Molecular epidemiological study on Infectious Pancreatic Necrosis Virus isolates from aquafarms in Scotland over three decades

Kristina Ulrich[1], Stefanie Wehner[2], Michaël Bekaert[1], Nicholas Di Paola[3], Meik Dilcher[4], Katherine Fiona Muir[1], John B Taggart[1], Iveta Matejusova[5], Manfred Weidmann[1]

1 Institute of Aquaculture, University of Stirling, Stirling, UK
2 Max Planck Institute of Psychiatry, Munich, Germany
3 Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil
4 Canterbury Health Laboratories, Christchurch, New-Zealand
5 Marine Science, Marine Laboratory Aberdeen, UK

Corresponding author:
Kristina Ulrich
Institute of Aquaculture
University of Stirling
FK9 4LA Stirling
United Kingdom

Tel: +4915111527452
e-mail: ulrich_kristina@yahoo.com

Contents category: RNA viruses

Running Title:
Keywords: Infectious Pancreatic Necrosis Virus, IPNV, Salmon, evolution, Codon adaptation, phylogeny, selection, vaccine, adaptation, Aquaculture

Word count abstract 181 words
Word count text 5161 words

Accepted for publication in *Journal of General Virology* published by Microbiology Society. The final published version will be available at:
http://jgv.microbiologyresearch.org/content/journal/jgv/about

Abstract

In order to obtain an insight into genomic changes and associated evolution and adaptation of Infectious Pancreatic Necrosis Virus (IPNV) the complete coding genomes of 57 IPNV isolates collected from Scottish aquafarms from 1982-2014 were sequenced and analysed.

Phylogenetic analysis of the sequenced IPNV strains showed separate clustering of genogroups I, II, III and V. IPNV isolates with genetic reassortment of segment A/B of genogroup III/II were determined. About 59% of the IPNV isolates belonged to the persistent type and 32% to the low virulent type and only one highly pathogenic strain (1.79%) was identified.

Codon adaptation index calculations indicated that the IPNV major capsid protein VP2 has adapted to its salmonid host. Underrepresentation of CpG dinucleotides in the IPNV genome to minimise detection by the innate immunity receptors, and observed positive selection in the virulence determination sites of VP2 embedded in the variable region of the main antigenic region, suggest an immune escape mechanism driving virulence evolution.

The prevalence of mostly persistent genotypes together with the assumption of adaptation and immune escape indicates that IPNV is evolving with the host.

Introduction

RNA viruses are economically important pathogens of fish, and Infectious Pancreatic Necrosis Virus (IPNV) is of particular concern for the aquaculture industry, especially for farmed Atlantic salmon (*Salmo salar*, Linnaeus 1758) and Rainbow trout (*Oncorhynchus mykiss*, Walbaum 1792). This non-enveloped aquatic virus belongs to the family of Birnaviridae and has a bi-segmented double-stranded RNA genome of about 6 kilobases (kb). IPNV causes high mortality in fry and lower mortality in adult fish. Surviving fish can become carriers, and persistently infected fish can spread IPNV by vertical transmission via oocytes or by shedding infective material into water via faeces and urine, and by milt and seminal fluids leading to clinical outbreaks [1-9].

The genome is composed of two segments. Segment A contains a large open-reading frame (ORF) encoding a polyprotein, composed of the major capsid viral protein 2 (VP2), the minor capsid protein VP3 and the Serine-Lysine protease VP4. A second ORF encodes the non-structural VP5 protein. Segment B contains only one ORF encoding the RNA-dependent RNA polymerase (RdRp/VP1) [10-14]. VP3 binds to RdRp and the dsRNA genome segments and seem to activate the RdRp [15, 16]. Furthermore, VP3 and VP4 inhibit IFNα (Interferon α) induction in salmon cells to block host immune response during infection [16]. VP5 is a non-structural arginine-rich anti-apoptosis protein, which shares homologous domains with Bcl-2 (B-cell lymphoma 2) and shuts off the host's apoptosis system by down regulation of Mcl-1 (Induced myeloid leukaemia cell differentiation protein 1). As Mcl-1 induces cells to undergo apoptotic cell death but Bcl-2 blocks apoptosis, VP5, mimicking Bcl-2, enhances cell survival for the benefit of IPNV proliferation [17].

The major capsid protein VP2 is more variable in the coding region than RdRp, VP3 and VP4 sequences and contains a variable domain in the central region between

the amino acid residues 183-335, which is an important antigenic site with two hypervariable regions at residues 239-257 and 271-284. Further studies revealed that mutations to threonine (T) from proline (P) at position 217 and Alanine (A) at position 221 are characteristic for virulent IPNV Sp isolates [18-20].

IPNV can have different virulence variants: virulent (T217, A221), persistent (P217, T221), avirulent (T217, T221) and low virulent (P217, A221) types. The virulent strains induce a high morbidity and mortality. Persistent strains induce a high morbidity (≤90%) and in general little to no mortality (less than 10%). Isolates with the residues T217, T221 induce a mortality of about 10%, isolates with the residues P217, A221 around 30%, and isolates with T217, A221 more than 60% [19-21].

IPNV is classified into 7 genogroups (I-VII) according to 10 known serotypes. Genogroup I contains the serotypes West Buxton, Buhl, Reno, VR299, Jasper-Dobos, Dry Mills and Jasper, genogroup II Abild, genogroup III Tellina, TV-1 and Canada1, genogroup IV Canada2 and Canada3, genogroup V Sparajub and N1, genogroup VI Hecht, and genogroup VII contains Yellowtail ascites virus and Marine Birnavirus [18, 22-24].

IPNV vaccines have been available since 1995 and include the main immunogenic protein VP2 and other components of the virus. The protection shown in the trials reached more than 80% but despite the introduction of IPNV vaccines to the farmed stock, IPNV outbreaks were still reported. DNA vaccines have been developed as well but currently cannot be administered on a bigger scale for regulatory reasons [25, 26]. IPNV outbreaks in Norwegian aquaculture continued up to 2010 and dropped from 2011 onward, when commercial salmon strains with an identified quantitative trait locus (QTL) for IPNV resistance became available [27]. IPNV resistant fish seem to be more effective for disease control than vaccines, as protective immune response in salmon can show big variations depending on their genetic background

and the IPNV strains involved [28]. IPNV vaccines used today are administered via intraperitoneal injection and are inactive whole virus vaccines or contain fragments of the major capsid protein VP2.

This project studied the molecular epidemiology of IPNV on aquafarms in Scotland by genome sequence analysis of IPNV isolates spanning more than three decades (1982-2014). Several phylogenetic tools were employed to infer phylogeny and to track the evolution of the virus.

Results

Sequencing results

Fifty-seven IPNV samples from an extensive diagnostic collection from 1982 – 2014 were randomly selected using equal probability of selection in MINITAB v.18.1.0 [29] and subjected to whole genome sequencing (WGS) according to sequencing protocols 1 or 2 (SP1, SP2) using high-throughput sequencing on the Illumina MiSeq (Illumina, USA) platform.

The library fragment sizes ranged from 154 bp to 473 bp as determined by capillary electrophoresis. We observed significant differences ($P<0.0001$) in the initial RNA concentration measured by spectrometry (NanoDrop ND-1000, Thermo Fisher Scientific, USA) and in the ds-cDNA concentration measured on the Qubit 2.0 (Life Technologies, USA) ($P<0.0001$) between SP1 and SP2 with higher concentrations in samples processed with SP1.

Sequencing reads were then used to estimate the genome coverage based on a reference genome for segment A and B (Figure 1A). There was no difference in the mean coverage between the two protocols ($P=0.5203$ segment A; $P=0.0695$ segment B). The maximum mean coverage for segment A was 18004.51 and for segment B 38575.39. The minimum mean coverage for segment A was 0.88, and for segment B 5.26 (Figure 1B).

After assembling the IPNV genomes with the *de novo* approach, nucleotides at the termini were missing in most samples (Figure 1C) with a significant difference between the SPs ($P=0.0143$). For segment A missing nucleotides ranged from 1 bp to 92 bp at the 5’-end and 1 bp to 157 bp at the 3’-end, and for segment B from 1-26 bp at the 5’-end and 1-37bp at the 3’-end. Therefore, for sequence comparison of all 57 IPNV isolates, only the ORFs were used.

Segment A showed a higher number of missing nucleotides than segment B.

Whole genome *de novo* assemblies with no missing nucleotides at the termini and therefore without manual editing were achieved in samples IPNV60, 011, 019, 021, 062, 066, 071 and 081 with both SPs.

Phylogenetic analysis

The 57 IPNV sequences determined from isolates in Scotland spread across genogroups I, II, III and V (Figure 2 and Figure 3).

Distribution of genogroups revealed that 78.95% of the 57 IPNV sequences belong to genogroup V and only 5.26% to genogroup I. Of the remaining isolates 15.79% of segment A sequences cluster within genogroup III and 8.77% of segment B sequences cluster within genogroup II and 7.02% within the genogroup III indicating reassortment.

VP2 analysis

To analyse and compare the VP2 proteins, IPNV segment A sequences were aligned with the virulent reference Sp strain NVI-001 with focus on known virulence markers at amino acid residues 217, 221, 247, 252, 281, 282 and 319 [19, 30, 31].

All genogroup III strains shared the same amino acid residue pattern (Table 1). Two genogroup I isolates, IPNV060 from 1983 and IPNV061 from 1986, shared the same residue pattern whereas IPNV081 from 1994 has different amino acids in positions 217, 247 and 282. Within the genogroup V strains, a range of amino acid variations in all decades and hosts was seen.

Thirty-three isolates (58.93%) were found to belong to the persistent type (P217, T221), followed by 18 (32.14%) of low virulent (P217, A221) type. Only one

sequence (1.79%) shows a virulent pattern and no avirulent isolates (T217, T221) were found. Previously not described patterns were determined in one (1.79%) isolate (P217, S221) and shared by three (5.36%) more isolates (A217, T221), previously found in isolates in Chile and Mexico [32, 33]. The persistent types (P217, T221) were found in samples from all three decades extracted from *S. salar* as well as from *O. mykiss*. Low virulent types were exclusively extracted from *S. salar* starting in the 1990's.

Previous studies have shown that IPNV attenuates in chinook salmon embryo (CHSE-214) cell culture after several passages, with amino acid changes at VP2 position 217 and 221 e.g. from virulent T217, A221 to avirulent T217, T221. All original diagnostic cultures had been performed on CHSE-214 cells (1-2 passages). To rule out, that the VP2 residues observed in the IPNV sequences determined here were not biased due to cell culture attenuation, 5 randomly selected IPNV isolates were additionally cultured for 3 passages on rainbow trout gonad (RTG-2) cells which are known not to introduce cell culture attenuation mutations in these specific VP2 amino acid positions [19, 20].

The analysis of the whole genome sequences obtained from the selected isolates cultured on RTG-2 cells (2-3 passages) showed no differences in amino acid positions P217 and T221 in the VP2 gene whether grown in CHSE-214 and TO or RTG-2 cells, and independent of the number of passages.

Network analysis

The network analysis of VP2 reveals a separate cluster representing early strains from the eighties connected to the main network cluster which is composed of persistent (P217, T221) and low virulent (P217, A221) IPNV isolates. Four samples, all of the persistent type (P217, T221), were located outside the main cluster (Figure

4). The only virulent T217, A221 isolate found in this study (sample IPNV053,) projects from the main cluster with only one amino acid difference. It derives directly from samples IPNV009 and IPNV013, with the amino acid composition PAAVTNA (Cluster 7, Table 1).

Codon adaptation index analysis

The codon adaptation index (CAI) was calculated for each IPNV protein to assess codon usage adaptation towards the salmon and the rainbow trout genome (Figure 5). Proteins VP2, VP3 and VP4 have a CAI above 1.00, indicating adaptation to both, salmon and rainbow trout genomes, although the adaptation in VP2 and VP4 is comparably higher than in VP3. Furthermore, VP2 showed a significantly higher adaptation towards the salmon codon usages than that of the rainbow trout ($P=4.867.10^{-5}$). There was also a significant difference in the CAI of VP5 between the two hosts ($P=1.544.10^{-8}$). However, a completely random codon usage for both hosts was observed in VP5 (CAI < 1.00) indicating that there is no evidence of adaption of VP5 to the codon choices of either host. The polymerase showed only a minor codon usage adaptation (mean nCAI=1.003 for both hosts).

Analysing CAI over time (Figure 6), an increase of codon usage adaptation was evident for both hosts in VP2 and VP4, with the highest increase in the 1980's and 1990's. In VP3 a fluctuation in codon adaptation over time for both hosts could be observed. The polymerase showed fluctuations over time but no signs of overall codon adaptation to either host.

Finally, CAI values for all IPNV proteins of the genomes of persistent (P217, T221) and low virulent (P217, A221) IPNV isolates were determined (see Supplementary Tables S1 and S2). Significant differences of the CAI for VP1 and VP2 were observed both in the genomes of the low virulent and persistent isolates towards the

salmon genome, whereas VP3 appeared to develop more against the rainbow trout genome.

Selection analysis

The results of the SLAC (Single Likelihood Ancestor Counting) analysis for positive selection show that only site 217 in the VP2 protein was under significant positive selection (P=0.087). Estimation of positively selected sites by applying MEME (Mixed Effects Model of Evolution) showed that all IPNV proteins have sites of positive selection (Supplement Table S3). In VP1 these sites were distributed across the gene. In VP2, sites 217, 248 and 315 were under positive selection. VP3 showed sites of selection within the RNA-binding domain. Sites under positive selection pressure in VP4 were not within the described active region of the protease. Sites 2-68 in VP5 are all within the truncated region of the protein.

Dinucleotide composition

The results of the dinucleotide composition analysis show that for both segments, the CpG dinucleotides were underrepresented, with an $CpG_{O/E}$ value of 0.72 for Segment A and 0.74 for Segment B.

Discussion

To analyse the phylogeny of IPNV samples from Scottish aquafarms from three decades, the whole genome of 57 IPNV isolates, without further clinical information, was sequenced by high-throughput sequencing. Only a few of the sequenced IPNV isolates from 1982-1992 belong to genogroups I, II or III. Over 70% of the sequenced IPNV isolates in this study belong to genogroup V, as also seen elsewhere [30, 34-36].

Segment A/B reassortment of genogroup III/II viruses was seen in five isolates. Natural reassortment of genogroup I/II viruses was observed in 1999 in Canada and Spain [37],[38]. Cases of natural re-assortment were also reported in IBDV, a Birnavirus infecting chicken [39].

Surveillance of IPNV prevalence since the seventies has accumulated evidence for transmission of IPNV from wild fish to aquafarm species and for spill back from farms to wild fish populations [40].

The detected IPNV re-assortants isolated from 1987-89, in the early aquaculture intensification phase in Scotland may have been generated in wild fish and then infected the farmed salmon and trout but were obviously not able to persist in the farmed salmonid populations.

