Highly Pathogenic H5N1 Avian Influenza Viruses Exhibit Few Barriers to Gene Flow in Vietnam

Margaret Carrel[1], **Xiu-Feng Wan**[2], **Tung Nguyen**[3], and **Michael Emch**[4]

[1]Department of Geography, University of Iowa, 316 Jessup Hall, Iowa City, IA 52242

[2]Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS 39762

[3]Department of Animal Health, National Center for Veterinary Diagnostics, Hanoi, Vietnam

[4]Department of Geography, University of North Carolina-Chapel Hill, Saunders Hall, Chapel Hill, NC 27599

Abstract

Locating areas where genetic change is inhibited can illuminate underlying processes that drive evolution of pathogens. The persistence of highly pathogenic H5N1 avian influenza in Vietnam since 2003, and the continuous molecular evolution of Vietnamese avian influenza viruses, indicates that local environmental factors are supportive not only of incidence but also of viral adaptation. This article explores whether gene flow is constant across Vietnam, or whether there exist boundary areas where gene flow exhibits discontinuity. Using a dataset of 125 highly pathogenic H5N1 avian influenza viruses, principal components analysis and wombling analysis are used to indicate the location, magnitude, and statistical significance of genetic boundaries. Results show that a small number of geographically minor boundaries to gene flow in highly pathogenic H5N1 avian influenza viruses exist in Vietnam, but that overall there is little division in genetic exchange. This suggests that differences in genetic characteristics of viruses from one region to another are not the result of barriers to H5N1 viral exchange in Vietnam, and that H5N1 avian influenza is able to spread relatively unimpeded across the country.

Keywords

Background

Avian influenza viruses can undergo dramatic changes in genetic sequence across short temporal and spatial distances. This ability for rapid molecular evolution enables avian influenza viruses to overcome host immunological responses and to more easily transmit from infected to susceptible host (Muramoto et al. 2006; Neumann and Kawaoka 2006; Smith et al. 2006). Tracking the evolution of avian influenza viruses is vital to define prevention and control measures, and detect areas where gene flow is blocked can illuminate how viral evolution is affected by landscape features.

Highly pathogenic H5N1 avian influenza has been continuously circulating in poultry populations and also caused sporadic human cases in Vietnam since 2001. Multiple cultural

and environmental factors in Southeast Asia, such as maintenance of backyard poultry flocks and the popularity of live bird markets, allow H5N1 to persist despite government vaccination campaigns and other containment efforts (Nguyen et al. 2005; Chen et al. 2006; Cristalli and Capua 2007). At least six new genotypes unique to Vietnam emerged from 2003 to 2008, and more new H5N1 subtypes are continuously emerging (Wan et al. 2008; Nguyen et al. 2009; Davis et al. 2010; FAO 2011).

Genetic diversity is reliant on limited genetic exchange among populations and either selection pressures in local-level environments, resulting in reassortment, which could lead to antigenic shift, or random mutations, which could lead to antigenic drift (Barbujani and Sokal 1990). Identifying boundaries to gene flow, common in ecology and in population genetics, is useful to generate hypotheses about underlying processes that drive viral incidence and evolution (Jacquez et al. 2000). Difference boundaries, hereafter referred to as either barriers or boundaries, are located in spatial zones of rapid genetic change, and indicate the presence of either a sharp environmental transition and/or a low rate of viral exchange (Barbujani et al., 1989).

Analysis of non-H5N1 avian influenza viruses in North America suggests that gene flow is restricted across migratory flyways (Lam et al. 2011). Previous studies of H5N1 avian influenza suggest that political borders can act as boundaries to gene flow, and that not all viral strains are able to spread across the Eurasian landscape (Wallace and Fitch 2008). The presence or the absence of similar filtering boundaries within Vietnam itself has never been established, however, leaving unanswered questions about how H5N1 influenza gene flow varies across the country. The genetic characteristics of H5N1 viruses in Vietnam have also been shown to have distinct spatial and temporal patterning, varying between northern and southern provinces (Carrel et al. 2010). The patterns observed were hypothesized to be driven by the distinct regional patterning of human and bird populations, with high population densities in the Red River and Mekong River deltas and the major cities of Hanoi and Ho Chi Minh City. However, whether the genetic differences observed between northern and southern isolates is driven by the blockage of genetic exchange due to comparatively low human and bird population densities in central Vietnam, or some other underlying variable, is still unknown.

Here we seek to explore whether there exist significant boundaries to H5N1 avian influenza genetic exchange. Specifically, is the genetic drift of H5N1 viruses in Vietnam impeded in space or time by barriers? Or do H5N1 avian influenza viruses drift unimpeded across the landscape?

Data and Methods

To investigate the potential presence of barriers to gene flow, a dataset consisting of 125 highly pathogenic H5N1 avian influenza viruses isolated across Vietnam from 2003 to 2007 was adapted from our previous study (see Table S1 in Carrel et al. 2010 for GenBank accession codes). These viruses were primarily found in domestic poultry, such as chickens and ducks, but were also sampled from other domestic bird species, such as quail, and from the environment, usually bird feces or cages containing birds. For each virus there exists location data on the province where the virus was found (representing 28 of Vietnam's 63 provinces) and the year in which the virus was sampled, as well as a full-length or at least 90% of the full-length genetic sequence for each segment. These viruses genetically belong to the Clade 1 and genotype VN3 (A/duck/HongKong/821/2002 (H5N1)-like virus), and no reassortment events were observed among these 125 isolates (Wan et al. 2008).

