LETTER

doi:10.1038/nature14460

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]

Understanding the spatiotemporal patterns of emergence and circulation of new human seasonal influenza virus variants is a key scientific and public health challenge. The global circulation patterns of influenza A/H3N2 viruses are well characterized[1–7], but the patterns of A/H1N1 and B viruses have remained largely unexplored. Here we show that the global circulation patterns of A/H1N1 (up to 2009), B/Victoria, and B/Yamagata viruses differ substantially from those of A/H3N2 viruses, on the basis of analyses of 9,604 haemagglutinin sequences of human seasonal influenza viruses from 2000 to 2012. Whereas genetic variants of A/H3N2 viruses did not persist locally between epidemics and were reseeded from East and Southeast Asia, genetic variants of A/H1N1 and B viruses persisted across several seasons and exhibited complex global dynamics with East and Southeast Asia playing a limited role in disseminating new variants. The less frequent global movement of influenza A/H1N1 and B viruses coincided with slower rates of antigenic evolution, lower ages of infection, and smaller, less frequent epidemics compared to A/H3N2 viruses. Detailed epidemic models support differences in age of infection, combined with the less frequent travel of children, as probable drivers of the differences in the patterns of global circulation, suggesting a complex interaction between virus evolution, epidemiology, and human behaviour.

Owing to the frequency and severity of human seasonal influenza A/H3N2 virus epidemics, recent work has focused on the global circulation dynamics of H3N2 viruses[1–7]. Studies have shown that, each year, H3N2 epidemics worldwide result from the introduction of new genetic variants from East and Southeast (E-SE) Asia, where viruses circulate via a network of temporally overlapping epidemics[1,2,4,5], rather than local persistence[1,3,6,7]. In addition to H3N2, H1N1 viruses and two antigenically diverged lineages of influenza B viruses, B/Victoria/2/1987-like (Vic) and B/Yamagata/16/1988-like (Yam), circulate among humans with lower but substantial disease burdens[8,9]. Despite their importance, the global circulation dynamics of former seasonal H1N1 viruses (preceding the 2009 pandemic) and B viruses have been largely neglected.

Given that influenza A and B viruses cause similar symptoms and evolve by similar mechanisms of immune escape, we hypothesized that each would follow similar patterns of global circulation, with new genetic variants originating in East and Southeast Asia that rapidly replace existing genetic variants. To test this hypothesis we compared the global circulation patterns of the haemagglutinin (HA) genes of H3N2, former seasonal H1N1, Vic, and Yam viruses. We assembled data sets of HA sequences with complete HA1 domains for each subtype from the World Health Organization Global Influenza Surveillance and Response System and the Influenza Research Database[10], covering 2000–2012. To reduce the impact of surveillance biases, we subsampled these data to more equitable spatiotemporal distributions, resulting in data sets comprising 4,006 H3N2, 2,144 H1N1, 1,999 Vic, and 1,455 Yam HA sequences (Extended Data Fig. 1). Although deficient in viruses from Africa and Eastern Europe, to our knowledge these are the most geographically and temporally comprehensive seasonal influenza virus data sets assembled to date.

By estimating temporally resolved phylogenetic trees for each subtype, we revealed faster rates of nucleotide mutation and amino acid substitution in H3N2 and H1N1 than in the B viruses (consistent with previous work[11,12]), but more genealogical diversity in B viruses than in A viruses (Extended Data Table 1). This inverse relationship between evolutionary rate and genealogical diversity is expected if increased mutation rate correlates with antigenic drift[13] and drives increased adaptive evolution, thus purging HA genetic diversity[14]. By inferring geographic ancestry using Bayesian phylogeographic methods[15], we found a consistent pattern for H3N2 viruses (Fig. 1a) in which viruses worldwide rapidly coalesce to the trunk of the tree (average time to trunk = 1.42 years), with trunk viruses mostly originating from East and Southeast Asia (Extended Data Fig. 2a). This finding is consistent with previously reported patterns[1,2,4,5], with East and Southeast Asia acting as the source population for epidemics worldwide.

In addition to China and Southeast Asia, India frequently contributed viruses to the trunk of the tree, suggesting that the global circulation of H3N2 viruses is maintained by an East and Southeast Asian network that includes India. India's role in the global dissemination of H3N2 viruses may have been similar historically, but India-wide influenza surveillance only began in 2004. There were brief periods, notably the 2007–2008 Northern Hemisphere winter, when regions outside

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [2]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London SW7 2AZ, UK. [3]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]World Health Organization (WHO) Collaborating Centre for Reference and Research on Influenza, Melbourne, Victoria 3000, Australia. [5]SGT Medical College, Hospital and Research Institute, Village Budhera, District Gurgaon, Haryana 122505, India. [6]National Institute of Virology, Pune 411001, India. [7]WHO Collaborating Center for Reference and Research on Influenza, Centers for Disease Control and Prevention, Atlanta, Georgia 30329, USA. [8]WHO Collaborating Center for Reference and Research on Influenza, Medical Research Council National Institute for Medical Research (NIMR), London NW7 1AA, UK. [9]King Institute of Preventive Medicine and Research, Guindy, Chennai 600032, India. [10]Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria 3010, Australia. [11]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [12]WHO Collaborating Center for Reference and Research on Influenza, National Institute for Viral Disease Control and Prevention, China CDC, Beijing 102206, China. [13]WHO Collaborating Center for Reference and Research on Influenza, National Institute of Infectious Diseases, Tokyo 208-0011, Japan. [14]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [15]Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, UK. [16]Department of Viroscience, Erasmus Medical Center, 3015 Rotterdam, The Netherlands. [17]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, California 90095, USA. [18]Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California 90095, USA. [19]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California 90095, USA. [20]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, 3000 Leuven, Belgium. [21]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK.

‡Deceased.

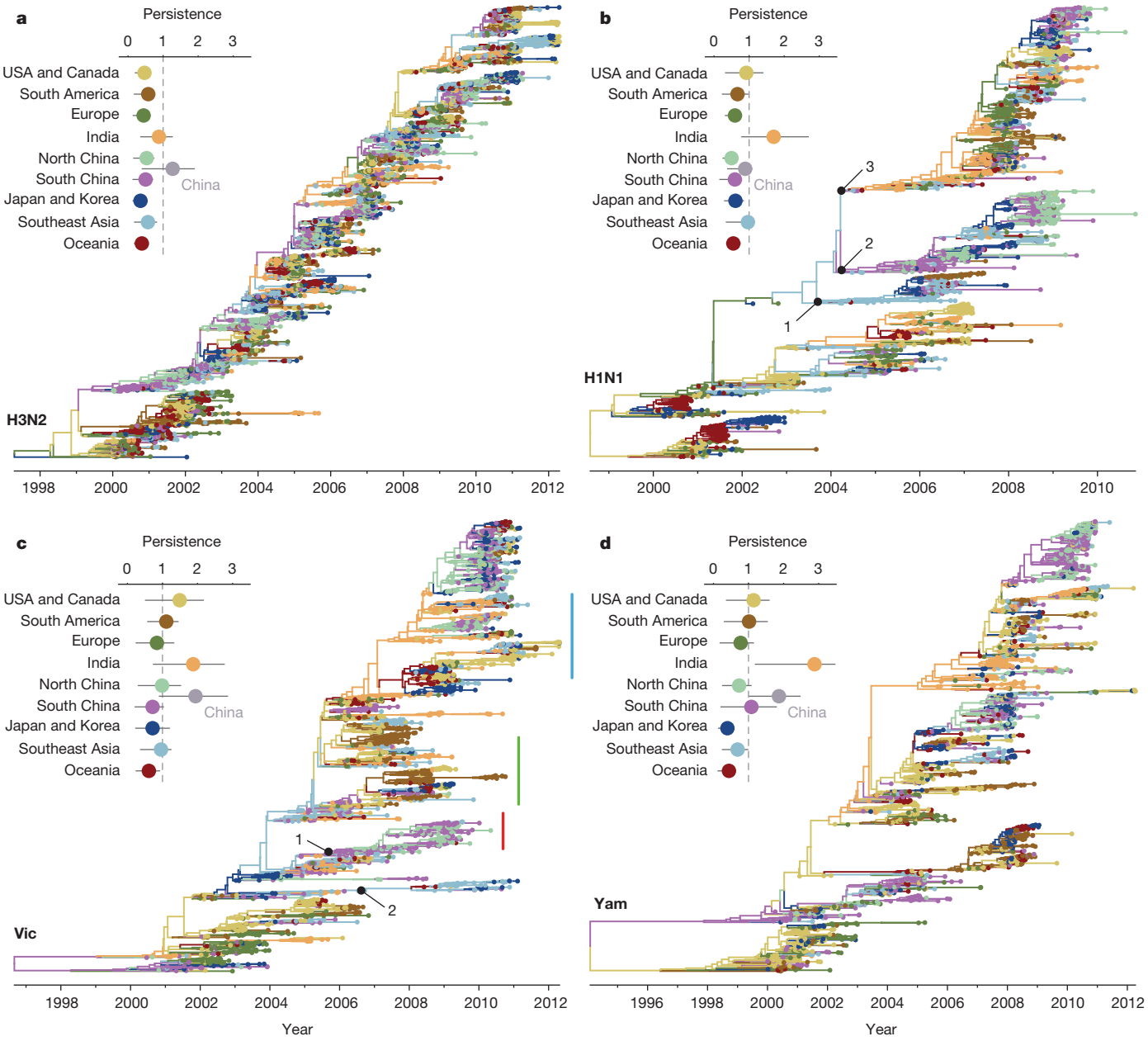


Figure 1 | Maximum clade credibility trees. a–d, Trees created with primary data sets of 4,006 H3N2 viruses (**a**), 2,144 H1N1 viruses (**b**), 1,999 Vic viruses (**c**) and 1,455 Yam viruses (**d**). Branch tips are coloured by geographic region of virus collection; internal branches are coloured by geographic region as inferred by Bayesian phylogeographic methods (region colours in persistence insets). In **b**, nodes 1–3 indicate co-circulating clades that diverged in 2004. In **c**, nodes 1 and 2 indicate divergent clades of viruses from Asia, coloured vertical bars indicate antigenic variants shown in Extended Data Fig. 5a (green, B/Malaysia/2506/2004-like; red, B/Hubei Songzi/52/2008-like; other post-2008 viruses, B/Brisbane/60/2008-like). The inset to the top left of each tree shows duration of region-specific persistence measured as the waiting time in years for a virus to leave its geographic region of origin. Circles represent mean persistence across sampled viruses, while lines show the interquartile range of persistence across sampled viruses. Region 'China' shows the combined persistence estimate for North China and South China together.

