Genomes reveal genetic diversity of Piscine orthoreovirus in farmed and free-ranging salmonids from Canada and USA

A. Siah,[1],[*],[†] R.B. Breyta,[2],[†] K.I. Warheit,[3] N. Gagne,[4] M.K. Purcell,[5] D. Morrison,[6] J.F.F. Powell,[1] and S.C. Johnson[7]

[1]British Columbia Centre for Aquatic Health Sciences, 871A Island Highway, V9W 2C2, Campbell River, BC, Canada, [2]School of Aquatic Fisheries Sciences, University of Washington, Western Fisheries Research Center, USGS, 6505 NE 65th Street Seattle, WA 98115-5016, USA, [3]Washington Department of Fish and Wildlife PO Box 43200, Olympia, WA 98504-3200, USA, [4]Gulf Fisheries Center, Fisheries & Oceans, 343 Université Ave, Moncton, NB E1C 5K4, Canada, [5]Western Fisheries Research Center, U.S. Geological Survey, 56505 NE 65th Street Seattle, WA 98115-5016, USA, [6]Mowi Canada West, Campbell River, BC, Canada and [7]Fisheries & Oceans Canada, Nanaimo, British Columbia, Canada

*Corresponding author: E-mail: ahmed.siah@cahs-bc.ca

[†]These authors contributed equally to this work.

Abstract

Piscine orthoreovirus (PRV-1) is a segmented RNA virus, which is commonly found in salmonids in the Atlantic and Pacific Oceans. PRV-1 causes the heart and skeletal muscle inflammation disease in Atlantic salmon and is associated with several other disease conditions. Previous phylogenetic studies of genome segment 1 (S1) identified four main genogroups of PRV-1 (S1 genogroups I–IV). The goal of the present study was to use Bayesian phylogenetic inference to expand our understanding of the spatial, temporal, and host patterns of PRV-1 from the waters of the northeast Pacific. To that end, we determined the coding genome sequences of fourteen PRV-1 samples that were selected to improve our knowledge of genetic diversity across a broader temporal, geographic, and host range, including the first reported genome sequences from the northwest Atlantic (Eastern Canada). Nucleotide and amino acid sequences of the concatenated genomes and their individual segments revealed that established sequences from the northeast Pacific were monophyletic in all analyses. Bayesian inference phylogenetic trees of S1 sequences using BEAST and MrBayes also found that sequences from the northeast Pacific grouped separately from sequences from other areas. One PRV-1 sample (WCAN_BC17_AS_2017) from an escaped Atlantic salmon, collected in British Columbia but derived from Icelandic broodstock, grouped with other S1 sequences from Iceland. Our concatenated genome and S1 analysis demonstrated that PRV-1 from the northeast Pacific is genetically distinct but descended from PRV-1 from the North Atlantic. However, the analyses were inconclusive as to the timing and exact source of introduction into the northeast Pacific, either from eastern North America or from European waters of the North Atlantic. There was no evidence that PRV-1 was evolving differently between free-ranging Pacific Salmon and farmed Atlantic Salmon. The northeast Pacific PRV-1 sequences fall within genogroup II based on the classification of Garseth, Ekrem, and Biering (Garseth, A. H., Ekrem, T., and Biering, E. (2013) 'Phylogenetic Evidence of Long Distance Dispersal and Transmission of Piscine Reovirus (PRV) between Farmed and Wild Atlantic Salmon', *PLoS One*, 8: e82202.), which also includes North Atlantic sequences from Eastern Canada, Iceland, and Norway. The additional full-genome sequences herein strengthen

our understanding of phylogeographical patterns related to the northeast Pacific, but a more balanced representation of full PRV-1 genomes from across its range, as well additional sequencing of archived samples, is still needed to better understand global relationships including potential transmission links among regions.

Key words: Piscine orthoreovirus; Atlantic salmon; Pacific salmon; HSMI; phylogenetic.

1. Introduction

Piscine orthoreovirus (PRV; formerly Piscine reovirus) belongs to the family Reoviridae and genus Orthoreovirus (Kibenge et al. 2013; Markussen et al. 2013). The PRV genome was first identified in archived Atlantic Salmon tissues obtained from Norwegian laboratory trials investigating the disease heart and skeletal muscle inflammation (HSMI) by high-throughput sequencing (Palacios et al. 2010). PRV has a non-enveloped, icosahedral virion with a double capsid and, based on sequence analysis, is considered non-fusogenic (Key et al. 2013). The genome consists of ten nucleic acid segments L1, L2, L3, M1, M2, M3, S1, S2, S3, and S4 encoding for λ3, λ2/p11, λ1, μ2, μ1, μNS, σ3/p13, σ2/p8, σNS, and σ1 proteins, respectively (Markussen et al. 2013). PRV differs from the other reoviruses by its 5′ terminal nucleotide sequence and the presence of the proteins σ3/p13 and σ2/p8 encoded by bicistronic genes on the S1 and S2 segments, respectively (Palacios et al. 2010; Markussen et al. 2013; Nibert and Duncan 2013).

There are currently three recognized PRV subtypes, PRV-1 (Palacios et al. 2010), PRV-2 (Takano et al. 2016), and PRV-3 (formerly called PRV-Om) (Dhamotharan et al. 2018). PRV-1 is the subtype that causes HSMI in Atlantic Salmon in Norway (Wessel et al. 2017) while PRV-2 and PRV-3 were discovered in association with other disease syndromes of Coho Salmon (*Oncorhynchus kisutch*) and Rainbow Trout (*O. mykiss*), respectively (Godoy et al. 2016; Takano et al. 2016; Dhamotharan et al. 2018). Molecular surveillance indicates that PRV-1 commonly infects wild and farmed salmonids belonging to the genera *Salmo* and *Oncorhynchus* in the UK, Ireland, Sweden, Faroe Islands, Iceland, Denmark, France, Germany, Norway, Chile, East and West coasts of the USA, and Canada and some non-salmonid species in Norway and West coast of Canada (Kibenge et al. 2013; Miller et al. 2014; Rodger, McCleary, and Ruane 2014; Marty et al. 2015; Siah et al. 2015; Godoy et al. 2016; Gunnarsdóttir 2017; Morton et al. 2017; Cartagena et al. 2018; Hrushowy 2018; Labrut et al. 2018; Purcell et al. 2018; Adamek et al. 2019; Laurin et al. 2019; Vendramin et al. 2019).

In the northeast Pacific Ocean, PRV-1 RNA was first detected by reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) in fish health audit samples of farmed Atlantic Salmon by the Animal Health Centre, Province of British Columbia (BC). PRV-1 RNA was reported to be commonly found in audit samples in BC as early as 1987 (Marty et al. 2015). PRV-1 has also been identified in Washington State (WA) and Alaska State (AK) (Purcell et al. 2018). This distribution will hereafter be referred to as northeast Pacific. Surveillance programs have reported that the northeast Pacific PRV-1 has a low to moderate prevalence in free-ranging wild Pacific Salmon species including Coho Salmon, Chinook Salmon (*Oncorhynchus tshawytscha*), Pink Salmon (*O. gorbuscha*), Chum Salmon (*O. keta*), Sockeye Salmon (*O. nerka*), Cutthroat Trout (*O. clarkii*), Steelhead/Rainbow Trout (*O. mykiss*), as well as a moderate to high prevalence in farmed Atlantic (*Salmo salar*) and Chinook Salmon (Kibenge et al. 2013; Miller et al. 2014, 2017; Marty et al. 2015; Siah et al. 2015; Bass et al. 2017; Di Cicco et al. 2017, 2018;

Morton et al. 2017; Teffer et al. 2017, 2018; Purcell et al. 2018; Polisnki and Garver 2020).

Although phylogenetic analyses of PRV have used different genome segments and concatenated-segment genomes to infer relationships within and between PRV-1 subtypes, the S1 segment has been sequenced most frequently (Garseth, Ekrem, and Biering 2013; Kibenge et al. 2013; Siah et al. 2015; Godoy et al. 2016; Di Cicco et al. 2017, 2018; Gunnarsdóttir 2017; Cartagena et al. 2018; Dhamotharan et al. 2019). Based on analysis of seventeen S1 segment sequences, Kibenge et al. (2013) proposed a single genogroup of PRV-1 in Norway that was made up of two sub-genogroups, 1a and 1b. Garseth, Ekrem, and Biering (2013) assigned sixty-seven segment S1 PRV-1 sequences from Norway into four genogroups I, II, III, and IV while Siah et al. (2015) identified the same four phylogenetic genogroups after adding twenty additional unique S1 sequences from the northeast Pacific. In both these later studies with larger numbers of taxa, genogroup I contained the sequences assigned to sub-genogroup 1b by Kibenge et al. (2013), while the 1a sequences were found in genogroup II. In addition, both studies reported that northeast Pacific sequences grouped together within genogroup II only, while sequences originating in Chile were reported in both genogroups I and II. The Norwegian sequences were found in all four genogroups.

Di Cicco et al. (2018) examined the relationship between S1 sequences of PRV-1 and reported that PRV-1 samples collected in BC formed a single phylogenetic cluster with sequences from Norway and Chile. These authors suggested that the grouping of PRV-1 into sub-genogroups 1a and 1b did not adequately describe the genetic patterns in their larger dataset. Kibenge et al. (2019) reported separation of sequences into the two genogroups 1a and 1b, with sequences from the northeast Pacific grouping with sequences from Iceland, Norway, and Chile within 1a.

More recently, Dhamotharan et al. (2019) analyzed thirty-one PRV-1 genomes and in all trees, the twenty northeast Pacific PRV-1 sequences grouped monophyletically apart from PRV-1 sequences from the north Atlantic and Chilean waters. Dhamotharan et al. (2019) also analyzed 240 partial S1 sequences and report that they separate into two distinct monophyletic clusters, one containing HSMI-associated strains (hypothesized high virulence) and the other cluster containing hypothesized low virulence strains. Based on this clustering, the authors identified distinct amino acid signatures in the S1 and M2 segments hypothesized to be associated with low and high virulence. All PRV-1 sequences from northeast Pacific, as well as some sequences from Norway including a pre-HSMI sequence NOR-1988 fell within the hypothesized low virulence genogroup (Dhamotharan et al. 2019).

The aim of this study was to expand the understanding of the genetic diversity and phylogenetic relationships within PRV-1 from the waters of the northeast Pacific. In the present study, we determined the coding genome of fourteen PRV-1 samples that were selected to improve our knowledge of genetic diversity across broader temporal, geographic, and host ranges. These new sequences and publicly available PRV-1 sequences

were subjected to Bayesian phylogenetic analyses to draw inferences regarding origin and divergence timing. Over the longer term, additional PRV-1 genome sequencing from farmed and free-ranging fish from different geographical areas will provide valuable information required to explore potential transmission links between and among regions.

2. Materials and methods

2.1 Tissue samples and RNA extraction

We obtained fourteen PRV-1 partial coding genome sequences from tissues sampled from a variety of free-ranging and farmed salmonids collected at different locations and sources; additional sequences were retrieved from publicly available resources (Table 1). RNA was extracted from ~30 mg of frozen tissues using Qiagen RNeasy Mini kit according to the manufacturer's recommendations (Qiagen, ON, Canada). Total RNA was eluted in 50 µl of RNAse-free water. The resulting RNA samples were stored at −80 °C until further analysis.

