Vol. 78: 97–104, 2007
doi: 10.3354/dao01864

DISEASES OF AQUATIC ORGANISMS
Dis Aquat Org

Published December 13

Prediction model for sequence variation in the glycoprotein gene of infectious hematopoietic necrosis virus in California, USA

Garry O. Kelley[1,*], Rebecca Garabed[1], Adam Branscum[2], Andres Perez[1,3], Mark Thurmond[1]

[1]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, California 95616, USA

[2]Departments of Biostatistics, Statistics, and Epidemiology, University of Kentucky, Lexington, Kentucky 40536, USA

[3]CONICET-INTA Balcarce, Balcarce, 7620 Buenos Aires, Argentina

ABSTRACT: The influence of spatio-temporal factors on genetic variation of infectious hematopoietic necrosis virus (IHNV) is an active area of research. Using host-isolate pairs collected from 1966 to 2004 for 237 IHNV isolates from California and southern Oregon, we examined genetic variation of the mid-G gene of IHNV that could be quantified across times and geographic locations. Information hypothesized to influence genetic variation was environmental and/or fish host demographic factors, viz. location (inland or coastal), year of isolation, habitat (river, lake, or hatchery), the agent factors of subgroup (LI or LII) and serotype (1, 2, or 3), and the host factors of fish age (juvenile or adult), sex (male or female), and season of spawning run (spring, fall, late fall, winter). Inverse distance weighting (IDW) was performed to create isopleth maps of the genetic distances of each subgroup. IDW maps showed that more genetic divergence was predicted for isolates found inland (for both subgroups: LI and LII) than for coastal watershed isolates. A mixed-effect beta regression with a logit link function was used to seek associations between genetic distances and hypothesized explanatory factors. The model that best described genetic distance contained the factors of location, year of isolation, and the interaction between location and year. Our model suggests that genetic distance was greater for isolates collected from 1966 to 2004 at inland locations than for isolates found in coastal watersheds during the same years. The agreement between the IDW and beta regression analyses quantitatively supports our conclusion that, during this time period, more genetic variation existed within subgroup LII in inland watersheds than within coastal LI isolates.

KEY WORDS: Infectious hematopoietic necrosis virus · IHNV · Rhabdovirus · Inverse distance weighting · Mixed-effect beta regression · Logit link function

INTRODUCTION

Infectious hematopoietic necrosis virus (IHNV) (family *Rhabdoviridae*) is a fish pathogen that affects farmed and wild salmonid fish including sockeye (*Oncorhynchus nerka*), Chinook (*O. tshawytscha*), chum (*O. keta*) salmon and rainbow and steelhead trout (*O. mykiss*) (Williams & Amend 1976, Wolf 1988, Bootland & Leong 1999). IHNV causes one of the most devastating diseases in US commercial fish farms and in state and federal hatcheries where losses of fry or fingerlings during an outbreak may reach 100%.

IHNV is an enveloped virus with a negative-sense, single-stranded RNA genome approximately 11 000 nucleotides in length. IHNV has a non-segmented genome that encodes for 6 proteins, viz. the nucleocapsid (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), non-virion proteins (NV), and polymerase (L) (Kurath & Leong 1985, Morzunov et al. 1995). There is evidence for 3 major genogroups, referred to as U, M, and L, that are related by the G and NV genes, wherein each genetic group may be associated with a distinct geographic distribution (Nichol et al. 1995, Emmenegger et al. 2000, Troyer et al. 2000, Kurath et

