

RESEARCH ARTICLE

Open Access



# Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry

Mathieu Fourment\*  and Edward C Holmes

## Abstract

**Background:** Wild birds are the major reservoir hosts for influenza A viruses, occasionally transmitting to other species such as domesticated poultry. Despite an abundance of genomic data from avian influenza virus (AIV), little is known about whether AIV evolves differently in wild birds and poultry, although this is critical to revealing the dynamics and time-scale of viral evolution. In particular, because environmental (water-borne) transmission is more common in wild birds, which may reduce the number of replications per unit time, it is possible that evolutionary rates are systematically lower in wild birds than in poultry.

**Results:** We estimated rates of nucleotide substitution in two AIV subtypes that are strongly associated with infections in wild birds – H4 and H6 – and compared these to rates in the H5N1 subtype that has circulated in poultry for almost two decades. Our analyses of three internal genes confirm that H4 and H6 viruses are evolving significantly more slowly than H5N1 viruses, suggesting that evolutionary rates of AIV are reduced in wild birds. This result was verified by the analysis of a poultry-associated H6 lineage that exhibited a markedly higher substitution rate than those H6 viruses circulating in wild birds. Interestingly, we also observed a significant difference in evolutionary rate between H4 and H6, despite frequent reassortment rate among them.

**Conclusions:** AIV experiences markedly different evolutionary dynamics between wild birds and poultry. These results suggest that rate heterogeneity among viral subtypes and ecological groupings should be taken into account when estimating evolutionary rates and divergence times.

## Background

Influenza A viruses infect a wide range of animals including humans, pigs, horses, and poultry. However, with the exception of a small number of viruses recently described in bats [1,2], wild (water) birds are the natural reservoir for all haemagglutinin (H) and neuraminidase (N) subtypes, harbouring 16 of the former and 9 of the latter [3-5]. With approximately 10,000 bird species present in terrestrial and aquatic environments, the often large sizes and densities of bird populations are a boon for the spread of infectious diseases such as influenza. Avian influenza (AIV) is believed to be largely asymptomatic in wild birds such as waterfowl, causing mild or no symptoms, such that most AIVs in these species can be considered to be low pathogenic avian influenza (LPAI). In contrast, poultry (chicken and turkey)

are not considered as natural hosts for AIV, but can be infected with both LPAI and high pathogenic (HPAI) strains, notably H5 and H7. HPAI outbreaks may result in high mortality through severe respiratory distress, with major economic costs to the agricultural industries. Strikingly, despite the diversity of H and N types in wild birds, only a small number – H1, H2, H3, N1 and N2 – have evolved sustained transmission in humans, indicating that there are major adaptive barriers associated with the bird to mammal species jump [6].

It is similarly noteworthy that AIV outbreaks in poultry have only been associated with a relatively small number of H and N subtypes, suggesting that there are important species barriers within birds. A question that has been less frequently addressed is whether evolutionary patterns and processes also differ between wild birds and poultry? Indeed, there are a number of factors that could create important evolutionary differences between AIV in wild birds and poultry, notably reflecting differences in the mode of transmission and population

\* Correspondence: [mfourment@sydney.edu.au](mailto:mfourment@sydney.edu.au)

Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia

structure. In particular, the transmission of AIV between wild birds largely occurs through an indirect faecal-oral route, in which, following excretion, the virus sits in an aquatic environment before infecting another host [2,7,8]. Importantly, there is growing evidence that AIV is able to remain in these water environments for several weeks, although this is dependent on the physico-chemical characteristics of the water source in question, and the presence of microorganisms which limit AIV infectiousness [7-11]. Critically, while present in the environment, AIV will effectively be in a “latent” state, such that there will be no viral replication and no mutational accumulation. In contrast, AIV transmission in poultry is more likely to be a function of close contact (including airborne) and rapid faecal-oral transmission, without an extended latent period in water. Hence, a simple prediction from these differing modes of transmission is that evolutionary rates in AIV will be lower in wild birds than in poultry. Similarly, it is possible that the rapidity of AIV outbreaks in poultry increases the number of viral replications per unit time which, depending on the fitness of the mutations involved, may also increase evolutionary rates. Finally, selection pressures may also differ between wild birds and poultry. High-density poultry populations are dominated by the presence of a single viral subtype (e.g. H5, N7) within which there may be strong selection pressure on the virus to escape from increasing host immunity [12,13]. Indeed, HP H5N1 has been characterized by steady antigenic change (drift) since its emergence [14]. In contrast, weaker antigenic drift might be expected in wild bird populations that experience concurrent infections with multiple antigenically distinct subtypes [13]. To date, however, these fundamental differences between wild birds and poultry have not been incorporated into evolutionary models, including molecular clock dating.

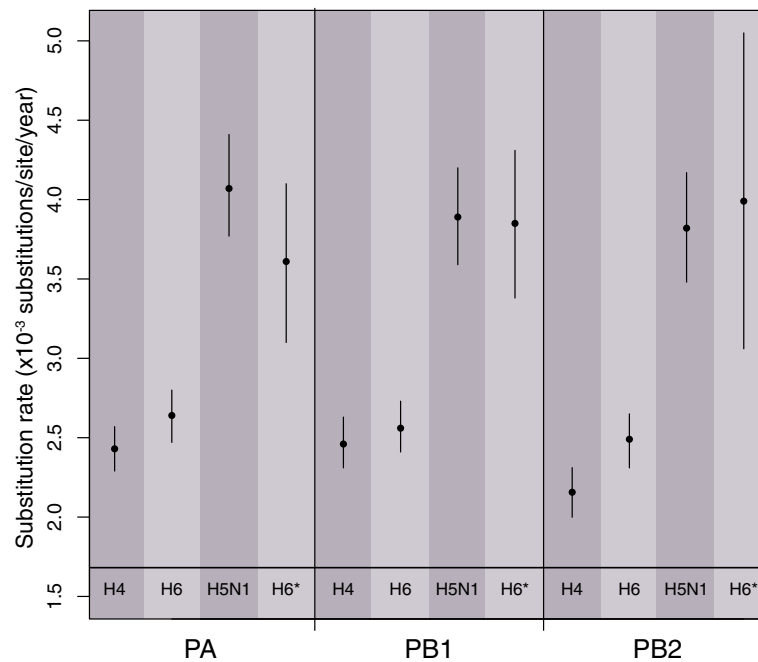
Most estimates of nucleotide substitution rates for influenza virus inferred using temporally sampled viruses point to rates in the order of  $10^{-3}$  substitution per site, per year (subs/site/year), in which genes coding for the external H and N proteins have higher evolutionary rates than internal genes [15-19]. The estimation of evolutionary rates in a statistical framework relies on the applicability of a molecular clock where the substitution rate can be constrained to be constant over time, or vary across time and among lineages. Methods that model rate heterogeneity include auto-correlation of rates among lineages [20], and uncorrelated models that allow rates assigned to lineages to be drawn from an underlying probability distribution [21]. Yoder and Yang (2000) modelled rate heterogeneity among lineage (heterotachy) using local clock models, a method that was recently reinvestigated using modern Bayesian-based sampling methods [22]. This clock

model allows phylogenetically related lineages to evolve at exactly the same rate. All these models are evidently oversimplifications of the substitution process and tend to provide different estimates, illustrating the need for the development of accurate model selection methods. Recently, Worobey et al. [18] used host-specific local clocks to infer the history of influenza A viruses and showed that the substitution rate of avian viruses were significantly higher than the human rate, although slightly lower or equal to the rate in swine. In this study, horses, birds, swine and humans are assumed to evolve at different rates, while viruses within each host are constrained to evolve at the same constant rate (i.e. a strict molecular clock). Critically, therefore, it was assumed that all AIVs evolve at the same rate, although the major ecological differences between wild birds and poultry noted above mean that this may be a simplification. Indeed, it is expected that the estimated substitution rates will be an average of multiple substitution rates that reflect the proportion of each viral subtype in the data set analysed.

