Migration and Persistence of Human Influenza A Viruses, Vietnam, 2001–2008

Mai Quynh Le, Ha Minh Lam, Vuong Duc Cuong, Tommy Tsan-Yuk Lam, Rebecca A Halpin, David E Wentworth, Nguyen Tran Hien, Le Thi Thanh, Hoang Vu Mai Phuong, Peter Horby, and Maciej F. Boni

Understanding global influenza migration and persistence is crucial for vaccine strain selection. Using 240 new human influenza A virus whole genomes collected in Vietnam during 2001–2008, we looked for persistence patterns and migratory connections between Vietnam and other countries. We found that viruses in Vietnam migrate to and from China, Hong Kong, Taiwan, Cambodia, Japan, South Korea, and the United States. We attempted to reduce geographic bias by generating phylogenies subsampled at the year and country levels. However, migration events in these phylogenies were still driven by the presence or absence of sequence data, indicating that an epidemiologic study design that controls for prevalence is required for robust migration analysis. With whole-genome data, most migration events are not detectable from the phylogeny of the hemagglutinin segment alone, although general migratory relationships between Vietnam and other countries are visible in the hemagglutinin phylogeny. It is possible that virus lineages in Vietnam persisted for >1 year.

Understanding influenza dynamics in tropical regions is crucial for understanding global influenza epidemiology because dynamics between temperate and tropical regions are closely linked. Phylogenetic studies have supported eastern Asia, Southeast Asia, and the tropics as potential ecological sources of global influenza circulation (*1,2*), but others have suggested a variety of geographic regions as potential sources (*3*–*5*). Consequently, the role played by the tropics in the global epidemiology of influenza is still uncertain. Viral gene sequence data from tropical countries are crucial for understanding virus migratory routes within the tropics and between tropical and temperate countries.

Vietnam is an example of a tropical country that may play a major role in global influenza dynamics but for which relatively little is known about influenza epidemiology and genetic population structure of the viruses. Sentinel surveillance suggests that in Vietnam, influenza peaks 1–2 times per year, but neither the influenza-like illness (ILI) data nor the virologic confirmation data show a simple seasonal pattern; the trends for confirmed influenza cases fluctuate more than trends for ILI (*6*,*7*). Serologic studies indicate that annual influenza incidence in Vietnam is between 17% and 26% (*8*). The population of Vietnam is relatively young; according to contact patterns, most cases should occur among younger persons (*9*,*10*). Given Vietnam's high population density and strong travel connections to eastern Asia, Southeast Asia, and Australia/New Zealand, Vietnam is as likely as any other country in eastern or Southeast Asia to support continuous, year-round circulation of a single influenza lineage (persistence) and potentially act as a global source of influenza viruses.

Previous global phylogenetic studies of influenza have demonstrated virus mixing globally (*3*,*4*,*11*), a lack of interseasonal persistence in temperate regions (*1*,*2*,*11*,*12*), and some evidence of persistence in subtropical regions (*5*,*13*). Time-series studies of confirmed influenza suggest (with exceptions [*14*]) that influenza does not exhibit the same strong and regular seasonality in

Author affiliations: National Institute for Hygiene and Epidemiology, Hanoi, Vietnam (M.Q. Le, V.D. Cuong, N.T. Hien, L.T. Thanh, H.V.M. Phuong); Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam (H.M. Lam, M.F. Boni); University of Edinburgh, Edinburgh, Scotland, UK (H.M. Lam); University of Oxford, Oxford, UK (T.T.-Y. Lam, P. Horby, M.F. Boni); The J. Craig Venter Institute, Rockville, Maryland, USA (R.A. Halpin, D. E. Wentworth); and Oxford University Clinical Research Unit, Hanoi (P. Horby)

DOI: http://dx.doi.org/10.3201/eid1911.130349

tropical countries as it does in temperate zones (*15*–*19*) and that it could be constantly circulating throughout the year (*20*,*21*); however, in the latter 2 studies, phylogenetic analyses were not performed. We analyzed 240 newly sequenced influenza virus whole genomes from Vietnam, sampled through the Vietnam National Sentinel Surveillance System during 2001–2008 (*6*). We determined the relative strength of influenza migratory connections between Vietnam and the rest of the world, and we interpreted these results in the context of a sampling bias that seems to affect all sequence-based studies aiming for phylogeographic interpretations. On the basis of frequent sampling in 2007 and 2008, we assessed whether influenza viral lineages persisted in Vietnam during this period. However, we could not definitively conclude whether Vietnam represents a sink or a source population for influenza transmission.

Understanding global influenza migration and persistence patterns is crucial for maintaining a coordinated and efficient biannual strain selection process for influenza vaccine. Choices for future vaccine components will depend on recent availability of samples, and understanding each region's contribution to global influenza circulation will help inform decisions based on viruses coming from highly connected or weakly connected regions.

Methods

Samples

During 2001–2008, as part of the Vietnam National Influenza Surveillance System, nasopharyngeal or throat swab samples were collected from patients seeking care for ILI at hospitals (*6*). Specimens were tested for influenza A and B viruses and were further subtyped for H1, H3, and H5 by reverse transcription PCR by using primers, probes, and reagents recommended by the Centers for Disease Control and Prevention and the World Health Organization (WHO). Samples that were positive for influenza A by PCR were selected for virus isolation, and isolates reaching a titer of 1:8 in hemagglutination assays were selected for sequencing analysis. All isolates were subtyped by using hemagglutination assays with reference antigens and antiserum from the WHO reagent kit. A total of 242 samples were shipped to the National Institutes of Health Influenza Genome Sequencing Project (USA) (*22*) for whole-genome sequencing at the J. Craig Venter Institute. Of the 242 samples, 2 were excluded from this analysis (1 that could not be sequenced and 1 from a patient with a mixed infection). The final dataset of the 240 whole-genome sequences comprised 145 influenza subtypes H3N2 and 95 H1N1 (GenBank accession nos. CY103972–CY105893). Table 1 shows the numbers and locations of the viruses.