*Email: gokelley@dfg.ca.gov

al. 2003). Viruses in the major genogroup U, which has been found in Alaska and through the coastal regions from British Columbia to Washington and Oregon, are believed to have evolved among sockeye salmon (Emmenegger et al. 2000, Emmenegger & Kurath 2002, Garver et al. 2003, Kurath et al. 2003). Viruses in the M genogroup (Troyer et al. 2000) have been found in rainbow trout in commercial farms in Idaho and in the Columbia River basin, USA (Garver et al. 2003). Isolates within the M genogroup are believed to have evolved into subgroups that are present in fish at many geographic locations (Troyer et al. 2000, Troyer & Kurath 2003, Kurath et al. 2003). Viruses in the L genogroup are found mainly in Chinook salmon in watersheds of northern California and southern Oregon (Ross et al. 1960, Parisot et al. 1965, Amend et al. 1969, Mulcahy et al. 1984, Kurath et al. 2003, Bendorf et al. in press, Kelley et al. 2007). Studies of virus isolates from northern California and southern Oregon fish hatcheries and waterways over a 34 yr period found a temporal trend in IHNV genetic divergence (Bendorf et al. in press, Kelley et al. 2007), as well as evidence of 2 major subgroups (LI and LII) of the L genogroup described by Kurath et al. (2003). Although Bendorf et al. (in press) and Kelley et al. (2007) reported evidence for a temporal trend in genetic divergence among California IHNV isolates, putative associations between host genetic variation and factors of the host or the environment have not been examined quantitatively. There are several factors that might explain in part the genetic diversity of IHNV isolates. The presence of 2 subgroups in the L genogroup may reflect selective constraints due to changes in the host environment. Environmental changes such as translocation of infected fish between hatcheries or through rivers could result in the appearance of distinct virus isolates evolving in certain river basins or in specific commercial hatcheries. Also, unknown host-related factors could favor selection of specific strains of the virus by individuals of a given fish species (Wingfield et al. 1970, Kurath et al. 2003), gender or age. Studies on other host-pathogen systems suggest ecological and host factors can influence pathogen selection and genetic divergence (Rodriguez et al. 1996, Lively & Dybdahl 2000). An understanding of the extent to which environmental or host factors could affect genetic change in these viruses would help provide insight into the basic biology of the host-virus relationship. Such information also could help guide management of fish populations in ways that would reduce the impact of disease caused by IHNV. For example, if clusters of similar virus isolates were identified in certain geographic regions, transplant of fish among regions with different isolates might be restricted to prevent emergence of new isolates in regions not previously known to have that strain of virus. Knowledge of seasonal, age- or gender-specific variations in the prevalence of IHNV could also aid in the selection of vaccine strategies aimed at controlling spread of the disease. Moreover, if IHNV evolution could be predicted for specific regions and locations, effectiveness of vaccines and diagnostic primers could be better anticipated. A fuller understanding of any macro geographic distributions of virus groups is prerequisite to studies specifically addressing host and environmental factors that might be driving virus evolution. The objective of this study was to examine genetic variation of the mid-G gene of IHNV that could be quantified and predicted from information on time and geographic location, using host-isolate pairs identified and collected from 1966 to 2004 for 237 IHNV isolates from California and southern Oregon (Nichol et al. 1995, Kurath et al. 2003, Bendorf et al. in press, Kelley et al. 2007).

MATERIALS AND METHODS

Viral sequence and host and demographic data. A 303 nucleotide region (referred to as the mid-G gene) for 237 IHNV isolates was obtained from GenBank (Table 1). All nucleotide sequence alignments and virus isolates representing the L genogroup from 1966 to 2004 are available upon request (Bendorf et al. in press, Kelley et al. 2007). Using Kimura's 2 parameter model (Kimura 1980), the genetic divergence within the L genogroup was estimated as the genetic distance from a presumed common ancestor to the L genogroup (Kelley et al. 2007), where genetic distance was measured as the proportion of nucleotides that differed between any 2 isolates. IHNV isolates and demographic and host-specific data were obtained for fish collected at 19 sites (between 1 and 90 fish site^{-1}, median = 4.5, mean = 11.9) (Kelley et al. 2007). In the 1990s, a freezer containing some IHNV samples failed and many samples collected prior to November of 1987 were lost (W. Cox pers. comm.). Consequently, the number of isolates available for study from fish sampled before 1988 was far fewer than for fish sampled subsequently.

Spatial distribution of the genetic distance. Inverse distance weighting (IDW, Isaaks & Srivastava 1989) was used to create isopleth maps of the genetic distances of each subgroup (LI and LII) of IHNV. The IDW technique predicts genetic distance (E) at any unsampled location (j) in the geographic area under study according to a weighted average of the genetic distance (O) across all sampling locations (i=1,...,n = 19) where nearby isolates are weighted more heavily in predicting E than distant isolates. Let geographic distance between locations i and j be denoted by d_{ij}.