2.2 Sequencing

Illumina sequencing was performed at the Genome Quebec Innovation Centre (Quebec, Canada). Total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.) and its integrity was assessed on a 2100 Bioanalyzer (Agilent Technologies). rRNA were depleted from 250 ng of total RNA using Ribo-Zero rRNA Removal kit specific for HMR RNA (Illumina). The remaining RNA was purified using the Agencourt RNACleanTM XP Kit (Beckman Coulter) and eluted in water. cDNA synthesis was achieved with the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs). The remaining steps of library preparation were done using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs). Adapters (read 1 sequence: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC and read 2 sequence: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) and PCR primers (C9 pair: TGGAAGCA, GTACTCTC; D9 pair: CAATAGCC, TGCGTAGA; E9 pair: CTCGAACA, GGAATTGC; F9 pair: GGCAAGTT, CTTCTGAG; G9 pair: AGCTACCA, CTTAGGAC; and H9 pair: CAGCATAC, TCTAACGC used for expected i7 and i5 index reads, respectively) were purchased from New England BioLabs. PCR conditions were 98 °C/30 s for denaturation, 98 °C/ 10 s and 65 °C/75 s with twelve cycle amplification and 65 °C/ 5 min for extension.

Libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). Average size fragment was determined using a LabChip GX (PerkinElmer) instrument.

For the HiSeq 2500 platform, the libraries were normalized at 2 nM, denatured in 0.05 N NaOH and then were diluted to 8 pM using HT1 buffer. Clustering was done on an Illumina cBot and the flowcell was run on a HiSeq 2000 for 2× 125 cycles following manufacturer's instructions. The Illumina control software was HCS 2.2.58, and the real-time analysis program was RTA v. 1.18.63. Reads were demultiplexed in Fastq files using bcl2fastq v1.8.4.

For the HiSeq 4000 platform, libraries were quantified and normalized, pooled, denatured in 0.05 N NaOH, and neutralized using HT1 buffer. ExAMP was added to the mix following manufacturer's instructions. The pool was loaded at 200 pM on an

Illumina cBot and the flowcell was run on a HiSeq 4000 for 2× 100 cycles (paired-end mode). A phiX library was mixed with libraries at 1 per cent. The Illumina control software was HCS HD 3.4.0.38, the real-time analysis program was RTA v. 2.7.7. Reads demultiplexing was performed using bcl2fastq2 v2.18.

Sequence assembly of HiSeq data was performed using CLC Genomic Workbench (v6.0). Adaptors and low-quality bases were discarded and sequence reads were aligned to the Canadian PRV-1 reference genome generated from sample VT06062012-358 (Kibenge et al. 2013). Mapping parameters were mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.7, and similarity fraction = 0.85. Consensus sequences were generated using the default parameters apart from the coverage threshold, which was set at 20.

The samples NORmDa314AS2018, NORmDa115AS2018, NORmGl118AS2018, and NORmGl118H8IsAS2018 (see Table 1) were analyzed at PatoGen AS, a commercial laboratory located in Aalesund (Norway). These samples were sequenced using an Ion Torrent S5 XL instrument (ThermoFisher Scientific). Ion Total RNA-Seq kit v2 was used to generate the libraries for further analysis as per manufacturer's recommendation (ThermoFisher Scientific). Raw reads were filtered and trimmed for quality using Trimmomatic (v0.36) (Bolger, Lohse, and Usadel 2014) with the parameters 'LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36'. PRV-1 sequences were identified by mapping filtered reads against the published PRV-1 genome (Palacios et al. 2010) using the mapper BWA (Li and Durbin 2009). Mapped reads were extracted from the resulting bam files and assembled using the same PRV-1 genome as a reference and CLC Main Workbench (v7.0.2) (QIAGEN Bioinformatics, Qiagen, Germany). Alignments were manually inspected, and consensus sequences exported for further analysis.

2.3 Genome concatenation and S1 sequences

Nucleotide sequences for all ten segments of thirty-four PRV-1 sequences were obtained from GenBank (access date August 2019) and added to the fourteen coding genome sequences in the present study (Table 1). Sequences from each segment were trimmed as necessary to obtain the same coding sequence length for all of the isolates, and aligned using ClustalW (Thompson, Higgins, and Gibson 1994). Individual segment sequences and the concatenated genomes from these forty-eight sequences were used in our phylogenetic analysis as described below. We created a segment S1 sequence dataset using sequences from GenBank and (Gunnarsdóttir 2017) as well as from the S1 portion of the fourteen genome sequences described above. A total of 330 sequences were aligned and trimmed to 829 nucleotides and used in our phylogenetic analysis as described below.

2.4 Phylogenetic analysis

The primary approach to infer PRV-1 evolutionary patterns was Bayesian inference using BEAST v1.8 (Drummond et al. 2012). The coalescent model of evolution was used, and under this model both the constant size and skyline demographic model priors were evaluated (Drummond et al. 2005). After performing a substitution model best-fit analysis in DataMonkey (Delport et al. 2010) and MEGA v7 (Kumar, Stecher, and Tamura 2016) based on the Akaike Information Criterion and Bayesian Information Criterion, we tested strict and uncorrelated log normal clock using the TamuraNei, 93 (TN93) + G4 + I substitution

model for concatenated genomes (Tamura et al. 2013) and Hasegawa-Kishino Yano (HKY) + G4 + I substitution model for S1 sequences (Hasegawa et al. 1985). Dated taxa allowed comparison of strict and uncorrelated log normal clock priors as well, after temporal significance was confirmed as described below (Drummond and Rambaut 2007; Drummond et al. 2012). The best demographic and clock model priors were then tested with and without two-part codon partition (Supplementary Table S1). All analyses used 100 million generations Markov Chain Monte Carlo chains resulting in 10,000 parameter estimates. Convergence was assessed using Tracer v1.6 and was based on the effective sampling size (ESS), defined as every parameter having an ESS > 200 (Drummond and Rambaut 2007). To assess the influence of the prior parameters on the analysis, each chosen prior parameter was analyzed with an empty alignment (no sequence input). Appropriate prior parameters were those that produced posterior probabilities that were significantly different and a poorer fit than the parallel analyses containing taxa sequence data (Supplementary Table S1). Overall model fit was assessed by log likelihood and likelihood ratio test.

The maximum clade credibility tree was recovered after removing burn-in using TreeAnnotator v1.8 and then drawn in FigTree v1.4.3. The final parameters used for the concatenated nucleotide genome dataset and the full S1 dataset were dated taxa with a strict clock (Continuous-Time Markov Chains (CTMC) initial 1e-4) and the constant population size demographic prior.

2.5 Dating and discrete state analyses

Since previous studies (Kibenge et al. 2013, 2017) have described dates for key internal nodes in the evolutionary tree of PRV-1, we first determined if dating internal nodes was statistically valid. Testing for temporal structure in the dataset consisted of a date randomizations of taxa dates (twenty replicates) using a python script. All replicates were analyzed using the same parameters as the true-date taxa set and the posterior probability (pp) of each replicate was compared to the analysis of true-date taxa. If all randomized date replicates produced posterior probabilities that were significantly different and a poorer fit than that of the true-date taxa, this would indicate a temporal signal in the dataset (Duchene et al. 2015).

In order to determine if there was a phylogenetic structure sufficient to support discrete state reconstruction for either infected host species or infection location, these traits were mapped onto trees using the basic parameters described above. If either fish host or the country of collection location grouped in a monophyletic manner, this trait was used for discrete state analyses. Discrete state analysis allows the phylogenetic algorithm to use the trait state as a cofactor in inferring the evolutionary history of the dataset and to make statistically rigorous estimates for this trait at all internal nodes. Location was encoded as a discrete state and ancestral node trait-state values were inferred under a symmetric model.

As a complimentary analysis, we used MrBayes v3.2.6 (Ronquist et al. 2005) to provide an alternative Bayesian inference of PRV-1 evolution. While BEAST analyses were conducted using all 330 sample sequences, MrBayes was conducted using the 118 unique haplotype sequences within this dataset. In this analysis, no initial assumptions were made about nucleotide substitutions, sampling across F81, HKY, and GTR substitution models, using Dirichlet priors equal to 1.0 for each of the six possible base substitutions (nst = mixed). We allowed a proportion of the nucleotide sites to be invariable, while the rate of change at the remaining sites followed a gamma distribution, with shape parameter (1.0) exponentially distributed (rates = invgamma). The proportion of the invariable sites followed a uniform distribution with (0.00, 1.00) interval. We assumed no molecular clock, using unconstrained branch lengths and with all tree topologies being equally probable as priors. Convergence was assessed by the average standard deviation falling below 0.01 of the split frequencies of two independent analysis chains. Burn-in was set as the initial 25 per cent of the final number of generations, and we sampled from the posterior distribution every 1,000 generations.

In cases where datasets analyzed under the simplest Bayesian BEAST parameters such as segment tree analysis from concatenated genomic sequences did not reach convergence, phylogenetic analyses were conducted under a maximum-likelihood framework. The nucleotide sequence of each segment for the forty-eight sequences was analyzed in MEGA (Kumar, Stecher, and Tamura 2016). These sequences were translated *in silico* into amino acids sequences and then analyzed using PhyML software under the IHVw + G + I + F model as determined by SMS: Smart Model Selection in PhyML (Guindon et al. 2010; Lefort, Longueville, and Gascuel 2017).

2.6 S1 segment tree nomenclature

Phylogenetic analysis of S1 nucleotide sequences performed by Garseth, Ekrem, and Biering (2013) and Siah et al. (2015) grouped PRV-1 into four genogroups (I–IV) and for the purposes of this paper we have decided to follow this tree topology nomenclature. As necessary we will reference to the 1a/1b nomenclature proposed by Kibenge et al. (2019) for continuity.

2.7 Amino acid alignments of S1 and M2

Amino acid sequences encoded by the S1 (σ3 and p13 proteins) and M2 (μ1 protein) regions were obtained from GenBank or generated for this study using the ExPASy (Expert Protein Analysis System) Translation Tool (https://web.expasy.org/translate/). Amino acid sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized in Jalview (version 2.0; http://www.jalview.org/) (Waterhouse et al. 2009; Sievers et al. 2011).

