



Frequent inter-species transmission and geographic subdivision in avian influenza viruses from wild birds

Rubing Chen^a, Edward C. Holmes^{a,b,*}

^a Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA

^b Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:

Received 14 August 2008

Returned to author for revision

22 September 2008

Accepted 1 October 2008

Available online 8 November 2008

Keywords:

Avian influenza virus

Migration

Phylogeny

Population structure

Epidemiology

ABSTRACT

Revealing the factors that shape the genetic structure of avian influenza viruses (AIVs) in wild bird populations is essential to understanding their evolution. However, the relationship between epidemiological dynamics and patterns of genetic diversity in AIV is not well understood, especially at the continental scale. To address this question, we undertook a phylogeographic analysis of complete genome sequences of AIV sampled from wild birds in North America. In particular, we asked whether host species, geographic location or sampling time played the major role in shaping patterns of viral genetic diversity. Strikingly, our analysis revealed no strong species effect, yet a significant viral clustering by time and place of sampling, as well as the circulation of multiple viral lineages in single locations. These results suggest that AIVs can readily infect many of the bird species that share breeding/feeding areas.

© 2008 Elsevier Inc. All rights reserved.

Introduction

Avian influenza viruses (AIVs) infect a wide variety of bird species in nature. Low pathogenic avian influenza (LPAI) viruses have been isolated from at least 105 wild bird species (Olsen et al., 2006), with only occasional disease (Fouchier et al., 2005; van Gils et al. 2007; Webster et al., 1992). AIVs are most commonly detected in those bird species that inhabit wetland and aquatic environments, namely *Anseriformes* (particularly ducks, geese and swans) and *Charadriiformes* (gulls, terns and shorebirds), which are thought to represent their natural reservoirs (Olsen et al., 2006). All of the viral subtypes (hemagglutinin H1-16; neuraminidase N1-9) and at least 84 serotypes (HA-NA combinations) have been found in these species (Dugan et al., 2008; Munster et al., 2007). It is likely that the relatively high prevalence of AIV in aquatic birds (about 9.5% in ducks; Olsen et al. 2006) is related to efficient transmission through fecally contaminated water (Webster et al., 1978, 1992). Sharing a common water resource may therefore facilitate the dissemination of AIVs among diverse bird species.

Long distance migration is a common phenomenon in *Anseriform* and *Charadriiform* birds (Cramp and Perrins, 1985–1994; Poole et al., 1993–2002), and has been proposed as a major factor determining the genetic structure of AIV populations (Garamszegi and Moller, 2007; Ito et al., 1995; Krauss et al., 2004; Olsen et al., 2006; Webster et al., 1992; Winker et al., 2007). Broadly speaking, bird migration normally

exhibits a north-and-south pattern, following coasts, mountain ranges and principal river valleys. In North America, four major flyways have been identified – the Atlantic, the Mississippi, the Centre and the Pacific – each of which is composed of numerous individual migration routes. With the exception of coastal regions, these flyways do not have clear boundaries and overlap to some extent in both the northern breeding and southern wintering lands (<http://www.birdnature.com/flyways.html>). As migrating animals often form foraging flocks of mixed species, with different migrants or residential species (Rappole and Jones, 2002), overlapping regions of migration pathway, such as breeding area, wintering area, as well as stop-over points, all provide the opportunity for virus transmission among a range of bird species (Garamszegi and Moller, 2007; Stallknecht et al., 1990).

Surveillance data have sketched broad-scale epidemiological patterns in AIV. In particular, there are strong species, temporal and spatial differences in prevalence. For example, dabbling ducks of the genus *Anas*, particularly mallards (*Anas platyrhynchos*), are associated with higher AIV prevalence than other birds (Olsen et al., 2006). This may be caused by different water salinity and their habit of surface water feeding (Garamszegi and Moller, 2007). As a particular case in point, in a surveillance study undertaken in Sweden, mallards and teals had a higher prevalence of virus than widgeons, pintails, gadwalls and shovelers combined (Munster et al., 2007). These pronounced differences in prevalence suggest that while many species can harbor AIVs, most contribute little to their evolution (Munster et al., 2007).

The prevalence of AIV also exhibits a temporal pattern relative to the timing of migration. In particular, the prevalence of AIVs in ducks

* Corresponding author. Fax: +1 814 865 9131.