To investigate the evolutionary rate heterogeneity of AIV, and particularly whether there is a systematic difference in rates between wild birds and poultry, we compiled data sets for the PB2, PB1, and PA internal genes representing AIV in these very different ecological contexts. These genes should be suitable for inferring substitution rates as they are the longest in the viral genome, are considered to be under moderate (predominantly negative) selection pressures, and different lineages often circulate in wild birds and poultry, facilitating the direct comparison of AIV in these two types of host species. Accordingly, for poultry, we examined sequences from the H5N1 (HPAI) subtype that has caused major outbreaks since 1997, while for wild birds we examined the commonly observed H4 and H6 subtypes (in combination with any other N type).

## Results and discussion

The MCMC-based (BEAST) analysis shows that the mean nucleotide substitution rate of AIV varies between  $1.87 \times 10^{-3}$  and  $4.2 \times 10^{-3}$  subs/site/year depending on the model used. These results consistently show significant lower rates (i.e. with non-overlapping confidence intervals) in those subtypes associated with wild birds (H4, H6) compared to poultry (H5N1) (Fig. 1). To further test this separation we analysed a monophyletic group of H6 viruses (denoted H6\*) that persisted in the poultry population rather than in their usual wild bird hosts. Strikingly, estimates of substitution rates in these poultry H6 viruses are similar to those in H5N1 from poultry and significantly higher than those of H6 from wild birds (Fig. 1), confirming that AIV evolutionary rates are elevated in poultry. Similarly, estimates of the mean  $\omega$  (a measure of overall selection pressure) and the



**Fig. 1** Evolutionary rates of avian influenza virus (AIV) among poultry and wild birds for three internal genes (PA, PB1, PB2). The nucleotide substitution rate per site of each gene was analysed using BEAST. Mean substitution rates and 95 % Bayesian confidence intervals estimated under strict (triangles) and lognormal (circles) molecular clocks are reported. \* denotes a lineage of H6 viruses from poultry

number of positively selected sites revealed consistent differences in selection pressures between H5N1 and the other subtypes (Table 1). In particular, the mean  $\omega$  of the H5N1 genes is substantially higher than those of H4 and H6, while no significant difference was observed between PA, PB1 and PB2.

This important discrepancy in evolutionary rates between those AIV subtypes that circulate in wild birds compared to poultry can be explained in at least two ways. First, this variation in rate could reflect underlying ecological differences. In particular, substitution rates may be reduced in wild birds because of the latent period associated with environmental (water-borne) transmission. In contrast, more frequent contact transmission in poultry may lead to more

replications, and hence a greater number of mutations, per unit time [23]. Selection pressures may also differ between wild birds and poultry, as revealed here (Table 1), and there is evidence for host-driven antigenic drift in H5N1 [12,13] which would be expected to increase evolutionary rates, especially at nonsynonymous sites. Second, because of heightened scientific attention, H5N1 has been more densely sampled than the other subtypes. It is possible that such intensive sampling may have increased the number of recently diverged sequences and hence the relative proportion of slightly deleterious mutations that are yet to be removed by purifying selection, in turn increasing rates toward the present [24]. Indeed, our analysis of selection pressures revealed that the mean values of  $\omega$  were higher for H5N1 than for H4 and H6 (Table 1). To determine whether this increase of  $\omega$  in H5N1 was indeed due to a higher frequency of transient deleterious mutations, or stronger positive selection, we examined  $\omega$  values on internal and external branches of the phylogenies (Table 1). Notably, in every data set there was a significantly higher  $\omega$  on external ( $\omega_e$ ) branches compared to internal ( $\omega_i$ ) branches (p-value < < 0.05). Importantly, however, the  $\omega_e/\omega_i$  ratio of the H5N1 data sets were consistently the lowest, such that this subtype has experienced proportionally more nonsynonymous mutations on internal branches and compatible with stronger positive selection. Hence, the elevated substitution rates in H5N1 are unlikely to be a function of sampling density.