Infections with the high virulent (T217, A221) IPNV strains induce mortalities after sea water transfer in Atlantic salmon stressed from smoltification [41], whereas persistent (P217, T221) strains cause subclinical infections [19], [31]. The two most prevalent variants in this study were the persistent P217, T221 and the low virulent P217, A221 types. Only one virulent isolate T217, A221 was found.

Similar results were obtained in studies by Ruane et al. [30] and Bain et al. [34]. Furthermore, it is known that variants with the persistent P217, T221 motif caused high mortality in rainbow trout [42].

Only two amino acid substitutions are required to move from the persistent to the virulent type in VP2 (Figure 7), and therefore there is a potential for virulent strains emerging in high density farmed salmonids due to stress or other selection pressures. Persistent P217, T221 types prevailed in the time before vaccine introduction in 1995. After 1995 a switch from persistent P217, T221 to low virulent P217, A221 occurred, with a continuous increase of the low virulent type over the decades (Figure 7). The IPNV resistant QTL has been shown to block the entry of the T217, A221 variant, in infected cells and in infected fish [43, 44]. To understand if in spite of this anymore T217, A221 IPNV types are currently evolving in QTL fish in Scotland, it would be necessary to study more recent isolates. In Norway the T217, A221 type was detected in the mid-90s, and dominated isolates from 1999. Recently persistent P217, T221 types have been isolated from groups of fish carrying the IPN-QTL in northern Norway (Oystein Evensen personal communication). Since there was no clinical information available for the isolates in this study, it is not possible to verify the pathogenic nature of the isolates.

A study in Norway in 2004 [45] demonstrated that amino acid substitutions can occur from the pathogenic to the low virulent type in IPNV isolates from field outbreaks with T217, A221 changing to T217, T221 in CHSE-214 cell culture after 2-3 passages and to P217, A221 after 13 passages post infection primarily describing cell adaptation.

A reversion of a non-virulent T217, T221 IPNV strain to a high virulent T217, A221 IPNV strain was found to be induced by stress in salmon [46].

The high number of persistent and low virulent strains indicates that the virus predominantly replicates at a level not harming the host, avoiding a strong immune response and possibly extended shedding of the virus. This was already suggested

for shrimp viruses, where it was shown that changes in the codon usage allowed a shift towards lower virus replication to avoid hosts immune response [47].

The efficiency of IPNV vaccines is mainly assessed in terms of their ability to suppress disease symptoms and improve fish survival [48]. In one study IPNV isolated from vaccinated farmed salmon in Norway were almost all of the high virulent T217, A221 type. When these isolates were used in challenge experiments in naïve fish they induced mortality of up to 56% and isolates of the low virulent P217, A221 type induced a mortality of more than 30%, showing that vaccinated fish harbour virulent strains. These observations indicate failure of the IPN-vaccines in the sense that they are non-sterilising vaccines, which do not completely inhibit IPNV replication and therefore permit low level virus evolution in vaccinated salmon [49, 50].

With a typical RNA virus mutation rate of 10^{-2} - 10^{-5} substitutions (mutations)/site/year, IPNV shows to acquire more non-synonymous than synonymous mutations [51, 52], the IPNV genome as a whole appears to be developing towards co-existence with the host, in spite of the specific selection pressure on the variable region of VP2 [45].

In order to analyse IPNV adaptation to salmonid host codon usage (*S. salar* and *O. mykiss*), CAI was calculated for each of the 5 IPNV proteins towards both hosts over time and for all persistent and low virulent strains.

The major capsid protein VP2 and the protease VP4 show the highest CAIs in both hosts. The adaptation of VP2 in *S. salar* is significantly higher than in *O. mykiss*.

The 20-fold higher annual production of salmon (162,817 tons in 2016) in comparison to rainbow trout (8,096 tons in 2016) farmed in Scotland [53], allows for a much higher number of IPNV replication cycles in the salmon population and consequently more scope for codon adaptation of VP2 to its salmon host.

The adaption of the codon usage in VP2 and VP4 ensures an efficient translation of the capsid protein, and the protease. Interestingly, the fastest increase of CAI over time for VP2 and VP4 correlates with the biggest increase of salmon farming in the 1980s (98.15%) and the 1990s (74.91%) [54]. The CAI for the VP1 (RdRp) is very low which is plausible, as the RdRp is strongly conserved within RNA viruses [55].

VP3 showed only minimal codon adaptation but has the highest mutation/kb rate of all IPNV proteins with preferred non-synonymous mutations [45]. VP3 blocks the hosts' immune response during infection by suppressing IFNα induction [16]. A non-optimised VP3 as observed here might be an important toggle to keep innate immune response at bay for just the right level of IPNV replication.

The CAI values for VP5 are lower than 1, indicating no codon adaptation. Previous research on IBDV showed that the ORF of VP5 has a different base usage than the ORF of VP2 suggesting that the original ORF was overprinted to create a *de novo* gene [56, 57]. Analysis of the IPNV VP5 sequences confirms this for IPNV. *De novo* genes are usually linked to viral pathogenicity and indeed VP5 is a homologue of the anti-apoptotic Bcl-2 family and shuts down hosts apoptosis system and together with VP4 inhibits IFN signalling [17, 58].

Some wild IPNV isolates express a truncated form of VP5 and some isolates do not even encode the VP5 protein but still replicate and induce mortality, thus the protein appears to be not essential for virus replication [61-64]. Due to its anti-apoptotic function VP5 might have been selected in densely farmed populations to maintain low level replication in vaccinated hosts.

Genes of mammalian viruses that encode structural proteins not involved in hosts immune system interaction show higher CAI values than non-structural proteins such as the RdRp [59]. The high CAI of VP2 indeed suggests adaptation to the host's codon usage in order to increase translational efficiency. VP2 however also contains

epitopes in the variable domain at position 183-335, which are under selection pressure (Table S3).

The CAI of Infectious salmon anaemia virus (ISAV) genes is high for structural genes N, M1 but highest for non-structural protein NS1 [60], suggesting that fish viruses may not follow the general adaptation models that were suggested for mammalian viruses. A study of CAI in shrimp viruses revealed higher CAI values only in non-structural proteins indicating that aquatic viruses of invertebrates also adapt to their host in a different way than shown for mammalian viruses [47].

Beside the CAI the CpG content can be considered as another adaptation feature, and invertebrate infecting dsRNA viruses have a higher CpG abundance than vertebrate infecting dsRNA viruses apparently because the $CpG_{O/E}$ of invertebrates themselves is 1.02 compared to 0.47 for vertebrates [61]. This suggests indeed a difference in host adaptation with invertebrates infecting dsRNA viruses. Other single strand DNA shrimp viruses also de-optimize their codon bias to not compete with the host translational system, thus possibly supporting a low replication strategy to avoid immune defence [47]. The CpG dinucleotide is under-represented in both IPNV segments with a $CpG_{O/E}$=0.72 and 0.74 for segment A and B respectively. These values are lower than the estimated mean $CpG_{O/E}$=0.88 for vertebrate infecting dsRNA viruses, representing a normal or an over-representation of CpG motifs [61]. It is possible that IPNV uses a strategy more similar to invertebrate dsRNA viruses, as viral replication on a low level ensures host survival and therefore, survival of the virus.

CpGs are also a pathogen associated molecular pattern (PAMP) recognised by innate immunity receptors such as Toll-like receptor 9. Small RNA viruses show CpG suppression in their genome, which sustains their replication as shown for IMNV with

a $CpG_{O/E}$ value of 0.61 [47, 62]. It is hypothesized that the IPNV $CpG_{O/E}$ of 0.72 and 0.74 indicates the same strategy. Possible viral immune escape in fish hosts is supported by the observation that Atlantic salmon leucocytes produce antiviral cytokines after stimulation with unmethylated CpG [63].

In order to determine sites of the IPNV genomes under selection pressure a positive selection analysis on each IPNV protein sequence alignment of the 57 sequenced IPNV samples was performed using two different algorithms.

The MEME calculation confirmed the SLAC calculation for VP2 but also revealed positive selection sites in each IPNV protein. Site 217 is one of the two virulence signature sites in the variable VP2 region. Four additional sites determined by MEME (241, 242, 248 and 315) are also within the variable VP2 domain (183-335). The immune pressure in this region can therefore result in changes to virulence.

The network analysis of non-homologous amino acid sequences within the VP2 hypervariable region of genogroup V isolates (Figure 4), shows that the main cluster of persistent and low virulent isolates evolves in the constraints of a certain sequence space from which, as shown by strains IPNV072 or 014, variants with specific new virulence amino acid signatures or even a virulent strain IPNV053 can emerge by acquiring the second amino acid change.

VP1(RdRp) sites calculated under positive selection pressure were mostly found outside conserved motifs I-VI. Only site 316 in motif I and site 467 in motif IV experienced positive selection pressure. These conserved motifs are crucial for the RdRp function [64].

VP3 showed one site under positive selection pressure within the self-binding domain and one within the dsRNA binding domain. Otherwise, no sites were under selection pressure within the RdRp binding domain, indicating that this region is quite conserved.

In VP4 no catalytic active sites are affected by positive selection. It is quite obvious that these sites are crucial to protease function and cannot undergo changes.

In VP5 the first 6 sites under positive selection pressure are within the truncated region of the protein indicating high local variability. As stated above VP5 appears not being essential, as some IPNVs have a truncated VP5, and some IPNVs lack VP5 but still are capable to induce mortality [65-68]. The role of VP5 remains to be elucidated.

Conclusion

The phylogenetic and genomic analyses of IPNV performed in this study indicate that selection pressures result in evolution of IPNV in farmed salmon and rainbow trout by (i) reassorting the dsRNA segments between different genogroups, (ii) changes within the hypervariable region in the VP2 protein, in favour of virulence beneficial to low level virus replication, (iii) adapting the codon usage of the major capsid protein VP2 and the VP4 for better maturation. These processes can lead to the emergence of new virulence amino acid signatures in VP2 and even to the emergence of fully virulent strains.

These mechanisms are possible due to (i) a high mutation rate of IPNV, (ii) high virus replication and enrichment in dense fish populations on farms and interactions with wild animals, (iii) the use of obviously non-sterilising IPNV vaccines not preventing low level IPNV replication in farmed vaccinated fish.

The high prevalence of persistent and low virulent strains shows that IPNV is tentatively evolving towards the host to lower replication levels with reduced disease impact on the host but sustained virus shedding.

Methods

Cell culture

CHSE-214 [69] and TO (Atlantic salmon head kidney leucocytes) cells [70] were used to propagate IPNV isolates. RTG-2 cells [71] were used to test that attenuation had not influenced IPNV sequence outcome. CHSE-214 and TO cell lines were grown at 22°C and 4% CO_2 in Eagle's Minimum Essential Medium (EMEM) with 10% Foetal Bovine Serum (FBS), 1x non-essential amino acids and 2 µM L-Glutamine. RTG-2 cells were grown at 22°C in Leibovitz-15 medium (L-15) with 10% Foetal Bovine Serum (FBS). In order to maintain a healthy and even monolayer cells were passaged regularly.

Fifty-seven IPNV isolates (Table S4) were defrosted and kept on ice. Simultaneous inoculation or inoculation via adsorption was performed using 24 well plate. HBSS+2% FBS was used as a diluent and as a negative control. The plate was incubated at 15°C and 1% CO_2 (CHSE-214, TO) or without CO_2 (RTG-2). Every day the plate was examined for cytopathic effect (CPE).

Supernatant of wells with a visible CPE was collected and centrifuged at 2500xg and 4°C for 15 min in order to be passaged into a 25 cm^2-tissue culture flask. The flasks were incubated for 1h at 15°C and 1% CO_2 (CHSE-214, TO) or without CO_2 (RTG-2) for inoculation. After incubation, the inoculum was discarded and 5 mL EMEM or Leibovitz L-15 medium +10% FBS media were added respectively. The flasks were incubated at 15°C and and 1% CO_2 (CHSE-214, TO) or without CO_2 (RTG-2) and examined every day for CPE.

When a full CPE was visible in the 25 cm^2-tissue culture flask, the supernatant was collected into a 15 mL falcon tube and centrifuged at 2500x*g* and 4°C for 15 min. The supernatant was filtered through a 0,2 µm sterile filter and stored at -70°C or used immediately for downstream applications.

RNA extraction, ds-cDNA synthesis and sequencing

Sequencing protocol 1 (SP1): IPNV RNA was extracted and processed as described [72]. Library preparation and normalization was prepared using Illumina Nextera XT DNA Library Preparation Kit (Illumina, UK). The library was checked on the Bionalyzer following the manufacturer's instructions before it was sequenced on the MiSeq (Illumina, UK). Paired-end sequencing with 2 x 75 bp was performed according to the MiSeq Reagent Kit (v3; 150 cycles) protocol.

Sequencing protocol 2 (SP2): IPNV RNA was extracted via Agencourt RNAdvance Blood Kit (Beckman Coulter, UK) according to Agencourt RNAdvance Blood 2 mL Tube Protocol with following changes: (i) 400 µL of filtered cell culture supernatant was used. (ii) The DNase step was performed by adding 100 µL of the DNase solution, containing 85 µL water, 10 µL 10x DNase buffer and 5 µL DNase I and incubating at 37°C for 5 min. Eluted RNA was stored at -70°C until further processing. RNA concentration was estimated using the NanoDrop.

10 µL of extracted RNA were converted into ss-cDNA using the Super Script III RT Kit (Life Technologies, UK) and ds-cDNA was synthesised using the NEBNext® mRNA Second Strand Synthesis Module (New England Biolabs, UK) by adding 20 µL of ss-cDNA to 60 µL of the ds-cDNA master mix and following the manufacturer's instructions.

Ds-cDNA was cleaned by adding 80 µL of Ampure XP beads (Beckman Coulter, UK) per sample (1x) and incubation for 5 min at room temperature. The sample was placed into a magnet rack, so the clear supernatant could be discarded. The magnetic beads pellet was washed twice with 200 µL of freshly prepared 70% Ethanol, while keeping the tube in the magnet and dried in the magnet for 5 min at room temperature to remove all ethanol traces. The pellet was eluted in 25 µL

molecular grade water and transferred into a new Eppendorf DNA LoBind 1.5 mL tube. The DNA was stored at -20°C until used for library preparation as described in SP1.

Several approaches to improve the sequencing yield in the ds-cDNA synthesis steps including use of single strand binding protein GP32, additional cDNA primers for the 3`-end, providing hexamers and dNTPs and incubation of RNA at 75°C and 95°C before synthesis did not significantly improve results.