E-mail address: ech15@psu.edu (E.C. Holmes).

in North America varies from less than 1% during spring migration, to 30% prior to and during the fall migration (Munster et al., 2007), the latter reflecting the birth of a large number of naïve birds during the breeding season. Indeed, there is a significantly higher prevalence of AIV in young compared to adult birds (Hanson et al., 2003; Hinshaw et al., 1978). For instance, the peak prevalence of AIV in blue-winged teals from coastal Louisiana was in September, while that of gadwalls and green-winged teals was in early November, evidently reflecting their different migration times (Stallknecht et al., 1990). Migration may also introduce new serotypes into specific geographic area. In Minnesota, USA, for example, observation between 1998 and 2000 revealed that some subtypes did not persist between years, while new serotypes emerged with the arrival of migratory birds (Hanson et al., 2003). However, it is unclear whether the observation of the same serotype over multiple years of sampling is due to its local persistence, or that it is imported newly each season. Finally, it is also evident that ecological and geographic barriers, including those associated with the division of the major migration flyways, also impact on the genetic structure of AIVs. This is most apparent in the major phylogenetic subdivision between the Eurasian/Australian and American AIV lineages, which experience little genetic exchange (Dugan et al., 2008; Krauss et al., 2007).

Despite extensive epidemiological work on the contribution of bird migration to the transmission and maintenance of AIVs, little is known about how these epidemiological dynamics relate to patterns of genetic diversity. In particular, the respective contributions of host species, geographical location and sampling time in shaping the genetic diversity of AIV have not been fully elucidated. Importantly, the increasing numbers of whole genome AIV sequences available on public data bases such as GenBank make it possible to address this question in a detailed manner. To this end, we used the available data base of AIV sequences sampled from North American bird species, totaling approximately 290 complete genomes, to explore key aspects of their phylogeography.

Results and discussion

In accord with surveillance data, most of the viral genomes analyzed here were sampled from mallard ($n=151$), pintail duck ($n=35$), blue-winged teal ($n=24$) and green-winged teal ($n=18$) species (Table 1). The ML tree of the representative (longest) PB2 segment, color-coded by each variable (host species, geographic region, time of sampling), is shown in Fig. 1, while the geographically-coded trees for both PB2 and HA are shown in Fig. 2. Phylogenetic trees for all other segments are available in the Supplementary Information.

The results of our analysis of how genetic variation is partitioned according to host species, geographical location and sampling time are summarized in Table 2 (we excluded subtype H13 and H16 viruses because all available samples of these subtypes have at least one segment of Eurasian/Australian origin – see Dugan et al., 2008). The most striking observation from this analysis was the absence of any strong species effect on genetic diversity. Specifically, the number of observed changes in character state based on host species in each segment tree – 66–72 for the internal genes and 67 and 70 for the HA and NA segments, respectively – either falls in, or approaches, the range of state changes observed in 1000 random trees (71–79 and 67–75 for internal genes and glycoproteins, respectively), so that there is no more clustering by host species than expected by chance alone. Hence, there appears to be no major barrier to AIV transmission among those bird species studied here. To ensure that these results were not biased by the inclusion of species that were represented only a few times in the data set (coot, canvasback, pheasant, pigeon, etc), we repeated the analysis using only those species for which more than three sequences were available. Again, the range of observed state changes (106–110 for internal genes and 100–102 for HA and NA, respectively) overlaps with that of random trees (106–119 for internal genes and 98–111 for HA and NA, respectively), supporting our conclusion that there is no strong species clustering in these data.

Table 1
Sizes of AIV data sets by geographic location, host species and year of sampling