**Table 1** Selection analyses of AIV among poultry and wild birds for three internal genes (PA, PB1, PB2)

	Mean $\omega$			Positively selected			Ratio $\omega_e/\omega_i$		
	PB2	PB1	PA	PB2	PB1	PA	PB2	PB1	PA
H4	0.3	0.28	0.36	0	0	0	3.17	3.21	3.31
H6	0.27	0.36	0.29	1	0	0	3.57	2.81	2.69
H5N1	0.48	0.55	0.73	0	1	5	1.53	2.24	1.84
H6*	1.32	1.15	1.31	0	0	0	2.71	3.24	2.98

H4 and H6 data sets contain wild birds only, while H5N1 and H6\* are poultry-only data sets. Mean  $\omega$  and the number of positively selected sites for each gene of each subtype were estimated using FUBAR. Separate ratios of nonsynonymous to synonymous substitution rates for external ( $\omega_e$ ) and internal ( $\omega_i$ ) branches were estimated using HyPhy

Our analysis further reveals that the mean substitution rate of the H6 subtype is consistently higher than that observed in H4 subtype. This disparity is surprising considering that both subtypes infect the same wild bird species and continuously reassort such that they often do not form distinct monophyletic groups, and have nearly identical selection profiles. Interestingly, however, chickens may be infected more frequently by the H6 than the H4 subtype. Taking the PA gene as an example, 7 % of all H6 sequences available on GenBank come from chickens, while the corresponding number for H4 is 0.3 % (although all poultry sequences were excluded from this analysis). While the more frequent circulation of the H6 lineage in the poultry population could explain the elevated rate in this subtype, this clearly needs to be confirmed on a larger and less biased data sample.

Also of note was that estimates of the mean substitution rate under an uncorrelated lognormal clock model are significantly higher than those estimated under a strict clock model. Considering the size of the data sets, the calculation of Bayes factors with computer intensive methods such as path sampling [25] is prohibitive. Instead we computed the AICM for each model (Additional files 1: Table S1, S2, and S3) and the results reveal the presence of rate heterogeneity in most data sets.

It is not straightforward to compare our estimates of the times to common ancestry to those of Worobey et al. [18] since the most recent common ancestor of all avian viruses (as analysed there) is not necessarily the MRCA of the subtypes analysed here. For example, H5N1 is the most recent lineage and originated between 12 and 16 years ago depending on the molecular clock model used, while the H4 and H6 subtypes have a tMRCA of at least 50 years but do not predate 1900 as suggested by the previous local clock analysis (Table 2). However, it is apparent that our estimates of root age are consistently younger than those of Worobey et al., suggesting that accounting for rate heterogeneity can lead to surprisingly different results [26].

## Conclusions

The estimates of nucleotide substitution rate presented here reveal a more complex picture of the evolutionary processes driving the evolution of avian influenza virus than previously thought. Most notably, it is clear that a single substitution rate cannot be applied to wild birds and poultry equally, as the evolutionary rates are consistently lower in the former and which may reflect a greater role for environmental transmission. Hence, it is evidently incorrect to impose a single substitution to AIV sampled from hosts as diverse as wild birds and poultry, and that molecular clock estimates based on a single rate may be erroneous. Similarly, rates differ significantly between the strict and relaxed clock models. It

