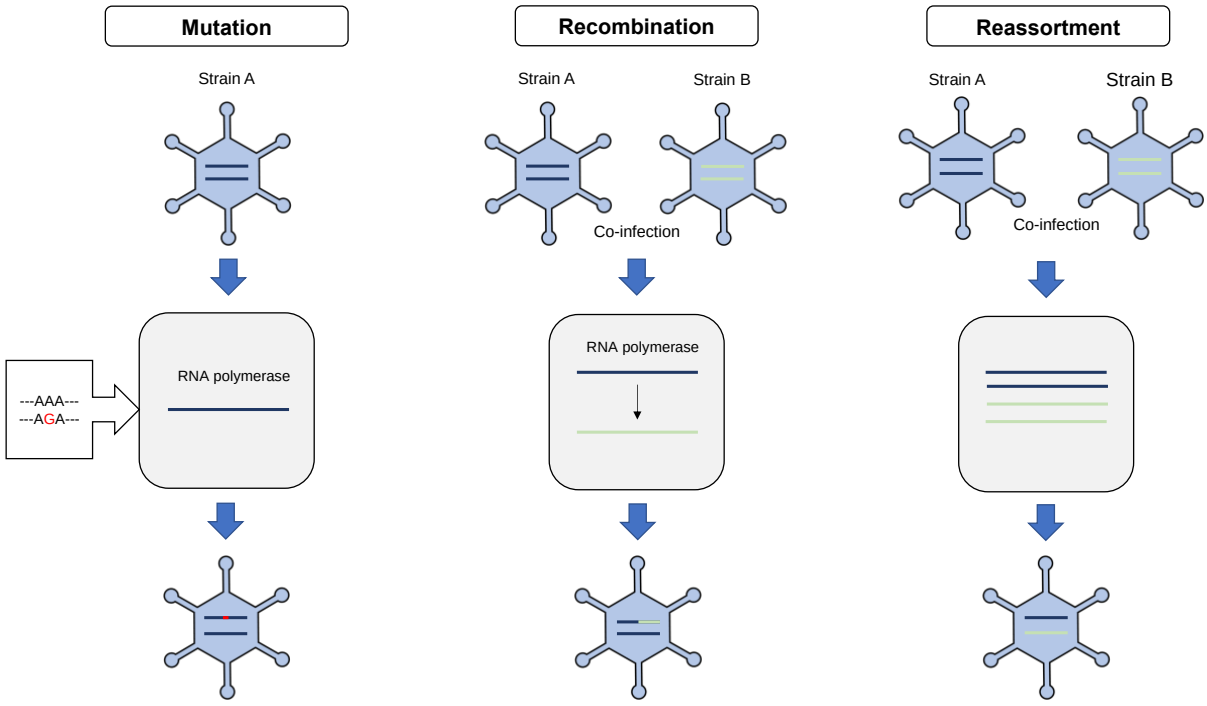


Figure 6. Mechanisms of viral genome evolution through point mutations, recombination and reassortment.

Most viral mutations confer no benefit to the virus or can be deleterious, while others can be advantageous. If the change is advantageous or indifferent for the virus, it will continue to replicate. This does not alter the other virus particles; they will also continue replication. Over time, selection pressure will optimize the virus' fitness to environmental changes, and it is a continuous way of evolution. In this way it is possible to get many different isolates that

originate from one virus particle. The rapid rate of viral mutation combined with natural selection allows viruses to quickly adapt to changes in their host environment. High fish density in sea pens in combination with the short generation time of viruses provides optimal conditions for viral replication and therefore facilities viral evolution.

Genogrouping: High and Low Virulent Isolates

Recently sampled PRV-1 isolates in Norwegian aquaculture have been identified and genogrouped through a system based on combinations of the five genomic segments putatively linked to virulence (S1, S4, M2, L1 and L2) (see table 3). Segment S1, M2, L1 and L2 were phylogenetically grouped into Group A or B, while S4 was analyzed with respect to single amino acid substitutions due to a conserved nucleotide sequence. Segments S1 and M2 displayed the same phylogenetic clustering, indicating a co-evolution for these two segments. The genogroups have been characterized as high virulent, low virulent or of unknown virulence based on comparison with PRV-1 reference isolates of known virulence (Vatne et al., 2021).

The genogrouping of the recently sampled PRV-1 isolates revealed that we today have a variety of both high virulent and low virulent isolates in Norway. Overall, both high virulent and low virulent isolates have been identified throughout the coast, but the high virulent isolates show a tendency to be more prevalent (Vatne et al., 2021).

Genogroup	Reference isolate(s)	S1 (σ3), M2 (μ1)	L1 (λ3)	L2 (λ2)	S4 (σ1)
High-1	NOR-2018/NL	B	A	A	D252N
High-2	NOR-2018/SF	B	A	A	V107A
Low-1	NOR-1997	B	B	B	
Low-2	NOR-1996, NOR-1988, CAN 16-005ND	A	A	A	
Unknown-1		B	B	A	
Unknown-2		B	A	A	
Unknown-3		B	B	A	D252N
Unknown-4		A	A	A	D252N

Table 3. Categorization of PRV-1 into genogroups. The grouping is based on the phylogenetic grouping (Group A or B) of S1 (σ3), M2 (μ1), L1 (λ3) and L2 (λ2) in addition to the presence of S4 (σ1) amino acid substitution V107A or D252N. Eight genogroups have currently been identified, including High-1, High-2 and Low-1, Low-2 with reference isolates of known virulence. In addition, four genogroups of unknown virulence have been identified (Unknown 1–4). Vatne, N., Stormoen, M., Lund, M., Devold, M., Rimstad, E. & Wessel, Ø. (2021). Genetic grouping and geographic distribution of Piscine orthoreovirus-1 (PRV-1) in farmed Atlantic salmon in Norway. *Veterinary Research,* 52. doi: 10.1186/s13567-021-01000-1.

Limited Sequence Data during the Emergence of HSMI

The growing amount of sequence data from recently sampled PRV-1 isolates shows that we today have a variety of both high virulent and low virulent isolates in Norway. A lot of the variation appears to be a consequence of reassortment events from the ancestral isolates prior to 1999 (NOR-1988, NOR-1996 and NOR-1997). However, there is only a limited number of isolates from before the first reported outbreak of HSMI in 1999 and there is a lack of sequence information from isolates in the recent years following 1999. Description of isolates from this critical time at which HSMI emerged have the potential to explain the evolution of PRV-1 isolates from before the emergence of HSMI to the isolates that circulates in Norway today.

Aims of Study

Main goal

The principal aim of the present study was to sequence PRV-1 isolates that circulated in farmed Norwegian Atlantic salmon before and after the first reported clinical outbreak of HSMI in 1999.

Sub goals

1) Screen and identify PRV-1 positive samples from archived material collected before and after 1999.

2) Sequence the PRV-1 positive samples, targeting segment S1, S4, M2, L1 and L2 because of their putative link to virulence.

3) Compare the sequenced isolates to existing reference isolates of known virulence (NOR-1988, NOR-1996, NOR-1997, NOR-2018/NL, NOR-2018/SF and CAN 16-005ND) through phylogenetic analysis.

Materials and Methods

Sample Overview

The study included 25 separate sample sets (see table 4) comprising a total of 204 single samples (see appendix 1). The samples were collected from 1987 to 2008 and included both whole fish, tissue, plasma and serum samples. Some samples were sampled in conjunction with disease outbreaks, others were sampled for unknown reasons. The samples had been preserved through freezing at different temperatures at Norwegian University of Life Sciences (NMBU) and the Norwegian Veterinary Institute (NVI) (see appendix 2).

Sample set	Year	Number of samples	Sample type	Purpose of sampling
1	1987	33	Serum	Unknown
2	1988	2	Plasma	Disease outbreak
3	1988	5	Unknown	Disease outbreak
4	1995	5	Plasma	Disease outbreak
5	1996	9	Plasma	Disease outbreak
6	1996	7	Plasma	Disease outbreak
7	1996	16	Heart, spleen	Suspected ISA
8	1997	4	Plasma	Disease outbreak
9	1998	12	Kidney, spleen	Suspected ISA
10	1998	20	Kidney, spleen	Suspected ISA
11	1998	2	Plasma	Disease outbreak
12	1999	10	Plasma	Disease outbreak
13	2001	5	Plasma	Disease outbreak
14	2004	6	Heart, head kidney, spleen	Suspected HSMI
15	2004	12	Heart, head kidney, spleen	Suspected HSMI
16	2004	3	Plasma	Disease outbreak
17	2004	12	Heart, head kidney, spleen	Suspected HSMI
18	2004	6	Heart, head kidney, spleen	Suspected HSMI
19	2005	6	Heart, head kidney, spleen	Suspected HSMI
20	2005	3	Heart, head kidney, spleen	Suspected HSMI
21	2005	5	Plasma	Disease outbreak
22	2007	9	Heart, head kidney, spleen	Suspected HSMI
23	2007	3	Heart, head kidney, spleen	Suspected HSMI
24	2007	6	Heart, head kidney, spleen	Suspected HSMI
25	2008	3	Heart, head kidney, spleen	Suspected HSMI

Table 4. Sample set overview. Overview of sample sets, number of samples in each sample set, year sampled, type of samples and the purpose of sampling.

Method Overview

The study was divided into three parts (see figure 7). The first part of the study aimed at identifying PRV-1 positive samples. To achieve that, RNA was isolated from all the samples. Some samples were stored as preprepared tissue samples, while others had to be retrieved from whole fish through dissection before RNA isolation. The samples were initially screened for PRV-1 through RT-qPCR. In the second part, cDNA was synthesized, amplified by PCR and separated through agarose gel electrophoresis. Thereafter, the amplicons were extracted and sequenced using Sanger sequencing. The third and final part of the study consisted of sequence analysis and phylogenetic analysis.

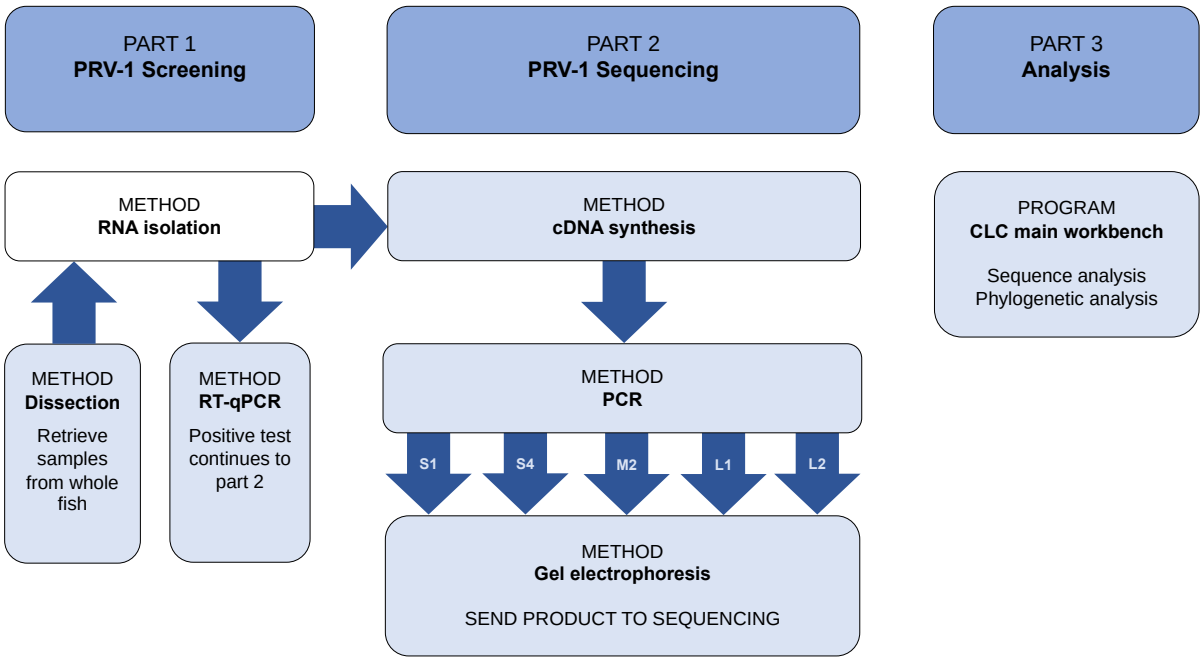


Figure 7. Method overview. The study was divided into three parts, consisting of PRV-1 screening, PRV-1 sequencing and analysis.

PART 1: PRV-1 SCREENING

RNA Isolation

Preparation for RNA Isolation: Dissection of Whole Fish

The study material included 22 whole fish (see appendix 3), stored at –20 °C, that needed to be dissected to prepare tissue samples for RNA isolation. Samples from the heart, head kidney and spleen were retrieved from each fish and fixated in RNAlater (Invitrogen, Waltham, Massachusetts, USA #AM7021) before proceeding with RNA isolation. To reduce the risk of cross contamination, the worktable was cleaned, disposable equipment such as sheets and gloves changed and reusable equipment such as scalpel holders and tweezers disinfected between dissection of different fish. The retrieved samples were stored at –80 °C.

Principle of RNA Isolation

In general, RNA isolation consists of cell lysis and homogenization followed by separation and retrieval of RNA. There are many techniques for RNA isolation, targeting total RNA or specific variants of RNA. The methods used in the present study are isolation of total RNA through organic extraction methods and filter-based spin basket formats. The first step requires effective cell lysis to release the RNA followed by homogenization. This can be achieved using mechanical methods, including high pressure and bead-beating, or non-mechanical methods, including physical, chemical or enzymatic methods (Islam et al., 2017). After homogenization, organic extractions methods, for example with phenol-chloroform-based extractions, use a three-phase partition to separate the RNA. Filter-based methods do not generate the three-phase separation but use silica-membrane based spin technology where spin columns capture RNA (Thermo Fisher, s.a.a).

RNA Isolation: Tissue

From tissue samples, preserved in RNAlater or through freezing at -80 °C, total RNA was isolated using RNeasy Mini QIAcube Kit (Qiagen, Hilden, Germany #74116). Pinpoint sized tissue pieces were dissected from the original samples and transferred to individual sample tubes. To reduce the risk of cross contamination, disposable equipment such as scalpels and petri dishes were exchanged between preparation of samples from different sample sets. Subsequently, a solution of 98 % lysis buffer RLT (Qiagen) and 2 % dithiothreitol (DTT) (Bio-Rad, Hercules, California, United States #1610611) and steel beads (Qiagen #69989) were added to each sample before cell lysis and homogenization using TissueLyser II (Qiagen) at 25 Hz for 8 minutes. After homogenization, the samples were centrifuged at 13 000 G for 3 minutes using Centrifuge 5424 R (Eppendorf, Hamburg, Germany). The supernatant was collected, and the following RNA isolation procedure was automated in QIAcube (Qiagen), following the manufacturer's instructions. The total RNA was eluted in 30 µl RNase free water. The concentration of the extracted RNA was quantified and the purity assessed using mySPEC spectrophotometer (VWR, Randor, Pennsylvania, United States). After extraction, the RNA samples were stored at -80 °C.

RNA Isolation: Plasma and Serum

Plasma and serum are cell-free fluids, and thus require a different protocol for RNA isolation compared to the tissue samples described above. Two different protocols were used, QIAmp Viral RNA mini kit and Trizol LS protocol.

QIAamp Viral RNA Mini Kit

From plasma and serum, total RNA was mainly isolated using QIAamp Viral RNA Mini kit (Qiagen #52904), according to the manufacturer's instructions. RNA was isolated from 10 µL

plasma diluted in 130 µL Dulbecco's phosphate buffered saline (DPBS) (Thermo Fisher Scientific, Waltham, Massachusetts, USA #14190094). Carrier RNA was used, which is useful if there is a small amount of RNA in the sample because it enhances binding of RNA to the membrane (Qiagen, 2020). The isolated RNA was eluted in 50 µL RNase free water and stored at –80 °C. The RNA concentration was not quantified in the plasma and serum samples because carrierRNA influences the measurement (Qiagen, 2020).

Trizol LS protocol

Total RNA was also isolated using a combination of Trizol LS protocol and RNeasy Mini QIAcube Kit (Qiagen #74116). This protocol was performed on a total of four samples (see table 18) because of their special interest due to their year of origin and after unsuccessfully PCR amplifications following isolation with QIAamp Viral RNA Mini kit. In brief, 50 µL of plasma was diluted in 150 µL DPBS (Thermo Fisher Scientific), followed by cell lysis with 600 µL Trizol LS (Invitrogen #15596018) and homogenization by pipetting. After homogenization, the samples were incubated at room temperature for 5 minutes. Chloroform was added and the mixture was shocked manually for 15 seconds. The samples were then incubated at room temperature for 2-3 minutes before centrifugation at 4 °C at 12 000 g for 15 minutes, separating the phases. The aqueous phase was collected and proceeded with the RNeasy Mini QIAcube Kit (Qiagen) as described by the manufacturer. The isolated RNA was eluted in 50 µL RNase free water and stored at –80 °C.

Quantitative Reverse Transcription PCR (RT-qPCR)

Principle of RT-qPCR

Quantitative reverse transcription PCR (RT-qPCR) is a method for detection and quantification of specific RNA molecules in a sample. The method consists of reverse transcription of the total RNA or mRNA to complimentary DNA (cDNA) by the enzyme reverse transcriptase, followed by amplification and detection of the cDNA using quantitative PCR (qPCR), also referred to as real-time PCR. RT-qPCR can be performed in one step or two step assays, depending on whether the reverse transcription and qPCR are performed combined or as separate reactions (Thermo Fisher, s.a.b).

The amplification consists of three steps, where the first step denatures the template, the second step allows optimal annealing of primers, and the third step permits the DNA polymerase to bind to the cDNA template and synthesize the PCR product (Lorenz, 2012). During each PCR cycle, the quantity of DNA can be measured in real-time by using a variety of fluorescent chemistries, for example through hydrolyzing probes. The fluorescence accumulates as the cycles are repeated and can be detected as a visible curve on software, representing real time amplification (SigmaAldrich, s.a.). Cycle threshold (Ct) is defined as the number of cycles required for the fluorescent signal to cross a set threshold and are inversely logarithmic to the amount of target nucleic acid in the sample.

One step RT-qPCR: Tissue

The isolated RNA from tissue was analyzed through a probe-based one step RT-qPCR, targeting RNA of PRV-1 σ3 and the elongation factor 1αB (EF1αB) of Atlantic salmon (see table 5). The EF1αB was included as a reference gene to control that the samples did not contain RNases or inhibitors and that the RNA was not degraded. External, positive and negative

controls were included. Diethyl pyrocarbonate (DEPC) water was used as a negative control to reveal contamination of the qPCR mix, and PRV-1 virus as a positive control to verify that the reagents were working, optimal conditions and consistency between different runs.

Primers/probes	**Primer sequence (5' → 3')**	**Accession number**
PRV-1 σ3 S1-659 Fwd	TGCGTCCTGCGTATGGCACC	GU994022
PRV-1 σ3 S1-801 Rev	GGCTGGCATGCCCGAATAGCA	
PRV-1 σ3 S1-693 Probe	*FAM*-ATCACAACGCCTACCT-*MGBNFQ*	
EF1αB Fwd	TGCCCCTCCAGGATGTCTAC	BG933897
EF1αB Rev	CACGGCCCACAGGTACTG	
EF1αB Probe	*6FAM*-AAATCGGCGGTATTGG-*MGBNFQ*	

Table 5. RT-qPCR primers and probes. The quencher and reporter dye of the probes are in italic. Accession numbers in GenBank (NCBI) are shown in the right column.