Table 1. Features of infectious hematopoietic necrosis virus (IHNV) isolates used in this study. Isolation sites, as described by Kelley et al. (2007), were Trinity River Hatchery (TR), Hoopa Fish Rearing Facility (HO), Mad River Hatchery (MA), Rowdy Creek Fish Hatchery (RC), Camp Creek (CC), and Eel River (EE). In addition, virus isolates were obtained from 11 inland sites: Feather River (FR), Lake Oroville (OR), Nimbus Hatchery (NB), Merced River Hatchery (MC), Mokelumne River Hatchery (MK), Yuba River lower site (YL), Yuba River middle site (YM), Yuba River upper site (YU), Clear Creek (CK), Battle Creek (BC), and Coleman National Fish Hatchery (CL). Isolates were also obtained from 3 coastal watersheds in southern Oregon: Elk River Hatchery (ER), Sixes River (SX), and Rogue River (RR). Host fish species included Chinook (C), coho (Co), steelhead (S), and rainbow trout (R). a: not available

Sub-group	Location	Isolation site(s)	Host(s) (n)	Life stage (n)	Year(s)	GenBank accession no.
I	Coastal and inland	ER, FR	C (1)	na	1969, 1976	AY598415
I	Coastal and inland	CC, ER, FR, HO, MC, RC, RR, TR	C (26), S (3), Co (2), na (1)	Adult (26), yearling (4), fry (2)	1969–1998	AY598416
I	Inland	MK	S	Adult	1971	DQ910915
I	Inland	FR	C	na	1971	AY598417
I	Inland	NB	C	Adult	1974	DQ910908
I	Coastal	TR	C	Adult	1991	DQ910909
II	Inland	NB	C	Adult	1966	L40881
II	Coastal and Inland	CL, MA	S (1), C (1)	Adult (2)	1979, 1985	DQ910916
II	Inland	CL	C	Adult	1980	L40873
II	Coastal	TR	C	na	1987	DQ910917
II	Coastal	MC, SX	C (2)	Adult (1), fingerling (1)	1986, 1988	DQ910910
II	Coastal	EE	S	Adult	1988	DQ910921
II	Inland	BC, CL, FR, MC, NB	C (13), S (5)	Adult (17), juvenile (1)	1989–2003	AY598418
II	Inland	CL, FR, MC, MK, NB	C (22), S (2)	Adult (24)	1990–1999	AY598419
II	Inland	CL	C	Adult	1995	DQ910911
II	Inland	NB	C	Adult	1998	DQ910918
II	Inland	CL	C	Adult	2001	DQ910922
II	Inland	BC, CK	C (4)	Adult (4)	2003	DQ910919
II	Inland	BC, CK	C (3)	Adult (3)	2003	DQ910920
II	Inland	BC, CK, CL, FR, MK, NB	C (37), S (11)	Adult (19), juvenile (1), fry (1)	1996–2003	AY598420
II	Inland	FR, MC, MK, NB, OR, YL, YM, YU	C (70), S (9), R (1)	Adult (71), yearling (2), fingerling (6) fry (1)	1999–2004	AY598421
II	Inland	FR	C	Fingerling	2001	AY598422
II	Inland	FR	C	Fingerling	2002	DQ910912
II	Inland	FR, YL	C (8)	Adult (8)	2003	DQ910913
II	Inland	FR	S	Adult	2004	DQ910914

Then

$$E_j = \frac{\sum_{i=1}^{n} \frac{O_i}{d_{ij}^p}}{\sum_{i=1}^{n} \frac{1}{d_{ij}^p}}$$

where the power $p > 0$ modifies the relative influence of sites i on predicted locations j so that the relation is not necessarily proportional. We used the value of p that minimized the root-mean-squared prediction error at sampled locations. The values O_i are averages of the genetic distances of sampled isolates collected from 1966 to 2004 at each site. Specifically, a weighted average genetic distance of all isolates collected from site i over the study time period was used, namely

$$O_i = \frac{\sum_{t,j} w_{t,j} O_{i,j,t}}{\sum_{t,j} w_{t,j}}$$

where $O_{i,j,t}$ denotes the observed genetic distance of the jth isolate collected during year t from site i, and

$$w_{t,j} \equiv w_t = \frac{x_{t+1} - x_{t-1}}{2(n_t)N}$$

is a function of the year (x_t) of isolation, the number of isolates collected in year t (n_t) from site i and the number of years represented by the samples (N). Numerators for the chronologically first and last isolate at each location were ($x_{t+1} - x_t + 1$) and ($x_t - x_{t-1} + 1$), respectively. For example, if a specific location had one isolate from 1977, one isolate from 1989, and 3 isolates from 2002, their respective genetic distances would be given the following weights: 0.25, 0.48, 0.09, 0.09, and 0.09. Calculations and mapping were done using ArcGIS software (ArcMap™ 9.0, ESRI).

