

# Geography and host species shape the evolutionary dynamics of U genogroup infectious hematopoietic necrosis virus

Allison Black,<sup>1,3</sup> Rachel Breyta,<sup>2,4</sup> Trevor Bedford,<sup>3,†</sup> and Gael Kurath<sup>2,4,\*</sup>

<sup>1</sup>Department of Epidemiology, University of Washington, Box 357236, Health Sciences Building F-262, Seattle, WA 98195, USA, <sup>2</sup>U.S. Geological Survey, Western Fisheries Research Center, 6505 NE 65th St, Seattle, WA 98115, <sup>3</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA and <sup>4</sup>Cary Institute for Ecosystems Studies, 2801 Sharon Turnpike, Millbrook, NY 12545, USA

\*Corresponding author: E-mail: gkurath@usgs.gov

<sup>†</sup><http://orcid.org/0000-0002-4039-5794>

## Abstract

Infectious hematopoietic necrosis virus (IHNV) is a negative-sense RNA virus that infects wild and cultured salmonids throughout the Pacific Coastal United States and Canada, from California to Alaska. Although infection of adult fish is usually asymptomatic, juvenile infections can result in high mortality events that impact salmon hatchery programs and commercial aquaculture. We used epidemiological case data and genetic sequence data from a 303 nt portion of the viral glycoprotein gene to study the evolutionary dynamics of U genogroup IHNV in the Pacific Northwestern United States from 1971 to 2013. We identified 114 unique genotypes among 1,219 U genogroup IHNV isolates representing 619 virus detection events. We found evidence for two previously unidentified, broad subgroups within the U genogroup, which we designated 'UC' and 'UP'. Epidemiologic records indicated that UP viruses were detected more frequently in sockeye salmon (*Oncorhynchus nerka*) and in coastal waters of Washington and Oregon, whereas UC viruses were detected primarily in Chinook salmon (*Oncorhynchus tshawytscha*) and steelhead trout (*Oncorhynchus mykiss*) in the Columbia River Basin, which is a large, complex watershed extending throughout much of interior Washington, Oregon, and Idaho. These findings were supported by phylogenetic analysis and by  $F_{ST}$ . Ancestral state reconstruction indicated that early UC viruses in the Columbia River Basin initially infected sockeye salmon but then emerged via host shifts into Chinook salmon and steelhead trout sometime during the 1980s. We postulate that the development of these subgroups within U genogroup was driven by selection pressure for viral adaptation to Chinook salmon and steelhead trout within the Columbia River Basin.

**Key words:** infectious hematopoietic necrosis virus; molecular epidemiology; fish virus evolution; rhabdovirus; phylogenetic.

## 1. Introduction

Infectious hematopoietic necrosis virus (IHNV) is a negative-sense RNA virus (family Rhabdoviridae, genus *Novirhabdoviridae*) that infects all six species of Pacific salmonids (*Oncorhynchus* sp.) as well as Atlantic salmon (*Salmo salar*) and some species of trout (Wolf 1988; Bootland and Leong 1999). In the Pacific Northwestern United States, IHNV is a significant burden on cultured Pacific salmon (Breyta et al. 2016). IHNV

infection in adult fish generally results in asymptomatic infections (Wolf 1988; Bootland and Leong 1999). However infections in juvenile fish can cause disease events with up to 90% mortality (Bootland and Leong 1999; Groberg 1983a,b; LaPatra et al. 1993; Breyta et al. 2016), severely impacting salmon hatcheries, enhancement programs, and commercial aquaculture facilities.

IHNV epidemics were first reported in the 1950s in juvenile sockeye salmon (*Oncorhynchus nerka*) raised at hatcheries in

Washington and Oregon (Rucker et al. 1953; Guenther et al. 1959; Wingfield et al. 1969), although the virus was not widespread during this time nor in the following two decades (Amend and Wood 1972; Mulcahy et al. 1980; Pilcher and Fryer 1980). In the early 1980s, IHNV emerged in salmonids in the Columbia River Basin (Groberg and Fryer 1983; Groberg 1983b), a large and complex watershed draining an area of 668,000 km<sup>2</sup> (Washington State Department of Ecology 2015) throughout much of Washington, Oregon, and Idaho. This watershed is also the most productive salmon watershed in the contiguous USA. Currently, North American IHNV is endemic throughout the west coast of North America from Alaska to California, including watersheds inland in Idaho (Wolf 1988).

In an effort to understand IHNV dynamics in the field, the Western Fisheries Research Center (WFRC) runs a genetic surveillance program to genotype and archive select IHNV field isolates. Genotyping data based on 303 nt of the IHNV glycoprotein gene (referred to as the variable midG region) has been used for epidemiological studies, to measure viral genetic diversity in the field, and to explore the IHNV phylogeny. Phylogenetic analysis of North American IHNV by Kurath et al. (2003) supports three distinct genogroups, referred to as U, M, and L. In this earlier work, the U genogroup had no distinct subgroups, in contrast to the phylogenetic substructure observed in the M and L genogroups. Although these genogroups are supported with high confidence (Kurath et al. 2003; Breyta et al. 2013), the overall genetic diversity of North American IHNV is relatively low, with maximum 8.6% nt diversity in the midG region (Kurath et al. 2003). The genogroups correlate strongly with distinct geographic ranges. Of all three North American IHNV genogroups, U genogroup has the largest geographic range yet has minimal genetic diversity and infects primarily sockeye salmon in Alaska, British Columbia, and Washington (Emmenegger et al. 2000; Kurath et al. 2003). However, in the Columbia River Basin, U genogroup virus also infects Chinook salmon (*O. tshawytscha*) and steelhead trout (*O. mykiss*) (Garver et al. 2003). Although Chinook salmon and steelhead trout occur throughout the entire range of IHNV, infection of these hosts with U genogroup IHNV outside the Columbia River Basin is infrequent.

We hypothesized that the non-trivial presence of U genogroup IHNV in Chinook salmon in the Columbia River Basin might correlate with viral population structure if host species and geographic range isolated distinct populations of viruses. To explore the evolutionary dynamics of U genogroup viruses we conducted a retrospective study in Washington, Oregon and Idaho, a sub-region of the U genogroup geographic range. We examined IHNV detection records and corresponding genetic sequence data for 1,219 isolates of U genogroup IHNV collected from 1971 through 2013. We classified isolates into two geographic ranges, the first representing all sites within the Columbia River Basin and the second representing smaller coastal watersheds that drain directly into the Pacific Ocean along the coasts of Washington (including Puget Sound) and Oregon. Within these geographic ranges the majority of U genogroup IHNV infections occurred in three species of Pacific salmonids: Chinook salmon (*O. tshawytscha*), steelhead and rainbow trout (ocean-going and resident freshwater forms of *O. mykiss*), and sockeye and kokanee salmon (ocean-going and land-locked freshwater forms of *O. nerka*). For simplicity we refer to these three species hereafter by the common names Chinook salmon, sockeye salmon, and steelhead trout, except where distinction of kokanee salmon or rainbow trout are relevant. We focus on these three host species as they are the most abundant species that occur naturally and are most commonly

cultured in our study area. Additionally, we use the following terms in our phylogenetic descriptions: genogroup refers to the three major clades of North American IHNV (U, M, and L). A subgroup is a well-supported clade within a genogroup. Finally, clade refers to smaller groups of related viruses within subgroups. We looked at patterns of U genogroup IHNV infection by geography and host species, and used viral genetic sequence data to explore the U genogroup phylogeny, test for population structure, and determine how the viral population size has changed over time.

## 2. Methods

### 2.1 Collection of virus isolates

IHNV isolates were originally provided by national (US Fish and Wildlife Service), state (Washington and Oregon Departments of Fish and Wildlife, Idaho Department of Fish and Game), and tribal (Northwest Indian Fisheries Commission) fish health agencies that perform active surveillance for IHNV at fish hatcheries, fish farms, or wild fish sites. Adult fish samples are collected annually for nearly all salmonid populations in the study area through routine surveillance of fish returning to the hatchery to spawn; generally, ovarian fluid or kidney and/or spleen tissues from 60 to 100 adult fish will be sampled (Thoesen 1994). In contrast, juvenile fish are usually only sampled, most often as whole fish, if they show mortality or visible signs suggesting infection or disease. Fish health diagnostic laboratories culture virus in fish cell lines using standardized procedures (Thoesen 1994) and a subset of these isolates, chosen at the discretion of the fish health diagnostic laboratory, is submitted to WFRC, along with accompanying case data, for genetic typing.

### 2.2 Generation of sequence data

Genetic typing is done by sequencing a 303 nt variable region (midG) of the glycoprotein gene corresponding to nt 686–988 (GenBank accession U50401) (Emmenegger and Kurath 2002; Kurath et al. 2003). Viral RNA extraction from field isolates, reverse-transcription PCR, and sequencing PCR were performed according to protocols presented in Breyta et al. (2013). We generated consensus sequences using Sequencher version 4.9 (Gene Codes Corporation) and compared them against all currently known IHNV midG genotypes. If a virus isolate sequence differed from all known IHNV midG genotypes, then a new USD number was assigned using the format mG### followed by genogroup U, M, or L (e.g., mG001U) (Breyta et al. 2013).