Raw reads were filtered in order to remove adapters and low-quality reads by using PRINSEQ v.0.20.4 [73] and TRIMMOMATIC v.0.35 [74]. DECONSEQ v.0.4.3 was used to remove contaminating reads from other organisms (databases: Salmon, Trout, PhiX) [75]. Down sampling was performed with KHMER v.2.0, which eliminates redundant short reads. After that the reads were used for a *de novo* assembly via SPADES v.3.7.1 [76]. In parallel, reference-based assemblies were created using TANOTI v.1.2 [77] and STAMPY v.1.0.28 [78] with NC_001915.1 (segment A) and NC_001916.1 (segment B) as reference sequences. The resulting sequences were compared using BLAST v.2.6.0+ (blastn, against reference NC_001915.1 and NC_001916.1) to find the most complete genome [79]. *De novo* assembled sequences were preferred as potential insertions, deletions and rearrangement of the genome are taken into account. The reference-based assembly was used to eventually extend the *de novo* sequences at the genome termini. BOWTIE2 v.2.3.4.1 [80] and SAMTOOLS v.1.3.1 [81] were used to map the MiSeq reads to the assembled genome to evaluate the coverage of the assembled segments.

For single protein analysis, the sequences were aligned to the reference genome using CLUSTAL OMEGA v.1.2.4 [82, 83] and the genome annotation was transferred from the reference sequence using RATT [84].

Phylogenetic analysis

BEAUTI v1.8.3 was used to analyze the sequences of segment A and segment B of 57 IPNV isolates and additional reference genomes (Supplemental Table S5 and S6) with the following parameters: 10,000,000 chains per cycle with every 1000th chain being recorded, using the HKY substitution model, a strict clock and a coalescent tree [85, 86]. BEAST v1.8.3 was used to perform the calculation of the tree [87].

In order to assess if IPNV has been adapting its codon usage in any of the 5 IPNV proteins to that of the host organisms salmon (*S. salar*) or rainbow trout (*O. mykiss*) to optimise virus protein translation, a codon adaptation analysis was performed with limit of significance of $P \leq 0.05$ according to the protocol [88], [89, 90].

To perform positive selection calculation on HYPHY v.2.2, aligned sequences of 57 IPNV protein sequence alignments with a corresponding tree file were used to perform MEME and SLAC analysis with limit of significance of $P \leq 0.1$ according to the protocol [91].

To estimate the abundance of dinucleotides in IPNV segment A and B sequences, EMBOSS COMPSEQ v.1.12.0 [92] was used. The expected frequencies were calculated based on the single base frequency of input sequence and not by assumption, that every dinucleotide has equal frequency. A CpG abundance is considered low when $CpG_{observed/expected}$ ($CpG_{O/E}$) ≤ 1 and high when $CpG_{O/E} \geq 1$.

SPLITSTREE v.4.14.6 was used to generate Neighbor Network of the IPNV VP2 protein using amino acid alignments containing only non-homologue amino acids [93].

Acknowledgements

This project was funded through the Scottish Government, The Strategic Partnership for Animal Science Excellence (SPASE) consortium and the Marine Alliance for Science and Technology Scotland.

We thank Dr. Thompson (MRC Virology, Glasgow University) for providing details on her sequencing protocol and Warren Murray (Marine Scotland Science) for culturing selected IPNV viruses for the study.

The authors declare no conflict of interests.

Abbreviations

A (alanine); Ab (Abild); Bcl-2 (B cell lymphoma 2); CAI (codon adaptation index); CHSE (chinook salmon embryo cells); CPE (cytopathic effect); CpG (CG site); EMEM (Eagle's minimum essential medium); FBS (foetal bovine serum); IBDV (infectious bursal disease virus); IFNα (interferon α); IMNV (infectious mynecrosis virus); HBSS (Hank's balanced salt solution); HVR (hypervariable region); IPNV (infectious pancreatic necrosis virus); ISAV (infectious salmon anaemia virus); Ja (Jasper); L-15 (Leibovitz's medium); Mcl-1 (Induced myeloid leukaemia cell differentiation protein 1); MEME (Mixed Effects Model of Evolution); N (asparagine); NS (non-structural protein); ORF (open reading frame); P (proline); PAMP (pathogen associated pattern); QTL (quantitative trait locus); RdRp (RNA dependent RNA polymerase); RTG (rainbow trout gonad); SLAC (Single Likelihood Ancestor Counting); Sp (Sparajub); SP (sequencing protocol); T (threonine); TO (head kidney leukocytes from Atlantic salmon); V (valine); VP (viral protein); WB (West Buxton); WGS (whole genome sequencing).

References

1. **Munro ES, Midtlyng PJ**. *Infectious pancreatic necrosis and associated aquatic birnaviruses* 2011. Vol. 3. Wallingford, UK: CABI
2. **Ronneseth A, Pettersen EF, Wergeland HI**. Flow cytometry assay for intracellular detection of Infectious Pancreatic Necrosis virus (IPNV) in Atlantic salmon (Salmo salar L.) leucocytes. *Fish Shellfish Immunol* 2012;33(6):1292-1302.
3. **Ahne W, Thomsen I**. Infectious pancreatic necrosis: detection of virus and antibodies in rainbow trout IPNV-carrier (Salmo gairdneri). *Zentralbl Vet Reihe B* 1986;33(7):552-554.
4. **Wood EM, Snieszko SF, Yasutake WT**. Infectious pancreatic necrosis in brook trout. *AMA Arch Pathol* 1955;60(1):26-28.
5. **Wolf K**. *Fish Viruses and Fish Virus Diseases*. Ithaca, NY: Cornell University Press; 1988.
6. **Smail DA, Bruno DW, Dear G, McFarlane LA, Ross K**. Infectious pancreatic necrosis (IPN) virus Sp serotype in farmed Atlantic salmon, Salmo salar L., post-smolts associated with mortality and clinical disease. *J Fish Dis* 1992;15(1):77-83.
7. **McKnight IJ, Roberts RJ**. The pathology of infectious pancreatic necrosis. I. The sequential histopathology of the naturally ocurring condition. *Br Vet J* 1976;132(1):76-85.
8. **Roberts RJ, McKnight IJ**. The pathology of infectious pancreatic necrosis. II. Stress-mediated recurrence. *Br Vet J* 1976;132(2):209-214.
9. **Wolf K, Dunbar CE, Pyle EA**. Infectious Pancreatic Necrosis of Trout: II. Experimental Infections with Brook Trout. *The Progressive Fish-Culturist* 1961;23(2):61-65.
10. **Dobos P**. The molecular biology of infectious pancreatic necrosis virus (IPNV). *Annu Rev Fish Dis* 1995;5:25-54.
11. **Dobos P**. Protein-primed RNA synthesis in vitro by the virion-associated RNA polymerase of infectious pancreatic necrosis virus. *Virology* 1995;208(1):19-25.
12. **Duncan R, Dobos P**. The nucleotide sequence of infectious pancreatic necrosis virus (IPNV) dsRNA segment A reveals one large ORF encoding a precursor polyprotein. *Nucleic Acids Res* 1986;14(14):5934.
13. **Duncan R, Nagy E, Krell PJ, Dobos P**. Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions. *J Virol* 1987;61(12):3655-3664.
14. **Petit S, Lejal N, Huet J-C, Delmas B**. Active residues and viral substrate cleavage sites of the protease of the birnavirus infectious pancreatic necrosis virus. *J Virol* 2000;74(5):2057-2066.
15. **Pedersen T, Skjesol A, Jørgensen JB**. VP3, a structural protein of infectious pancreatic necrosis virus, interacts with RNA-dependent RNA polymerase VP1 and with double-stranded RNA. *J Virol* 2007;81(12):6652-6663.
16. **Lauksund S, Greiner-Tollersrud L, Chang CJ, Robertsen B**. Infectious pancreatic necrosis virus proteins VP2, VP3, VP4 and VP5 antagonize IFNa1 promoter activation while VP1 induces IFNa1. *Virus Res* 2015;196:113-121.
17. **Hong J-R, Gong H-Y, Wu J-L**. IPNV VP5, a novel anti-apoptosis gene of the Bcl-2 family, regulates Mcl-1 and viral protein expression. *Virology* 2002;295(2):217-229.

18. **Blake S, Ma J-Y, Caporale D, Jairath S, Nicholson B**. Phylogenetic relationships of aquatic birnaviruses based on deduced amino acid sequences of genome segment A cDNA. *Dis Aquat Organ* 2001;45(2):89-102.
19. **Santi N, Vakharia VN, Evensen Ø**. Identification of putative motifs involved in the virulence of infectious pancreatic necrosis virus. *Virology* 2004;322(1):31-40.
20. **Song H, Santi N, Evensen Ø, Vakharia VN**. Molecular determinants of infectious pancreatic necrosis virus virulence and cell culture adaptation. *J Virol* 2005;79(16):10289-10299.
21. **Mutoloki S, Munang'andu H, Evensen Ø**. Clinical and subclinical forms of infectious pancreatic necrosis virus infections show specific viral genetic fingerprints that link differences in virulence to immunogenicity. *Fish Shellfish Immun* 2013;34(6):1667.
22. **Hill B, Way K**. Serological classification of infectious pancreatic necrosis (IPN) virus and other aquatic birnaviruses. *Annu Rev Fish Dis* 1995;5:55-77.
23. **Underwood B, Smale C, Brown F, Hill B**. Relationship of a virus from Tellina tenuis to infectious pancreatic necrosis virus. *J Gen Virol* 1977;36(1):93-109.
24. **Zhang C, Suzuki S**. Aquabirnaviruses isolated from marine organisms form a distinct genogroup from other aquabirnaviruses. *J Fish Dis* 2004;27(11):633-643.
25. **Dorson M**. Vaccination against infectious pancreatic necrosis. *Fish Vaccination London, Academic Press Limited* 1988:162-171.
26. **Rimstad E**. Vaccination against Infectious Pancreatic Necrosis. *Fish Vaccination* 2014:303-312. W. R. Gudding, A. Lillehaug, Oe. Evensen (eds), Wiley & Sons Ltd, Chichester
27. **Olsen A, Hellberg H**. Fiskehelserapporten 2011. *Norwegian Veterinary Institute* 2012. (ed. by A.B. Olsen), p. 5, Norwegian Veterinary Institute, Oslo.
28. **Ramstad A, Midtlyng P**. Strong genetic influence on IPN vaccination - and - challenge trials in Atlantic salmon, Salmo salar L. *J Fish Dis* 2008;31(8):567-578.
29. *Minitab reference manual : Macintosh version, release 8*: State College, PA : Minitab, [1991]; 1991.
30. **Ruane NM, McCleary SJ, McCarthy LJ, Henshilwood K**. Phylogenetic analysis of infectious pancreatic necrosis virus in Ireland reveals the spread of a virulent genogroup 5 subtype previously associated with imports. *Arch Virol* 2015;160(3):817-824.
31. **Mutoloki S, Jossund TB, Ritchie G, Munang'andu HM, Evensen O**. Infectious Pancreatic Necrosis Virus Causing Clinical and Subclinical Infections in Atlantic Salmon Have Different Genetic Fingerprints. *Front Microbiol* 2016;7:1393.
32. **Tapia D, Eissler Y, Torres P, Jorquera E, Espinoza JC et al.** Detection and phylogenetic analysis of infectious pancreatic necrosis virus in Chile. *Dis Aquat Organ* 2015;116(3):173-184.
33. **Salgado-Miranda C, Rojas-Anaya E, Garcia-Espinosa G, Loza-Rubio E**. Molecular characterization of the VP2 gene of infectious pancreatic necrosis virus (IPNV) isolates from Mexico. *J Aquat Anim Health* 2014;26(1):43-51.
34. **Bain N, Gregory A, Raynard R**. Genetic analysis of infectious pancreatic necrosis virus from Scotland. *J Fish Dis* 2008;31(1):37-47.
35. **Holopainen R, Eriksson-Kallio AM, Gadd T**. Molecular characterisation of infectious pancreatic necrosis viruses isolated from farmed fish in Finland. *Arch Virol* 2017;162(11):3459-3471.
36. **Mutoloki S, Evensen O**. Sequence similarities of the capsid gene of Chilean and European isolates of infectious pancreatic necrosis virus point towards a common origin. *J Gen Virol* 2011;92(Pt 7):1721-1726.

37. **Romero - Brey I, Bandín I, Cutrín J, Vakharia V, Dopazo C**. Genetic analysis of aquabirnaviruses isolated from wild fish reveals occurrence of natural reassortment of infectious pancreatic necrosis virus. *J Fish Dis* 2009;32(7):585-595.
38. **Moreno P, Olveira JG, Labella A, Cutrín JM, Baro JC et al.** Surveillance of viruses in wild fish populations in areas around the Gulf of Cadiz (South Atlantic Iberian Peninsula). *Appl Environ Microb* 2014:AEM 80, 6560–6571. 02090-02014.
39. **Wei Y, Li J, Zheng J, Xu H, Li L et al.** Genetic reassortment of infectious bursal disease virus in nature. *Biochem Biophys Res Commun* 2006;350(2):277-287.
40. **Wallace IS, McKay P, Murray AG**. A historical review of the key bacterial and viral pathogens of Scottish wild fish. *J Fish Dis* 2017;40(12):1741-1756.
41. **Julin K, Johansen LH, Sommer AI, Jørgensen J**. Persistent infections with infectious pancreatic necrosis virus (IPNV) of different virulence in Atlantic salmon, Salmo salar L. *J Fish Dis* 2015;38(11):1005-1019.
42. **Ahmadivand S, Soltani M, Behdani M, Evensen Ø, Alirahimi E et al.** VP2 (PTA motif) encoding DNA vaccine confers protection against lethal challenge with infectious pancreatic necrosis virus (IPNV) in trout. *Mol Immunol* 2018;94:61-67.
43. **Moen T, Torgersen J, Santi N, Davidson WS, Baranski M et al.** Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics* 2015;200(4):1313-1326.
44. **Houston RD, Haley CS, Hamilton A, Guy DR, Mota-Velasco JC et al.** The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity*, Omit 'Original article' 2009;105:318.
45. **Skjesol A, Skjæveland I, Elnæs M, Timmerhaus G, Fredriksen BN et al.** IPNV with high and low virulence: host immune responses and viral mutations during infection. *Virol J* 2011;8(1):396.
46. **Gadan K, Sandtro A, Marjara IS, Santi N, Munang'andu HM et al.** Stress-induced reversion to virulence of infectious pancreatic necrosis virus in naive fry of Atlantic salmon (Salmo salar L.). *PLoS One* 2013;8(2):e54656.
47. **Tyagi A, Kumar BTN, Singh NK**. Genome dynamics and evolution of codon usage patterns in shrimp viruses. *Arch Virol* 2017: 162 3137-3142.
48. **Heininger U, Bachtiar N, Bahri P, Dana A, Dodoo A et al.** The concept of vaccination failure. *Vaccine* 2012;30(7):1265-1268.
49. **Read AF, Baigent SJ, Powers C, Kgosana LB, Blackwell L et al.** Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens. *PLoS Biol* 2015;13(7):e1002198.
50. **Ganusov VV, Antia R**. Imperfect vaccines and the evolution of pathogens causing acute infections in vertebrates. *Evolution* 2006;60(5):957-969.
51. **Hanada K, Suzuki Y, Gojobori T**. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol* 2004;21(6):1074-1080.
52. **Silva FM, Vidigal PM, Myrrha LW, Fietto JL, Silva A et al.** Tracking the molecular epidemiology of Brazilian Infectious bursal disease virus (IBDV) isolates. *Infect Genet Evol* 2013;13:18-26.
53. Munro AL, Wallace IS. SCOTTISH FISH FARM PRODUCTION SURVEY 2016. Edinburgh, UK: Marine Scotland Science2017.
54. **Government S**. 1979-2016. Scottish Fish Farm Production Surveys. http://www.gov.scot/Topics/marine/Publications/stats/FishFarmProductionSurveys/OlderSurveys [accessed.