Two genetic distance measures were generated from the dataset via a maximum likelihood phylogenetic tree constructed using nucleotide sequences. First, patristic genetic distance for

each of the eight gene segments in the 125 viruses from the ancestral Hong Kong virus (a "straight line" genealogical distance) was calculated using PATRISTIC (Fourment and Gibbs 2006). Patristic distance indicates a total amount of genetic change that exists between genetic sequences by summing the lengths of branches in the phylogenetic tree between two viruses. Second, a 125×125 matrix of genetic distances between viruses was generated, also using PATRISTIC (Fourment and Gibbs 2006). Thus, for each virus we have not only its genetic distance from the putative progenitor virus but also the distance from every other virus in the dataset.

The eight straight line genetic distances were used as variables in a principal components analysis (PCA). PCA is used to explore the underlying structure of a dataset, and works by converting the observed values on a set number of possibly correlated variables into a new set of values for uncorrelated principal component variables. This data transformation results in a first principal component (or factor) that accounts for as much variance in the observed dataset as possible, a second principal component that is uncorrelated with the first and accounts for as much remaining variability as possible, and so on. The number of principal components that exist is always less than or equal to the original number of variables in the dataset. Each data observation was assigned a score for each factor detected.

PCA on the 125 observations of eight genetic distance measures was conducted in SPSS using a correlation method, varimax rotation, returning only principal components with Eigenvalues >1. Factor scores for each of the 125 viruses were graphed according to the region in which each virus was isolated (northern Vietnam versus southern Vietnam) and the year of incidence. This allowed for exploration of whether sharp distinctions in PCA scores could be observed across these categories, indicating the presence of potential regional or temporal genetic dissimilarity and, by extension, the existence of genetic discontinuity due to boundaries.

Previous studies of genetic discontinuity have indicated that areas of sharp change in a smoothed map of PCA factor scores can signal the presence of boundaries to gene flow (Brierley et al. 1995; Piertney et al. 1998). Factor scores for each of the 125 viruses from the first two principal components were mapped across Vietnam using a geographic information system (GIS) and an inverse distance weighted (IDW) interpolation method. IDW is a more appropriate interpolation method than kriging for this dataset given the irregular spacing of the data. Viruses were grouped according to the province of isolation, and then a province-level factor score for each principal component was calculated by averaging across all viruses in each province. These average factor scores were then assigned to the geographic center of the province in the GIS. IDW interpolation methods were then used to generate a smoothed map of average factor scores across Vietnam, enabling observation of areas of steep changes in PCA scores.

Wombling is one of the original methods used for delineation of barriers to gene flow. Womble (1951) proposed that when examining genetic cline across geographic space, boundaries to genetic exchange exist where the cline exhibits steep gradients (Womble 1951). Detecting areas of steep slopes of genetic change can illuminate places where gene flow is inhibited by environmental or other features. In wombling, there is no constraint on the position, orientation, or shape of detected boundaries (Barbujani and Sokal 1990). Neither wombling nor mapping of PCA scores to determine boundaries to gene flow have previously been applied to the genetics of H5N1 avian influenza, although these methods have been used to study genetic features of other infectious diseases such as rabies and non-infectious diseases such as leukemia (Real et al. 2005; Wheeler and Waller 2008; Jacquez 2010).

The geographic scale at which the influenza data were collected led to several different viruses being assigned to the same geographic location at the center of the province of isolation. In the PCA analysis, the scores for viruses taking place in the same province were averaged before being mapped. In contrast, before conducting the wombling boundary analysis, the average genetic distance for each province on all eight gene segments was calculated by collapsing the 125×125 distance matrices into 28×28 distance matrices, with one row/column for each of the 28 provinces in Vietnam where viruses were isolated. Each of the eight 28×28 distance matrices was then associated with the 28 latitude and longitude coordinates of the province centroids as calculated in the GIS. Collapsing the matrices was preferred over randomly assigning viruses to geographic locations within the provinces as to avoid falsely detecting genetic boundaries between viruses located in the same provinces.

In wombling, magnitude indicates the average absolute slope of surfaces at a point, and direction indicates the orientation of the slope (Barbujani et al. 1989). A triangulation wombling procedure was implemented in PASSaGE to account for the irregular spatial sampling of the influenza datapoints (Rosenberg and Anderson 2010). Under triangulation wombling, potential boundary elements are located at the center of Delaunay triangles created among the observed datapoints. For each boundary element, the magnitude is calculated as

$$m = \sqrt{\left[\frac{\partial f(x, y)}{\partial x}\right]^2 + \left[\frac{\partial f(x, y)}{\partial y}\right]^2}$$

where $f(x, y) = ax + by + c$ and constants a, b, and c are calculated from

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} x_A & y_A & 1 \\ x_B & y_B & 1 \\ x_C & y_C & 1 \end{bmatrix}^{-1} \begin{bmatrix} Z_A \\ Z_B \\ Z_C \end{bmatrix}$$

The direction of the slope of change is calculated at each boundary element as (Rosenberg and Anderson 2010)

$$\theta = \arctan\left(\frac{\partial f(x, y)/\partial x}{\partial f(x, y)/\partial y}\right)$$

Once the surface of genetic change, with magnitude and direction calculated at each boundary element, is created, neighboring boundary elements are linked into boundaries if they satisfied three criteria. The first is that they fall into the top 10% of all calculated magnitudes. The second is that the directions of the two boundary elements must be $<90°$ different, so that boundary elements of opposing gradients are not connected (Figure 1). Finally, boundary elements are connected into boundaries only if the angle of their connection is $>30°$ different than the direction of the slope. The so-called singleton boundaries consist of barrier elements that have magnitudes of change large enough to meet the first criteria, but that are not located next to other boundary elements that meet the criteria. The statistical significance ($P < 0.10$) of all detected boundaries was evaluated based on 999 random permutations of the input data 28×28 matrix. Boundaries were mapped onto province centroids in the dataset to visualize their spatial patterns.