East and Southeast Asia contributed to the trunk of the H3N2 tree. However, these instances were rare and trunk viruses from outside East and Southeast Asia descended directly from viruses within East and Southeast Asia (Fig. 1a). Quantifying the average ancestry of strains from each geographic region in the 3 years before sampling showed prominent roles for China, India, and Southeast Asia in seeding epidemics in all regions (Extended Data Fig. 3).

Surprisingly, the global circulation patterns of former seasonal H1N1 viruses differed substantially from those observed for H3N2 viruses (Fig. 1). Like H3N2, most lineages of H1N1 viruses eventually coalesced with viruses from East and Southeast Asia and India. However, this coalescence was slower than for H3N2 viruses with prolonged co-circulation of geographically segregated H1N1 lineages (Fig. 1b, Extended Data Figs 3 and 4). Geographic segregation of H1N1 viruses was particularly pronounced beginning in 2004/2005, with the emergence of three co-circulating genetic lineages (Fig. 1b, nodes 1–3) that each independently acquired HA mutations leading to antigenic evolution from the A/New Caledonia/20/1999-like phenotype to the A/Solomon Islands/3/2006-like phenotype. These lineages circulated in Southeast Asia (node 1), China (node 2) and India (node 3), with the Indian lineage eventually spreading worldwide before the emergence of H1N1pdm09 viruses.

Phylogeographic analyses of B Vic and Yam viruses revealed further differences from H3N2 viruses with lineages frequently circulating

outside of East and Southeast Asia for several years without evidence of seeding from East and Southeast Asia (Fig. 1c, d). Prominent examples include the seeding of the North American 2006/2007 Vic season directly from 2005/2006 North American viruses and the seeding of the North American 2001/2002 Yam season directly from 2000/2001 North American viruses (Extended Data Fig. 4). Similarly, lineages of viruses within East and Southeast Asia commonly circulated exclusively in East and Southeast Asia for more than 1 year. These long circulating East and Southeast Asian lineages were most apparent for Vic viruses where two lineages (Fig. 1c, nodes 1 and 2) persisted independently in China and Southeast Asia for over 5 years without spreading to other regions and led to the co-circulation of three distinct Vic antigenic variants in different parts of the world during 2007–2008 (Extended Data Fig. 5a).

Patterns of persistence of genetic variants differed by subtype and region, with H3N2 viruses persisting regionally for an average of ~6 months, H1N1 for ~9 months, Vic for ~13 months and Yam for ~12 months. H3N2 viruses showed comparably short durations of persistence across the world (Fig. 1), with the exceptions of India and China. Patterns within China were characterized by North and South lineages contributing jointly to persistence, as combining North and South phylogeny nodes resulted in substantially greater persistence estimates than from North or South lineages alone (Fig. 1). For H3N2, evidence for joint contributions to persistence by region pairs that exclude China is comparatively weak (Extended Data Fig. 6a, Supplementary Information). For Vic and Yam, the mean duration of persistence was longer than for H3N2 or H1N1 in most regions, particularly in India and China where mean durations were >2 years (Fig. 1, Extended Data Fig. 4). Duration of regional persistence correlated with the proportion of virus originating from that region (Extended Data Fig. 6b) and observed phylogeographic patterns were robust to subsampling assumptions (Supplementary Information, Extended Data Table 2).

To investigate differences in the global migration patterns of H3N2, H1N1 and B viruses, we used the spatiotemporally resolved phylogenies to estimate the amounts of virus movement between regions (Fig. 2). Rates of movement between pairs of regions were highly correlated between viruses with Spearman correlation coefficients ranging from 0.65 (H3N2 vs Yam) to 0.75 (H3N2 vs H1N1), suggesting similar global connectivity networks for all viruses. However, while the overall structure of the migration network was similar, H3N2 viruses moved between regions more frequently than H1N1 and B viruses (migration events per lineage per year H3N2 = 1.96, H1N1 = 1.27, Vic = 0.93, Yam = 0.97, Extended Data Table 1).

We hypothesized a relationship between rates of global movement and rates of antigenic drift: although rates of genetic evolution were similar for H3N2 and H1N1 viruses, both H1N1 and B viruses evolved antigenically more slowly than H3N2 viruses[13] (Extended Data Table 1). We also hypothesized that lower rates of immune escape for B and H1N1 compared with H3N2 would lead to younger average ages of infection, as children increasingly comprise the largest pool of susceptible individuals, and smaller, less frequent epidemics owing to smaller populations of susceptible individuals[13]. These differences are consistent with results from several community-based cohort studies that found that children were more frequently infected with B viruses than adults[8,16,17]. Age of infection data covering 2002–2011 from Australia show that H1N1 and B viruses infect younger individuals than H3N2 viruses (Extended Data Fig. 5b–d, median age of infection H3N2 = 30 years, H1N1 = 20 years, B = 16 years) and epidemiological data from Australia and the United States show reduced size and frequency of H1N1 and B epidemics compared to H3N2 (Extended Data Fig. 5f–i).

Differences in age of infection may explain differences in global circulation as children travel long distances much less frequently than adults (Extended Data Fig. 5e). A previous study hypothesized that age-specific patterns of infection could lead to differences in contact rates and the spread of influenza types within the United States over the course of a single season[18]. Here, we hypothesized that differential global air travel provides a plausible mechanism by which H1N1 and B

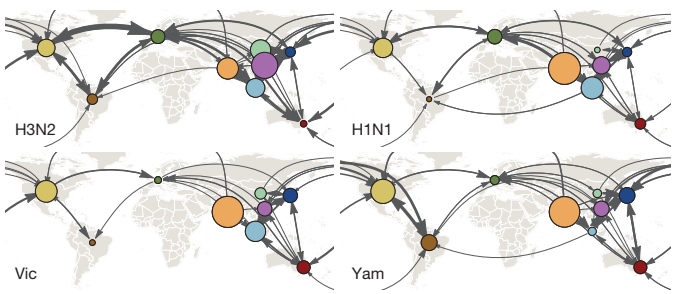


Figure 2 | Estimates of mean pairwise virus migration rate. Line thickness between regions indicates average number of migration events per lineage per year. Arrowhead size indicates the strength of directionality of migration. For clarity, only arrows corresponding to migration rates greater than 0.25 events per lineage per year are shown. Circle area indicates the global proportion of ancestry deriving from each region.

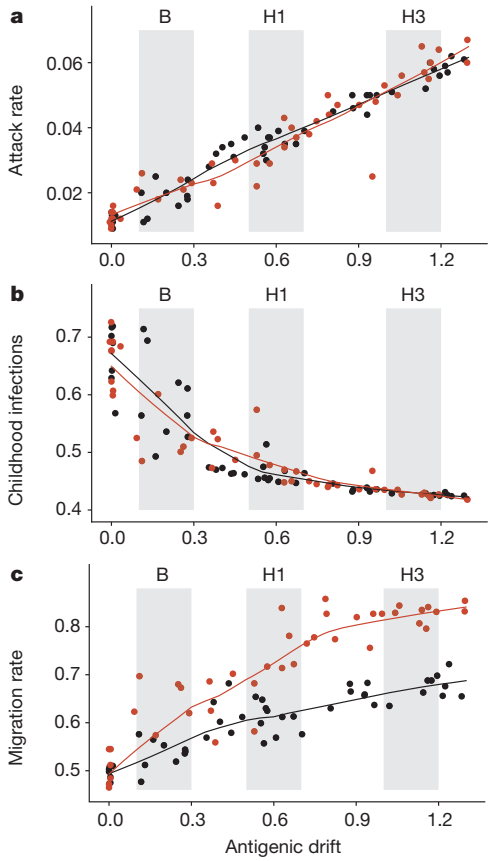


Figure 3 | Relationship of antigenic drift to incidence (a)**, proportion of childhood infections** (b)**, and geographic migration rate** (c)**, in a multi-strain multi-region model of influenza transmission.** Black points represent outcomes from a model in which children and adults travel between regions at equal rates. Red points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children (Extended Data Fig. 5e). Solid black and red lines represent LOESS fits to the data. With 2 travel scenarios, 7 mutation rates and 8 replicates, there are 112 individual stochastic simulations (Extended Data Fig. 7). Antigenic drift was measured in cartographic units[13] per year (see Methods). In **a**, attack rate was measured as proportion of the total population infected yearly. In **c**, migration rate was measured in terms of migration events per lineage per year.

viruses show increased genetic differentiation and reduced rates of global migration across multiple seasons, compared to H3N2 viruses.

To test the impact of differences in age distribution of infection on global patterns of virus movement, we constructed a multi-patch transmission model. We modelled two scenarios for host movement: (1) age-independent mixing between patches; (2) age-stratified mixing with host movement derived from air travel passenger age data (Extended Data Fig. 5e). In the age-independent scenario, model parameters only differed in rate of antigenic mutation, leading to differences in observed rates of antigenic drift among viruses and hence epidemic size and frequency (Extended Data Fig. 7). Faster antigenic drift resulted in greater incidence and more adult infections (Fig. 3a, b), but only modest differences in virus lineage movement (Fig. 3c, B-like viruses differ from H3-like viruses by a factor of 1.2), consistent with slightly faster spread of antigenically novel strains. However, age-stratified mixing between patches intensified the effect of antigenic drift on migration rate and created differences in rates of movement between patches more consistent with those observed for H3N2 vs H1N1 and B (Fig. 3c, B-like viruses differ from H3-like by a factor of 1.6). In the scenario with faster antigenic drift, infections were more mobile owing to greater frequency of adult infection, causing a knock-on effect on rates of viral movement. The model also suggests that the differences in patterns of regional persistence observed in the phylogenies might be shaped by a combination of differences in rates of antigenic evolution and variation in amplitude of epidemic seasonality, with slowly evolving viruses persisting longer than rapidly evolving viruses at low amplitudes of seasonal forcing (Extended Data Fig. 8a, Supplementary Information).

In the model, varying transmission rate rather than antigenic mutation rate also resulted in differences in the observed rate of antigenic drift, with higher transmission resulting in faster drift (Extended Data Fig. 8b). The relationship between antigenic drift rate and migration rate is similar, regardless of whether drift is modulated by mutation rate or transmission rate (Extended Data Fig. 8b). This finding is in line with theoretical work showing that epidemiological processes can influence influenza virus evolution[19,20]. However, there are important virological differences between influenza viruses that are likely to affect the efficiency and tempo at which antigenic variation is generated and fixed, which could in turn affect epidemiology[21–24] (Supplementary Information).