**Table 2** Estimates of time to the most recent common under the best model for each AIV data set

Segment	Subtype	Population	Clock	Mean	Lower	Upper
PA	H5N1	Skyride	Lognormal	1999	1998	2000
	H4	Skyride	Lognormal	1962	1960	1964
	H6 wild birds	Skyride	Lognormal	1961	1959	1964
	H6 poultry	Skyride	Strict	1994	1993	1995
PB1	H5N1	Skyride	Lognormal	2000	2000	2001
	H4	Skyride	Lognormal	1963	1961	1965
	H6 wild birds	Constant	Lognormal	1942	1931	1951
	H6 poultry	Constant	Strict	1994	1993	1995
PB2	H5N1	Skyride	Lognormal	2000	2000	2001
	H4	Constant	Lognormal	1905	1874	1934
	H6 wild birds	Constant	Lognormal	1909	1876	1934
	H6 poultry	Constant	Lognormal	1993	1988	1998

Mean and 95 % Bayesian confidence intervals are reported for each gene. The population size and clock prior are reported

is therefore likely that the evolutionary dynamics of avian influenza virus can only be captured by new models that allow rates to vary in a more complex manner along lineages and which reflect changes in the underlying ecology.

## Methods

### Data

To determine whether there are significant differences in evolutionary rates between AIV from wild birds and poultry we focused on internal rather than external (i.e. H, N) proteins because the former are longer and less subject to positive selection that will likely have a major impact on evolutionary dynamics. Accordingly, full-length nucleotide sequences encoding the PB2, PB1, and PA proteins of AIV were obtained from the Influenza Virus Resource database at NCBI (GenBank) [27]. Four separate data sets were created for each gene based on their subtypes (H5N1, H4, and H6) and hosts (wild birds versus poultry). The H5N1 data sets were constructed using sequences isolated exclusively from poultry (i.e. chicken and turkey), while in the H4 and H6 data sets the small numbers of chicken and turkey (i.e. poultry) sequences were removed to obtain data sets composed exclusively of wild bird samples. An additional data set was created using H6 sequences isolated from chickens that formed a distinct monophyletic group, and hence which could be used to independently assess whether evolutionary rates are elevated in poultry. Multiple sequence alignments were generated using MAFFT version 7 [28] and manually edited using Seqotron [29], with final alignment lengths of 2151 bp, 2271 bp and 2280 bp for PA, PB1 and PB2, respectively. Total data set sizes were: H5N1: PA = 479 sequences, PB1 = 411, PB2 = 391;



H4: PA = 811, PB1 = 840, PB2 = 797; H6 (wild bird only): PA = 557, PB1 = 550, PB2 = 545; H6 (poultry only): PA = 36, PB1 = 35, PB2 = 32. For each gene, H4, H6, and H5N1 final data sets were combined and the phylogenetic history was investigated using the neighbor joining method. The tree files are available as supplementary files (Additional files 2, 3, and 4). Finally, for each gene, we provide neighbor-joining trees containing every H4, H6, and H5N1 sequence available on GenBank, and which shows the distribution of poultry versus wild bird viruses among subtypes (Additional files 5, 6, and 7).

### Phylogenetic analyses

Phylogenetic trees were inferred using maximum likelihood available in Physher [26] employing the generalised time-reversible (GTR) substitution model and a discretised gamma distribution (4 categories) of rates across sites ( $\Gamma_4$ ). Using the year of isolation as sampling time, temporal outliers possibly indicative of sequencing errors were removed by inspecting the correlation coefficient between regressions of root-to-tip divergence and sampling times using Path-O-Gen [30]. Correlation coefficients ranged between 0.66 and 0.9 in all cases.

Estimates of nucleotide substitution rate were performed using the Markov chain Monte Carlo (MCMC) method available in BEAST [31] with the BEAGLE library [32]. All analyses were performed under the GTR+ $\Gamma_4$  model of nucleotide substitution, with the exception of the H6 poultry data sets where the simpler HKY+ $\Gamma_4$  model was used due to convergence issues. Every data set were analysed using the non-parametric Bayesian skyride or constant population size prior [33], and either a strict or uncorrelated lognormal relaxed clock [21]. An exponential prior was placed on the substitution rate with mean  $3 \times 10^{-3}$  subs/site/year. For each data set at least two chains of 100 million generations were run and an appropriate number of samples were discarded as burn-in after visually assessing the convergence of chains using Tracer version 1.6 [34]. Finally, independent runs were combined using LogCombiner. Means and 95 % highest posterior density (HPD) intervals of the substitution rate and divergence time parameters were calculated from the posterior sample. We compared models using the approximate AICM [25] implemented in Tracer.