Prediction model. A mixed-effect beta regression analysis with a logit link function (which extends the beta regression model of Ferrari & Cribari-Neto (2004) for independent data to correlated data) was used to seek associations between genetic distances and putative explanatory variables. These variables were the environmental or demographic factors of location (inland or costal), year of isolation, and habitat (river, lake, or hatchery), the agent factors of subgroup (LI or LII), and serotype (1, 2, or 3), and the host factor of fish age (juvenile or adult) (Kelley et al. 2007). The model is a variation of linear regression wherein genetic distances are modeled according to beta distributions with means that depend on explanatory variables. A beta distribution is particularly suitable for modeling the distribution of genetic distances (which are measured as proportions) because it takes values between 0 and 1 and is very flexible (e.g. accommodating left or right skewed, and symmetric distributions). In addition, a random effect variable was used to account for a latent 'site effect'; specifically, site effects are clustering effects of isolates from the same sites that are expected to be genetically more similar than those from different sites. The random effects were modeled as independent from a normal $(0, \sigma^2)$ distribution.

Year-of-isolation was used as an explanatory variable in the model to account for an assumed background annual mutation rate of the virus for all isolates obtained after 1987. We grouped isolates collected before 1987 because there was a small number of available isolates sparsely distributed throughout the years 1966–1987. Therefore, a model was used that grouped all isolates collected prior to 1987 into year zero (i.e. the time period 1966 to 1987 was coded as 0, and years $1988, 1989, \ldots, 2004$ were coded as $1, 2, \ldots, 17$). The consequence of aggregating data collected prior to 1987 into one single temporal point is that the genetic variation cannot be estimated longitudinally before that date. This procedure is analogous to establishing a baseline to the genetic variation observed until 1986, which is given by the average of the values observed prior to 1987. Thus, associations estimated here are relative to the baseline given by the values observed prior to 1987. Location (i.e. inland or coastal) described the site at which the host was collected. Age and habitat were included as categorical variables, and an interaction term for year and location was included to examine for an effect of year on genetic distance that could be conditional on location, where an effect of year (or location) on genetic distance would depend on the specific location (or year). Because the rate of genetic change may not be linear, various transformations of time were considered including power ($year^2$), inverse (1/year), and exponential (e^{year}) transformations.

Inferences for regression coefficients for the putative explanatory variables (β) were derived from a Bayesian analysis, which was implemented using WinBUGS© software (WinBUGS© 1.4, Imperial College and Medical Research Council, UK). Independent prior distributions were placed on the regression coefficients, the precision parameter ($\tau = 1/\sigma^2$) of the random effects distribution, and the scale parameter of the beta distribution (ψ). Specifically, the β coefficients were modeled with normal (0, 1) priors and τ and ψ were modeled with gamma priors, namely $\tau \sim$ gamma (1.33, 0.1) and $\psi \sim$ gamma (1, 1). This prior structure resulted in induced priors for mean genetic distances that were diffuse and, as a result, inferences were primarily influenced by the data.