RT-qPCR was performed using the Brilliant III Ultra-Fast qRT-PCR master mix kit (Agilent Technologies, Santa Clara, California, United States #600885). The total input of isolated total RNA per reaction was set to standard 100 ng (5 µL of 20 ng/µL). The template RNA was diluted in DEPC water to obtain the correct RNA concentration, centrifuged and denatured at 95 °C for 5 minutes in Veriti 96-well Thermal Cycler (Thermo Fisher Scientific) to separate the strands of the dsRNA. The same type of thermal cycler was used throughout qPCR, cDNA synthesis and PCR. The master mixes of both PRV-1 σ3 and EF1αB were prepared after the protocol described in table 6 and table 7.

Reagent	**Final concentration**
Brilliant III Ultra-Fast QRT-PCR Master Mix (2X) (Agilent Technologies)	1X
RT/RNase Block (Agilent Technologies)	1:20*
Reference dye (1 mM) (Agilent Technologies)	0,03 µM
DTT (100 mM) (Agilent Technologies)	1000 µM
PRV-1 σ3 S1-659 Fwd (10 µM)	0,4 µM
PRV-1 σ3 S1-801 Rev (10 µM)	0,4 µM
PRV-1 σ3 S1-693 Probe (10 µM)	0,3 µM

Table 6. RT-qPCR reagents targeting PRV-1 σ3 and their final concentration for tissue samples. The final concentrations are based on earlier assays. *RT/RNase Block are provided as a ratio.

Reagent	Final concentration
Brilliant III Ultra-Fast QRT-PCR Master Mix (2X) (Agilent Technologies)	1X
RT/RNase Block (Agilent Technologies)	1:20*
Reference dye (1 mM) (Agilent Technologies)	0,03 μM
DTT (100 mM) (Agilent Technologies)	1000 μM
EF1αB Fwd (10 μM)	0,3 μM
EF1αB Rev (10 μM)	0,3 μM
EF1αB Probe (10 μM)	0,2 μM
RNase free water (Qiagen)	1:100*

Table 7. RT-qPCR reagents targeting EF1αB and their final concentration for tissue samples. The final concentrations are based on earlier assays. *RT/RNase Block and RNase free water are provided as ratios.

All samples were run in duplicates, in addition to duplicates of positive and negative controls, on AriaMx 96 well plates (Agilent Technologies, #401419). qPCR was performed using AriaMx Real-time PCR System (Agilent Technologies) using the protocol illustrated in table 8.

Step	Temperature/time	Cycle(s)
RT	50 °C /10 min	1X
Denaturation	95 °C/3 min	1X
Denaturation	95 °C/5 sec	40X
Annealing	60 °C/10 sec	

Table 8. RT-qPCR cycle parameters for tissue samples.

The results were analyzed using the software AriaMx (Agilent). The fluorescence threshold was set to 0.01 to assess which samples were positive for PRV-1. The average Ct value was calculated for the duplicate Ct values, and samples were defined positive if both parallel samples had a Ct value < 35.

One step RT-qPCR: Plasma and Serum

Detection of PRV-1 RNA from plasma and serum samples were performed by a probe-based one step RT-qPCR, targeting PRV-1 σ3. Primers and probes are listed in table 5.

RT-qPCR was performed using QIAGEN OneStep RT-PCR Kit (Qiagen #210212). The reagent mixture was prepared with the conditions described in table 9. The plasma and serum samples were centrifuged, preheated in a thermal cycler at 95 °C for 5 minutes and then centrifuged again. The reagent mixture and template were added to AriaMx 96 well plates (Agilent Technologies). PRV-1 σ3 were set up as duplicates, in addition to external controls as described under the tissue protocol. The plates were centrifuged, and then analyzed in AriaMx Real-time PCR System (Agilent Technologies), following the protocol in table 10. The results were analyzed as described under RT-qPCR for tissue samples.

Reagent	**Final concentration**
QIAGEN OneStep RT-PCR Enzyme Mix (Qiagen)	1X
QIAGEN OneStep RT-PCR Buffer (5X) (Qiagen)	1X
dNTP Mix (10 mM) (Qiagen)	400 μM
Reference dye (Agilent Technologies)	0,03 μM
$MgCl_2$ (Invitrogen)	1260 μM
PRV-1 σ3 S1-659 Fwd (10 μM)	0,4 μM
PRV-1 σ3 S1-801 Rev (10 μM)	0,4 μM
PRV-1 σ3 S1-693 Probe (10 μM)	0,3 μM
RNase free water (Qiagen)	17:100*

Table 9. RT-qPCR reagents targeting PRV-1 σ3 and their final concentration for plasma and serum samples. *RNase free water is provided as ratio.

Step	**Temperature/time**	**Cycle(s)**
RT	50 °C /30 min	1X
Denaturation	95 °C/15 min	1X
Denaturation	94 °C/15 sec	40X
Annealing	60 °C/30 sec	
Extension	72°C/30 sec	

Table 10. RT-qPCR cycle parameters for plasma and serum samples.

PART 2: PRV-1 SEQUENCING

Selection Criteria for Sequencing

A subset of samples was subjected to sequencing based on the results from RT-qPCR and RNA quality. A total of 24 samples, including two duplicates isolated with different RNA isolation protocols described earlier, were selected based on the lowest Ct values (see table 18). For tissue samples originating from the same sample set where all samples had similar Ct values, the RNA concentration was taken into consideration. The highest RNA concentration was then prioritized over the lowest Ct value. A few selected samples had Ct values > 30 but were chosen because they were sampled from a period of special interest, years before and close to 1999. These samples had the lowest Ct value from that specific time period.

cDNA Synthesis

Principle of Single Stranded cDNA Synthesis from dsRNA

On the molecular level cDNA can be single stranded or double stranded. For amplifying a gene sequence through PCR, a single stranded cDNA is sufficient. Through reverse transcription, using the RNA as a template, a cDNA strand is transcribed. The synthesis can be summed up in three main steps all facilitated in a thermal cycler: primer annealing, DNA polymerization and enzyme deactivation (Thermo Fisher, s.a.c).

cDNA Synthesis

cDNA synthesis was performed on all samples selected for sequencing, using Superscript III (Invitrogen #18080085). The primer chosen for the synthesis was random hexamers (Thermo Fisher Scientific #N8080127). A total of 20 µl cDNA was synthesized for all samples with a concentration of 50 ng/µl.

Firstly, mixture 1 was prepared in two steps (see table 11). The first step consisted of preparation and denaturation of the RNA template. For tissue samples, the total input RNA was set to 1000 ng and DEPC water was added to obtain a total reaction volume of 13 µL. For plasma samples, with unknown RNA concentration, 11 µL undiluted template was used. The RNA template was denatured at 95°C for 5 minutes. The second step consisted of adding random hexamers and dNTPs to each template mixture, before incubating at 65 °C for 5 minutes. After incubation, mixture 2 (see table 12) was added to mixture 1. The RNA mixtures were then incubated in a thermal cycler at 25 °C for 5 min, 50 °C for 60 min, 70 °C for 15 min and then cooled down to 4 °C before stored at –20 °C.

		Tissue	**Plasma and serum**
	Reagent	**Added volume**	**Added volume**
Step 1	RNA template	1000 ng	11 µL
	DEPC water	x µL	-
Step 2	Random hexamers (250 ng) (Thermo Fisher Scientific)	1 µL	1 µL
	dNTP Mix (10 mM) (Qiagen)	1 µL	1 µL
	Total volume	13 µL	13 µL

Table 11. Mixture 1. Volume of added random hexamers, dNTPs, template and DEPC water for each tissue and plasma/serum sample. In the tissue samples, the total input RNA was set to 1000 ng, and the volume to be added was calculated based on earlier calculated RNA concentrations. The volume DEPC water was then added to obtain a total volume of 13 µL.

Reagent	**Volume**
5X First-strand buffer (Invitrogen)	4 µL
DTT (Invitrogen)	1 µL
RNaseOUT (Invitrogen)	1 µL
Superscript III Reverse Transcriptase (Invitrogen)	1 µL

Table 12. Mixture 2. Overview of the reagents added to each sample after incubation at 65°C for 5 minutes.

Polymerase Chain Reaction (PCR)

Principle of PCR

Polymerase chain reaction (PCR) is conducted in a thermal cycler to amplify DNA. The DNA goes through denaturation, annealing and extension stepwise in repeated cycles with the goal of amplifying itself. Amplification is necessary to generate enough DNA for detection in gel electrophoresis and sequencing.

PCR

The five genomic segments were amplified by PCR, using kit Platinum™ SuperFi™ DNA Polymerase (Invitrogen #12351010). PCR was conducted on all five selected genomic segments, using the primers listed in table 13. The segments S1, L1 and L2 had to be divided into two parts to retrieve high quality sequences.

Target gene	Primers	Primer sequence (5' → 3')	Accession number
PRV-1 S1 (part 1)	S1 orfF	GATCAAAGACTTCTGTACGTGAAACC	MK675858.1
	S1 intR 611	GTTCAGCTCCTCCATGTCTTC	MW260151.1
PRV-1 S1 (part 2)	S1 intF 404	CGCGGTTCAAACGACAGACC	MW260151.1
	S1 orfR	GATGAATAAGACCTCCTTCCCC	MW260151.1
PRV-1 S4	S4 orfF	ATGCATAGATTTACCCAAGAAGACCATG	MW260154.1
	S4 orfR	CTAGATGATGATCACGAAGTCTCCA	MW260154.1
PRV-1 M2	M2 orfF	ATGGGTAACTATCAGACAAGTAACAACC	MW260149.1
	M2 orfR	CTATTTTTGGCCTCGACGTGAGTC	MW260149.1
PRV-1 L1 (part 1)	L1 orfF	ATGGAGAAACCTAAAGCGCTTGTCAA	MW260145.1
	L1 intR 2284	GACGACCATCATTAGTGGGTAG	MW260145.1
PRV-1 L1 (part 2)	L1 intF 2151	CCTGATGCACTGGTTACCTAAT	MW260145.1
	L1 orfR	CTAAAAATCGGACACCATCCAATTAAGG	MW260145.1
PRV-1 L2 (part 1)	L2 orfF	ATGGCTACGCTTTATGGGCTACG	MW260146.1
	L2 intR 2497	GCAGGACCACTACCAACATCAA	MW260146.1
PRV-1 L2 (part 2)	L2 intF 2218	ACCATCACCATCTCACGTAATC	MW260146.1
	L2 orfR	CTATGGCAACTTCAAGAACGTGTA	MW260146.1

Table 13. Primers used for PCR and sequencing targeting PRV-1 segment S1, S4, M2, L1 and L2. Accession numbers in GenBank (NCBI) are shown in the right column.

The total reaction volume was set to 50 µL, comprising 47 µL reagent mixture and 3 µL template. The mixture used in each reaction contained the reagents listed in table 14. PCR was conducted in a thermal cycler using the cycle parameters described in table 15. Separate protocols were used for each segment to optimize temperature, time and number of cycles.

Reagent	**Final concentration**
Platinum SuperFi DNA polymerase (2 U/µL) (Invitrogen)	0,02 U/µL
SuperFi Buffer (5X) (Invitrogen)	1X
SuperFi GC Enhancer (5X) (Invitrogen)	1X
dNTP Mix (10 mM) (Qiagen)	200 µM
Primer Fwd (10 µM)	0,5 µM
Primer Rev (10 µM)	0,5 µM
RNase free water	41:100*

Table 14. Reagents for PCR and their final concentration. *RNase free water is provided as a ratio.

	Temperature/time (cycles)				
	S1	**S4**	**M2**	**L1**	**L2**
Initial denaturation	98 °C/30 sec (1X)	98 °C/30 sec (1X)	98 °C/30 sec (1X)	98 °C/30 sec (1X)	98 °C/30 sec (1X)
Denaturation	98 °C/10 sec	98 °C/10 sec	98 °C/10 sec	98 °C/10 sec	98 °C/10 sec
Annealing	58 °C/10 sec	60 °C/10 sec	58 °C/10 sec	56 °C/10 sec	56 °C/10 sec
Extension	72 °C/30 sec (35X)	72 °C/30 sec (40X)	72 °C/1 min (40X)	72 °C/1,5 min (42X)	72 °C/2 min (42X)
Final extension	72 °C/5 min (1X)	72 °C/5 min (1X)	72 °C/5 min (1X)	72 °C/5 min (1X)	72 °C/5 min (1X)

Table 15. PCR cycle parameters used for each genomic segment.

Agarose Gel Electrophoresis

Principle of Agarose Gel Electrophoresis

Agarose gel electrophoresis is a method for separating macromolecule fragments based on their size and/or charge to distinguish and identify segments of known size. The macromolecules could be proteins or nucleic acids, as in this study (DNA). By casting a gel block, loading it with PCR product and then running an electric current through the gel, negatively charged molecules will migrate towards the positive electrode, at a rate determined by fragment length. Many variable methods of staining exist, which are needed for detection. A marker with molecules of known fragment lengths is usually run through the gel at the same time as the PCR product. After staining and separation the gel is put in an imager and the PCR product can be detected as bands, identified (owing to the marker as reference) and then cut out before cleansing of PCR product (Bio-Rad, s.a.).

Agarose Gel Electrophoresis

To visualize the amplified PCR products, agarose gel electrophoresis was conducted. A 1% agarose gel was made, by dissolving agar powder (Sigma Aldrich, Saint-Louis, Missouri, USA #A9539) in 1X Tris-acetate-EDTA (TAE) buffer (Thermo Fisher Scientific #B49) and adding SYBR Safe DNA Gel Stain (Invitrogen, #S33102). The gel was poured into an appropriate casting tray and solidified for 30 minutes. The solidified agarose gel was put in a gel electrophoresis chamber and covered with 1X TAE buffer. PCR products were mixed with either Orange DNA Loading Dye (6X) (Thermo Fisher Scientific #R0631) or TriTrack DNA Loading Dye (6X) (Thermo Fisher Scientific #R1161). The first well was then filled with GeneRuler 1kb DNA ladder (Thermo Fisher Scientific #SM031) or GeneRuler 1kb Plus DNA ladder (Thermo Fisher Scientific #SM1331), as a molecular weight marker, followed by duplicates of each PCR product with empty wells between different samples. The gel was run

at 95 Volts for 60 minutes in Model 200/2.0 Power Supply (Bio Rad). The gel was imaged using ChemiDoc XRS+ Imager (Bio-Rad) (see figure 8) and then examined by the Safe Imager Blue-Light Transilluminator (Invitrogen), with care taken to not expose the gel to UV light. Sufficiently amplified PCR products, visualized as single strong bands of appropriate size on the gel, were cut out from the gel with separate, sterile scalpels and stored at -20° C before cleansing of PCR product. For simplification, the term classic PCR will be used to describe PCR and agarose gel electrophoresis hereafter.

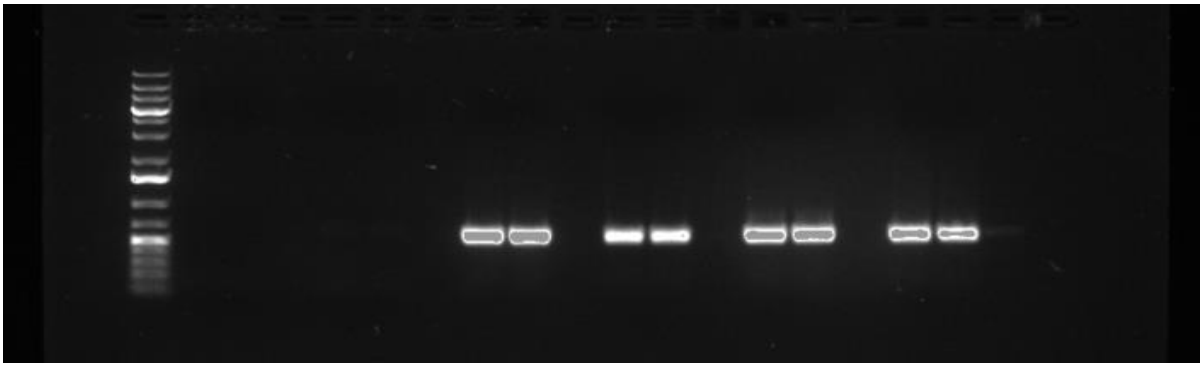


Figure 8. Agarose gel electrophoresis. The figure illustrates an example of agarose gel with distinct bands. GeneRuler 1kb Plus DNA was applied as reference.

Cleansing of PCR Products

Principle of Cleansing of PCR Products

Gel bands from agarose gel electrophoresis containing PCR products need to be extracted and cleansed before sequencing. This procedure is done in steps: First, the gel slice is solubilized. Then the DNA is bound to the silica membrane, before washing and drying of the membrane. DNA is then eluted, and concentration is measured.

Cleansing of PCR Products

PCR products, evaluated on the gel as sufficiently amplified, were purified using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany #740609.50), prior to sequencing. Buffer NTI, a water bath holding the temperature of 50 °C and a vortex were used to solubilize the gel slices. Each gel band was weighed to determine the volume of buffer NTI to be added, following the manufacturer's description. Next, binding of the cDNA to the silica membrane was done by passing the solvent through the membrane of a column by centrifugation for one minute at 11000 G. Washing was performed by adding buffer NT3 to the column and centrifugation in repeated steps. Thereafter drying by centrifugation. The cDNA was then eluted by adding buffer NE to the column and incubating for one minute at room temperature. The column was centrifuged, and the eluted cDNA was collected. The samples were stored at 4 °C until submission for sequencing.

DNA Sequencing

DNA sequencing is the process of determining the nucleic acid sequence. There exist different methods of sequencing, but only Sanger sequencing will be further described as this method was used in this study.

Principle of Sanger Sequencing

There are four nitrogen base types in DNA in the form of deoxynucleotide triphosphates (dNTPs). In Sanger sequencing dideoxynucleotide triphosphates (ddNTPs) are a part of the reaction, these are different from dNTPs. If a ddNTP binds a DNA strand, no further dNTPs can be added and it therefore functions as a DNA chain elongating inhibitor. ddNTPs of the four nitrogen bases are marked with fluorescent labels, one color for each nitrogen base. dNTPs, ddNTPs with labels, DNA polymerase and primers are mixed with PCR-amplicon template.

Primers bind the template and DNA polymerase elongates DNA. However, segments of different lengths are produced since the binding of dNTPs and ddNTPs is random. It can be visualized a color in every sequence segment since ddNTPs are labeled with fluorescent light. When separating the DNA molecules (segments) in order of size, the ddNTP of each nitrogen base can be lit up, which creates a ladder of segments, differing by one nucleotide. With every segment illuminating a color this ladder can be visualized as a chromatogram (Hawkins, 2017). Sanger sequencing does not give complete sequences, typically lacking some sequence in the beginning and in the end. This is taken into consideration in the following analysis.

Sanger Sequencing

Sanger sequencing through Eurofins genomics was performed to identify the PRV-1 sequences from the samples. The primers used for sequencing were the same as for PCR (see table 13). For samples with unsatisfactory quality, synthesis of supplementary primers was ordered for internal sequencing (see table 16).

Target gene	Primers	Primer sequence (5' → 3')	Accession number
PRV-1 M2	M2 intF_657	GCTCTGCAAGTGATTCGAGCTA	MW260149.1
PRV-1 L1 (part 1)	L1_intR_1496	GAGCTCTTAGCGCCAGATTT	MW260145.1
	L1_intR_815	AAGCATCATACGTACCGTTGG	MW260145.1
PRV-1 L2 (part 2)	L2 intF_700	TTCGACAATCCAACTTACGCT	MW260146.1

Table 16. Primers used for internal sequencing. Accession numbers in GenBank (NCBI) are shown in the right column.