Regardless of the underlying drivers, there is a remarkable correspondence in model behaviour, quantified as a stable relationship between observable rate of antigenic drift and global circulation patterns. The patterns of epidemic spread observed here suggest that differences in ages of infection could explain patterns of global circulation across a variety of human viruses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October; accepted 26 March 2015.
Published online 8 June 2015.

1. Russell, C. A. *et al.* The global circulation of seasonal influenza A (H3N2) viruses. *Science* **320,** 340–346 (2008).
2. Lemey, P. *et al.* Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* **10,** e1003932 (2014).
3. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453,** 615–619 (2008).
4. Bedford, T., Cobey, S., Beerli, P. & Pascual, M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathog.* **6,** e1000918 (2010).
5. Chan, J., Holmes, A. & Rabadan, R. Network analysis of global influenza spread. *PLOS Comput. Biol.* **6,** e1001005 (2010).
6. Nelson, M. I., Simonsen, L., Viboud, C., Miller, M. A. & Holmes, E. C. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog.* **3,** e131 (2007).
7. Nelson, M. I. *et al.* Stochastic processes are key determinants of short-term evolution in influenza a virus. *PLoS Pathog.* **2,** e125 (2006).
8. Glezen, P. W., Schmier, J. K., Kuehn, C. M., Ryan, K. J. & Oxford, J. The burden of influenza B: a structured literature review. *Am. J. Public Health* **103,** e43–e51 (2013).
9. Thompson, W. W. *et al.* Influenza-associated hospitalizations in the United States. *J. Am. Med. Assoc.* **292,** 1333–1340 (2004).
10. Squires, R. B. *et al.* Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir. Viruses* **6,** 404–416 (2012).
11. Chen, R. & Holmes, E. C. The evolutionary dynamics of human influenza B virus. *J. Mol. Evol.* **66,** 655–663 (2008).
12. Cox, N. J. & Bender, C. A. The molecular epidemiology of influenza viruses. *Semin. Virol.* **6,** 359–370 (1995).
13. Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3,** e01914 (2014).
14. Bedford, T., Cobey, S. & Pascual, M. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evol. Biol.* **11,** 220 (2011).
15. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5,** e1000520 (2009).
16. Fox, J. P., Hall, C. E., Cooney, M. K. & Foy, H. M. Influenzavirus infections in Seattle families, 1975–1979. I. Study design, methods and the occurrence of infections by time and age. *Am. J. Epidemiol.* **116,** 212–227 (1982).
17. Longini, I. M. Jr, Koopman, J. S., Monto, A. S. & Fox, J. P. Estimating household and community transmission parameters for influenza. *Am. J. Epidemiol.* **115,** 736–751 (1982).
18. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312,** 447–451 (2006).
19. Recker, M., Pybus, O. G., Nee, S. & Gupta, S. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc. Natl Acad. Sci. USA* **104,** 7711–7716 (2007).
20. Zinder, D., Bedford, T., Gupta, S. & Pascual, M. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS Pathog.* **9,** e1003104 (2013).
21. Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J. Virol.* **80,** 3675–3678 (2006).
22. Neuzil, K. M. *et al.* Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naive 5–8-year-old children. *J. Infect. Dis.* **194,** 1032–1039 (2006).
23. Matrosovich, M. N. *et al.* Probing of the receptor-binding sites of the H1 and H3 influenza A and influenza B virus hemagglutinins by synthetic and natural sialosides. *Virology* **196,** 111–121 (1993).
24. Hensley, S. *et al.* Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* **326,** 734–736 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank National Influenza Centres worldwide for their contributions to influenza virus surveillance. T.B. was supported by a Newton International Fellowship from the Royal Society and through National Institutes of Health (NIH) U54 GM111274. S.R. was supported by Medical Research Council (UK, Project MR/J008761/1), Wellcome Trust (UK, Project 093488/Z/10/Z), Fogarty International Centre (USA, R01 TW008246-01), Department of Homeland Security (USA, RAPIDD program), National Institute of General Medical Sciences (USA, MIDAS U01 GM110721-01) and National Institute for Health Research (UK, Health Protection Research Unit funding). The Melbourne WHO Collaborating Centre for Reference and Research on Influenza was supported by the Australian Government Department of Health and thanks N. Komadina and Y.-M. Deng. The Atlanta WHO Collaborating Center for Surveillance, Epidemiology and Control of Influenza was supported by the US Department of Health and Human Services. NIV thanks A.C. Mishra, M. Chawla-Sarkar, A. M. Abraham, D. Biswas, S. Shrikhande, B. AnuKumar, and A. Jain. Influenza surveillance in India was expanded, in part, through US Cooperative Agreements (5U50C1024407 and U51IP000333) and by the Indian Council of Medical Research. M.A.S. was supported through National Science Foundation DMS 1264153 and NIH R01 AI 107034. Work of the WHO Collaborating Centre for Reference and Research on Influenza at the MRC National Institute for Medical Research was supported by U117512723. P.L., A.R. & M.A.S were supported by EU Seventh Framework Programme [FP7/2007-2013] under Grant Agreement no. 278433-PREDEMICS and ERC Grant agreement no. 260864. C.A.R. was supported by a University Research Fellowship from the Royal Society.

Author Contributions C.A.R. and T.B. conceived the research. C.A.R. and T.B. drafted the manuscript with substantial support from P.L. and S.R. I.G.B., S.B., M.C., N.J.C., R.S.D., C.P.G., A.C.H., A.K., A.Kl. X.L., J.W.M., T.O., V.P., Y.S., M.T., D.W. and X.X. coordinated and produced the influenza surveillance data. T.B. performed the modeling and data analyses along with C.A.R., S.R., P.L., M.A.S. and A.R. T.B. created the figures. All authors discussed the results and contributed to the revision of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.R. (car44@cam.ac.uk).

METHODS

Sequence data. Haemagglutinin (HA) coding sequences for influenza A H3N2 viruses, former seasonal H1N1 viruses (preceding the 2009 pandemic), and influenza B virus lineages Victoria (Vic) and Yamagata (Yam) collected by the World Health Organization (WHO) Global Influenza Surveillance and Response Network including the National Institute of Virology, Pune, India between 2000 and 2012 were combined with human seasonal influenza virus sequences (minimum length = 984 base pairs) covering 2000 to 2012 from the Influenza Research Database[10]. After removing duplicate strains and strains overly divergent based on root-to-tip distances, the data set contained 9,139 H3N2 sequences, 3,789 H1N1 sequences, 2,577 Vic sequences and 1,821 Yam sequences. Sampling locations for these sequences were parsed from strain names. Sequences were grouped into 9 geographic regions: USA/Canada, South America, Europe, India, North China, South China, Japan/Korea, Southeast Asia and Oceania. Specifics of this partitioning are shown in Extended Data Fig. 1. Groups were chosen to maximize available sequences within each region while still providing enough geographic diversity to ensure nearly global coverage. Sequences from Africa, Central America, the Middle East and Russia were excluded because of a lack of sufficient numbers of sequences to provide comparable estimates to other regions.

In the raw sequence data, some regions, such as the USA, were over-represented. Additionally, more recent years were over-represented compared to years at the start of the study period. In order to control for these sampling biases, we subsampled the raw data randomly by location and time to create a more equitable spatiotemporal distribution. The USA had consistently more sequences available every year from 2000 to 2012, thus in order to maintain similar total numbers of sequences for each region across the entire study period it was necessary to sample fewer sequences per year from the USA. We selected 50 sequences per region per year (40 for USA/Canada) for H3N2 and 80 sequences per region per year (45 for USA/Canada) for H1N1, Vic and Yam. This subsampling resulted in largely similar sequence counts across years and across regions for each virus, but overall more H3N2 sequences than H1N1 or B sequences, with 4,006 H3N2 sequences, 2,144 H1N1 sequences, 1,999 Vic sequences and 1,455 Yam sequences (Extended Data Fig. 1). When selecting subsampled sequences we first selected sequences with full day-month-year collection dates and then longer sequences over sequences with less precise dates or shorter sequences. HA sequence data for 1,630 H3N2 isolates, 1,600 H1N1 isolates, 1,394 Vic isolates and 881 Yam isolates have been deposited in the Influenza Research Database[10] and accession numbers for all sequences used provided as Supplementary Information.

Phylogeographic inference. Time-resolved phylogenetic trees were estimated for H3N2, H1N1, Vic and Yam using BEAST v1.8.1[25] and incorporated the SRD06 nucleotide substitution model[26], a coalescent demographic model with constant effective population size and a strict molecular clock across branches. A strict molecular clock was chosen based on finding strong correlations between date of sampling and evolutionary distance in all data sets, as estimated by Path-O-Gen v1.4 (http://tree.bio.ed.ac.uk/software/pathogen/). Using a strict clock also reduced the risk of model over-parameterization (for example, for the complete H3N2 data set with a relaxed clock, there would be $2 \times 4{,}006 - 2 = 8{,}010$ branch-specific rates). Samples with imprecise dates (known only to the month or to the year) had their dates of sampling estimated assuming a uniform prior within the known temporal bounds[27]. Markov Chain Monte Carlo (MCMC) was run for 600 million steps and trees were sampled every 5 million steps after allowing a burn-in of 100 million steps, yielding a total sample of 100 trees for H1N1, Vic and Yam. With significantly more samples, H3N2 required a longer chain to converge. Here, MCMC was run in parallel for 2 chains, each with 650 million steps sampled every 3 million steps with a burn-in of 500 million steps and samples across chains combined, yielding a total of 100 sampled trees. These trees were treated as independent draws from the posterior space of trees when subsequently used in the robust counting and phylogeographic analyses[28]. Evolutionary rates in Extended Data Table 1 were estimated using the 'renaissance' counting methods of Lemey *et al.*[29].

Phylogeographic patterns were estimated using a discrete-state continuous time Markov chain (CTMC) model, in which transition rates were estimated between each pair of regions[15]. We assumed a non-reversible transition model[30] consisting of 72 separate rate parameters, each with a Bayesian stochastic search variable selection (BSSVS) indicator variable, and a separate overall rate of geographic transition. We assumed an exponential prior with mean of 1 for each transition rate, a negative binomial prior with mean of 9 and standard deviation of 9 for the total number of non-zero rates and an exponential prior with mean of 1 migration event per lineage per year for the overall geographic transition rate. MCMC was run for 12 million steps with a burn-in of 2 million steps, and parameters sampled every 10,000 steps and trees sampled every 100,000 steps, yielding a total sample of 1,000 parameter states and 100 trees on which estimates were based. Pairwise migration rate estimates had an effective sample size (ESS) of 350 at the minimum and most had ESS greater than 500.