Datasets

For phylogeographic analysis, we compiled influenza virus sequences of subtypes H1N1 and H3N2 into 2 datasets: a regional dataset of whole-genome sequences from Asia and Australia/New Zealand and a global dataset of geographically subsampled sequences (50 replicates) of the hemagglutinin (HA) segment (online Technical Appendix, wwwnc.cdc.gov/EID/article/19/11/13-0349-Techapp1.pdf). For subsampling, we randomly sampled 12 sequences per geographic region per year.

Phylogenetic Inference

Sequences were aligned by using the MUSCLE program, version 3.8 (*23*). Maximum-likelihood trees were inferred by using RAxML version 7.3.0 with 2,000 bootstrap replicates (*24*,*25*). For the regional HA datasets, phylogenetic trees with sampling date information were inferred by using BEAST version 1.6.2 (*26*) and a relaxed molecular clock (uncorrelated lognormal). The nucleotide substitution model was SRD06 + HKY85 + Γ, and the demographic models used were constant population size and Bayesian skyride (online Technical Appendix).

Analysis of Regional Migration

Migration analysis was conducted by using a straightforward parsimony method in the PAUP* program (*27*,*28*). The 2,000 bootstrap trees and the best maximum-likelihood tree inferred by RAxML were read into PAUP*, and nucleotide sequences were replaced by single-letter location codes assigned to a set of predefined global regions (online Technical Appendix Figures 5, 6). Changes in location code were mapped onto the branches of the trees

Table 1. Influenza viruses successfully sequenced, Vietnam, 2001–2008

Virus subtype, geographic region	Year							
	2001	2002	2003	2004	2005	2006	2007	2008
H3N2								
Northern	1	0	4	18	4	0	66	0
Central	0	0	9	2	19	1	8	13
Southern	0	0	0	0	0	0	0	0
H1N1								
Northern	6	2	24	0	1	6	0	6
Central	0	0	1	0	2	4	0	36
Southern	0	0	0	0	0	6	0	1

by using an ACCTRAN parsimony criterion (*28*). Analysis was performed on all segments to determine whether migration histories differed among them, which could have been caused by reassortment among influenza virus RNA segments. A strong reassortment signal was verified when standard mosaic and phylogenetic criteria were used (*29*) (online Technical Appendix Figure 2).

Analysis of Global Migration

For comparison, we performed the same migration analysis on the global dataset of 50 subsampled replicates for influenza virus subtypes H3N2 and H1N1. Migration matrices were built describing numbers of connections between 27 subtype H3N2 or 29 subtype H1N1 predefined geographic regions. We used Gephi software (*30*) to visualize the connections in the matrix. For regions with sufficient samples, we computed minimum distances to the trunk of the rooted phylogeny for all 50 subsampled trees (subtype H3N2 only) to determine whether viruses from different regions could be described as ancestral (close to the trunk) or derived (far from the trunk).

Results

Regional Migration of Influenza Virus (H3N2) HA

The relationship between the subtype H3N2 HA sequences from Vietnam and other viruses sampled in the region is shown in Figure 1. Representative samples from Vietnam are available for 2003–2008 but not for 2006, when subtype H1N1 predominated. Viruses isolated in Vietnam show close relationships to viruses isolated in Hong Kong, Taiwan, Singapore, and Australia/New Zealand. Inferred from this tree were 20 parsimony-unambiguous migration events, 9 showing Vietnam–Hong Kong migration, 7 showing Vietnam–Australia/New Zealand migration, 2 showing Vietnam–Taiwan migration, 1 showing Vietnam–South Korea migration, and 1 showing Vietnam–Qatar migration. Clearly, because Hong Kong (68 sequences) and Australia/New Zealand (>500 sequences) were overrepresented in the regional dataset, most Vietnam migrations were associated with these 2 locations. Because we were initially uncertain how well a geographically regional phylogenetic tree would reflect the true migration patterns of influenza

