Evolutionary Dynamics of Avian Influenza A Virus in the Natural Reservoir

J Bahl[1], D Vijaykrishna[1], GJD Smith[1], RG Webster[2], Y Guan[1]

[1]State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; [2]Virology Division, Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

Avian influenza viruses were thought to exist in a form of evolutionary stasis within their natural reservoirs, i.e. waterfowls. However, a recent study demonstrated very high evolutionary rates, with epidemic-like population growth, for individual influenza subtypes in both aquatic birds and poultry, suggesting the stasis theory may be incorrect. Yet the evolutionary dynamics of the influenza gene pool within one species of migratory waterfowl remains unclear. We therefore tested influenza virus population behavior by estimating rates of nucleotide substitution of the internal genes from different subtypes of influenza viruses exclusively from mallard ducks in North America. This dataset represents a co-circulating influenza gene pool in the natural reservoir from, importantly, a region without long-term prevalence of influenza in poultry populations. Our results showed high substitution rates for each internal gene examined ($>10^{-3}$ substitutions/site/year), in line with the results of previous studies. Furthermore, the majority of these substitutions were neutral with little evidence of positive selection. To evaluate the evolutionary stasis hypothesis, we compared the evolutionary dynamics of influenza hemagglutinin genes from the gene pool (mallards) that are not associated with disease in poultry, human or other animal populations (H4 subtypes) or with those that may be subsequently introduced and linked with disease outbreaks in poultry (H6 subtypes). Even though both subtypes share the same host species, analyses revealed that the mean evolutionary rate of the HA of H6 viruses was almost two-fold greater than that seen in H4. Although there was little site-specific positive selection in both H4 and H6 subtypes, the ratio of d_N/d_S was much greater for H6 isolates than those of H4 isolates, indicating that a higher proportion of mutations in the H6 subtype were nonsynonymous. These findings suggest that H4 isolates are under strong purifying selection, while H6 isolates are under Darwinian selection. Thus, high substitution rates seen in influenza viruses in mallard likely represent the neutral rate of evolution of the influenza gene pool in its natural host and that evolutionary stasis does not describe the dynamics of all influenza viruses in their natural reservoirs.

Introduction
Avian influenza A viruses (AIV) infect a variety of host species including humans, pigs, poultry, shorebirds, whales and seals. Migratory waterfowl are believed to be the natural hosts and AIV were thought to exist in a form of evolutionary stasis within this natural reservoir [1,2]. However, a recent study demonstrated very high evolutionary rates, with epidemic-like

population behavior, for individual influenza subtypes from multiple hosts, suggesting the stasis theory is incorrect [3]. Yet the evolutionary dynamics of the influenza gene pool in the natural reservoir remains unclear. In the present study we took two approaches to examine the stasis hypothesis. First we examined the gene pool behavior by comparing the population dynamics of AIV in its natural gene pool with that of disease associated poultry populations. We then investigated the evolutionary dynamics of the hemagglutinin sequences from H4 and H6 subtype viruses isolated from mallards. H4 subtype viruses are rarely isolated from poultry, humans or other animal species and have not been linked to disease outbreaks, while it is commonly isolated from the natural reservoir. In contrast, H6 subtypes are commonly isolated from mallard hosts but they have also been linked to disease outbreaks in poultry.

Methods
All data analyzed for this study was downloaded from GenBank. More than 2,000 sequences were analyzed for this study. For the analysis of the population dynamics of the influenza gene pool in its natural hosts, datasets for each internal gene (PB1, PB2, PA, M, NP and NS) isolated from all wild waterfowl were compared with results from datasets consisting of viral sequences from each host. In the current study the results of population dynamics of wild waterfowl and mallards are discussed, and compared with the population dynamics of disease associated AIV. To test the evolutionary dynamics of the surface hemagglutinin protein in its natural hosts, datasets of the H4 and H6 hemagglutinin subtypes were also analyzed. Dated phylogenies, Bayesian skyline demographic plots and substitution rates with associated highest posterior density (HPD) were determined with BEAST 1.4.2 [5,6,]. Demographic growth models were tested using the dated phylogenies generated in BEAST with GENIE [6]. Growth models tested were constant, exponential, expansion, logistic, and piecewise constant. The best-fit model was determined using the Akaike information criterion. DATAMONKEY [7] and HYPHY [8] were used to test for site-specific positive selection and determine the global d_N/d_S based on an appropriate nucleotide substitution model for each dataset analyzed.