This procedure yielded posterior trees with the geographic states of internal nodes resolved. We analysed these posterior trees using the program PACT v0.9.5 (https://github.com/trvrb/PACT) to compute the following summary statistics: (1) genealogical diversity[14], measuring the average time it takes for two randomly chosen contemporaneous lineages to coalesce, (2) time to the most recent common ancestor (TMRCA)[14], measuring the average time it takes for all contemporaneous lineages to find a common ancestor, (3) genealogical F_{ST}, measuring the degree of population structure in contemporaneous lineages calculated as $F_{ST} = (\pi_b - \pi_w)/\pi_b$, where π_w is genealogical diversity between randomly sampled lineages from the same geographic region and π_b is genealogical diversity between randomly sampled lineages from different geographic regions, (4) persistence, measuring the average number of years for a tip to leave its sampled location, walking backwards up the phylogeny, (5) migration rate, measuring the average number of migration events over the phylogeny divided by total tree length to give migration events per lineage per year, (6) trunk location through time[4], measuring the posterior distribution across sampled phylogenies of the trunk geographic state, where the trunk is defined as all branches ancestral to viruses sampled within 1 year of the most recent sample, (7) region-specific ancestral geographic history, measuring the distribution of geographic locations of tips belonging to a particular region traced backwards in time through the phylogeny averaged across sampled phylogenies. Statistics (1), (2), (3), (6), and (7) were calculated across 0.1 year genealogical windows. These procedures gave an estimate of credible intervals for inferred ancestral locations across posterior phylogeographic reconstructions.

Code and data availability. Sequence data has been deposited with the Influenza Research Database[10] and accession numbers provided as Supplementary Data. The entire bioinformatic pipeline, including data subsampling, preparing XML files for BEAST, setting up PACT analyses and rendering figures is available at https://github.com/blab/global-migration. Analysis and data files are archived on the Dryad Digital Repository under DOI http://dx.doi.org/10.5061/dryad.pc641.

Surveillance, travel and age-structure data. We investigated epidemic size and frequency using virological isolation data between 2000 and 2012 collected by the WHO Collaborating Centre for Reference and Research on Influenza at the Victorian Infectious Diseases Reference Laboratory (VIDRL), Melbourne, Australia and the Centers for Disease Control and Prevention, Atlanta, USA (Extended Data Fig. 5f–i). These isolations were categorized by date of sampling and by virus type: H3N2, H1N1, Vic, or Yam. The data from VIDRL also contained information on patient age. The age structure of incidence was estimated by constructing a distribution of age of infection from individuals > 5 years (owing to the overrepresentation of < 5 year old patients for all subtypes) (Extended Data Fig. 5b–d). Median age of infection was 30 years (H3N2), 20 years (H1N1) and 16 years (B) and mean age of infection was 33.9 years (H3N2), 23.1 years (H1N1) and 23.2 years (B). Median age of infection was significantly different for H3N2 vs H1N1 ($P = 4.6 \times 10^{-29}$, Mann–Whitney U test), H3N2 vs B ($P = 1.2 \times 10^{-62}$) and H1N1 vs B ($P = 0.041$). The patient age data from VIDRL were potentially biased by testing strategy and the generally higher severity of H3N2 virus infections. Children and working age adults were more likely to be tested than the elderly but the greater severity of H3N2 virus infections might spread and flatten the patient age distribution. For this reason we additionally tested excluding individuals > 65 years and recalculating summary statistics, finding median ages of infection of 27 years (H3N2), 19 years (H1N1) and 15 years (B) and mean age of infection as 28.0 years (H3N2), 22.2 years (H1N1) and 20.3 years (B). We classified children as 0–15 years and adults as 16 years and older, and estimated proportion of childhood infections as 30% (H3N2), 52% (H1N1) and 60% (B). There are potentially other biases specific to individual sentinel physicians and hospitals that could affect sample collection. However, the estimate derived from the VIDRL data that ~60% of influenza B virus infections occur in children is consistent with other estimates (reviewed in Glezen *et al.*[8]). Other studies similarly corroborate the estimates of lower age of infection for H1N1 viruses as compared to H3N2[31,32].

Additionally, we analysed the distribution of ages of ~102.5 million air passengers travelling through London Heathrow and London Gatwick airports in 2011 (Extended Data Fig. 5E) reported by Civil Aviation Authority of the UK (http://www.caa.co.uk/docs/81/2011CAAPaxSurveyReport.pdf). Assuming that children of ages 0 to 15 make up 17% of the UK population (Office of National Statistics), this distribution suggests that children engage in air travel at 19% the rate of adults.

For the modelling described below, we estimated age-structured contact rates following the empirical mixing data provided by Mossong *et al.*[33]. These contact matrices were previously validated in modelling pertussis epidemiology[34]. We simplified the Mossong *et al.* mixing matrices to record child-to-child contacts, child-to-adult contacts, adult-to-child contacts and adult-to-adult contacts, where

children were defined to be 0 to 15 and adults to be 16 or over. This resulted in the following mixing matrix

$$\alpha = \begin{pmatrix} 1.0 & 0.21 \\ 0.21 & 0.26 \end{pmatrix},$$

where rates are relative to child-to-child contact rates.

Epidemiological modelling. An individual-based model of influenza evolution and epidemiology was constructed following methods presented in Bedford *et al.*[35]. The model used here is identical to Bedford *et al.* except where specified below. The present implementation used a linear-strain space[36,37], in which virus phenotype is represented by a continuous variable and cross-immunity between viruses is a function of distance between viruses in this space. We parameterized the model to compare scenarios of age-structured mixing between regions and to compare viruses with different rates of antigenic drift.

The model was simulated for 120 years with daily time steps and the first 100 years discarded to allow equilibrium to be reached. We modelled a metapopulation with individuals equally divided into three regions (North, Tropics, South). Individual's ages were tracked throughout the simulation and those less than 16 years old were classified as children and those 16 or older were classified as adults. Transmission occurred by mass action, with transmission rates modified by regional compartment and by age compartment. Thus, for example, the force of infection into children in the Tropics followed

$$\lambda_{ct} = \sum_{i\in(\mathrm{a,c})} \beta_t \alpha_{\mathrm{ic}} I_{it} \frac{S_{ct}}{N_t} + \sum_{i\in(\mathrm{a,c})} \sum_{j\in(n,s)} \beta_t \alpha_{\mathrm{ic}} m_i I_{ij} \frac{S_{ct}}{N_t},$$

where β_j is the seasonally forced contact rate in region j, α_{ac} represents adult-to-child transmission, m_i represents between-region transmission in age class i, I_{ij} represents the number of persons infected in age class i in region j, S_{ij} represents the number of susceptible persons in age class i in region j, and N_j represents the total number of hosts in region j. The northern and southern regions were seasonally forced in opposite phase with a sinusoidal function following ε, while the tropics had no seasonal forcing.

Each virus possessed a one-dimensional antigenic phenotype ϕ_v and after recovery a host 'remembered' its infecting phenotype. For each contact event, the Euclidean distance from infecting phenotype ϕ_v was calculated to each of the phenotypes in the host immune history $\phi_{h_1},\ldots,\phi_{h_n}$. Here, one unit of antigenic distance was designed to roughly correspond to a twofold dilution of antiserum in a haemagglutination inhibition (HI) assay[38]. The probability ρ that infection occurred after exposure was proportional to the distance d to the closest phenotype in the host immune history, following $\rho = \min\{d\,s, 1\}$. Each day there was a chance μ that an infection mutates to a new phenotype. This mutation rate represents a phenotypic rate, rather than genetic mutation rate, and can be thought of as arising from multiple genetic sources. When a mutation occurred, the virus's phenotype was moved either left or right randomly and mutation size sampled from an exponential distribution with mean step size σ_{avg}. Epidemiological parameters for the baseline epidemiological scenario with notation following Bedford *et al.*[35] were:

- Base transmission rate $\beta = 0.88$ per day
- Duration of infection $1/\nu = 5$ days
- Birth/death rate $= 1/50$ years
- Total population size $N = 45$ million
- Seasonal forcing in north and south $\varepsilon = 0.15$
- Antigenic scaling $s = 0.07$
- Antigenic mutation rate $\mu = 0.5$ to 6.5×10^{-4} per day
- Average mutation size $\sigma_{\mathrm{avg}} = 0.3$ units
- Child-to-child transmission $\alpha_{\mathrm{cc}} = 1.00$
- Child-to-adult transmission $\alpha_{\mathrm{ca}} = 0.21$
- Adult-to-child transmission $\alpha_{\mathrm{ac}} = 0.21$
- Adult-to-adult transmission $\alpha_{\mathrm{aa}} = 0.26$
- Child between-region transmission $m_{\mathrm{c}} = 0.0020$
- Adult between-region transmission $m_{\mathrm{a}} = 0.0020$