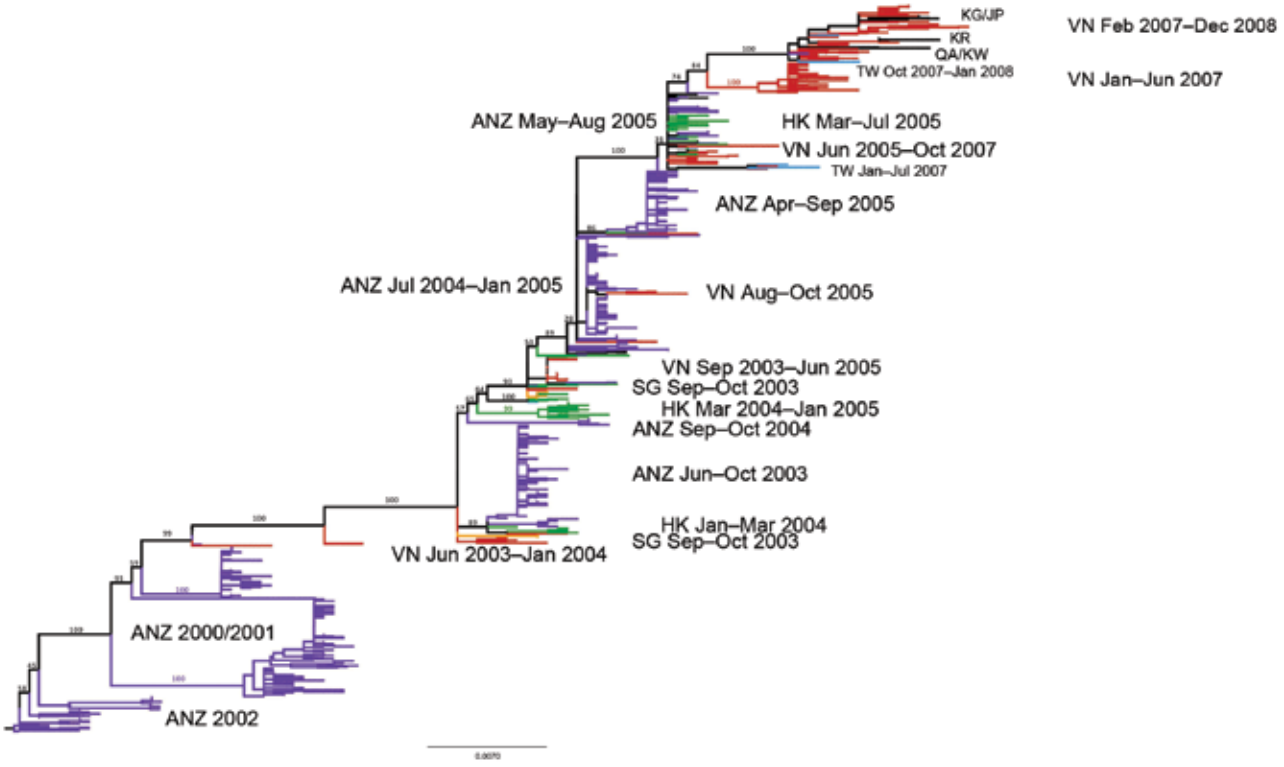


Figure 1. Maximum-likelihood phylogenetic tree (hemagglutinin segment) of the 787 sequences that comprise the regional influenza (H3N2) dataset. Tree is rooted on A/Canterbury/179/1999, and bootstrap values are shown on key nodes. Branches are colored by location: red, Vietnam; purple, Australia or New Zealand; green, Hong Kong; blue, Taiwan; orange, Singapore. Labels are shown directly to the left or right of the clades they are describing, with 2 exceptions: the label "VN Jun 2003–Jan 2004" refers to the viruses directly above it, and the label "ANZ 2000/2001" refers to the 2 major clades below it and above it. KG, Kyrgyzstan; VN, Vietnam; JP, Japan; KR, South Korea; QA, Qatar; KW, Kuwait; TW, Taiwan; HK, Hong Kong; ANZ, Australia and New Zealand. Scale bar indicates nucleotide substitutions per site.

viruses in Vietnam, we performed a validation exercise to determine what proportion of global migration of influenza subtype H3N2 virus from Vietnam could be detected in a regional phylogenetic tree (online Technical Appendix). Approximately 70% of Vietnam influenza migrations from a global analysis were also observed in the regional tree.

Regional Migration of Influenza Virus (H1N1) HA

The relationship between influenza subtype H1N1 HA sequences from Vietnam and other regional viruses is shown in Figure 2. Representative samples from Vietnam are available for 2001–2008 but not for 2004 and 2007. Inferred from this tree are 10 parsimony-unambiguous migration events, 6 showing Vietnam–Taiwan migration, 2 showing Vietnam–Japan migration, and 2 showing Vietnam–Australia/New Zealand migration. As in the analysis for subtype H3N2, these migration links correspond with the viruses that were sequenced from the region during 2001–2008.

Whole-Genome Migration Patterns

Because the regional trees included only sequences for which whole genomes were available, migration patterns were compared systematically across all 8 influenza segments. Because influenza viruses reassort, different event histories should be visible in phylogenies inferred separately for the 8 virus segments. Indeed, for the subtype H3N2 dataset, we observed a median of 14 parsimony-unambiguous migration events for the neuraminidase segment and a median of 41 for the matrix protein segment; the other segments fell somewhere in between (Table 2). Again, most migrations were with Australia/New Zealand and Hong Kong, indicating that the pattern of migration is similar across segments, although different numbers of migrations and different individual migration events are visible when different segments are analyzed. For the matrix protein and nonstructural protein segments of subtype H3N2 viruses, the large number of migration events may result from the larger number of topologically uncertain and polytomic nodes in these trees that had to be randomly resolved to compute the number of migration events; that is, sequences from one country could be mistakenly mixed with sequences from other countries, thus generating some artificial migration events in the parsimony analysis. The low confidence in the