657 55. **Shwed PS, Dobos P, Cameron LA, Vakharia VN, Duncan R**. Birnavirus VP1
658 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a
659 GDD motif. *Virology* 2002;296(2):241-250.
660 56. **Do Yew T, Bejo MH, Ideris A, Omar AR, Meng GY**. Base usage and dinucleotide
661 frequency of infectious bursal disease virus. *Virus Genes* 2004;28(1):41-53.
662 57. **Sabath N, Wagner A, Karlin D**. Evolution of viral proteins originated de novo by
663 overprinting. *Mol Biol Evol* 2012;29(12):3767-3780.
664 58. **Skjesol A, Aamo T, Hegseth MN, Robertsen B, Jørgensen JB**. The interplay
665 between infectious pancreatic necrosis virus (IPNV) and the IFN system: IFN
666 signaling is inhibited by IPNV infection. *Virus Res* 2009;143(1):53-60.
667 59. **Bahir I, Fromer M, Prat Y, Linial M**. Viral adaptation to host: a proteome-based
668 analysis of codon usage and amino acid preferences. *Mol Syst Biol* 2009;5:311.
669 60. **Tello M, Vergara F, Spencer E**. Genomic adaptation of the ISA virus to Salmo salar
670 codon usage. *Virology J* 2013;10(1):223.
671 61. **Cheng X, Virk N, Chen W, Ji S, Ji S et al.** CpG usage in RNA viruses: data and
672 hypotheses. *PloS one* 2013;8(9):e74109.
673 62. **Karlin S, Doerfler W, Cardon L**. Why is CpG suppressed in the genomes of
674 virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J*
675 *Virol* 1994;68(5):2889-2897.
676 63. **Jørgensen JB, Johansen A, Stenersen B, Sommer A-I**. CpG oligodeoxynucleotides
677 and plasmid DNA stimulate Atlantic salmon (Salmo salar L.) leucocytes to produce
678 supernatants with antiviral activity. *Dev Comp Immunol* 2001;25(4):313-321.
679 64. **Graham SC, Sarin LP, Bahar MW, Myers RA, Stuart DI et al.** The N-terminus of
680 the RNA polymerase from infectious pancreatic necrosis virus is the determinant
681 of genome attachment. *PLoS Pathog* 2011;7(6):e1002085.
682 65. **Santi N, Song H, Vakharia VN, Evensen Ø**. Infectious pancreatic necrosis virus
683 VP5 is dispensable for virulence and persistence. *J Virol* 2005;79(14):9206-9216.
684 66. **Heppell J, Tarrab E, Berthiaume L, Lecomte J, Arella M**. Characterization of the
685 small open reading frame on genome segment A of infectious pancreatic necrosis
686 virus. *J Gen Virol* 1995;76(8):2091-2096.
687 67. **Weber S, Fichtner D, Mettenleiter TC, Mundt E**. Expression of VP5 of infectious
688 pancreatic necrosis virus strain VR299 is initiated at the second in-frame start
689 codon. *J Gen Virol* 2001;82(4):805-812.
690 68. **Santi N, Sandtrø A, Sindre H, Song H, Hong J-R et al.** Infectious pancreatic
691 necrosis virus induces apoptosis in vitro and in vivo independent of VP5
692 expression. *Virology* 2005;342(1):13-25.
693 69. **Fryer JL, Yusha A, Pilcher KS**. The in vitro cultivation of tissue and cells of pacific
694 salmon and steelhead trout*. *Ann NY Acad Sci* 1965;126(1):566-586.
695 70. **Wergeland HI, Jakobsen RA**. A salmonid cell line (TO) for production of infectious
696 salmon anaemia virus (ISAV). *Dis Aquat Organ* 2001;44(3):183-190.
697 71. **Officer JE**. Ability of a Fish Cell Line to Support the Growth of Mammalian Viruses.
698 *Proc Soc Exp Biol Med* 1964;116(1):190-194.
699 72. **Dilcher M, Hasib L, Lechner M, Wieseke N, Middendorf M et al.** Genetic
700 characterization of Tribeč virus and Kemerovo virus, two tick-transmitted human-
701 pathogenic Orbiviruses. *Virology* 2012;423(1):68-76.
702 73. **Schmieder R, Edwards R**. Quality control and preprocessing of metagenomic
703 datasets. *Bioinformatics* 2011;27(6):863-864.
704 74. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina
705 sequence data. *Bioinformatics* 2014;30(15):2114-2120.

706 75. **Schmieder R, Edwards R**. Fast identification and removal of sequence
707 contamination from genomic and metagenomic datasets. *PloS one*
708 2011;6(3):e17288.
709 76. **Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A et al.**
710 Assembling single-cell genomes and mini-metagenomes from chimeric MDA
711 products. *J Comput Biol* 2013;20(10):714-737.
712 77. **Thomson E, Ip CL, Badhan A, Christiansen MT, Adamson W et al.** Comparison
713 of Next-Generation Sequencing Technologies for Comprehensive Assessment of
714 Full-Length Hepatitis C Viral Genomes. *J Clin Microbiol* 2016;54(10):2470-2484.
715 78. **Lunter G, Goodson M**. Stampy: a statistical algorithm for sensitive and fast
716 mapping of Illumina sequence reads. *Genome Res* 2011;21(6):936-939.
717 79. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. Basic local alignment
718 search tool. *J Mol Biol* 1990;215(3):403-410.
719 80. **Langmead B, Salzberg SL**. Fast gapped-read alignment with Bowtie 2. *Nat*
720 *Methods* 2012;9(4):357-359.
721 81. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al.** The Sequence
722 Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078-2079.
723 82. **Goujon M, McWilliam H, Li W, Valentin F, Squizzato S et al.** A new
724 bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res*
725 2010;38(suppl_2):W695-W699.
726 83. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K et al.** Fast, scalable
727 generation of high - quality protein multiple sequence alignments using Clustal
728 Omega. *Mol Syst Biol* 2011;7(539): 1-6.
729 84. **Otto TD, Dillon GP, Degrave WS, Berriman M**. RATT: Rapid Annotation Transfer
730 Tool. *Nucleic Acids Research* 2011;39(9):e57-e57.
731 85. **Kingman JFC**. The coalescent. *Stoch Proc Appl* 1982;13(3):235-248.
732 86. **Kingman JF**. On the genealogy of large populations. *J Appl Probab* 1982;19(A):27-
733 43.
734 87. **Drummond A, Suchard MA, Xie D, Rambaut A**. *Bayesian phylogenetics with*
735 *BEAUti and the BEAST 1.7* 2012; *Mol Biol Evol* 29:1969–1973.
736 88. **Di Paola N, de Melo Freire CC, de Andrade Zanotto PM**. Does adaptation to
737 vertebrate codon usage relate to flavivirus emergence potential? *PloS one*
738 2018;13(1):e0191652.
739 89. **Puigbo P, Bravo IG, Garcia-Vallve S**. E-CAI: a novel server to estimate an
740 expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008;9:65.
741 90. **Puigbo P, Bravo IG, Garcia-Vallve S**. CAIcal: a combined set of tools to assess
742 codon usage adaptation. *Biol Direct* 2008;3:38.
743 91. **Pond SLK, Frost SDW, Muse SV**. HyPhy: hypothesis testing using phylogenies.
744 *Bioinformatics* 2005;21(5):676-679.
745 92. **Rice P, Longden I, Bleasby A**. EMBOSS: the European Molecular Biology Open
746 Software Suite. *Trends Genet* 2000;16(6):276-277.
747 93. **Huson DH, Bryant D**. Application of phylogenetic networks in evolutionary
748 studies. *Mol Biol Evol* 2005;23(2):254-267.
749 94. **Glenney GW, Barbash PA, Coll JA, Quartz WM**. Isolation and molecular
750 characterization of a novel infectious pancreatic necrosis virus strain in returning
751 Atlantic salmon Salmo salar from the Connecticut River, USA. *J Aquat Anim Health*
752 2012;24(2):63-72.
753 95. **Mohr PG, Moody NJ, Williams LM, Hoad J, St JCM**. Molecular characterization of
754 Tasmanian aquabirnaviruses from 1998 to 2013. *Dis Aquat Organ* 2015;116(1):1-9.

755 96. **Cutrin JM, Barja JL, Nicholson BL, Bandin I, Blake S et al.** Restriction Fragment
756 Length Polymorphisms and Sequence Analysis: an Approach for Genotyping
757 Infectious Pancreatic Necrosis Virus Reference Strains and Other Aquabirnaviruses
758 Isolated from Northwestern Spain. *Appl Environ Microbiol* 2004;70(2):1059-1067.
759 97. **Dadar M, Peyghan R, Memari HR, Shapouri MR, Hasanzadeh R et al.** Sequence
760 analysis of infectious pancreatic necrosis virus isolated from Iranian reared
761 rainbow trout (Oncorhynchus mykiss) in 2012. *Virus Genes* 2013;47(3):574-578.
762 98. **Havarstein LS, Kalland KH, Christie KE, Endresen C**. Sequence of the large
763 double-stranded RNA segment of the N1 strain of infectious pancreatic necrosis
764 virus: a comparison with other Birnaviridae. *J Gen Virol* 1990;71 (Pt 2):299-308.
765 99. **Heppell J, Berthiaume L, Corbin F, Tarrab E, Lecomte J et al.** Comparison of
766 Amino Acid Sequences Deduced from a cDNA Fragment Obtained from Infectious
767 Pancreatic Necrosis Virus (IPNV) Strains of Different Serotypes. *Virology*
768 1993;195(2):840-844.
769 100. **Chung H, Lee S, Lee H, Lee D, Kim Y**. Nucleotide sequence analysis of the VP2-NS-
770 VP3 genes of infectious pancreatic necrosis virus DRT strain. *Mol Cells 1993; 4, 341-*
771 *354*.
772 101. **Lee H-H, Chung H-K, Lee S-H**. Nucleotide sequence analysis of the RNA-
773 dependent RNA polymerase gene of infectious pancreatic necrosis virus DRT strain.
774 *J Microbiol Biotechn* 1994;4(4):264-269.
775 102. **Yao K, Vakharia VN**. Generation of infectious pancreatic necrosis virus from
776 cloned cDNA. *J Virol*1998;72(11):8913-8920.
777 103. **Hirayama T, Nagano I, Shinmoto H, Yagyu Ki, Oshima Si**. Isolation and
778 characterization of virulent yellowtail ascites virus. *Microbiol Immunol*
779 2007;51(4):397-406.
780 104. **Zhao Z, Ke F, Li Z, Gui J, Zhang Q**. Isolation, characterization and genome
781 sequence of a birnavirus strain from flounder Paralichthys olivaceus in China. *Arch*
782 *Virol* 2008;153(6):1143-1148.
783 105. **Zhang CX, Suzuki S**. Comparison of the RNA polymerase genes of marine
784 birnavirus strains and other birnaviruses. *Arch Virol* 2003;148(4):745-758.
785 106. **Bandin I, Souto S, Cutrin JM, Lopez-Vazquez C, Olveira JG et al.** Presence of
786 viruses in wild eels Anguilla anguilla L, from the Albufera Lake (Spain). *J Fish Dis*
787 2014;37(7):597-607.
788 107. **Galloux M, Chevalier C, Henry C, Huet JC, Costa BD et al.** Peptides resulting from
789 the pVP2 C-terminal processing are present in infectious pancreatic necrosis virus
790 particles. *J Gen Virol* 2004;85(Pt 8):2231-2236.
791 108. **Dixon PF, Ngoh GH, Stone DM, Chang SF, Way K et al.** Proposal for a fourth
792 aquabirnavirus serogroup. *Arch Virol* 2008;153(10):1937-1941.
793 109. **Duncan R, Mason CL, Nagy E, Leong J-A, Dobos P**. Sequence analysis of infectious
794 pancreatic necrosis virus genome segment B and its encoded VP1 protein: a
795 putative RNA-dependent RNA polymerase lacking the Gly-Asp-Asp motif. *Virology*
796 1991;181(2):541-552.
797 110. **Shivappa R, Song H, Yao K, Aas-Eng A, Evensen Ø et al.** Molecular
798 characterization of Sp serotype strains of infectious pancreatic necrosis virus
799 exhibiting differences in virulence. *Dis Aquat Organ* 2004;61(1-2):23-32.
800 111. **Jorquera E, Morales P, Tapia D, Torres P, Eissler Y et al.** Chilean IPNV isolates:
801 Robustness analysis of PCR detection. *Electronic J Biotechnol* 2016;19(2):28-32.
802 112. **Nobiron I, Galloux M, Henry C, Torhy C, Boudinot P et al.** Genome and
803 polypeptides characterization of Tellina virus 1 reveals a fifth genetic cluster in the
804 Birnaviridae family. *Virology* 2008;371(2):350-361.

Tables

Table 1 VP2 characteristic variable residue pattern of sequenced IPNV isolates compared to the virulent reference strain excluding IPNV032.

Cluster	Sequence ID	Collection Year	Genogroup	VP2 Amino acid position						
				217	**221**	**247**	**252**	**281**	**282**	**319**
1	IPNVV006, 011, 019, 063, 064, 065, 066, 068, 069	1982, 1986, 1987, 1988, 1989, 1992	III	P	T	P	A	T	A	A
2	IPNV007, 008, 021, 062, 026, 070, 071, 073, 075, 076, 081, 083, 084, 107, 001, 003, 039	1984, 1987, 1990, 1992, 1993, 1994, 1997, 1999, 2000, 2002, 2008	I, V	P	T	A	N	T	N	A
3	IPNV027	1989	V	P	T	A	N	T	N	V
4	IPNV024, 031, 057, 047, 048, 050	2005, 2006, 2013, 2014	V	P	T	A	V	T	N	A
5	IPNV072	1992	V	P	A	A	T	T	N	A
6	IPNV087, 088, 090, 111	1997, 1998, 1999	V	P	A	A	I	T	N	A
7	IPNV009, 010, 013, 018, 030, 035, 051, 052, 054, 055, 056, 049, 058	2003, 2004, 2006, 2007, 2010, 2011, 2012, 2014	V	P	A	A	V	T	N	A
8	IPNV060, 061, 082	1983, 1986, 1996	I, V	A	T	E	N	T	A	A
9	IPNV014	2004	V	P	S	A	V	T	N	A
10	IPNV053	2011	V	T	A	A	V	T	N	A
11	Santi et al., 2004 [19]	2000	V	T	A	T	V	T	N	A

Figures

Figure 1 A. Forward and reverse reads recovered from IPNV samples processed with the SP1 or the SP2. **B.** Mean coverage of Segment A and B for 57 processed IPNV samples with the SP1 or SP2. **C** Missing terminal nucleotides of both segments of IPNV samples processed with the SP1 or the SP2 after *de novo* assembly. Values were analysed using the Mann-Whitney test.