To evaluate the strength of natural selection across sites, we used the fast unconstrained Bayesian approximation (FUBAR) method [35]. The Muse-Gaut (MG4) codon model of substitution [36] incorporating general nucleotide substitution biases (MG94  $\times$  GTR) and the phylogeny estimated from the nucleotide data was used to investigate site-specific selection. The sign of selection acting on the sequences was assessed using the ratio  $\omega$  of the relative rates of nonsynonymous ( $\beta$ ) to

synonymous ( $\alpha$ ) nucleotide substitutions per site. In addition, we analysed differences in selective pressures between internal and external lineages (i.e. any branch leading to a taxon) using a two-ratio model implemented in HyPhy [37], and where an elevated  $\omega$  on internal versus external branches is indicative of stronger adaptive evolution. The difference in goodness-of-fit between the two-ratio and one-ratio model (i.e. where this is a single  $\omega$  for all branches) was evaluated using the likelihood ratio test with one degree of freedom.

### Availability of supporting data

The data sets supporting the results of this article are available in the Zenodo repository, <http://dx.doi.org/10.5281/zenodo.18622>.

### Additional files

#### Additional file 1: Supplementary Table 1 to 3.

**Additional file 2: Neighbour-joining tree of the PA gene segment data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This can be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6 wild bird: black, H6 poultry: green).

**Additional file 3: Neighbour-joining tree of the PB1 gene segment data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This can be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6 wild bird: black, H6 poultry: green).

**Additional file 4: Neighbour-joining tree of the PB2 gene segment data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This can be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6 wild bird: black, H6 poultry: green).

**Additional file 5: Neighbour-joining tree of the PA gene segment complete data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This file can be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6: black).

**Additional file 6: Neighbour-joining tree of the PB1 gene segment complete data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This file can be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6: black).

**Additional file 7: Neighbour-joining tree of the PB2 gene segment complete data set.** Neighbour-joining tree in newick format containing the H4, H6, and H5N1 viruses analysed in this study. This file should be opened with FigTree (free at: <http://tree.bio.ed.ac.uk/software/figtree/>). Taxa are colour-coded by subtype (H5N1: red, H4: blue, H6: black).

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MF and ECH conceived and contributed to study design. MF performed the analysis. MF and ECH were involved in the writing of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

MF is currently supported by a postdoctoral research fellowship from the University of Sydney. ECH is supported by an NHMRC Australia Fellowship (AF30).

