Invasions by Eurasian Avian Influenza Virus H6 Genes and Replacement of the Virus’ North American Clade

Heinrich zu Dohna, Jinling Li, Carol J. Cardona, Joy Miller, and Tim E. Carpenter

The spread of highly pathogenic avian influenza virus (AIV) (H5N1) underlines the potential for global AIV movement through birds. The phylogenies of AIV genes from avian hosts usually separate into Eurasian and North American clades, reflecting limited bird migration between the hemispheres. However, mounting evidence that some H6 sequences from North America cluster with Eurasian subtype H6 sequences calls the strict hemispheric divide into question. We conducted a comprehensive phylogenetic analysis of the extent and timing of cross-hemisphere movements by the H6 gene. Results suggested that Eurasian H6 subtype has invaded North America several times, with the first invasions occurring 10 years before the first detection of invading isolates. The members of the North American clade decreased from 100% in the 1980s to 20% in the 2000s among H6 isolates from North America. Unraveling the reasons for this large-scale gene movement between hemispheres might identify drivers of global AIV circulation.
Methods

Nucleotide Sequence Data

Nucleotide sequences and information on host, sampling date, and location were obtained for all H6 nucleotide sequences from the National Center for Biotechnology Information Influenza Sequence Database and BioHealthBase (24,25) by using the following keywords: influenza A, H6, and avian host. All H6 sequences above the length of 1,650 nt were retained, and duplicate sequences of the same isolates were removed, leading to a total of 291 H6 sequences. These sequences were aligned with AlignX (Vector NTI version X; Invitrogen, Carlsbad, CA, USA), and a maximum likelihood phylogenetic tree was built using the R package ape (26). The tree was based on a substitution model that was first proposed by Felsenstein (27) and is now standard for maximum-likelihood estimation in most phylogenetic software packages (e.g., see p. 104 in 26). This substitution model enables different rates for transitions and transversions, unequal base frequencies, and different substitution rates for wobble positions (26). The oldest H6 isolate was used as the root. The support of each bipartition was determined from 100-bootstrap samples. Other substitution models and neighbor-joining methods to construct phylogenetic trees produced very similar trees (H. zu Dohna, unpub. data).

Cross-Hemisphere Movement

Within the phylogenetic tree for the H6 gene, the 2 largest clades were determined and labeled as either North American or Eurasian, depending on where most isolates in each clade were found. All exceptions to this pattern, i.e., all isolates that clustered in a clade from a hemisphere that differed from their sampling location, were analyzed further. Following the analysis of Krauss et al. (6), these isolates were considered to have invaded a new hemisphere and will be referred to as invading isolates. The hemisphere to which the invading isolates belong genetically, as determined by the clade in which they cluster, will be referred to as the old hemisphere and the hemisphere from which they were isolated as the new hemisphere. All invading isolates were grouped into monophyletic invading subclades to estimate the times of invasion events.

For each invading subclade, 2 time points that bracket the most parsimonious invasion time were estimated. An invasion event that gave rise to a subclade must have happened after the last branching that led to descendents in the old hemisphere and before the time of the most recent common ancestor (TMRCA) for all descendents in the new hemisphere (Figure 1). These 2 time points were approximated for each subclade by estimating the TMRCAs for 2 sets of sequences, either including or excluding the most closely related sequence from the old hemisphere (Figure 1). This method assumes that invading isolates that are more closely related to each other than they are to other isolates from the old hemisphere descended from the same invasion event. This assumption is most parsimonious since there should be a low probability that the descendents of an invading virus reach a high enough density to be detected in bird samples several years later.

TMRCAs were obtained by fitting a coalescent process to the gene sequence data, by using the Bayesian Monte Carlo Markov Chain package BEAST 1.4 (28). A general time-reversible substitution model with variable population size (29) was fitted to nucleotide data. To maximize conformity with model assumptions, only the wobble positions of conserved amino acids with 4-fold degenerate codons were chosen. The program returns sampled posterior distributions of each TM RCA estimate. The Markov chain was run for 5 million steps and parameters were sampled every 1,000th step.