Given the temporal range of the dataset, spanning 5 years of viral isolation, the potential existed for year-specific genetic boundaries to be obscured in an analysis of the overall dataset. To detect the presence of such boundaries, the 125 viruses were divided into three separate temporally contiguous groups: those that occurred between 1/1/2003–12/31/2004 (55 viruses), 1/1/2004–12/31/2005 isolates (82 viruses), and 1/1/2005–12/31/2007 isolates (70 viruses). The 2005–2007 group is considered temporally contiguous in light of the lack of reported H5N1 in Vietnam in 2006. Eight genetic distance matrices (one for each gene segment) were then generated for these viral groupings, again with the genetic distances for isolates from the same province being averaged. Wombling was then performed on all eight genetic matrices for each of the three groupings, and the results were mapped onto the province centroids in the dataset.

Results

Two principal components with eigenvalues of >1 were detected in the analysis of the eight straight-line influenza genetic distances. The first component accounted for 47.2% of the variance in the dataset, the second accounted for 37.3%, for a total explanation of 84.5% of the variance. When the PCA scores for these two factors were graphed for all 125 viruses based on the location of incidence (in either northern or southern Vietnam) some regional distinctions in PCA scores were observed (Figure 2). Northern and southern viral PCA scores were mixed across the upper and lower left quadrants of the graph (negative Factor 1 scores, both positive and negative Factor 2 scores). A cluster of southern H5N1 viruses is seen at the positive end of the Factor 1 scale, this cluster also includes one northern virus.

Smoothing the PCA scores across a map of Vietnam (Figure 3) further indicates the extension of southern viruses across the full range of Factor 1 and the position of both northern and southern viruses across the range of Factor 2. Areas of both high and low Factor 1 PCA scores are seen in southern Vietnam, indicating potential barrier zones for gene flow. Both positive and negative PCA Factor 2 scores are seen in northern and southern Vietnam, also indicating the possibility of barriers.

Stratifying the PCA scores by year of viral incidence indicated that the temporal characteristics of the viruses might also be strongly connected to the genetic characteristics of the viruses (Figure 4). Viruses from 2003 to 2004 exhibit primarily negative PCA scores for both factors, while 2005 viruses cluster in the negative Factor 1 and positive Factor 2 quadrant. Viruses isolated in 2007 are located solely in the cluster of highly positive Factor 1 scores.

Results from womble analyses of the eight 28×28 genetic distance matrices detected multiple boundaries to gene flow on each gene segment (Table 1). Of these genetic boundaries, only those detected for the PB1 and the NS gene segments are statistically significant in the randomization trials.

The location, magnitude, and direction of the PB1 and NS boundary elements and boundaries are mapped in Figure 5. A single genetic boundary connects three high-magnitude (high gradient of genetic change) PB1 boundary elements in a small area of northern Vietnam, with directions of change to the north. A singleton boundary element is located in the southwestern portion of the map, outside of the Vietnam study area, with a genetic slope towards the west. The location of this boundary element is due to the concave nature of Vietnam's boundaries with Cambodia and Laos, such that when the Delaunay triangulation of points was generated some vertices fall outside of the political boundary of Vietnam. NS boundaries are found in both northern and southern Vietnam. A northern NS boundary consists of a singleton boundary element with a large magnitude and is in a

southwestern direction away from the locations of viral incidence. Three southern NS genetic boundary elements, with gradient directions towards the west, are connected into a single boundary. Again, this boundary extends outside of the border of Vietnam, and is due to the connection of data points via straight lines without political constraints.

Boundary detection for viruses stratified by year of incidence (into three temporally contiguous sets) found statistically significant barriers to gene flow only in 2003–2004 viruses on the PB1 and NP gene segments (Table 2). While boundaries were detected in other years and on all other gene segments, randomization analysis indicated that they were not unexpected.

The location, direction, and magnitude of these significant barriers can be seen in Figure 6. Nearly identical boundaries were detected for PB1 and NP, located in northern Vietnam and connecting boundary elements with high magnitude and directed towards the northeast (the PB1 boundary also connects to a boundary element directed towards the southwest). A singleton boundary element with a high magnitude was detected in southern Vietnam among 2003–2004 NP genetic distances, and is directed from southwest to northeast.

Despite the lack of statistical significance for the bulk of detected boundary elements and connected boundaries, the majority of high-magnitude boundary elements identified across all years and all eight gene segments were located in northern Vietnam. The direction of gene flow along these northern boundary elements was almost exclusively to the north and northwest, indicating that rates of genetic difference drop from south to north in northern Vietnam. Boundary elements in southern Vietnam were detected only in the 2003–2004 and 2004–2005 datasets, and there were no overall patterns to the direction of these magnitudes as are observed in the northern boundary elements. Additionally, no boundaries or boundary elements were found that divided the gene flow between northern and southern viruses.

Discussion

Areas of genetic discontinuity, where there are barriers to gene flow, do exist among highly pathogenic H5N1 avian influenza viruses in Vietnam. However, these genetic boundaries vary temporally and by gene segment rather than existing statically in space. Rather, viral evolution in Vietnam is driven by the amount of time that viruses have had to mix and mutate, and viruses appear to be able to spread across the landscape with few impediments.