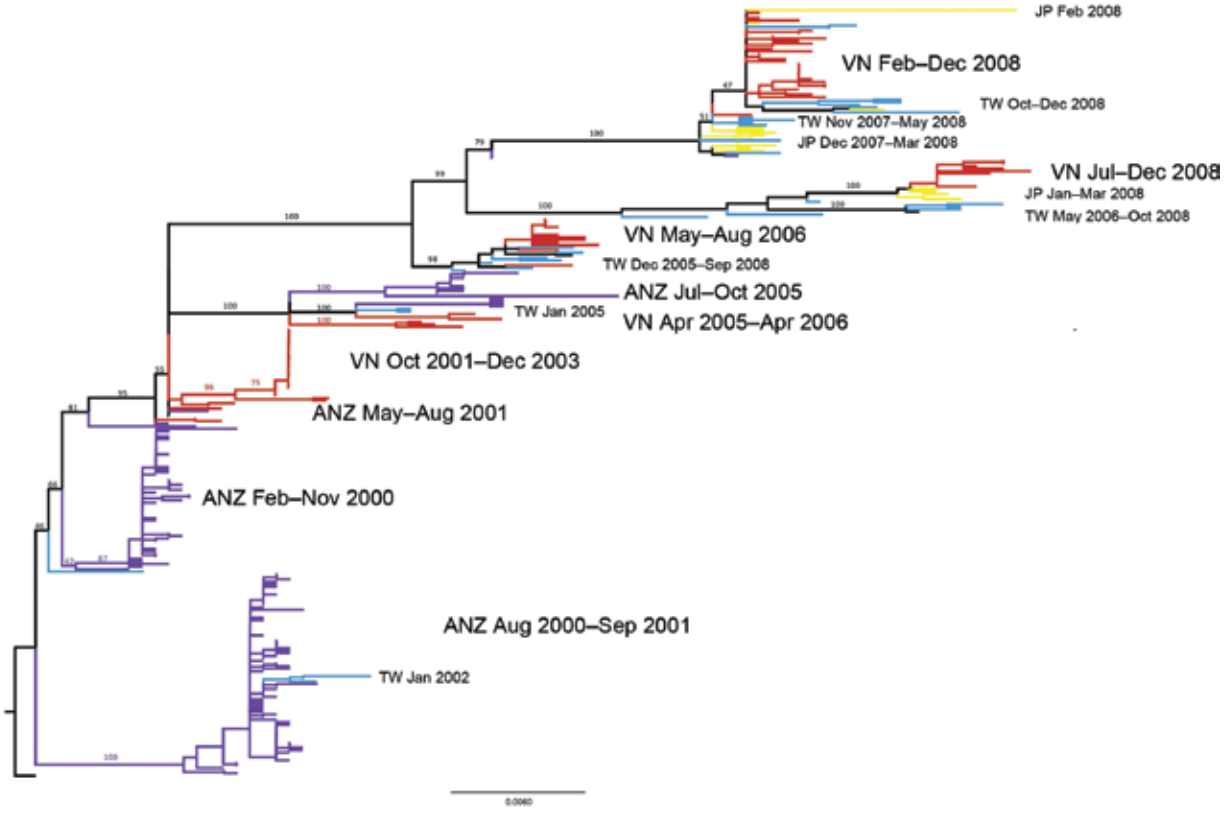


Figure 2. Maximum-likelihood phylogenetic tree (hemagglutinin segment) of the 300 sequences that comprise the regional influenza (H1N1) dataset. Tree is rooted on A/New Caledonia/20/1999, and bootstrap values are shown on key nodes. Branches are colored by location: red, Vietnam; purple, Australia or New Zealand; yellow, Japan; blue, Taiwan. Labels are shown directly to the left or right of the clades they are describing. JP, Japan; VN, Vietnam; TW, Taiwan; ANZ, Australia and New Zealand. Scale bar indicates nucleotide substitutions per site.

Table 2. Observed migration of influenza virus between Vietnam and other countries*

Virus subtype, country	Median no. events (95% range)							
	PB2	PB1	PA	HA	NP	NA	MP	NS
H3N2								
Aus/NZ	9 (5–13)	9 (5–13)	7 (4–10)	7 (3–10)	11 (6–16)	5 (2–8)	26 (19–33)	27 (18–36)
Hong Kong	4 (1–7)	4 (1–8)	6 (2–10)	5 (2–8)	6 (2–10)	7 (3.5–12)	12 (6–18)	5 (1–10.5)
Taiwan	2 (1–3)	3 (2–4)	3 (2–4)	3 (1–4)	2 (1–4)	1 (0–3)	3 (1–5)	2 (1–5)
Singapore	1 (0–3)	1 (0–3)	0 (0–2)	1 (0–3)	1 (0–6)	1 (0–3)	0 (0–2)	1 (0–3)
H1N1								
Taiwan	3 (1–6)	3 (1–5)	4 (2–7)	5 (2–7)	5 (2–8)	4 (2–8)	5 (2–10)	5 (2–8)
Aus/NZ	2 (0–5)	1 (0–4)	2 (0–5)	2 (0–5)	4 (1–8)	2 (0–4)	5 (2–9)	2 (0–6)
Japan	2 (1–4)	2 (1–4)	2 (1–4)	3 (1–4)	4 (1–8)	1 (0–2)	6 (2–9)	4 (2–7)

*Data from 2,000 bootstrapped trees for all 8 segments of regional datasets for influenza subtypes H3N2 and H1N1. PB, polymerase basic protein; PA, polymerase acidic protein; HA, hemagglutinin; NP, nucleocapsid protein; MP, matrix protein; NS, nonstructural protein; Aus/NZ, Australia/New Zealand.

Vietnam–Singapore migration link for subtype H3N2 may result from the small number of whole-genome sequences available from Singapore, all of which were collected in 2003.

For subtype H1N1 viruses, we observed 6–16 migration events across the trees inferred for the 8 segments (Table 2). In the bootstrapped data, the Vietnam–Taiwan and Vietnam–Japan migratory connections seem to be approximately equal, despite the fact that the best maximum-likelihood tree showed 6 Vietnam–Taiwan connections and 2 Vietnam–Japan connections. The migratory connection between Vietnam and Australia/New Zealand seems to be somewhat weaker, possibly because of substantially less sampling of Australia/New Zealand viruses in the subtype H1N1 dataset. In the subtype H1N1 dataset, the number of migration events for the matrix protein and nonstructural protein segments did not increase.

Migration in Subsampled Global HA Trees

Because it is clear that the presence and number of samples from different regions influence migration analysis, migratory patterns were reanalyzed on the global subsampled dataset to reduce the geographic bias present from having higher numbers of samples available from some regions than others. Using global HA trees for subtype H3N2 and H1N1 sequences, we constructed full migration matrices including all parsimony-unambiguous migration events among our predefined regions (online Technical Appendix Figures 5, 6). These migration networks are shown in Figure 3, where the United States is a major hub of influenza migration and eastern Asia and Australia/New Zealand play major roles. The subtype H3N2 data show Vietnam connected with most other countries in eastern and Southeast Asia, with the United States, and weakly with southern Asia. The subtype H1N1 data show Vietnam connected with the United States and Europe but weakly with other Asian countries. For both subtypes, the total number of migration events associated with each node in the network is correlated with the number of samples available for that node (all p values were $<10^{-5}$; Kendall and Spearman tests). Note that sample numbers are not identical for each node because for some regions <12 sequences per year were available, and these regions did not need to be subsampled for those years. Hence, undersampling and oversampling can generate this correlation. Despite our attempt to reduce geographic bias in the global dataset, inference on migration events is still closely associated with regional availability of samples; this bias appears to affect all phylogeographic studies.