Figure 2. Phylogenetic classification of 57 IPNV isolates based on segment A sequences. Genogroup VII consists of AY98 , YTAV_Y6 and GC1. Yellowtail ascites virus is currently grouped as a species distinct from IPNV in the genus Aquabirnavirus by ICTV. However even ICTV states: “The extremely close genetic relationship between the species, suggests that there may be a case for combining them into one“. The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000th chain being recorded. Posterior probability values are shown for different genogroups at corresponding nodes in percent. The scale represents 0.02 substitutions per nucleotide site.

Figure 3. Phylogenetic classification of 57 IPNV isolates based on segment B sequences. The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000th chain being recorded. Posterior probability values are shown for different genogroups at corresponding nodes in percent. The scale represents 0.06 substitutions per nucleotide site.

Figure 4. Neighbor Network of the VP2 protein for genogroup V IPNV samples based on non-homologous amino acid alignment. The scale bar represents the split support for the edges. Sample IPNV053 (red) is the only virulent type (T217, A221). Green samples represent persistent type (P217, T221), purple low virulent (P217, A221) and blue not assigned types (A217, T221 and P217, S221).

Figure 5. IPNV proteins have different codon usage adaptation indexes towards their hosts of Atlantic salmon and rainbow trout. Normalized CAI values above 1.00 support evidence of codon usage adaptation to that specific host. Asterisks indicate a significant difference (p<0.05) between hosts.

Figure 6. Temporal CAI analysis of each IPNV open reading frame shows differences of codon usage adaptation indexes between the IPNV host genome of Atlantic salmon and rainbow trout from 1980 to 2010. Normalized CAI values above 1.00 support evidence of codon usage adaptation to that specific host. Asterisks indicate a significant difference (p<0.05) between hosts.

Figure 7 Distribution of variants of 56 sequenced IPNV strains in each decade. Due to not assigned residues one isolate was excluded. Persistent type (P217, T221) 33 strains (58.93%), low virulent (P217, A221) 18 strains (32.14%) virulent type 1 strain (1.79%).

Supplement

Table S1 Significant differences in CAI values between hosts against a sample variant persistent (P) type and low virulent (LV) type; significance level p ≤ 0.05.

Protein	Host	Comparison	P=value
RdRp	*O. mykiss*	P - LV	0.00037
	S. salar	P - LV	0.00043
VP2	*O. mykiss*	P - LV	0.014
	S. salar	P - LV	0.01627
VP3	*O. mykiss*	P - LV	0.048

Table S2 Significant differences in CAI values persistent (P) and low virulent (LV) type of samples against host species; significance level p ≤ 0.05.

Protein	Virulence type	Comparison	P=value
VP2	P	*O. mykiss – S. salar*	0.01589
	LV	*O. mykiss – S. salar*	$8.17\ 10^{-6}$
VP5	P	*O. mykiss – S. salar*	$7.26\ 10^{-6}$
	LV	*O. mykiss – S. salar*	0.00088

Table S3 Results of site-specific positive and diversifying selection analysis of all IPNV proteins for all 57 sequenced samples using MEME on HYPHY, applying the MG94xREV codon model. P describes the P=value (p≤0.1) for positive selection at one site.

RdRp/VP1		VP2		VP3		VP4		VP5	
Codon	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**
50	0.01	195	0.02	66	0.10	37	0.09	2	0.01

81	0.00	208	0.02	145	0.10	46	0.01	13	0.19
82	0.00	217	0.00			131	0.04	17	0.03
140	0.04	241	0.00			212	0.09	36	0.09
226	0.05	242	0.03			218	0.05	53	0.07
232	0.04	248	0.00					68	0.09
316	0.00	315	0.09					98	0.08
358	0.07	333	0.07					106	0.04
368	0.05							111	0.08
467	0.01								
687	0.07								
699	0.07								
737	0.09								

Table S4 Sequenced IPNV samples. NA=not assigned, Atlantic salmon (*Salmo salar)*, rainbow trout (*Oncorhynchus mykiss), brown trout (Salmo trutta), turbot (Scophthalmus maximus).*

Sequence ID	Collection Year	Confirmed Serotype	Cell line	Host	Origin	Region
IPNV006	1982	NA	CHSE-214	*S. salar*	UK	Scotland, Argyll
IPNV060	1983	NA	CHSE-214	NA	NA	NA
IPNV007	1984	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV008	1984	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV011	1986	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV061	1986	Sp	CHSE-214	*S. maximus*	NA	NA
IPNV019	1987	NA	CHSE-214	*O. mykiss*	UK	Scotland, Highlands
IPNV021	1987	NA	CHSE-214	*O. mykiss*	UK	Scotland, Strathclyde
IPNV062	1987	NA	CHSE-214	*S. salar*	NA	NA
IPNV063	1987	Ab	CHSE-214	*S. salar*	NA	NA
IPNV064	1987	Ab	CHSE-214	NA	UK	Scotland, Highlands
IPNV065	1987	Ab	CHSE-214	NA	UK	Scotland, Highlands
IPNV066	1988	Ab	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV027	1989	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV032	1989	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV068	1989	Ab	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV026	1990	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV069	1992	Ab	CHSE-214	*S. salar*	NA	NA
IPNV070	1992	Sp	CHSE-214	NA	Norway	Norway
IPNV071	1992	Sp	TO	NA	UK	Scotland, Argyll
IPNV072	1992	Sp	CHSE-214	NA	UK	Scotland, Shetland
IPNV073	1993	Sp	CHSE-214	*S. salar*	NA	NA

IPNV075	1993	Sp	TO	*S. salar*	Norway	Norway
IPNV076	1994	Sp	TO	*S. salar*	NA	NA
IPNV081	1994	NA	CHSE-214	*S. salar*	NA	NA
IPNV082	1996	Sp	CHSE-214	*S. salar*	NA	NA
IPNV083	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV084	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV087	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV088	1997	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV090	1998	Sp	CHSE-214	*S. salar*	NA	NA
IPNV107	1999	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV111	1999	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV001	2000	Sp	TO	*S. salar*	NA	NA
IPNV003	2002	Sp	TO	*S. salar*	NA	NA
IPNV009	2003	Sp	CHSE-214	*S. salar*	UK	Scotland, Argyll
IPNV010	2003	Sp	TO	*S. salar*	NA	NA
IPNV013	2004	Sp	TO	*S. salar*	UK	Scotland, Argyll
IPNV014	2004	Sp	CHSE-214	NA	NA	NA
IPNV018	2004	Sp	TO	NA	NA	NA
IPNV024	2005	Sp	TO	*S. salar*	NA	NA
IPNV030	2006	Sp	TO	*S. salar*	NA	NA
IPNV031	2006	Sp	TO	*S. trutta*	UK	Scotland, Shetland
IPNV035	2007	Sp	CHSE-214	*S. trutta*	UK	Scotland, Shetland
IPNV039	2008	Sp	CHSE-214	*S. salar*	NA	NA
IPNV051	2010	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV052	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV053	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Strathclyde
IPNV054	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV055	2012	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV056	2012	NA	CHSE-214	*S. salar*	UK	Scotland, Orkney
IPNV057	2013	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV047	2014	Sp	CHSE-214	NA	NA	NA
IPNV048	2014	Sp	CHSE-214	NA	NA	NA
IPNV049	2014	Sp	CHSE-214	NA	NA	NA
IPNV050	2014	Sp	CHSE-214	NA	NA	NA
IPNV058	2014	NA	CHSE-214	*S. salar*	UK	Scotland, Outer Hebrides

Table S5 Reference strains used for phylogenetic analysis of segment A; NA means not assigned.

Strain	GenBank	Geographical Origin	Host	Reference
Ca3	AF342734	Canada	*Salvelinus alpinus*	[18]

Connecticut1	JF440810	USA	*Salmo spec.*	[94]
Ca2	AF342733	Canada	*Oncorhynchus* spec	[18]
TABV2013	KP268663.1	Australia	*Oncorhynchus mykiss*	[95]
TABV1998	NC_028252.1	Australia	*Gadus* spec	[95]
1146	AJ489222.1	Spain	Trout (not further specified)	[96]
Sp	KF279643	Iran	Trout (not further specified)	[97]
Gre	NA	Turkey	*Oncorhynchus mykiss*	unpublished
SDF4	NA	Turkey	*Oncorhynchus mykiss*	unpublished
1054	NA	Turkey	*Oncorhynchus mykiss*	unpublished
Sp	AF342728	Denmark	Trout (not further specified)	[18]
Ireland	HQ457197.1	Ireland	*Salmon salar*	[36]
DPL	AY026485.1	Thailand	Snakehead fish (*Channidae*)	[18]
Fr21	AY026483.1	France	Trout (not further specified)	[18]
OV2	AY026484.1	England, UK	Oyster (*Ostreidae*)	[18]
N1	D00701	Norway	*Salmon salar*	[98]
Chile	HQ457169	Chile	Salmon (not further specified)	[36]
Norway	HQ457199	Norway	*Salmon salar*	[36]
Sp vir.	AY374435.1	Norway	*Salmon salar* / *Oncorhynchus mykiss*	[19]
Ab	AF342729	Denmark	Trout (not further specified)	[18]
2284	AJ489223	Spain	Trout (not further specified)	[96]
Ab	L40580.1	Denmark	NA	[99]
CVHB1	AY026489.1	Taiwan	Clam (not further specified)	[18]
S	AY026487	Taiwan	Eel (*Anguilliformes*)	[18]
V	AY026486.1	Japan	Eel (not further specified)	[18]
PV	AY026488.1	Taiwan	Perch (*Perca*)	[18]
Ca1	AF342732	Canada	Trout (not further specified)	[18]
ASV	AY026490.1	Canada	Salmon (not further specified)	[18]
Te2	AF342731	England, UK	Tellina (*Tellina*)	[18]
AY98	AY283785.1	Japan	Ayu (*Plecoglossus*)	[24]
YTAV_Y6	AY283781	Japan	Yellowtail (*Seriola*)	[24]
GC1	AY064396.1	Korea	Rockfish (not further specified)	unpublished
Ja	AF342735.1	Canada	Trout (not further specified)	[18]
Reno	AY026345	USA	Trout (not further specified)	[18]
Ja2310	AJ489225	Spain	Salmon (not further specified)	[96]
HL1	D26526.1	Korea	NA	[100]
Ja	NC_001915.1	Canada	Trout (not further	[13]

			specified)	
VR299	AF343572	USA	Trout (not further specified)	[18]
93	AY026346.1	USA	Trout (not further specified)	[18]
11	AY026347.1	USA	Trout (not further specified)	[18]
Mexico	JX174178	Mexico	Trout (not further specified)	unpublished
114	AY026348	USA	Trout (not further specified)	[18]
91	AF343570	USA	Trout (not further specified)	[18]
Buhl	AF343573	USA	Trout (not further specified)	[18]
DryMills	AF343571.1	USA	Trout (not further specified)	[18]
WB	AF342727	USA	Trout (not further specified)	[18]
VTAB	NC_030242.1	Australia	*Oncorhynchus mykiss*	[95]
He	AF342730	Germany	*Esox lucius*	[18]

Table S6 Reference strains used for phylogenetic analysis of segment B; NA means not assigned.

Strain	GenBank	Geographical Origin	Host	Reference
Mexico	EU665685.1	Mexico	Trout (*Oncorhynchus*)	unpublished
HL1	D26527.1	Korea	NA	[101]
2310	AJ489241	Spain	Salmon (*Salmo*)	[96]
Ja Duncan	NC_001916	Canada	Trout	[13]
WB	AF078669.1	USA	NA	[102]
YTAV-06	AB281674	Japan	Yellowtail (*Seriola*)	[103]
POBV	EU161286.1	China	Japanese Flounder (*Paralychthys olivaceus*)	[104]
YTAV-NC1	AY129666	Korea	Japanese Flounder (not further specified)	[105]
Y-6	AY129662	Japan	Yellowtail (not further specified)	[105]
AY-98	AY123970	Japan	Ayu (*Plecoglossus*)	unpublished
VTAB	NC_030244.1	Australia	Rainbow Trout	[95]
He	JF734351.1	Spain	Pike (*Esox Lucius*)	[106]
Ca2	JF734354	Canada	Trout (not further specified)	[106]
Ca3	JF734355	Canada	Trout (not further specified)	[106]
Ca1	JF734353	Canada	Trout (not further specified)	[106]
Ab1	AM114033.1	NA	NA	[96]
Ab2	JF734350.1	NA	Trout (not further specified)	[106]
2284	AJ489239.1	Spain	Rainbow Trout (*Oncorhynchus mykiss*)	[96]
Sp1	AJ622823	France	Trout	[107]
IRIPNV	KC900161.1	Iran	Rainbow Trout	[97]
Gre	NA	Turkey	Rainbow Trout	unpublished
SDF4	NA	Turkey	Rainbow Trout	unpublished
1054	NA	Turkey	Rainbow Trout	unpublished
Sp3	AM889221	Denmark	Trout (not further specified)	[108]
Sp2	M58757.1	NA	NA	[109]
Sp122	AY354524.1	Norway	NA	[110]
Chile	KU609618	Chile	Salmon (not further specified)	[111]
Sp116	AY354523	Norway	NA	[110]
Sp103	AY354522.1	Norway	NA	[110]
TABV2013	KP268678.1	Australia	Rainbow Trout	[95]
TABV1998	NC_028253.1	Australia	Cod (*Gadus*)	[95]
Te	AJ920336	UK	Tellina (*Tellina tenuis*)	[112]

Supplement

Table S1 Significant differences in CAI values between hosts against a sample variant persistent (P) type and low virulent (LV) type; significance level $p \leq 0.05$.

Protein	Host	Comparison	P=value
RdRp	*O. mykiss*	P - LV	0.00037
	S. salar	P - LV	0.00043
VP2	*O. mykiss*	P - LV	0.014
	S. salar	P - LV	0.01627
VP3	*O. mykiss*	P - LV	0.048

Table S2 Significant differences in CAI values persistent (P) and low virulent (LV) type of samples against host species; significance level $p \leq 0.05$.