Results from the PCA indicate that genetic characteristics of H5N1 viruses are heavily influenced by the year in which they occur, such that viruses isolated in 2003 and 2004 have very different PCA scores than those isolated in 2007. Viruses isolated in southern Vietnam had greater diversity in PCA scores than did those isolated in northern Vietnam, and smoothed maps of the PCA scores indicated that, for Factor 1, genetic boundaries were more likely to be found in southern Vietnam than northern. Mapping Factor 2 scores suggested that boundaries could also be located in northern Vietnam. While using PCA for studies of underlying population structure has been cautioned recently, due to the emergence of seemingly non-random sinusoidal patterns even when migration is spatiotemporally homogeneous, we believe that the PCA results obtained are still representative of underlying genetic structure of the dataset (Novembre and Stephens 2008). Figure 4 evidences no sinuosity along geographically connected observations, but rather a mixing of both northern and southern viruses at the lower end of the first PC.

Few statistically significant barriers were detected in the womble analyses, either on the overall dataset or the temporally stratified datasets. The location of the significant boundary on the NS gene segment in the overall womble analysis seems to indicate that there is a barrier to gene flow between southern Vietnam and the west, i.e., Cambodia, with genetic

diversity falling from east to west, as indicated by the arrows signaling gradient direction. While we did not explicitly test for barriers between Vietnamese and Cambodian isolates, these results suggest further study and also confirm previous work examining gene flow across borders (Wallace and Fitch 2008). The significant boundaries for the all-years PB1 gene segment, the 2003–2004 PB1 gene segment, and the 2003–2004 NP gene segment were located in northern Vietnam, dividing isolates in the provinces surrounding Hanoi, with gene flow decreasing from south to north. This area was also the site of the majority of boundary elements detected in the non-significant womble analyses, rather than the south, as the PCA Factor 1 map would have suggested. These results suggest that northern Vietnam, more so than southern, is the site of small geographic areas with high rates of genetic change.

The gene segments for which boundaries were detected, PB1, NP, and NS, are all internally located in influenza viruses. Each of these three gene segments encode for different characteristics of influenza infections. Mutations on the PB1 gene segment are reported to be responsible for viruses in ducks being pathogenic versus non-pathogenic viruses, and both the PB1 and NP affect replication strengths of viruses (Hulse-Post et al. 2007; Wasilenko et al. 2008). The NP gene is also linked to high pathogenicity of H5N1 viruses in chickens, and NP mutations may influence the adaptation of viruses from ducks to chicken populations (Wasilenko et al. 2008; Tada et al. 2011). The NS protein is believed to boost H5N1 virulence by affecting host immune responses in two ways: blocking interferons, thus hiding infection from immune system sensors, and boosting proinflammatory cytokines, encouraging systemic rather than localized infection (Geiss et al. 2002; Neumann and Kawaoka 2006; Gambotto et al. 2008). Exploration of genetic structure among AIV in migratory flyways also found limited gene flow among NP and NS gene segments, although not for PB1 genes (Lam et al. 2011). Locations of genetic discontinuity only on these internal segments, rather than the surface HA and NA gene segments, indicate that perhaps the location of gene segments within viruses affect the likelihood of boundaries to gene flow occurring.

A general pattern of H5N1 viral dispersal has been suggested for Vietnam, wherein viruses are introduced across the Chinese border into northern Vietnam. In northern Vietnam, viruses begin to interact and mutate, before gradually spreading southward (Wan et al. 2008). Viruses in southern Vietnam then begin to evolve and adapt to local conditions and diverge from the northern seed populations. This north to south movement is likely driven by movement of people and birds between the major population centers around Hanoi and Ho Chi Minh City. This general model is supported by our findings, that Hanoi and the surrounding province are the site of numerous boundaries to gene flow, and also high rates of genetic distance, and that southern Vietnam is also a site for divergent PCA scores and significant boundaries and boundary elements, but that central Vietnam is not a barrier zone for gene flow. While we had few samples from central Vietnam, womble analysis would still have detected rapid rates of genetic change from northern to central or central to southern regions, or would have indicated that central Vietnam, with its relatively lower human and poultry population density, acts as a regional sink for genetic change.

Our study is limited by the number of H5N1 samples used, and by the necessity of assigning those samples to only 28 geographic locations, the centroids of the province of isolation. This means that no barriers to genetic drift within a province can be detected. Additionally, genetic diversity within a province is necessarily reduced to one aggregate measure. Perhaps if more detailed geographic locations of viral incidence were available we would see different patterns of barrier presence or absence across northern and southern provinces. It is unlikely, however, that these data constraints are responsible for the lack of barriers to gene drift found in central Vietnam.

The lack of widespread or temporally continuous boundaries in the analyses is consistent with previous research on the genetic character of H5N1 avian influenza in Vietnam. Trend surface analyses of H5N1 influenza's movement across Asia from 2003 to 2009 found no evidence that land-based physical environmental barriers, such as mountains, affected the direction or speed of viral movement, although oceans did (Li et al. 2011). While our work did not correlate the presence or the absence of barriers with underlying population and environment features, our findings do suggest that no temporally static and spatially unvarying barriers exist in Vietnam to slow genetic drift of influenza viruses. Thus, while genetic characteristics of viruses do vary across Vietnam, it is not in such a way as to suggest the driving force of boundaries to gene flow. Rather, we hypothesize that genetic differences between northern and southern viruses previously observed are likely the result of gradual changes over time and space unimpeded by barriers and facilitated by human and poultry population movements. Surveys of sellers in live bird markets found that inexperienced poultry traders were more likely than experienced vendors to purchase birds from areas with prior H5N1 outbreaks, that poultry traders drew upon bird stocks across multiple geographic areas, and that birds were allowed to leave liver bird markets and be re-introduced into different flocks than they had originated, supporting this hypothesis about gradual genetic change over space/time (Magalhães et al. 2010). It also suggests that vaccination of poultry may be of less importance for H5N1 prevention and containment than control measures undertaken in Thailand to limit movement of live birds between farm and market (Paul et al. 2010). If there are few barriers to the movement of H5N1 across the Vietnamese landscape, and this movement is facilitated by population mobility, then patchy vaccination would be rendered ineffective by the introduction of exposed or contaminated humans and poultry from surrounding areas. Further research is needed to correlate such population–environment mobility variables with genetic change, since as yet the majority of population–environment research related to H5N1 has considered viral incidence rather than viral genetic change (Gilbert et al. 2006, 2008; Paul et al. 2010; Li et al. 2011; Martin et al. 2011).