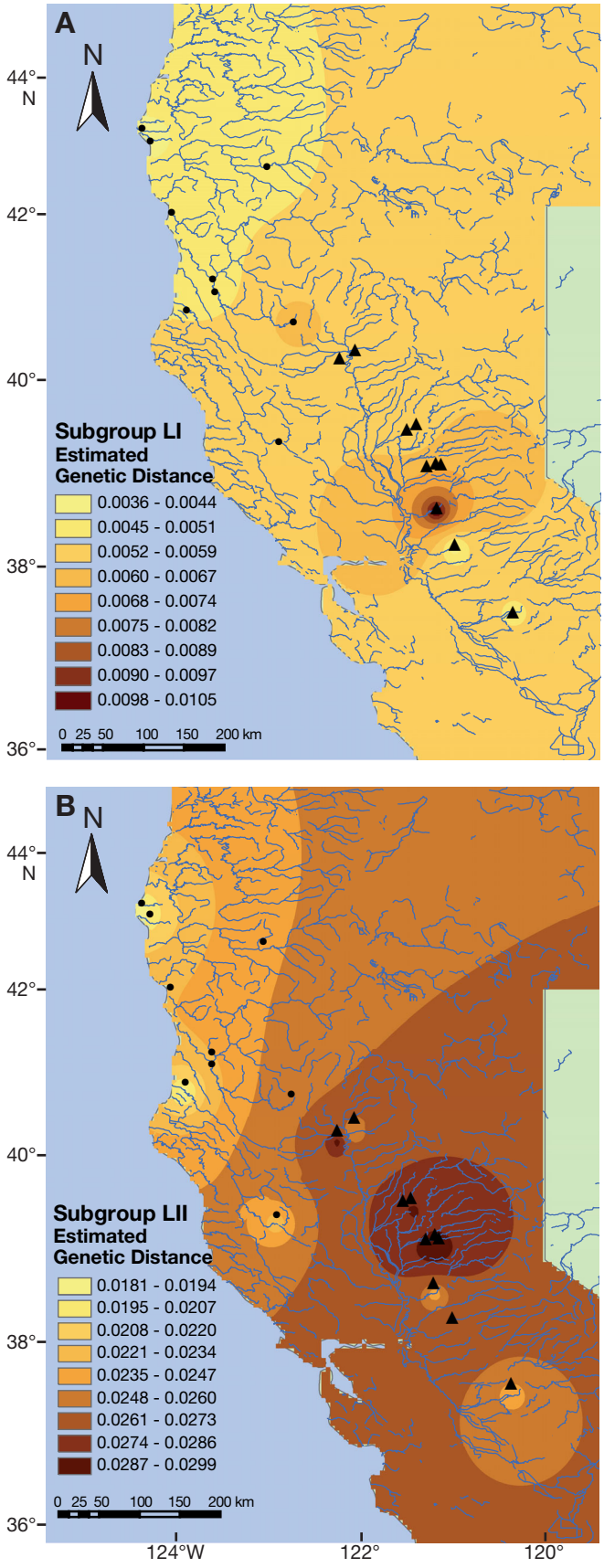
Sensitivity analysis. The model was executed 5 times using 3 different prior distributions for τ, 3 for β, and 2 for ψ to determine whether estimates of the coefficients would be significantly different with different prior information.

All models were run for 50 000 iterations with a burn-in of 4000 iterations to ensure sampling from a stable distribution. Convergence of the posterior sampling procedure (i.e. the Gibbs sampler) was assessed using the Gelman-Rubin convergence statistic (Brooks & Gelman 1998). The WinBUGS© code used to fit the models is available upon request. The relative accuracy of predictions from models using different explanatory variables was assessed by calculating mean absolute prediction errors (MAPE). MAPEs were calculated as the average absolute value of the error (predicted genetic distance minus observed genetic distance) for all isolates. Variable selection proceeded by choosing the model with the smallest MAPE.

RESULTS

Spatial distribution of the genetic distance

The IDW map (Fig. 1) depicts the relationship between location and genetic distance by indicating the predicted genetic distances of IHNV isolates in California and southern Oregon based on the collected samples (Kelley et al. 2007) for subgroups LI and LII, respectively. The objective of the exploratory IDW analysis of genetic distance was to provide a descriptive, graphical statistic for the relationship between genetic divergence and geographical location, without controlling for other factors. IDW provided a smoothed estimate of the geographical variation of the genetic divergence of the virus, which was estimated as a function of the geographical distance among isolates and of the genetic distance between each isolate and the earliest isolate analyzed. The purpose of this smoothed



representation was to provide a simple visualization of the general spatial pattern of the genetic divergence in the area based on values observed in point locations sampled. Compared with lighter colors, darker colors represent greater genetic distances relative to the earliest isolate available. Visual inspection of the IDW map suggests that isolates from both virus subgroups collected at inland locations diverged genetically more from the earliest isolate than isolates collected at coastal locations.

Prediction model: factors predicting genetic distance

The model that best described the genetic distance of IHNV isolates in California (Kelley et al. 2007) included predictor variables for location (i.e. inland or coastal), year of isolation, and the interaction between location and year. The MAPE for this model was 0.0047, which is the average amount that the model's predicted genetic distances differed from the observed genetic distances; observed genetic distances ranged from 0.0023 to 0.0337. Though including the other information (fish age and spawning run) slightly improved the MAPE, the number of missing values in these variables prevented them from being used in a meaningful way. The addition of habitat to the model did not improve the MAPE.