PART 3: SEQUENCE ANALYSIS

Sequence and Phylogenetic Analysis

The software used for sequence and phylogenetic analysis were CLC Main Workbench (Qiagen) and MEGA X (available at www.megasoftware.net).

Reference Isolates

Six isolates (NOR-1988, NOR-1996, NOR-1997, NOR-2018/NL, NOR-2018/SF and CAN 16-005ND) with previously determined virulence, were used as references in the analysis.

Sequence Analysis

The first part included quality control of the raw nucleotide sequences received from Eurofins with respect to trace and quality data. Quality control was done through visualizing the sequences as four-color chromatograms, where each of the nitrogenous bases was indicated by a specific color (adenine (green), cytosine (blue), guanine (black), and thymine (red)). A high-quality chromatogram was defined as a chromatogram with evenly spaced peaks with a single color, similar peak amplitude and minimal baseline noise. Sequences with poor-quality chromatograms or significant amounts missing were re-sequenced, further described in the previous paragraph.

After quality checking, sequences were assembled and analyzed for ambiguous nucleotides. Ambiguous nucleotides are defined as positions where two or more nucleotides gave fluorescent signal at the same time during sequencing, resulting in secondary peaks in the chromatogram. When a strand had two or more well defined peaks, a letter code (IUPAC nucleotide codes) was inserted to signalize an ambiguous nucleotide. To be defined as a

secondary peak, the peak amplitude had to be higher than half of the primary peak (see figure 9).

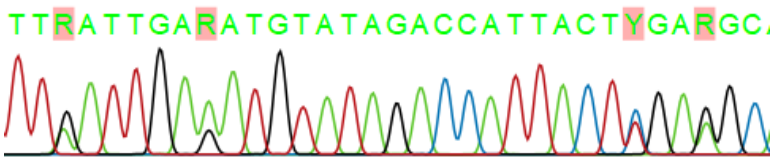


Figure 9. Secondary peaks. The figure illustrates examples of secondary peaks in a chromatogram.

Consensus nucleotide sequences were translated into the corresponding amino acid sequence through identification of the correct reading frame. The correct reading frame was identified by the absence of early stop codons and by comparing them with references of each protein. Both the nucleotide sequences and amino acid sequences were aligned with reference isolates into nucleotide and amino acid alignments.

Based on results from sequence analysis some nucleotide and amino acid sequences were blasted to a registered database in NCBI, with the aim of comparing the queries to existing sequences. Further description of the results and which sequences were blasted and why can be found under “Results”.

Phylogenetic Analysis

Phylogenetic Grouping: S1 (σ3), M2 (μ1), L1 (λ3) and L2 (λ2)

Phylogenetic analysis was performed using CLC Main Workbench and Mega X, using partial nucleotide and corresponding amino acid sequences from the four genomic segments S1, M2, L1 and L2.

Phylogenetic trees were made for all analyzed partial nucleotide sequences and amino acid sequences to investigate the phylogenetic affiliation of the field isolates into the established grouping system described earlier (Group A and B). Phylogenetic grouping is defined as sorting the sequences to groups based on which reference they resemble. The trees were made based on differences in nucleotide and amino acid sequence, as observed in the manual analysis of the alignments. Maximum likelihood was used to construct phylogenetic trees for nucleotide sequences and neighbor joining was used for amino acid sequences, as statistic models. For all trees, the best-fitting substitution model suggested by the software was used. PRV-3 NOR/060214 was selected as outgroup in all trees. Bootstrap values were calculated from 1000 replicates and values above 70 considered significant.

Phylogenetic Analysis: S4 (σ1)

Segment S4 and the encoded σ1 were analyzed by aligning the isolates with the references.

Genogrouping

Finally, the field isolates were categorized into the genogroups described in the introduction based on combinations of the five segments analyzed.

Results

Study Overview

In brief, the study material consisted of 25 separate sample sets collected form Norwegian Atlantic salmon from 1987 to 2008, comprising a total of 204 samples. All samples were screened for PRV-1 by RT-qPCR and PRV-1 was detected in 20 out of the 25 sample sets including 110 positive individual samples (see table 17). A subset of 24 samples were subjected to classic PCR to generate material for sequencing of the five segments linked to PRV-1 virulence (S1, S4, M2, L1 and L2) (see table 18). From six samples, high-quality sequences were obtained from all five segments (see table 19). These results are illustrated in figure 10.

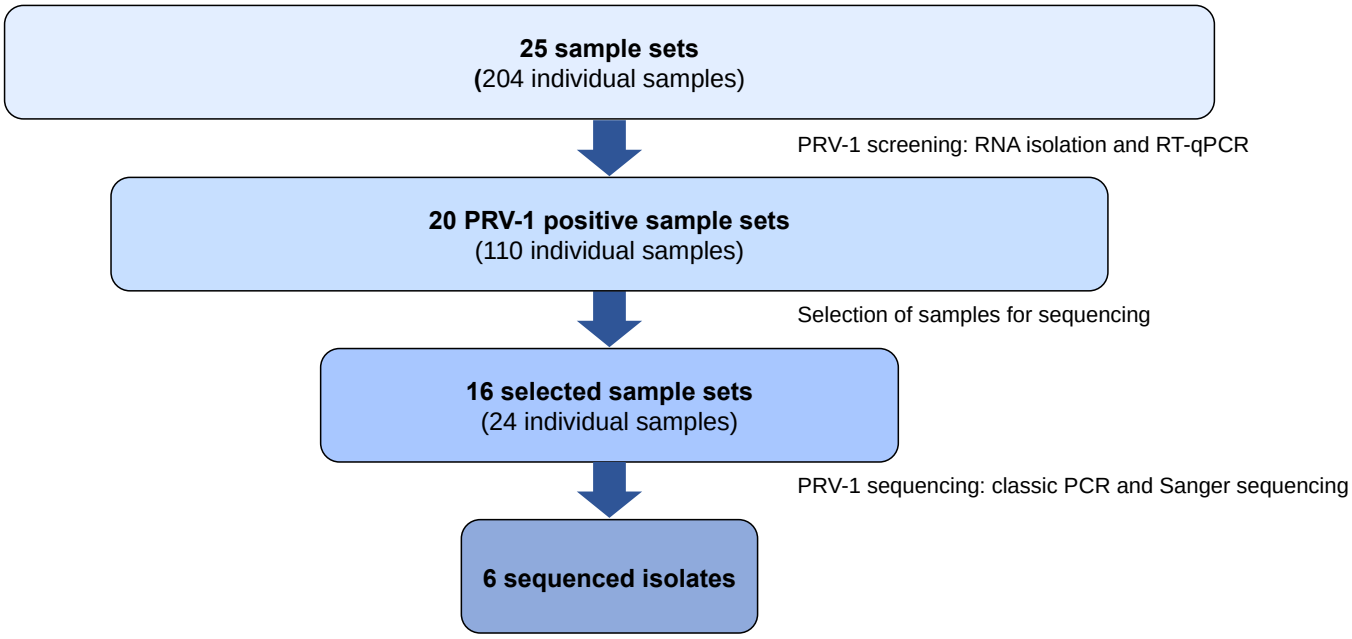


Figure 10. Study overview. Illustration of the total number sample sets screened for PRV-1 by RT-qPCR. Among the 110 PRV-1 positive samples, 24 samples from different sample sets were selected for sequencing, resulting in six sequenced isolates from different sample sets.

Sample set	**Year**	**Samples PRV-1 positive/Total**	**Ct range**
1	1987	0/33	35.7-35.7
2	1988	0/2	No Ct
3	1988	5/5	28.2-32.6
4	1995	5/5	27.8-34.1
5	1996	0/9	36.9-36.9
6	1996	5/7	27.5-36.4
7	1996	4/16	32.8-37.7
8	1997	1/4	31.5-31.5
9	1998	2/12	33.6-38.0
10	1998	16/20	19.9-29.3
11	1998	1/2	22.8-22.8
12	1999	2/10	17.7-21.6
13	2001	3/5	31.5-36.2
14	2004	6/6	19.1-25.4
15	2004	12/12	20.8-25.4
16	2004	0/3	No Ct
17	2004	12/12	19.7-30.5
18	2004	6/6	21.7-25.6
19	2005	6/6	21.7-26.6
20	2005	3/3	27.8-30.8
21	2005	0/5	No Ct
22	2007	9/9	22.1-30.5
23	2007	3/3	26.4-27.3
24	2007	6/6	24.5-28.3
25	2008	3/3	26.5-27.5

Table 17. Overview of the number of PRV-1 positive samples and Ct range in each sample set. The Ct range displays the lowest and highest Ct-value in each sample set. Highlighting in grey indicates positive results, while white indicates negative results.

PRV-1 Screening: Virus Detected in 20 out of 25 Sample Sets

The measured RNA concentration of the isolated tissue samples varied strongly, both within and between sample sets, probably reflecting variation in preservation of RNA. The RNA concentration varied from as little as 0,6 ng/µL to as much as 1520.4 ng/µL. The 260/280 nm ratio varied as well but was ~2 for the samples selected for sequencing.

PRV-1 positive samples were identified by RT-qPCR in 20 out of 25 sample sets, comprising 110 positive individual samples. In general, the Ct values varied both within and between different sample sets. Samples from some sample sets displayed a wide range in Ct values (Sample set 17: Ct: 19.7-30.5), while other sample sets had narrow Ct ranges (Sample set 23: Ct: 26.4-27.3). For an overview of the RNA concentration, 260/280 nm ratio and Ct value of all samples, see appendix 1.

PRV-1 Sequencing

Based on the results from RT-qPCR and RNA quality, a subset of 24 samples from 16 different sample sets were subjected to classic PCR. Sixteen of the amplified samples were successfully amplified for a varying number of the segments, visible as single distinct bands in agarose gel electrophoresis (see table 18). Eight samples from three different sample sets did not obtain bands on any segments. PCR amplification was attempted for all five segments on two of these samples. A total of ten samples from nine different sample sets were not successfully amplified for one or more segments. A total of six samples from six different sample sets were successfully amplified on each analyzed segment. In general, samples with the lowest Ct values gave successful PCR amplifications, while the ones that did not had higher Ct values. All successfully PCR amplifications were sequenced.

Sample set	Sample name	Ct value	RNA ng/ul	S1 1	S1 2	S4	M2	L1 1	L1 2	L2 1	L2 2
4	NOR-1995-001	31.4	na			-					
	NOR-1995-002*	32.9	na			-					
	NOR-1995-003	31.6	na			-					
	NOR-1995-004	27.8	na	-	-	-	-	-	-	-	-
	NOR-1995-004*	27.8	na			-					
6	NOR-1996-013*	29.0	na			-					
	NOR-1996-014	27.5	na	-	-	-	-	-	-	-	-
	NOR-1996-014*	27.5	na			-					
10	NOR-1998-ROG-3	19.9	107.1	x	x	x	x	x	x	x	x
	NOR-1998-ROG-017	24.8	232.0	x	x	-	-	-	-	-	x
11	NOR-1998-5	22.8	na	x	x	x	x	x	x	x	x
12	NOR-1999-003	21.6	na	x	x	x	x	-	-	-	-
	NOR-1999-007	17.7	na	x	x	x	x	x	x	-	-
13	NOR-2001-003	31.5	na			-					
	NOR-2001-004	32.2	na			-					
14	NOR-2004-MRO-9	19.1	250.1	x	x	x	x	x	x	x	x
	NOR-2004-003	24.4	82.3	x	x	-	-	-	-	-	x
15	NOR-2004-007	20.8	333.6	x	x	x	x	-	x	-	x
17	NOR-2004-028	19.7	123.9	x	x	x	x	-	x	-	x
18	NOR-2004-12	21.7	652.7	x	x	x	x	x	x	x	x
19	NOR-2005-13	21.7	131.4	x	x	x	x	x	x	x	x
20	NOR-2005-008	27.8	80.0	x	x	-	-	-	-	-	x
21	NOR-2007-010	26.7	160.8	x	x	-	x	-	x	-	x
22	NOR-2007-15	22.1	266.7	x	x	x	x	x	x	x	x
24	NOR-2007-017	24.5	307.3	x	x	-	x	x	x	-	x
25	NOR-2008-003	26.5	106.2	x	x	-	-	-	x	-	x

Table 18. Overview of the RNA concentration, Ct value and successful PCR amplifications of the 24 samples selected for sequencing. RNA concentration is not applicable (na) for cell-free material. X indicates successful PCR amplification in given segment. – indicates unsuccessful PCR amplification in given segment. Empty fields indicate gel not attempted in given segment. * indicates sample isolated with Trizol LS protocol, including two duplicates. The six sequenced isolates are highlighted in grey.

Overview of the Six Sequenced PRV-1 Isolates

High-quality sequence of each of the five selected genomic segments (S1, S4, M2, L1 and L2) putatively linked to virulence were obtained for a total of six PRV-1 isolates collected from farmed Atlantic salmon in Norway in 1998, 2004, 2005 and 2007 (see table 18 and 19). Two of the samples were collected before 1999 and thus the emergence of HSMI. One of these was collected for unknown reasons, while the other was collected from a suspected ISA outbreak. The remaining four samples were collected from 2004-2007 from suspected HSMI outbreaks. The geographic origin was known for two of the samples. All sequenced isolates had Ct values ≤ 22.8.

Name	Sample set	County	Purpose of sampling	Ct value
NOR-1998-ROG-3	10	Rogaland	Suspected ISA	19.9
NOR-1998-5	11	Unknown	Unknown	22.8
NOR-2004-MRO-9	14	Møre og Romsdal	Suspected HSMI	19.1
NOR-2004-12	18	Unknown	Suspected HSMI	21.7
NOR-2005-13	19	Unknown	Suspected HSMI	21.7
NOR-2007-15	22	Unknown	Suspected HSMI	22.1

Table 19. Overview of the six sequenced isolates.

Phylogenetic and Sequence Analysis

Phylogenetic analysis was performed on each separate segment (S1, M2, L1 and L2) including both the nucleotide sequence and the encoded amino acid sequence (σ3, p13, μ1, λ3 and λ2) (see table 20). In general, similar tree topologies were observed for all nucleotide sequences and corresponding amino acid sequences, but the trees based on nucleotide sequences displayed a greater degree of subgrouping reflecting more variation in the sequences. The present study will mainly focus on the amino acid sequence, and the nucleotide trees can be found in appendix 5. All observed amino acid differences can be found in appendix 4.

Segment	Protein	Analyzed coding sequence bp (aa)
S1	σ3	29-1021 (1-330)
	p13	108-482 (1-124)
S4	σ1	78-967 (14-309)
M2	μ1	77-2020 (18-664)
L1	λ3	1100-3788 (365-1260)
L2	λ2	1652-3772 (546-1251)

Table 20. Overview of the coding sequences and amino acids analyzed.

S1 (σ3 and p13)

Phylogenetic analysis of the amino acid sequence of σ3 (S1) separated the PRV-1 isolates into two distinct phylogenetic groups, σ3 (S1) Group A and B. One of the isolates from the present study, NOR-1998-ROG-3, clustered to σ3 (S1) Group A and had a highly similar sequence to the low virulent reference isolates NOR1998, NOR-1996 and CAN 16-005ND. Three isolates, NOR-2004-MRO-9, NOR-2004-12 and NOR-2004-15, clustered to σ3 (S1) Group B with a bootstrap value of 99. These isolates had identical or highly similar sequences to the high virulent reference isolates NOR-2018/NL and NOR-2018/SF as well as the low virulent isolate NOR-1997. Isolate NOR-1998-5 and NOR-2005-13 were excluded from the phylogenetic tree due to a significant number of unknown amino acids during sequence analysis. These findings will be addressed later in the results section. A similar tree topology was observed by phylogenetic analysis of the amino acid sequence of p13 (see appendix 6).

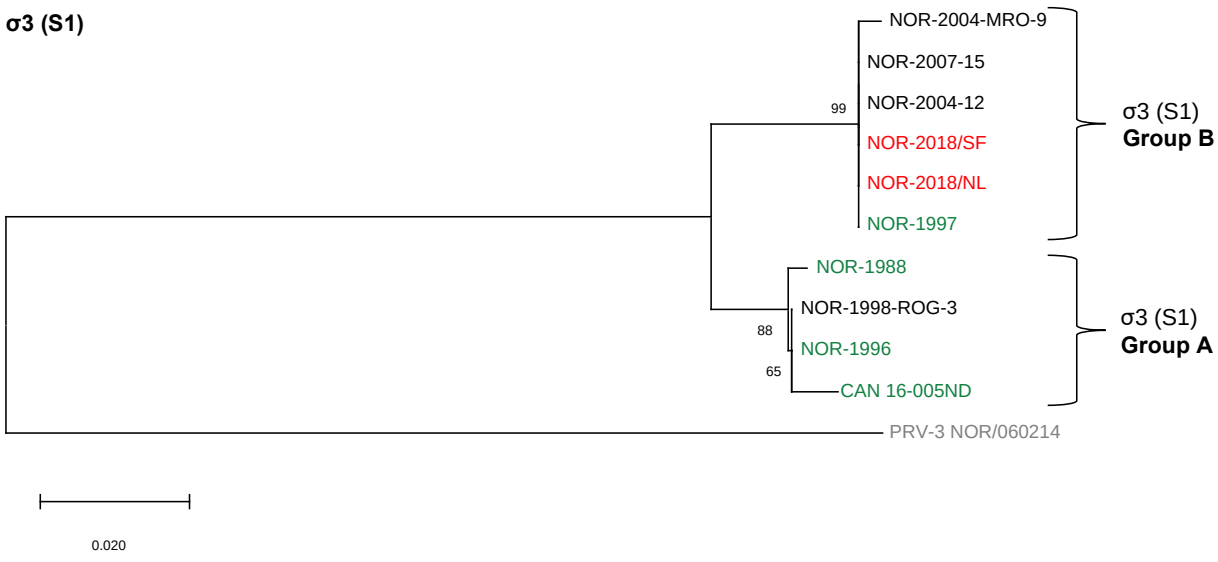


Figure 11. Phylogenetic analysis of σ3 (S1). Phylogenetic tree constructed from the amino acid sequence σ3 (aa 1-330), using neighbor joining. The analysis included four isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253816) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as σ3 (S1) Group A and B. Bootstrap values were calculated from 1000 replicates.

M2 (µ1)

Phylogenetic analysis of the partial amino acid sequence of µ1 (M2) separated the PRV-1 isolates into the same two distinct phylogenetic groups observed for S1 (σ3), which is natural as co-segregation of these two segments have been confirmed earlier. In contrast to in segment S1, there were no secondary peaks for NOR-1998-5 in M2 and the isolate could therefore be placed in Group B. NOR-2005-13 was excluded from the phylogenetic tree due to a significant number of unknown amino acids, described in detail under sequence analysis.