Received: 5 March 2015 Accepted: 2 June 2015

Published online: 26 June 2015

## References

- Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, et al. New world bats harbor diverse influenza A viruses. *PLoS Pathog.* 2013;9(10), e1003657.
- Vong S, Ly S, Mardy S, Holl D, Buchy P. Environmental contamination during influenza A virus (H5N1) outbreaks, Cambodia, 2006. *Emerg Infect Dis.* 2008;14(8):1303–5.
- W.H.O. A Revision of the System of Nomenclature for Influenza-Viruses - a Who Memorandum. *B World Health Organ.* 1980;58(4):585–91.
- Webster RG, Bean Jr WJ. Genetics of influenza virus. *Annu Rev Genet.* 1978;12:415–31.
- Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, Smith D, et al. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J Virol.* 2005;79(5):2814–22.
- Causey D, Edwards SV. Ecology of avian influenza virus in birds. *J Infect Dis.* 2008;197 Suppl 1:S29–33.
- Brown JD, Swayne DE, Cooper RJ, Burns RE, Stallknecht DE. Persistence of H5 and H7 avian influenza viruses in water. *Avian Dis.* 2007;51(1 Suppl):285–9.
- Domanska-Blicharz K, Minta Z, Smietanka K, Marche S, van den Berg T. H5N1 high pathogenicity avian influenza virus survival in different types of water. *Avian Dis.* 2010;54(1 Suppl):734–7.
- Meixell BW, Borchardt MA, Spencer SK. Accumulation and inactivation of avian influenza virus by the filter-feeding invertebrate *Daphnia magna*. *Appl Environ Microbiol.* 2013;79(23):7249–55.
- Roche B, Drake JM, Brown J, Stallknecht DE, Bedford T, Rohani P. Adaptive evolution and environmental durability jointly structure phylogenetic patterns in avian influenza viruses. *Plos Biol.* 2014;12(8), e1001931.
- Zhang H, Li Y, Chen J, Chen Q, Chen Z. Perpetuation of H5N1 and H9N2 avian influenza viruses in natural water bodies. *J Gen Virol.* 2014;95(Pt 7):1430–5.
- Chen R, Holmes EC. Hitchhiking and the population genetic structure of avian influenza virus. *J Mol Evol.* 2010;70(1):98–105.
- Dugan VG, Chen R, Spiro DJ, Sengamaly N, Zaborsky J, Ghedin E, et al. The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* 2008;4(5), e1000076.
- Cattoli G, Milani A, Temperton N, Zecchin B, Buratin A, Molesti E, et al. Antigenic Drift in H5N1 Avian Influenza Virus in Poultry Is Driven by Mutations in Major Antigenic Sites of the Hemagglutinin Molecule Analogous to Those for Human Influenza Virus. *J Virol.* 2011;85(17):8718–24.
- Fourment M, Wood JT, Gibbs AJ, Gibbs MJ. Evolutionary dynamics of the N1 neuraminidases of the main lineages of influenza A viruses. *Mol Phylogenet Evol.* 2010;56(2):526–35.
- Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. *Nature.* 2008;453(7195):615–9.
- Smith GJ, Bahl J, Vijaykrishna D, Zhang J, Poon LL, Chen H, et al. Dating the emergence of pandemic influenza viruses. *Proc Natl Acad Sci U S A.* 2009;106(28):11709–12.
- Worobey M, Han GZ, Rambaut A. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature.* 2014;508(7495):254–7.
- Worobey M, Han GZ, Rambaut A. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc Natl Acad Sci U S A.* 2014;111(22):8107–12.
- Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 1998;15(12):1647–57.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *Plos Biol.* 2006;4(5):699–710.
- Drummond AJ, Suchard MA. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* 2010;8.
- Hanada K, Suzuki Y, Gojobori T. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol.* 2004;21(6):1074–80.
- Duchene S, Holmes EC, Ho SY. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proceedings Biological Sciences/The Royal Society*, 2014;281(1786). <http://www.ncbi.nlm.nih.gov/pubmed/24850916>
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Mol Biol Evol.* 2012;29(9):2157–67.
- Fourment M, Holmes EC. Novel non-parametric models to estimate evolutionary rates and divergence times from heterochronous sequence data. *BMC Evol Biol.* 2014;14:163.
- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, et al. The influenza virus resource at the National Center for Biotechnology Information. *J Virol.* 2008;82(2):596–601.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
- Seqotron v1.0 [<http://github.com/4ment/seqotron>]
- Path-o-Gen [<http://tree.bio.ed.ac.uk/software/pathogen/>]
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:214.
- Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. *Bioinformatics.* 2009;25(11):1370–6.
- Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* 2008;25(7):1459–71.
- Tracer [<http://tree.bio.ed.ac.uk/software/tracer/>]
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* 2013;30(5):1196–205.
- Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 1994;11(5):715–24.
- Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;21(5):676–9.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