Because the best model included so few predictors, we informally tested several additional putative predictors and found that they did not improve upon the best model. Specific data on the environments surrounding the isolates (hatchery nearby and tributary versus mainstream) were not recorded, and hence not tested in the regression equation. However, inspection of the location-specific random effects from the best model showed no appreciable difference in the effect of un-modeled factors in locations near hatcheries as opposed to those greater than 10 or 20 km from hatcheries, in mainstream versus tributary locations, or in any combination of these. The numbers of isolates from each host species were: 198 from Chinook salmon, 35 from steelhead trout, 2 from coho salmon, 1 from rainbow trout, and 1 unknown. The differences in the average genetic distances for inland versus coastal isolates from steelhead versus Chinook were insignifi-

Fig. 1. Genetic distance of IHNV in California and southern Oregon estimated using observed genetic distances of viruses isolated between 1966 and 2004. Color shades represent spatial-distance-weighted smoothed genetic distances. (A) predicted genetic distances for subgroup LI; (B) predicted genetic distances for subgroup LII. (●): coastal sample locations; (▲): inland sample locations previously described by Bendorf et al. (in press) and Kelley et al. (2007). Several collection sites appear as one triangle at this resolution due to their proximity

Table 2. Posterior summaries of the regression coefficients, mean (with 95% probability interval) and posterior probability that the coefficient is <0 (or >0 in the case of the coefficient for Year)

Parameter	Mean	Probability < (or >) 0
Intercept	−3.95 (−4.22; −3.69)	1.00
Geolocation	−0.54 (−1.04; −0.06)	0.99
Year	0.04 (0.02; 0.06)	(1.00)
Interaction	−0.08 (−0.20; 0.03)	0.91

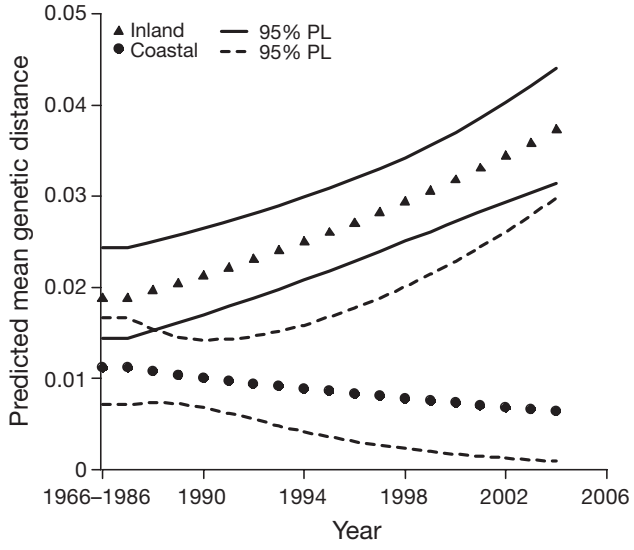


Fig. 2. Estimated mean genetic distance among IHNV isolates over time in coastal and inland locations illustrating the interaction between location and year of isolation. Estimated means and 95% probability limits (PL) are based on a mixed-effects beta regression model with random effect set to 0, which corresponds to a 'typical' site

cant in a simple contingency table analysis. The numbers of isolates from coho and rainbow trout or even a lumped 'other,' category were insufficient for meaningful statistical analysis.

The posterior estimates of the regression coefficients that defined the best model (Table 2) suggest that genetic distance was greater for isolates collected during 1988–2004 at inland locations (estimated median genetic distances ranged from 0.020 in 1988 to 0.037 in 2004) than for isolates found in coastal watersheds (estimated mean genetic distances ranged from 0.011 in 1988 to 0.006 in 2004) from northern California and southern Oregon during the same years. Using the parameter estimates in Table 2, we can calculate the model-predicted genetic distances for each location at each time as shown in Fig. 2.

The estimated mean genetic distance among isolates at coastal locations prior to 1988 (0.011) was slightly less than that of viruses isolated inland (0.018), as indicated by the negative regression coefficient for geolocation (geolocation was coded such that the value 1 corresponded to coastal locations, and 0 corresponded to inland locations) in Table 2. In the period 1988–2004 (Fig. 2), the estimated mean genetic distances increased from 0.020 to nearly 0.037 among inland locations. The mean genetic distances among coastal isolates decreased slightly from 0.01 during the period 1988–2004 (Fig. 2). In Fig. 2, the 95% probability band was constructed by interpolating adjacent individual 95% intervals for each year. In a given year, the interval contains the true mean genetic distance with probability 0.95.

Transformations of the time variable [power ($year^2$), inverse (1/year), and exponential (e^{year})] did not decrease the MAPE compared to the model with untransformed year values. Estimates of the regression coefficients for location, year, and the interaction between location and year were insensitive to moderate/large changes of the prior specification.