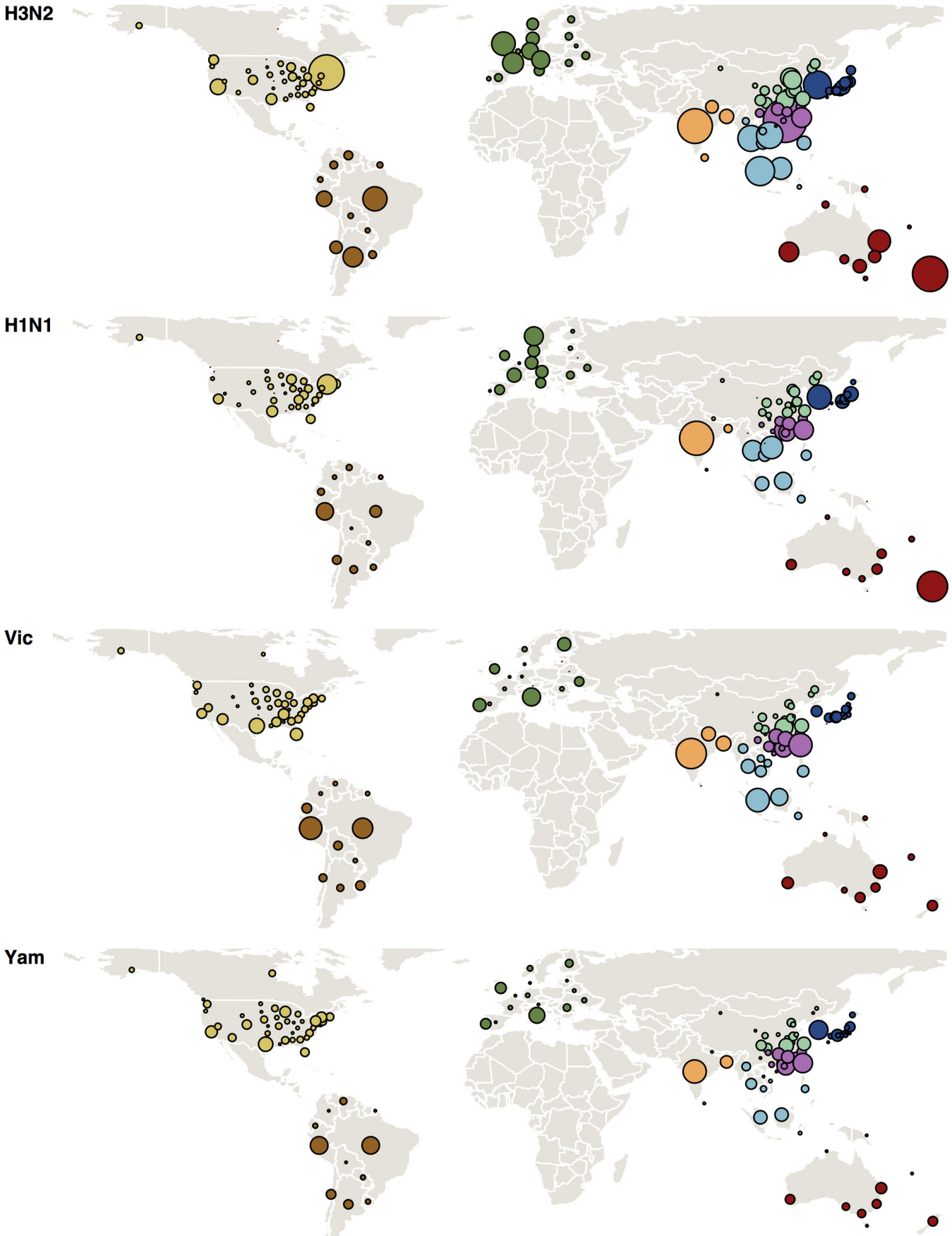
In the model with age-stratified mixing with host movement derived from air travel passenger age data, child between-region transmission m_{c} was 0.0011 and adult between-region transmission m_{a} was 0.0060.

In the course of the simulation, the underlying infection history of who infects whom was recorded and output as a complete infection tree. Without ample within-host diversity owing to chronic infection, the complete infection tree also generated a fully observed phylogenetic tree. Examining geographic location across the phylogenetic tree allowed us to directly calculate migration rate as total migration events observed (transitions from one region to another) divided by total opportunity (tree length).

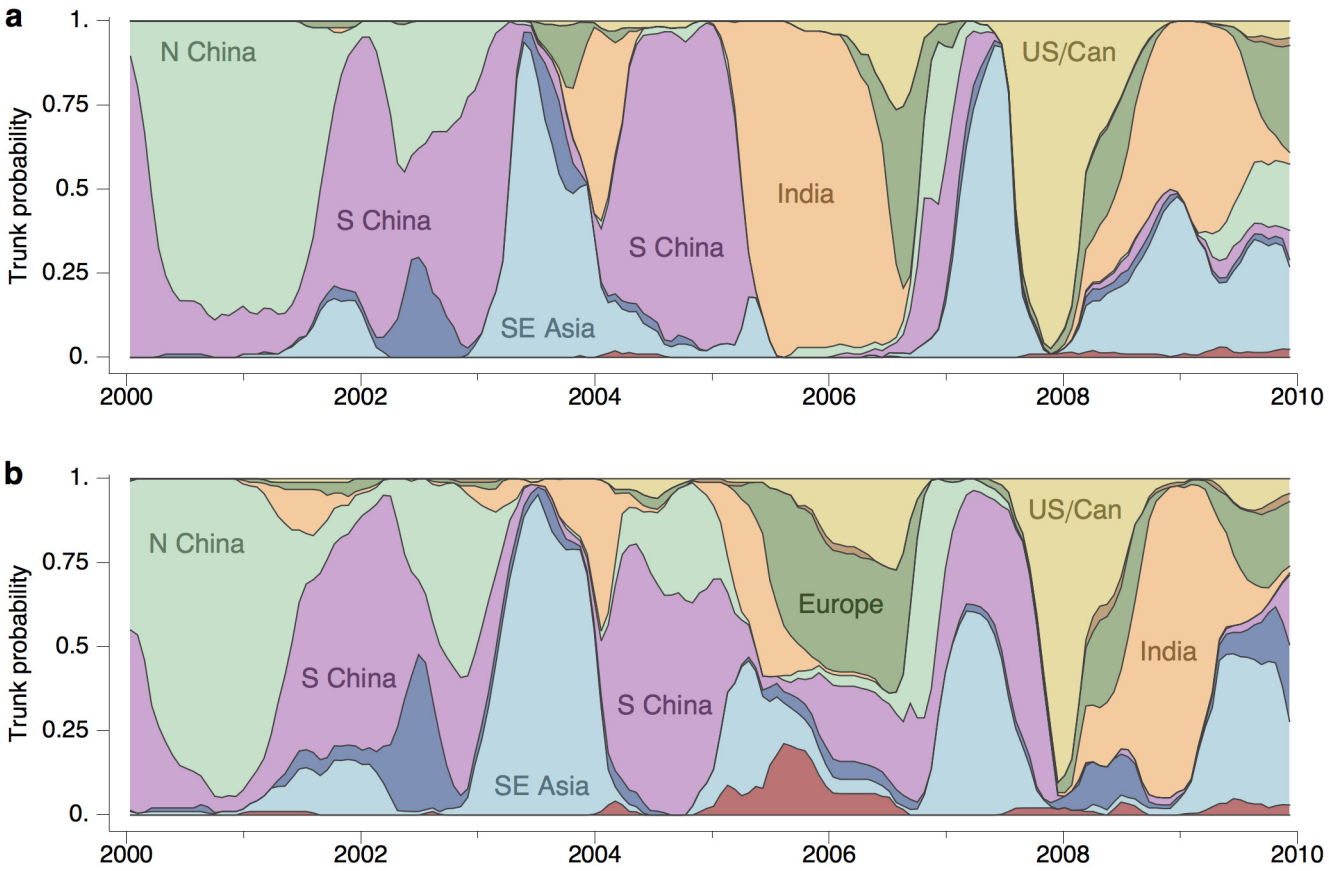
The simulation was parameterized to model H3-like, H1-like and B-like behaviour (Extended Data Fig. 7) by modulating antigenic mutation rate μ in the primary analysis (Fig. 3) or transmission rate β as a secondary analysis (Extended Data Fig. 8b). Values for μ and β were chosen based on observed attack rate, proportion of childhood infections, and antigenic drift rate.

Source code for the simulation is available at https://github.com/trvrb/antigen/tree/global-migration and parameter and results files are available at https://github.com/blab/global-migration/tree/master/model.

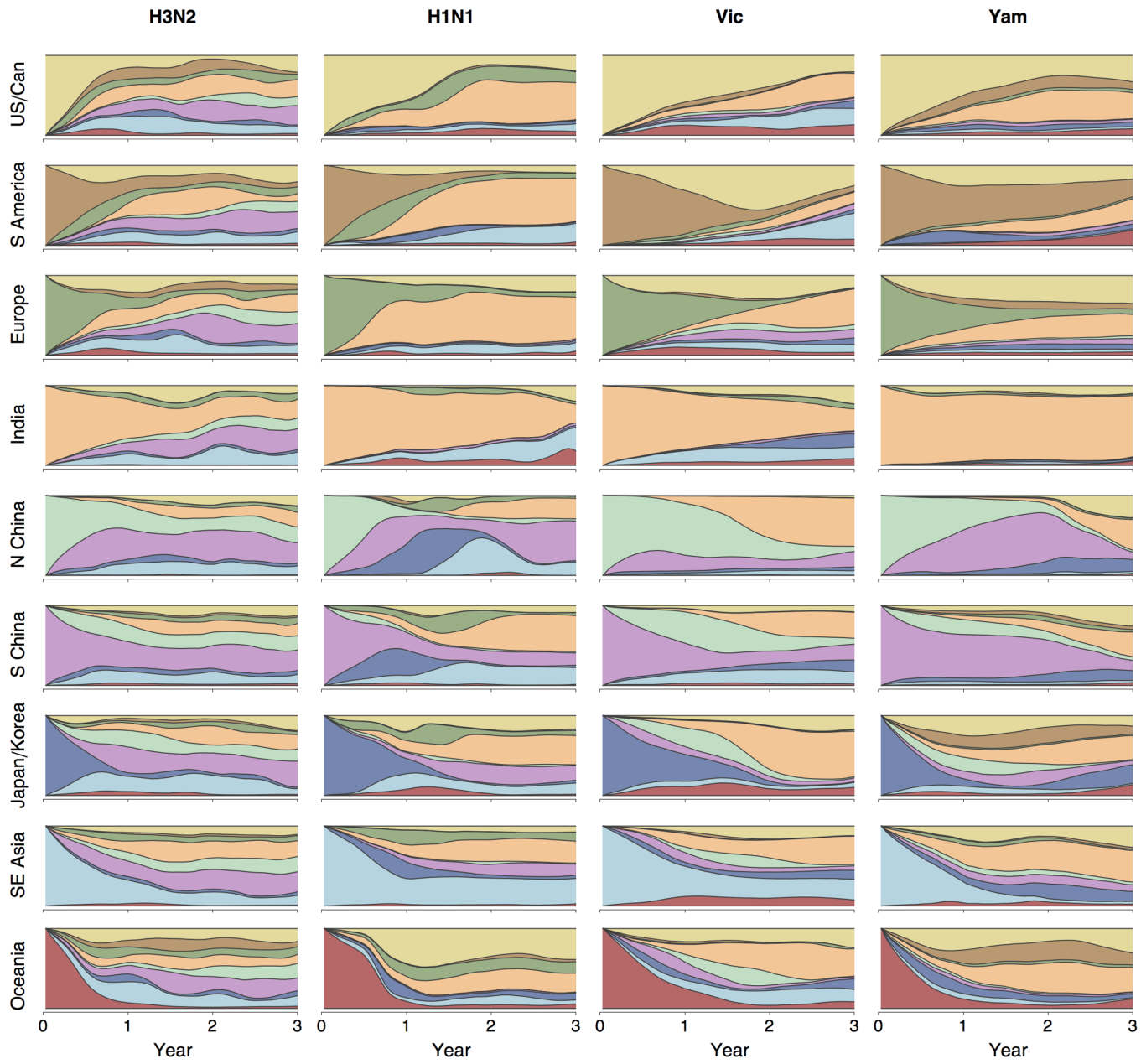
25. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
26. Shapiro, B., Rambaut, A. & Drummond, A. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).
27. Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* **28,** 879–887 (2011).
28. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53,** 673–684 (2004).
29. Lemey, P., Minin, V. N., Bielejec, F., Pond, S. L. K. & Suchard, M. A. A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. *Bioinformatics* **28,** 3248–3256 (2012).
30. Edwards, C. J. *et al.* Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* **21,** 1251–1258 (2011).
31. Kelly, H. A., Grant, K., Williams, S., Fielding, J. & Smith, D. Epidemiological characteristics of pandemic influenza H1N1 2009 and seasonal influenza infection. *Med. J. Aust.* **191,** 146–149 (2009).
32. Khiabanian, H., Farrell, G., St George, K. & Rabadan, R. Differences in patient age distribution between influenza A subtypes. *PLoS ONE* **4,** e6832 (2009).
33. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5,** e74 (2008).
34. Rohani, P., Zhong, X. & King, A. A. Contact network structure explains the changing epidemiology of pertussis. *Science* **330,** 982–985 (2010).
35. Bedford, T., Rambaut, A. & Pascual, M. Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biol.* **10,** 38 (2012).
36. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99,** 17209–17214 (2002).
37. Lin, J., Andreasen, V., Casagrandi, R. & Levin, A. S. Traveling waves in a model of influenza A drift. *J. Theor. Biol.* **222,** 437–445 (2003).
38. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305,** 371–376 (2004).