The posterior distribution \( f(t) \) of the time of invasion \( t \) was determined from all sampled pairs of the time of the earlier node and later node, \( t_{el} \) and \( t_{li} \), respectively:

\[
f(t) = \frac{1}{c} \sum_{i} I(t_{el} \leq t \leq t_{li})
\]

where the sum was taken over all pairs \( i \), \( I \) denotes an indicator variable that equals 1 if \( t_{el} \leq t \leq t_{li} \) and 0 otherwise, and \( c \) is a normalizing constant ensuring that \( f(t) \) sums to 1 when summed over all times \( t \). Additional analysis showed that the results were very similar when all nucleotide positions were included or when the HKY 85 substitution model (30) was chosen. However, the method described above produced the tightest posterior distributions for TMRCAs.

To determine whether invasion events had an effect on clade composition, the temporal change of the probability that an isolate was sampled from the North American clade was fitted by logistic regression using the statistical software R (31). The data were grouped by hemisphere.
and host type (wild versus domestic birds) and analyzed for each group separately since wild and domestic birds were sampled differently. The wild bird analysis was based on individual isolates and the domestic bird analysis on the number of introductions into the domestic poultry species. Because the analysis relied on publicly available data, only introductions into domestic poultry that led to a sequence submission were analyzed. H6 sequences from domestic bird isolates in different years were counted as separate introductions. For all neuraminidase (NA) sequences that were part of a virus isolate whose H6 gene was classified as invader, the clade membership of the NA gene was determined as well. In addition, the proportion of H6 among all hemagglutinin sequences (any length, but only 1 sequence per isolate) from long-term wild bird survey sites in Alberta, Canada, and Ohio and Delaware, USA (25), was calculated per decade. These sites were chosen because AIV isolates were sampled from these sites following a standardized sampling regimen for more than a decade (32).

Results

The phylogenetic tree of all 291 full-length H6 gene sequences shows a clear division between the Eurasian and North American clades (online Technical Appendix Figure 1, available from www.cdc.gov/EID/content/15/7/1040-Techapp.pdf). Within the Eurasian clade there were 7 monophyletic subclades composed of strains isolated from North America (Figure 2). Six of these 7 subclades had bootstrap support above 90%. Five of these 7 subclades were closely related to isolates in Eurasia. Of these 5 subclades, 3 were most closely related to isolates in East Asia, and 2 were most closely related to clades in Northern Europe. All North American isolates of the 2 subclades most closely related to clades in Europe were found on the East Coast or in the Midwest region of the United States. Most of the isolates from the subclades closely related to isolates in Asia were found in Alberta, Canada, and the West Coast of the United States. The exceptions were 1 isolate from New Jersey and 1 from Delaware, which clustered in subclades with close relatives in Asia (Figure 2).

The estimated entry times of these 7 subclades overlapped in part, with the earliest times starting in the late 1970s and the latest around 2005 (Figure 3). The proportion of North American clade members among isolates from North American wild birds decreased from 100% in the 1980s to 20% in the 2000s (Figure 4; p<0.0001, logistic regression, residual deviance = 75, df = 106). Similarly, all H6 introductions into poultry species in North America with sequences submitted to the public databases were caused by viruses from the North American clade before 1998 and by viruses from the Eurasian clade after 1999 (Figure 4). All 38 H6 sequences isolated in North America after 2002 belong to the Eurasian clade. The change in clade composition among North American viruses is visible across the entire North American continent (online Technical Appendix Figures 2, 3). One subclade invaded Australia from North America but did not result in a significant impact on the clade composition over time in Eurasia (Figure 4, panel B). The increase of Eurasian H6 among North American H6 isolates did not coincide with an overall increase of H6 prevalence among AIV isolates reported in wild birds from Eurasia (results not shown). All NA genes that were combined with invading H6 genes belonged to North American clades (online Technical Appendix Figures 4–7).

Discussion

A comprehensive phylogenetic analysis of all currently available full-length H6 gene sequences showed a major change in clade composition among isolates from North America. Seven subclades within the Eurasian clade were composed of isolates from North America. Isolates from these subclades started to appear in North America in the 1990s and became highly dominant among North
American isolates by the early 2000s. This pattern provides strong evidence for multiple invasions of North America by Eurasian H6 strains that led to the replacement of the North American H6 clade. Previous studies that addressed cross-hemisphere movement detected evidence for infrequent cross-hemisphere movement (6, 7), but did not show the change of clade composition over time.