For sequences from 26 of the 619 virus detection events (see below for definition of ‘events’) more than 1 nt was detected at the same site in the sense and antisense strands. Due to pooling of tissue samples from up to five fish (Thoesen 1994) we could not determine whether these heterogeneous sites represented infections of individual fish with different genotypes or a single fish co-infected with multiple genotypes. Heterogeneous sites were coded using IUPAC ambiguity codes and labeled with the two component midG USDs or ‘HH’ if more than one heterogeneous site was observed.

### 2.3 Dataset creation and virus detection event coding

The IHNV sequencing database at WFRC contains midG sequences for over 3,000 field isolates representing the entire spatial range of IHNV in North America (Kurath 2012; Breyta and Kurath, unpublished data). However, the majority of routine surveillance comes from Washington, Oregon, and Idaho, thus we limited our dataset to field isolates from those three states

to improve reporting consistency. Given our specific interest in evolutionary dynamics of U genogroup IHNv, we limited the dataset to U genogroup isolates, despite frequent detection of M genogroup viruses in this region. The dataset includes isolates collected between 1971 and 2013 (inclusive). Records without data on sampling year, fish species, or viral genotype were removed. These steps resulted in a dataset of 1,219 U genogroup viral field isolates.

We corrected for sampling bias in the numbers of isolates received via event coding (Breyta et al. 2013). Although isolates represent the primary unit of IHNv genetic surveillance, individual IHNv detection events may be represented by variable numbers of isolates (one to thirty here). We corrected for over-sampling by limiting our final dataset to one isolate record per detection event, as described in Breyta et al. (2013). To accomplish this, unique fish populations were defined according to year of sampling, sampling site, fish species, fish age (adult, yearling, or juvenile), and seasonal run timing. Fish populations were considered unique if they differed from another population by any of these variables, and an event was recorded if IHNv was detected in a distinct fish population. Separate events were also recorded if a different viral genotype was detected in an already-positive fish population.

The final dataset featured 619 U genogroup IHNv events occurring in Washington, Oregon, or Idaho from 1971 to 2013 for which there were 114 unique genotypes. Although this events dataset corrects for oversampling, some sampling bias likely remains due to under-reporting of events, under-sampling of events, and variation in the numbers of isolates a fish health diagnostic laboratory chooses to submit for genotyping.

## 2.4 Data and code availability

MidG sequences and epidemiological data for over 2,400 viral field isolates are maintained in the publicly available MEAP-IHNv database at <http://gis.nacse.org/ihnv/>. A FASTA format file for the 619 events used in the analyses presented here, and code for data extraction, preparation, and figure creation, are available at <https://github.com/alliblk/ihnv>.

## 2.5 Estimation of nucleotide diversity

Mean intrapopulation nucleotide diversity ( $\pi$ ) was calculated using events data in the PopGenome 2.1.6 package for R (Pfeifer et al. 2014). Maximum nucleotide diversity was estimated in MEGA version 6.06 (Tamura et al. 2013).

## 2.6 Maximum likelihood phylogenetic inference and test of molecular clocks

Maximum likelihood trees were inferred in RAxML version 8.2.3 (Stamatakis 2014) using GTR as the evolutionary model for flexibility and an M genogroup virus (mG139M) as the outgroup. Using these trees we tested for a molecular clock in Path-O-Gen version 1.4 (Rambaut 2010).

## 2.7 Bayesian coalescent phylogenetic analysis

A dated coalescent phylogeny for all 619 U genogroup events was inferred using BEAST version 1.8.2 (Drummond et al. 2012) using a GTR substitution model with gamma-distributed rate variation between sites, a strict molecular clock across all branches of the tree, a Bayesian skyline demographic prior (Drummond et al. 2005), and the continuous-time Markov chain rate reference prior (Ferreira and Suchard 2008) on the evolutionary rate. We selected these

priors because they were relatively non-informative and did not over-parameterize the model. All other priors were set to default. The year that an event occurred was used to date each event in the analysis. Markov chain Monte Carlo (MCMC) was run for 200 million steps logging trees every 100,000 steps after allowing 40 million steps for burn-in. MCMC convergence was assessed in Tracer version 1.6.0 (Rambaut et al. 2014). Estimates were taken from runs with effective sample sizes (ESSs) of 100 or greater. A maximum clade credibility (MCC) tree was inferred from the 1,600 sampled posterior trees. Coalescent phylogenies for UC ( $n = 475$  events) and UP ( $n = 144$  events) subgroups were reconstructed separately using the same priors described earlier to enable comparison of population dynamics. These trees were treated as independent draws from the posterior space of trees when subsequently used in discrete trait analyses.

## 2.8 Discrete trait analysis

Using the samples of 1,600 UC trees and 1,600 UP trees from the phylogenetic analyses described earlier we modeled the phylogenetic history of geographic range and host species. We treated geographic range and host species as discrete evolutionary traits (Lemey et al. 2009) and assumed a non-reversible transition matrix (Edwards et al. 2011). We coded geographic range as coastal watersheds or Columbia River Basin. The species trait was coded as: Chinook salmon (*O. tshawytscha*), sockeye/kokanee salmon (*O. nerka*), steelhead/rainbow trout (*O. mykiss*) or non-dominant, representing events in coho salmon, chum salmon, or Atlantic salmon. We used an exponential distribution with mean of 1 as our prior distribution for transition rates between trait states.

We sampled posterior trees, transition rates and ancestral states via MCMC. We ran the MCMC for 20 million steps with trees and transition rates sampled every 5,000 steps after allowing 2 million steps for burn-in. All transition rate estimates had ESSs of 3,000 or greater. MCC trees were inferred from 3,600 sampled posterior trees and colored according to the inferred geographic and host species reconstructions.

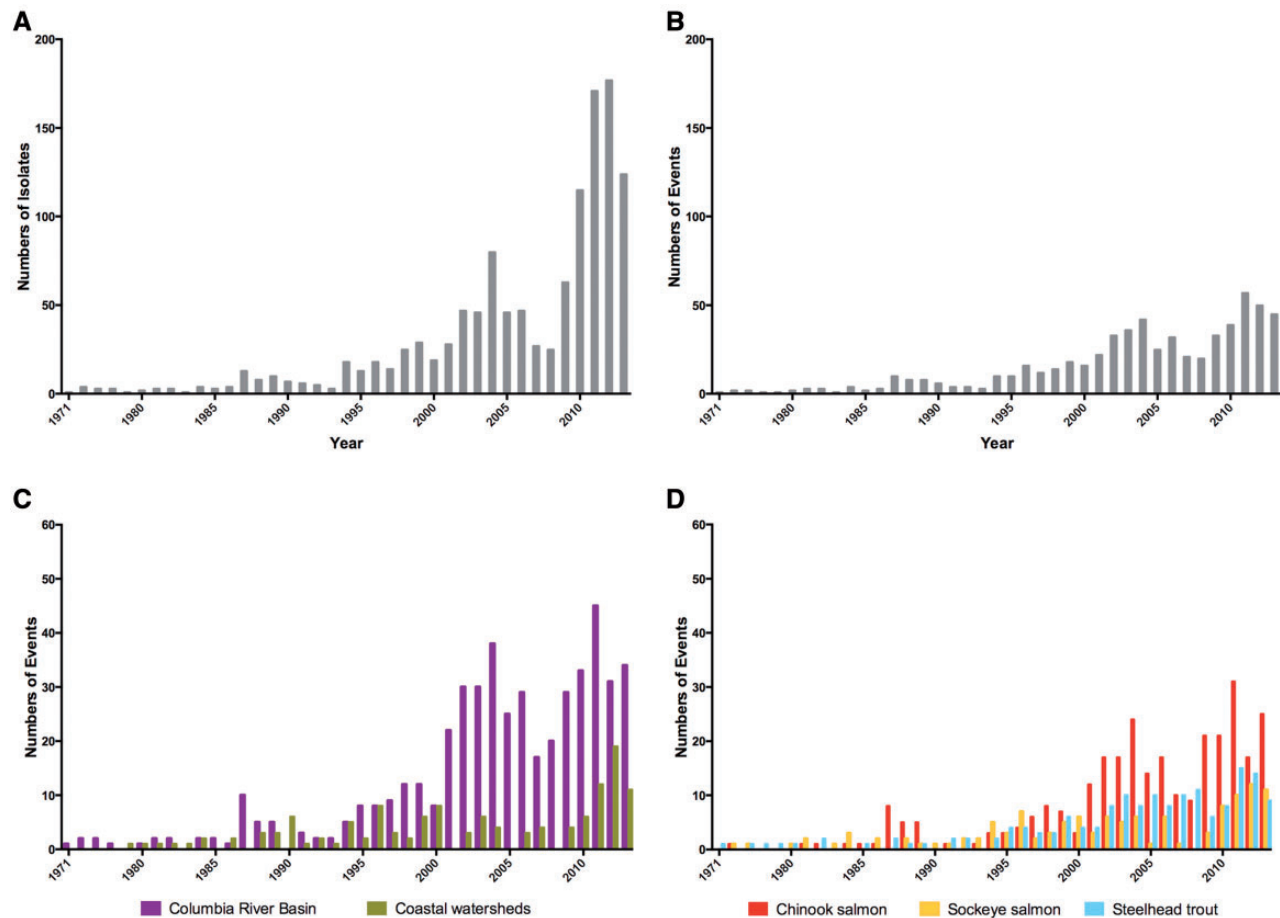
## 2.9 Tests of population subdivision

We assessed the degree of population structure using tests for compartmentalization implemented in HyPhy version 2.2.3 (Kosakovsky Pond et al. 2005). Pairwise distances between sequences were estimated using maximum likelihood methods, fixed rates across branches, and JC69 as the substitution matrix (Jukes and Cantor 1969).  $F_{ST}$  was calculated according to Hudson et al. (1992). These analyses were performed using the isolates data and the events data (Black 2015), although we present only analyses performed on events here.