Location	Internal segments (290)	HA/NA (281)	Species	Internal segments (290)	HA/NA (281)	Year	Internal segments (290)	HA/NA (281)
Alberta	121	116	Mallard	151	149	2005	32	32
Ohio	92	92	Pintail	35	33	2004	5	5
Maryland	18	18	Blue wing teal	24	23	2003	7	7
New York	15	15	Green-winged teal	18	18	2002	23	23
Delaware	12	11	(American) black duck	8	8	2001	5	5
Minnesota	11	9	Chicken	7	6	2000	5	5
Alaska	10	10	Shorebird	5	5	1999	12	11
New Jersey	4	4	Ruddy turnstone	5	5	1998	11	11
Mexico	3	2	Turkey	5	4	1997	3	3
Tennessee	2	2	Laughing gull	3	3	1996	4	4
Pennsylvania	1	1	Snow goose	3	3	1995	7	6
Missouri	1	1	(Northern) shoveler	3	3	1994	2	2
			Duck	3	2	1993	7	7
			Longtail duck	2	2	1992	4	3
			Gadwall	2	2	1991	5	4
			Guinea fowl	2	2	1990	7	7
			Widgeon	2	2	1989	12	12
			Redhead duck	2	2	1988	15	14
			(Red) knot	2	1	1987	15	15
			Common scoter	1	1	1986	17	17
			Bufflehead	1	1	1985	19	18
			Coot	1	1	1984	4	4
			Canvasback	1	1	1983	10	9
			Pheasant	1	1	1982	11	11
			Pigeon	1	1	1981	3	2
			Sanderling	1	1	1980	6	6
			Wood duck	1	1	1979	12	12
						1978	9	9
						1977	17	16
						1976	1	1