The closest native Eurasian relatives of the invading subclades indicated movement of H6 to North America from Northern Europe and East Asia. In addition, some evidence indicates possible movement back from North America to Eurasia (e.g., subclades 6 and 7 as shown in Figure 2 could have resulted from 1 movement from Europe into North America and a second movement from North America back to Europe). Although the phylogenetic analysis cannot assess the actual number of invasion events, it nevertheless indicates that several invasion events occurred. A main limitation of the data is that the sampling system was not consistent across time, space, and host species. For example, no H6 samples from the West Coast of North America are available before 2000; therefore, some of the decline in the proportion of the North American clade among H6 viruses could be attributed to an increase of samples from the West Coast, which contained only members of the Eurasian clade. However, even when isolates are grouped by region, the decline of the North American clade can still be observed (online Technical Appendix Figure 2).

Overall, the patterns suggest frequent movement between the hemispheres in the 1990s and 2000s. The deep division between the Eurasian and North American clades would not be expected if this frequent cross-hemisphere movement had occurred in the distant past. We therefore propose that cross-hemisphere virus movement increased in the last few decades. A formal test of this hypothesis would require estimating migration rates over time from phylogenetic data (33–35), while taking the limitations of the data into account, which is beyond the scope of the present paper. Comparing estimated time-varying virus migration rates with data on bird trade and wild bird migration might shed light on the mechanisms of cross-hemisphere movement.

That the sequences of the invading viruses did not cluster in a single monophyletic group suggests that the invasion events were not triggered by a single genetic change in the H6 gene. The pattern observed in the H6 gene could have been caused by the H6 gene spreading along with the expansion of another gene segment that reassorts rarely with the H6 gene. The analysis of the NA genes shows that no concurrent movement of NA genes occurred. However, we have not analyzed the internal genes. For a more complete understanding of AIV movement, the analysis developed here must be applied to all subtypes and gene segments.

Whether similar invasion events occurred for other gene segments or subtypes is unknown. Another study found that among AIVs, viruses of the H6 subtype have wider host ranges than other AIV subtypes (13). This could

Figure 3. Posterior distribution of estimated invasion times of the 7 subclades of Eurasian avian influenza virus subtype H6 that invaded North America. Numbering of subclades corresponds to the order within the phylogeny of Figure 2 (top = 1, bottom = 7).

Figure 4. Proportion of viruses from the North American clade of avian influenza virus among wild bird isolates (blue) or poultry outbreaks (pink) in North America (A) or Eurasia (B). All poultry outbreaks in Eurasia were caused by viruses from the Eurasian clade (results not shown). Wild bird isolates are grouped by decade. Black circles show proportions of all isolates per decade (number of isolates per decade shown next to circles), black lines show fitted logistic regressions, and red circles show individual poultry outbreaks.
make H6 virus more prone to spread between and within hemispheres than other AIV subtypes.

Our study has shown that cross-hemisphere AIV movement can lead to a dramatic change in strain composition within a decade that affects an entire continent. We do not know whether expansion of the Eurasian H6 led to a change in risk for the poultry industry since H6 poultry isolates may not be well represented in public gene sequence databases. For example, H6 outbreaks have been documented in poultry operations in Minnesota in 12 different years from 1978 to 2005 (36), yet gene sequences of only 1 virus isolate from poultry outbreaks in Minnesota are available in GenBank. Although it is currently not possible to assess the risk posed to poultry by invading H6 AIV, the change in risk for the domestic poultry flocks clearly would be substantial if a similar invasion did occur for the highly pathogenic AIVs that are currently restricted to Eurasia.

This study corroborates a previous study (6) that found no evidence of cross-hemisphere invasions by entire virus genomes but detected only invasions of single gene segments that reassorted with other segments found in the new hemisphere. Thus, invasions can lead to new combinations of host range determining factors on different gene segments that were geographically separated before an invasion event. In general, invasions of the scope and speed as the instance documented here have the potential to strongly affect the risk posed by AIV to poultry and humans.

This work was funded by the Director for National Intelligence, Postdoctoral Research Fellowship Program (Washington, DC, USA).

Dr zu Dohna is a disease ecologist, interested in the spatial dynamics of diseases and their vectors, and a postdoctoral fellow at the Center for Animal Disease Modeling and Surveillance at the University of California, Davis, where he is working on foot-and-mouth disease and avian influenza.