# 3. Results

## 3.1 Isolate and event distribution over time, geography and host

Figure 1A shows the number of virus isolates that were genotyped ( $n = 1,219$ ) by year over the study period. The distribution was strongly skewed toward more recent isolates with a sharp increase beginning in 2009. To eliminate bias due to higher numbers of IHNv isolates being genotyped per positive fish cohort in recent years, the data were normalized via event coding, a process where we ensure that a virus detection event is represented by only one isolate record (see 'Methods' section), resulting in a distribution of 619 events over time (Fig. 1B). Although there were fewer events than isolates, the events distribution is



**Figure 1.** (A) Distribution of numbers of genotyped U genogroup IHN virus isolates, by year, collected between 1971 and 2013 ( $n = 1219$ ). (B) Distribution of numbers of genotyped U genogroup IHN virus detection events that occurred between 1971 and 2013 ( $n = 619$ ). (C) Numbers of genotyped U genogroup events that occurred by year in either the Columbia River Basin (purple) or in coastal watersheds (green) between 1971 and 2013 ( $n = 619$ ). (D) Numbers of genotyped U genogroup events that occurred by year in Chinook salmon (red), sockeye salmon (including kokanee) (yellow), or steelhead trout (including rainbow trout) (blue) between 1971 and 2013 ( $n = 581$ ).

similar to the isolates distribution, with a rise in events in the mid-1990s to a peak in 2004, and a second rise in events beginning in 2009 and peaking in 2011–2012. As expected, the rise in virus detection events after 2009 is not as sharp as the rise in virus isolates after 2009 (compare Fig. 1A and B).

These distributions do not directly represent changes in virus incidence and may be influenced by three factors: the actual incidence of IHN virus infection in fish in the field; the amount of testing conducted to detect virus in fish (virology surveillance intensity); and the proportion of isolates from virus-positive cohorts that were genotyped (genetic surveillance intensity). Prior to 1980 virology surveillance was opportunistic, and few isolates were available for genotyping. However, since the mid 1980s, virology surveillance programs have been reasonably consistent. Genetic surveillance has been reasonably consistent since the mid-1990s to present, providing genotype data associated with 50–65% of all IHN virus detection events (Breyta and Kurath, unpublished data). Therefore, it is likely that the fluctuations in numbers of virus events in the latter part of the distribution (Fig. 1B) do reflect, at least in part, changes in actual virus incidence.

In terms of the two defined geographic ranges, greater numbers of genotyped events occurred in the Columbia River Basin than in coastal watersheds, and that difference increased over time (Fig. 1C). The number of events in coastal watersheds was relatively constant between 1990 and 2010 and then rose during

2011–13, while Columbia River Basin events increased earlier, from 2002 to present. By host, genotyped events occurred at a similar frequency in Chinook salmon, sockeye salmon, and steelhead trout prior to the early 2000s (Fig. 1D), after which the numbers of events in Chinook salmon outpaced those in sockeye salmon and in steelhead trout. Sockeye salmon generally had the fewest numbers of events per year. The peaks in genotyped events in Chinook salmon correlated with peaks in events in the Columbia River Basin.

### 3.2 Descriptive epidemiology of U genogroup IHN virus in the Pacific Northwest

During the study period there were 488 events at sites in the Columbia River Basin and 131 events in coastal watersheds (Table 1). Of the 619 events, 295 (47.7%) occurred in Chinook salmon, 163 (26.3%) occurred in steelhead/rainbow trout, and 123 (19.9%) occurred in sockeye/kokanee salmon (Table 1). Close to 80% of events occurred in adult fish (Table 1), indicating that most of the data were collected as routine annual surveillance rather than outbreak surveillance. Previous studies report sockeye as the primary host for U genogroup IHN virus throughout coastal Washington, British Columbia, and Alaska (Emmenegger et al. 2000; Emmenegger and Kurath 2002; Kurath et al. 2003), and many factors support the hypothesis that ancestral U



**Table 1.** Descriptive analysis of U genogroup events occurring in Washington, Oregon, and Idaho between 1971 and 2013.

	All U Events (n = 619)	UC Events (n = 475)	UP Events (n = 144)
<b>Geography</b>			
Coastal watersheds (37 sites)	131 (21.2%)	15 (3.2%)	116 (80.6%)
Columbia River Basin (77 sites)	488 (78.8%)	460 (96.8%)	28 (19.4%)
<b>Host</b>			
Sockeye Salmon	92 (14.9%)	9 (1.9%)	83 (57.6%)
Kokanee Salmon <sup>1</sup>	31 (5.0%)	22 (4.6%)	9 (6.3%)
Chinook Salmon	295 (47.7%)	282 (59.4%)	13 (9.0%)
Steelhead Trout	153 (24.7%)	141 (29.7%)	12 (8.3%)
Rainbow Trout <sup>2</sup>	10 (1.6%)	6 (1.3%)	4 (2.8%)
Coho Salmon	20 (3.2%)	12 (2.5%)	8 (5.6%)
Chum Salmon	14 (2.3%)	3 (0.6%)	11 (7.6%)
Atlantic Salmon	4 (0.6%)	N/A	4 (2.8%)
<b>Lifestage</b>			
Juveniles	119 (19.2%)	86 (18.1%)	33 (22.9%)
Adults	491 (79.3%)	386 (81.3%)	105 (72.9%)
Unknown Age	9 (1.5%)	3 (0.6%)	6 (4.2%)

<sup>1</sup>Kokanee salmon are a freshwater form of sockeye salmon, *O. nerka*.

<sup>2</sup>Rainbow trout are a freshwater form of steelhead trout, *O. mykiss*.

genogroup virus was specifically associated with sockeye salmon. Although sockeye salmon in the Columbia River Basin also carry almost exclusively U genogroup IHN, they are less abundant than the Chinook salmon and steelhead trout that also serve as major U genogroup hosts in the basin (Garver et al. 2003). Thus our restriction of the study region to the three-state area including the Columbia River Basin is a factor in the relatively low number of events in sockeye salmon in this study.

### 3.3 Distribution of viral genotypes suggests geographic division

The increase in available data after the mid-1990s (Fig. 1A and B) allowed us to detect landscape-level patterns of circulating viral genotypes. We initially noticed geographic divisions in IHN detections for mG001U and mG002U, two viral genotypes that historically caused the greatest number of U genogroup IHN events. Despite the wide range of hatcheries where these two genotypes were detected, mG001U was detected almost exclusively within the Columbia River Basin (334 out of 342 isolates), whereas mG002U was detected almost exclusively at non-Columbia River Basin sites, in watersheds that drained directly to the Pacific coast or Puget Sound (sixty-eight out of seventy-three isolates) (Black 2015). Similar spatial separation was noted for other genotypes that represented multiple isolates (data not shown).