References

Barbujani G, Sokal RR. Zones of sharp genetic change in Europe are also linguistic boundaries. Proceedings of the National Academy of Sciences of the United States of America. 1990; 87:1816. [PubMed: 2308939]

Barbujani G, Oden NL, Sokal RR. Detecting regions of abrupt change in maps of biological variables. Systematic Biology. 1989; 38:376.

Brierley AS, Thorpe JP, Pierce GJ, Clarke MR, Boyle PR. Genetic variation in the neritic squid *Loligo forbesi* (Myopsida: Loliginidae) in the northeast Atlantic Ocean. Marine Biology. 1995; 122:79–86.

Carrel MA, Emch M, Jobe RT, Moody A, Wan XF. Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza viruses in Vietnam. PLoS One. 2010; 5:e8631. [PubMed: 20072619]

Chen H, Smith GJ, Li KS, Wang J, Fan XH, Rayner JM, et al. Establishment of multiple sublineages of H5N1 influenza virus in Asia: implications for pandemic control. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:2845–2850. [PubMed: 16473931]

Cristalli A, Capua I. Practical problems in controlling H5N1 high pathogenicity avian influenza at village level in Vietnam and introduction of biosecurity measures. Avian Diseases. 2007; 51:461–462. [PubMed: 17494607]

Davis CT, Balish AL, O'Neill E, Nguyen CV, Cox NJ, Xiyan X, et al. Detection and characterization of clade 7 high pathogenicity avian influenza H5N1 viruses in chickens seized at ports of entry and live poultry markets in Vietnam. Avian Diseases. 2010; 54:307–312. [PubMed: 20521651]

FAO. Bird flu rears its head again. Food and Agriculture Organization of the United Nations Media Centre; 2011. http://www.fao.org/news/story/en/item/87196/icode/ [Accessed 19 January 2012]

Fourment M, Gibbs M. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evolutionary Biology. 2006; 6:1. [PubMed: 16388682]

Gambotto A, Barratt-Boyes SM, de Jong MD, Neumann G, Kawaoka Y. Human infection with highly pathogenic H5N1 influenza virus. Lancet. 2008; 371:1464–1475. [PubMed: 18440429]

Geiss GK, Salvatore M, Tumpey TM, Carter VS, Wang X, Basler CF, et al. Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:10736–10741. [PubMed: 12149435]

Gilbert M, Chaitaweesub P, Parakamawongsa T, Premashthira S, Tiensin T, Kalpravidh W, et al. Free-grazing ducks and highly pathogenic avian influenza, Thailand. Emerging Infectious Diseases. 2006; 12:227–234. [PubMed: 16494747]

Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, Czarnecki C, et al. Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:4769–4774. [PubMed: 18362346]

Hulse-Post D, Franks J, Boyd K, Salomon R, Hoffmann E, Yen HL, et al. Molecular changes in the polymerase genes (PA and PB1) associated with high pathogenicity of H5N1 influenza virus in mallard ducks. Journal of Virology. 2007; 81:8515. [PubMed: 17553873]

Jacquez GM. Geographic boundary analysis in spatial and spatio-temporal epidemiology: perspective and prospects. Spatial and Spatio-temporal Epidemiology. 2010; 1:207–218. [PubMed: 21218153]

Jacquez GM, Maruca S, Fortin MJ. From fields to objects: a review of geographic boundary analysis. Journal of Geographical Systems. 2000; 2:221–241.

Lam TTY, Ip HS, Ghedin E, Wentworth DE, Halpin RA, Stockwell TB, et al. Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among North American birds. Ecology Letters. 2011; 15:24–33. [PubMed: 22008513]

Li XH, Tian HD, Heiner M, Li DM. Global occurrence and spread of highly pathogenic avian influenza virus of the subtype H5N1. Avian Diseases. 2011; 55:21–28. [PubMed: 21500631]

Magalhães RS, Ortiz-Pelaez A, Thi K, Dinh Q, Otte J, Pfeiffer D. Associations between attributes of live poultry trade and HPAI H5N1 outbreaks: a descriptive and network analysis study in northern Vietnam. BMC Veterinary Research. 2010; 6:10. [PubMed: 20175881]

Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, et al. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. PLoS Pathogens. 2011; 7:e1001308. [PubMed: 21408202]

Muramoto Y, Le TQ, Phuong LS, Nguyen T, Nguyen TH, Sakai-Tagawa Y, et al. Pathogenicity of H5N1 influenza A viruses isolated in Vietnam between late 2003 and 2005. The Journal of Veterinary Medical Science. 2006; 68:735–737. [PubMed: 16891788]

Neumann G, Kawaoka Y. Host range restriction and pathogenicity in the context of influenza pandemic. Emerging Infectious Diseases. 2006; 12:881–886. [PubMed: 16707041]