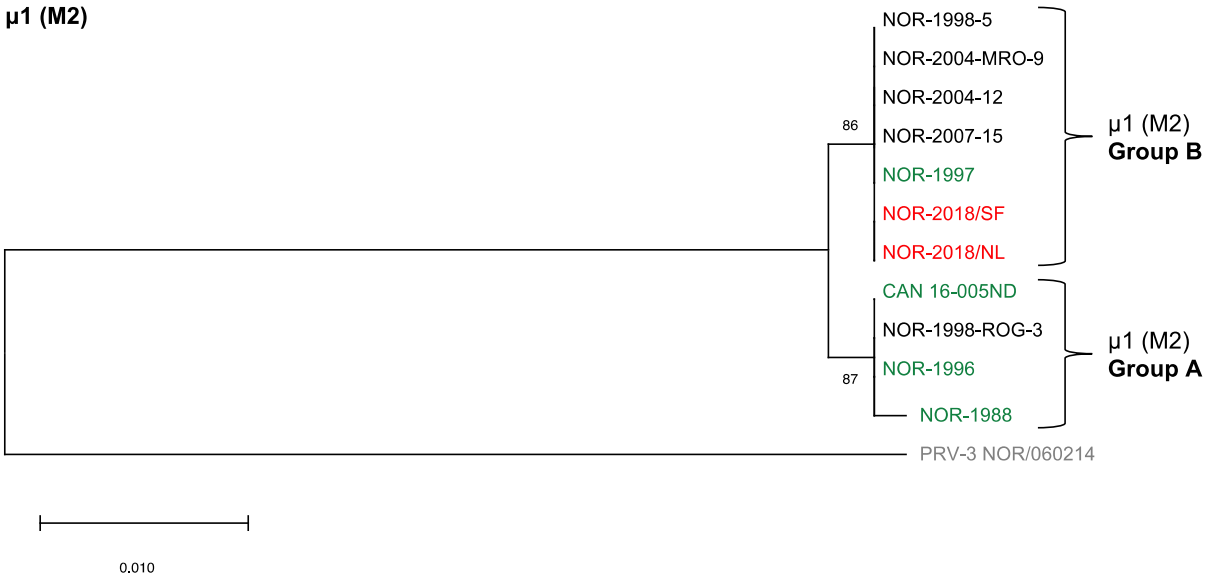


Figure 12. Phylogenetic analysis of µ1 (M2). Phylogenetic tree constructed from the partial amino acid sequence µ1 (aa 18-664), using neighbor joining. The analysis included five isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253811) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as µ1 (M2) Group A and B. Bootstrap values were calculated from 1000 replicates.

L1 (λ3)

Phylogenetic analysis of the partial amino acid sequence of λ3 (L1) clustered the PRV-1 isolates into two different distinct phylogenetic groups, named λ3 (L1) Group A and B. However, these two clusters were composed of different isolates than that observed for S1 (σ3) and M2 (μ1). The PRV-1 field isolates, NOR-1998-ROG-3, NOR-1998-5, NOR-2004-MRO-9 and NOR-2005-13 clustered to λ3 (L1) Group A, together with both high virulent and low virulent reference isolates. No field isolates clustered together with NOR-1997 in λ3 (L1) Group B. NOR-2004-12 and NOR-2007-15 were excluded due to inconclusive results, described in detail under sequence analysis.

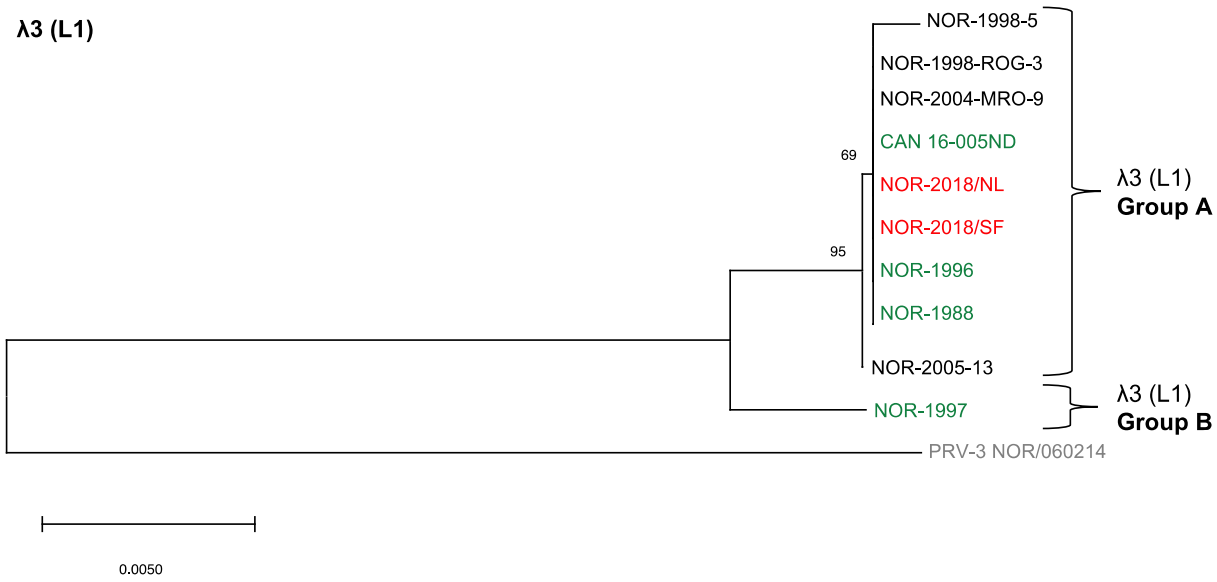


Figure 13. Phylogenetic analysis of λ3 (L1). Phylogenetic tree constructed from the partial amino acid sequence λ3 (aa 365-1260), using neighbor joining. The analysis included four isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253807) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as λ3 (L1) Group A and B. Bootstrap values were calculated from 1000 replicates.

L2 (λ2)

Phylogenetic analysis of the partial amino acid sequence of λ2 (L2) separated the PRV-1 isolates into the same two distinct groups observed for λ3. All the six PRV-1 field isolates, NOR-1998-ROG-3, NOR-1998-5, NOR-2004-MRO-9, NOR-2004-12, NOR-2005-13 and NOR-2005-13 clustered to λ2 (L2) Group A, together with both high virulent and low virulent reference isolates. λ2 (L2) Group A had a bootstrap value of 95 and was subdivided into two sub-groups, but with a low bootstrap value support (>70). No field isolates clustered together with NOR-1997 in λ2 (L2) Group B.

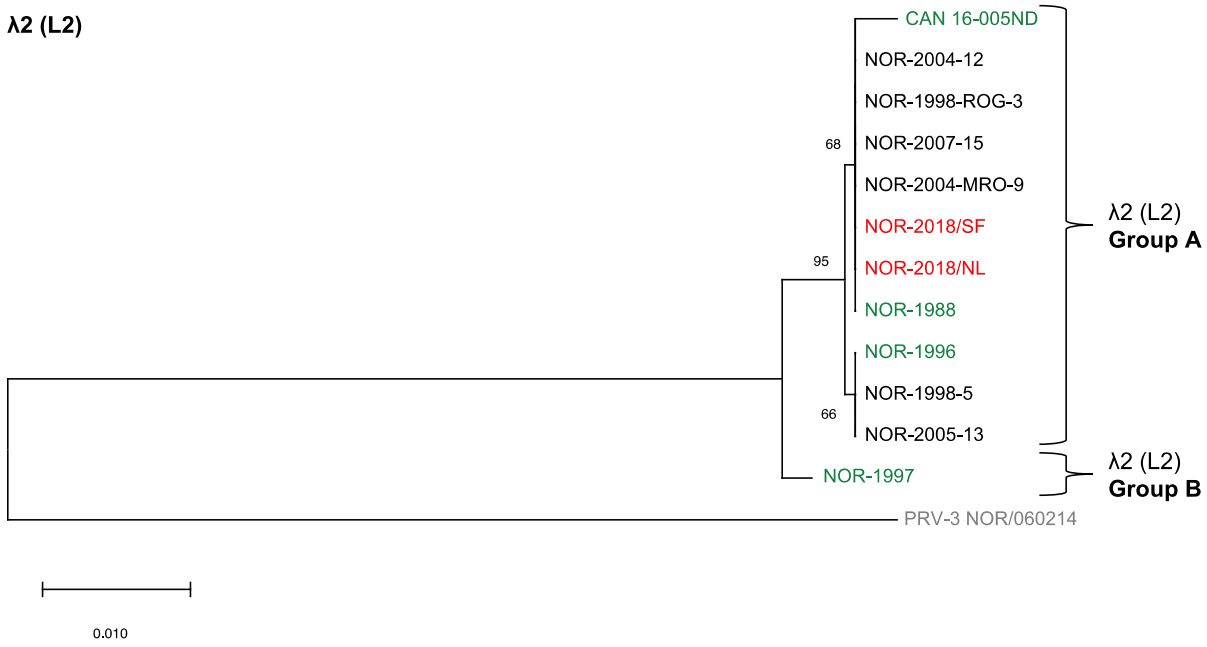


Figure 14. Phylogenetic analysis of λ2 (L2). Phylogenetic tree constructed from the partial amino acid sequence λ2 (aa 546-1251), using neighbor joining. The analysis included six isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253808) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as λ2 (L2) Group A and B. Bootstrap values were calculated from 1000 replicates.

S4 (σ1)

Segment S4 did not cluster into distinct phylogenetic groups, and the partial amino acid sequence σ1, encoded by S4, were therefore analyzed with respect to single amino acid substitutions putatively linked to virulence. NOR-2007-15 shared the amino acid substitution D252N with the high virulent reference isolate NOR-2018/NL. NOR-2004-MRO-9 and NOR-2004-12 had an amino acid substitution in position 254, where glutamic acid was replaced by lysine, due to a single point mutation in bp 798 (G798A). Neither this amino acid substitution nor any variation was observed in this position in the reference isolates. The amino acid sequences of these isolates were blasted to the database of GenBank (NCBI) for comparison with publicly available DNA sequences, which displayed other described PRV-1 isolates with variation in this position, but none with lysine. All observed amino acid substitutions are shown in table 21.

Isolate	107	252	254	Isolate	107	252	254
NOR-1988	V	D	E	NOR-1998-ROG-3	V	D	E
NOR-1996	V	D	E	NOR-1998-5	V	D	E
NOR-1997	V	D	E	NOR-2004-MRO-9	V	D	**E254K**
CAN 16-005ND	V	D	E	NOR-2004-12	V	D	**E254K**
NOR-2018/SF	**V107A**	D	E	NOR-2005-13	V	D	E
NOR-2018/NL	V	**D252N**	E	NOR-2007-15	V	**D252N**	E

Table 21. Observed amino acid variations in the partial analyzed amino acid sequence of σ1 (S4) for all PRV-1 isolates in the present study. The reference isolate NOR-2018/SF, with demonstrated high virulence, had a V107A substitution. The reference isolate NOR-2018/NL, also with demonstrated high virulence, had a D252N substitution. The same substitution was observed in NOR-2007-15. NOR-2004-MRO-9 and NOR-2004-12 had a E254K substitution.

Genetic Grouping of Each Isolate

NOR-1998-ROG-3: Genogroup Low-2

The phylogenetic analysis demonstrated that isolate NOR-1998-ROG-3 clustered to Group A in all segments (S1, M2, L1 and L2) and had no single amino acid substitutions in σ1 (S4). This is consistent with the genetic composition of isolates in genogroup Low-2, which includes the low virulent reference isolates NOR-1988, NOR-1996 and CAN 16-005ND. This strongly suggests that isolate NOR-1998-ROG-3 is of low virulence.

NOR-2004-MRO-9: Genogroup Unknown-5

NOR-2004-MRO-9 clustered to Group B in segment S1 and M2 and Group A in L1 and L2, which is similar to genogroup High-1 and High-2. However, the isolate differed from these groups in S4 (σ1). Interestingly, the isolate contained an amino acid substitution in amino acid position 254 σ1 (E254K). There is no reference isolate with a similar genetic composition, and the findings placed the isolate in a new genogroup with unknown virulence, Unknown-5.

NOR-1998-5: Potential Co-infection

During sequence analysis of isolate NOR-1998-5, a number of secondary peaks in the chromatograms was observed in all segments, except M2, resulting in ambiguous nucleotides and unknown amino acids in the nucleotide and amino acid sequences. The total number secondary peaks were 18 in S1, 5 in S4, 25 in L1 and 16 in L2. The high number of secondary peaks throughout the segments rendered them non-suitable for phylogenetic analysis and indicated a potential co-infection with two different isolates.

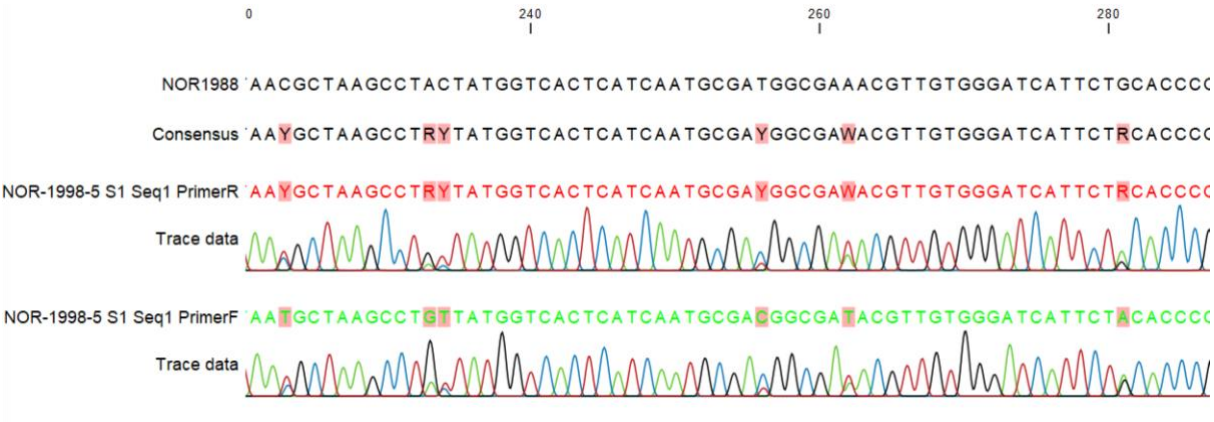


Figure 15. Illustration of six of the secondary peaks found in the nucleotide sequence of S1 of NOR-1998-5.

In segment S1 and the encoded σ3 and p13, a high number of the secondary peaks were located in nucleotide positions defining the grouping into Group A and Group B. Interestingly, the isolate had ambiguous nucleotides and unknown amino acids in the positions defining grouping in the first part of S1, while the remaining half mostly sorted into Group A. In positions with ambiguous nucleotides it was evident that one of the isolates was similar to the reference isolates in Group A, while another was similar to the references in Group B. S1 was sequenced in two parts with two different primer sets, and the shift from unknown nucleotides and amino acids to clear grouping coincided with the overlapping section of the sequences from the two different primer sets, which may indicate that only the first primer set detected both isolates. As the isolate could not be grouped in this segment, it was therefore excluded from the phylogenetic trees of S1, σ3 and p13.

In segment M2, no secondary peaks were observed and NOR-1998-5 clustered to Group B. In segment, L1 and L2, the secondary peaks were to a greater extent present in positions not defining grouping, and NOR-1998-5 could therefore be grouped in these segments. NOR-1998-5 was grouped in Group A in L1 and A in L2 and had no amino acid substitution in S4. Nevertheless, due to the observation of potential co-infection, the isolate could not be assigned to a genogroup.

NOR-2005-13: Potential Co-infection

In comparison with NOR-1998-5, NOR-2005-13 also displayed a significant number of secondary peaks in the chromatograms of all segments, indicating a co-infection. 22 secondary peaks were observed in S1, 3 in S4, 48 in M2, 44 in L1 and 20 in L2, interfering with the grouping in some of the segments.

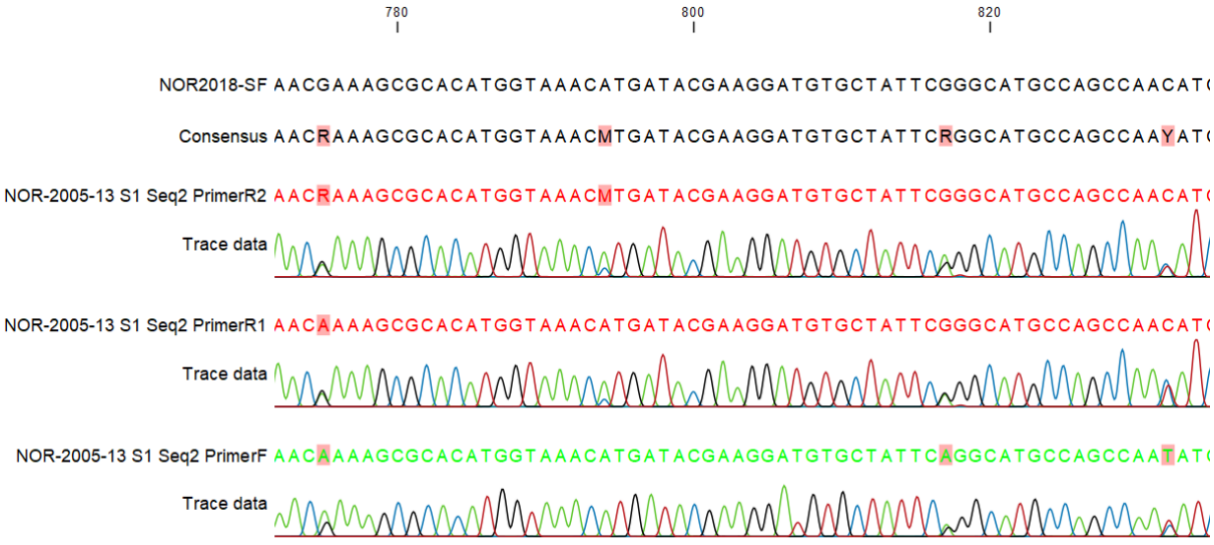


Figure 16. Illustration of six of the secondary peaks found in the nucleotide sequence of S1 of NOR-2005-13.

Secondary peaks and following ambiguous nucleotides and unknown amino acids interfered with the grouping of the isolate in segment S1 and the encoded σ3, but the isolate showed clear affiliation to group B in p13. In comparison to NOR-1998-5, the isolate had ambiguous nucleotides and unknown amino acids in half of S1 and clear grouping in the remaining half. All positions with ambiguous nucleotides and unknown amino acids were found in the second half of S1, while the first half grouped in Group B. Interestingly, this group shift also found place in the overlapping section between the two primer sets, indicating a possible technical error. As the isolate could not be grouped in S1 or σ3, it was therefore excluded from these phylogenetic trees.

Secondary peaks also interfered with the grouping of NOR-2005-13 in M2 and the encoded μ1, and the isolate was therefore also excluded from these phylogenetic trees. In segment, L1 and L2, the secondary peaks were to a greater extent present in positions not defining grouping, and NOR-2005-13 could therefore be grouped in Group A in L1 and A in L2. The isolate had no amino acid substitution in S4. As the isolate could not be grouped in all segments, it was not genogrouped.

NOR-2004-12: Inconclusive in L1

NOR-2004-12 showed a clear affiliation to one group in three of four analyzed segments: Group B in S1 and M2 and Group A in L2. Due to inconclusive results in L1 the isolate could not be grouped on this particular segment and consequently not assigned to a genogroup. The isolate had an amino acid substitution in σ1, E254K, which was also observed in NOR-2004-MRO-9. The substitution was explained by a point mutation in position 798 in the nucleotide sequence of S4.

In the partial analyzed nucleotide sequence of L1 and the encoded amino acid sequence of λ3, signs of recombination was observed. In both sequences, NOR-2004-12 was identical to Group B in the first positions defining grouping, while it was identical to Group A in the remaining positions defining grouping. The observed group shift indicated a possible recombination, and to investigate this further the nucleotide sequence was blasted to study if this recombination was detected in earlier described isolates. Blasting revealed that the recombination was not found in any other existing isolates. Due to the inconclusive results with respect to grouping, NOR-2004-12 was excluded from the phylogenetic tree of both the nucleotide sequence of L1 and the amino acid sequence of λ3. The group shift was located to an overlapping section

between the two sequenced part sequences, indicating a possible sample confusion, or that primer design was not optimal. Regardless of the reason, the results were inconclusive.