3. Results

3.1 Genome sequencing

Here, we obtained PRV-1 genome sequences from fourteen samples from the northeast Pacific and the north Atlantic (New Brunswick (NB), Canada and Norway). Also, we report on a previously unpublished genome sequence from an escaped Atlantic Salmon collected in British Columbia (WCan_BC17_AS_2017, accession numbers MT580965–MT580974). These were compared to publicly available PRV-1 sequences (Table 1 and Supplementary Table S2). The coding regions of each of the ten segments were concatenated to the same length (21,432 bp) for all samples (hereafter referred to as 'concatenated genome'). Nucleotide similarity across the forty-eight concatenated genomes (Fig. 1) ranged from 97.5 to 100 per cent with the lowest nucleotide similarity between the northeast Pacific and two Norwegian sequences (NOR_AS_1997 and NOR_mDa314_AS_2018) (Supplementary Table S3). Viruses sampled from northeast Pacific (WCAN and US) had 99.7–100 per cent nucleotide similarity, except for WCan_BC17_AS_2017,
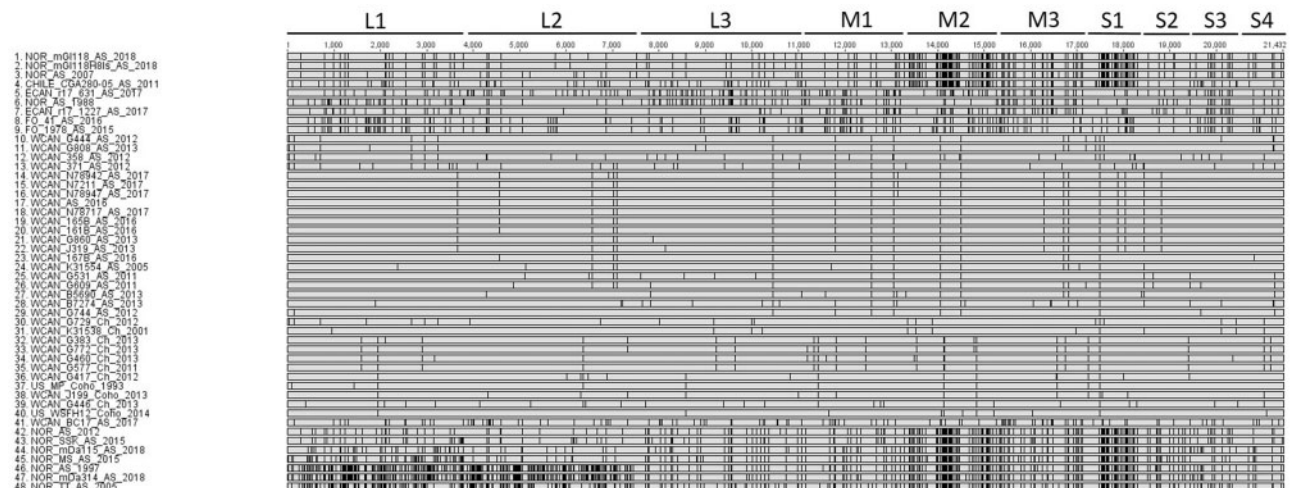
Figure 1. Sequences from each segment were trimmed to coding regions, concatenated, and aligned against published PRV-1 genome sequences. Genome segments are labeled across the top, sample names down the left, and global nucleotide differences are shown as black lines. Sequences obtained in this study are indexed #1-2-5-7-14-15-16-17-18-24-31-37-44-47 and the remaining sequences are publicly available sequences. Location of origin, host species, and sampling year is indicated in the sample name.

which was obtained in 2017 from an escaped Atlantic Salmon that is presumed to have its broodstock origin in Iceland (hereafter referred to as 'escapee_BC17'). The escapee_BC17 genome sequence, along with sequences from Eastern Canada (ECAN), Faroe Islands (FO), and the 1988 sequence from Norway (NOR_AS_1988), had lower shared nucleotide similarity with the northeast Pacific sequences (98.7–99.3%). Except for the 1988 sequence, PRV-1 genome sequences from Norway shared 97.5–99.4 per cent nucleotide similarity with northeast Pacific sequences (Fig. 1 and Supplementary Table S3).

A higher genetic diversity can be visualized for segments M2 and S1 for sequences from Norway/Chile in comparison to the sequences from northeast Pacific, Eastern Canada, Faroe Islands, and Norway 1988 (Fig. 1). In addition, segments L1 and L2 from Norwegian sequences NOR_TT_AS_2005, NOR_mDa314_AS_2018, and NOR_AS_1997 had a greater genetic diversity in comparison to the other sequences including Norway and elsewhere (Fig. 1).

3.2 Phylogenetic analysis of concatenated genome sequences

Bayesian phylogenetic analysis of concatenated genomes using BEAST ran to consensus (Supplementary Fig. S1). Twenty replicate date-randomization tests of the concatenated genomes produced posterior probabilities that were significantly different and a poorer fit than that of the true-date taxa, suggesting that there was significant temporal signal in the data (see Section 2 and Supplementary Table S1). We evaluated the goodness of fit of a strict molecular clock and the relaxed uncorrelated lognormal clock. The comparison showed the strict clock was a better fit to our data (Supplementary Table S4). No monophyletic association with species of infected host was apparent, so only location was used as a discrete trait for further analysis. The topology of the tree generated with dated taxa and discrete state spatial locations (Fig. 2) was identical to the sequence-only tree (Supplementary Fig. S1), indicating that date and location were not artifactually changing tree topology.

All phylogenetic analyses here describe monophyly as internal nodes (nodes with more than two descendants and are not the inferred root) with high statistical support, either pp or

bootstrap values ≥70. All concatenated genome sequences from the northeast Pacific were monophyletic at node B, pp = 0.96 (Fig. 2). The node B genogroup included the escapee_ BC17 sample and node A. Node A was also monophyletic (pp = 1.0) and contained all other genome sequences representing the established northeast Pacific variants and was genetically distinct from escapee_BC17 (Fig. 2). The model predicted a common ancestor node B in Eastern Canada with a median date of year = 1833 (pp = 0.84, 95% HPD 1711–1907) while node A was predicted to have diverged in northeast Pacific with a median date of year = 1950 (pp = 1.0, 95% HPD 1900–1975). Within node A, viruses from Pacific Salmon hosts group separately from those of Atlantic Salmon. However, their ancestral node (node *, Fig. 2) is not monophyletic (pp = 0.57), which is consistent with the failure of host species to improve model fit. Node C contained nodes A and B, as well as an archived sequence from Norway (Nor_As_1988) and two contemporary New Brunswick sequences (Fig. 2). The model predicted the common ancestor of node C was located in Eastern Canada or Norway with a median date of year = 1790 (pp = 0.55 and pp = 0.32 respectively; 95% HPD 1636–1877). The combined analyses of temporal and spatial signal infer that the common ancestor of all forty-eight samples of PRV-1 genomes was located in the north Atlantic waters with a median date of year = 1313 (pp = 0.83, 95% HPD 868–1593) (root, Fig. 2).

3.3 Phylogenetic analysis of S1 segment

Here, we used 330 partial (829 bp) S1 segment sequences to conduct Bayesian inference phylogenetic trees using BEAST and MrBayes (Figs 3 and 4, respectively). The BEAST analysis of these data followed a similar parameter optimization process to that used in our BEAST analysis of the concatenated genomes (used dated taxa, strict molecular clock, constant population size, and HKY + G + I substitution model with CTMC rate prior, 1e-4 initial value). There was no signal for host specificity but there was one for spatial origin, so we modeled location as a discrete trait (Fig. 3). The MrBayes analysis converged on a slightly modified HKY model.

Both models of Bayesian inference S1 phylogenetic trees recovered all four of the genogroups I–IV previously described by
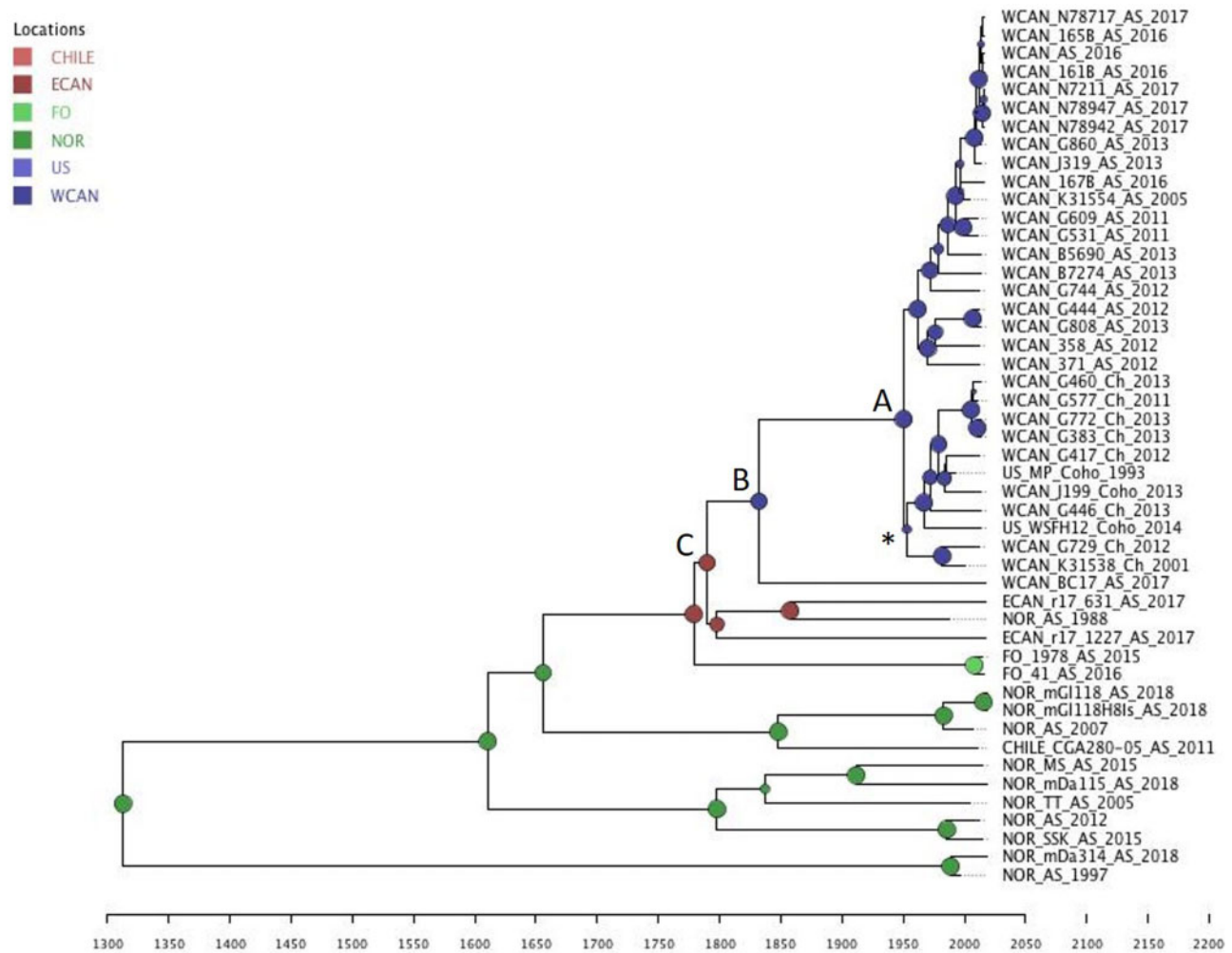
Figure 2. Bayesian coalescent phylogeny of 48 concatenated-segment genome sequences. Circles at nodes are scaled to posterior probability. Each node is an inferred ancestor, so the circles are also color coded for the most probable spatial location of the ancestor. Legend for spatial locations of sample origin: Chile, Chile; ECan, New Brunswick, Canada; FO, Faroe Islands, Denmark; Nor, Norway; US, Washington State, USA; WCan, British Columbia, Canada. See methods for inference parameters. Lettered nodes described in text, and where appropriate correspond to the same nodes in Figs 3 and 4.

Garseth, Ekrem, and Biering (2013). Node L was consistent with genogroup I, node C was consistent with genogroup II, node F was consistent with genogroup III, and node G was consistent with genogroup IV (Figs 3 and 4; see Supplementary Fig. S3 for subtrees with legible taxa names of these key nodes).

Node L (genogroup I) was monophyletic in both the BEAST (pp = 0.85, Fig. 3) and MrBayes (pp = 0.73, Fig. 4) analyses. Node L grouped PRV-1 sequences from Norway and Chile. The ancestor of this genogroup was inferred to have originated in Norway with a model predicted median date of year = 1995 (95% HPD 1988–1997). Node F (genogroup III) was monophyletic in the BEAST analysis, with strong support (pp = 1.0, Fig. 3 and Supplementary Fig. S3), and in the MrBayes analysis, but with weak support (pp = 0.57, Fig. 4). Node F included sequences from Norway and the Faroe Islands with an inferred ancestor of Norway and model predicted median date of year = 2005 (95% HPD 2002–2007). Similarly, node G (genogroup IV) was monophyletic in the BEAST analysis (pp = 0.85, Fig. 3 and Supplementary Fig. S3) but was not monophyletic in the MrBayes analyses (Fig. 4). Node G also grouped sequences from Norway and Faroe Islands, with an inferred ancestor of Norway and model predicted date of year = 2002 (95% HPD 1997–2005).

