Supplementary Information:
Global circulation patterns of seasonal influenza viruses vary with antigenic drift

Trevor Bedford[1], Steven Riley[2,3], Ian G. Barr[4], Shobha Broor[5], Mandeep Chadha[6],
Nancy J. Cox[7], Rodney S. Daniels[8], C. Palani Gunasekaran[9], Aeron C. Hurt[4,10],
Anne Kelso[4], Alexander Klimov[7], Nicola S. Lewis[11], Xiyan Li[12], John W. McCauley[8],
Takato Odagiri[13], Varsha Potdar[6], Andrew Rambaut[3,14,15], Yuelong Shu[12], Eugene
Skepner[11], Derek J. Smith[11,16], Marc A. Suchard[17,18,19], Masato Tashiro[13], Dayan
Wang[12], Xiyan Xu[7], Philippe Lemey[20] & Colin A. Russell[21]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA,
[2]Department of Infectious Disease Epidemiology, Imperial College London, London, UK, [3]Fogarty
International Center, National Institutes of Health, Bethesda, MD, USA, [4]World Health Organization
(WHO) Collaborating Centre for Reference and Research on Influenza, Melbourne, Australia, [5]SGT
Medical College, Hospital and Research Institute, Village Budhera, District Gurgaon, Haryana, India,
[6]National Institute of Virology, Pune, India, [7]WHO Collaborating Center for Influenza, Centers for Disease
Control and Prevention, Atlanta, GA, USA, [8]WHO Collaborating Centre for Influenza, National Institute
for Medical Research (NIMR), London, UK, [9]King Institute of Preventive Medicine and Research, Guindy,
Chennai, India, [10]Melbourne School of Population and Global Health, University of Melbourne, Parkville
VIC 3010, Australia, [11]Department of Zoology, University of Cambridge, Cambridge, UK, [12]WHO
Collaborating Center for Influenza, China Centers for Disease Control, Beijing, China, [13]WHO
Collaborating Center for Influenza, National Institute for Infectious Diseases, Tokyo, Japan, [14]Institute of
Evolutionary Biology, University of Edinburgh, Edinburgh, UK, [15]Centre for Immunology, Infection and
Evolution, University of Edinburgh, Edinburgh, UK, [16]Department of Viroscience, Erasmus Medical
Center, Rotterdam, The Netherlands, [17]Department of Biostatistics, UCLA Fielding School of Public
Health, University of California, Los Angeles CA, USA, [18]Department of Biomathematics David Geffen
School of Medicine at UCLA, University of California, Los Angeles, CA, USA, [19]Department of Human
Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA,
[20]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven,
Leuven, Belgium, [21]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

Supplementary Discussion

Sensitivity analyses

In our primary analyses we used approximately twice as many H3N2 viruses as H1N1 or B viruses, largely due to the greater availability of H3N2 sequence data. To assess the robustness of our results to sampling patterns and numbers of sequences used, we repeated our entire analysis pipeline with smaller datasets (1241 to 1394 viruses) designed to have similar sample counts in each region for each virus. This was accomplished by sampling at most 14 sequences per region per year (13 for USA/Canada) for H3N2, 30 sequences per region per year (28 for USA/Canada) for H1N1, 30 sequences per region per year (24 for USA/Canada) for Vic and 40 sequences per region per year (25 for USA/Canada) for Yam from the full datasets (9139 H3N2, 3789 H1N1, 2577 Vic and 1821 Yam sequences). This resulted in 1391 H3N2 sequences, 1372 H1N1 sequences, 1394 Vic sequences and 1240 Yam sequences with approximately 100–180 sequences per region. Even with substantially altered sample sizes we obtained results highly consistent with the primary analysis, with mean persistence estimates differing by at most 8% compared with the primary dataset (Extended Data Table 2). The biggest difference in the secondary analysis was the degree of certainty, i.e. posterior probability, in H3N2 trunk location through time (Extended Data Fig. 2). With reduced sample size, there was less certainty, e.g. 2006 was not assigned confidently to any one location in the smaller secondary dataset, while it was assigned confidently to India in the larger primary dataset (93% posterior probability). In this case, the larger dataset contained Indian samples that lay very close to the trunk of the phylogeny that were absent in the smaller dataset. We observed no occurrences of conflicting support in assignment of the source region, i.e. where one dataset supported one region and the other dataset supported a different region as source; suggesting that our statistical approach was robust to the sampling strategies we employed.

The analyses were also repeated with an alternative geographic assignment of sequences to 10 distinct regions: USA/Canada, South America, Europe, India, Japan/Korea, Southeast Asia, Oceania, China, Central America and Africa. For this geographic breakdown, subsampling resulted in 1967 H3N2 sequences, 1439 H1N1 sequences, 1756 Vic sequences and 1223 Yam sequences. Extended Data Table 2 shows summary statistics for primary, secondary and alternative datasets with highly consistent results regardless of the sampling strategy used. The secondary and alternative datasets gave highly similar results to the primary analysis with H1N1 and influenza B showing substantially more regional persistence and population structure than H3N2. Persistence estimates differed by at most 20% between primary, secondary and alternative datasets.