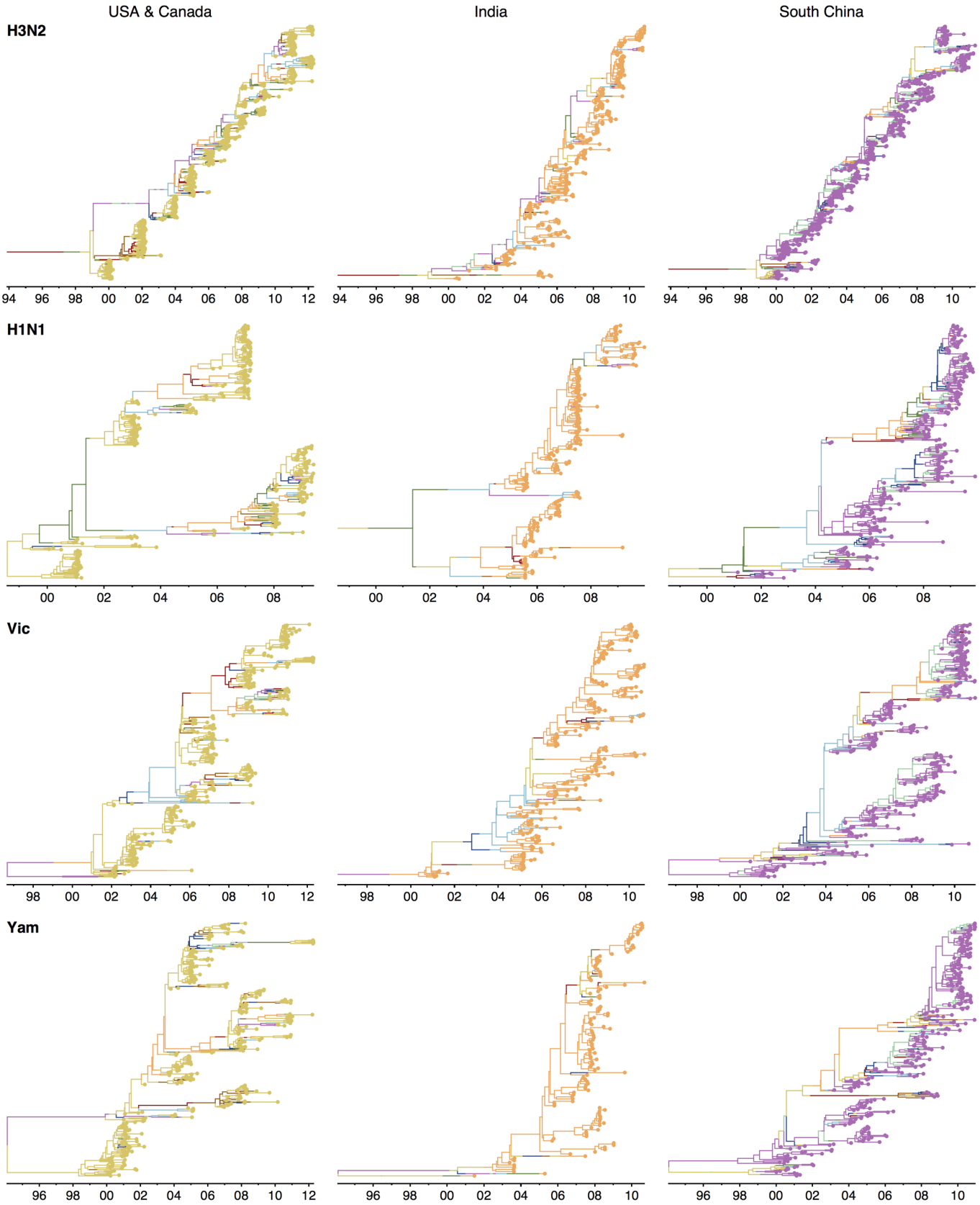
Extended Data Figure 1 | Spatial distribution of 4,006 H3N2, 2,144 H1N1, 1,999 Vic and 1,455 Yam samples. Circle area is proportional to the number of sequenced viruses originating from a location. Colour indicates assignment to one of 9 geographic regions.



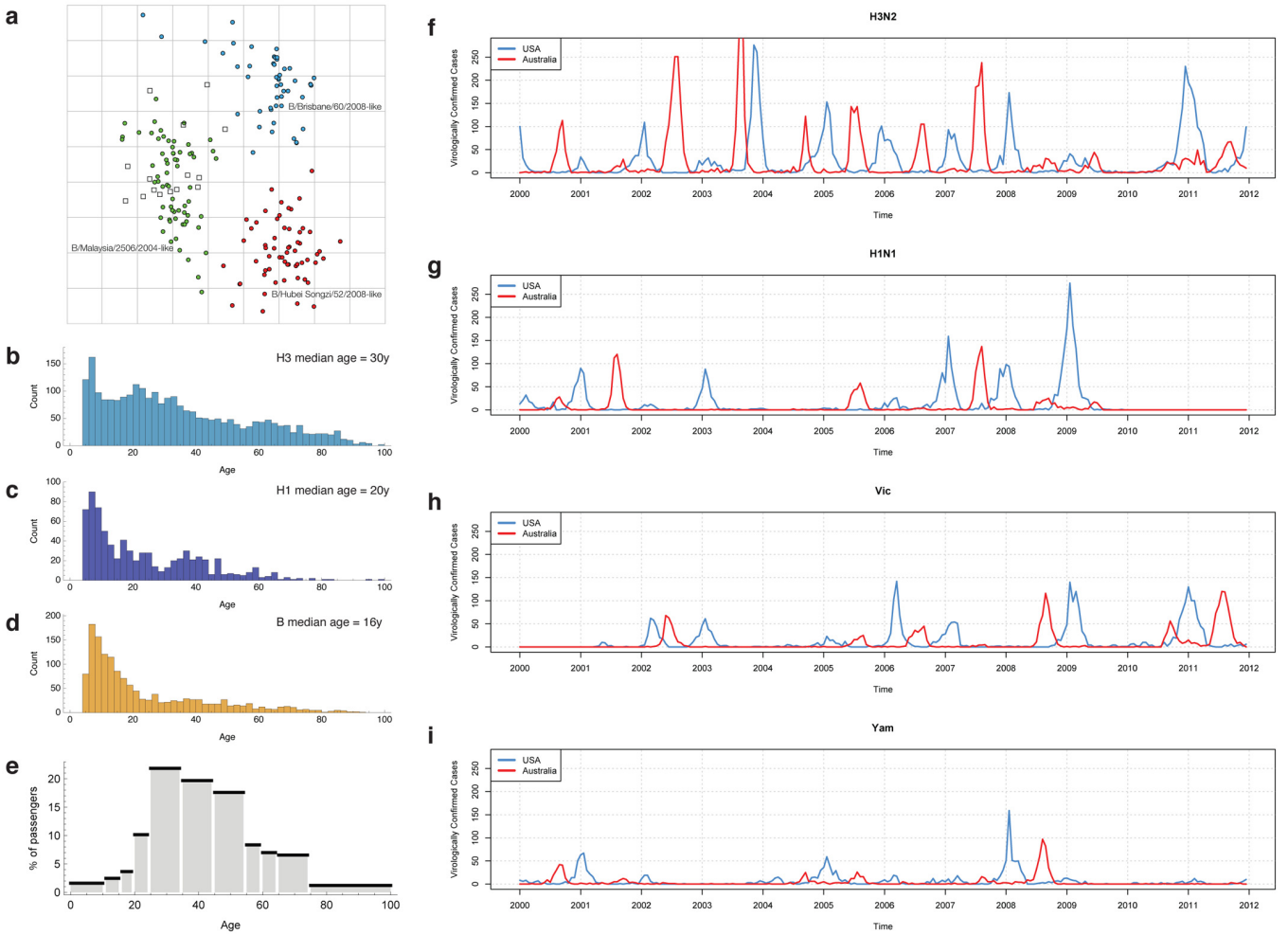
Extended Data Figure 2 | Inferred location of the trunk of H3N2 tree through time in the primary data set (a) **and in a smaller secondary data set** (b)**.** Coloured width at each time point indicates the posterior support for viruses from a particular geographic location comprising the trunk of the phylogenetic tree. Colours correspond to coloured circles in persistence insets in Fig. 1. The secondary data sets consist of 1,391 H3N2 viruses, 1,372 H1N1 viruses, 1,394 Vic viruses and 1,240 Yam viruses.