For H3N2 sequences, to determine whether any region has the characteristics of an ecological source, we computed the phylogenetic distance of sequences from 6 well-sampled regions (China, Hong Kong, Japan, Vietnam, Australia/New Zealand, and the United States) to the trunk of the global maximum-likelihood phylogenetic tree (Figure 4). In 2003, for example, across all 50 subsampled trees, sequences isolated in China were typically closest to the trunk of the phylogenetic tree, indicating that these sequences are ancestral to other viral sequences sampled in 2003; this finding is consistent with the global replacement of subtype H3N2 viruses by the A/Fujian/411/2002-lineage that occurred in 2003. In general, for the years 2003–2007, in no region were sequences consistently ancestral, indicating that it is unlikely that there is a single global source of human influenza viruses. The more likely global migration model involves periodic global strain replacements originating in different regions in different years (*3,4*). There were not sufficient samples from all regions/years to perform this analysis on the subtype H1N1 dataset.

Lineage Persistence

Figure 5 shows a Bayesian subtype H3N2 phylogenetic tree inferred from the time-stamped regional sequence data. The insets of this figure detail the 2007–2008 part of the phylogeny (87 sequences) and the coalescent times for the tips of these branches. It is difficult to draw a complete persistence picture for these viruses because of undersampling during the second quarter of 2007 and the first half of 2008 despite PCR-confirmed evidence of influenza virus activity during these periods (*6*). The median coalescent time for viruses from Vietnam sampled during this period

is 37 days (interquartile range 21–72 days), and the insets in Figure 5 suggest that one of the lineages persisted in Vietnam for the 13 months from January 2007 through January 2008. An absence of samples from February through May 2008 makes it impossible to determine conclusively if this lineage persisted in Vietnam for the entire 2-year period.

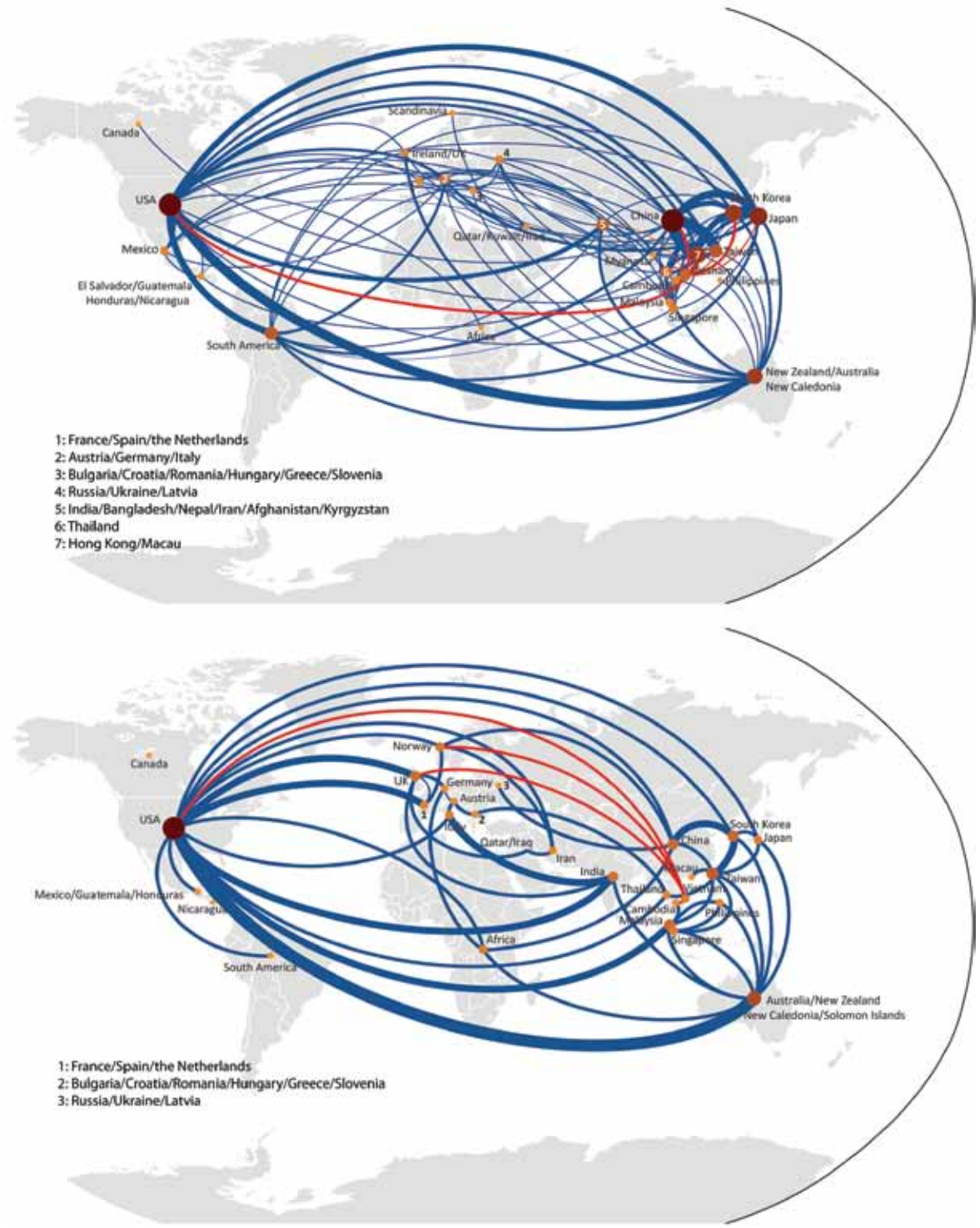


Figure 3. Global migration maps from fully subsampled global hemagglutinin tree for A) influenza (H3N2), based on 1,140 sequences, and B) influenza (H1N1), based on 554 sequences. The size and color of the nodes corresponds to the number of migration events associated with that location (median from 50 subsamples). The thickness of the lines corresponds to the number of migration events between 2 nodes. Red lines join Vietnam to other locations; blue lines join other locations. UK, United Kingdom; USA, United States.