We further confirmed our phylogeographic results by analyzing F_{ST} (a measure of population heterogeneity[42]) across viruses, finding that H3N2 shows less geographic population structure than H1N1 or B viruses (Extended Data Table 1). Crucially, F_{ST} does not depend on phylogeographic reconstruction and instead compares genealogical diversity between isolates from the same geographic region to genealogical diversity between isolates from different geographic regions, so that $F_{ST} = (\pi_b - \pi_w)/\pi_b$, where π_w represents within-region genealogical diversity and π_b represents between-region genealogical diversity. F_{ST} estimates were also highly consistent across datasets (Extended Data Table 2).

Results were also robust to specific choice of prior distribution for phylogeographic rate parameters.

Defining regional persistence

Regional persistence was defined as continuous circulation of specific genetic variants (lineages) of a particular virus within a particular geographic region. It is necessary to specify persistence of "genetic variants" as most regions of the world that have good surveillance will collect influenza viruses in most, if not all, months of each year. These inter-epidemic viruses will often be antigenically similar to viruses that caused the previous local epidemic. Thus, purely on epidemiological or antigenic data, one might infer continuous circulation of viruses in many regions. However, we have previously shown for H3N2 viruses[1] that these inter-epidemic isolates are always as closely related, or more, to viruses circulating in other parts of the world than to viruses from the previous local epidemic. This suggests that genetic variants go extinct locally at the end of an epidemic and viruses collected between epidemics were more likely to be the product of short-lived external introductions. Our results in the present study corroborate these patterns for H3N2 viruses, but we find evidence that genetic variants (lineages) of H1N1, Vic, and Yam viruses circulated continuously after local epidemics and subsequently seeded epidemics in the same location.

Persistence time estimates for combined regions

Defining optimal epidemiological regions for seasonal influenza viruses remains challenging and necessarily subjective. As stated in the Methods section, nine regions were defined to maximize available sequences within each region while still providing enough geographic diversity to ensure nearly global coverage. In our primary analyses, China was split into North China and South China owing to the wealth of available sequence data and because of potential differences in virus seasonality and circulation between North and South China[43].

When estimating average persistence times for H3N2 viruses (Fig. 1), we found similar durations of persistence of less than one year for most regions of the world, including North China (0.53 years) and South China (0.50 years). However, the estimated persistence time for North and South China combined (1.27 years) was substantially greater than for North or South China alone, suggesting that virus persistence was enabled by movement of viruses between North and South China. To investigate the effect of combining other regions, we estimated the mean persistence times for all pairs of regions from our primary phylogeographic analyses. Extended Data Figure 6a shows persistence times for combined pairs of regions. Here, the diagonals of each panel are equivalent to the point estimates in Figure 1 and show the mean persistence times within each region (plus China) for all four viruses. For H3N2 viruses, combining most region pairs did very little, if anything, to increase persistence, except for North and South China as mentioned above.

Expanding beyond pairs of regions for H3N2 viruses, combining North China, South China and Southeast Asia gave a mean persistence time of 1.50 years, longer than the 1.26 years estimated for North and South China combined. Collectively, North China, South China, Southeast Asia, India, and Japan/Korea gave a mean persistence time of 2.61 years. However, excluding China from this group (leaving Southeast Asia, India, and Japan/Korea) gave a persistence estimate of 0.78 years, highlighting the importance of China for the evolution and circulation of H3N2 viruses. For comparison, combining regions outside of Asia shows a much smaller signature of joint persistence (persistence estimated for USA, Canada, South America, Europe and Oceania was 1.06 years).

For H1N1 viruses, joint persistence estimates for most region pairs highlight the substantial contributions of India to mean durations of H1N1 persistence (Extended Data Fig. 6a). Similarly,

North and South China, along with Southeast Asia, Japan and Korea, jointly contribute to H1N1 persistence in East and Southeast Asia (combined persistence estimate of 2.59 years). For the B viruses, joint estimates of persistence times highlight contributions by India and China to duration of persistence of viruses in all regions (Extended Data Fig. 6a). Additionally, the joint persistence estimates for North and South America for Vic and Yam viruses suggest that B virus persistence in the Americas could have arisen from movement between North and South America (combined persistence estimates of 1.90 years for Vic and 1.57 years for Yam).

Generally, our analyses suggest that the nine regions used in the primary analyses, with the possible exception of North and South China, could be considered epidemiologically distinct entities, particularly for H3N2 viruses. Genetic population structure is clearly evident across these regions and ancestry patterns show clear geographic structuring. However, our analyses do not indicate that the specific partitioning considered here is the single optimal epidemiological or genetic partitioning of this complex geographic structure and there remains some ambiguity regarding the definition of meaningful epidemiological regions. Regardless, the regions considered here provide a useful frame of reference to compare influenza viruses.