Node C (genogroup II) was monophyletic in both the BEAST (pp = 0.85, Fig. 3) and MrBayes (pp = 0.94, Fig. 4) S1 tree analyses just as it was in the concatenated genome tree (Fig. 2). The monophyly of node C across all three trees indicated the topology for this genogroup is particularly robust. Node C included all northeast Pacific and all Eastern Canada isolates, and a subset of PRV-1 from Norway and Chile including the archived Norwegian sample Nor_As_1988 (red asterisk in Fig. 2).

Several well-supported nodes were resolved descending from node C that included sequences relevant to North America. Node K included the majority (146/162) of PRV-1 S1 sequences originating from fish sampled in the northeast Pacific (pp = 0.99, Fig. 3 and Supplementary Fig. S3) representing the established PRV-1 sequence types circulating from Alaska to Washington in both Pacific and Atlantic salmon. Node K also included sequences from Chile (Fig. 3, black arrowhead). Node K was predicted to have occurred in the northeast Pacific with a median year = 1990 (95% HPD 1986–1993). Node K was also recovered by the MrBayes analysis (pp = 0.97, Fig. 4).

Node M included sequences from Norway, Canada, Iceland, and WA with a model predicted origin in Norway with a median date of year = 1999 (pp = 0.94, 95% HPD 1991–2004; Fig. 3). Two
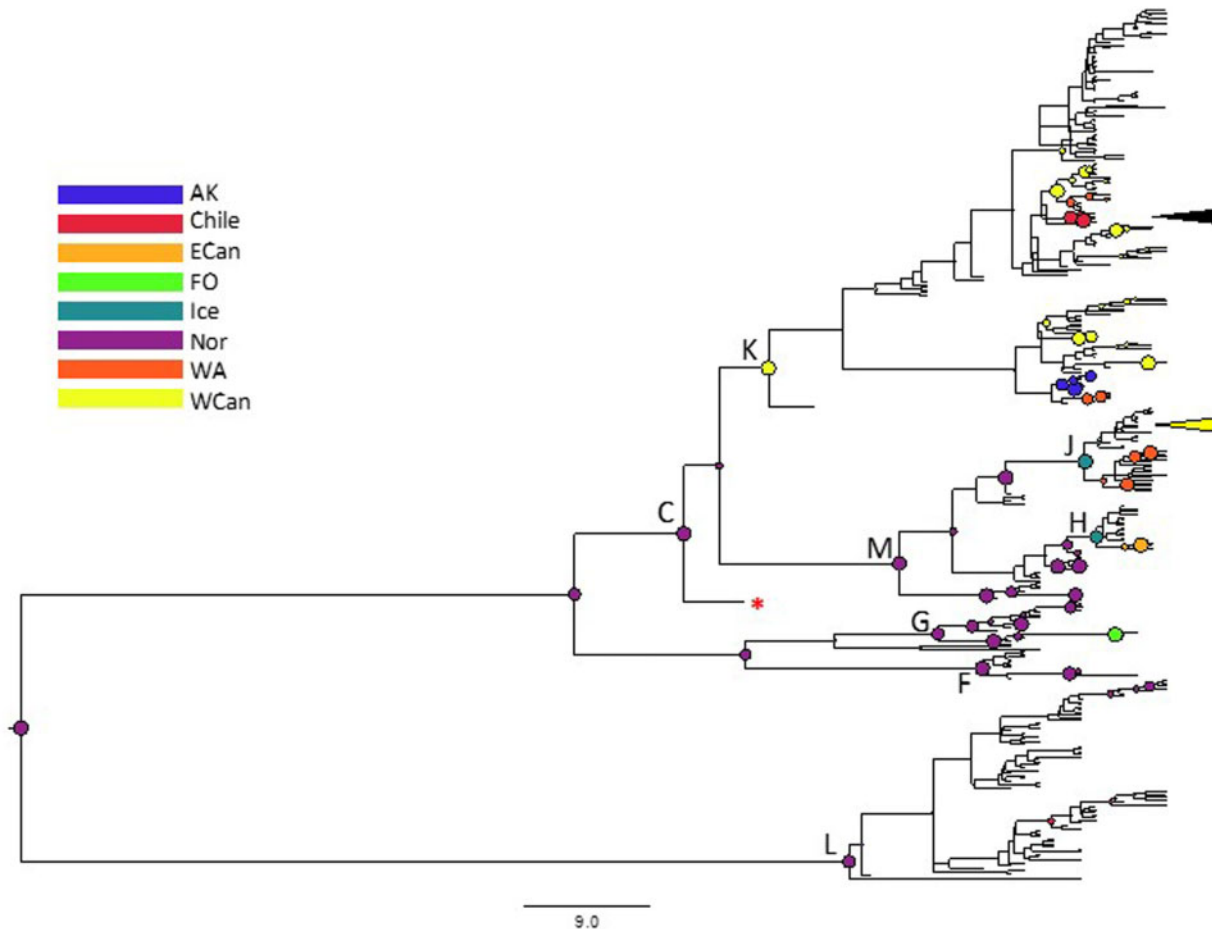
Figure 3. Bayesian coalescent phylogeny of 330 S1 sequences. Circles at nodes are scaled to posterior probability and colored for the most probable spatial location of that ancestor. Legend for spatial locations are as follows: AK, Alaska State, USA; Chile, Chile; ECan, New Brunswick, Canada; FO, Faroe Islands, Denmark; Ice, Iceland; Nor, Norway; WA, Washington State, USA; WCan, British Columbia, Canada. See methods for inference parameters. Lettered nodes described in text, and where appropriate correspond to the same nodes in Figs 2 and 4. Red asterisk indicates sample Nor_AS_1988, black arrowhead indicates transfer from northeast Pacific to Chile, and yellow arrowhead indicates samples identical to BC17.

well-supported nodes were resolved from within node M (nodes J and H; Supplementary Fig. S3). Node J included PRV-1 variants from Iceland and sixteen escaped Atlantic Salmon caught in WA and BC in 2017 and 2018 (Kibenge et al. 2019) and escapee_BC17. The escapees were derived from 2016 egg imports into WA; BEAST predicted the median year for node J = 2012 (pp = 0.76; 95% HPD 2010–2014). Node H grouped PRV-1 sequences from Iceland and eastern Canada (pp = 0.94; estimated median year = 2013; 95% HPD 2012–2014). Node H also included a sequence of unclear provenance (MF946300 labeled only as Atlantic salmon, Canada). Node J was also recovered by the MrBayes analysis (pp = 0.97, Fig. 4); but node H was not recovered.

The major difference in topology between the BEAST and MrBayes S1 tree is node Y (pp = 0.97, Fig. 4), which is ancestral to node K (pp = 0.97) and does not appear in the BEAST tree (Fig. 3). Node Y also contains polytomous sequence types from Norway, including NOR_AS_1988 (red asterisk in Fig. 4). This taxon is embedded deeply within node C in Fig. 3.

3.4 Phylogenetic trees of individual PRV-1 segments

Bayesian phylogenetic analysis of all individual genome segments could not achieve consensus. Thus, maximum-

likelihood phylogenetic inference for each genome segment was performed to assess whether certain individual segments were the primary drivers of the concatenated genome tree topology, as well as evaluate the potential of past segment reassortment (Table 2 and Supplementary Fig. S2 and Table S3). Since no taxa from node F (genogroup III) or node G (genogroup IV) had genome data available, only formation of node L (genogroup I) and part of node C (genogroup II) were assessed, as was the monophyletic grouping of the established sequences from the northeast Pacific (nodes A and K from the concatenated genome and S1 trees, respectively) (Table 2). The segment trees performed poorly in recovering both genogroup I (node L, detected in segments M2 and S1) and genogroup II (node C, detected in segment S1 only) (Table 2). This indicates that the forty-eight segment sequences do not contain as many phylogenetically informative differences compared to the forty-eight genomes, 330 partial S1 sequences or 118 partial S1 haplotypes (Figs 2, 3 and 4, respectively). Northeast Pacific taxa (node A) formed a detectable genogroup in nine/ten segment trees, all of which were monophyletic. This suggests that the genetic relatedness of this genogroup is particularly robust.

Identification of reassortment events is possible when taxa fall into different monophyletic groups among segment trees. This can be limited by small taxa number, low genetic diversity,
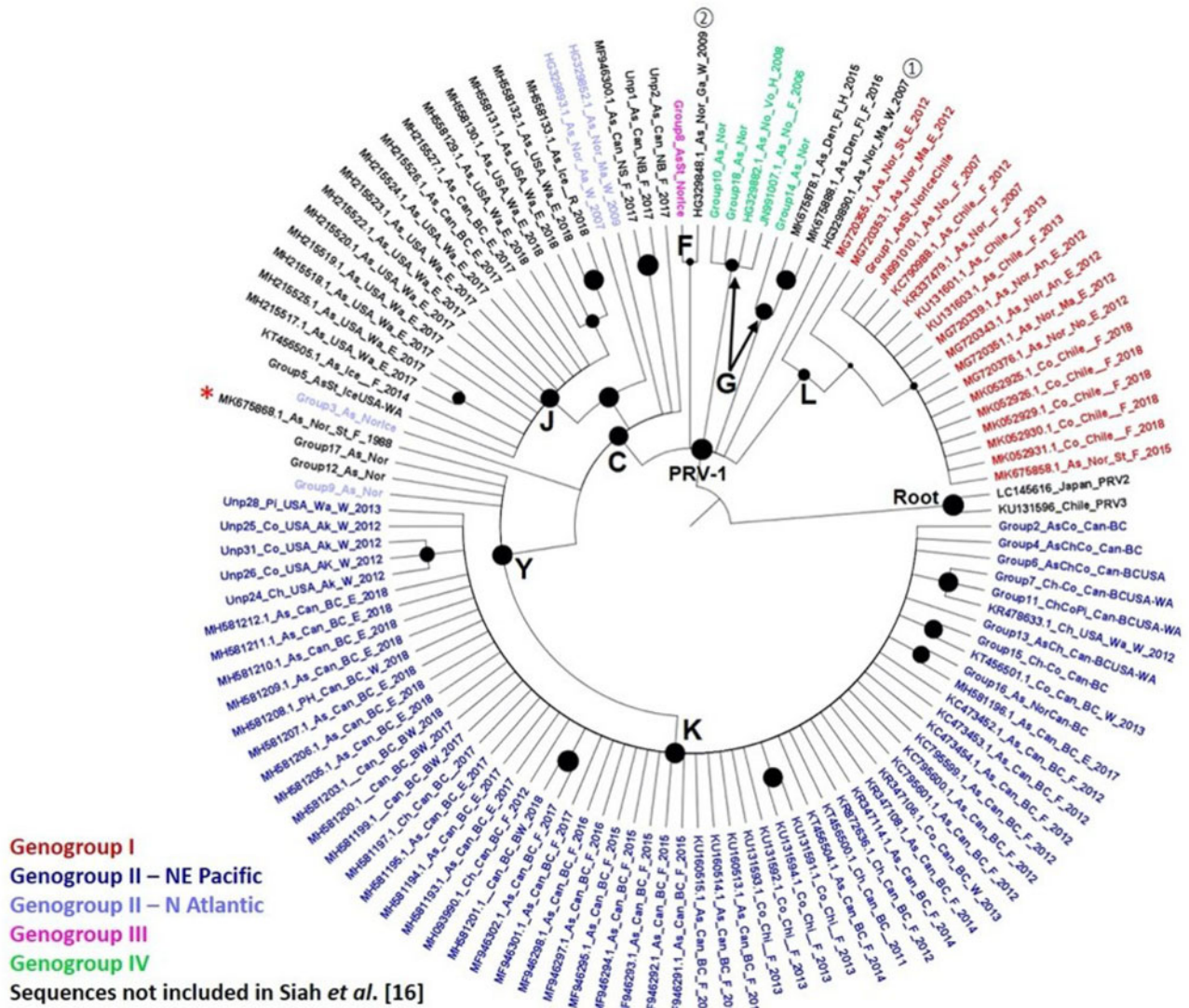
Figure 4. Bayesian phylogeny of the 118 haplotypes of 330 S1 sequences using MrBayes (Ronquist et al. 2005). Circles at nodes are scaled to their posterior probabilities. Samples are color coded for genogroups I–IV, as per legend, with three exceptions. All samples from northeast Pacific are color coded as genogroup II—NE Pacific, although not all these samples were included in Siah et al. (2015). Samples designated as ① and ② are included in Siah et al. (2015) but were not placed within a genogroup in that publication; both sit between genogroups II and IV. Sample marked with a red asterisk is sample Nor_AS_1988 as in Fig. 3. Terminal branches label with genogroups 1–18 are aggregates of samples with identical sequences. See Supplementary Table S6 for list of samples included in each genogroup. Lettered nodes described in text, and where appropriate correspond to the same nodes in Figs 2 and 3. Tree is rooted using PRV-2 and PRV-3 as indicated.