### 3.4 Phylogenetic analysis of genotypes indicates preliminary support for subgroups within the U genogroup

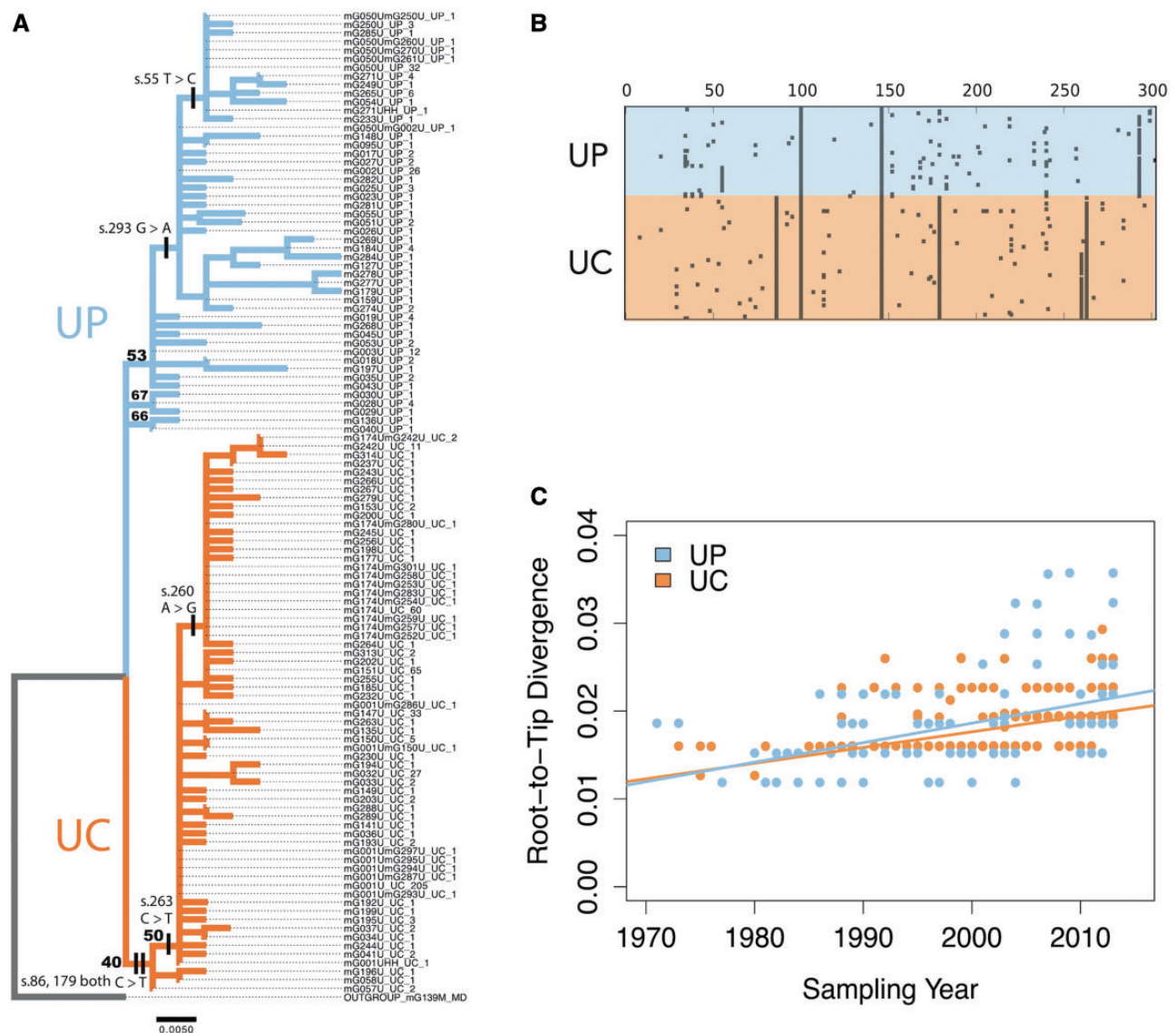
Our dataset included 114 unique midG genotypes. A maximum likelihood phylogeny of these genotypes (Fig. 2A) indicated two broad subgroups, which we designated 'UC' (U Columbia River Basin) and 'UP' (U Pacific). Bootstrap support values on the maximum likelihood tree were low: forty for the most basal UC node and fifty-three, sixty-seven, and sixty-six on UP basal nodes. We attribute the low bootstrap support to the low number of subgroup defining nucleotide differences. UC and UP subgroups

are best separated by four sites (Fig. 2B) which may not be captured in bootstrap replicates. We also noticed 'wandering' (Bryant 2003) of eight taxa between UC and UP subgroups. Although the majority of the tree maintained a consistent topology, the movement of these taxa prevented global consistency. We found wandering taxa by visually inspecting a posterior subsample of eighty coalescent trees inferred in BEAST. Although removing wandering taxa resulted in slightly higher bootstrap support for the UC and UP basal nodes (fifty-one and sixty, respectively) (Supplementary Fig. 1), the support remains low, likely because bootstrap resampling continued to miss informative sites. Given this low support, we did not define UC and UP as taxonomic units, since other partitions would have been possible. Rather, we used the observed partitioning as an operational definition for putative UP and UC subgroups in order to further explore viral population dynamics.

Sequences from the 114 genotypes were aligned and inspected for shared nucleotide mutations using the oldest genotype detected during the study (mG018U) as the reference sequence (Fig. 2B). The alignment supported our operational subgroup division; genotypes within each subgroup were separated without exception by two mutations (Fig. 2B): genotypes within the UP subgroup had C at positions 86 and 179, while UC genotypes had T. A third mutation at position 263 separated UP genotypes from all but two UC genotypes (mG057U and mG058U). A fourth mutation at position 293 separated most UP genotypes (A) from UC genotypes and some basal UP genotypes (G). Two other mutations described smaller clades within the UC and UP subgroups. Further clade substructure was apparent with mutations at sites 55 and 260. We observed 83 nt changes across the alignment which resulted in thirty-six amino acid changes, thus 43% of observed mutations were non-synonymous.

### 3.5 Genetic diversity and molecular clock of U genogroup IHN in the Pacific Northwest

As 114 viral genotypes were found across 619 U genogroup events, certain viral genotypes were detected multiple times. To



**Figure 2.** Analysis of UC and UP subgroups. (A) Maximum likelihood tree of all 114 U genotypes detected within the study geographic range between 1971 and 2013. Blue branches indicate genotypes within the UP subgroup and orange branches indicate genotypes within the UC subgroup. Numbers of events attributable to each genotype are indicated as the last argument in the taxon name. Bolded numbers indicate bootstrap support values based on 1,000 replicates. Black bars represent clade defining mutations, whereby site number (s. no.) corresponds to the alignment figure in panel B, and the nucleotide substitution is stated. (B) Alignment of 113 UP and UC genotypes against the earliest genotype in the dataset, a UP subgroup virus. Coloring indicates which subgroup each genotype belongs to. Genotypes are arranged chronologically within their respective subgroups from earliest to most recent detection. (C) Root-to-tip divergence (phylogeny not shown) plotted against sampling date for all 619 U genogroup events occurring within the study region between 1971 and 2013. Orange indicates UC events and blue indicates UP events. Linear regression lines approximate the observed rate of evolution over the study time period. For UP events: slope =  $2.22 \times 10^{-4}$  subs per site per year,  $R^2 = 0.2093$ ,  $P < 0.001$ . For UC events: slope =  $1.79 \times 10^{-4}$  subs per site per year,  $R^2 = 0.2184$ ,  $P < 0.001$ .

appropriately represent the diversity of the virus in the field, we examined nucleotide diversity of U genogroup events rather than distance between unique genotypes. The maximum pairwise distance between any two U genotypes was 13 nt (4.29%) and  $\pi$  for all 619 U events was 0.0052 nt differences per site (nt/site). For UP events  $\pi$  was 0.0066 nt/site, and for UC events  $\pi$  was 0.0023 nt/site. The maximum pairwise distance between UP viruses was 12 and 6 nt for UC viruses. Thus although UC events were more abundant than UP events, they were overall less genetically diverse.

### 3.6 UC and UP subgroup viruses are detected in different geographic ranges and host species

Epidemiologic case data was used to quantify patterns of detection stratified by host and geography. Just over 80% of the events attributable to UP subgroup viruses occurred at sites within the coastal watersheds geography (Table 1). UC subgroup viruses showed an even sharper geographic constraint; 97% of UC events occurred at sites within the Columbia River Basin (Table 1). We found that 64% of UP events occurred in sockeye

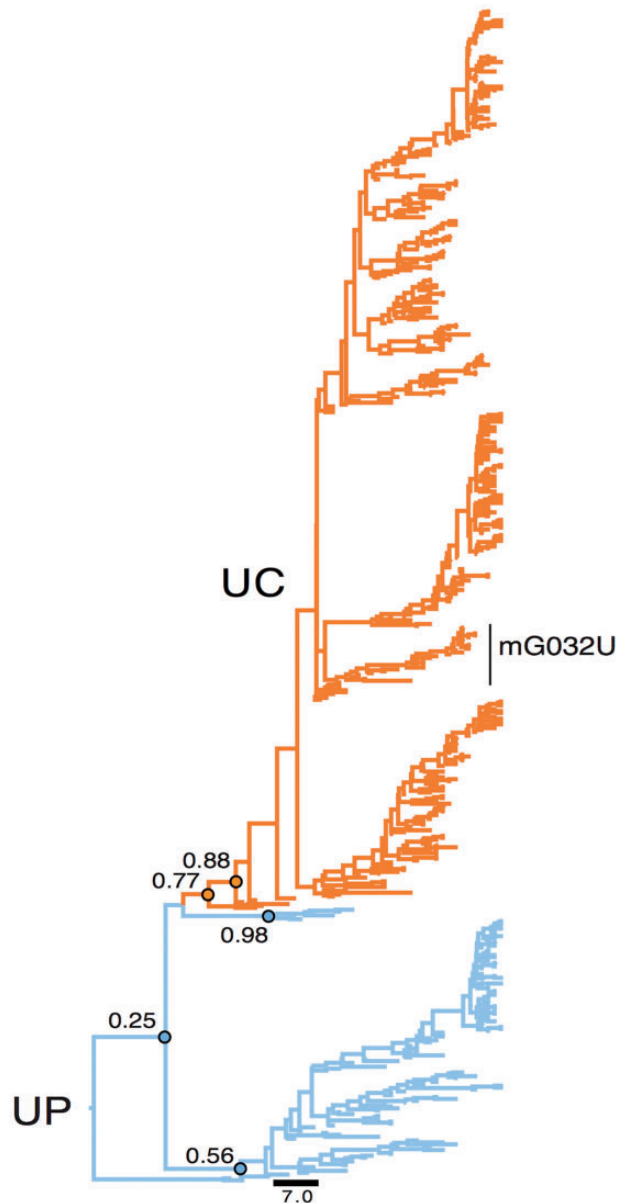


Figure 3. Coalescent phylogenetic tree showing U genogroup topology. UC viruses are shown in orange and UP viruses are shown in blue. Posterior support values are given at key nodes. Scale represents number of substitutions per site per year.

or kokanee salmon (both *O. nerka*) with the balance distributed among multiple hosts, each with less than 10% of UP events (Table 1). Thus the majority of UP events occurred in the species traditionally associated with U genogroup infection. UC events demonstrated different host preferences; 59% of UC events occurred in Chinook salmon and 31% of UC events occurred in steelhead or rainbow trout (both *O. mykiss*), with only 6.5% of UC events occurring in sockeye or kokanee salmon (Table 1).