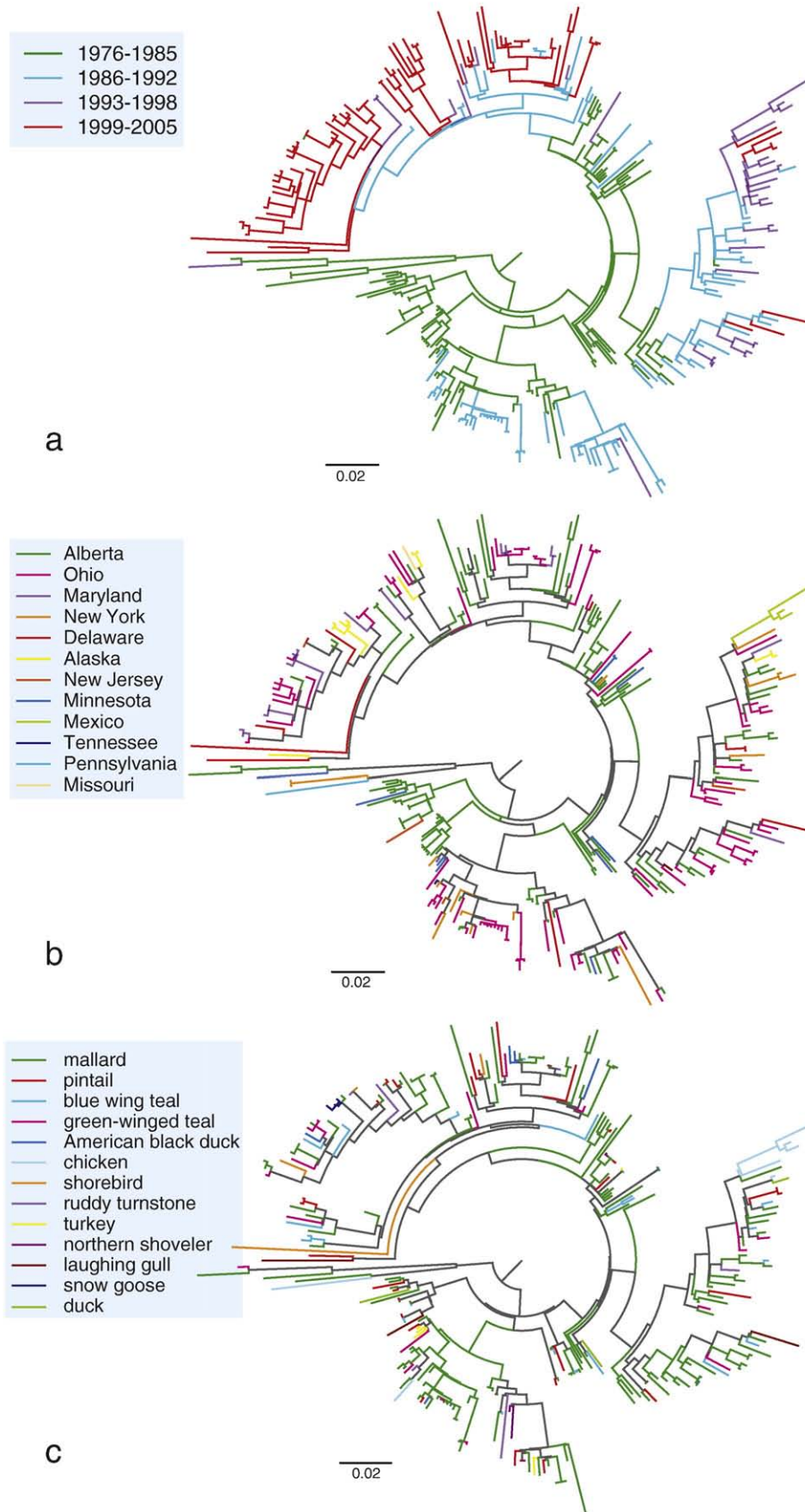


Fig. 1. Distribution (color-coded) of sampling (a) times, (b) locations, and (c) species in the maximum likelihood tree of the representative PB2 segment of avian influenza A viruses. Similar patterns are seen in the other gene segments (see Supplementary Information). All trees are unrooted, drawn in circular orientation, and all horizontal branch lengths have been drawn to a scale of nucleotide substitutions/site.