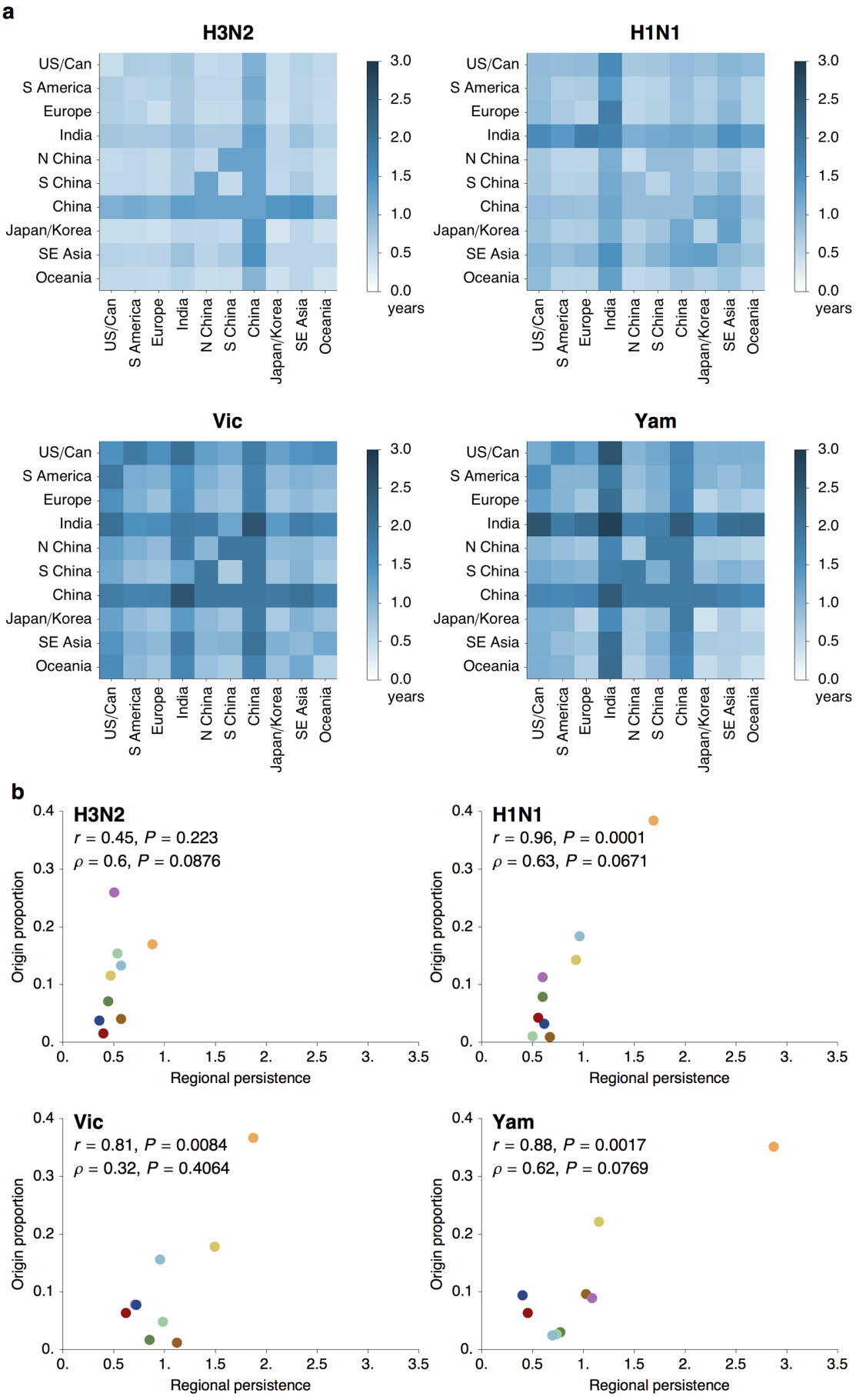
Extended Data Figure 3 | Average inferred geographic history of region-specific samples for H3N2, former seasonal H1N1, Vic and Yam viruses from 2000 to 2012. In each panel, phylogeny tips belonging to a particular region were collected and their phylogeographic histories traced backwards in time averaging across the phylogenetic tree to combine all viruses within each region. The *x*-axis shows number of years backward in time from phylogeny tips from a particular region and the *y*-axis shows the geographic make up as stacked histogram of the ancestors of these tips, where region colour-coding corresponds to the legend in Fig. 1. For example, the top left panel shows the ancestry of USA and Canadian H3N2 viruses. At $x = 0$, all of these viruses are still in the USA or Canada and so an unbroken yellow band takes up the entire *y*. However, at $x = 1$ year, a number of different geographic regions appear on the *y*. This indicates that, 1 year back, ancestors of USA and Canadian viruses are primarily found in Southeast Asia, India and South China. The pattern in the top right panel shows that the ancestors of USA and Canadian Yam viruses more often remain in the USA or Canada with approximately 50% of ancestors remaining 1 year back. Each panel is constructed by averaging across region-specific tips within a tree, but also across sampled posterior trees.



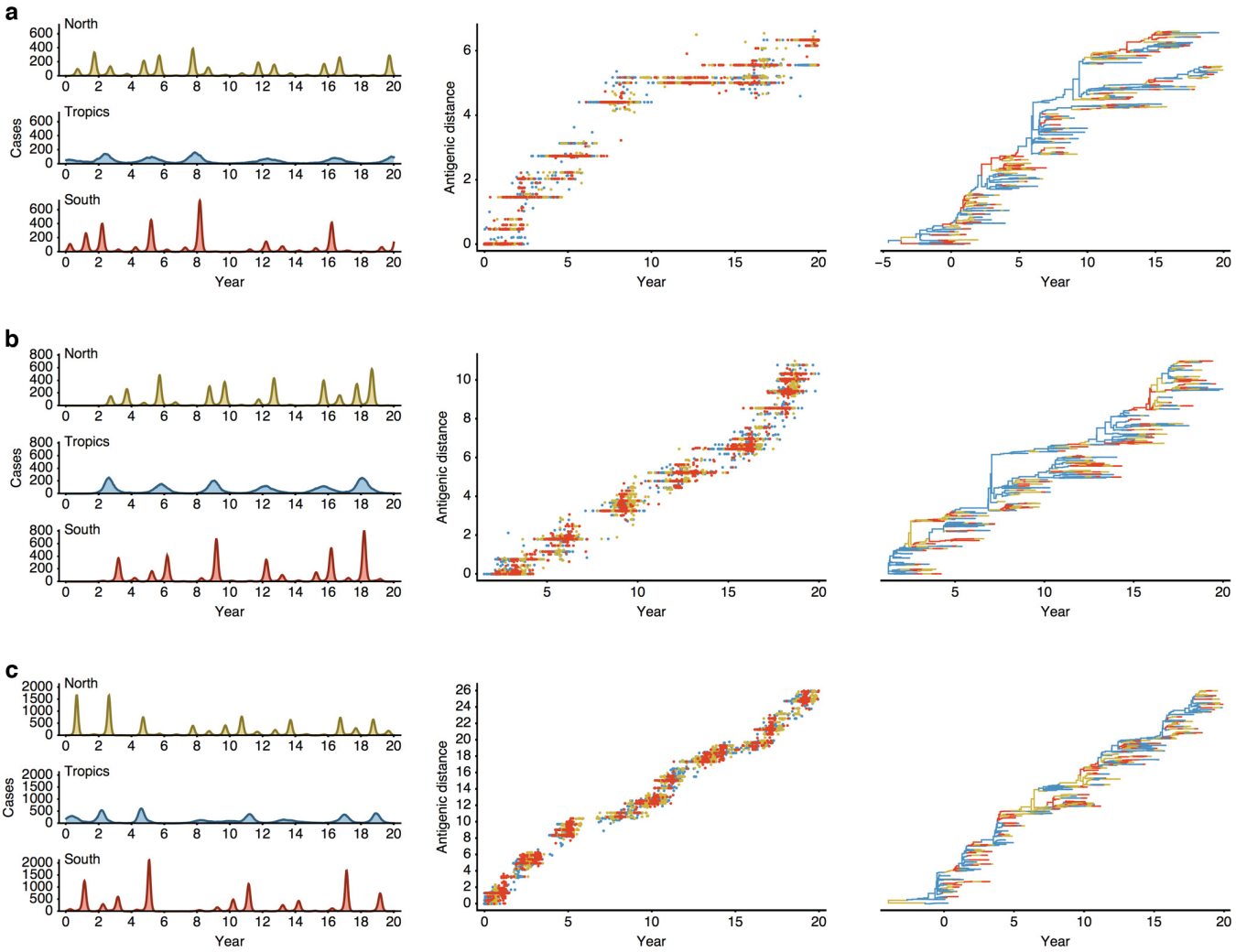
Extended Data Figure 4 | Maximum clade credibility (MCC) trees for region-specific samples from USA/Canada, India and South China for H3N2, H1N1, Vic and Yam viruses. Each tree only contains viruses from a particular geographic region and thus tips are all a single colour within a tree. Branch and trunk colouring have been retained from Fig. 1 to highlight the inferred geographic ancestry of each lineage.