or lopsided selection criteria, all of which were present in our dataset. Despite this, there were some signs of reassortment. The Nor_As_1988 and two ECAN taxa showed topological variability, compared to the relative stability of other sequences from node C—genogroup II (Table 2). Virus sample Nor_As_1988 was found in a polytomous topology in six of the segment trees (L2, M1, M3, S1, S2, and S4), indicating a lack of resolution for its genetic relatedness to the other taxa. Some of these polytomous tree structures included a well-supported node with the two Faroe Islands samples, and in the remaining four segment trees (L1, L3, M2, and S3), this sample fell into a monophyletic genogroup that also contained the Faroe Islands samples. One of these monophyletic groups in the segment M2 analysis also contained the monophyletic genogroup II taxa from the northeast Pacific. The two samples from New Brunswick, Eastern Canada, were found in polytomous positions within seven of

the segment trees (ECAN SJR 17631 polytomous in segments L2, M2, M3, S1, S2, S3, and S4; while ECAN SJR 17227 was polytomous in segments L2, L3, M2, M3, S1, S2, and S4). When they were found in supported association with other taxa, ECAN SJR 17631 grouped with Norwegian samples three times and Faroe Islands samples three times, while ECAN SJR 17227 grouped with Norwegian samples four times and Faroe Islands samples three times.

3.5 Concatenated genome amino acid sequence phylogenetic analysis

We conducted a maximum-likelihood phylogenetic analysis of amino acid sequences encoded by the concatenated-segment genomes (Fig. 5). As with the concatenated nucleotide genome sequence tree, the S1 node L (genogroup I) taxa from Norway

Table 1. The 48 PRV-1 genome sequences used in this study. Samples were either sequenced as part of this study or taken from GenBank as part of cited studies. Segment-specific GenBank accession numbers and year of collection are included for each virus in Supplementary Table S2. No data on fish health were available.

Concatenated genome name[a]	Citation	Isolate name	Host species[b]	Source[c]	Location
NOR_mGl118_AS_2018	**This study**	**mGl118**	**Atlantic**	**Farmed**	**Norway**
NOR_mGl118H8Is_AS_2018	**This study**	**mGl118H8Is**	**Atlantic**	**Farmed**	**Norway**
ECAN_r17_631_AS_2017	**This study**	**r17_631**	**Atlantic**	**Farmed**	**Canada (NB)**
ECAN_r17_1227_AS_2017	**This study**	**r17_1227**	**Atlantic**	**Farmed**	**Canada (NB)**
WCAN_N78942_AS_2017	**This study**	**N78942**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_N7211_AS_2017	**This study**	**N7211**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_N78947_AS_2017	**This study**	**N78947**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_AS_2016	**This study**	**NA**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_N78717_AS_2017	**This study**	**N78717**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_K31554_AS_2005	**This study**	**K31554**	**Atlantic**	**Farmed**	**Canada (BC)**
WCAN_K31538_Ch_2001	**This study**	**K31538**	**Chinook**	**Wild[c]**	**Canada (BC)**
US_MP_Coho_1993	**This study**	**NA**	**Coho**	**Wild[c]**	**USA (WA)**
NOR_mDa115_AS_2018	**This study**	**mDa115**	**Atlantic**	**Farmed**	**Norway**
NOR_mDa314_AS_2018	**This study**	**mDa314**	**Atlantic**	**Farmed**	**Norway**
WCAN_G609_AS_2011	Di Cicco et al. (2018)	A.3.2-36_G609	Atlantic	Farmed	Canada (BC)
WCAN_G531_AS_2011	Di Cicco et al. (2018)	A.3.2-69_G531	Atlantic	Farmed	Canada (BC)
WCAN_G808_AS_2013	Di Cicco et al. (2018)	A.3.2-153_G808	Atlantic	Farmed	Canada (BC)
WCAN_G744_AS_2012	Di Cicco et al. (2018)	A.3.3-19_G744	Atlantic	Farmed	Canada (BC)
WCAN_G444_AS_2012	Di Cicco et al. (2018)	A.3.3-133_G444	Atlantic	Farmed	Canada (BC)
WCAN_G860_AS_2013	Di Cicco et al. (2018)	A.3.5-168_G860	Atlantic	Farmed	Canada (BC)
WCAN_G577_Ch_2011	Di Cicco et al. (2018)	P.2-1_G577	Chinook	Farmed	Canada (BC)
WCAN_G460_Ch_2013	Di Cicco et al. (2018)	P.2-3_G460	Chinook	Farmed	Canada (BC)
WCAN_G383_Ch_2013	Di Cicco et al. (2018)	P.2-95_G383	Chinook	Farmed	Canada (BC)
WCAN_G772_Ch_2013	Di Cicco et al. (2018)	P.2-99_G772	Chinook	Farmed	Canada (BC)
WCAN_G729_Ch_2012	Di Cicco et al. (2018)	P.3-3_G729	Chinook	Farmed	Canada (BC)
WCAN_G446_Ch_2013	Di Cicco et al. (2018)	P.3-37_G446	Chinook	Farmed	Canada (BC)
WCAN_G417_Ch_2012	Di Cicco et al. (2018)	P.3-120_G417	Chinook	Farmed	Canada (BC)
WCAN_B5690_AS_2013	Di Cicco et al. (2017)	B5690	Atlantic	Farmed	Canada (BC)
WCAN_B7274_AS_2013	Di Cicco et al. (2017)	B7274	Atlantic	Farmed	Canada (BC)
WCAN_J199_Co_2013	Siah et al. (2015)	BCJ19943_13	Coho	Wild[c]	Canada (BC)
WCAN_J319_AS_2013	Siah et al. (2015)	BCJ31915_13	Atlantic	Farmed	Canada (BC)
US_WSKFH12_CO_2014	Siah et al. (2015)	WSKFH12_14	Coho	Wild[c]	USA (WA)
WCAN_358_AS_2012	Kibenge et al. (2013)	VT06062012-358	Atlantic	Retail	Canada (BC)
WCAN_371_AS_2012	Kibenge et al. (2013)	VT06202012-371	Atlantic	Retail	Canada (BC)
NOR_AS_2007	Haatveit et al. (2017)	50607	Atlantic	Farmed	Norway
NOR_AS_2012	Wessel et al. (2017)	NOR2012-V3621	Atlantic	Farmed	Norway
CHILE_CGA280-05_AS_2011	Kibenge et al. (2013)	CGA280-05	Atlantic	Farmed	Chile
WCAN_161B_AS_2016 WCAN_165B_AS_2016	Polinski et al. (2019)	16-005	Atlantic	Farmed	Canada (BC)
WCAN_167B_AS_2016	Polinski et al. (2019)	16-011	Atlantic	Farmed	Canada (BC)
NOR_AS_1988	Dhamotharan et al. (2019)	NOR-1988	Atlantic	Farmed	Norway
NOR_AS_1997	Dhamotharan et al. (2019)	NOR-1977	Atlantic	Farmed	Norway
NOR_TT_AS_2005	Dhamotharan et al. (2019)	NOR-2005/TT	Atlantic	Farmed	Norway
NOR_SSK_AS_2015	Dhamotharan et al. (2019)	NOR-2015/SSK	Atlantic	Farmed	Norway
NOR_MS_AS_2015	Dhamotharan et al. (2019)	NOR-2015/MS	Atlantic	Farmed	Norway
FO-1978_AS_2015	Dhamotharan et al. (2019)	FO-1978/16	Atlantic	Farmed	Faroe Islands
FO-41_AS_2016	Dhamotharan et al. (2019)	FO/41/16	Atlantic	Farmed	Faroe Islands
WCAN_BC17_AS_2017	This study	R2BC17	Atlantic	Escaped	Canada (BC)

The sequences from this study are highlighted in bold.

and Chile was not recovered as a monophyletic genogroup. Both nodes A and C (genogroup II) were recovered, with node A containing the established northeast Pacific variants and node P containing samples from Eastern Canada, Faroe Islands, as well as the escapee BC17 and Nor_As_1988 samples. Node P is similar to genome tree node C except for the positions of the two Faroe Island taxa, which are now paraphyletic with Eastern and Western Canada, and BC17, which is now a sister taxa of the eastern Canada samples.

3.6 S1 and M2 amino acid sequence analysis

Dhamotharan et al. (2019) identified amino acid (AA) sites in the S1-encoded σ3 and p13 and M2-encoded μ1 genes, which the authors hypothesized as being conserved in HSMI-associated sequence (node L in our study) and absent in Nor_AS_1988 (archived PRV-1 variant present in Norway before the emergence of HSMI). Here, all node L sequences (Norway and Chile) possessed the previously identified residues (Dhamotharan et al. 2019) (Table 3, footnote a) with four infrequently observed

Table 2. Comparison of segment trees to previously published genogroups I and II topology,[a] and whether northeast Pacific viruses form a monophyletic genogroup.[b] Numbers in parentheses are the bootstrap value for the node. Nodes that are not recovered are indicated as not detected ND.

Segment	Genogroup I	Genogroup II	Northeast Pacific
L1	ND	ND	Monophyletic (98)
L2	ND	ND	Monophyletic (98)
L3	ND	ND	Monophyletic (92)
M1	ND	ND	Monophyletic (85)
M2	Monophyletic (77)	ND	Monophyletic (79)
M3	ND	ND	Monophyletic (99)
S1	Monophyletic (100)	Monophyletic (89)	Monophyletic (76)
S2	ND	ND	Monophyletic (97)
S3	ND	ND	Monophyletic (87)
S4	ND	ND	ND

[a]Genogroup nomenclature according to Garseth, Ekrem, and Biering (2013), where genogroup I includes taxa from Norway and Chile (recovered as node L in the Bayesian S1 tree; Fig. 3) and genogroup II includes taxa from Norway and northeast Pacific (recovered as node C in the Bayesian concatenated genome and S1 trees, Figs 2 and 3).
[b]The established northeast Pacific PRV-1 variants were recovered as nodes A and K from the concatenated genome and S1 trees, respectively.

variant residues (observed in ≤2 samples; Table 3). All sequences outside node L had AA sequence highly similar to Nor_AS_1988 (Nor1988, Table 3). The Nor1988-like AA pattern was observed in PRV-1 from a range of species over the period of 1993–2018 in northeast Pacific, Eastern Canada, Chile (farmed Pacific salmon), Iceland, and Faroe Islands, as well as non-node L and Norway/Chile (non-node L). Northeast Pacific sequences (node K) were identical to Nor1988 at the previously identified residues (Dhamotharan et al. 2019) with only one infrequent variant (Table 3). However, node K sequences had frequent variants (≥3 samples) at other amino acid residues, including σ3 (L_{180}, A_{230}, and D_{315}) and μ1 (S_{313} and I_{547}); these node K μ1 variants were observed in farmed and free-ranging Chinook and Coho Salmon (1993–2013) but not farmed Atlantic salmon samples (Table 3). Node J, which included the WA Atlantic salmon escapees, had three unique σ3 amino acid residues (A_{69}, R_{74}, and E_{118}) and one unique p13 AA residues (D_{48}) not present in other sequences in our analysis (node J, Table 3). A second fixed p13 variant in node J (K_{92}) was also observed in PRV-1 sequences from free-ranging Atlantic salmon (node F).