NOR-2007-15: Inconclusive in L1

NOR-2007-15 belonged to Group B in segment S1 and M2, and Group A in L2. In L1 the results were inconclusive, and the isolate could therefore not be grouped in this segment nor genogrouped. The isolate had an amino acid substitution in σ1, encoded by S4, D252N, also observed in NOR-2018/NL. The substitution was explained by a point mutation in position 792 in the nucleotide sequence.

In L1, the same signs of recombination were observed as for NOR-2004-12. In the amino acid sequence, the isolate was identical to group B in the positions from amino acid 372-490, then shifted to group A between amino acid 491 and 758 from where it continued to sort in group A the rest of the amino acid sequence. In the nucleotide sequence, NOR-2007-15 was highly similar to Group B from position 1121-2176, then it shifted to group A between base pair 2177-2260 and uninterruptedly continued the sorting into Group A until the end of the nucleotide sequence. The nucleotide sequence was blasted, revealing that this recombination was not known. The group shift was located to an overlapping section between the two sequenced part sequences, indicating a possible sample confusion, or that primer design was not optimal. Regardless of the reason, the results were inconclusive. As the isolate could not be grouped in this segment, it was excluded from the phylogenetic trees of L1 and λ3.

Overview of Genogrouping

Genogroup	Reference isolate(s)	S1 (σ3), M2 (μ1)	L1 (λ3)	L2 (λ2)	S4 (σ1)	Isolates from present study
High-1	NOR-2018/NL	B	A	A	D252N	
High-2	NOR-2018/SF	B	A	A	V107A	
Low-1	NOR-1997	B	B	B		
Low-2	NOR-1988, NOR-1996, CAN16-005 ND	A	A	A		NOR-1998-ROG-3
Unknown-1		B	B	A		
Unknown-2		B	A	A		
Unknown-3		B	B	A	D252N	
Unknown-4		A	A	A	D252N	
Unknown-5		B	A	A	E254K	NOR-2004-MRO-9
Inconclusive		-	A	A		NOR-1998-5
Inconclusive		B	-	A	E254K	NOR-2004-12
Inconclusive		-	A	A		NOR-2005-13
Inconclusive		B	-	A		NOR-2007-15

Table 22. Overview of the genogrouping of the isolates, based on the phylogenetic grouping (Group A or B) of S1, M2, L1 and L2 and the presence of the amino acid substitution V107 or D252N in S4. NOR-1998-ROG-3 sorted into the genogroup Low-2. NOR-2004-MRO-9 sorted into a new genogroup of unknown virulence, Unknown-5. The remaining four isolates could not be genogrouped due to unclear grouping in at least one segment. Two isolates (NOR-1998-5, NOR-2005-13) had a significant number of secondary peaks in at least one segment. The remaining two isolates (NOR-2004-12, NOR-2007-15) had inconclusive results in segment L1 but showed a distinct affiliation to one group in the other segments. Hyphens indicate segments that could not be grouped.

Discussion

The aim of the present study was to study PRV-1 isolates that circulated in farmed Norwegian Atlantic salmon before and after the first reported HSMI outbreak in 1999. The genomic segments S1, S4, M2, L1 and L2 and their encoded proteins were chosen as a basis for the analysis because of their putative link to virulence. Sequences from six PRV-1 isolates from farmed Norwegian Atlantic salmon sampled in 1998 to 2007 were obtained and analyzed, to provide information about a critical time period around the emergence of HSMI (see figure 17). Based on phylogenetic analysis of the five segments, the isolates were categorized into genogroups according to the definition established by Vatne et al (2021).

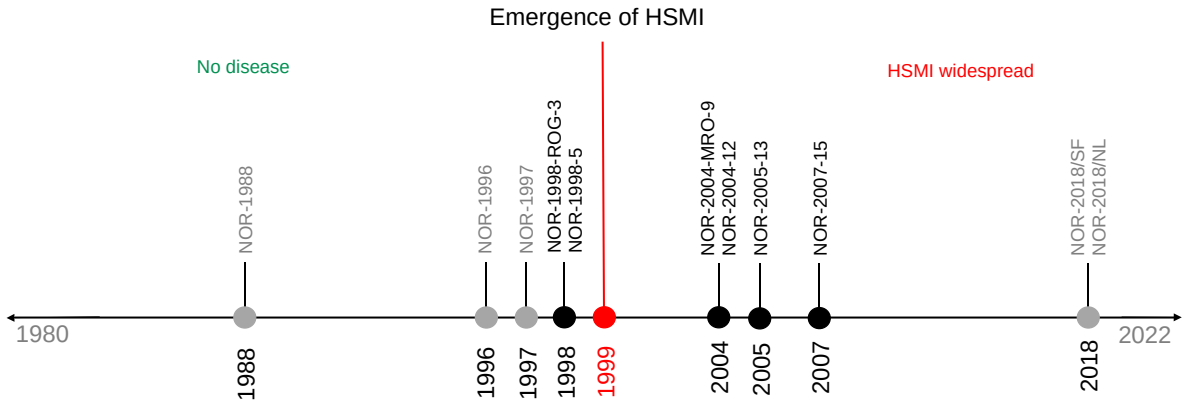


Figure 17. Timeline illustrating the emergence of HSMI, the year of origin of the Norwegian reference isolates and the sequenced isolates from the present study. The sequenced isolates are marked in black. The reference isolates are marked in grey.

PRV-1 Isolate Collected Prior to 1999 is of Low Virulence

All Norwegian PRV-1 isolates sampled prior to the emergence of HSMI (NOR-1988, NOR-1996 and NOR-1997) have previously been confirmed to be of low virulence (Wessel et al., 2020). NOR-1998-ROG-3, from the present study, was genogrouped in the genogroup Low-2, together with the reference isolates NOR-1988, NOR-1996 and CAN 16-005ND, which indicates that NOR-1998-ROG-3 is likely also of low virulence. This substantiates the

hypothesis that PRV-1 isolates present in Norwegian farmed salmon aquaculture in the years before the emergence of HSMI predominantly were of low virulence.

PRV-1 Isolates Collected after 1999 Resemble High Virulent Variants

In the present study, four isolates were sampled after 1999 within the time period 2004-2007. None of these were categorized as low virulent. In general, the isolates grouped with B in S1 and M2 and with A in L1 and L2, a segment combination that has only been found in high virulent isolates. Although they could not be classified as high-virulent, they generally displayed a higher similarity with the reference isolates of demonstrated high virulence, NOR-2018/NL and NOR-2018/SF (Wessel et al., 2020). In addition, co-infections that most likely included low and high virulent isolates, was observed. Altogether this indicates that there was a mix of low and high virulent isolates in the time period after 1999. Earlier studies also illustrate genetic variation in samples originating from 2010, but only one segment (S1) was investigated (Løvoll et al., 2012).

Evolution from Low Virulent to High Virulent: Reassortment

The fact that the current dominating PRV-1 isolates in Norwegian salmon farming aquaculture differ from the isolates pre-dating the first reported clinical HSMI outbreak reflects viral evolution. PRV-1 evolves through mutations, recombination and reassortment, and for viruses with segmented genomes reassortment is of particular importance. Viral evolution, possibly through reassortment and mutation, has generated new genotypes with increased virulence that induced HSMI, and this event occurred about the time of the first reported outbreak of the disease.

A large genetic pool increases viral evolution. The wide-spread wild Atlantic salmon population as well as the farmed salmon in Norway have been a contributor to the size of the PRV-1 genetic pool (Garseth et al., 2013). Genetics and environmental conditions of wild and farmed Atlantic salmon differ. Stocking density, stress and almost unlimited number of susceptible hosts will provide different selection criteria for a virus compared to that in the wild. The development of the PRV-1 isolates is likely related to the rapid expansion of the farming industry, i. e. large farms, high stocking density, transportation of fish and equipment provides optimal conditions for spread of infectious diseases and evolution of pathogens.

The Norwegian reference isolates sampled prior to 1999 either belonged to Group A (NOR-1988, NOR-1996) or Group B (NOR-1997) in all segments. The high virulent reference isolates (NOR-2018/NL and NOR-2018/SF) that are prevalent today are combinations of Group A and B. Based on this, the evolution from low virulent to high virulent can tentatively be explained by reassortment of the segments S1 and M2 of genogroup Low-1 (NOR-1997) and L1 and L2 of genogroup Low-2 (NOR-1996 and NOR-1988) and accumulations of mutations, which have formed High-1 (NOR-2018/NL) and High-2 (NOR-2018/SF) (see figure 18). In contrast, PRV-1 strains from the North American Pacific Coast and Faroe Islands does not appear to have undergone this evolution. These isolates all appear to be closely related to genogroup Low-2 and not associated with clinical HSMI. Interestingly, the NOR1997 variant (genogroup Low 1) have not been observed in North American Pacific salmon, which according to the model proposed above is a prerequisite for the reassortment event that generated the high virulent variants observed in Norway (Dhamotharan et al., 2019).

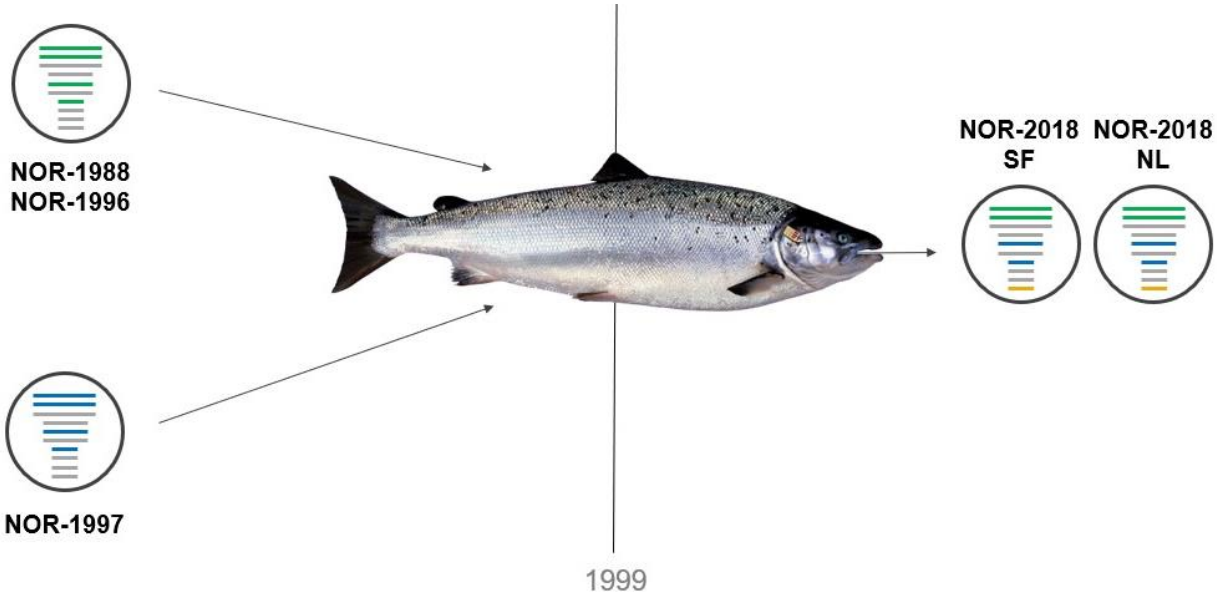


Figure 18. PRV-1 evolution model from low virulent to high virulent through reassortment and mutations. Segment S1, M2, L1 and L2 originating from NOR-1988 and NOR-1996 (Group A) are marked in green. The same segments originating from NOR-1997 (Group B) are marked in blue. S4 with an amino acid substitution linked to high virulence are marked in yellow.

Discovery of Co-infections: A Prerequisite for Reassortment

Co-infection is simultaneous infection in a host by multiple viruses. If two PRV-1 virus particles infect the same host cell, they can exchange segments during replication, and thereby create new hybrid viruses. In the present study we observed signs of co-infection with PRV-1 in two samples, in one sample from 1998 (NOR-1998-5), one year prior to the first description of HSMI, and in one sample from 2005 (NOR-2005-13).

The sequences of NOR-1998-5 and NOR-2005-13 displayed a significant number of secondary peaks on the chromatograms, interpreted as co-infections. A sample containing two virus isolates that differ in nucleotide sequence, will occur if the primers anneal to both isolates. RNA samples of bad quality can also give secondary peaks, but these typically occur intensely in the beginning and end of the sequence as well as randomly throughout the sequence, and often with small amplitudes. Our observations with well-defined secondary peaks with sufficient amplitude indicated that they represented co-infections.

Although the data strongly indicates co-infection in these two samples, secondary peaks were not detected in all segments. In segment M2 in NOR-1998-5 and one partial sequence of S1 for both NOR-1998-5 and NOR-2005-13, the isolates did not show secondary peaks. This could be explained by at least two theories: (1) The primers had higher affinity to one isolates' segment, or (2) this segment originally belongs to the same phylogenetic group for both isolates. In segments L1 and L2 the secondary peaks were in arbitrary positions that did not define the grouping of NOR-1998-5 and NOR-2005-13 and therefore did show clear affiliation to one group in these segments. This can most likely be explained by theory two (2) above.

Interestingly, NOR-1998-5 and NOR-2005-13 had features of both Group A and B. A co-infection with this combination of isolates is especially interesting, due to the theory of Low-1 donating S1 and M2 to Low-2, contributing to the high virulent genogroups High-1 and High-2. Even more interesting, is that one of these isolates was sampled prior to the first reported outbreak of HSMI in 1999, and thus can represent an example of how high virulent isolates occurred.

All currently described high virulent isolates, and the co-infected samples NOR-1998-5 and NOR-2005-13 in this study, had group A in segment L1 and L2. This may indicate that gGoup B in these segments is not advantageous for the virus. This is substantiated by the fact that Group A dominates in segment L1 and L2 in recently described Norwegian PRV-1 field isolates.

New Amino Acid Substitution in σ1

Variation in viral receptor binding proteins is of particular interest when it comes to investigating virulence, as it can impose consequences for the virus' cell tropism and therefore

also virulence. In general, there is little variation in the assumed PRV-1 receptor-binding protein σ1. However, a single amino acid substitution is observed in both reference isolates of confirmed high-virulence, NOR-2018/NL and NOR-2018/SF, D252N and V107A respectively (Vatne et al., 2021). If these amino acids substitutions are implicated in virulence, they impose a common phenotypic effect.

Interestingly, a new variation in σ1 was found with an amino acid substitution in position 254, not being present in any of the reference isolates. The amino acid change in position 254 is only two positions from the amino acid change in position 252 in NOR-2018/NL. The amino acid substitution in 254 was observed in two isolates, NOR-2004-MRO-9 and NOR-2004-12 and gave an exchange from glutamic acid to lysine (E254K). The E254K substitution was not observed in any other sequences in GenBank (NCBI), but interestingly other variations in the same amino acid position were found in the deposited sequences of GenBank (NCBI), indicating a possible hot spot for mutations. S4 has a highly conserved sequence, and the fact that two different isolates displayed the same amino acid substitution suggests a common origin.

The observed variation in other isolates in the same position together with E254K is only two amino acid positions away from the known D252N substitution makes it interesting with respect to the receptor binding function of σ1. The importance of the E254K substitution for virulence should therefore be further investigated in a challenge trial. It would also be interesting to visualize the predicted secondary structure of the protein of the different variants, in order to investigate the spatial consequence of the amino acid changes.

Methodological Considerations: Degradation of RNA

Out of the 24 selected samples, high quality sequences were obtained from six samples. It is important to recognize that the results from amplification and sequencing can be negatively influenced by poor integrity of RNA and it is natural to question the RNA quality of the samples. RNA is sensitive to spontaneous degradation due to its chemical structure and highly susceptible to cleavage catalyzed by ubiquitous RNases. Caution must therefore be taken when handling and preserving RNA to avoid contamination and ensure optimal storage conditions for stability (Fleige et al., 2006).

The archived samples used in the present study were not originally stored for the purpose of sequencing. All the investigated samples had been stored for a long time, from between 14 and 35 years, and had been stored in different ways, in general not favorable with respect to preservation of RNA. RNases are less active in lower temperatures. and non-fixated samples should be stored at –80 °C, however many samples of the samples had been stored at -20°C, which increases the risk of RNA degradation by RNases significant over time. None of the samples were stored in RNAlater or similar fixatives, as these were not yet available at the time when many of the samples were collected. The storage methods and time contributed to the challenges with obtaining acceptable sample quality for classic PCR amplification and subsequent amplicon sequencing.

In addition to ensuring the right preservation method, it is important to ensure stability and avoid contamination of the samples both before, during and after isolation. Both the sampling, handling methods before preservation and the preservation were out of our hands in regards of this. From storage to the extraction of RNA and further on preservation, general principles for ensuring stability of RNA and reducing the risk of contamination were implemented. To avoid

spontaneous degradation, the samples were thawed as few times as possible, always thawed on ice and kept on ice during experiments. To avoid contamination, workspaces were disinfected with ethanol 70 % and gloves were used when handling samples.

As general principles for avoiding degradation were used during the present study, the variable RNA quality and challenges with sequencing were most likely related to the storage of the samples.

Conclusion

Sequence analysis of PRV-1 isolates pre- and post-emergence of HSMI have brought us a step closer in understanding the emergence of the disease through evolution of the virus. Today PRV-1 is considered ubiquitous. The results from the present study revealed that the prevalence of PRV-1 was high throughout the study period (1987-2008), as virus was detected in 20 out of 25 sample sets. This confirms that PRV-1 was common in farmed Atlantic salmon prior to the emergence of HSMI.

The analyzed sample prior to 1999 was likely of low virulence, while the ones sampled after 1999 were more similar the high-virulent isolates described from recently sampled Norwegian farmed Atlantic salmon, indicating viral evolution. Even though HSMI was a notifiable disease until 2014, there were no specific measures taken from the Norwegian Food Safety Authority at this time. Thus, PRV-1 spread freely and adapted to changes and expansion in the farming situation over time. Evolutionary selection for isolates that gave high viral loads and were effective spreaders was inevitable, since this is the nature of evolution. Co-infection and reassortment are a part of this evolution and two co-infections with isolates belonging to different monophyletic clusters were observed in this study, illustrating that reassortment can

occur. For PRV-1 it looks like the virus has evolved towards a genotype resembling the isolates with confirmed high virulence.

Even though this is the conclusion for now, one cannot know what the future brings. PRV-1 will keep adapting and it may evolve to a virus inducing less or more severe pathology, or stay indifferent, since it is the ability to infect and spread that is essential for viruses. For this reason, it is imperative to recognize PRV-1 as a ubiquitous virus that needs to be kept under surveillance. However, the findings are based on a few sequenced isolates. More knowledge regarding PRV-1 variants in Norwegian aquaculture before and after the emergence of HSMI is of central importance in understanding the evolution of PRV-1 and emergence of disease. Further attempts to sequence the remaining PRV-1 positive isolates would be beneficial. The properties of genogroup Unknown 1-5 should be further investigated, both in challenge trials to understand their virulence as well as investigation of the secondary structure and phenotypic property of the E254K substitution discovered in this study.

Acknowledgements

The study was funded by the Norwegian University of Life Sciences (NMBU) and conducted in collaboration with PhD project of Nina Askim Vatne (PRV variants and virulence) and Research Council of Norway project #301477 (Virus entry and receptors in fish).