### 3.7 Coalescent phylogenetic analysis further supports UC and UP subgroups

Coalescent phylogenetic analysis of the 619 U genogroup events indicated the same subgroup partitioning as the maximum likelihood tree of the 114 unique genotypes. Notably, the coalescent

analysis demonstrated greater support for the UC subgroup. UC viruses grouped together with posterior support of 0.77 at the most basal node and 0.88 for the next most basal node (Fig. 3). For all U virus events analyzed here the estimated rate of evolution was  $3.01 \times 10^{-4}$  subs per site per year (95% HPD:  $2.07 \times 10^{-4} - 4.04 \times 10^{-4}$ ), while UP subgroup viruses evolved at a rate of  $5.42 \times 10^{-4}$  subs per site per year (95% HPD:  $3.37 \times 10^{-4} - 7.75 \times 10^{-4}$ ) and UC viruses evolved at a rate of  $2.76 \times 10^{-4}$  subs per site per year (95% HPD:  $1.72 \times 10^{-4} - 3.86 \times 10^{-4}$ ). The most recent common ancestor of all U virus events in our dataset circulated in ~1,948 (95% HPD: 1927–65), and in approximately 1966 for UC genotype viruses (95% HPD: 1957–73). Because UP viruses are not monophyletic, we do not report a UP-specific estimate of the time to the most recent common ancestor.

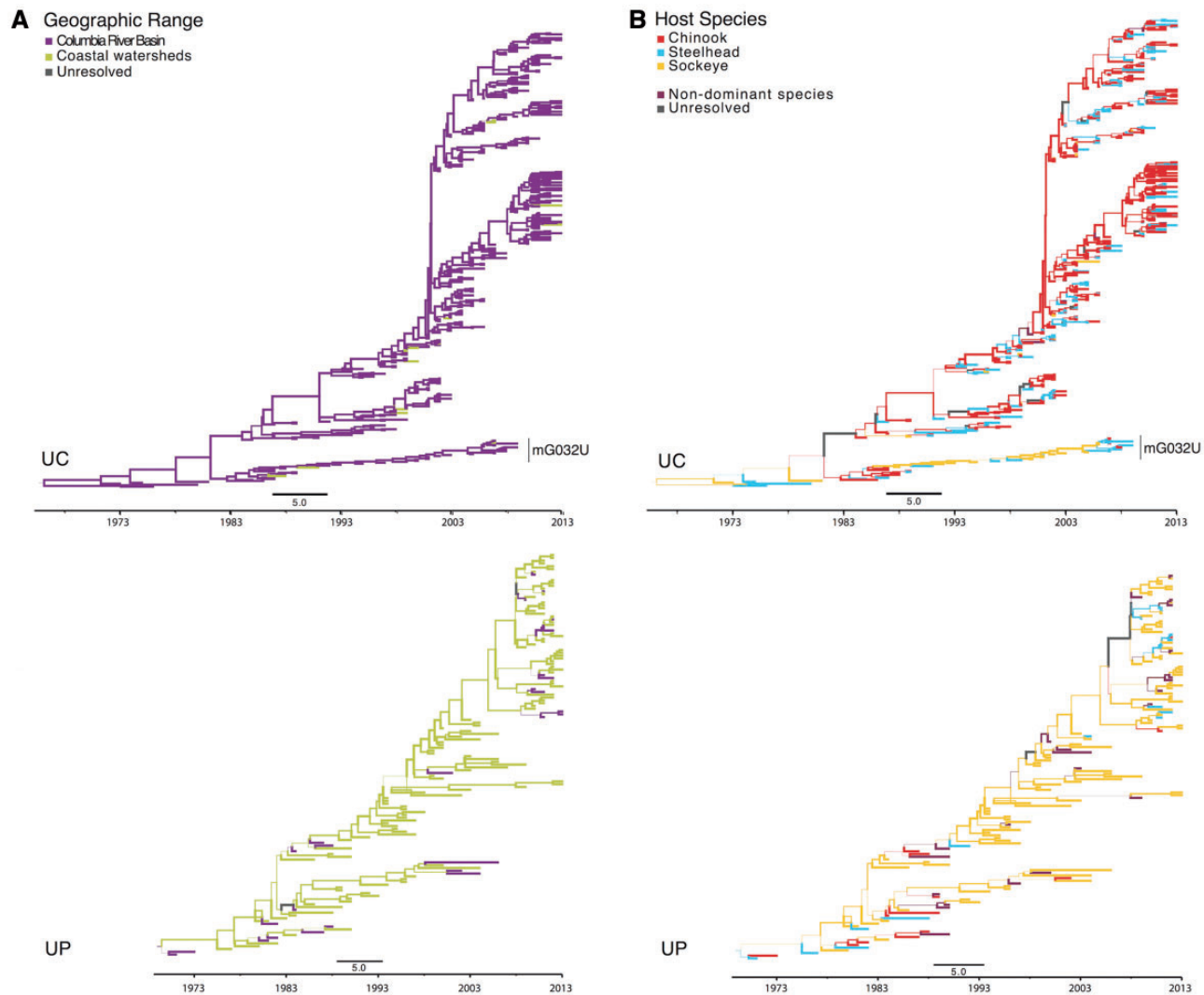
### 3.8 Ancestral state reconstruction of UC and UP subgroups indicates geographic and host structure

We reconstructed ancestral geographic and host states and mapped this information to separate UC and UP coalescent phylogenies (Fig. 4A and B). Inferred ancestral geographic states reiterated patterns in the epidemiologic case data; UC events occurred primarily in the Columbia River Basin (purple) and UP events occurred mainly in coastal watersheds (green), although we observed exceptions to this general trend (Fig. 4A). It was unclear whether the coastal watersheds or the Columbia River Basin was the ancestral geographic range: UC viruses were firmly established in the Columbia River Basin and UP viruses were largely in coastal watersheds by the time our study began in 1971. This association remained similarly strong in an analysis of subsampled data representing more equitable numbers of events from the two geographic ranges (Supplementary Fig. 3) (see Supplemental Methods for a description of the subsampling scheme).

We found that UC viruses moved from coastal watersheds to the Columbia River Basin at a rate of 1.9 events per lineage per year (95% HPD: 0.051–4.5), roughly twenty-four times more than UC viruses transitioned from the Columbia River Basin to coastal watersheds (rate:  $7.9 \times 10^{-2}$  events per lineage per year, 95% HPD:  $1.2 \times 10^{-3} - 2.0 \times 10^{-1}$ ). The opposite relationship was seen in UP viruses, which transitioned from the Columbia River Basin to coastal watersheds at a rate of 1.6 events per lineage per year (95% HPD: 0.036–3.8), roughly four times the rate of UP transitions from coastal watersheds into the Columbia River Basin (rate: 0.41 events per lineage per year, 95% HPD: 0.0088–1.0) (Supplementary Fig. 2). Phylogeographic analysis of the equitably subsampled data showed similar transition rates as inferred from the full data and overall our conclusions did not change between the analyses. All estimates from the subsampled data fell within the 95% HPDs from the full data (Supplementary Fig. 4), and the rate estimates from the analysis on the full data and the subsampled data were highly correlated (Pearson correlation coefficient was 1.0 for both UC and UP rates).

The ancestral host analysis indicated that many UC events prior to the 1980s and most UP events occurred in sockeye salmon (yellow) (Fig. 4B). Most UC events after the 1980s occurred in Chinook salmon (red) and to a lesser degree in steelhead trout (light blue) (Fig. 4B). These findings did not change with more equitable subsampling (Supplementary Fig. 3). Although relatively high numbers of UC events occurred in steelhead trout, the phylogeny did not show specific clades within UC that were associated more with infections in trout. Rather, events in steelhead trout were scattered throughout the UC subgroup





**Figure 4.** UC and UP coalescent phylogenetic trees showing ancestral geographic range and ancestral host species. (A) Separate UC and UP trees showing inferred ancestral geographic states. Inferred detection in the Columbia River Basin is indicated in purple and inferred detection in coastal watersheds is indicated in green. Rarely, the ancestral state had the same probability of occurring in both ranges; these branches are indicated as unresolved and are colored grey. Line thickness indicates the probability of being in the colored state; thicker lines indicate higher probabilities. (B) Separate UC and UP trees showing inferred ancestral host states. Red denotes Chinook salmon, yellow denotes sockeye/kokanee salmon, and light blue denotes steelhead/rainbow trout. Only rarely ancestral states had equal probability of detection in multiple host species. These are indicated as unresolved and colored grey. Non-dominant hosts for IHNV such as coho salmon, chum salmon, and Atlantic salmon are all indicated by maroon. As in panel A, line thickness correlates with probability of being in a state.

with no evidence of further structure (Fig. 4B). Within UC one clade was an exception to the general pattern of Chinook salmon and steelhead trout hosts after the 1980s. Between 1988 and 2006, viral genotype mG032U (indicated in Figs. 3 and 4A and B) caused events mainly in kokanee salmon, the landlocked form of sockeye salmon. Despite this clear signal of long-term infection in *O. nerka*, the mG032U clade was detected almost exclusively in the Columbia River Basin (Fig. 4A). The presence of mG032U in kokanee salmon is likely due to founder effects as this kokanee population inhabits a reservoir upstream of an unpassable dam that blocks these fish from mixing with migratory populations of Chinook salmon and steelhead trout (Anderson et al. 2000).