Nguyen DC, Uyeki TM, Jadhao S, Maines T, Shaw M, Matsuoka Y, et al. Isolation and characterization of avian influenza viruses, including highly pathogenic H5N1, from poultry in live bird markets in Hanoi, Vietnam, in 2001. Journal of Virology. 2005; 79:4201–4212. [PubMed: 15767421]

Nguyen T, Davis CT, Stembridge W, Shu B, Balish A, Inui K, et al. Characterization of a highly pathogenic avian influenza H5N1 virus sublineage in poultry seized at ports of entry into Vietnam. Virology. 2009; 387:250–256. [PubMed: 19342072]

Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. Nature Genetics. 2008; 40:646–649. [PubMed: 18425127]

Paul M, Tavornpanich S, Abrial D, Gasqui P, Charras-Garrido M, Thanapongtharm W, et al. Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1: prospects from a spatial-based model. Veterinary Research. 2010; 41:28. [PubMed: 20003910]

Piertney SB, Maccoll ADC, Bacon PJ, Dallas JF. Local genetic structure in red grouse (*Lagopus lagopus scoticus*): evidence from microsatellite DNA markers. Molecular Ecology. 1998; 7:1645–1654. [PubMed: 9859199]

Real LA, Henderson JC, Biek R, Snaman J, Jack TL, Childs JE, et al. Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:12107. [PubMed: 16103358]

Rosenberg MS, Anderson CD. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods in Ecology and Evolution. 2010; 2:229–232.

Smith GJ, Naipospos TS, Nguyen TD, de Jong MD, Vijaykrishna D, Usman TB, et al. Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. Virology. 2006; 350:258–268. [PubMed: 16713612]

Tada T, Suzuki K, Sakurai Y, Kubo M, Okada H, Itoh T, et al. NP body domain and PB2 contribute to increased virulence of H5N1 highly pathogenic avian influenza viruses in chickens. Journal of Virology. 2011; 85:1834. [PubMed: 21123376]

Wallace RG, Fitch WM. Influenza A H5N1 immigration is filtered out at some international borders. PLoS ONE. 2008; 3:e1697. [PubMed: 18301773]

Wan XF, Nguyen T, Davis CT, Smith CB, Zhao ZM, Carrel M, et al. Evolution of highly pathogenic H5N1 avian influenza viruses in Vietnam between 2001 and 2007. PLoS ONE. 2008; 3:e3462. [PubMed: 18941631]

Wasilenko JL, Lee CW, Sarmento L, Spackman E, Kapczynski DR, Suarez DL, et al. NP, PB1, and PB2 viral genes contribute to altered replication of H5N1 avian influenza viruses in chickens. Journal of Virology. 2008; 82:4544. [PubMed: 18305037]

Wheeler DC, Waller LA. Mountains, valleys, and rivers: the transmission of raccoon rabies over a heterogeneous landscape. Journal of Agricultural, Biological, and Environmental Statistics. 2008; 13:388–406.

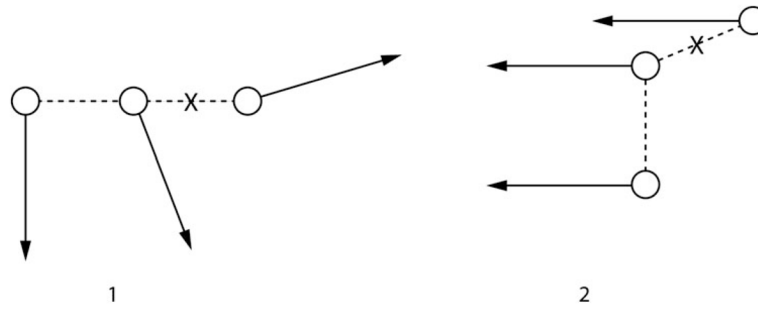Womble WH. Differential systematics. Science. 1951; 114:315–322. [PubMed: 14883851]