Foremost, we would like to express our sincere gratitude to our supervisors Professor Espen Rimstad, Dr. Øystein Wessel and Nina Askim Vatne at NMBU. A special thanks to Prof. Espen Rimstad for giving us this opportunity, we are forever grateful. Thanks to supervisor Dr. Øystein Wessel for your constant mentoring, motivation and support from the very beginning. Your knowledge and organized approach in handling any research problem is priceless, and your passion has truly inspired us in the field of virology and research. Thanks to co-supervisor Nina Askim Vatne for all your support, discussions and ideas.

We would also like to praise the virology lab engineers Elisabeth, Stine, Ingvild and Ingrid for helping us with methodical guidance. A special thanks to Elisabeth for teaching us all the methods, listening to all the challenges and encouraging us throughout the whole study. All our newly found skills in lab work, we owe to you.

We would like to thank all the members of the virus unit for welcoming us and creating a warm and cheerful environment.

Last but not least, we would like to thank Anne Berit Olsen for providing us with archived samples as well as help finding data from HSMI outbreaks in Norway.

References

Aldrin, M., Storvik, B., Frigessi, A., Viljugrein, H. & Jansen, P. A. (2009). A stochastic model for the assessment of the transmission pathways of heart and skeleton muscle inflammation, pancreas disease and infectious salmon anaemia in marine fish farms in Norway. *Preventive Veterinary Medicine*, 93 (1): 51–61. doi: 10.1016/j.prevetmed.2009.09.010.

Becker, M. M., Peters, T. R., & Dermody, T. S. (2003). Reovirus sigma NS and mu NS proteins form cytoplasmic inclusion structures in the absence of viral infection. *Journal of virology*, 77 (10): 5948–5963. doi:10.1128/jvi.77.10.5948-5963.2003.

Bio-Rad (s.a.). *Horizontal electrophoresis*. Available at: https://www.bio-rad.com/en-no/applications-technologies/horizontal-electrophoresis?ID=LUSOQX9JY (accessed: 23.09.2022).

Bjørgen, H., Wessel, Ø., Fjelldal, P. G, Hansen, T. J, Sveier, H., Sæbø, H. R., Enger, K. B., Monsen, E., Kvellestad, A. S., Rimstad, E. & Koppang, E.O. (2015). Piscine orthoreovirus (PRV) in red and melanized foci in white muscle of Atlantic salmon (*Salmo salar*). *Veterinary Research*, 46 (1). doi: 10.1186/s13567-015-0244-6.

Bouziat, R., Hinterleitner, R., J. Brown, J. J., Stencel-Baerenwald, J. E., Ikizler, M., Mayassi, T., Meisel, M., M. Kim, S. M., Discepolo, V., Pruijssers, A. J., Ernest, J. D., Iskarpatyoti, J. A., Costes, L. M. M., Lawrence, I., Palanski, B. A., Varma, M., Zurenski, M. A., Khomandiak, S., McAllister, N., Aravamudhan, P., Boehme, K. W., Hu, F., Samsom, J. N., Reinecker, H., Kupfer, S. S., Guandalini, S., Semrad, C. E., Abadie, V., Khosla, C., Barreiro, L. B, Xavier, R. J., Ng, A., Dermody, T. S. & Jabri1, B. (2017). Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science*, 356 (6333): 44-50. doi: 10.1126/science.aah5298.

Brentano, L., Noah, D. L., Brown, E. G., & Sherry, B. (1998). The reovirus protein mu2, encoded by the M1 gene, is an RNA-binding protein. *Journal of virology*, 72 (10), 8354–8357. doi:10.1128/JVI.72.10.8354-8357.1998.

Dhamotharan, K., Tengs, T., Wessel, Ø., Braaen, S., Nyman, I. B., Hansen, E. F., Christiansen, D. H., Dahle, M. K., Rimstad, E. & Markussen, T. (2019). Evolution of the Piscine orthoreovirus Genome Linked to Emergence of Heart and Skeletal Muscle Inflammation in Farmed Atlantic Salmon (Salmo salar). *Viruses*, 11 (5). doi: 10.3390/v11050465.

Di Cicco, E., Ferguson, H. W., Schulze, A. D., Kaukinen, K. H., Li, S., Vanderstichel, R., Wessel, Ø., Rimstad, E., Gardner, I. A., Hammell, K. L. & Miller, K. M. (2017). Heart and skeletal muscle inflammation (HSMI) disease diagnosed on a British Columbia salmon farm through a longitudinal farm study. *PLoS ONE*, 12 (2). doi: 10.1371/journal.pone.0171471.

The Norwegian Directorate of Fisheries. (2022). *Matfiskproduksjon: Salg av laks og regnbueørret, og salg per sysselsatt 1994-2021.* Available at: https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Laks-regnbueoerret-og-oerret/Salg%20av%20laks%20og%20regnbue%C3%B8rret (accessed 26.08.2022).

Dryden, K. A., Wang, G., Yeager, M., Nibert, M. L., Coombs, K. M., Furlong, D. B., Fields, B. N., & Baker, T. S. (1993). Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. *The Journal of cell biology*, 122 (5): 1023–1041. doi:10.1083/jcb.122.5.1023.

FAO. (2022). *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation*. The State of World Fisheries and Aquaculture (SOFIA). Available at: https://www.fao.org/documents/card/en/c/cc0461en (accessed 05.09.2022).

Ferguson, H. W., Kongtorp, R. T., Taksdal, T., Graham, D. & Falk, K. (2005). An outbreak of disease resembling heart and skeletal muscle inflammation in Scottish farmed salmon, Salmo salar L., with observations on myocardial regeneration. *Journal of fish diseases*, 28 (2): 119–123. doi: 10.1111/j.1365-2761.2004.00602.x.

Finstad, Ø. W., Falk, K., Løvoll, M., Evensen, Ø., & Rimstad, E. (2012). Immunohistochemical detection of piscine reovirus (PRV) in hearts of Atlantic salmon coincide with the course of heart and skeletal muscle inflammation (HSMI). *Veterinary Research*, 43 (1). doi: 10.1186/1297-9716-43-27.

Finstad, Ø. W., Dahle, M. K., Lindholm, T. H., Nyman, I. B., Løvoll, M., Wallace, C., Olsen, C. M., Storset, A. K. & Rimstad, E. (2014). Piscine orthoreovirus (PRV) infects Atlantic salmon erythrocytes. *Veterinary Research*, 45 (1). doi: 10.1186/1297-9716-45-35.

Fleige, S. & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine*, 27 (2-3): 126–139. doi:10.1016/j.mam.2005.12.003.

Garver, K. A., Johnson, S. C., Polinski, M. P., Bradshaw, J. C., Marty, G. D., Snyman, H. N., Morrison, D. B. & Richard, J. (2016). Piscine Orthoreovirus from Western North America Is Transmissible to Atlantic Salmon and Sockeye Salmon but Fails to Cause Heart and Skeleton Muscle Inflammation. *PLoS ONE*, 11 (1). doi: 10.1371/journal.pone.0146229.

Garseth, Å. H., Fritsvold, C., Opheim, M., Skjerve, E. & Biering, E. (2012). Piscine reovirus (PRV) in wild Atlantic salmon, Salmo salar L., and sea-trout, Salmo trutta L., in Norway. *Journal of fish diseases*, 36 (5): 483–493. doi: 10.1111/j.1365-2761.2012.01450.x.

Garseth, Å. H., Ekrem, T. & Biering, E. (2013). Phylogenetic evidence of long distance dispersal and transmission of piscine reovirus (PRV) between farmed and wild Atlantic salmon. *PLoS ONE*, 8 (12). doi: 10.1371/journal.pone.0082202.

Godoy, M. G., Kibenge, M. J. T., Wang, Y., Suarez, R., Leiva, C., Vallejos, F. & Kibenge, F. S. B. (2016). First description of clinical presentation of piscine orthoreovirus (PRV) infections in salmonid aquaculture in Chile and identification of a second genotype (Genotype II) of PRV. *Virology Journal,* 13. doi: 10.1186/s12985-016-0554-y.

Gudding, R. & Van Muiswinkel, W. B. (2013). A history of fish vaccination: Science-based disease prevention in aquaculture. *Fish & Shellfish Immunology,* 35 (6): 1683-1688. doi: 10.1016/j.fsi.2013.09.031.

Guglielmi, K. M., Johnson, E.M., Stehle, T. & Dermody, T.S. (2006). Attachment and Cell Entry of Mammalian Orthoreovirus. *Current Topics in Microbiology and Immunology,* 309: 1-38. doi: 10.1007/3-540-30773-7_1.

Gravningen, K. & Berntsen, J. O. (2008, September 1). Fish vaccination: Success with salmon expanding. *Global Seafood Alliance.* Available at: https://www.globalseafood.org/advocate/fish-vaccination-success-with-salmon-expanding/ (accessed: 04.09.2022).

Haatveit, H. M., Nyman, I. B., Markussen, T., Wessel, Ø., Dahle, M. K., & Rimstad, E. (2016). The non-structural protein μNS of piscine orthoreovirus (PRV) forms viral factory-like structures. *Veterinary research*, 47 (1). doi: 10.1186/s13567-015-0302-0.

Hawkins, G. A. (2017). Analysis of human Genetic Varaitions Using DNA sequencing. In Jalali, M., Saldanha, F. Y. L., Jalali M. (ed.) *Basic Science Methods for Clinical Researchers*, pp. 77-98. London, United Kingdom: Elsevier.

International Committee of Taxonomy of Viruses, (s.a.). Orthoreovirus. In *Current ICTV Taxonomy Release.* Available at: https://ictv.global/taxonomy (accessed 13.09.2022).

Islam, M. S., Aryasomayajula, A. & Selvaganapathy, P. R. (2017). A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*, 8 (3). doi: 10.3390/mi8030083.

Joklik, W. K. (1981). Structure and function of the reovirus genome. *Microbiological reviews,* 45 (4): 483-501. doi: 10.1128/mr.45.4.483-501.1981.

Key, T., Read, J., Nibert, M. L. & Duncan, R. (2013). Piscine reovirus encodes a cytotoxic, non-fusogenic, integral membrane protein and previously unrecognized virion outer-capsid proteins. *Journal of General Virology,* 94 (5): 1039-1050. doi: 10.1099/vir.0.048637-0.

Kongtorp, R. T., Kjerstad, A., Taksdal, T., Guttvik, A. & Falk, K. (2004a). Heart and skeletal muscle inflammation in Atlantic salmon, Salmo salar L: a new infectious disease. *Journal of Fish Diseases*, 27 (6): 351-358. doi: 10.1111/j.1365-2761.2004.00549.x.

Kongtorp, R. T., Taksdal, T. & Lyngøy, A. (2004b). Pathology of heart and skeletal muscle inflammation (HSMI) in farmed Atlantic salmon Salmo salar. *Diseases of aquatic organisms*, 59 (3): 217-224. doi: 10.3354/dao059217.

Kongtorp, R. T. & Taksdal, T. (2009). Studies with experimental transmission of heart and skeletal muscle inflammation in Atlantic salmon, Salmo salar L. *Journal of fish diseases*, 32(3): 253–262. doi: 10.1111/j.1365-2761.2008.00983.x.

Lorenz T. C. (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of visualized experiments*, (63): 1-14. doi:10.3791/3998.

Løvoll, M., Alarcón, M., Bang Jensen, B., Taksdal, T., Kristoffersen, A. B., & Tengs, T. (2012). Quantification of piscine reovirus (PRV) at different stages of Atlantic salmon Salmo salar production. *Diseases of aquatic organisms*, 99 (1): 7–12. doi:10.3354/dao02451.

Malik, M. S., Bjorgen, H., Dhamotharan, K., Wessel, O., Koppang, E. O., Di Cicco, E., Hansen, E., F., Dahle, M., K. & Rimstad, E. (2019). Erythroid progenitor cells in Atlantic Salmon (*Salmo salar)* may be persistently and productively infected with piscine orthoreovirus (PRV). *Viruses*, 11 (9). doi: 10.3390/v11090824.

Markussen, T., Dahle, M.K., Tengs, T., Løvoll, M., Finstad, Ø.W., Wiik-Nielsen, C.R., Grove, S., Lauksund, S., Robertsen, B. & Rimstad, E. (2013). Sequence Analysis of the Genome of Piscine Orthoreovirus (PRV) Associated with Heart and Skeletal Muscle Inflammation (HSMI) in Atlantic Salmon (Salmo salar). *PLoS ONE*, 8 (7). doi: 10.1371/journal.pone.0070075.

McDonald, S. M., Nelson, M. I., Turner, P. E., & Patton, J. T. (2016). Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature reviews Microbiology*, 14: 448–460. doi: 10.1038/nrmicro.2016.46.

Nibert, M. L., Odegard, A. L., Agosto, M. A., Chandran, K., & Schiff, L. A. (2005). Putative autocleavage of reovirus mu1 protein in concert with outer-capsid disassembly and activation for membrane permeabilization. *Journal of molecular biology*, 345 (3): 461–474. doi: 10.1016/j.jmb.2004.10.026.

Norwegian Seafood Council. (2020). *Celebrating 50 years of modern aquaculture*. Available at: https://en.seafood.no/news-and-media/news-archive/celebrating-50-years-of-modern-aquaculture/ (accessed 25.08.2022)

Norwegian Veterinary Institute. (2012). *The Health Situation in Norwegian Aquaculture 2011*. Norwegian Veterinary Institute Report. Available at: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2012/fish-health-report-2011 (accessed 05.10.2022).

Norwegian Veterinary Institute. (2015). *The Health Situation in Norwegian Aquaculture 2014.* Norwegian Veterinary Institute Report. Available at: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2015/fish-health-report-2014 (accessed 25.08.2022)

Norwegian Veterinary Institute. (2018). *The Health Situation in Norwegian Aquaculture 2017.* Norwegian Veterinary Institute Report, series 1b/2018. Available at:

https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2018/fish-health-report-2017 (accessed 25.08.2022)

Norwegian Veterinary Institute. (2022). *Norwegian Fish Health report 2021*. Norwegian Veterinary Institute Report, series 2a/2022. Available at: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2022/fish-health-report-2021 (accessed 25.08.2022).

Olsen, A. B., Hjortaas, M., Tengs, T., Hellberg, H., & Johansen, R. (2015). First Description of a New Disease in Rainbow Trout (*Oncorhynchus mykiss (Walbaum)*) Similar to Heart and Skeletal Muscle Inflammation (HSMI) and Detection of a Gene Sequence Related to Piscine Orthoreovirus (PRV). *PloS one*, 10 (7). doi: 10.1371/journal.pone.0131638.

Palacios, G., Løvoll, M., Tengs, T., Hornig, M., Hutchison, S., Hui, J., Kongtorp, R. T., Savji, N., Bussetti, A. V., Solovyov, A., Kristoffersen, A. B., Celone, C., Street, C., Trifonov, V., Hirschberg, D. L., Rabadan, R., Egholm, M., Rimstad, E. & Lipkin, W. I. (2010). Heart and Skeletal Muscle Inflammation of Farmed Salmon is Associated with Infection with a Novel Reovirus. *PloS one,* 5 (7): 483-493. doi: 10.1371/journal.pone.0011487.

Polinski, M. P., Marty, G. D., Snyman, H. N. & Garver, K. A. (2019). Piscine orthoreovirus demonstrates high infectivity but low virulence in Atlantic salmon of Pacific Canada. *Scientific reports*, 9 (1). doi: 10.1038/s41598-019-40025-7.

Polinski, M. P., Vendramin, N., Cuenca, A. & Garver, K. A. (2020). Piscine orthoreovirus: Biology and Distribution in farmed and wild fish. *Journal of fish diseases*, 43 (11): 1331–1352. doi: 10.1111/jfd.13228.

Qiagen, 2020. *QIAamp® Viral RNA Mini Handbook.* Available at: https://www.qiagen.com/us/resources/resourcedetail?id=c80685c0-4103-49ea-aa72-8989420e3018&lang=en (accessed 2.9.2022)

Reinisch, K. M., Nibert, M. L. & Harrison, S. C. (2000). Structure of the reovirus core at 3.6 Å resolution. *Nature,* 404 (6781): 960-967. doi: 10.1038/35010041.

Rimstad, E. (2022) *Interview with Professor Espen Rimstad from NMBU*. Ås (18.10.2022)

Sanjuán, R. & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and molecular life sciences,* 73 (23): 4433–4448. doi: 10.1007/s00018-016-2299-6.

Schiff L. A., Nibert M. L. & Tyler K. L. (2007). Orthoreoviruses and their replication. In Knipe D. M., Howley P. M., Fields B. N. (ed.) *Fields virology Volume 2*, 5th ed, pp. 1853-1915. Philadelphia, USA: Lippincott Williams & Wilkins.

SigmaAldrich. (s.a.) *How qPCR works*. Available at: https://www.sigmaaldrich.com/NO/en/technical-documents/technical-article/genomics/qpcr/how-qpcr-works?gclid=Cj0KCQjwguGYBhDRARIsAHgRm4-Dduuu_bvfG_OvKyslbPEzZOSsBWWo-eS31TYdGJZyF8G5ADinwR4aApEEEALw_wcB (accessed 7.9.2022)

Simon-Loriere, E. & Holmes, E. C. (2011). Why do RNA viruses recombine? *Nature Reviews Microbiology*, 9: 617-626. doi: 10.1038/nrmicro2614.

Statista. (2022). *Production of Atlantic salmon worldwide from 2000 to 2021*. Available at: https://www.statista.com/statistics/1179062/global-atlantic-salmon-production/ (accessed 24.08.2022).

Takano T., Nawata A., Sakai T., Matsuyama T., Ito T., Kurita J., Terashima, S., Yasuike, M., Nakamura, Y., Fujiwara, A., Kumagai, A. & Nakayasu, C. (2016). Full-Genome Sequencing and Confirmation of the Causative Agent of Erythrocytic Inclusion Body Syndrome in Coho Salmon Identifies a New Type of Piscine Orthoreovirus. *PLoS ONE* 11 (10). doi: 10.1371/journal.pone.0165424.

Thermo fisher. (s.a.a). *The Basics: RNA Isolation.* Availiable at: https://www.thermofisher.com/no/en/home/references/ambion-tech-support/rna-isolation/general-articles/the-basics-rna-isolation.html (accessed 27.09.2022)

Thermo Fisher (s.a.b). *Basic principles of RT-qPCR.* Available at: https://www.thermofisher.com/no/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/spotlight-articles/basic-principles-rt-qpcr.html (accessed: 24.09.2022)

Thermo Fisher. (s.a.c). *5 steps to optimal cDNA Synthesis*. Available at: https://www.thermofisher.com/no/en/home/life-science/pcr/reverse-transcription/5steps-cDNA.html (accessed: 23.09.2022).

Vatne, N. A., Stormoen, M., Lund, M., Devold, M., Rimstad, E. & Wessel Ø. (2021). Genetic grouping and geographic distribution of Piscine orthoreovirus-1 (PRV-1) in farmed Atlantic salmon in Norway. *Veterinary Research*, 52. doi: 10.1186/s13567-021-01000-1.

Wessel, Ø., Nyman, I. B., Markussen, T., Dahle, M. K. & Rimstad, E. (2015). Piscine orthoreovirus (PRV) σ3 protein binds dsRNA. *Virus research*, 198: 22–29. doi: 10.1016/j.virusres.2015.01.001.

Wessel, Ø., Braaen, S., Alarcon, M., Haatveit, H., Roos, N., Markussen, T., Tengs, T., Dahle, M. K. & Rimstad, E. (2017). Infection with purified Piscine orthoreovirus demonstrates a causal relationship with heart and skeletal muscle inflammation in Atlantic salmon. *PloS ONE*, 12 (8). doi: 10.1371/journal.pone.0183781.