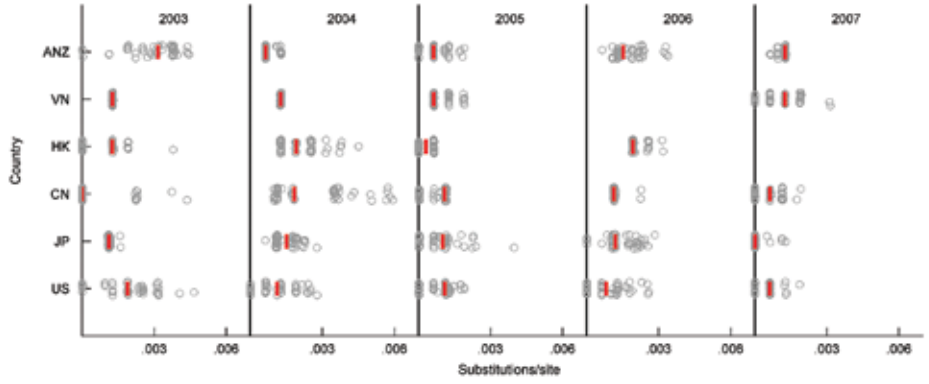


Figure 4. Minimum phylogenetic distance to the trunk, computed for the 50 subsampled global influenza (H3N2) phylogenies. Minimum distances are shown by year and by region, for 6 regions with sufficient sampling during 2003–2007. ANZ, Australia/New Zealand; VN, Vietnam; HK, Hong Kong; CN, China; JP, Japan; US, United States. Red lines show medians across 50 subsamples. For Vietnam in 2006 and Hong Kong in 2007, there were insufficient virus sequences.

To determine whether the lack of samples from other countries created an artifactual picture of lineage persistence in Vietnam during 2007–2008, we assembled a sequence set of all 672 viruses from Asia and Australia/ New Zealand from 2006 through 2008. The maximum clade credibility tree of these sequences (online Technical Appendix Figure 4) indicates that the Vietnam lineages separate into >10 distinct lineages when viewed in the context of all Asian/ Australia/New Zealand influenza viruses. One of these lineages persisted in Vietnam for 15 months (online Technical Appendix Figure 1 panel A), and another persisted for 10–12 months (online Technical Appendix Figure 1, panel B), suggesting that lineage persistence of >1 year may have occurred in Vietnam during 2007–2008. However, this type of analysis is very sensitive to phylogenetic uncertainty because the individual lineages (or subclades) contain few sequences and may not be robust to small changes in tree topology.

Discussion

According to our analysis, the major migratory routes of influenza virus pass through the United States, eastern Asia, and Australia/New Zealand. Europe—despite its population density and consistency of wintertime influenza epidemics—was slightly less connected to other parts of the world when compared with the United States. These results are consistent with those of previous studies that showed eastern Asia (*2*) and tropical Asia (*1*) as key influenza source populations and the United States as a major contributing region (*3*). The new sequence data in this analysis support strong migratory connections between Vietnam and neighboring countries, the United States, and Europe. Our regional phylogenetic analysis supports a strong connection between Vietnam and Australia/New Zealand, but the global analysis reveals that Australia/New Zealand sequences are more closely related to sequences from Asian countries other than Vietnam. In addition, the inferred phylogenies provide evidence of virus persistence in Vietnam for >1 year. This is a major finding because strong migratory links and persistence are the 2 key features for a proposed source region for influenza transmission; long-term persistence in tropical regions may be associated with more antigenic evolution and immune escape if it can be shown that longer persistence gives the virus population more time to accumulate and fix antigenic changes (*2,31,32*).

In general, persistence analyses are difficult even with regular sequence sampling and weekly virologic confirmations. When attempting to assess the likelihood of influenza persistence in a focal region (e.g., Vietnam), we must sample outside the focal region to determine whether local viruses have been reintroduced from elsewhere. However, the more sampling in the nonfocal region, the more likely it becomes that we sample nonfocal viruses similar to focal viruses and that more diversity is detected in the nonfocal region, making it seem basal (closer to the root) to the focal region. There are no clear criteria for whether we have undersampled or oversampled the focal or the nonfocal region; thus, it is extremely difficult to state with certainty that an apparently local lineage has persisted in the same location. For the 2007–2008 Vietnam influenza sequences, viruses were sampled for most of this period and coalescence times were generally short, indicating that most of these viruses have a relatively recent ancestor in Vietnam. These data are consistent with and provide evidence for lineage persistence in Vietnam during this time. However, we know of no unbiased test that can reject the possibility of virus introduction. The perfect dataset for demonstrating lineage persistence would seem to be 52 viruses sampled in 52 weeks, with consecutive viruses differing at 0 or 1 nt positions.

A major limitation of all migration analyses performed with sequence data is geographic sampling bias: undersampling and oversampling. The more sequences that are available for a given location, the more likely it is that 1 of these sequences will be a recent immigrant, identifiable by the presence of similar sequences from other locations. To overcome this bias, subsampling is typically conducted (*3,5*) to ensure that the same numbers of sequences are used