Protein	Virulence type	Comparison	P=value
VP2	P	*O. mykiss – S. salar*	0.01589
	LV	*O. mykiss – S. salar*	$8.17\ 10^{-6}$
VP5	P	*O. mykiss – S. salar*	$7.26\ 10^{-6}$
	LV	*O. mykiss – S. salar*	0.00088

Table S3 Results of site-specific positive and diversifying selection analysis of all IPNV proteins for all 57 sequenced samples using MEME on HYPHY, applying the MG94xREV codon model. P describes the P=value (p≤0.1) for positive selection at one site.

RdRp/VP1		**VP2**		**VP3**		**VP4**		**VP5**	
Codon	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**	**Codon**	**P=value**
50	0.01	195	0.02	66	0.10	37	0.09	2	0.01
81	0.00	208	0.02	145	0.10	46	0.01	13	0.19
82	0.00	217	0.00			131	0.04	17	0.03
140	0.04	241	0.00			212	0.09	36	0.09
226	0.05	242	0.03			218	0.05	53	0.07
232	0.04	248	0.00					68	0.09
316	0.00	315	0.09					98	0.08
358	0.07	333	0.07					106	0.04
368	0.05							111	0.08
467	0.01								
687	0.07								
699	0.07								
737	0.09								

Table S4 Sequenced IPNV samples. NA=not assigned, Atlantic salmon (*Salmo salar)*, rainbow trout (*Oncorhynchus mykiss), brown trout (Salmo trutta), turbot (Scophthalmus maximus).*

Sequence ID	Collection Year	Confirmed Serotype	Cell line	Host	Origin	Region
IPNV006	1982	NA	CHSE-214	*S. salar*	UK	Scotland, Argyll
IPNV060	1983	NA	CHSE-214	NA	NA	NA
IPNV007	1984	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV008	1984	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV011	1986	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV061	1986	Sp	CHSE-214	*S. maximus*	NA	NA
IPNV019	1987	NA	CHSE-214	*O. mykiss*	UK	Scotland, Highlands
IPNV021	1987	NA	CHSE-214	*O. mykiss*	UK	Scotland, Strathclyde
IPNV062	1987	NA	CHSE-214	*S. salar*	NA	NA
IPNV063	1987	Ab	CHSE-214	*S. salar*	NA	NA
IPNV064	1987	Ab	CHSE-214	NA	UK	Scotland, Highlands
IPNV065	1987	Ab	CHSE-214	NA	UK	Scotland, Highlands
IPNV066	1988	Ab	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV027	1989	NA	CHSE-214	*O. mykiss*	UK	Scotland, Borders
IPNV032	1989	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV068	1989	Ab	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV026	1990	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV069	1992	Ab	CHSE-214	*S. salar*	NA	NA
IPNV070	1992	Sp	CHSE-214	NA	Norway	Norway
IPNV071	1992	Sp	TO	NA	UK	Scotland, Argyll
IPNV072	1992	Sp	CHSE-214	NA	UK	Scotland, Shetland
IPNV073	1993	Sp	CHSE-214	*S. salar*	NA	NA
IPNV075	1993	Sp	TO	*S. salar*	Norway	Norway
IPNV076	1994	Sp	TO	*S. salar*	NA	NA
IPNV081	1994	NA	CHSE-214	*S. salar*	NA	NA
IPNV082	1996	Sp	CHSE-214	*S. salar*	NA	NA
IPNV083	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV084	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV087	1997	Sp	CHSE-214	*S. salar*	NA	NA
IPNV088	1997	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV090	1998	Sp	CHSE-214	*S. salar*	NA	NA
IPNV107	1999	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV111	1999	Sp	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV001	2000	Sp	TO	*S. salar*	NA	NA
IPNV003	2002	Sp	TO	*S. salar*	NA	NA
IPNV009	2003	Sp	CHSE-214	*S. salar*	UK	Scotland, Argyll
IPNV010	2003	Sp	TO	*S. salar*	NA	NA

IPNV013	2004	Sp	TO	*S. salar*	UK	Scotland, Argyll
IPNV014	2004	Sp	CHSE-214	NA	NA	NA
IPNV018	2004	Sp	TO	NA	NA	NA
IPNV024	2005	Sp	TO	*S. salar*	NA	NA
IPNV030	2006	Sp	TO	*S. salar*	NA	NA
IPNV031	2006	Sp	TO	*S. trutta*	UK	Scotland, Shetland
IPNV035	2007	Sp	CHSE-214	*S. trutta*	UK	Scotland, Shetland
IPNV039	2008	Sp	CHSE-214	*S. salar*	NA	NA
IPNV051	2010	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV052	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV053	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Strathclyde
IPNV054	2011	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV055	2012	NA	CHSE-214	*S. salar*	UK	Scotland, Highlands
IPNV056	2012	NA	CHSE-214	*S. salar*	UK	Scotland, Orkney
IPNV057	2013	NA	CHSE-214	*S. salar*	UK	Scotland, Shetland
IPNV047	2014	Sp	CHSE-214	NA	NA	NA
IPNV048	2014	Sp	CHSE-214	NA	NA	NA
IPNV049	2014	Sp	CHSE-214	NA	NA	NA
IPNV050	2014	Sp	CHSE-214	NA	NA	NA
IPNV058	2014	NA	CHSE-214	*S. salar*	UK	Scotland, Outer Hebrides

Table S5 Reference strains used for phylogenetic analysis of segment A; NA means not assigned.

Strain	GenBank	Geographical Origin	Host	Reference
Ca3	AF342734	Canada	*Salvelinus alpinus*	[18]
Connecticut1	JF440810	USA	*Salmo spec.*	[94]
Ca2	AF342733	Canada	*Oncorhynchus* spec	[18]
TABV2013	KP268663.1	Australia	*Oncorhynchus mykiss*	[95]
TABV1998	NC_028252.1	Australia	*Gadus* spec	[95]
1146	AJ489222.1	Spain	Trout (not further specified)	[96]
Sp	KF279643	Iran	Trout (not further specified)	[97]
Gre	NA	Turkey	*Oncorhynchus mykiss*	unpublished
SDF4	NA	Turkey	*Oncorhynchus mykiss*	unpublished
1054	NA	Turkey	*Oncorhynchus mykiss*	unpublished
Sp	AF342728	Denmark	Trout (not further specified)	[18]
Ireland	HQ457197.1	Ireland	*Salmon salar*	[36]
DPL	AY026485.1	Thailand	Snakehead fish (*Channidae*)	[18]
Fr21	AY026483.1	France	Trout (not further specified)	[18]
OV2	AY026484.1	England, UK	Oyster (*Ostreidae*)	[18]
N1	D00701	Norway	*Salmon salar*	[98]
Chile	HQ457169	Chile	Salmon (not further specified)	[36]
Norway	HQ457199	Norway	*Salmon salar*	[36]
Sp vir.	AY374435.1	Norway	*Salmon salar* / *Oncorhynchus mykiss*	[19]
Ab	AF342729	Denmark	Trout (not further specified)	[18]
2284	AJ489223	Spain	Trout (not further specified)	[96]
Ab	L40580.1	Denmark	NA	[99]
CVHB1	AY026489.1	Taiwan	Clam (not further specified)	[18]
S	AY026487	Taiwan	Eel (*Anguilliformes*)	[18]
V	AY026486.1	Japan	Eel (not further specified)	[18]
PV	AY026488.1	Taiwan	Perch (*Perca*)	[18]
Ca1	AF342732	Canada	Trout (not further specified)	[18]
ASV	AY026490.1	Canada	Salmon (not further specified)	[18]
Te2	AF342731	England, UK	Tellina (*Tellina*)	[18]
AY98	AY283785.1	Japan	Ayu (*Plecoglossus*)	[24]
YTAV_Y6	AY283781	Japan	Yellowtail (*Seriola*)	[24]
GC1	AY064396.1	Korea	Rockfish (not further specified)	unpublished

Ja	AF342735.1	Canada	Trout (not further specified)	[18]
Reno	AY026345	USA	Trout (not further specified)	[18]
Ja2310	AJ489225	Spain	Salmon (not further specified)	[96]
HL1	D26526.1	Korea	NA	[100]
Ja	NC_001915.1	Canada	Trout (not further specified)	[13]
VR299	AF343572	USA	Trout (not further specified)	[18]
93	AY026346.1	USA	Trout (not further specified)	[18]
11	AY026347.1	USA	Trout (not further specified)	[18]
Mexico	JX174178	Mexico	Trout (not further specified)	unpublished
114	AY026348	USA	Trout (not further specified)	[18]
91	AF343570	USA	Trout (not further specified)	[18]
Buhl	AF343573	USA	Trout (not further specified)	[18]
DryMills	AF343571.1	USA	Trout (not further specified)	[18]
WB	AF342727	USA	Trout (not further specified)	[18]
VTAB	NC_030242.1	Australia	*Oncorhynchus mykiss*	[95]
He	AF342730	Germany	*Esox lucius*	[18]

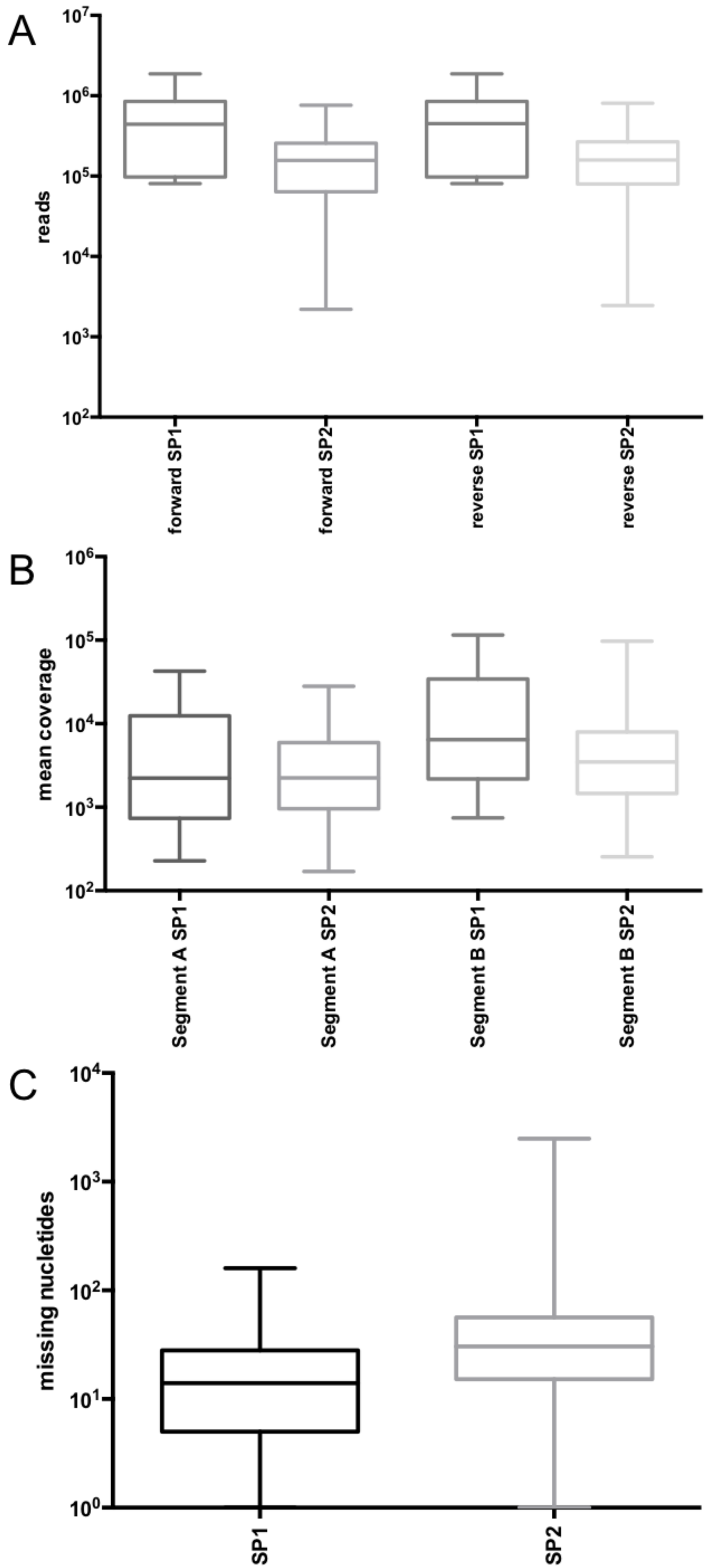
Table S6 Reference strains used for phylogenetic analysis of segment B; NA means not assigned.