Results
Population dynamics in natural hosts vs. poultry. Utilizing data from wild birds isolated from North America [4], we estimated and compared the population dynamics and substitution rates of all AIV isolated from mallard hosts. Analysis of the population dynamics for the internal genes from all waterfowl is similar to those isolated only from mallards. This indicates that AIV isolated from mallards are part of the same natural gene pool found in all aquatic waterfowl and provides a suitable model for the entire natural gene pool (data not shown). However, the populations of AIV from poultry hosts over the thirty-year time period sampled, are not stable and show high degrees of fluctuation (data not shown). These results show that the AIV population in domestic poultry is not stable and likely

results from repeated introductions from the natural gene pool and therefore can be excluded from subsequent analyses. *Population growth models.* Our results show that the internal genes of AIV have high substitution rates (>10^{-3} substitutions/site/year) in its natural reservoir hosts (Table 1). This result is similar to previous studies for influenza viruses [3], and for all RNA viruses [9]. However, most of these mutations are neutral; as indicated by the low number of nonsynonymous changes (d_N) over the thirty-year sampling period. The population growth models show that, while the population is stable, there are still changes over time.

Table 1. Population dynamics of AIV in Mallard.

Gene Segment	Population growth model	Mean substitution rate (10^{-3} substitutions/site/year)	HPD range	d_N/d_S
H4	Constant	5.268	2.627-8.083	0.079
H6	Constant	9.508	6.604-12.68	0.144
PB2	Logistic	2.526	1.712-3.383	0.03
PB1	Logistic	3.986	3.012-4.946	0.02
PA	Constant	3.088	2.207-3.975	0.02
NP	Piecewise constant	3.104	2.181-3.967	0.02
M	Logistic	2.076	1.475-2.682	0.02
NS	Logistic	3.791	2.489-5.162	N/A*

Population dynamics of the H4 and H6 hemagglutinin genes isolated from mallard show the same population growth patterns. However, the mean substitution rate is much higher for H6 subtypes and the d_N/d_S is nearly twice as high as compared to H4 subtypes. Positive selection analysis of the H4 dataset failed to detect any sites under positive selection and only one site (aa 154) for the H6 dataset (p = 0.1).
Dated phylogeny. Figure 1 shows a dated phylogeny reconstructed from the analysis of the hemagglutinin (HA) of H4 and H6 subtype viruses isolated from mallards. H6 viruses show periods of rapid diversification and high evolutionary rates (Fig. 2A). The tree topology and the constant population growth model suggest that H6 viruses are subjected to continual Darwinian evolution with high rates of lineage extinction and that surviving lineages are regularly selected for. This result is contrary to that seen in the dated phylogeny of H4 subtype viruses where substitution rates along lineages are low (Fig. 2B), suggesting H4 viruses have been adapted to mallard hosts for a longer period.

Discussion
The dataset analyzed represents a co-circulating influenza gene pool in the natural reservoir from, importantly, a region with no long-term endemicity of influenza viruses in terrestrial poultry populations. Our results showed high substitution rates for each gene examined (>10^{-3} substitutions/site/year), in line with the results of previous studies [3,4]. These substitutions were neutral and mutations along a lineage showed little evidence of positive selection. Results indicate that populations of H4 and H6 subtypes of influenza viruses in mallard duck were under strong purifying selection pressure (d_N/d_S <0.2); in contrast to the high background mutation rate. Variation in tree topologies,

d_N/d_S and evolutionary rate suggests that antigenic drift may explain the evolution of the faster evolving H6 subtypes, but is inadequate to explain the changes of H4 viral population within its natural host. The naturally high substitution rate of H6 subtypes in mallard results in a rapid accumulation of genetic changes and therefore accumulatation of many more nonsynonymous changes over time. These results indicate that the high substitution rates seen in influenza viruses in mallard, which are consistent with other RNA viruses, likely represent the neutral rate of evolution of the influenza gene pool in its natural host.

References
1. Gorman OT, et al. Evolution of the nucleoprotein gene of influenza A virus. *J Virol.* 1990;64:1487-1497.
2. Webster RG Predictions for future human influenza pandemics. *J Infect Dis.* 1997;176(Suppl 1):S14-19.
3. Chen R, Holmes EC Avian influenza exhibits rapid evolutionary dynamics. *Mol Biol Evol.* 2006;23:2336-2341.
4. Obenauer JC, et al. Large-Scale Sequence Analysis of Avian Influenza Isolates. *Science.* 2006;311:1576-1580
5. Drummond AJ, et al. Relaxed phylogenetics and dating with confidence. *PLoS Biology.* 2006;4:e88.
6. Drummond AJ, et al. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 2005;22:1185-1192.
7. Kosakovsky Pond SL, Frost SDW. Not so Different after all: a comparison of methods for detecting amino-acid sites under selection. *Mol Biol Evol.* 2005;22(5):1208-1222.
8. Kosakovsky Pond SL, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21(10):2531-2533.
9. Jenkins GM, et al. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol.* 2002;54:156-165.

Figure 1. Dated phylogeny of the HA gene of 1A) H6 and 1B) H4 subtype viruses. grey bars represent 95% HPDs. Numbers shown on branches represent the evolutionary rates along each lineage (10^{-3} Substitutions/Site/Year). Thickened branches shown in black indicate evolutionary rates along diversifying branches of H6 that are much greater than evolutionary rates seen in H4 during similar time periods. Numbers in bold and italicized are evolutionary rates significantly greater than the mean evolutionary rates for H4 shown in Table 1.