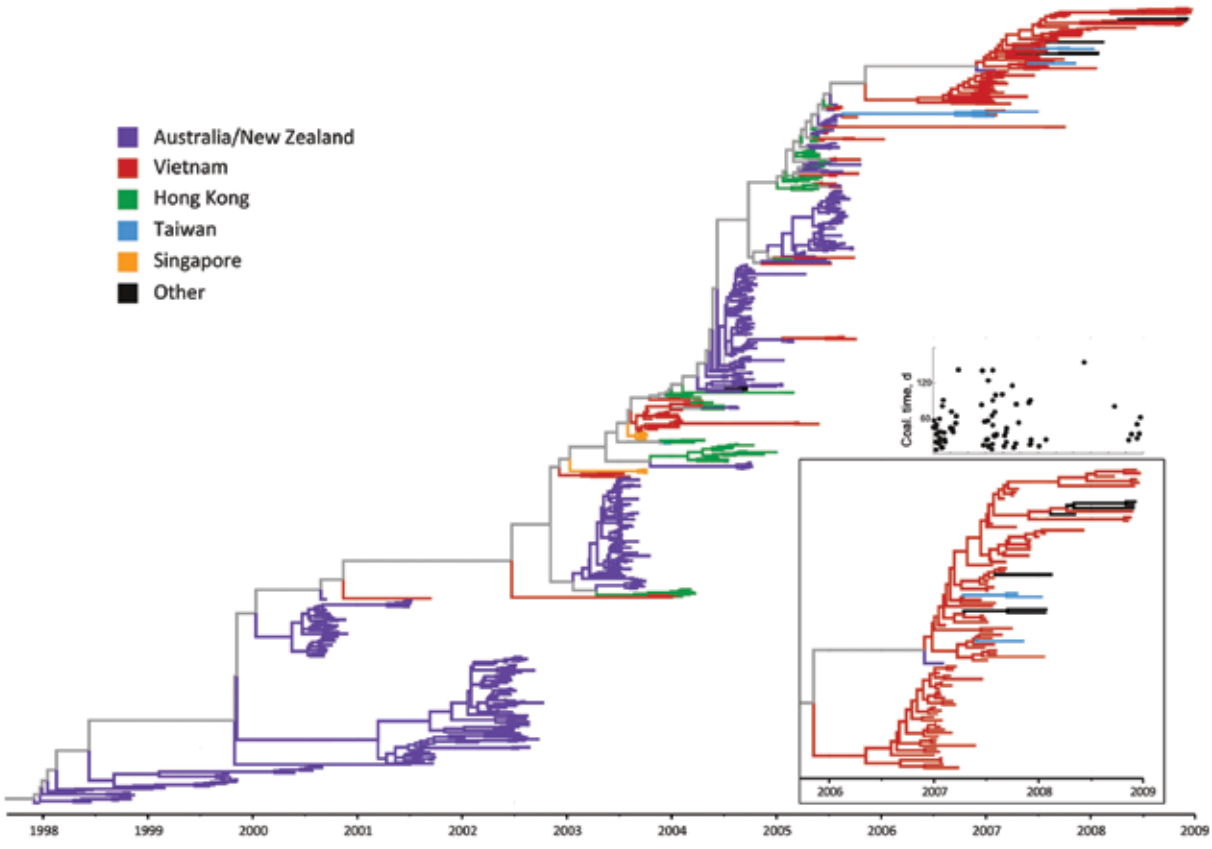


Figure 5. Maximum clade credibility tree for regional influenza (H3N2) hemagglutinin data, generated by BEAST version 1.6.2 (*26*) under a constant population model; these are the same sequences as shown on Figure 1, except 2 sequences from Vietnam were removed because of missing sampling dates (n = 785 sequences). Inset on the bottom right shows a magnification of the tree for the 2007–2008 Vietnam sequences, to highlight persistence during this time. The smaller inset above shows coalescent (coal.) times for the Vietnam sequences in the larger inset. The 2 *x*-axes on the insets coincide, and each black circle showing a coalescence time corresponds to the tip of a branch of Vietnam virus in the magnified-tree inset.

from each region. In the situation when too few sequences are available from a particular location, a smaller number of migratory links will be able to be inferred for that location. This second bias cannot be corrected with a subsampling strategy.

Our analysis of the global subsampled dataset showed that sample counts and strength of migratory connections were highly correlated. It has so far been impossible to determine the causal direction in this correlation. A migration signal can be weak because of a dearth of samples. Conversely, the small number of samples can be the result of low influenza activity and a corresponding weak migratory connection with other regions. The directionality of causation cannot be determined from sequence data alone. A sequence sampling strategy must be devised in the context of an influenza surveillance system, and the epidemiologic data and sequence data must be analyzed jointly. Disease prevalence and sequence data should be directly linked to provide a denominator to help determine whether undersampling or oversampling are truly occurring, which would allow for correction of sampling numbers across regions.

Despite this seemingly obvious point about oversampling, the counterpoint is that oversampling in influenza sequence data occurs with a high degree of pseudoreplication. Influenza sequence sampling in most scientific studies and public health contexts is conducted in such a way that each additional sequence sample is not an independent observation but, rather, is an observation with a high degree of correlation to recently collected samples (*33*). These pseudoreplicated samples should not, in principle, generate additional artificial migration events into the analysis because the dependency structure of the samples is entirely accounted for in the phylogeny. Nevertheless, a correlation between sample number and migration strength persists in the data, partially, at least, because a larger number of samples increases the probability that a distant recently introduced lineage is sampled.

New approaches are needed in order to fully account for all spatial, evolutionary, and epidemiologic dependencies in phylogeographic analyses. For recent phylogeographic studies, Bayesian approaches have been the method of choice (*1,3,4,34–37*), primarily because of their ability to account for uncertainty in evolutionary, demographic, and migratory parameters, but especially because of their ability to incorporate topological uncertainty into phylogenetic analyses. If these methods can be further developed to incorporate representativeness uncertainty—essentially, a prior distribution on the size of the sampling pool to account for the fact that some parts of the phylogeny will be oversampled while others will be undersampled—then this type of Bayesian analysis could serve as a powerful auxiliary tool in phylogeography, enabling us to determine whether sampling bias has a larger effect in some regions than others. Another role for Bayesian analysis of influenza sequences will be the application of Bayesian phylogeographic methods on whole-genome sequence data (*1*). For highly reassortant datasets, the presence of independent migration signals in 8 phylogenies (for the influenza virus 8 RNA segments) should act to reduce uncertainty for the inferred migration parameters.

We intended to elucidate the migratory pathways of influenza into and out of Vietnam and the likelihood of virus persistence in Vietnam. For each of these objectives, we recommend that future studies link phylogenetic analysis with prevalence data, allowing for correction of known biases and providing crucial complementary epidemiologic evidence for migration and persistence. If the source–sink framework is an oversimplification of global influenza circulation (*3–5*), Vietnam probably plays both roles on different occasions, given its close connections to other countries in Asia, Europe, and the United States.

Acknowledgments

We thank WHO Vietnam, the US Centers for Disease Control and Prevention, H. Suzuki, and R. Saito for helping establish influenza sentinel surveillance in Vietnam. Thanks to E.C. Holmes for critical review of this manuscript.