References


Invasions by Eurasian Avian Influenza Virus H6 Genes and Replacement of Its North American Clade

Technical Appendix

Technical Appendix Figure 1. Complete H6 tree. Gray branches indicate isolates from Eurasia; black branches, isolates from North America.
Technical Appendix Figure 2. Proportion of viruses from the North American clade among wild bird isolates in 3 regions of North America grouped by decade. Lines show fitted logistic regressions. Triangles and dashed line show proportions among samples in the Northeast (Delaware and Maryland, \( p = 0.63, n = 17 \)), crosses and dotted lines show samples from the Midwest (Ohio and Minnesota, \( p < 0.01, n = 34 \)), circles and solid line show samples from Alberta (\( p < 0.05, n = 33 \)).

Technical Appendix Figure 3. Maps of H6 distributions in North America at different decades: 1980s (top), 1990s (middle), and 2000s (bottom). H6 from North American clade is shown in blue, from Eurasian clade in red.
Technical Appendix Figure 4. Complete phylogenetic tree of all N1 nucleotide sequences. The tree was built by the neighbor-joining algorithm. Green branches show isolates whose H6 gene invaded North America from Eurasia, red branches show isolates from Eurasia, and blue branches show isolates from North America. Virus names are not shown because of the large tree size.
Technical Appendix Figure 5. Complete phylogenetic tree of all N2 nucleotide sequences. The tree was built by the neighbor-joining algorithm. Green branches show isolates whose H6 gene invaded North America from Eurasia, red branches show isolates from Eurasia and blue branches show isolates from North America.
Technical Appendix Figure 6. Complete phylogenetic tree of all N5 nucleotide sequences. The tree was built by the neighbor-joining algorithm. Green branches show isolates whose H6 gene invaded North America from Eurasia, red branches show isolates from Eurasia, and blue branches show isolates from North America.
Technical Appendix Figure 7. Complete phylogenetic tree of all N8 nucleotide sequences. The tree was built by the neighbor-joining algorithm. Green branches show isolates whose H6 gene invaded North America from Eurasia, red branches show isolates from Eurasia and blue branches show isolates from North America.
Technical Appendix Table. Abbreviations of bird names and virus sampling locations used in avian influenza virus strain names

<table>
<thead>
<tr>
<th>Bird full name</th>
<th>Abbreviation</th>
<th>Location full name</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruddy turnstone</td>
<td>RT</td>
<td>Delaware</td>
<td>DE</td>
</tr>
<tr>
<td>Shorebird</td>
<td>SB</td>
<td>Sweden</td>
<td>SWD</td>
</tr>
<tr>
<td>Duck</td>
<td>DK</td>
<td>Hong Kong</td>
<td>HK</td>
</tr>
<tr>
<td>Laughing gull</td>
<td>LG</td>
<td>New Jersey</td>
<td>NJ</td>
</tr>
<tr>
<td>Mallard</td>
<td>MLD</td>
<td>Alberta</td>
<td>ALB</td>
</tr>
<tr>
<td>Pintail duck</td>
<td>PTD</td>
<td>Ohio</td>
<td>OH</td>
</tr>
<tr>
<td>Green-winged teal</td>
<td>GWT</td>
<td>California</td>
<td>CA</td>
</tr>
<tr>
<td>Chicken</td>
<td>CK</td>
<td>Hokkaido</td>
<td>HOK</td>
</tr>
<tr>
<td>Japanese quail</td>
<td>JQL</td>
<td>Alaska</td>
<td>AK</td>
</tr>
<tr>
<td>Swan</td>
<td>SW</td>
<td>Shimane</td>
<td>SHM</td>
</tr>
<tr>
<td>American widgeon</td>
<td>AWG</td>
<td>Maryland</td>
<td>MD</td>
</tr>
<tr>
<td>Ring-necked duck</td>
<td>RND</td>
<td>Minnesota</td>
<td>MN</td>
</tr>
<tr>
<td>Dunlin</td>
<td>DL</td>
<td>The Netherlands</td>
<td>NETH</td>
</tr>
<tr>
<td>Gadwall</td>
<td>GDW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoveler</td>
<td>SHV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue-winged teal</td>
<td>BWT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snow goose</td>
<td>SGS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guillemot</td>
<td>GM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>