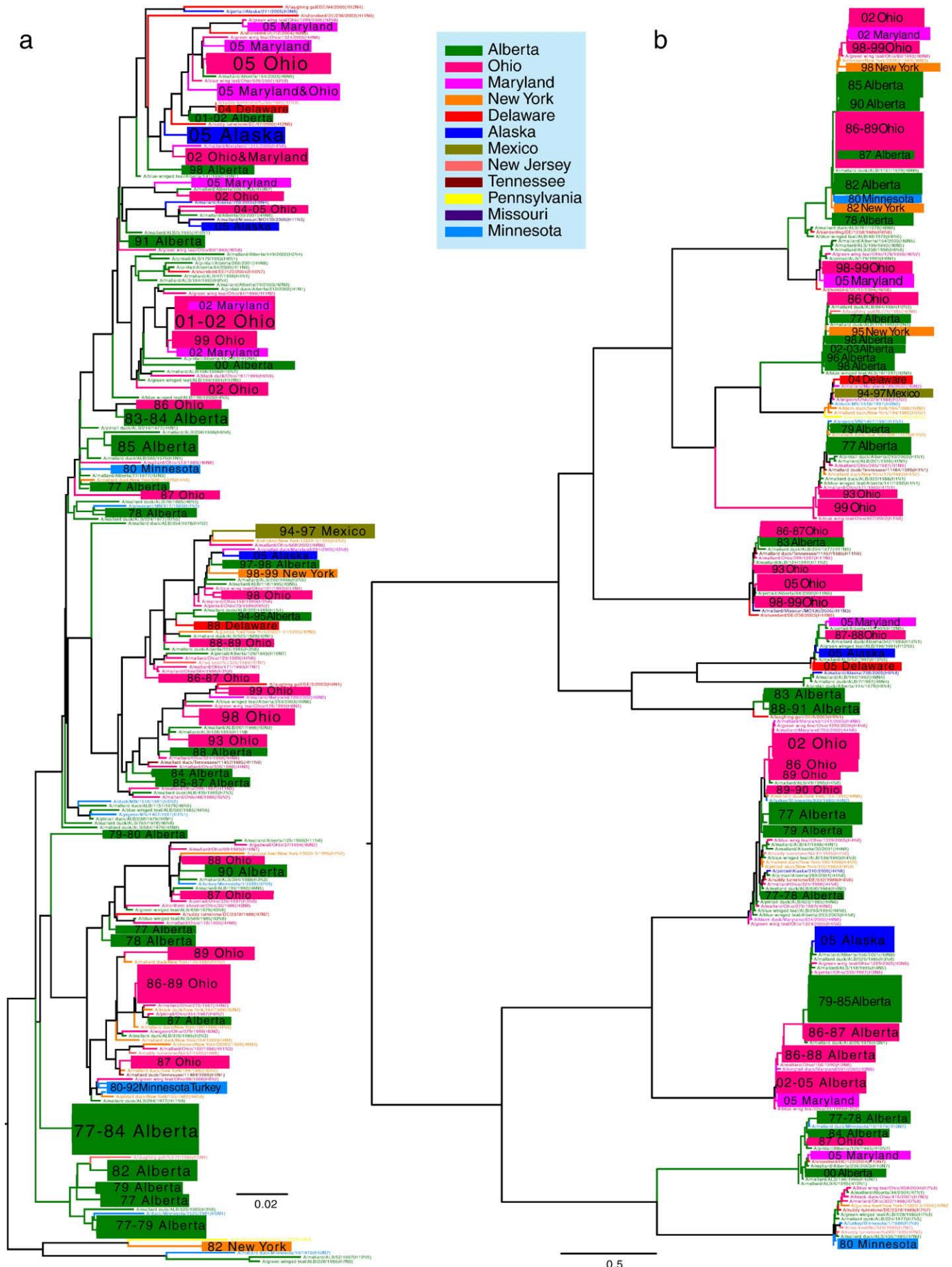


Table 2

Number of observed and expected (random) character state changes by sampling time, sampling location and host species

Factor (Character no.)	Data set	No. of state changes	
		ML tree range	Random tree range
Time (30)	Internal segments	143–163	220–246
	HA/NA	157–160	216–237
Location (12)	Internal segments	85–97	133–157
	HA/NA	84–90	129–151
Species (27)	Internal segments	66–72	71–79
	HA/NA	67–70	67–75
Main host species ^a (13/12)	Internal segments	106–110	106–119
	HA/NA	100–102	98–111

^a Species that are represented by more than three sequences.

In marked contrast, these AIV sequences were significantly clustered according to both year of sampling and geographic location as the number of observed changes in character state was far fewer than that expected by chance alone ($p < 0.001$ in both cases, calculated as $p < 1/1000$ as there was no overlap between the numbers of state changes between the ML and 1000 random trees) (Table 1). In addition, as the ratio of observed to expected state changes are essentially identical for both time and space (0.54–0.73), both factors seem to have had a similar effect on the genetic structure of AIV. Hence, viral strains sampled in the same place or the same time (year) tend to be related. The strong clustering by sampling time further indicates that there is a marked turnover of viral lineages through time, supporting suggestions that AIVs exhibit relatively rapid evolutionary dynamics (Chen and Holmes, 2006).

As there was an evident bias in how the viral samples analyzed here were collected, with substantial variation between specific places and years, it is important to exclude that the significant clustering in time and space we observed was not simply a consequence of uneven sampling. To test this, we repeated the analysis using only those samples collected between 1986–2000, which depict a more even temporal and spatial sampling (Table 1). We obtained results highly consistent with those for the analysis of the whole data set, suggesting that our conclusions are robust (results not shown, available from the authors on request).

Due to the relatively small sample size of AIV genomes available for analysis, our phylogenetic analysis does not provide sufficient resolution to reveal the geographical direction of viral transmission, although some patterns are apparent. For example, the strongest signals of (bi-directional) viral migration were between Ohio and Maryland (observed – expected number of character state changes = 15.46), Alberta and Minnesota (7.31), and Delaware and New Jersey (4.94) (Supplementary Table 1). These results therefore suggest that the strongest phylogenetic signal for migration is among geographic localities that are relatively close in space. However, it is clear that sequence data from a far larger sample of localities across North America are required to fully understand the spatial dynamics of AIV.

Finally, it is important to note that most of the geographic clusters observed in our trees covered samples collected during time-spans of only 1 or 2 years. This strongly suggests that viruses do not persist within specific localities, but rather are regularly imported into populations. A similar dominance of migration over persistence has recently been observed in human influenza A virus (Nelson et al., 2007), although within the context of a source-sink ecological model (Rambaut et al., 2008; Russell et al., 2008). Hence, despite the very

different mechanisms, humans and avian influenza viruses both exhibit very fluid dynamics. Similarly, multiple AIV lineages are able to co-circulate within individual geographic localities during single years. For example, in the PB2 phylogeny, four separate clusters of viruses are detected in both Ohio 2002 and in Maryland during 2005 (Fig. 2). Such co-circulation of different viral strains provides opportunity for both mixed infection and reassortment, which appears to be common in AIV (Dugan et al., 2008). More generally, our study suggests that when an influenza virus is introduced into a specific location at a specific time it can infect many of the species breeding/feeding there. This is analogous to the 'epizootiological compartments' that have been observed in the phylogeography of rabies virus (Bourhuy et al., 1999).