DISCUSSION

This is the first analytical, model-based study of the spatio-temporal genetic variation of IHNV. Our analysis of IHNV genetic distance data from 1966–2004 indicated that location (i.e. inland or coastal) and year of isolation were predictive of genetic distance. Although the IDW maps in Fig. 1 do not permit visualization of change in genetic distance over time, they are consistent with results of the beta regression analysis in that inland isolates were more genetically divergent than coastal isolates. Although Kelley et al. (2007) noted possible relationship between subgroups (LI and LII) and geographical location, the current study is, as far as we know, the first statistical evaluation that addresses the macro temporal-spatial pattern of IHNV genetic diversity, which when explored further might provide insight into possible underlying forces affecting IHNV genetic variation.

The model presented here suggests that genetic distance was greater for isolates collected from 1966–2004 at inland locations (ranging from 0.020 in 1988 to 0.037 in 2004), compared with isolates found in coastal watersheds (ranging from 0.011 in 1988 to 0.006 in 2004) from California and southern Oregon during the same years (Fig. 2). Although associations between specific factors unique to each collection site and genetic distance were not examined here, the IDW maps (Fig. 1 A and B) highlight the geographic differences in genetic distances, where a greater genetic distance occurs among isolates in subgroup LII (ranging from 0.0181 to 0.0299) than among isolates in subgroup LI (ranging from 0.0035 to 0.0105). The pre-

dicted genetic distances for LI and LII presented visually in the IDW maps support previous findings (Kelley et al. 2007) suggesting that the rate of divergence in subgroup LII (1×10^{-3} nucleotide substitutions site^{-1} yr^{-1}) was greater than that in subgroup LI (1×10^{-5} nucleotide substitutions site^{-1} yr^{-1}). The agreement of results from the IDW and beta regression analyses supports the conclusion that more genetic variation existed during this time period within the LII subgroup of IHNV isolates obtained in inland watersheds of California than within coastal watersheds of northern California and southern Oregon.

The choice of a common ancestor as a reference to measure genetic differences among isolates somewhat limits the degree to which the model can be expected to predict genetic difference because specific nucleic or amino acid differences among isolates could not be considered. Further, we explored use of regression models to predict the subgroups and serotypes of isolates, but the lack of variability in the data beyond that predicted for year and location precluded computational evaluation of other variables. To avoid these analytical difficulties, future studies should attempt to obtain specific data that characterize habitats and management.

Although the temporal component of genetic change could not be addressed by the IDW analysis, the IDW maps (Fig. 1) offer a method to identify geographic areas with significantly different genetic strains of IHNV. Identification of these areas represents an important initial step in identifying specific explanations for such geographic differences, which should be the subject of subsequent studies. For example, Rodriguez et al. (1996) suggest that genetic divergence in vesicular stomatitis virus–New Jersey may reflect adaptations of the virus to different vectors in highland versus lowland endemic areas within Costa Rica. Here, there may be specific factors unique to management or habitat of either the inland or of the coastal watersheds that are fostering or retarding genetic diversity. The predicted high difference in genetic distance of isolates from inland areas compared with coastal areas may be related to unknown environmental factors affecting exposure of fish to the virus. We speculate that the presence of intensive salmonid hatcheries and altered water flows and temperatures more common for inland areas may have contributed to differences in genetic change among inland and coastal isolates during the 34 yr time period. Studies examining specific factors that may influence viral mutation within the fish host are needed; specifically, the influence of temperature on transmission rates among overlapping migratory runs of adult chinook salmon and virus amplification due to an increase in susceptible fish biomass.

The model developed in this study provides a quantitative approach to predict IHNV temporal and spatial IHNV genetic variation and should prove useful in predicting genetic variation attributable to other putative factors influencing variation within habitats and management systems. To refine the models developed in this study, additional studies of other disease systems that include agent nucleic acid sequence data, a stringent sampling scheme for both the pathogen and host, and spatial and temporal covariates are warranted.

Acknowledgements. We gratefully acknowledge funding and support from both the Intelligence Community Postdoctoral Research Fellowship Program and the Armed Forces Medical Intelligence Center. We kindly thank R. Hedrick, W. Cox, and C. Bendorf for their technical assistance.