4. Discussion

The primary goal of this study was to explicitly evaluate the spatial, temporal, and host patterns of PRV-1 in the northeast Pacific by conducting a detailed analysis of genetic variation in the viral genome using Bayesian inference methods. The strength of this approach is that it infers properties of the evolving population beyond tree topology like ancestral date and location, but this also means that the analyses must be extensively parameterized and statistically validated. Without this validation process, the risk of obtaining biologically implausible inferences increases.

Analyses of nucleotide and amino acid concatenated genomes, S1 nucleotide sequences, and individual segment sequences were all consistent indicating that established northeast Pacific PRV-1 variants were a monophyletic genogroup and therefore phylogenetically distinct from PRV-1 in other
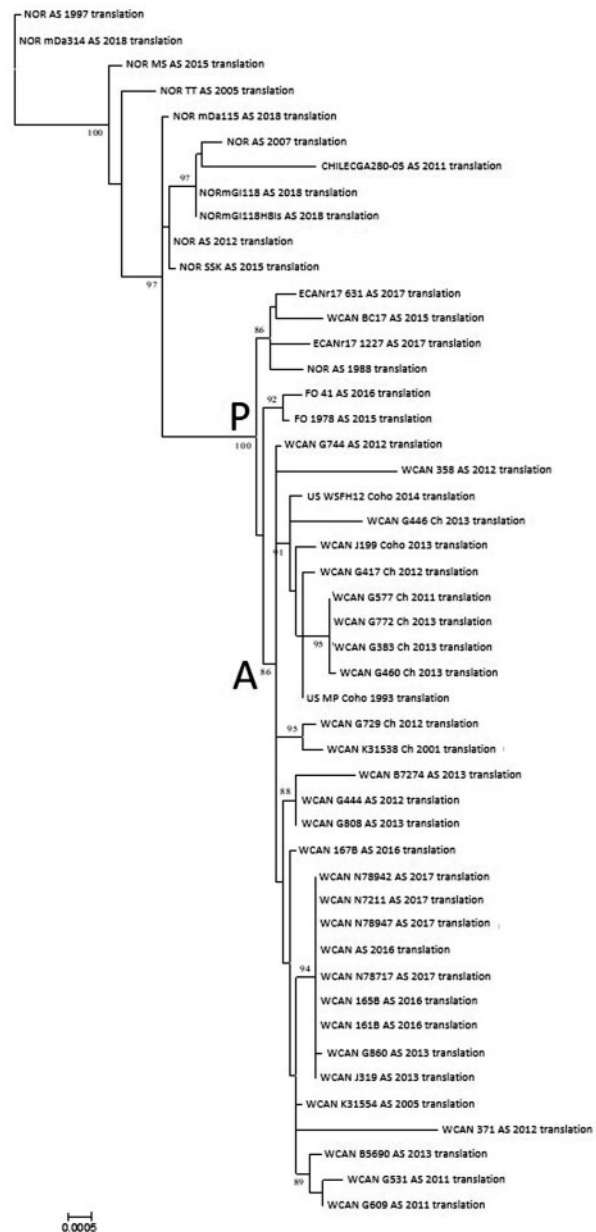
Figure 5. Maximum-likelihood phylogenetic analyses of each segment of forty-eight genomes coding sequences, translated *in silico* to amino acids. Analysis was conducted in MEGA using JTT + G4 + I substitution model and 1,000 bootstrap replicates. Nodes with less than seventy bootstrap support are collapsed, and bootstrap support is indicated at each node. See Table 1 for topology summary.

locations (Figs 3–5 and Table 2). While PRV-1 in the northeast Pacific is distinct compared to other regions, there was no statistical support for further subdivisions within the region. Furthermore, there was no detectable signal for the virus evolving differently among different hosts in the northeast Pacific. The established northeast Pacific lineage, node K, was descended from node C in both the genome and S1 trees; this node also included sequences from Eastern Canada, Iceland, and Norway. Both S1 and genome phylogenetic analyses predicted that PRV-1 had its origins in the North Atlantic.

Previous studies have mainly relied on the diverse PRV-1 S1 segment sequence, which encodes the outer capsid protein σ3,

Table 3. Piscine orthoreovirus 1 (PRV-1) segments S1 (σ3 and p13) and M2 (μ1) amino acids residues previously hypothesized to be associated with heart and skeletal muscle disease (HSMI) samples by Dhamotharan et al. (2019) corresponding to node L in this study (Fig. 3). The amino acid residues identified as being distinct from a putatively low virulence Nor_AS_1988 (Nor1988) sample are shown below,[a] along with other amino acid variants identified in this study with a special emphasis on northeast Pacific informative nodes (K and J). The Nor1988 amino acid pattern was observed in all North Atlantic samples outside node L, with the exceptions in footnotes.[b,c] Non-bolded variants were infrequent (≤2 samples) while bolded variants were frequent (≥3 samples).

Segment–gene	AA position[a]	Amino acid residues (variants[a])			
		Node L	Nor1988[b,c]	Node K[d]	Node J
S1–σ3	69[a]	V (I)	T	T	A
	74	Q	Q	Q	R
	78[a]	D (N, E)	E	E (K)	E
	85[a]	T	A	A	A (T)
	117[a]	N	T	T	T
	118	D	D	D	E
	137[a]	V	I	I	I
	156[a]	T (P)	A	A	A
	157[a]	A	S (P, D[b])	S	S
	174[a]	E	K	K	K
	180	S	S	S (L)	S
	206[a]	A	V	V	V
	218[a]	V	I	I	I
	230	V	V	V (**A**)	V (M)
	315	G	G	G (**D**)	G
S1–p13	16[a]	A (V)	V	V	V
	23	P	P	H (P)	P
	39[a]	M	T	T	T
	48	N	N	N	D
	50[a]	T	M	M	M
	52[a]	I	K	K	K
	74[a]	V	A	A	A
	76	Q	R (**L**, Q[c])	R	R
	81[a]	R	Q	Q	Q
	91[a]	M	L	L	L (P)
	92	T	T (K)	T	K
M2–μ1	184[a]	T	S	S	S
	262[a]	S	A	A	A
	313	Y	Y	Y (**S**[d])	Y
	370[a]	D	N	N	N
	547	V	V	V (**I**[d])	V

[a]Amino acids residues identified by Dhamotharan et al. (2019) as being distinguished between node L (HSMI-associated) and Nor1988.
[b]One Eastern Canada (New Brunswick, NB) had a S1–σ3 residue 157 substitution of P, a second NB sequence had the common S residue. One Faroe Islands (FO) had a D substitution at this position.
[c]Fifty sequences from Chile, Denmark, Faroe Islands, Iceland, and Norway had a S1–p13 residue 76 of Q. Five sequences from Norway and one from Denmark had L residues at this position.
[d]In the NE Pacific sequences, M2–μ1 residue 313 and 547 were a Y and V in Atlantic salmon, respectively, while Chinook and Coho Salmon samples had S and I in these respective positions.

for phylogenetic inferences (Garseth, Ekrem, and Biering 2013; Kibenge et al. 2013; Siah et al. 2015; Godoy et al. 2016; Di Cicco et al. 2017, 2018; Gunnarsdóttir 2017; Cartagena et al. 2018; Dhamotharan et al. 2019). Among previous phylogenetic studies, two systems of nomenclature for describing the relationships among PRV-1 have been commonly used. Kibenge et al. (2013) assigned PRV-1 into two major sub-genogroups (sub-

genogroups Ia and Ib). However, Garseth, Ekrem, and Biering (2013) and Siah et al. (2015) recovered four genogroups (I–IV) from S1 segment analysis. In the present study, both Bayesian S1 phylogenetic methods (Figs 3 and 4) recovered three of the four previously described S1 genogroups: I, node L; II, node C; and III, node F. In Fig. 3, genogroup IV—node G was recovered, but it was not recovered in Fig. 4. All established northeast Pacific PRV-1 variants belong to monophyletic genogroup II—node C and this monophyletic genogroup was also recovered in the concatenated genome analysis (Fig. 2). Genogroup I—node L corresponded to genogroup 1 b of Kibenge et al. (2013) and this genogroup has been hypothesized to be a higher virulence form of the virus associated with the emergence of HSMI in Norway (Dhamotharan et al. 2019). However, genogroup I was not recovered in the concatenated genome analysis. Instead, these Norwegian and Chilean sequences were paraphyletic and formed a series of three separate monophyletic genogroups. Concatenated genomes were not available for Norway and Faroe Islands samples associated with genogroups III and IV corresponding to nodes F and G in the S1 tree (Figs 3 and 4).

In contemporary samples from the northeast Pacific, PRV-1 is consistently detected at low prevalence (commonly <10%) in free-ranging Chinook and Coho Salmon from Alaska to Washington (Marty et al. 2015; Purcell et al. 2018; Polisnki and Garver 2020). When farmed populations of Atlantic Salmon were tested throughout rearing, PRV-1 infections reached 100 per cent by intermediate stages of rearing after 100 degree days post transfer to sea cages (Polisnki and Garver 2020). In our analysis, the 146 S1 sequences and thirty-one concatenated genomes representing these established northeast Pacific PRV-1 variants formed a well-supported genogroup, which shared a common ancestor (nodes A and K in the genome and S1 trees, respectively) (Figs 3–5). Node K also included four S1 sequences from Chilean farmed Coho Salmon (black arrowhead Fig. 3). This node represents an inferred transfer event from the northeast Pacific to Chile. The movement of PRV-1 from the northeast Pacific to Chile is consistent with past commercial trade of Coho Salmon from the northeast Pacific (Crawford and Muir 2008). A fifth sequence identified as being from a Norwegian Atlantic salmon is 100 per cent identical to a sequence from BC Atlantic salmon analyzed in the same study [43] (GenBank accession number MF946299 and MF946290); the validity of the sequence metadata requires further investigation. Further genome sequencing could potentially resolve additional substructure within nodes A and K. However, it remains to be seen if partial S1 segment sequences from additional taxa will improve resolution of regional differences within the northeast Pacific, if they exist.