Methods

Because of the very large number of AIV complete genome sequences now available, we used those viruses sampled from wild bird species in North America as a model system to investigate viral phylogeography. These sequence data are same as those used in the recent study of Dugan et al. (2008) which considered other aspects of AIV evolution. Hence, any strains which have a Eurasian/Australia origin (as determined by provisional phylogenetic analysis) were excluded to avoid automatic clustering by geographic segregation. All AIV genome sequences were downloaded from NCBI Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>). Sequence alignments for each gene segment were created using MUSCLE (Edgar, 2004) and then adjusted manually using Se-Align (Rambaut, 1996) according to amino acid sequences. This resulted in data sets of the following size: PB2, PB1, PA, M1/2, NP, NS1/2 = 290 sequences; HA and NA = 281 sequences. In total, these genome data sets represent 12 HA subtypes (H1–H12), 9 NA subtypes (N1–N9), and 56 serotypes (HA–NA combinations). It also includes samples from 27 species, 24 locations (US States, Canadian Provinces, Mexico), and 30 years (1976 to 2005) (Table 1). Unrooted maximum likelihood (ML) phylogenetic trees were then inferred using the PAUP* (version 4.0b10) package (Swofford, 2003), based on the best-fit models of nucleotide substitution determined by MODELTEST (Posada and Crandall, 1998) (details available from the authors on request).

To determine which biological factors play the most important roles in determining the observed phylogenetic patterns – namely host species, place of sampling and time of sampling – we utilized a parsimony-based character-mapping approach (Nakano et al., 2004). For a particular data set, each viral isolate was first assigned a character symbol according to the variable (species, place, time) in question. Given the (unrooted) ML phylogeny for these sequences (as determined above), the minimum number of changes in character state needed to give rise to the observed distribution of character states was then estimated using a parsimony procedure (with all ambiguous changes excluded). To determine the number of changes expected under the null hypothesis of complete mixing among states by space or time, the states of all isolates were randomized on the ML trees 1000 times, and for each randomization the number of changes in state was calculated in the manner described above. The difference between the mean number of observed and expected changes for each pair of states indicates the level of species, geographic or temporal isolation, with statistical significance calculated by comparing the total number of observed state changes to the number expected under random mixing. All analyses were undertaken using PAUP* (Swofford, 2003).

Fig. 2. Maximum likelihood trees of North American AIVs based on the representative (a) PB2 segment ($n = 290$) and (b) HA ($n = 281$) segments. Taxa and terminal branches are colored by sample location. Colored blocks represent the clusters in which samples come from the same location within a certain range of time. The tree is mid-point rooted (for purposes of clarity only) and all horizontal branch lengths have been drawn to a scale of nucleotide substitutions/site.