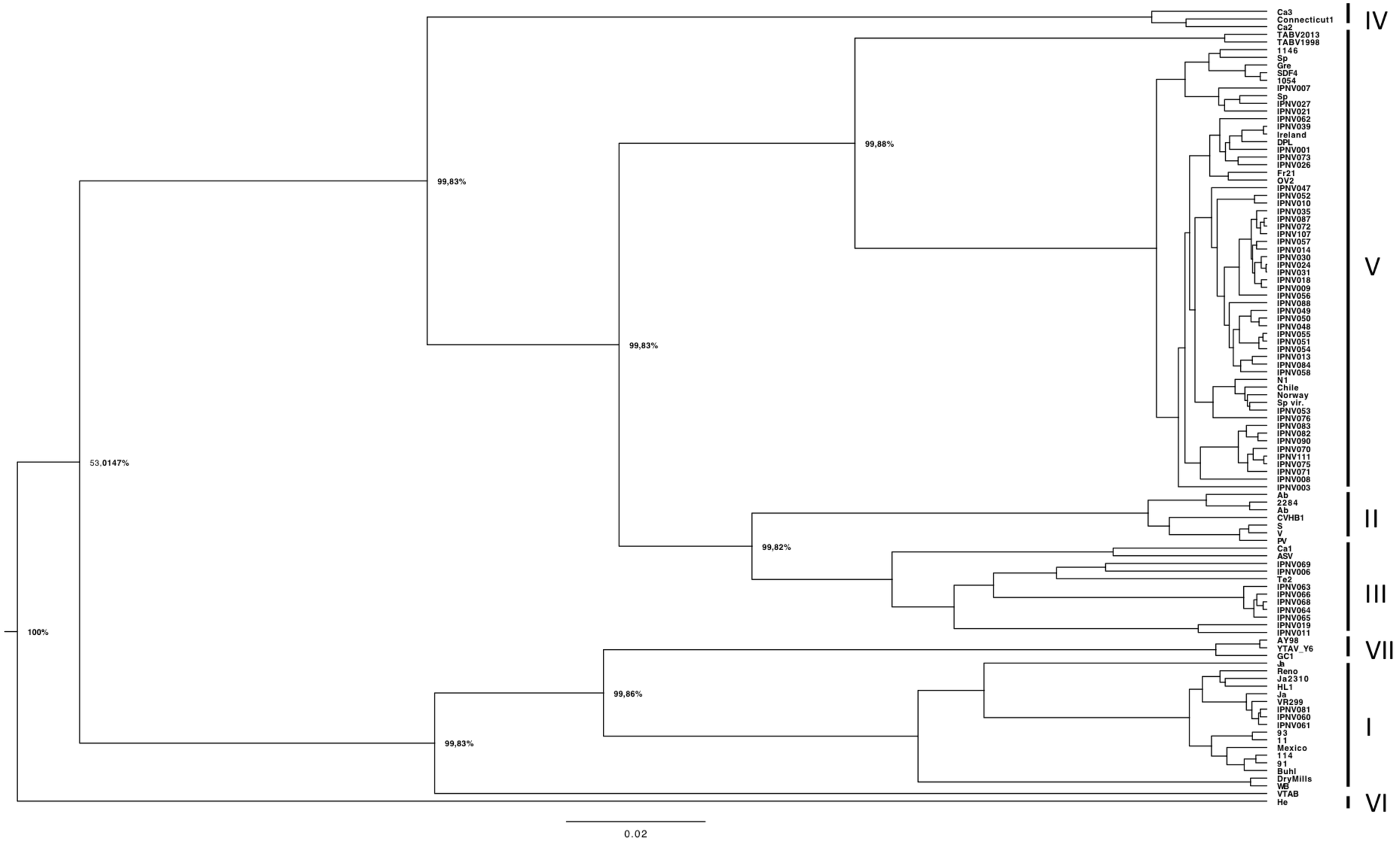
Strain	GenBank	Geographical Origin	Host	Reference
Mexico	EU665685.1	Mexico	Trout (*Oncorhynchus*)	unpublished
HL1	D26527.1	Korea	NA	[101]
2310	AJ489241	Spain	Salmon (*Salmo*)	[96]
Ja Duncan	NC_001916	Canada	Trout	[13]
WB	AF078669.1	USA	NA	[102]
YTAV-06	AB281674	Japan	Yellowtail (*Seriola*)	[103]
POBV	EU161286.1	China	Japanese Flounder (*Paralychthys olivaceus*)	[104]
YTAV-NC1	AY129666	Korea	Japanese Flounder (not further specified)	[105]
Y-6	AY129662	Japan	Yellowtail (not further specified)	[105]
AY-98	AY123970	Japan	Ayu (*Plecoglossus*)	unpublished
VTAB	NC_030244.1	Australia	Rainbow Trout	[95]
He	JF734351.1	Spain	Pike (*Esox Lucius*)	[106]
Ca2	JF734354	Canada	Trout (not further specified)	[106]
Ca3	JF734355	Canada	Trout (not further specified)	[106]
Ca1	JF734353	Canada	Trout (not further specified)	[106]
Ab1	AM114033.1	NA	NA	[96]
Ab2	JF734350.1	NA	Trout (not further specified)	[106]
2284	AJ489239.1	Spain	Rainbow Trout (*Oncorhynchus mykiss*)	[96]
Sp1	AJ622823	France	Trout	[107]
IRIPNV	KC900161.1	Iran	Rainbow Trout	[97]
Gre	NA	Turkey	Rainbow Trout	unpublished
SDF4	NA	Turkey	Rainbow Trout	unpublished
1054	NA	Turkey	Rainbow Trout	unpublished
Sp3	AM889221	Denmark	Trout (not further specified)	[108]
Sp2	M58757.1	NA	NA	[109]
Sp122	AY354524.1	Norway	NA	[110]
Chile	KU609618	Chile	Salmon (not further specified)	[111]
Sp116	AY354523	Norway	NA	[110]
Sp103	AY354522.1	Norway	NA	[110]
TABV2013	KP268678.1	Australia	Rainbow Trout	[95]
TABV1998	NC_028253.1	Australia	Cod (*Gadus*)	[95]
Te	AJ920336	UK	Tellina (*Tellina tenuis*)	[112]

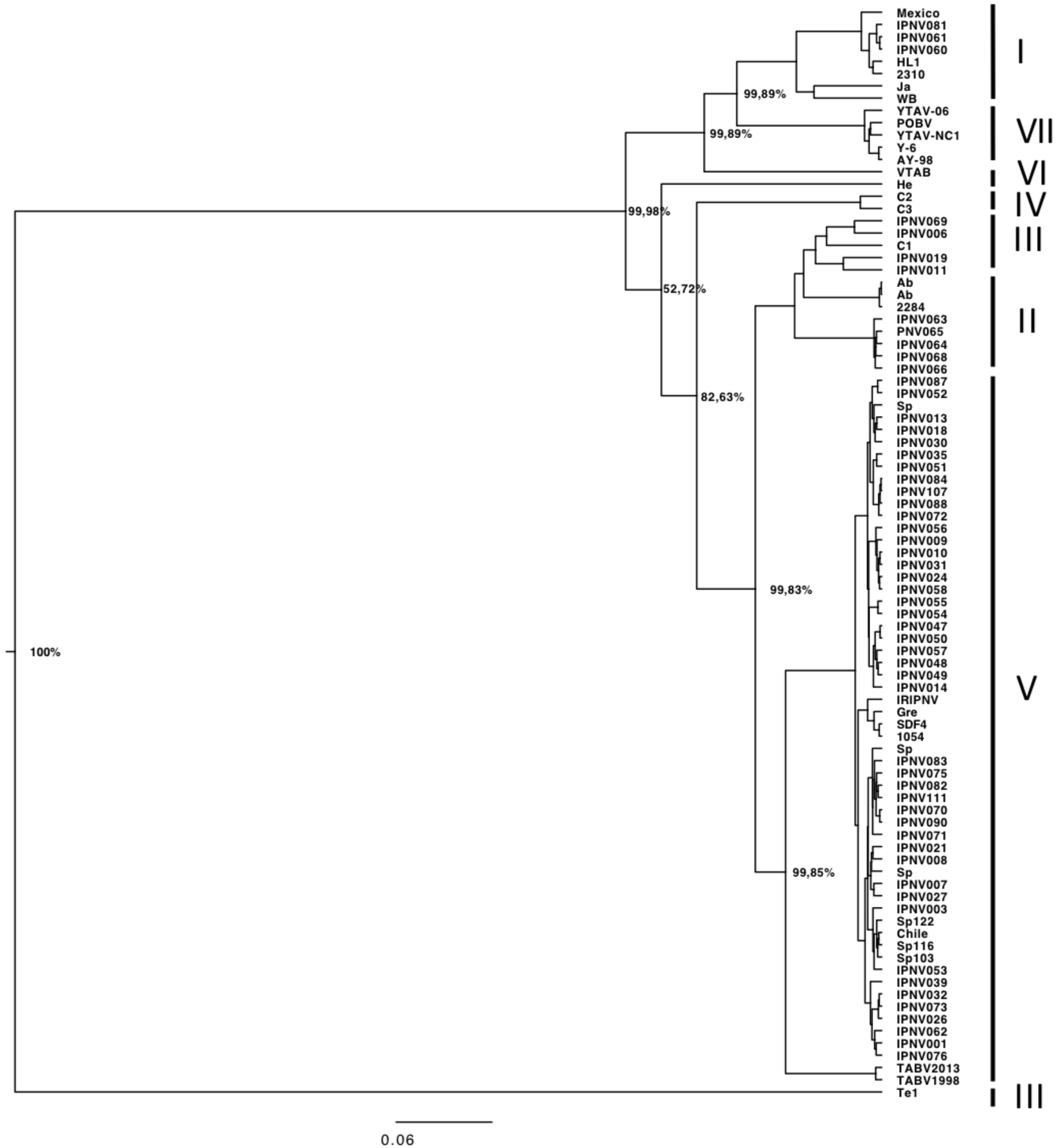
Accepted for publication in Journal of General Virology published by Microbiology Society. The final published version will be available at:
http://jgv.microbiologyresearch.org/content/ journal/jgv/abou**Tables**

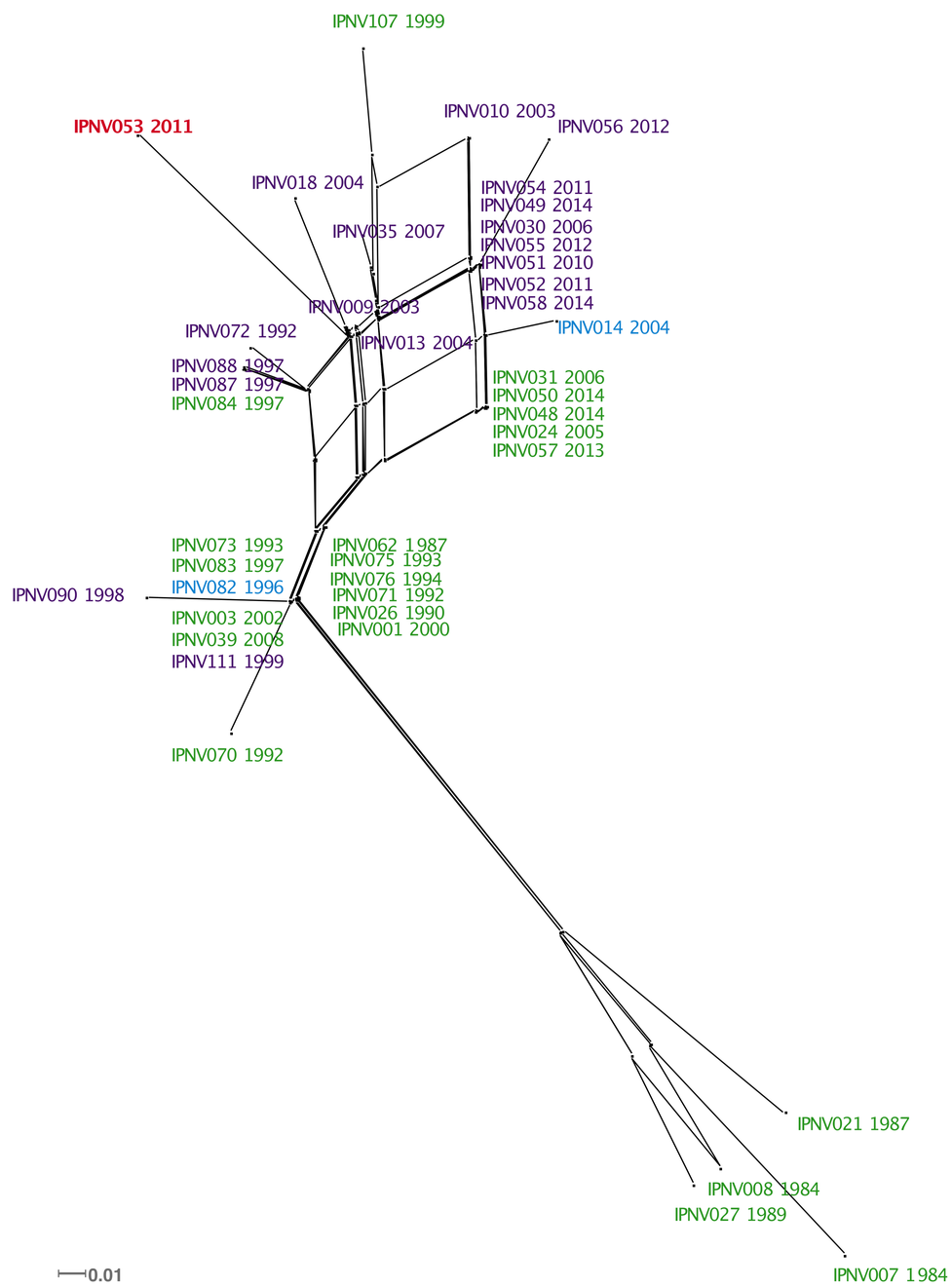
Table 1 VP2 characteristic variable residue pattern of sequenced IPNV isolates compared to the virulent reference strain excluding IPNV032.

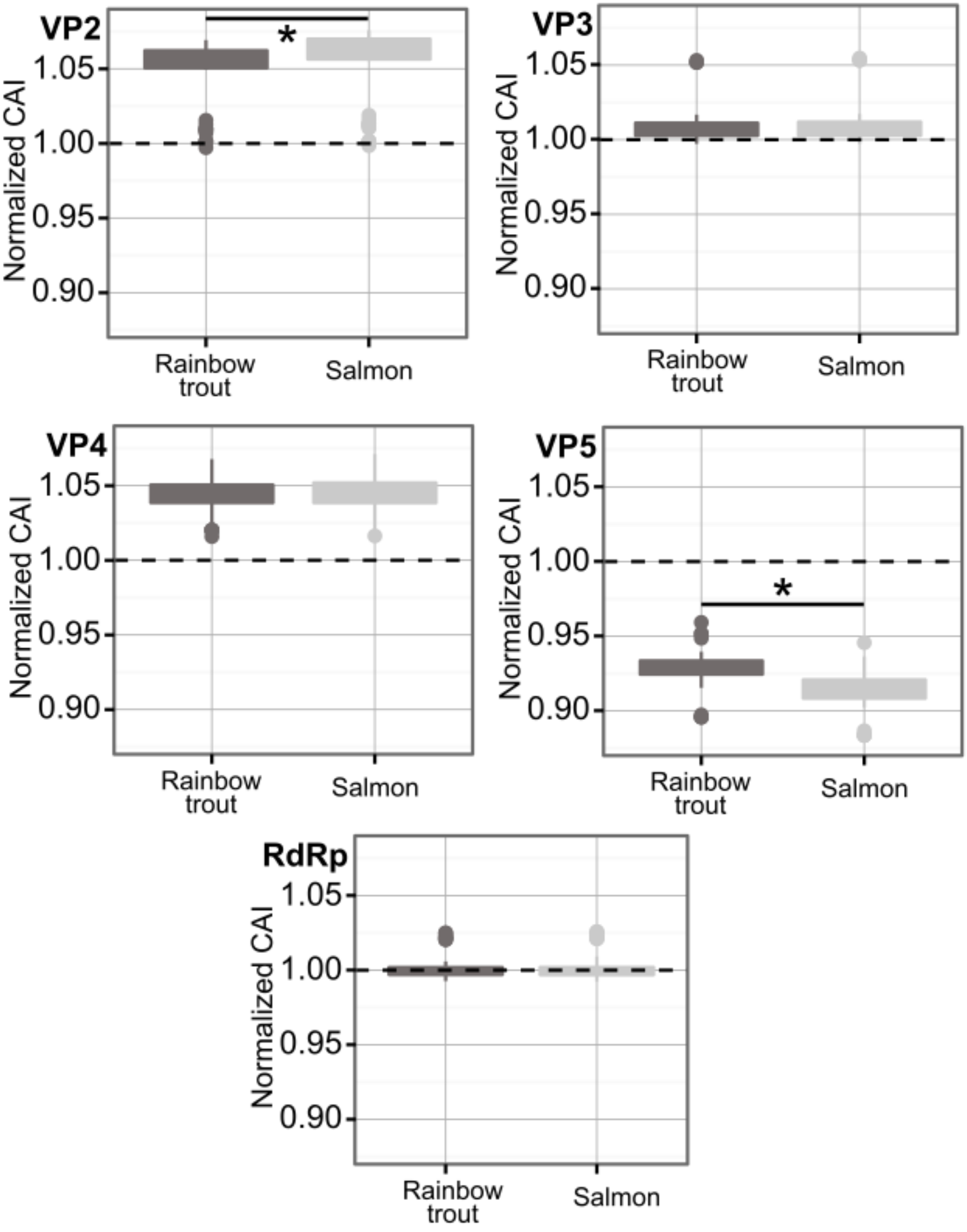
Cluster	Sequence ID	Collection Year	Genogroup	VP2 Amino acid position						
				217	**221**	**247**	**252**	**281**	**282**	**319**
1	IPNVV006, 011, 019, 063, 064, 065, 066, 068, 069	1982, 1986, 1987, 1988, 1989, 1992	III	P	T	P	A	T	A	A
2	IPNV007, 008, 021, 062, 026, 070, 071, 073, 075, 076, 081, 083, 084, 107, 001, 003, 039	1984, 1987, 1990, 1992, 1993, 1994, 1997, 1999, 2000, 2002, 2008	I, V	P	T	A	N	T	N	A
3	IPNV027	1989	V	P	T	A	N	T	N	V
4	IPNV024, 031, 057, 047, 048, 050	2005, 2006, 2013, 2014	V	P	T	A	V	T	N	A
5	IPNV072	1992	V	P	A	A	T	T	N	A
6	IPNV087, 088, 090, 111	1997, 1998, 1999	V	P	A	A	I	T	N	A
7	IPNV009, 010, 013, 018, 030, 035, 051, 052, 054, 055, 056, 049, 058	2003, 2004, 2006, 2007, 2010, 2011, 2012, 2014	V	P	A	A	V	T	N	A
8	IPNV060, 061, 082	1983, 1986, 1996	I, V	A	T	E	N	T	A	A
9	IPNV014	2004	V	P	S	A	V	T	N	A
10	IPNV053	2011	V	T	A	A	V	T	N	A
11	Santi et al., 2004 [1]	2000	V	T	A	T	V	T	N	A