In August 2017, ~250,000 Atlantic salmon were released during a net-pen failure near Cyprus Island, WA (Clark et al. 2017). These salmon were descended from ~900,000 eggs transferred from Iceland to Washington in 2015 and ~370,000 smolts were transferred into the net pens in May 2016. The PRV-1 S1 sequences from these escaped Atlantic salmon shared a common ancestor (node J) with sequences from Iceland. Our results were concordant with Kibenge et al. (2019) that also reported escapees carried a PRV-1 variant similar to Iceland. The node J sequences shared unique S1 amino acid residues for σ3 (A_{69}, R_{74}, and E_{118}) and p13 (D_{48}, K_{92}) genes, which were not seen in samples from other geographical areas including the northeast Pacific (Table 3). The functional significance of these amino acid changes, if any, is unknown. The observation that PRV-1 in the escapees was linked epidemiologically to the Icelandic broodstock suggests these salmon were most likely infected during

early rearing, rather than during the 15 months in the net pen prior to release. Haatveit et al. (2017) showed that PRV-1 causes an acute infection in Atlantic salmon (peaking at 5–7 weeks post-initial infection) before the virus enters a long-term, persistent carrier state. Thus, the escapees were more likely persistent carriers of PRV-1. Previous studies indicate persistently infected Atlantic salmon (>5 months post-initial infection) had diminished capacity to transmit the virus (Garver et al. 2016). This is an important risk consideration as to the likelihood that the exotic node J variant will become established. Nonetheless, ongoing management surveillance programs for PRV-1 in WA includes S1 sequencing to assess the possibility that the node J variant has established. Since nodes J and K were monophyletic in analyses, this indicates that this segment contains sufficient genetic variability to discriminate the established northeast Pacific variants from the escapee-associated variant, as well as exotic variants from other geographical regions.

The potential that the established northeast Pacific variant PRV-1 was introduced from the North Atlantic and, if so, the date of that introduction has been a topic of interest. Kibenge et al. (2013, 2017) first proposed that PRV-1 was recently introduced into BC from Norway sometime between 2006 and 2008. However, there have been confirmed molecular detections of PRV-1 in archived samples dating back to the mid-1980s in Pacific salmon (Marty et al. 2015) and the present study includes a genome sequence from an archived WA Coho Salmon ca. 1993. While this manuscript was in review, partial PRV-1 S1 and S3 sequences from a 1977 Steelhead Trout collected in BC (Marty et al. 2015) were made available in GenBank (MT506522– MT506523) further supporting longer term presence of the virus in the northeast Pacific. Our concatenated genomes analysis predicted sequences from northeast Pacific lineage descended from a common ancestor that occurred between 1900 and 1975 (median 1950, node A; Fig. 2). In contrast, the S1 analysis predicted a more recent and non-overlapping range of dates for the northeast Pacific lineage (1986–1993, median 1990, node K; Fig. 3). Both the concatenated genome and S1 analyses predicted that the most recent common ancestor of the northeast Pacific sequences was in the Atlantic Basin (node C in Figs 2–4) without agreement on a more precise location (Eastern Canada or Norway). The analyses recovered a different median age of node C that differs by 194 years. This difference is likely due to how well the genome and S1 datasets represent the four PRV-1 genogroups (e.g. the genome tree under-represents genogroups I–IV). However, the nodes representing the northeast Pacific samples (node A in genome tree and node K in S1 tree) are more equitable, as is the narrower range of median dates, which are 40 years apart. We are inclined to think this inference of a translocation event in the early twentieth century is likely.

Predicted dates of divergence of PRV-1 in this study, as well as in other published studies, need to be interpreted cautiously taking into consideration the assumptions made and the limitation of these types of analysis. The ancestral date estimation in Bayesian analysis is related to the date of sample collection and rate of substitution of the sequences used as prior parameters in the analysis. The concatenated genome has the advantage to provide a deeper analysis of the substitution rate, whereas the S1 tree analysis contained more sequences, in particularly, S1 sequences from genogroup II–IV that were not available for the concatenated genome analysis. The differences in input data feeding the two genome and S1 analyses likely explain the discrepancies of the dates and origin of transfer. These predicted dates should also be considered in the context of fish movements between Atlantic and Pacific waters, which are

documented in historical records. Transfer of Atlantic Salmon from the North Atlantic into the northeast Pacific were recorded from the late nineteenth century (MacCrimmon and Gots 1979). Many of the early introduction attempts into BC and Washington were from stocks derived from New Brunswick, Quebec, and Maine starting in 1904. Approximately 8.6 million Atlantic Salmon were introduced to Vancouver Island waters between 1905 and 1935, although these fish did not become established (Klinkenberg 2018). In 1932–1935, half a million of brown trout were released in Vancouver Island waters, where they naturalized (Ginetz 2002). We are unaware of any PRV-1 surveillance performed on these introduced populations of brown trout. These historical dates were not represented in the concatenated genome and S1 tree analysis as the oldest sequence that was available for our analysis was the pre-HSMI Norwegian form of PRV-1 (Nor_AS_1988 sample).

The topological position and biological features of the sample Nor_AS_1988 was of interest due to the fact that this sample was collected prior to the emergence of HSMI in Norway and has been theorized to represent a low virulence lineage of PRV-1 (Dhamotharan et al. 2019). In our concatenated genome analysis, the sample grouped within the genogroup II—node C and was most closely related two sequences from farmed Atlantic Salmon collected in New Brunswick, Canada, in 2017, where HSMI, to our knowledge, has not been reported. In the S1 trees, Nor_AS_1988 is found within the genogroup II—node C. All northeast Pacific PRV-1 variants are descended from this node, along with other variants from Eastern Canada, Iceland, and Norway. The location of the NOR_AS_1988 position in the genome and S1 trees suggests that the established northeast Pacific PRV-1 lineage was more closely related to NOR_AS_1988 than to other sequences. Thus, the viruses circulating in the northeast Pacific are more closely related to the 'pre-HSMI' variant than they are to the variants associated with HSMI emergence in Norway.

Dhamotharan et al. (2019) 'HSMI-associated' PRV-1 sequences fall into the S1 genogroup I—node L in our analysis, while the samples designated as low HSMI virulence included all the non-node L sequences including the 1988 pre-HSMI sample discussed above. Based on these assumed phenotypes, the authors identified amino acid residues in the S1 (σ3 and p13) and M2 (μ1) proteins that were unique to the HSMI-associated variants. Our analysis of a larger dataset largely corroborated that these amino acid residues are unique to node L found in only Norway and Chile (Table 3). An important caveat to the high and low virulence phenotype hypothesis scheme is the fact that information on disease state from field samples is rarely uniformly assessed or reported.

In the northeast Pacific, infections with PRV-1 are considered to have the potential to exacerbate instances of cardiopathy in farmed Atlantic Salmon, as well as to play a role in the development of jaundice syndrome in farmed Chinook Salmon (Di Cicco et al. (2018) reviewed by Polisnki and Garver (2020)). Experimental challenge trials using PRV-1 from the northeast Pacific have resulted in extreme PRV blood infections in Atlantic and Pacific salmon but failed to generate notable disease or physiological impairment (Garver et al. 2016; Polinski et al. 2019; Zhang et al. 2019; Polisnki and Garver 2020; Purcell et al. 2020). More recently, preliminary laboratory challenge results of PRV-1 demonstrated that a northeast Pacific PRV-1 and the pre-HSMI Nor_AS_1988 isolate caused lower pathology in Atlantic salmon relative to a genogroup I—node L Norwegian PRV-1 (Wessel et al. 2018). The high and low HSMI virulence hypothesis and significance of the 'HSMI-associated' amino acid residues

proposed by Dhamotharan et al. (2019) require additional empirical validation. Nonetheless, these particular amino acid residues were not found in either the established northeast Pacific PRV-1 variants or the Icelandic node J variants associated with the Atlantic salmon escapees. More work is needed to fully understand the physiological and disease consequences of PRV-1 across its host and geographic range using controlled infection trials. Therefore, firm classification of isolates into virulence genogroups is likely premature.

Based on our concatenated genome analysis, PRV-1 in the North Atlantic displays a higher amount of genetic diversity when compared to the northeastern Pacific. Low sequence diversity has been reported previously for PRV-1 sequences originating from samples collected in the northeast Pacific (Siah et al. 2015; Di Cicco et al. 2018). In this study, established PRV-1 from the northeast Pacific showed low levels of nucleotide and amino acid sequence diversity over a 25-year period (ca. 1993–2018). The archived Nor_AS_1988 also shared high genome similarity with the northeast Pacific variants (99.0–99.2%). This low diversity likely explains why individual segments trees of the forty-eight genomes did not run to consensus under a Bayesian framework, since there was likely insufficient genetic diversity for the algorithm to resolve the coalescent process. It is possible that if segment termini had been included, segment only trees may have reached consensus, although termini also may be too highly conserved to add informative diversity. Siah et al. (2015) reported that temporal homogeneity is not unique to PRV-1 and provide examples of other salmon RNA viruses, which display this feature such as Pacific salmon paramyxovirus. Several hypotheses could explain genetic homogeneity in an RNA virus across a geographical and/or temporal range, including recent epidemiological linkages or an established host–pathogen relationship at a relatively stable fitness peak in the reservoir population. The reasons for this high homogeneity have not been adequately explored.

Sequencing of the S1 segment has provided insight into the global evolution of PRV-1 but this segment may have limited power to resolve some relationships among PRV-1 variants at finer scales. Additionally, analysis of a single segment of a multi-segmented genome can only provide the evolution history for that one segment. As an example, two Norwegian sequences (NOR_AS_1997 and NOR_mDa314_AS_2018) clustered independently in the concatenated genome analysis primarily due to a relatively high number of nucleotide differences in their L1 and L2 segments, but the uniqueness of these variants was not reflected in the S1 segment analysis. At present thirty-two of forty-eight (67%) PRV-1 full-genome sequences have been obtained from samples collected in the northeastern Pacific. Full-genome sequences are not available from many regions in the North Atlantic. Additionally, all available genomes from Norway, with the exception of NOR_AS_1988, belong to the HSMI-associated variants as proposed by Dhamotharan et al. (2019). To better understand the relationships between genetic variants and processes that have shaped global PRV-1 evolution, additional full-genome sequences are needed, particularly from North Atlantic wild or free-ranging fish.

Since viruses with segmented genomes are capable of reassortment, individual segments can have different evolutionary histories and therefore different tree topologies (Svinti, Cotton, and McInerney 2013). If different segment trees have stronger evolutionary signal (in the form of better supported internal nodes), these segments may drive the tree topology of a concatenated genome analysis, making a tree based on that

dominant sequence a poor representation of the virus as a whole. There were several virus samples that showed topological variation among the segment trees including Nor_As_1988 and two variants from Eastern Canada. These variants had polytomous topology on some segment trees, while forming well-supported nodes with either Faroe Islands or Norwegian variants on other segments. Dhamotharan et al. (2019) also reported that several of the 'HSMI-associated' variants (S1 genogroup I—node L), as well as Nor_As_1988, changed genogroup affiliation depending on the segment tree. Based on the positioning of Nor_As_1988, the authors suggest that M3 and S3 may have undergone reassortment. If true, reassortment may be hypothesized to explain the emergence of HSMI in Norwegian aquaculture (Dhamotharan et al. 2019). At present, these results cannot be interpreted as conclusive of reassortment because these analyses are hindered by low genetic diversity and limited number of taxa in our datasets. As discussed above, there is a relatively small number of full PRV-1 genome sequences available for much of the North Atlantic outside Norwegian Atlantic salmon aquaculture.

The established northeast Pacific PRV-1 variant was derived from a single introduction from North Atlantic waters. The timing and exact source of introduction, either from eastern North America or from European waters of the North Atlantic, was inconclusive. The northeast Pacific variant was genetically distinct from the PRV-1 variant associated with an Atlantic salmon escape event in Washington State and variation in the S1 segment was sufficient to distinguish this exotic variant. A more balanced representation of full PRV-1 genomes across its range, as well as additional sequencing of archived samples, will provide a better understanding of phylogeographical patterns and transmission routes.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

References

Adamek, M. et al. (2019) 'Detection of Piscine Orthoreoviruses (PRV-1 and PRV-3) in Atlantic Salmon and Rainbow Trout

Farmed in Germany', *Transboundary and Emerging Diseases*, 66: 14–21.