References

- Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J.F., Holmes, E.C., 1999. Ecology and evolution of rabies virus in Europe. *J. Gen. Virol.* 80, 2545–2558.
- Chen, R., Holmes, E.C., 2006. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* 23, 2336–2341.
- Cramp, S., Perrins, C.M., 1985–1994 Cramp, S., Perrins, C.M., (Eds.) 1985–1994. *The Birds of the Western Palearctic*. Oxford, UK: Oxford University Press.
- Dugan, V.G., Chen, R., Spiro, D.J., Sengamalay, N., Zaborsky, J., Ghedin, E., Nolting, J., Swayne, D.E., Runstadler, J.A., Happ, G.M., Senne, D.A., Wang, R., Slemmons, R.D., Holmes, E.C., Taubenberger, J.K., 2008. The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* 4, e1000076.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Fouchier, R.A., Munster, V., Wallensten, A., Bestebroer, T.M., Herfst, S., Smith, D., Rimmelzwaan, G.F., Olsen, B., Osterhaus, A.D., 2005. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* 79, 1822–2814.
- Garamszegi, L.Z., Moller, A.P., 2007. Prevalence of avian influenza and host ecology. *Proc. R. Soc. B* 274, 2003–2012.
- Hanson, B.A., Stallknecht, D.E., Swayne, D.E., Lewis, L.A., Senne, D.A., 2003. Avian influenza viruses in Minnesota ducks during 1998–2000. *Avian Dis.* 47, 867–871.
- Hinshaw, V.S., Webster, R.G., Turner, B., 1978. Novel influenza A viruses isolated from Canadian feral ducks: including strains antigenically related to swine influenza (Hsw1N1) viruses. *J. Gen. Virol.* 41, 115–127.
- Ito, T., Okazaki, K., Kawaoka, Y., Takada, A., Webster, R.G., Kida, H., 1995. Perpetuation of influenza-A viruses in Alaskan waterfowl reservoirs. *Arch. Virol.* 140, 1163–1172.
- Krauss, S., Obert, C.A., Franks, J., Walker, D., Jones, K., Seiler, P., Niles, L., Pryor, S.P., Obenauer, J.C., Naeve, C.W., Widjaja, L., Webby, R.J., Webster, R.G., 2007. Influenza in migratory birds and evidence of limited international virus exchange. *PLoS Pathog.* 3, e167.
- Krauss, S., Walker, D., Pryor, S.P., Niles, L., Li, C.H., Hinshaw, V.S., Webster, R.G., 2004. Influenza A viruses of migrating wild aquatic birds in North America. *Vector Borne Zoonotic Dis.* 4, 177–189.
- Munster, V.J., Baas, C., Lexmond, P., Waldenstrom, J., Wallensten, A., Fransson, T., Rimmelzwaan, G.F., Beyer, W.E., Schutten, M., Olsen, B., Osterhaus, A.D., Fouchier, R.A., 2007. Spatial, temporal, and species variation in prevalence of influenza A virus in wild migratory birds. *PLoS Pathog.* 3, e61.
- Nakano, T., Lu, L., Liu, P., Pybus, O.G., 2004. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J. Infect. Dis.* 190, 1098–1108.
- Nelson, M.I., Simonsen, L., Viboud, C., Miller, M.A., Holmes, E.C., 2007. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog.* 3, e131.
- Olsen, B., Munster, V.J., Wallensten, A., Waldenstrom, J., Osterhaus, A.D., Fouchier, R.A., 2006. Global pattern of influenza A virus in wild birds. *Science* 312, 384–388.
- Poole, A., Stettenheim, P., Gill, F., 1993–2003 Poole, A., Stettenheim, P., Gill, F., (Eds.) 1993–2003. *The Birds of North America*. Philadelphia, PA: The American Ornithologists' Union and The Academy of Natural Sciences of Philadelphia.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rambaut, A., 1996. Se-AI: Sequence Alignment Editor. Available at <http://evolve.zoo.ox.ac.uk/>.
- Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453, 615–619.
- Rappole, J.H., Jones, P., 2002. Evolution of old and new world migration systems. In: Both, C., Piersma, T. (Eds.), *The Avian Calendar: Exploring Biological Hurdles in the Annual Cycle*. Proc. 3rd Conf. European Orn. Union, Groningen, 2001. *Ardea* 90, pp. 525–537.
- Russell, C.A., Jones, T.C., Barr, I.G., Cox, N.J., Garten, R.J., Gregory, V., Gust, I.D., Hampson, A.W., Hay, A.J., Hurt, A.C., de Jong, J.C., Kelso, A., Klimov, A.I., Kageyama, T., Komadina, N., Lapedes, A.S., Lin, Y.P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A.D., Rimmelzwaan, G.F., Shaw, M.W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R.A., Smith, D.J., 2008. The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320, 340–346.
- Stallknecht, D.E., Shane, S.M., Zwank, P.J., Senne, D.A., Kearney, M.T., 1990. Avian influenza viruses from migratory and resident ducks of coastal Louisiana. *Avian Dis.* 34, 398–405.
- Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*And Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
- van Gils, J.A., Munster, V.J., Radersma, R., Liefhebber, D., Fouchier, R.A., Klaassen, M., 2007. Hampered foraging and migratory performance in swans infected with low-pathogenic avian influenza A virus. *PLoS ONE* 2, e184.
- Webster, R.G., Yakhno, M., Hinshaw, V.S., Bean, W.J., Murti, K.G., 1978. Intestinal influenza: replication and characterization of influenza viruses in ducks. *Virology* 84, 268–278.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawoka, Y., 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56, 152–179.
- Winker, K., McCracken, K.G., Gibson, D.D., Pruett, C.L., Meier, R., Huettmann, F., Wege, M., Kulikova, I.V., Zhuravlev, Y.N., Perdue, M.L., Spackman, E., Suarez, D.L., Swayne, D.E., 2007. Movements of birds and avian influenza from Asia into Alaska. *Emerg. Infect. Dis.* 13, 547–552.