Wessel, Ø., Hansen, E. F., Dahle, M. K., Alarcon, M., Vatne, N. A., Nyman, I. B., Soleim, K. B., Dhamotharan, K., Timmerhaus, G., Markussen, T., Lund, M., Aanes, H., Devold, M., Inami, M., Løvoll, M., & Rimstad, E. (2020). Piscine Orthoreovirus-1 Isolates Differ in Their Ability to Induce Heart and Skeletal Muscle Inflammation in Atlantic Salmon (Salmo salar). *Pathogens*, 9 (12). doi: 10.3390/pathogens9121050.

Appendices

Appendix 1. Total sample overview

Appendix 2. Storage overview

Appendix 3. Whole fish overview

Appendix 4. Differences in amino acid sequence

Appendix 5. Phylogenetic trees from segment S1, M2, L1 and L2

Appendix 6. Phylogenetic tree from protein p13

Appendix 1. Total sample overview. Overview of all 204 screened samples with respect to sample set, sample date, sampling county, sample type, [RNA], 260/280 nm ratio and Ct value. Selected samples for sequencing are colored grey; light grey to mark amplified samples, dark grey to mark sequenced isolates.

Sample set	Sample name	Sample date	County	Type	RNA ng/µL	260/280 nm ratio	Ct value
1	NOR-1987-001	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-002	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-003	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-004	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-005	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-006	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-007	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-008	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-009	16.11.1987	Unknown	Serum	-	-	-
	NOR-1987-010	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-011	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-012	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-013	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-014	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-015	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-016	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-017	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-018	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-019	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-020	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-021	17.11.1987	Unknown	Serum	-	-	-
	NOR-1987-022	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-023	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-024	18.11.1987	Unknown	Serum	-	-	35.7
	NOR-1987-025	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-026	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-027	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-028	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-029	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-030	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-031	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-032	18.11.1987	Unknown	Serum	-	-	-
	NOR-1987-033	18.11.1987	Unknown	Serum	-	-	-
2	NOR-1988-001	27.04.1988	Unknown	Plasma	-	-	-
	NOR-1988-002	28.04.1988	Unknown	Plasma	-	-	-
3	NOR-1988-003	1988	Unknown	Unknown	0.8	1.38	-
	NOR-1988-004	1988	Unknown	Unknown	2.6	1.39	32.1
	NOR-1988-005	1988	Unknown	Unknown	8.8	1.43	32.6
	NOR-1988-006	1988	Unknown	Unknown	10.0	1.54	30.4
	NOR-1988-007	1988	Unknown	Unknown	2.8	1.33	28.2

4	NOR-1995-001	17.10.1995	Unknown	Plasma	-	-	31.4
	NOR-1995-002	17.10.1995	Unknown	Plasma	-	-	32.9
	NOR-1995-003	17.10.1995	Unknown	Plasma	-	-	31.6
	NOR-1995-004	17.10.1995	Unknown	Plasma	-	-	27.8
	NOR-1995-005	17.10.1995	Unknown	Plasma	-	-	34.1
5	NOR-1996-001	1996	Unknown	Plasma	-	-	-
	NOR-1996-002	1996	Unknown	Plasma	-	-	-
	NOR-1996-003	1996	Unknown	Plasma	-	-	-
	NOR-1996-004	1996	Unknown	Plasma	-	-	36.9
	NOR-1996-005	1996	Unknown	Plasma	-	-	-
	NOR-1996-006	1996	Unknown	Plasma	-	-	-
	NOR-1996-007	1996	Unknown	Plasma	-	-	-
	NOR-1996-008	1996	Unknown	Plasma	-	-	-
	NOR-1996-009	1996	Unknown	Plasma	-	-	-
6	NOR-1996-010	1996	Unknown	Plasma	-	-	29.4
	NOR-1996-011	1996	Unknown	Plasma	-	-	35.3
	NOR-1996-012	1996	Unknown	Plasma	-	-	36.4
	NOR-1996-013	1996	Unknown	Plasma	-	-	29.0
	NOR-1996-014	1996	Unknown	Plasma	-	-	27.5
	NOR-1996-015	1996	Unknown	Plasma	-	-	30.9
	NOR-1996-016	1996	Unknown	Plasma	-	-	31.2
7	NOR-1996-017	1996	Unknown	Heart	63.6	1.96	36.5
	NOR-1996-018	1996	Unknown	Heart	107.2	1.80	33.4
	NOR-1996-019	1996	Unknown	Heart	27.6	1.98	32.8
	NOR-1996-020	1996	Unknown	Heart	37.8	1.99	36.0
	NOR-1996-021	1996	Unknown	Heart	26.2	1.91	36.6
	NOR-1996-022	1996	Unknown	Heart	160.3	2.00	37.7
	NOR-1996-023	1996	Unknown	Heart	129.3	2.03	35.3
	NOR-1996-024	1996	Unknown	Heart	81.5	2.00	37.6
	NOR-1996-025	1996	Unknown	Heart	26.7	1.90	35.6
	NOR-1996-026	1996	Unknown	Spleen	6.7	1.85	37.4
	NOR-1996-027	1996	Unknown	Spleen	8.2	1.67	35.2
	NOR-1996-028	1996	Unknown	Spleen	5.4	1.34	37.1
	NOR-1996-029	1996	Unknown	Spleen	34.9	1.55	34.7
	NOR-1996-030	1996	Unknown	Spleen	23.7	1.95	35.6
	NOR-1996-031	1996	Unknown	Spleen	2.2	2.04	34.1
	NOR-1996-032	1996	Unknown	Spleen	3.5	2.57	35.8
8	NOR-1997-001	1997	Unknown	Plasma	-	-	31.5
	NOR-1997-002	1997	Unknown	Plasma	-	-	-
	NOR-1997-003	1997	Unknown	Plasma	-	-	-
	NOR-1997-004	1997	Unknown	Plasma	-	-	-
9	NOR-1998-001	12.02.1998	Unknown	Kidney	2.2	2.03	37.2
	NOR-1998-002	12.02.1998	Unknown	Kidney	17.2	1.90	-
	NOR-1998-003	12.02.1998	Unknown	Kidney	2.1	1.69	37.0

	NOR-1998-004	12.02.1998	Unknown	Kidney	1.4	1.07	35.6
	NOR-1998-005	12.02.1998	Unknown	Kidney	5.2	1.83	36.7
	NOR-1998-006	12.02.1998	Unknown	Kidney	3.8	2.60	37.0
	NOR-1998-007	12.02.1998	Unknown	Spleen	369.7	1.81	37.6
	NOR-1998-008	12.02.1998	Unknown	Spleen	40.1	2.01	33.8
	NOR-1998-009	12.02.1998	Unknown	Spleen	280.7	1.94	38.0
	NOR-1998-010	12.02.1998	Unknown	Spleen	144.4	2.03	37.8
	NOR-1998-011	12.02.1998	Unknown	Spleen	57.4	2.03	33.6
	NOR-1998-012	12.02.1998	Unknown	Spleen	23.0	2.02	35.5
10	NOR-1998-ROG-013	11.03.1998	Rogaland	Kidney	116.0	2.02	-
	NOR-1998-ROG-014	11.03.1998	Rogaland	Kidney	76.5	2.05	25.5
	NOR-1998-ROG-015	11.03.1998	Rogaland	Kidney	57.5	2.00	22.7
	NOR-1998-ROG-016	11.03.1998	Rogaland	Kidney	54.0	1.97	21.2
	NOR-1998-ROG-017	11.03.1998	Rogaland	Kidney	232.0	2.04	24.8
	NOR-1998-ROG-018	11.03.1998	Rogaland	Kidney	0.6	0.75	-
	NOR-1998-ROG-019	11.03.1998	Rogaland	Kidney	21.6	1.95	29.2
	NOR-1998-ROG-020	11.03.1998	Rogaland	Kidney	12.8	2.15	23.8
	NOR-1998-ROG-021	11.03.1998	Rogaland	Kidney	28.2	2.14	26.6
	NOR-1998-ROG-022	11.03.1998	Rogaland	Kidney	7.9	1.79	21.9
	NOR-1998-ROG-023	11.03.1998	Rogaland	Spleen	99.9	1.99	-
	NOR-1998-ROG-024	11.03.1998	Rogaland	Spleen	62.3	2.03	23.2
	NOR-1998-ROG-025	11.03.1998	Rogaland	Spleen	116.2	1.99	26.4
	NOR-1998-ROG-026	11.03.1998	Rogaland	Spleen	84.0	1.99	23.1
	NOR-1998-ROG-027	11.03.1998	Rogaland	Spleen	60.0	2.01	21.5
	NOR-1998-ROG-028	11.03.1998	Rogaland	Spleen	40.6	1.92	-
	NOR-1998-ROG-029	11.03.1998	Rogaland	Spleen	46.0	1.88	27.5
	NOR-1998-ROG-3	11.03.1998	Rogaland	Spleen	107.1	1.98	19.9
	NOR-1998-ROG-031	11.03.1998	Rogaland	Spleen	54.3	2.04	29.3
	NOR-1998-ROG-032	11.03.1998	Rogaland	Spleen	25.0	2.05	22.4
11	NOR-1998-5	16.04.1998	Unknown	Plasma	-	-	22.8
	NOR-1998-034	16.04.1998	Unknown	Plasma	-	-	-
12	NOR-1999-001	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-002	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-003	17.02.1999	Unknown	Plasma	-	-	21.6
	NOR-1999-004	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-005	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-006	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-007	17.02.1999	Unknown	Plasma	-	-	17.7
	NOR-1999-008	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-009	17.02.1999	Unknown	Plasma	-	-	-
	NOR-1999-010	17.02.1999	Unknown	Plasma	-	-	-
13	NOR-2001-001	11.12.2001	Unknown	Plasma	-	-	36.2
	NOR-2001-002	11.12.2001	Unknown	Plasma	-	-	32.6
	NOR-2001-003	11.12.2001	Unknown	Plasma	-	-	31.5

	NOR-2001-004	11.12.2001	Unknown	Plasma	-	-	32.2
	NOR-2001-005	11.12.2001	Unknown	Plasma	-	-	35.6
14	NOR-2004-001	2004	Unknown	Spleen	7.8	3.27	22.2
	NOR-2004-002	2004	Unknown	Kidney	23.8	2.20	25.4
	NOR-2004-003	2004	Unknown	Heart	82.3	2.15	24.4
	NOR-2004-MRO-9	2004	Møre og Romsdal	Spleen	250.1	2.09	19.1
	NOR-2004-MRO-005	2004	Møre og Romsdal	Kidney	75.6	2.11	22.7
	NOR-2004-MRO-006	2004	Møre og Romsdal	Heart	108.1	2.11	25.2
15	NOR-2004-007	06.10.2004	Unknown	Spleen	333.6	2.09	20.8
	NOR-2004-008	06.10.2004	Unknown	Kidney	1374.4	1.98	25.4
	NOR-2004-009	06.10.2004	Unknown	Heart	120.6	2.12	23.5
	NOR-2004-010	06.10.2004	Unknown	Spleen	181.5	1.99	23.5
	NOR-2004-011	06.10.2004	Unknown	Kidney	193.7	2.02	24.5
	NOR-2004-012	06.10.2004	Unknown	Heart	2.2	*	22.5
	NOR-2004-013	06.10.2004	Unknown	Spleen	116.0	2.12	20.9
	NOR-2004-014	06.10.2004	Unknown	Kidney	463.9	1.95	23.4
	NOR-2004-015	06.10.2004	Unknown	Heart	104.2	2.09	23.7
	NOR-2004-016	06.10.2004	Unknown	Spleen	119.1	2.12	21.5
	NOR-2004-017	06.10.2004	Unknown	Kidney	34.5	2.06	22.4
	NOR-2004-018	06.10.2004	Unknown	Heart	43.0	2.06	21.2
16	NOR-2004-019	25.10.2004	Unknown	Plasma	-	-	-
	NOR-2004-020	25.10.2004	Unknown	Plasma	-	-	-
	NOR-2004-021	25.10.2004	Unknown	Plasma	-	-	-
17	NOR-2004-022	28.10.2004	Unknown	Spleen	27.5	1.84	22.1
	NOR-2004-023	28.10.2004	Unknown	Kidney	370.7	2.01	24.6
	NOR-2004-024	28.10.2004	Unknown	Heart	162.6	2.04	28.8
	NOR-2004-025	28.10.2004	Unknown	Spleen	270.9	1.94	22.4
	NOR-2004-026	28.10.2004	Unknown	Kidney	80.3	2.00	23.9
	NOR-2004-027	28.10.2004	Unknown	Heart	336.2	2.03	26.8
	NOR-2004-028	28.10.2004	Unknown	Spleen	123.9	2.03	19.7
	NOR-2004-029	28.10.2004	Unknown	Kidney	379.5	1.90	24.7
	NOR-2004-030	28.10.2004	Unknown	Heart	1520.4	2.04	29.0
	NOR-2004-031	28.10.2004	Unknown	Spleen	51.2	1.93	30.5
	NOR-2004-032	28.10.2004	Unknown	Kidney	1581.0	1.88	24.9
	NOR-2004-033	28.10.2004	Unknown	Heart	199.8	2.07	26.0
18	NOR-2004-12	29.10.2004	Unknown	Spleen	652.7	1.82	21.7
	NOR-2004-035	29.10.2004	Unknown	Kidney	414.7	1.89	25.4
	NOR-2004-036	29.10.2004	Unknown	Heart	334.6	2.04	24.5
	NOR-2004-037	29.10.2004	Unknown	Spleen	309.2	1.87	23.2
	NOR-2004-038	29.10.2004	Unknown	Kidney	332.1	1.89	24.2
	NOR-2004-039	29.10.2004	Unknown	Heart	611.8	1.97	25.6
19	NOR-2005-13	09.06.2005	Unknown	Spleen	131.4	2.01	21.7

	NOR-2005-002	09.06.2005	Unknown	Kidney	437.2	1.86	24.8
	NOR-2005-003	09.06.2005	Unknown	Heart	206.7	1.96	26.4
	NOR-2005-004	09.06.2005	Unknown	Spleen	762.0	1.70	26.6
	NOR-2005-005	09.06.2005	Unknown	Kidney	543.2	1.78	24.9
	NOR-2005-006	09.06.2005	Unknown	Heart	545.2	0.55	23.6
20	NOR-2005-007	19.10.2005	Unknown	Spleen	103.2	2.03	30.8
	NOR-2005-008	19.10.2005	Unknown	Kidney	80.0	2.02	27.8
	NOR-2005-009	19.10.2005	Unknown	Heart	115.0	1.98	30.6
21	NOR-2005-010	16.12.2005	Unknown	Plasma	-	-	-
	NOR-2005-011	16.12.2005	Unknown	Plasma	-	-	-
	NOR-2005-012	16.12.2005	Unknown	Plasma	-	-	-
	NOR-2005-013	16.12.2005	Unknown	Plasma	-	-	-
	NOR-2005-014	16.12.2005	Unknown	Plasma	-	-	-
22	NOR-2007-001	06.06.2007	Unknown	Spleen	281.1	2.04	24.8
	NOR-2007-002	06.06.2007	Unknown	Kidney	90.3	2.05	25.3
	NOR-2007-003	06.06.2007	Unknown	Heart	149.3	2.10	30.5
	NOR-2007-004	06.06.2007	Unknown	Spleen	229.5	2.01	22.9
	NOR-2007-005	06.06.2007	Unknown	Kidney	320.2	2.02	26.8
	NOR-2007-006	06.06.2007	Unknown	Heart	114.4	2.10	28.7
	NOR-2007-15	06.06.2007	Unknown	Spleen	266.7	2.07	22.1
	NOR-2007-008	06.06.2007	Unknown	Kidney	344.5	1.99	25.5
	NOR-2007-009	06.06.2007	Unknown	Heart	271.2	2.08	25.6
23	NOR-2007-010	19.06.2007	Unknown	Spleen	160.8	1.94	26.7
	NOR-2007-011	19.06.2007	Unknown	Kidney	117.2	2.01	26.4
	NOR-2007-012	19.06.2007	Unknown	Heart	168.5	2.01	27.3
24	NOR-2007-013	22.06.2007	Unknown	Spleen	154.0	1.95	27.6
	NOR-2007-014	22.06.2007	Unknown	Kidney	574.2	1.96	25.6
	NOR-2007-015	22.06.2007	Unknown	Heart	616.6	2.00	24.8
	NOR-2007-016	22.06.2007	Unknown	Spleen	956.7	1.89	28.3
	NOR-2007-017	22.06.2007	Unknown	Kidney	307.3	2.01	24.5
	NOR-2007-018	22.06.2007	Unknown	Heart	555.4	1.97	24.6
25	NOR-2008-001	27.03.2008	Unknown	Spleen	213.6	2.03	27.5
	NOR-2008-002	27.03.2008	Unknown	Kidney	37.4	1.98	27,0
	NOR-2008-003	27.03.2008	Unknown	Heart	106.2	2.04	26.5

Appendix 2. Storage overview. Overview of sampling year, storage location, form and temperature for each sample set prior to RNA isolation.

Sample set	Year	Storage location	Storage form	Storage temperature (°C)
1	1987	NMBU	Serum	-20
2	1988	NVI	Plasma	-80
3	1988	NVI	Unknown	-80
4	1995	NVI	Plasma	-80
5	1996	NVI	Plasma	-80
6	1996	NVI	Plasma	-80
7	1996	NVI	Heart, spleen	-20
8	1997	NVI	Plasma	-80
9	1998	NMBU	Kidney, spleen	-20
10	1998	NMBU	Kidney, spleen	-20
11	1998	NVI	Plasma	-80
12	1999	NVI	Plasma	-80
13	2001	NVI	Plasma	-80
14	2004	NMBU	Whole fish	-20
15	2004	NMBU	Whole fish	-20
16	2004	NVI	Plasma	-80
17	2004	NMBU	Whole fish	-20
18	2004	NMBU	Whole fish	-20
19	2005	NMBU	Whole fish	-20
20	2005	NMBU	Whole fish	-20
21	2005	NVI	Plasma	-80
22	2007	NMBU	Whole fish	-20
23	2007	NMBU	Whole fish	-20
24	2007	NMBU	Whole fish	-20
25	2008	NMBU	Whole fish	-20

Appendix 3. Whole fish overview. Overview of weight of dissected fish and samples originating from each fish.