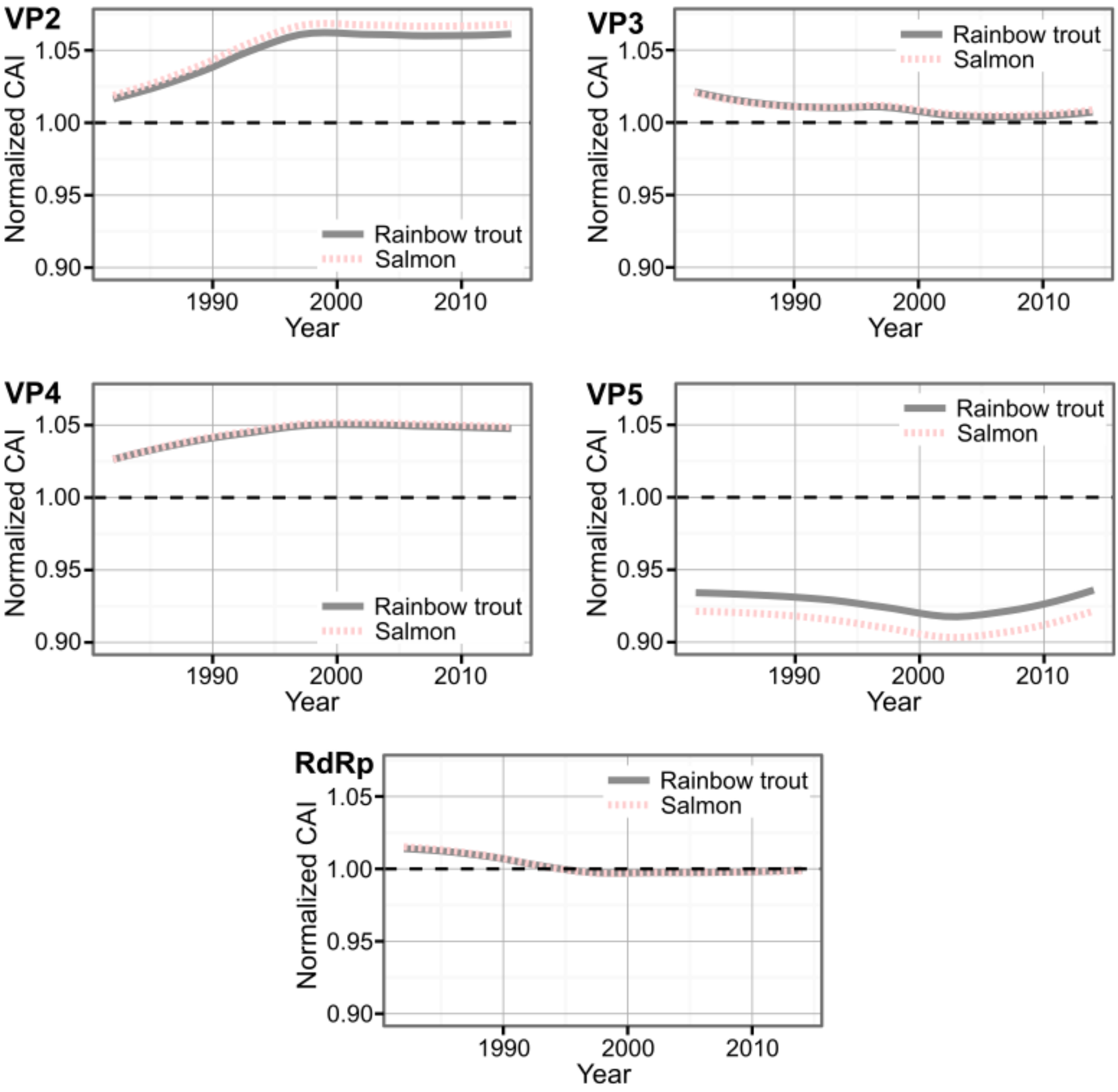

A
reads
10^7
10^6
10^5
10^4
10^3
10^2
forward SP1
forward SP2
reverse SP1
reverse SP2
B
mean coverage
10^6
10^5
10^4
10^3
10^2
Segment A SP1
Segment A SP2
Segment B SP1
Segment B SP2
C
missing nucletides
10^4
10^3
10^2
10^1
10^0
SP1
SP2

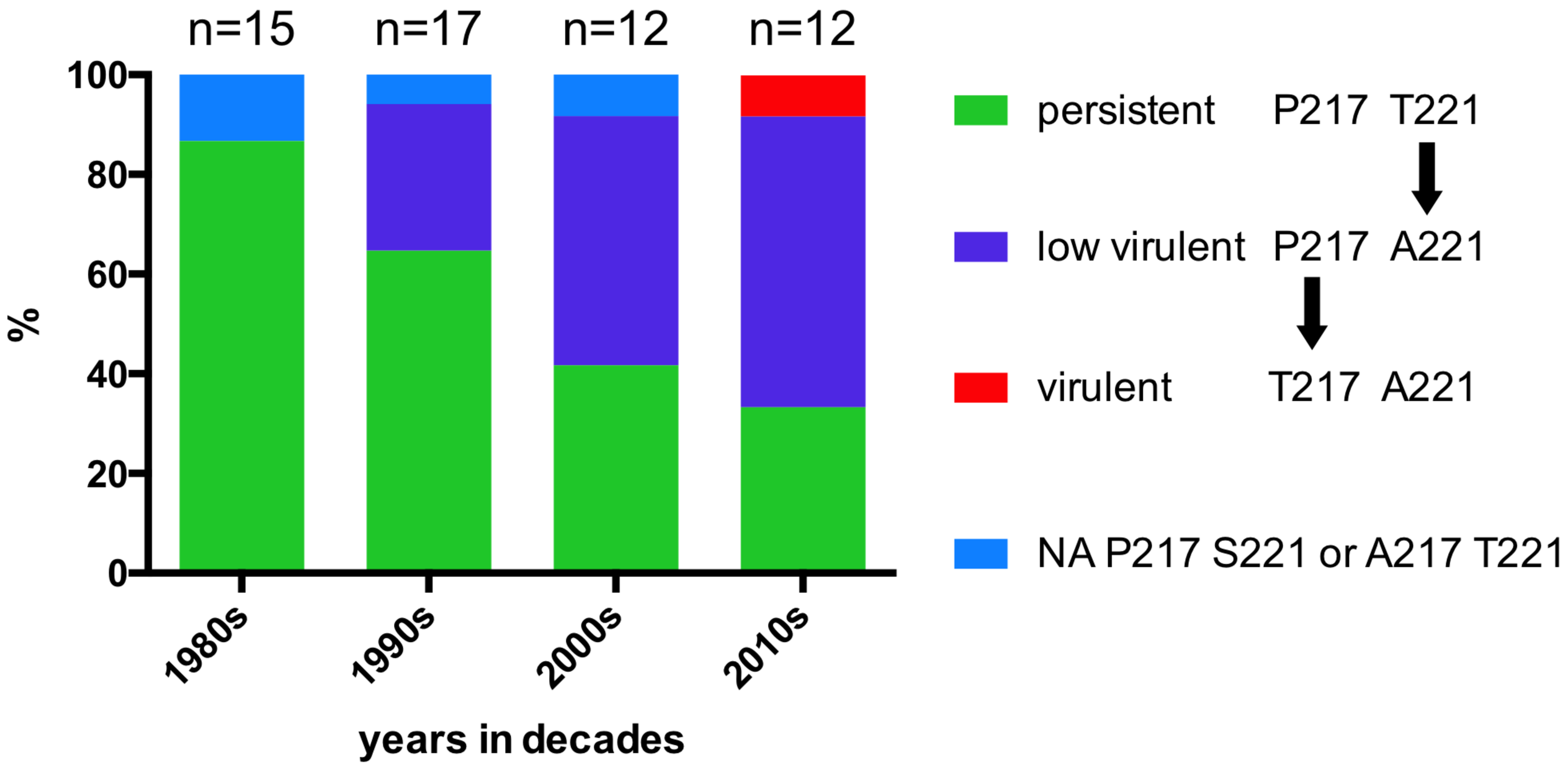

Ca3
Connecticut1
Ca2
TABV2013
TABV1998
1146
Sp
Gre
SDF4
1054
IPNV007
Sp
IPNV027
IPNV021
IPNV062
IPNV039
Ireland
DPL
IPNV001
IPNV073
IPNV026
Fr21
OV2
IPNV047
IPNV052
IPNV010
IPNV035
IPNV087
IPNV072
IPNV107
IPNV057
IPNV014
IPNV030
IPNV024
IPNV031
IPNV018
IPNV009
IPNV056
IPNV088
IPNV049
IPNV050
IPNV048
IPNV055
IPNV051
IPNV054
IPNV013
IPNV084
IPNV058
N1
Chile
Norway
Sp vir.
IPNV053
IPNV076
IPNV083
IPNV082
IPNV090
IPNV070
IPNV111
IPNV075
IPNV071
IPNV008
IPNV003
Ab
2284
Ab
CVHB1
S
V
PV
Ca1
ASV
IPNV069
IPNV006
Te2
IPNV063
IPNV066
IPNV068
IPNV064
IPNV065
IPNV019
IPNV011
AY98
YTAV_Y6
GC1
Ja
Reno
Ja2310
HL1
Ja
VR299
IPNV081
IPNV060
IPNV061
93
11
Mexico
114
91
Buhl
DryMills
WB
VTAB
He
IV
V
II
III
VII
I
VI
99,88%
99,83%
99,83%
99,82%
99,86%
99,83%
53,0147%
100%
0.02


Mexico
IPNV081
IPNV061
IPNV060
HL1
2310
Ja
WB
YTAV-06
POBV
YTAV-NC1
Y-6
AY-98
VTAB
He
C2
C3
IPNV069
IPNV006
C1
IPNV019
IPNV011
Ab
Ab
2284
IPNV063
PNV065
IPNV064
IPNV068
IPNV066
IPNV087
IPNV052
Sp
IPNV013
IPNV018
IPNV030
IPNV035
IPNV051
IPNV084
IPNV107
IPNV088
IPNV072
IPNV056
IPNV009
IPNV010
IPNV031
IPNV024
IPNV058
IPNV055
IPNV054
IPNV047
IPNV050
IPNV057
IPNV048
IPNV049
IPNV014
IRIPNV
Gre
SDF4
1054
Sp
IPNV083
IPNV075
IPNV082
IPNV111
IPNV070
IPNV090
IPNV071
IPNV021
IPNV008
Sp
IPNV007
IPNV027
IPNV003
Sp122
Chile
Sp116
Sp103
IPNV053
IPNV039
IPNV032
IPNV073
IPNV026
IPNV062
IPNV001
IPNV076
TABV2013
TABV1998
Te1
99,89%
99,89%
99,98%
52,72%
82,63%
99,83%
99,85%
100%
I
VII
VI
IV
III
II
V
III
0.06


IPNV107 1999
IPNV053 2011
IPNV010 2003
IPNV056 2012
IPNV018 2004
IPNV054 2011
IPNV049 2014
IPNV035 2007
IPNV030 2006
IPNV055 2012
IPNV051 2010
IPNV052 2011
IPNV058 2014
IPNV009 2003
IPNV014 2004
IPNV072 1992
IPNV013 2004
IPNV088 1997
IPNV087 1997
IPNV084 1997
IPNV031 2006
IPNV050 2014
IPNV048 2014
IPNV024 2005
IPNV057 2013
IPNV073 1993
IPNV083 1997
IPNV082 1996
IPNV062 1987
IPNV075 1993
IPNV076 1994
IPNV071 1992
IPNV026 1990
IPNV001 2000
IPNV090 1998
IPNV003 2002
IPNV039 2008
IPNV111 1999
IPNV070 1992
IPNV021 1987
IPNV008 1984
IPNV027 1989
IPNV007 1984
0.01


VP2
*
Normalized CAI
1.05
1.00
0.95
0.90
Rainbow trout
Salmon
VP3
Normalized CAI
1.05
1.00
0.95
0.90
Rainbow trout
Salmon
VP4
Normalized CAI
1.05
1.00
0.95
0.90
Rainbow trout
Salmon
VP5
*
Normalized CAI
1.05
1.00
0.95
0.90
Rainbow trout
Salmon
RdRp
Normalized CAI
1.05
1.00
0.95
0.90
Rainbow trout
Salmon


VP2
VP3
VP4
VP5
RdRp
Normalized CAI
1.05
1.00
0.95
0.90
1990
2000
2010
Year
Rainbow trout
Salmon


n=15
n=17
n=12
n=12
100
80
60
40
20
0
%
1980s
1990s
2000s
2010s
years in decades
persistent P217 T221
low virulent P217 A221
virulent T217 A221
NA P217 S221 or A217 T221