LITERATURE CITED

Amend DF, Yasutake WT, Mead W (1969) A hematopoietic virus disease of rainbow trout and sockeye salmon. Trans Am Fish Soc 98:796–804

Bendorf CM, Kelley GO, Yun SC, Kurath G, Andree KB, Hedrick RP (in press) Genetic diversity of infectious hematopoietic necrosis virus (IHNV) from Feather River and Lake Oroville, California and virulence of selected isolates for Chinook salmon and rainbow trout. J Aquat Anim Health

Bootland LM, Leong JC (1999) Infectious hematopoietic necrosis virus. In: Woo PTK, Bruno, DW (eds) Fish diseases and disorders. CAB International, Wallingford, p 57–112

Brooks SP, Gelman A (1998) Alternative methods for monitoring convergence of iterative simulations. J Comput Graph Statist 7:434–455

Emmenegger EJ, Kurath G (2002) Genetic characterization of infectious hematopoietic necrosis virus of coastal salmonid stocks in Washington state. J Aquat Anim Health 14: 25–34

Emmenegger EJ, Meyers TR, Burton TO, Kurath G (2000) Genetic diversity and epidemiology of infectious hematopoietic necrosis virus in Alaska. Dis Aquat Org 40: 163–176

Ferrari SLP, Cribari-Neto F (2004). Beta regression for modeling rates and proportions. J Appl Stat 31:799–815

Garver KA, Troyer RM, Kurath G (2003) Two distinct phylogenetic clades of infectious hematopoietic necrosis virus overlap within the Columbia River basin. Dis Aquat Org 55:187–203

Isaaks EH, Srivastava RM (1989) Inverse distance methods. In: Isaaks EH, Srivastava RM (eds) An introduction to applied geostatistics. Oxford University Press, New York, NY, p 257–259

Kelley GO, Bendorf C, Yun SC, Kurath G, Hedrick RP (2007) Genotypes and phylogeographical relationships of infectious hematopoietic necrosis virus in California, USA. Dis Aquat Org 77:29–40

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kurath G, Garver KA, Troyer RM, Emmenegger EJ, Einer-Jensen K, Anderson ED (2003) Phylogeography of infectious haematopoietic necrosis virus in North America. J Gen Virol 84:803–814

Kurath G, Leong JC (1985) Characterization of infectious hematopoietic necrosis virus mRNA species reveals a non-virion rhabdovirus protein. J Virol 53:462–468

Lively CM, Dybdahl MF (2000) Parasite adaptation to locally common host genotypes. Nature 405:679–681

Morzunov SP, Winton JR, Nicho ST (1995) The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. Virus Res 38:175–192

Mulcahy D, Pascho R, Jenes CK (1984) Comparsion of *in vitro* growth characteristics of ten isolates of infectious haematopoietic necrosis virus. J Gen Virol 65:2199–2201

Nichol ST, Rowe JE, Winton JR (1995) Molecular epizootiology and evolution of the glycoprotein and non-virion protein genes of infectious hematopoietic necrosis virus, a fish rhabdovirus. Virus Res 38:159–173

Parisot TJ, Yasutake WT, Klontz GW (1965) Virus diseases of Salmonidae in Western United States. I. Etiology and epizootiology. Ann NY Acad Sci 126:502–519

Rodriguez LL, Walter, MF, Nichol ST (1996) Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. Proc Natl Acad Sci USA 93: 13030–13035

Ross AJ, Pelnar J, Rucker RR (1960) A virus-like disease of Chinook salmon. Trans Am Fish Soc 89:160–163

Troyer RM, Kurath G (2003) Molecular epidemiology of infectious hematopoietic necrosis virus reveals complex virus traffic and evolution within southern Idaho aquaculture. Dis Aquat Org 55:175–185

Troyer RM, LaPatra SE, Kurath G (2000) Genetic analyses reveal unusually high diversity of infectious haematopoietic necrosis virus in rainbow trout aquaculture. J Gen Virol 81:2823–2832

Williams IV, Amend DF (1976) A natural epizootic of infectious hematopoietic necrosis in fry of sockeye salmon (*Oncorhynchus nerka*) at Chilko Lake, British Columbia. J Fish Res Board Can 33:1564–1567

Wingfield WH, Nims L, Fryer JL, Pilcher KS (1970) Species specificity of the sockeye salmon virus (Oregon strain) and its cytopathic effects in salmonid cell lines. In: Snieszko SF (ed) A symposium on diseases of fishes and shellfishes. American Fisheries Society, Washington, DC, p 319–326

Wolf K (1988) Infectious hematopoietic necrosis virus. In: Wolf K (ed) Fish viruses and fish viral diseases. Cornell University Press, Ithaca, NY, p 83–114

Editorial responsibility: Julie Bebak, Auburn, Alabama, USA

Submitted: May 5, 2007; Accepted: August 17, 2007
Proofs received from author(s): September 14, 2007