Within UC subgroup viruses, the two highest host state transition rates were from steelhead trout to Chinook salmon (4.8 events per lineage per year, 95% HPD: 2.2–8.0) and from Chinook salmon to steelhead trout (2.6 events per lineage per year, 95%

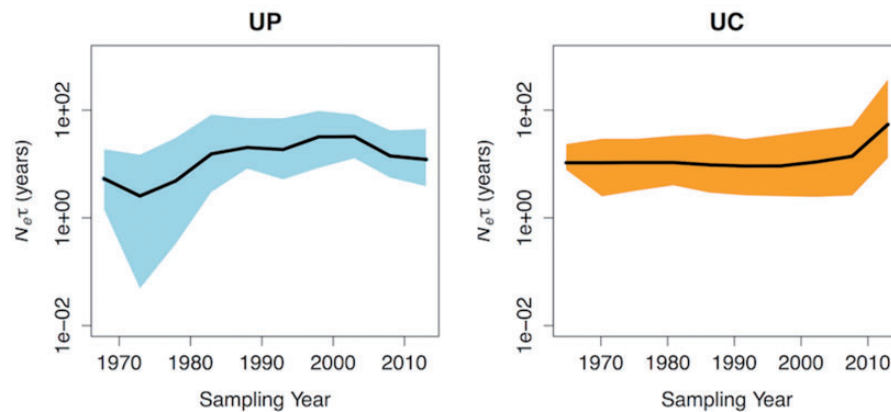
HPD: 1.1–4.4). The highest UP transition rates were from non-dominant species into sockeye salmon at 2.7 events per lineage per year (95% HPD: 0.40–5.4) and from steelhead trout into sockeye salmon at 2.3 events per lineage per year (95% HPD:  $2.3 \times 10^{-3}$ –5.1) (Supplementary Fig. 2).

Host transition rates from the phylogeographic analysis of the subsampled data all fell within the 95% HPDs of estimates from the full data (Supplementary Figure 4). Again the rates from the analysis on the subsampled data were highly correlated with rates inferred from the full data, supporting our original conclusions (Pearson's correlation coefficient was 0.93 for UC host transition rates and 0.97 for UP host transition rates).

### 3.9 Population dynamics of UC and UP

We compared how scaled effective population size  $N_e\tau$  for UP and UC subgroups changed over time. Both UP and UC





**Figure 5.** Changes in effective population size over time for UP and UC subgroup viruses. The median effective population size is indicated in black and the shaded area represents the 95% HPD around the estimate.

subgroups had median posterior estimates of  $N_e\tau$  that appeared relatively constant (Fig. 5). For UP viruses the median values rose between the early 1970–80s but with a broad posterior density for the earlier years. Toward the end of the study the median value of  $N_e\tau$  for UP declined, while for UC the median value of  $N_e\tau$  increased, although these apparent trends are not statistically significantly different from zero. The median posterior estimate of  $N_e\tau$  for UC ranged from 9.2 to 54.2 years over the study period. For UP, median  $N_e\tau$  ranged between 2.6 and 32.3 years.

We estimated serial interval  $\tau$  from experimental data on fish-to-fish transmission of IHN in juvenile rainbow trout (freshwater *O. mykiss*). Upon IHN infection, roughly 3 days are required before viral shedding appears, followed by 2–4 days for IHN prevalence in the susceptible fish to peak (Ogut and Reno 2004). We therefore considered  $\tau$  to be 6 days (3 days of latency after infection and on average 3 more days for the secondary infection to occur). Using this value of  $\tau$  we estimated that the effective population size of UP viruses ranged from 163 infections to 2,019 infections and UC viruses ranged between 575 infections and 3,388 within the study area over the study period.

### 3.10 Population structure due to geographic range and host species of detection

Given that UC viruses were detected primarily in the Columbia River Basin and UP viruses were detected primarily in coastal watersheds, we hypothesized that geographic separation played a role in subdividing U genogroup viruses. Events occurred across 114 sampling locations (Fig. 6A). The Columbia River Basin partition included 488 (78.8%) events and the coastal watersheds partition included 131 (21.2%). The distribution of these events across their geographic range is shown in Figure 6C. Pairwise comparisons of the midG sequence data yield an  $F_{ST}$  statistic of 0.379 (95% CI: 0.334–0.422,  $P < 0.001$ ), suggesting strong evidence of population structure due to geographic range (Table 2).

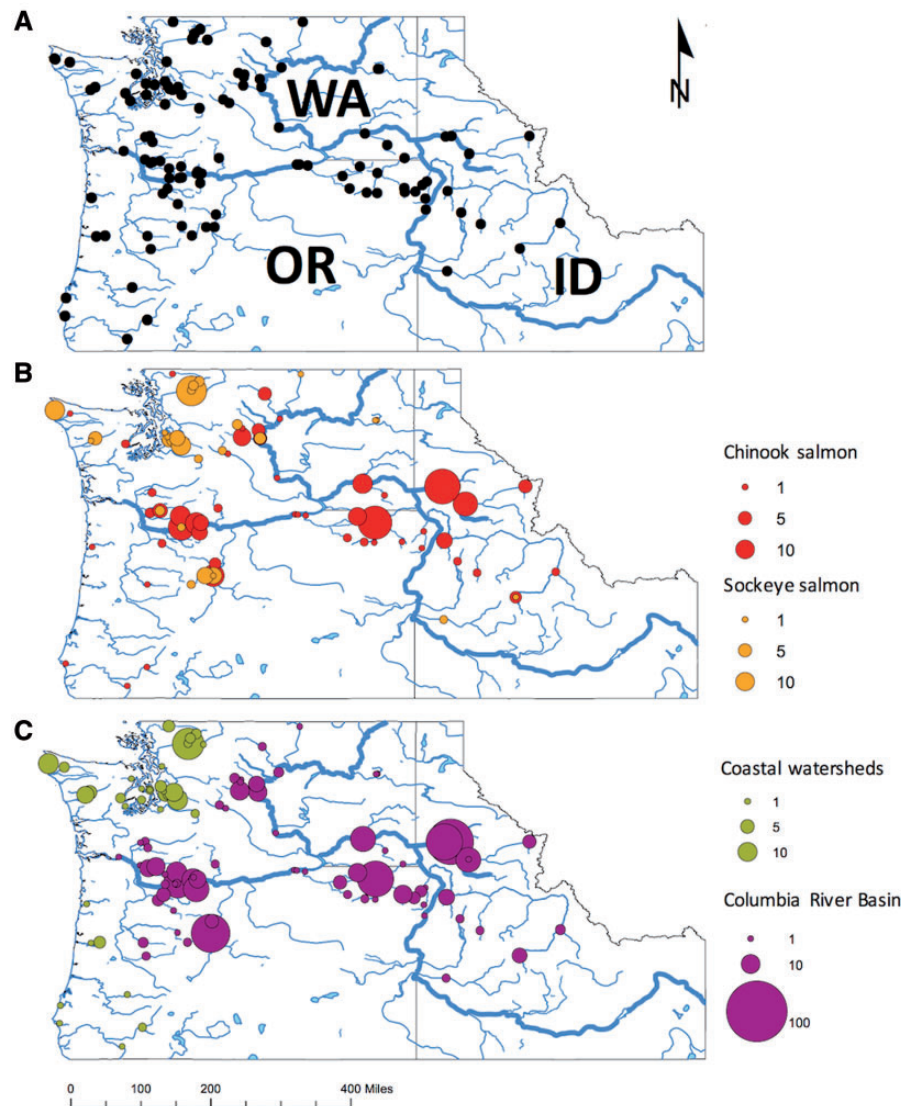
Because UP viruses were detected primarily in sockeye salmon and UC viruses were detected primarily in Chinook salmon, we hypothesized that viral adaptation to different hosts might also structure U genogroup viruses. Although UC viruses were often detected in steelhead trout, the phylogeny did not support a distinction between UC viruses detected in Chinook salmon versus steelhead trout (Fig. 4B). Therefore, we consider it unlikely that UC viruses were adapting separately to both Chinook salmon and steelhead trout. Tests for population structure

indicated substantially less structure between steelhead trout and Chinook salmon ( $F_{ST} = 0.047$ , 95% CI: 0.015–0.087,  $P = 0.001$ ) (Black 2015). Therefore, the by-host analysis was done by comparing sockeye salmon to Chinook salmon only. Here we found significant population structure, with  $F_{ST}$  of 0.406 (95% CI: 0.358–0.458,  $P < 0.001$ ), strongly supporting host species as a factor that subdivides U genogroup viruses (Table 2). Although host species and geographic range are not independent ( $P < 0.0001$ ,  $\chi^2$  test of independence; see Supplementary Table 2 for contingency table), logistic modeling of the subgroup of an infection (UC or UP) was best predicted by an additive model including both host species and geographic range as predictors (Supplementary Table 3).