Figure 1.
Two criteria for boundary element connections. In drawing 1, two boundary elements (represented by *circles*) may be connected into a boundary because the directions of their associated slopes are <90° different. The third boundary element in drawing 1 has an opposing direction and will not be combined into a boundary. Two boundary elements in drawing 2 will be connected into a boundary because the bearing of the connection is >30° different than the direction of the slope. The third boundary element in drawing 2 has similar bearing for both the boundary connection and the direction of change, so connecting this would not accurately reflect the presence of a boundary. Adapted from Rosenberg and Anderson (2010).
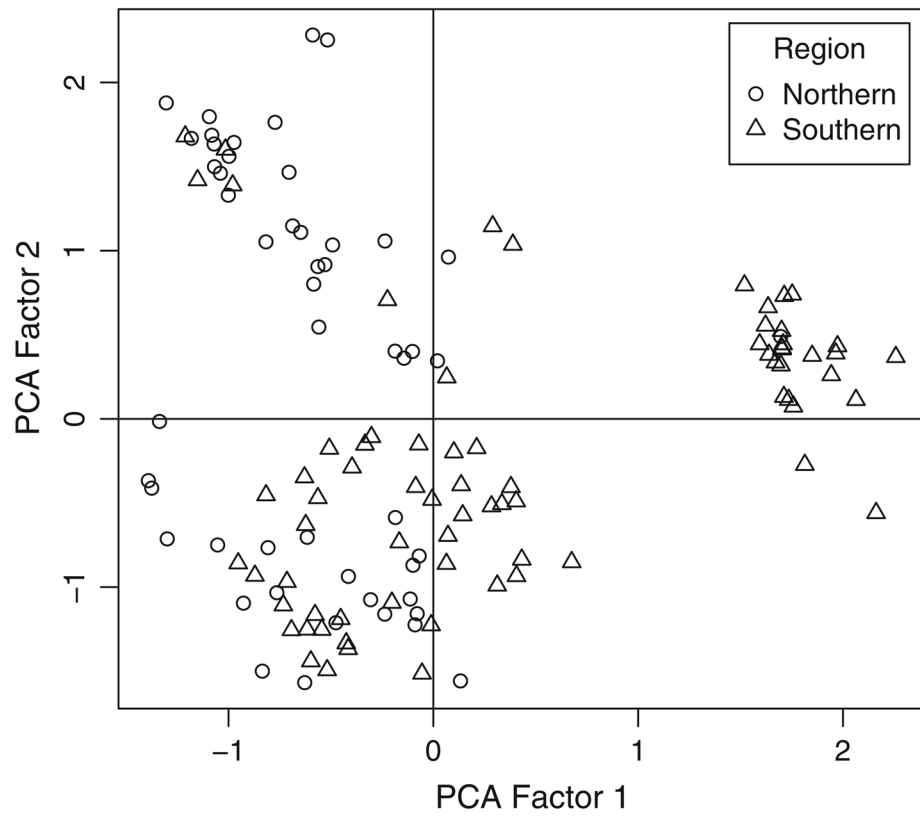
Figure 2.
PCA scores for the first two factors, plotted according to region of viral incidence.
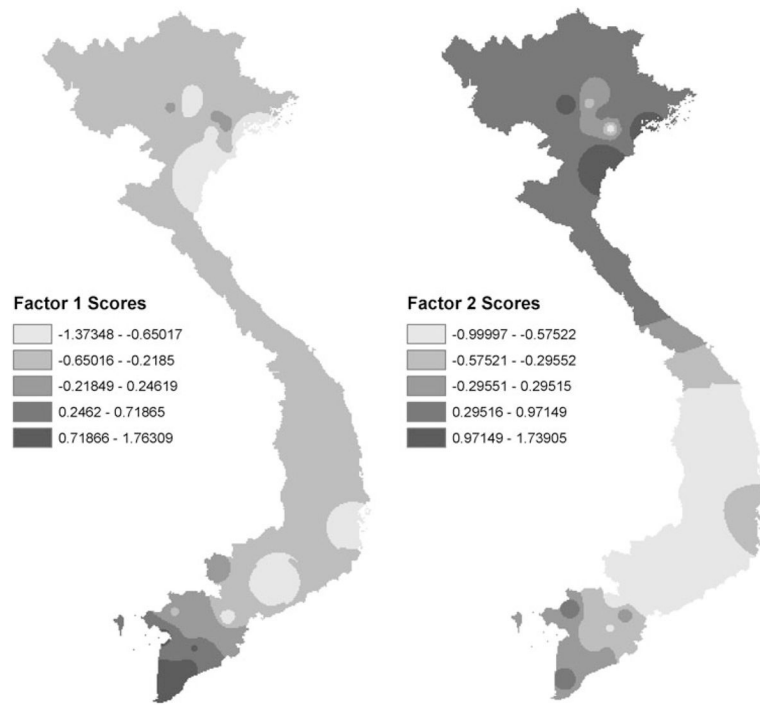
Figure 3.
Interpolated factor scores for the first and second principal components. Average factor scores were created for provinces with more than one H5N1 influenza virus.
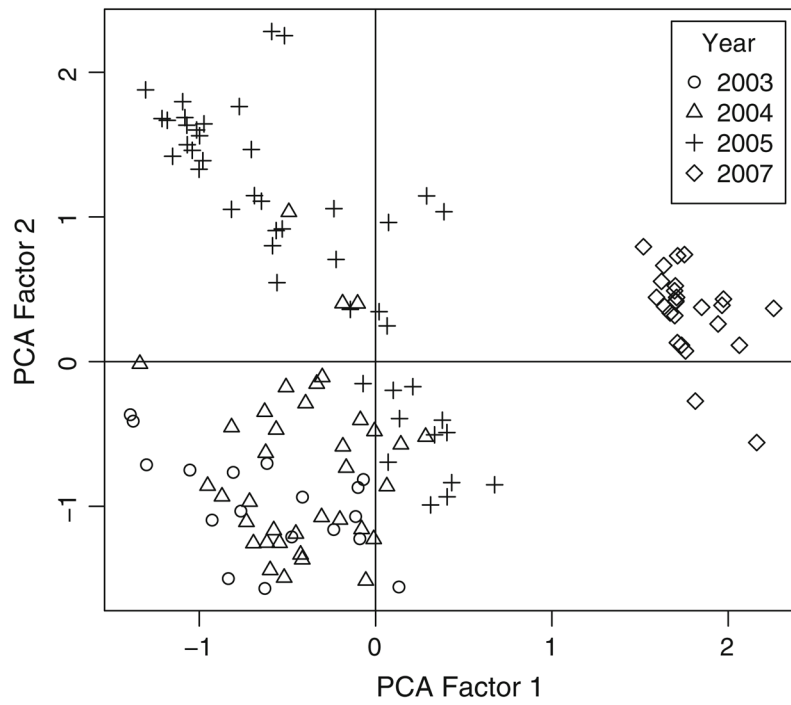
Figure 4.
PCA factor scores according to year of viral isolation.
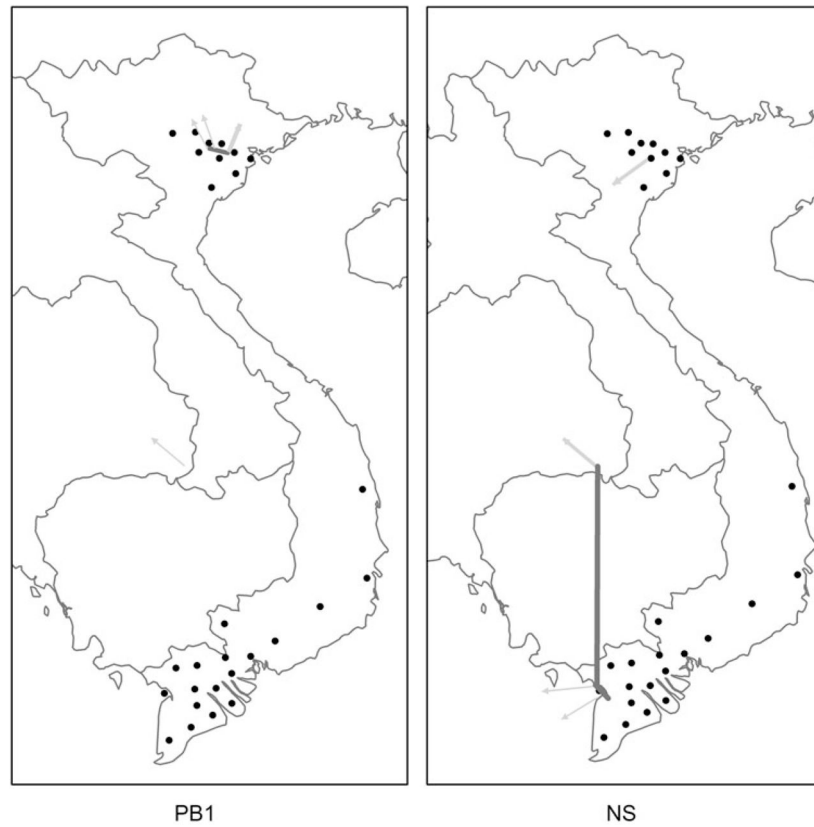
PB1 NS