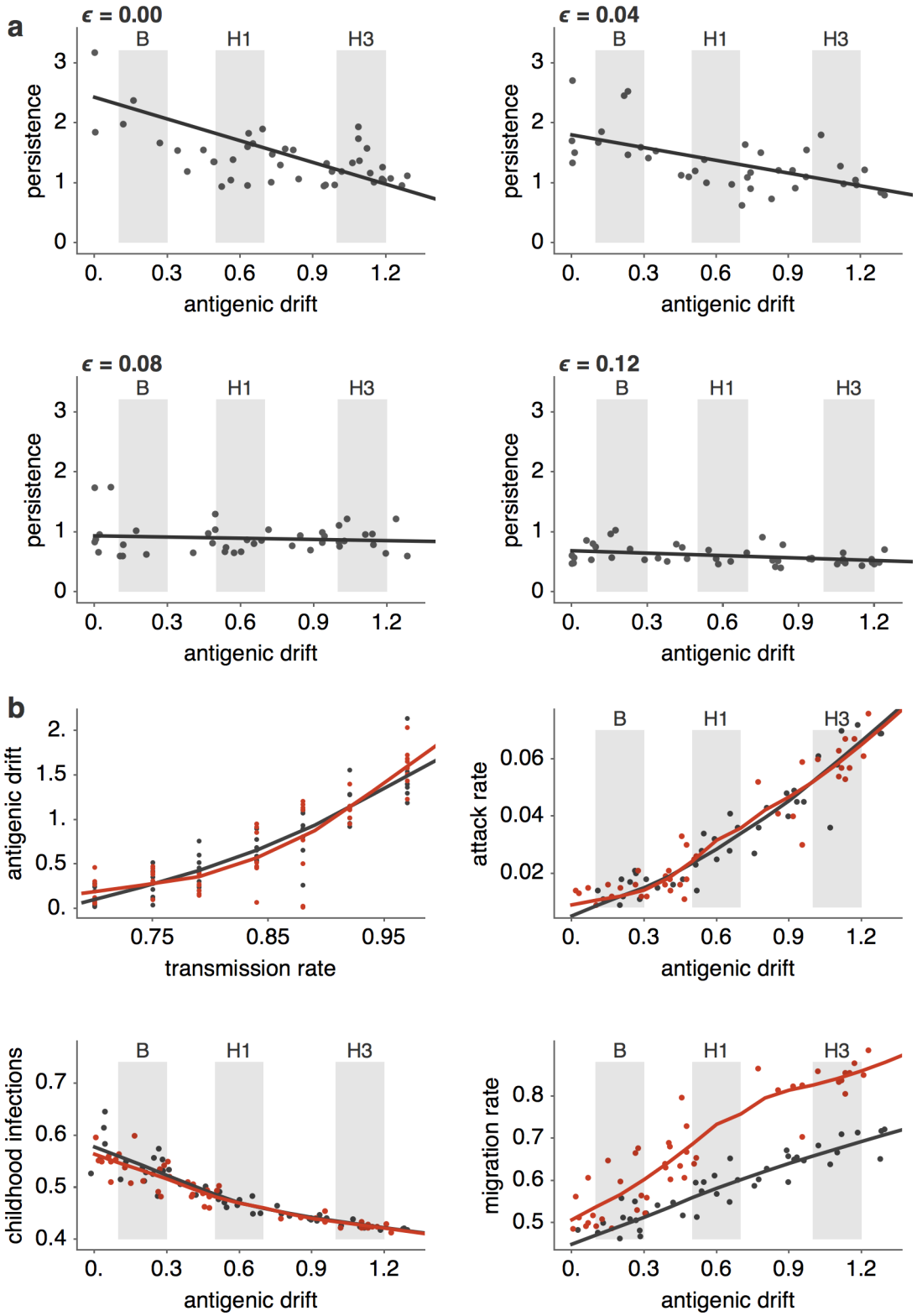
Extended Data Figure 5 | Antigenic map of Vic viruses primarily collected in 2008 (a)**, age distribution of infections for H3N2** (b)**, H1N1** (c) **and B** (d) **in Australia 2000–2011, age distribution of ~102.5 million passengers at London Heathrow and London Gatwick airports during 2011** (e)**, time series of virological characterizations from 2000 to 2012 of viruses from the USA by US CDC and from Australia by VIDRL for H3N2** (f)**, H1N1** (g)**, Vic** (h) **and Yam** (i)**.** In **a**, the positions of strains (coloured circles) and antisera (uncoloured squares) are fit such that the distances between strains and antisera in the map represent the corresponding haemagglutination inhibition (HI) measurements with the least error following Smith *et al.*[38] using data on Vic viruses from the WHO Collaborating Centre for Reference and Research on Influenza at the Centers for Disease Control and Prevention, Atlanta, Georgia, USA. Strains are coloured by antigenic cluster. Genetic clades corresponding to each antigenic cluster are marked with coloured vertical bars in Fig. 1c. The spacing between grid lines is one unit of antigenic distance corresponding to a twofold dilution of antiserum in the HI assay. In **f**–**i**, virological characterizations are a surrogate for epidemiological activity that allow for accurate discrimination among H3N2, H1N1, Vic, and Yam viruses. These data generally reflect the relative magnitudes and frequencies of epidemics but in some cases will inflate magnitudes of very small epidemics due to preferential characterization of subtypes circulating at low levels.



Extended Data Figure 6 | Combined persistence estimates across pairs of regions for H3N2, H1N1, Vic and Yam (a) **and Spearman correlation of a region's persistence vs the region's contribution to phylogenetic ancestry for H3N2, H1N1, Vic and Yam** (b)**.** In **a** and **b**, persistence is measured as the average waiting time in years for a sample to leave its origin backwards in time in the phylogeny, with waiting time averaged across tips within a tree and across sampled posterior trees. In each panel of **a**, the diagonal shows persistence within each of the 9 study regions and within the combined region of 'China', for which nodes in North China and in South China were considered to belong to a single region. The estimates along the diagonal are equivalent to the means shown in Fig. 1. Off-diagonal elements show persistence estimates for pairwise combinations of regions. For example, the off-diagonal for North and South China is exactly equivalent to the diagonal element for 'China' and the off diagonal for 'China' and India represents mean persistence when combining nodes from North China, South China and India. In **b**, origin proportion is measured as the proportion of the time that a region is represented when tracing back 2 or more years from each tip in the phylogeny, averaged across tips within a tree and across sampled posterior trees. Spearman's ρ is not significant for any individual virus. However, the probability of observing 4 instances where each virus has a ρ of at least 0.32 is significant ($P = 0.0017$, bootstrap resampling test).



Extended Data Figure 7 | Simulation results for a model parameterized for slow antigenic drift (a)**, moderate antigenic drift** (b)**, and fast antigenic drift** (c)**.** Colours represent geographic regions with tropics in blue, north in yellow and south in red. Region-specific incidence patterns are shown in terms of cases per 100,000 individuals per week, patterns of antigenic drift in terms of increasing antigenic distance (roughly proportional to $\log_2$ HI units) over time and in the geographically labelled phylogeny. The parameterized antigenic mutation rate is 0.00015 antigenic mutations per infection per day in **a**, 0.00035 in **b** and 0.00055 in **c**, while the realized antigenic drift rate is 0.29 antigenic units per year in **a**, 0.58 in **b** and 1.19 in **c**. Between-region mixing is 5.26× faster in adults. Each panel shows output from a single simulation selected from the 112 shown in Fig. 3, and is intended to show model behaviours over a range of parameters, not necessarily the behaviour of particular viruses.



Extended Data Figure 8 | Simulation results showing relationship between antigenic drift and persistence as a function of seasonality (a) **and simulation results showing the effects of modulating transmission rate β on model behaviour** (b)**.** In **a**, the seasonal forcing parameter ε follows $\varepsilon = 0.00$ (no forcing), $\varepsilon = 0.04$, $\varepsilon = 0.08$ and $\varepsilon = 0.12$ (moderate seasonal forcing). Points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children. Solid black lines represent linear fits to the data. With 4 seasonality scenarios, 7 mutation rates and 8 replicates, there are 224 individual simulations shown. Persistence is measured as the average time in years taken for a tip to leave its region of origin going backwards in time, up the tree. In **b**, transmission rate β in contacts per day is varied and compared to its effect on observed antigenic drift (in antigenic units per year), attack rate per year, proportion of childhood infections and migration rate between regions (in events per viral lineage per year). One antigenic unit is roughly equivalent to one $\log_2$ HI unit. Black points represent outcomes from a model in which children and adults travel between regions at equal rates. Red points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children. Solid black and red lines represent LOESS fits to the data. With 2 travel scenarios, 7 transmission rates and 8 replicates, there are 112 individual simulations shown.

Extended Data Table 1 | Posterior mean estimates (and 95% highest posterior density intervals) across viruses for evolutionary and phylogeographic parameters

Statistic	H3N2	H1N1	Vic	Yam
Total nucleotide rate*	5.0 (4.8–5.2)	4.4 (4.2–4.6)	2.7 (2.6–2.9)	2.8 (2.6–3.0)
Nonsynonymous rate*	2.2 (2.2–2.3)	1.9 (1.9–2.0)	1.0 (0.9–1.1)	1.0 (0.9–1.0)
Synonymous rate*	2.8 (2.7–2.9)	2.6 (2.5–2.7)	1.8 (1.8–1.9)	1.8 (1.8–1.9)
Antigenic drift rate†	1.01 (0.98–1.04)	0.62 (0.56–0.67)	0.42 (0.32–0.52)	0.32 (0.25–0.39)
Diversity‡	3.03	4.59	5.46	6.83
TMRCA§	3.89	4.53	5.22	7.62
F_{ST}‖	0.30	0.36	0.37	0.36
Persistence¶	0.50 (0.48–0.54)	0.79 (0.73–0.85)	1.07 (0.98–1.16)	1.03 (0.88–1.21)
Migration rate#	1.99 (1.85–2.10)	1.27 (1.18–1.37)	0.93 (0.86–1.02)	0.98 (0.83–1.14)

* Evolutionary rates are measured in terms of 10^{-3} substitutions per site per year.
† Antigenic drift rates are from table 2 in Bedford *et al.*[13], and measures cartographic drift per year in terms of twofold dilution of antiserum in a haemagglutination inhibition (HI) assay.
‡ Diversity of contemporaneous lineages is measured as average time in years for two randomly sampled lineages to share a common ancestor.
§ Time to the most recent common ancestor (TMRCA) of contemporaneous lineages is measured as the average time in years for all lineages to find a common ancestor.
‖ F_{ST} compares diversity within regions to diversity between regions, so that $F_{ST} = (\pi_b - \pi_w)/\pi_b$.
¶ Persistence is calculated as the average number of years for a tip to leave its sampled location, walking backwards up the phylogeny.
# Migration rate is calculated as migration events per lineage per year between any two regions.

Extended Data Table 2 | Posterior mean estimates across viruses and data sets of regional persistence, migration rate and geographic population structure

Statistic	Dataset	H3N2	H1N1	Vic	Yam
Persistence*	Primary§	0.51	0.79	1.07	1.03
Persistence*	Secondary‖	0.53	0.75	1.16	1.11
Persistence*	Alternative¶	0.50	0.76	1.28	1.12
Migration rate†	Primary§	1.96	1.27	0.93	0.97
Migration rate†	Secondary‖	1.89	1.33	0.86	0.90
Migration rate†	Alternative¶	2.00	1.32	0.78	0.89
F_{ST}‡	Primary§	0.30	0.36	0.37	0.36
F_{ST}‡	Secondary‖	0.29	0.35	0.36	0.37
F_{ST}‡	Alternative¶	0.29	0.34	0.36	0.35

* Regional persistence is measured as the average waiting time in years for a sample to leave its origin backwards in time in the phylogeny.
† Migration rate is measured as migration events per lineage per year.
‡ F_{ST} compares diversity within regions to diversity between regions, so that $F_{ST} = (\pi_b - \pi_w)/\pi_b$.
§ The primary data sets consist of 4006 H3N2 viruses, 2144 H1N1 viruses, 1999 Vic viruses and 1455 Yam viruses.
‖ The secondary data sets consist of 1391 H3N2 viruses, 1372 H1N1 viruses, 1394 Vic viruses and 1240 Yam viruses.
¶ The alternative data sets consist of 1967 H3N2 viruses, 1439 H1N1 viruses, 1756 Vic viruses and 1223 Yam viruses divided into 10 geographic regions (USA/Canada, South America, Europe, India, Japan/Korea, Southeast Asia, Oceania, China, Central America and Africa).