Bass, A. L. et al. (2017) 'A Survey of Microparasites Present in Adult Migrating Chinook Salmon (*Oncorhynchus tshawytscha*) in South-Western British Columbia Determined by High-Throughput Quantitative Polymerase Chain Reaction', *Journal of Fish Diseases*, 40: 453–77.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

Cartagena, J. et al. (2018) 'Detection of Piscine Orthoreovirus in Farmed Rainbow Trout from Chile', *Aquaculture*, 493: 79–84.

Clark, D. et al. 2017. *Cypress Island Atlantic Salmon Net Pen Failure: An Investigation and Review*. Olympia, WA: Washington Department of Natural Resources.

Crawford, S. S., and Muir, A. M. (2008) 'Global Introductions of Salmon and Trout in the Genus Oncorhynchus: 1870–2007', *Reviews in Fish Biology and Fisheries*, 18: 313–44.

Delport, W. et al. (2010) 'DataMonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology', *Bioinformatics*, 26: 2455–7.

Dhamotharan, K. et al. (2018) 'Molecular and Antigenic Characterization of Piscine Orthoreovirus (PRV) from Rainbow Trout (*Oncorhynchus mykiss*)', *Viruses*, 10: 170.

——— et al. (2019) 'Evolution of the Piscine Orthoreovirus Genome Linked to Emergence of Heart and Skeletal Muscle Inflammation in Farmed Atlantic Salmon (*Salmo salar*)', *Viruses*, 11: 465.

Di Cicco, E. et al. (2017) 'Heart and Skeletal Muscle Inflammation (HSMI) Disease Diagnosed on a British Columbia Salmon Farm through a Longitudinal Farm Study', *PLoS One*, 12: e0171471.

——— et al. (2018) 'The Same Strain of Piscine Orthoreovirus (PRV-1) is Involved with the Development of Different, but Related, Diseases in Atlantic and Pacific Salmon in British Columbia', *Facets*, 3: 599–641.

Drummond, A. J. et al. (2005) 'Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences', *Molecular Biology and Evolution*, 22: 1185–92.

——— et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29: 1969–73.

———, and Rambaut, A. (2007) 'BEAST: Bayesian Evolutionary Analysis by Sampling Trees', *BMC Evolutionary Biology*, 7: 214.

Duchene, S. et al. (2015) 'The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data', *Molecular Biology and Evolution*, 32: 1895–906.

Garseth, A. H., Ekrem, T., and Biering, E. (2013) 'Phylogenetic Evidence of Long Distance Dispersal and Transmission of Piscine Reovirus (PRV) between Farmed and Wild Atlantic Salmon', *PLoS One*, 8: e82202.

Garver, K. A. et al. (2016) 'Piscine Orthoreovirus from Western North America is Transmissible to Atlantic Salmon and Sockeye Salmon but Fails to Cause Heart and Skeletal Muscle Inflammation', *PLoS One*, 11: e0146229.

Ginetz, R. M. J. 2002. *On the Risk of Colonization by Atlantic Salmon in BC Waters*. Campbell River, BC: BC Salmon Farmers Association.

Godoy, M. G. et al. (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: 98.

Guindon, S. et al. (2010) 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML', *Systematic Biology*, 59: 307–21.

Gunnarsdóttir, H. M. (2017) 'Screening for Viruses in Atlantic Salmon in Iceland ISAV, PMCV and PRV', MSc thesis, 87 pp.

Haatveit, H. M. et al. (2017) 'Viral Protein Kinetics of Piscine Orthoreovirus Infection in Atlantic Salmon Blood Cells', *Viruses*, 9: 49.

Hasegawa, M. et al. (1985) 'Phylogenetic Relationships among Eukaryotic Kingdoms Inferred from Ribosomal RNA Sequences', *Journal of Molecular Evolution*, 22: 32–8.

Hrushowy, S. 2018. 'A Molecular Investigation of the Dynamics of Piscine Orthoreovirus in a Wild Sockeye Salmon Community on the Central Coast of British Columbia', MSc thesis, Simon Fraser University, Vancouver, 137 pp.

Key, T. et al. (2013) 'Piscine Reovirus Encodes a Cytotoxic, Non-Fusogenic, Integral Membrane Protein and Previously Unrecognized Virion Outer-Capsid Proteins', *Journal of General Virology*, 94: 1039–50.

Kibenge, M. J. et al. (2013) 'Whole-Genome Analysis of Piscine Reovirus (PRV) Shows PRV Represents a New Genus in Family Reoviridae and Its Genome Segment S1 Sequences Group It into Two Separate Sub-Genotypes', *Virology Journal*, 10: 230.

Kibenge, M. J. T. et al. (2017) 'Formal Comment on: Piscine Reovirus: Genomic and Molecular Phylogenetic Analysis from Farmed and Wild Salmonids Collected on the Canada/US Pacific Coast', *PLoS One*, 12: e0188690.

——— et al. (2019) 'Piscine Orthoreovirus Sequences in Escaped Farmed Atlantic Salmon in Washington and British Columbia', *Virology Journal*, 16: 41.

Klinkenberg, B. 2018. *E-Fauna BC: Electronic Atlas of the Fauna of British Columbia* <www.efauna.bc.ca>. Vancouver: Lab for Advanced Spatial Analysis, Department of Geography, University of British Columbia.

Kumar, S., Stecher, G., and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular Biology and Evolution*, 33: 1870–4.

Labrut, S. et al. 2018. 'Emergence of the heart and skeletal muscle inflammation syndrome in France' in *World Aquaculture Society Meeting*, Montpellier, France, 25–29 August. World Aquaculture Society.

Laurin, E. et al. (2019) 'Histopathological and Novel High-Throughput Molecular Monitoring Data from Farmed Salmon (*Salmo salar* and *Oncorhynchus* spp.) in British Columbia, Canada, from 2011–2013', *Aquaculture*, 499: 220–34.

Lefort, V., Longueville, J. E., and Gascuel, O. (2017) 'SMS: Smart Model Selection in PhyML', *Molecular Biology and Evolution*, 34: 2422–4.

Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform', *Bioinformatics*, 25: 1754–60.

MacCrimmon, H. R., and Gots, B. L. (1979) 'World Distribution of Atlantic Salmon, *Salmo salar*', *Journal of the Fisheries Research Board of Canada*, 36: 422–57.

Markussen, T. et al. (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: e70075.

Marty, G. D. et al. (2015) 'Piscine Reovirus in Wild and Farmed Salmonids in British Columbia, Canada: 1974–2013', *Journal of Fish Diseases*, 38: 713–28.

Miller, K. M. et al. (2014) 'Infectious Disease, Shifting Climates, and Opportunistic Predators: Cumulative Factors Potentially Impacting Wild Salmon Declines', *Evolutionary Applications*, 7: 812–55.

—— et al. (2017) 'Molecular Indices of Viral Disease Development in Wild Migrating Salmon', *Conservation Physiology*, 5: cox036.

Morton, A. et al. (2017) 'The Effect of Exposure to Farmed Salmon on Piscine Orthoreovirus Infection and Fitness in Wild Pacific Salmon in British Columbia', *PLoS One*, 12: e0188793.

Nibert, M. L., and Duncan, R. (2013) 'Bioinformatics of Recent Aqua- and Orthoreovirus Isolates from Fish: Evolutionary Gain or Loss of FAST and Fiber Proteins and Taxonomic Implications', *PLoS One*, 8: e68607.

Palacios, G. et al. (2010) 'Heart and Skeletal Muscle Inflammation of Farmed Salmon is Associated with Infection with a Novel Reovirus', *PLoS One*, 5: e11487.

Polisnki, M., and Garver, K. (2020) 'Characterization of Piscine Orthoreovirus (PRV) and Associated Diseases to Inform Pathogen Transfer Risk Assessments in British Columbia', *Canadian Science Advisory Secretariat Research Document*, 35 pp.

Polinski, M. P. et al. (2019) 'Piscine Orthoreovirus Demonstrates High Infectivity but Low Virulence in Atlantic Salmon of Pacific Canada', *Scientific Reports*, 9: 3297.

Purcell, M. K. et al. (2018) 'Molecular Testing of Adult Pacific Salmon and Trout (*Oncorhynchus* spp.) for Several RNA Viruses Demonstrates Widespread Distribution of Piscine Orthoreovirus in Alaska and Washington', *Journal of Fish Diseases*, 41: 347–55.

—— et al. (2020) 'Consequences of Piscine Orthoreovirus Genotype 1 (PRV-1) Infections in Chinook Salmon (*Oncorhynchus tshawytscha*), Coho Salmon (*O. kisutch*) and Rainbow Trout (*O. mykiss*)', *Journal of Fish Diseases*, 43: 719–28.

Rodger, H. D., McCleary, S. J., and Ruane, N. M. (2014) 'Clinical Cardiomyopathy Syndrome in Atlantic Salmon, *Salmo salar* L', *Journal of Fish Diseases*, 37: 935–9.

Ronquist, F., Huelsenbeck J. P. J. P., and , and van der Mark P. 2005. MrBayes3 manual. *http://mrbayes.csit.fsu.edu/mb3.1_-manual.pdf* .

Siah, A. et al. (2015) 'Piscine Reovirus: Genomic and Molecular Phylogenetic Analysis from Farmed and Wild Salmonids Collected on the Canada/US Pacific Coast', *PLoS One*, 10: e0141475.

Sievers, F. et al. (2011) 'Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega', *Molecular Systems Biology*, 7: 539.

Svinti, V., Cotton, J. A., and McInerney, J. O. (2013) 'New Approaches for Unravelling Reassortment Pathways', *BMC Evolutionary Biology*, 13: 1.

Takano, T. et al. (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: e0165424.

Tamura, K. et al. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, 30: 2725–9.

Teffer, A. K. et al. (2017) 'Capture Severity, Infectious Disease Processes and Sex Influence Post-Release Mortality of Sockeye Salmon Bycatch', *Conservation Physiology*, 5: 1–33.

—— et al. (2018) 'Infections, Fisheries Capture, Temperature, and Host Responses: Multistressor Influences on Survival and Behaviour of Adult Chinook Salmon', *Canadian Journal of Fisheries and Aquatic Sciences*, 75: 2069–83.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) 'CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice', *Nucleic Acids Research*, 22: 4673–80.

Vendramin, N. et al. (2019) 'Presence and Genetic Variability of Piscine Orthoreovirus Genotype 1 (PRV-1) in Wild Salmonids in Northern Europe and North Atlantic Ocean', *Journal of Fish Diseases*, 42: 1107–18.

Waterhouse, A. M. et al. (2009) 'Jalview Version 2—A Multiple Sequence Alignment Editor and Analysis Workbench', *Bioinformatics*, 25: 1189–91.

Wessel, O. et al. (2017) 'Infection with Purified Piscine Orthoreovirus Demonstrates a Causal Relationship with Heart and Skeletal Muscle Inflammation in Atlantic Salmon', *PLoS One*, 12: e0183781.

—— et al. 2018. 'PRV1: Virulence differences in Atlantic salmon' in *8th International Symposium on Aquatic Animal Health*, Charlottetown, PEI, Canada.

Zhang, Y. et al. (2019) 'High-Load Reovirus Infections Do Not Imply Physiological Impairment in Salmon', *Frontiers in Physiology*, 10: 114.