Figure 5.
Statistically significant boundaries detected in the overall 2003–2007 dataset. *Black circles* indicate province centroids. Boundary elements are represented in *light gray*, thickness indicates degree of magnitude and *arrows* indicate the direction of the gradient. Boundary connections are represented in *dark gray*.
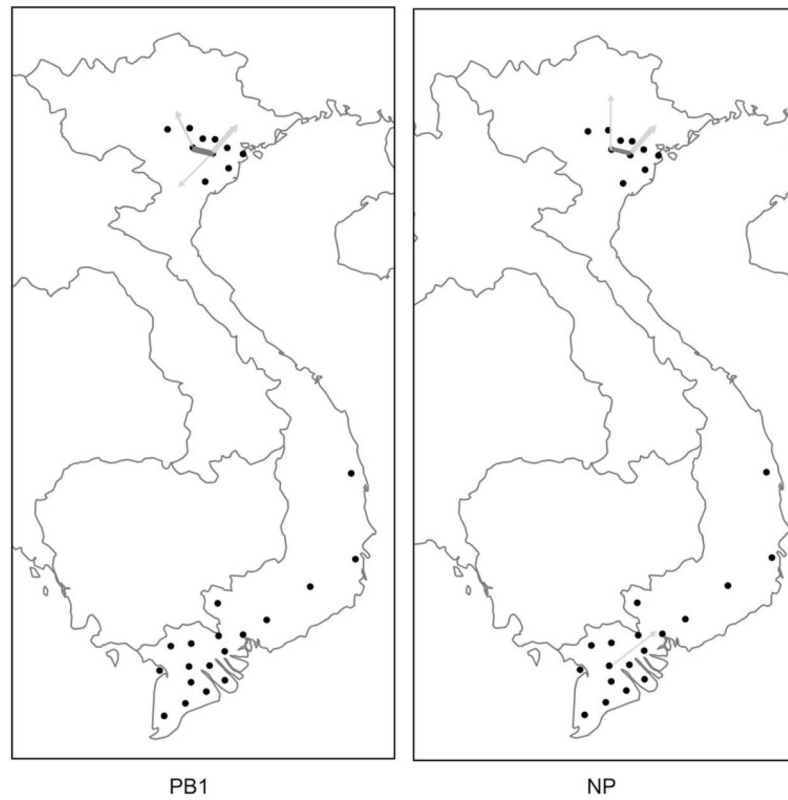
PB1   NP

Figure 6.
Statistically significant barriers detected in 2003–2004 viruses. Province centroids are shown in *black circles*, individual boundary elements in *light gray*, and connected boundaries in *dark gray*. Magnitude is indicated by thickness, while direction of genetic gradient is indicated by an *arrow*.

Table 1

Number of Barriers Detected Among Viral Distance on Each Gene Segment and Their Associated Significance Based on a Randomization Test

	PB2	PB1	PA	HA	NP	NA	MP	NS
No. of barriers	3	2	4	3	4	3	3	2
P value	0.447	0.094	1	0.45	1	0.428	0.412	0.082

Table 2

Number of Barriers Detected for Each Year Category According to Gene Segment

No. of barriers	PB2	PB1	PA	HA	NP	NA	MP	NS
2003–2004	3	2	3	3	2	3	2	2
2004–2005	4	4	4	4	4	4	4	4
2005–2007	2	2	2	2	2	2	2	2

Only boundaries for PB1 and NP gene segments in 2003–2004 were statistically significant at the $P < 0.10$ level, all others were insignificant based on randomization trials.