## 4. Discussion

We analyzed the evolutionary dynamics of U genogroup IHN within the Pacific Northwestern United States from 1971 to 2013 using molecular sequence data and epidemiologic surveillance data. We found evidence for two previously unidentified, broad subgroups within the U genogroup, which we designated ‘UC’ and ‘UP’. UP viruses were detected more frequently in sockeye salmon and in coastal watersheds, whereas UC viruses were detected primarily in Chinook salmon and steelhead trout in the Columbia River Basin. This is consistent with previous analyses of the host associations of U genogroup viruses from throughout the North American geographic range (Kurath et al. 2003), and within the Columbia River Basin (Garver et al. 2003). These initial patterns were supported by ancestral state reconstructions of UC and UP phylogenetic trees (Fig. 4A and B) and  $F_{ST}$  inferred for geographic and host partitions (Table 2).

UC viruses were established in the Columbia River Basin prior to the beginning of the study period. After their appearance in the Columbia River Basin these viruses moved from their ancestral hosts, sockeye salmon, into Chinook salmon and steelhead trout. Given this emergence we expected to see changes in viral population size consistent with epidemic growth as seen with landscape expansion of rabies virus (Lemey et al. 2010). However, there was no indication of large changes in viral population size around the time of emergence (Fig. 5). This constancy could be attributable to low phylogenetic signal. Our work is based on relatively short sequences (303 nt) with relatively little genetic diversity, which may inhibit our ability to accurately infer changes in population size. Selective sweeps might also reduce viral diversity and keep effective



**Figure 6.** Maps indicating sampling locations and event counts. (A) Location of sample collection sites where one or more events occurred within this study. WA, Washington; OR, Oregon; ID, Idaho. (B) Locations of all events that occurred in either sockeye salmon (including kokanee salmon) or in Chinook salmon. Circle scale indicates the number of events that occurred. This host delineation also represents the partitions used in the by-host analysis of population structure. (C) Locations of all events occurring in either the Columbia River Basin or in coastal watersheds. Circles scale according to numbers of events. These event counts correspond to the partitions defined for tests of population structure due to geographic subdivision.

population sizes relatively constant. Additionally, the entire geographic range of U genogroup extends with high prevalence into British Columbia and Alaska (Kurath et al. 2003; Meyers et al. 2003); incorporating further U events could change measures of population size.

We attribute the emergence of UC viruses in Chinook salmon and steelhead trout within the Columbia River Basin to a possible host adaptation. Unfortunately, we could not look for causal adaptive variation in such a small fraction of the genome with limited mutations. Only 43% of mutations in our alignment were non-synonymous, an indication of selective constraint that is consistent with sequencing a random region of a viral RNA genome (Pybus et al. 2007). However, the landscape patterns of IHN virus infections support the possible development of host specificity. For example, the low frequency of UP detections in Chinook salmon is not attributable to a lack of Chinook salmon in coastal watersheds where UP viruses predominate. Indeed, Chinook salmon are the most commonly cultured

species in both coastal watersheds and in the Columbia River Basin (Pacific States Marine Fisheries Commission 2015). Thus despite abundant populations of Chinook salmon in coastal watersheds, we did not see strong evidence of UP viruses emerging in this highly prevalent host. The lack of IHN virus detection in coastal Chinook salmon is not due to lack of surveillance; diagnostic data indicate similar surveillance intensity of returning coastal Chinook populations yet with fewer detections of IHN virus (Breyta and Kurath, unpublished data). Similarly, in Alaska IHN virus detections are nearly all in sockeye salmon, despite the co-occurrence and surveillance of Chinook salmon populations (Meyers, pers. comm.).

Geographic subdivision also contributes to viral population structure with implications for the timing of viral transmission. It has been proposed that the majority of IHN virus transmission occurs during the freshwater portions of the salmonid host life cycle, at spawning and after hatching (Bootland and Leong 1999). However previous molecular epidemiology studies found

Table 2. Detection of population structure by host and by geography.

<b>By Host: Chinook salmon and sockeye salmon<sup>3</sup></b>	
Mean interpopulation diversity	0.0161
Mean subpopulation diversity	0.00608
Metapopulation diversity	0.0102
$F_{ST}$ (95% CI), P-value	0.406 (0.358–0.458), $P < 0.001$
<b>By Geography: Columbia River Basin and coastal watersheds</b>	
Mean interpopulation diversity	0.0171
Mean subpopulation diversity	0.00606
Metapopulation diversity	0.00975
$F_{ST}$ (95% CI), P-value	0.379 (0.334–0.422), $P < 0.001$

<sup>3</sup>Sockeye includes both sockeye and kokanee salmon (both *O. nerka*)

identical genotypes of U virus in North America and eastern Russia, suggesting that IHN virus transmission may also occur in the marine environment (Rudakova et al. 2007), when Asian and North American salmonids co-mingle in the Alaskan gyre and the Bearing Sea gyre (Healy 1991, 366 Fig. 28). Ancestral state reconstructions indicated that by the late 1960s the two U subgroups were already partitioned into two large geographic regions. This geographic partitioning provides evidence that the majority of contacts resulting in transmission occur when hosts are separated into the Columbia River Basin or coastal watersheds. Additionally, the evolution of a subgroup specific to the Columbia River Basin provides the first clear example of watershed-specific IHN virus evolution, which would support the original hypothesis that a large component of viral transmission occurs in freshwater. Further work should focus on inferring whether freshwater transmission generally occurs in an upstream direction (correlating with transmission by returning adult populations) or in a downstream direction (correlating with transmission by out-migrating juvenile populations).

Given the presented data, we postulate that the development of two separate subgroups within U genogroup was driven by a unique selection pressure for viral adaptation to Chinook salmon and steelhead trout within the Columbia River Basin. We cannot rule out the possibility that the UC subgroup developed separately from the UP subgroup due to founder effects. However, it is interesting to consider a speculative alternative hypothesis based on documented shifts in relative abundance of IHN virus host species in the Columbia River Basin. Historically, all salmonid populations, including sockeye salmon, were more numerous in the Columbia River Basin than they are today (Fulton 1970) due to major population declines that occurred in the early 1900s. As hatchery programs developed in the Columbia River Basin starting in the 1950s, severe IHN epidemics occurred specifically in sockeye salmon (Rucker et al. 1953; Guenther et al. 1959; Wingfield et al. 1969), contributing to management decisions to culture Chinook salmon and steelhead trout in substantially larger numbers than sockeye salmon (Pacific States Marine Fisheries Commission 2015). Although purely speculative, the resulting change in host species composition of the Columbia River Basin away from the ancestral host of U genogroup IHN virus, sockeye salmon, to previously less susceptible but more abundant hosts, Chinook salmon and steelhead trout, may have provided a selection pressure for U genogroup IHN virus to adapt to alternative host(s), resulting in the evolution of the UC subgroup. It is possible that this same adaptation was not seen in coastal watersheds because coastal hatchery programs continued to culture sockeye salmon in

higher numbers (Pacific States Marine Fisheries Commission 2015).

In summary, this work has revealed a previously unrecognized subgroup within the U genogroup of IHN virus that occurs in Chinook salmon and steelhead trout in the Columbia River Basin. Human impact on salmonid populations and the ecosystems of the Columbia River Basin has been extensively described (Weitkamp 1994; Lichatowich 2001; Dauble et al. 2003). We present evidence that both geographic and host factors were major contributors to the evolution of the UC subgroup of IHN virus. We further suggest one potential hypothesis for how anthropogenic influence may have profoundly impacted the evolution of this aquatic virus, but this and other possible explanations for divergence of the UC subgroup remain to be investigated.

## Funding

AB was supported by the US Fish and Wildlife Service FONS program project IFHC-USGS IA2012 and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256082. The work was supported by the US Geological Survey Western Fisheries Research Center and the Washington Cooperative Fish and Wildlife Research Unit at the University of Washington. Financial support for this analysis was also provided through USDA grant 2012-67015-19960 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Disease program, and National Institutes of Health (NIH) U54 GM111274. Any mention of trade names is for descriptive purposes only and does not imply U.S. government endorsement.

## Data availability

Data available on the Molecular Epidemiology of Aquatic Pathogens—Infectious Hematopoietic Necrosis Virus (MEAP-IHN virus) database at <http://gis.nacse.org/ihnv/>.

## Supplementary data

Supplementary data are available at Virus Evolution online.

Conflict of interest: None declared.

## Acknowledgements

We thank Bill Batts for contributing sequences to the analysis, Jim Winton for essential historical knowledge and



Debbie Reusser for technical assistance with GIS. We are grateful to the continued participation of the fish health professionals throughout the Pacific Northwest who contribute their time in providing virus isolates and epidemiologic information. Virus isolates were provided by: B. Stewart, M. House, J. Bertolini, C. Olsen, and J. Gleck of the Northwest Indian Fisheries Commission; H. M. Engelking, J. Kaufman, W. Groberg, G. Claire, and S. Onjukka of the Oregon Department of Fish and Wildlife; J. Thomas of the Washington Department of Fish and Wildlife; S. Landon, D. Munson, and K. Johnson of the Idaho Department of Fish and Game; R. Brunson, S. Gutenberger, T. London, K. Clemens, S. Lutz, S. Mumford, C. Patterson, M. Blair, and C. Samson of the US Fish and Wildlife Service.