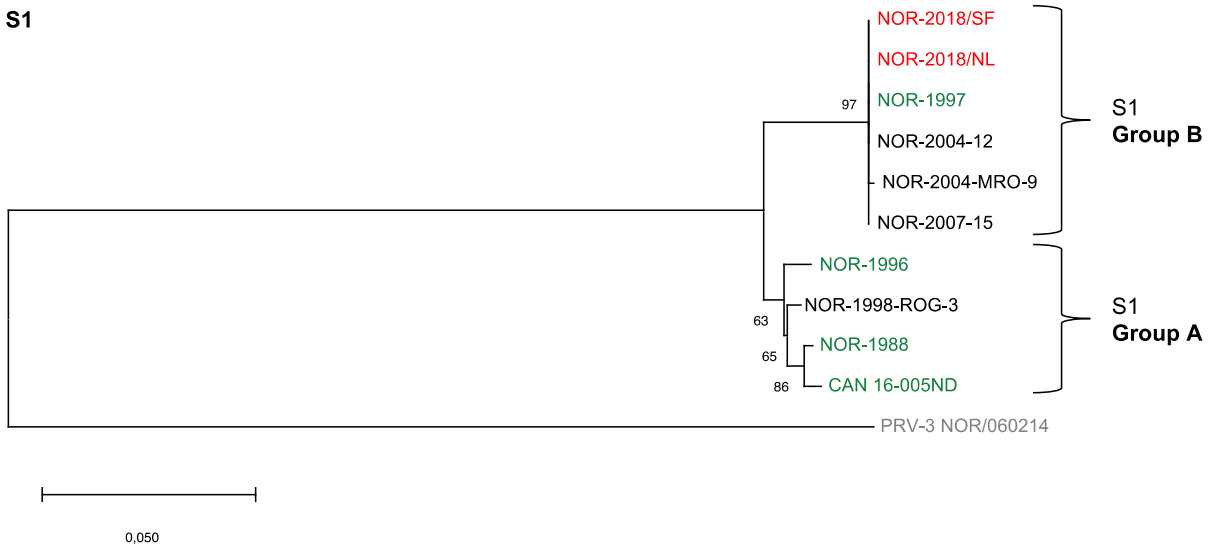
Sample set	Fish number	Weight (gram)	Samples
14	1	65	NOR-2004-001, NOR-2004-002, NOR-2004-003
	2	203	NOR-2004-MRO-9, NOR-2004-MRO-005, NOR2004-MRO-006
15	3	210	NOR-2004-007, NOR-2004-008, NOR-2004-009
	4	220	NOR-2004-010, NOR-2004-011, NOR-2004-012
	5	230	NOR-2004-013, NOR-2004-014, NOR-2004-015
	6	220	NOR-2004-016, NOR-2004-017, NOR-2004-018
17	7	820	NOR-2004-022, NOR-2004-023, NOR-2004-024
	8	655	NOR-2004-025, NOR-2004-026, NOR-2004-027
	9	816	NOR-2004-028, NOR-2004-029, NOR-2004-030
	10	846	NOR-2004-031, NOR-2004-032, NOR-2004-033
18	11	711	NOR-2004-12, NOR-2004-035, NOR-2004-036
	12	571	NOR-2004-037, NOR-2004-038, NOR-2004-039
19	13	502	NOR-2005-004, NOR-2005-005, NOR-2005-006
	14	689	NOR-2005-002, NOR-2005-003, NOR-2005-004
20	15	220	NOR-2005-007, NOR-2005-008, NOR-2005-009
22	16	511	NOR-2007-15, NOR-2007-008, NOR-2007-009
	17	491	NOR-2007-004, NOR-2007-005, NOR-2007-006
	18	279	NOR-2007-001, NOR-2007-002, NOR-2007-003
23	19	690	NOR-2007-010, NOR-2007-011, NOR-2007-012
24	20	465	NOR-2007-013, NOR-2007-014, NOR-2007-015
	21	385	NOR-2007-016, NOR-2007-017, NOR-2007-018
25	22	719	NOR-2008-001, NOR-2008-002, NOR-2008-003

Appendix 4. **Differences in amino acid sequence.** Differences in amino acid sequence between the six reference isolates and the six sequenced isolates from the present study in the analyzed coding sequences. Amino acids differing from the consensus sequence or three reference isolates differing at a site (i.e. σ3, p13 and μ1) are shown in red color. Unknown amino acids are shown in grey color.

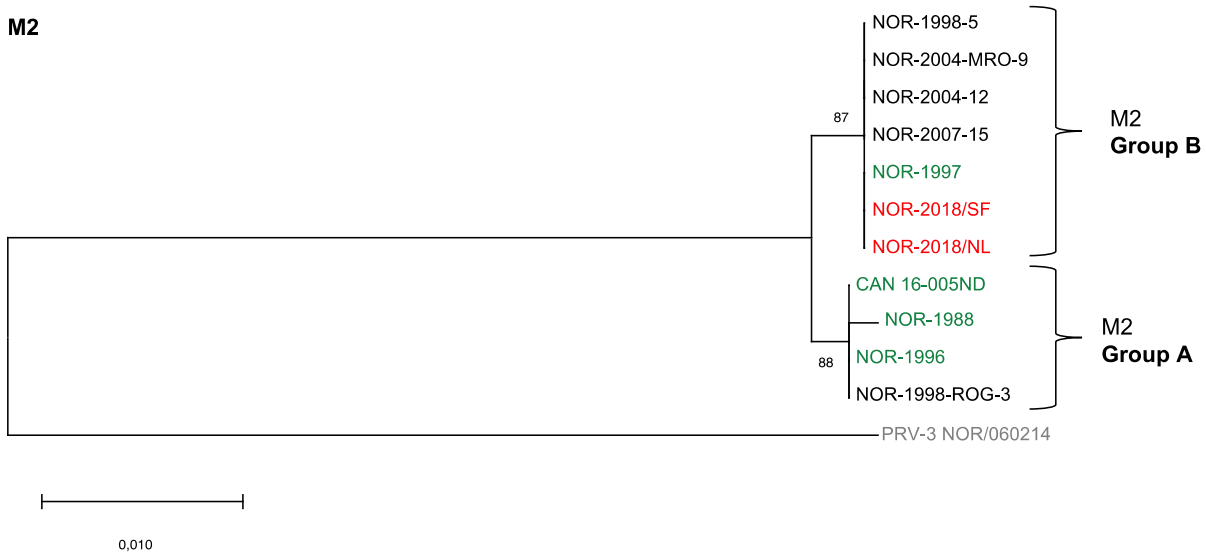
	NOR-1988	NOR-1996	NOR-1997	CAN 16-005ND	NOR-2018/NL	NOR-2018/SF	NOR-1998-ROG-3	NOR-1998-5	NOR-2004-MRO-9	NOR-2004-12	NOR-2005-13	NOR-2007-15
σ3 (S1)	S39	P39	P39	P39	P39	P39	P39	P39	P39	P39	P39	P39
	T69	T69	V69	T69	V69	V69	T69	X69	V69	V69	V69	V69
	E78	E78	D78	E78	D78	D78	E78	X78	D78	D78	D78	D78
	A85	A85	T85	A85	T85	T85	A85	X85	T85	T85	T85	T85
	T117	T117	N117	T117	N117	N117	T117	X117	N117	N117	N117	N117
	I137	I137	V137	I137	V137	V137	I137	X137	V137	V137	X137	V137
	A156	A156	T156	A156	T156	T156	A156	A156	T156	T156	X156	T156
	S157	S157	A157	S157	A157	A157	S157	S157	A157	A157	X157	A157
	K174	K174	E174	K174	E174	E174	K174	K174	E174	E174	X174	E174
	S180	S180	S180	L180	S180	S180	S180	S180	S180	S180	S180	S180
	K199	K199	K199	K199	K199	K199	K199	K199	E199	K199	K199	K199
	V206	V206	A206	V206	A206	A206	V206	V206	A206	A206	X206	A206
	I218	I218	V218	I218	V218	V218	I218	X218	V218	V218	X218	V218
	V230	V230	V230	A230	V230	V230	V230	V230	V230	V230	V230	V230
	M256	M256	M256	M256	M256	M256	M256	M256	M256	M256	X256	M256
p13 (S1)	V16	V16	A16	V16	A16	A16	V16	X16	A16	A16	A16	A16
	P23	P23	P23	H23	P23	P23	P23	P23	P23	P23	P23	P23
	T39	T39	M39	T39	M39	M39	T39	X39	M39	M39	M39	M39
	M50	M50	T50	M50	T50	T50	M50	X50	T50	T50	T50	T50
	K52	K52	I52	K52	I52	I52	K52	X52	I52	I52	I52	I52
	A74	A74	V74	A74	V74	V74	A74	X74	V74	V74	V74	V74
	R76	R76	Q76	R76	Q76	Q76	R76	X76	Q76	Q76	Q76	Q76
	Q81	Q81	R81	Q81	R81	R81	Q81	X81	R81	R81	R81	R81
	L91	L91	M91	L91	M91	M91	L91	X91	M91	M91	M91	M91
	N93	S93	N93	N93	N93	N93	N93	N93	N93	N93	N93	N93
	A117	G117	A117	A117	A117	A117	A117	A117	A117	A117	X117	A117
σ1 (S4)	V107	V107	V107	V107	V107	A107	V107	V107	V107	V107	V107	V107
	T113	T113	T113	T113	T113	T113	T113	X113	T113	T113	T113	T113
	D252	D252	D252	D252	N252	D252	D252	D252	D252	D252	D252	N252
	E254	E254	E254	E254	E254	E254	E254	E254	K254	K254	E254	E254
μ1 (M2)	S184	S184	T184	S184	T184	T184	S184	T184	T184	T184	X184	T184
	A262	A262	S262	A262	S262	S262	A262	S262	S262	S262	X262	S262
	N370	N370	D370	N370	D370	D370	N370	D370	D370	D370	D370	D370
	A389	V389	V389	V389	V389	V389	V389	V389	V389	V389	V389	V389
λ3 (L1)	N372	N372	D372	N372	N372	N372	N372	N372	N372	D372	N372	D372
	V490	V490	I490	V490	V490	V490	V490	V490	V490	I490	X490	I490
	S657	S657	S657	S657	S657	S657	S657	X657	S657	S657	S657	S657
	M719	M719	M719	M719	M719	M719	M719	M719	M719	M719	M719	T719
	N758	N758	S758	N758	N758	N758	N758	N758	N758	N758	N758	N758
	F863	F863	F863	F863	F863	F863	F863	L872	F863	F863	F863	F863
	K872	K872	K872	K872	K872	K872	K872	X872	K872	K872	K872	K872
	D937	D937	A937	D937	D937	D937	D937	D937	D937	D937	D937	D937
	L941	L941	L941	L941	L941	L941	L941	L941	L941	L941	L941	S941
	V962	V962	I962	V962	V962	V962	V962	V962	V962	V962	V962	V962
	N996	N996	S996	N996	N996	N996	N996	N996	N996	N996	N996	N996
	D1226	D1226	D1226	D1226	D1226	D1226	D1226	X1226	D1226	D1226	D1226	D1226
	L1233	L1233	L1233	L1233	L1233	L1233	L1233	X1233	L1233	L1233	L1233	L1233

	T_{1254}	T_{1254}	T_{1254}	T_{1254}	T_{1254}	T_{1254}	T_{1254}	X_{1254}	T_{1254}	T_{1254}	T_{1254}	T_{1254}
λ2	R_{876}	R_{876}	R_{876}	R_{876}	R_{876}	R_{876}	X_{876}	R_{876}	R_{876}	R_{876}	R_{876}	R_{876}
(L2)	T_{932}	T_{932}	A_{932}	T_{932}	T_{932}	T_{932}	T_{932}	T_{932}	T_{932}	T_{932}	T_{932}	T_{932}
	T_{1043}	N_{1043}	S_{1043}	T_{1043}	T_{1043}	T_{1043}	T_{1043}	N_{1043}	T_{1043}	T_{1043}	N_{1043}	T_{1043}
	V_{1055}	V_{1055}	I_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}	V_{1055}
	D_{1071}	D_{1071}	E_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}	D_{1071}
	I_{1075}	I_{1075}	I_{1075}	V_{1075}	I_{1075}	I_{1075}	I_{1075}	I_{1075}	I_{1075}	I_{1075}	I_{1075}	I_{1075}
	A_{1128}	A_{1128}	A_{1128}	T_{1128}	A_{1128}	A_{1128}	A_{1128}	A_{1128}	A_{1128}	A_{1128}	A_{1128}	A_{1128}
	P_{1141}	P_{1141}	L_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}	P_{1141}

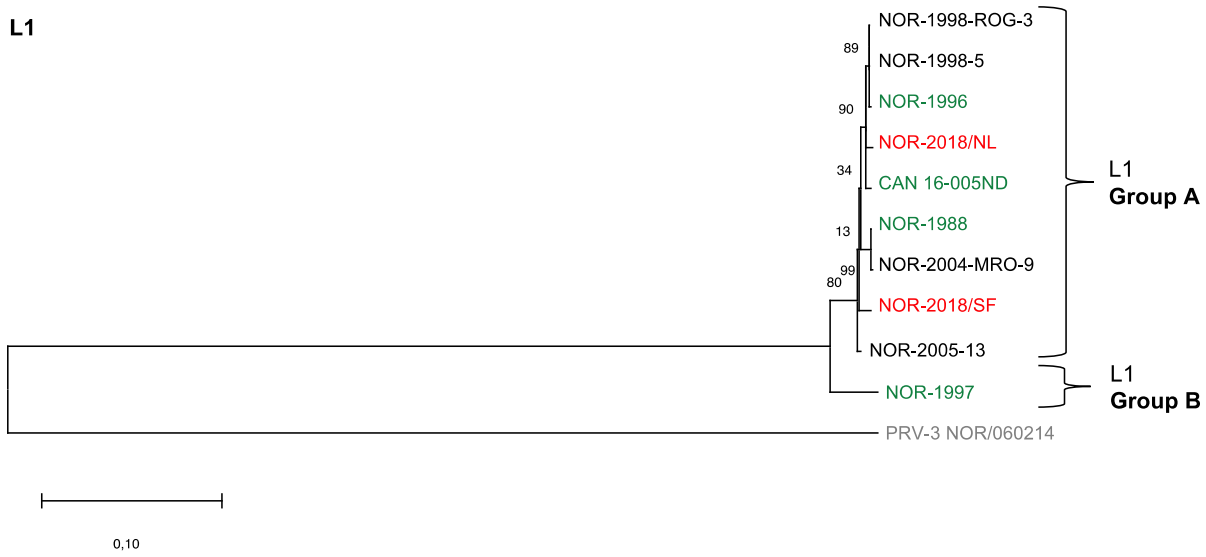
Appendix 5. Phylogenetic trees from segment S1, M2, L1 and L2. Phylogenetic trees constructed from the partial analyzed sequences from gene segment S1, M2, L1 and L2 using maximum likelihood (ML). Bootstrap values were calculated from 1000 replicates and values above 70 were considered significant.



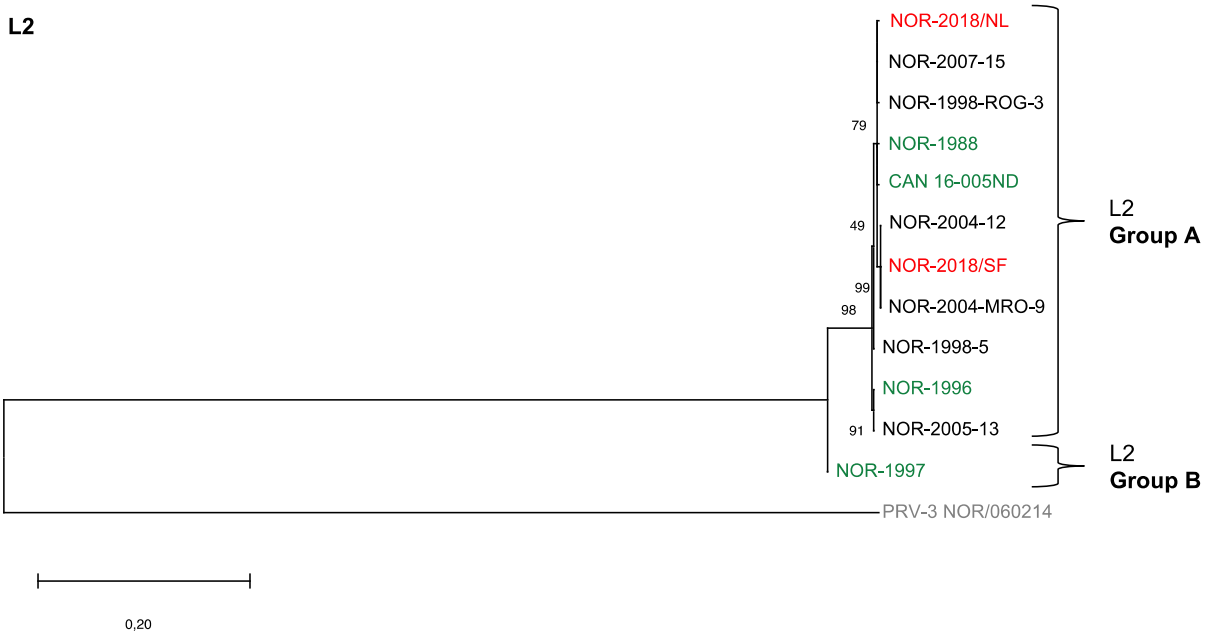
Phylogenetic tree of S1. Phylogenetic tree constructed from the nucleotide sequence of S1 (bp 21-1021). The analysis included four isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253816) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as S1 Group A and B.



Phylogenetic tree of M2. Phylogenetic tree constructed from the nucleotide sequence of M2 (bp 77-2020). The analysis included five isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253811) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as M2 Group A and B.

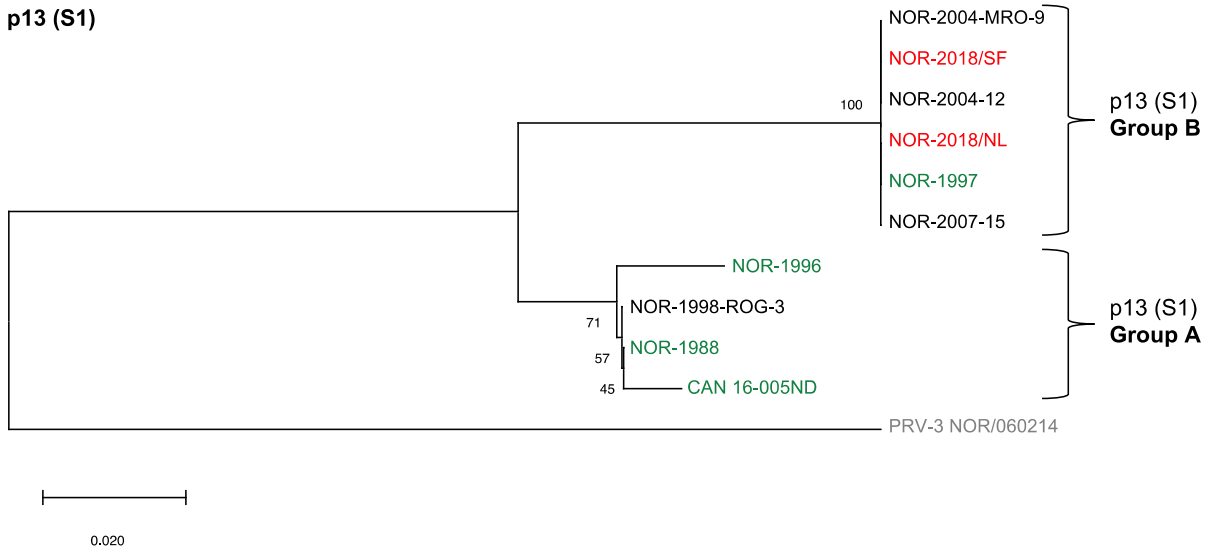


Phylogenetic tree of L1. Phylogenetic tree constructed from the nucleotide sequence of L1 (bp 1100-3788). The analysis included four isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253807) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as L1 Group A and B.



Phylogenetic tree of L2. Phylogenetic tree constructed from the nucleotide sequence of M2 (bp 1652-3772). The analysis included six isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253808) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as L2 Group A and B.

Appendix 6. Phylogenetic tree from protein p13 (S1). Phylogenetic tree constructed from the amino acid sequence p13 (aa 1-124) using neighbor joining. The analysis included four isolates from the present study (black) and six reference isolates of known virulence (low virulent in green, high virulent in red). PRV-3 NOR/060214 (MG253816) (grey) was used as outgroup. The field isolates clustered into two distinct groups annotated as p13 (S1) Group A and B. Bootstrap values were calculated from 1000 replicates and values above 70 were considered significant.



Norges miljø- og
biovitenskapelige
universitet

Postboks 5003
NO-1432 Ås
67 23 00 00
www.nmbu.no