Disease dynamics and persistence

In the epidemiological model, within deme persistence times were affected by rates of antigenic evolution and by amplitude of seasonal forcing, ϵ (Extended Data Fig. 8a). At high amplitudes of seasonal forcing, within-deme persistence was rare for all rates of antigenic evolution. However, at low amplitudes of seasonal forcing, rates of antigenic evolution had strong effects: slow evolution resulted in long (i.e. ~2 years) within-deme persistence while rapid evolution resulted in substantially reduced within-deme persistence (i.e. ~1 year). For $\epsilon = 0.00$, we found a linear relationship where 1 additional unit/year of antigenic drift results in a decrease of average persistence by 1.21 years ($P = 0.0002$), while for $\epsilon = 0.04$, we found a linear relationship where 1 additional unit/year of antigenic drift results in a decrease of average persistence by 0.71 years ($P = 0.00006$).

In the model, amplitudes of seasonal forcing and rates of antigenic evolution directly affect R_t. High amplitudes of seasonal forcing and rapid evolution result in relatively short but intense epidemics followed by local extinction. Conversely, low amplitudes of seasonal forcing and slow evolution result in small epidemics that slowly deplete the susceptible population and thus epidemics have the potential to continue outside of the influenza season giving rise to persistent lineages. Low amplitudes of seasonal forcing and rapid evolution result in epidemics of intermediate size and duration that are frequently followed by local extinction. The variations in amplitude of seasonality from our model appear to be inline with recently described patterns of influenza virus epidemiology in China[43] where there are strong latitudinal gradients in amplitude of annual epidemic periodicity with substantially weaker periodicity in southern latitudes.

Virological differences

Given that influenza H1N1 and B viruses primarily infect children, and children are likely to exert less immune pressure than adults, epidemiology potentially shapes virus evolution. Previous theoretical work has shown that epidemiological processes can influence influenza virus evolution[19,20]. Influenza virus evolution in swine provides an valuable comparative example. In swine, H3N2 viruses evolve very slowly antigenically despite accumulating genetic mutations at roughly the

same rate as H3N2 viruses in humans[44]. Ecological differences between humans and pigs, particularly the very short lifespan of the average commercially-reared pig and thus potential lack of immune pressure to drive evolution, are likely to be the cause of evolutionary differences.

While epidemiology has the potential to shape evolution, there are important virological differences between influenza viruses, particularly A and B viruses, that are likely to constrain virus evolution and thus shape epidemiology. First, influenza B viruses have a higher fidelity polymerase than influenza A viruses[21], which results in slower rates of synonymous evolution (Extended Data Table 1) and fewer opportunities for antigenic change. Second, influenza B viruses induce a weaker humoral immune response as measured by hemagglutination inhibition titer than influenza A viruses[22]. Third, influenza B viruses have lower receptor binding affinity than A viruses[23] which might compromise their ability to evolve antigenically[24].

Other recent studies have highlighted other potential evolutionary constraints on virus evolution that are likely to be more strongly mediated by virological than epidemiological factors. Koelle et al.[45] put forth an elegant hypothesis regarding the evolution of influenza viruses along neutral networks where antigenic change is limited by the accessibility of substitutions that simultaneously alter antigenic phenotype while preserving protein function. Koel et al.[46] found that single amino acid substitutions at 7 positions near the receptor-binding site were capable of causing antigenic cluster transitions. They hypothesized that given the frequency at which these substitutions were likely to occur given polymerase fidelity alone, the lack of regular appearance of new antigenic variants and relatively infrequent changes in these 7 positions was due to an intrinsic fitness cost (changes in replication, receptor binding) for the virus to make these changes. One key finding from Koel et al.[46] is that the introduction of specific substitutions at these 7 positions in some viruses lead to the inability to rescue those viruses. This suggests that these antigenic-change inducing substitutions can intrinsically cripple a virus and thus viruses are not 'free' to vary. Importantly, these factors could easily differ among viruses. The findings of Koelle et al.[45] and Koel et al.[46] are also very much inline with recent work from Gong, Suchard, and Bloom[47] showing that epistasis is a critical limiting factor in the evolution of influenza virus nucleoprotein.

Sequence data limitations

For this study, we focused exclusively on the influenza HA, as HA is the primary target of the human immune response, the primary component of influenza virus vaccines and the main focus of most seasonal influenza virus surveillance efforts, it is therefore the gene segment for which most sequence data exists. Studies based on whole genome sequences are likely to provide insights into the genesis and spread of reassortant viruses and the co-circulation of multiple lineages. The continued expansion of whole genome sequencing will offer additional markers for tracking the global movement of viruses. In this study, we lacked data from Africa, Central America, and the Middle East and their roles in the global circulation of seasonal influenza viruses are yet to be determined.

References

42. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).

43. Yu, H. *et al.* Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data. *PLOS Med.* **10**, e1001552 (2013).

44. de Jong, J. *et al.* Antigenic and genetic evolution of swine influenza A (H3N2) viruses in Europe. *J. Virol.* **81**, 4315–4322 (2007).

45. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006).

46. Koel, B. F. *et al.* Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976–979 (2013).

47. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).