## References

- Amend, D. F., and Wood, J. F. (1972) 'Survey for Infectious Hematopoietic Necrosis (IHN) Virus in Washington Salmon', *Progressive Fish-Culturist*, 34/3: 143–7.
- Anderson, E. D. et al. (2000) 'Molecular Epidemiology Reveals Emergence of a Virulent Infectious Hematopoietic Necrosis (IHN) Virus Strain in Wild Salmon and Its Transmission to Hatchery Fish', *Journal of Aquatic Animal Health*, 12/2: 85–99.
- Black, A. (2015) Geographic and Host Factors Shape the Evolution of a Newly Recognized Subgroup within the U Genogroup of the Fish Rhabdovirus IHNV. University of Washington. <<https://digital.lib.washington.edu/researchworks/handle/1773/34160>>.
- Bootland, L. M., and Leong, J. C. (1999) 'Infectious Hematopoietic Necrosis Virus', in P.T.K., Woo and D.W., Bruno(eds.) *Fish Diseases and Disorders*, pp. 57–112. Wallingford UK: CAB International.
- Breyta, R. et al. (2013) 'Emergence of MD Type Infectious Hematopoietic Necrosis Virus in Washington State Coastal Steelhead Trout', *Diseases of Aquatic Organisms*, 104/3: 179–95.
- et al. (2016) 'Successful Mitigation of Viral Disease Based on a Delayed Exposure Rearing Strategy at a Large-Scale Steelhead Trout Conservation Hatchery', *Aquaculture*, 450: 213–24.
- Bryant, D. (2003) 'A Classification of Consensus Methods for Phylogenetics', *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 61: 163–84.
- Dauble, D. D. et al. (2003) 'Impacts of the Columbia River Hydroelectric System on Main-Stem Habitats of Fall Chinook Salmon', *North American Journal of Fisheries Management*, 23/3: 641–59.
- Drummond, A. J. et al. (2005) 'Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences', *Molecular Biology and Evolution*, 22/5: 1185–92.
- et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29/8: 1969–73.
- Edwards, C. J. et al. (2011) 'Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline', *Current Biology*, 21/15: 1251–8.
- Emmenegger, E. J., and Kurath, G. (2002) 'Genetic Characterization of Infectious Hematopoietic Necrosis Virus of Coastal Salmonid Stocks in Washington State', *Journal of Aquatic Animal Health*, 14/1: 25–34.
- , and — et al. (2000) 'Genetic Diversity and Epidemiology of Infectious Hematopoietic Necrosis Virus in Alaska', *Diseases of Aquatic Organisms*, 40/3: 163–76.
- Ferreira, M. A. R., and Suchard, M. A. (2008) 'Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains', *Canadian Journal of Statistics*, 36/3: 355–68.
- Fulton, L. A. (1970) 'Spawning Areas and Abundance of Steelhead Trout and Coho, Sockeye, and Chum Salmon in the Columbia River Basin - Past and Present'. Washington D.C.
- Garver, K. A., Troyer, R. M., and Kurath, G. (2003) 'Two Distinct Phylogenetic Clades of Infectious Hematopoietic Necrosis Virus Overlap within the Columbia River Basin', *Diseases of Aquatic Organisms*, 55/3: 187–203.
- Groberg, W. J. (1983a) 'Priority Research Needs Concerning Fish Viruses Prevalent among Columbia River Basin Salmonids', in G. C., Leong and T. Y., Barila (eds) *Proceedings of a Workshop on Viral Diseases of Salmonid Fishes in the Columbia River Basin, Special Publication*, pp. 159–67. Portland, OR: Bonneville Power Administration.
- (1983b) 'The Status of Viral Fish Diseases in the Columbia River Basin', in J. C., Leong and T., Barila (eds) *Proceedings of a Workshop on Viral Diseases of Salmonid Fishes in the Columbia River Basin, Special Publication*. Portland, OR: Bonneville Power Administration.
- and Fryer, J. L. (1983) 'Increased Occurrence of Infectious Hematopoietic Necrosis Virus in Fish at Columbia River Basin Hatcheries, 1980–1982', Technical paper no. 6620. Corvallis, OR.
- Guenther, R. W., Watson, S. W., and Rucker, R. R. (1959) 'Etymology of Sockeye Salmon 'Virus' Disease', *U.S. Fish and Wildlife Service Special Scientific Report on Fisheries*, 296: 1–10.
- Healy, M. C. (1991) 'Life History of Chinook Salmon', in C., Groot and L., Margolis (eds) *Pacific Salmon Life Histories*, pp. 311–94. Vancouver, BC: University of British Columbia Press.
- Hudson, R. R., Boos, D. D., and Kaplan, N. L. (1992) 'A Statistical Test for Detecting Geographic Subdivision', *Molecular Biology and Evolution*, 9/1: 138–51 <<http://mbe.oxfordjournals.org/content/9/1/138.short>>.
- Jukes, T. H., and Cantor, C. R. (1969) 'Evolution of Protein Molecules', in H. N., Munro (ed.) *Mammalian Protein Metabolism*, pp. 21–132. New York: Academic Press.
- Kosakovsky Pond, S. L. et al. (2005) 'HyPhy: Hypothesis Testing Using Phylogenies', *Bioinformatics*, 21/5: 676–9.
- Kurath, G. et al. (2003) 'Phylogeography of Infectious Haematopoietic Necrosis Virus in North America', *Journal of General Virology*, 84/4: 803–14.
- (2012) 'An Online Database for IHN Virus in Pacific Salmonid fish—MEAP-IHNV', *U.S. Geological Survey Fact Sheet 2012-3027*. <<http://pubs.usgs.gov/fs/2012/3027/>>.
- LaPatra, S. E. et al. (1993) 'Early Life Stage Survival and Susceptibility of Brook Trout, Coho Salmon, Rainbow Trout, and Their Reciprocal Hybrids to Infectious Hematopoietic Necrosis Virus', *Journal of Aquatic Animal Health*, 5/4: 270–74.
- Lemey, P. et al. (2010) 'Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time', *Molecular Biology and Evolution*, 27/8: 1877–85.
- et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5/9: e1000520.
- Lichatowich, J. (2001) *Salmon Without Rivers: A History of the Pacific Salmon Crisis*. Island Press, Washington.
- Meyers, T. R. et al. (2003) 'Infectious Hematopoietic Necrosis Virus (IHNV) in Alaskan Sockeye Salmon Culture from 1973 to 2000: Annual Virus Prevalences and Titers in Broostocks Compared with Juvenile Losses', *Journal of Aquatic Animal Health*, 15/1: 21–30.
- Mulcahy, D. M. et al. (1980). *The Occurrence and Distribution of Salmonid Viruses in Oregon*. Technical paper No. 5504, Oregon Agricultural Experiment Station, Corvallis, OR, ORESU-T-80-004.

- Ogut, H., and Reno, P. W. (2004) 'Early Kinetics of Infectious Hematopoietic Necrosis Virus (IHNV) Infection in Rainbow Trout', *Journal of Aquatic Animal Health*, 16/3: 152–60.
- Pacific States Marine Fisheries Commission. (2015) 'Regional Mark Information System', <<http://www.rmmpc.org/>>.
- Pfeifer, B. et al. (2014) 'PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R', *Molecular Biology and Evolution*, 31/7: 1929–36.
- Pilcher, K. S., and Fryer, J. L. (1980) 'The Viral Diseases of Fish, a Review through 1978. Part I, Diseases of Proven Viral Etiology. CRC', *Critical Review of Microbiology*, 7: 287–364
- Pybus, O. G. et al. (2007) 'Phylogenetic Evidence for Deleterious Mutation Load in RNA Viruses and Its Contribution to Viral Evolution', *Molecular Biology and Evolution*, 24/3: 845–52.
- Rambaut, A. (2010) 'Path-O-Gen Version 1.4', <http://tree.bio.ed.ac.uk/software/pathogen/>.
- et al. (2014) 'Tracer 1.6', <<http://beast.bio.ed.ac.uk/Tracer>>.
- Rucker, R. R. et al. (1953) 'A Contagious Disease of Sockeye Salmon Possibly of Virus Origin', *US Fish and Wildlife Service Fisheries Bulletin*, 54: 35–46.
- Rudakova, S. L., Kurath, G., and Bochkova, E. V. (2007) 'Occurrence and Genetic Typing of Infectious Hematopoietic Necrosis Virus in Kamchatka, Russia', *Diseases of Aquatic Organisms*, 75/1: 1–11
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies', *Bioinformatics*, 30: 1312–3.
- Tamura, K. et al. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, 30/12: 2725–9.
- Thoese, J.C., ed. (1994) *Suggested Procedures for the Detection and Identification of Certain Finfish and Shellfish Pathogens*. 4th ed. Bethesda, MD: Fish Health Section, American Fisheries Society.
- Washington State Department of Ecology. (2015) 'Columbia River Facts and Maps', <<http://www.ecy.wa.gov/Programs/wr/cwp/cwpfactmap.html>>.
- Weitkamp, L. A. (1994) 'A Review of the Effects of Dams on the Columbia River Estuarine Environment, with Special Reference to Salmonids', <[http://www.nwfsc.noaa.gov/assets/26/6867\\_03112012\\_231538\\_Weitkamp.1994b.pdf](http://www.nwfsc.noaa.gov/assets/26/6867_03112012_231538_Weitkamp.1994b.pdf)>.
- Wingfield, W. H., Fryer, J. L., and Pilcher, K. S. (1969) 'Properties of the Sockeye Salmon Virus (Oregon Strain)', *Proceedings of the Society for Experimental Biology and Medicine*, 130: 1055–9.
- Wolf, K. (1988) *Fish Viruses and Fish Viral Diseases*. Ithaca, New York, USA: Cornell University Press.