H.M.L., P.H., M.F.B. are supported by the Wellcome Trust (089276/B/09/7, 098511/Z/12/Z, WT/093724). T.T.L. is supported by a Royal Society Newton International Fellowship (UK). Computing and hardware costs were funded by the Li Ka Shing, University of Oxford Global Health Programme (LG05). This project has been funded in part with federal funds from the Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, under contract no. HHSN272200900007C.

Dr Le is head of the Virology Department at the Vietnam National Institute for Hygiene and Epidemiology, Hanoi, Vietnam. Her research interests are virology, public health, and human and avian influenza viruses.

References

1. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008;453:615–19. http://dx.doi.org/10.1038/nature06945
2. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global circulation of seasonal influenza A (H3N2) viruses. Science. 2008;320:340–6. http://dx.doi.org/10.1126/science.1154137
3. Bedford T, Cobey S, Beerli P, Pascual M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). PLoS Pathog. 2010;6:e1000918. http://dx.doi.org/10.1371/journal.ppat.1000918
4. Bahl J, Nelson MI, Chan KH, Chen R, Vijaykrishna D, Halpin RA, et al. Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. Proc Natl Acad Sci U S A. 2011;108:19359–64. http://dx.doi.org/10.1073/pnas.1109314108
5. Cheng X, Tan Y, He M, Lam TT-Y, Lu X, Viboud C, et al. Epidemiological dynamics and phylogeography of influenza virus in southern China. J Infect Dis. 2013;207:106–14. http://dx.doi.org/10.1093/infdis/jis526
6. Nguyen HT, Dharan NJ, Le MTQ, Nguyen NB, Nguyen CT, Hoang DV, et al. National influenza surveillance in Vietnam 2006–2007. Vaccine. 2009;28:398–402. http://dx.doi.org/10.1016/j.vaccine.2009.09.139
7. Oxford University Clinical Research Unit. Influenza-like illness (ILI) in Ho Chi Minh City. 2013 [cited 2013 Mar 4]. http://www.ili.vn.
8. Horby P, Mai Le Q, Fox A, Thai PQ, Thi Thu Yen N, Thanh le T, et al. The epidemiology of interpandemic and pandemic influenza in Vietnam, 2007–2010: The Ha Nam Household Cohort Study I. Am J Epidemiol. 2012;175:1062–74. http://dx.doi.org/10.1093/aje/kws121
9. Boni MF, Manh BH, Thai PQ, Farrar J, Hien TT, Hien NT, et al. Modelling the progression of pandemic influenza A(H1N1) and the opportunities for reassortment with other influenza viruses. BMC Med. 2009;7:43. http://dx.doi.org/10.1186/1741-7015-7-43
10. Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD, et al. Social contact patterns in Vietnam and implications for the control of infectious diseases. PLoS ONE. 2011;6:e16965. http://dx.doi.org/10.1371/journal.pone.0016965
11. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. PLoS Pathog. 2007;3:1220-8. http://dx.doi.org/10.1371/journal.ppat.0030131
12. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, St. George K, et al. Stochastic processes are key determinants of short-term evolution in influenza A virus. PLoS Pathog. 2006;2:e125. http://dx.doi.org/10.1371/journal.ppat.0020125
13. Tang JW, Ngai KLK, Lam WY, Chan PKS. Seasonality of influenza A (H3N2) virus: a Hong Kong perspective (1997–2006). PLoS ONE. 2008;3:e2768. http://dx.doi.org/10.1371/journal.pone.0002768
14. Doraisingham S, Goh KT, Ling AE, Yu M. Influenza surveillance in Singapore: 1972–86. Bull World Health Organ. 1988;66:57–63.
15. Simmerman JM, Uyeki TM. The burden of influenza in East and South-East Asia: a review of the English language literature. Influenza Other Respi Viruses. 2008;2:81–92. http://dx.doi.org/10.1111/j.1750-2659.2008.00045.x
16. Wong CM, Yang L, Chan KP, Leung GM, Chan KH, Guan Y, et al. Influenza-associated hospitalization in a subtropical city. PLoS Med. 2006;3:e121. http://dx.doi.org/10.1371/journal.pmed.0030121
17. Yap FHY, Ho P-L, Lam K-F, Chan PKS, Cheng Y-H, Peiris JSM. Excess hospital admissions for pneumonia, chronic obstructive pulmonary disease, and heart failure during influenza seasons in Hong Kong. J Med Virol. 2004;73:617–23. http://dx.doi.org/10.1002/jmv.20135

18. Shek LPC, Lee BW. Epidemiology and seasonality of respiratory tract virus infections in the tropics. Paediatr Respir Rev. 2003;4:105–11. http://dx.doi.org/10.1016/S1526-0542(03)00024-1
19. Finkelman BS, Viboud C, Koelle K, Ferrari MJ, Bharti N, Grenfell BT. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. PLoS ONE. 2007;2:e1296. http://dx.doi.org/10.1371/journal.pone.0001296
20. Viboud C, Alonso WJ, Simonsen L. Influenza in tropical regions. PLoS Med. 2006;3:e89. http://dx.doi.org/10.1371/journal.pmed.0030089
21. Alonso WJ, Viboud C, Simonsen L, Hirano EW, Daufenbach LZ, Miller MA. Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. Am J Epidemiol. 2007;165:1434–42. http://dx.doi.org/10.1093/aje/kwm012
22. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature. 2005;437:1162–6. http://dx.doi.org/10.1038/nature04239
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. http://dx.doi.org/10.1093/nar/gkh340
24. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 2008;57:758–71. http://dx.doi.org/10.1080/10635150802429642
25. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90. http://dx.doi.org/10.1093/bioinformatics/btl446
26. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. http://dx.doi.org/10.1186/1471-2148-7-214
27. Swofford DL, Maddison WP. Reconstructing ancestral character states under Wagner parsimony. Math Biosci. 1987;87:199–229. http://dx.doi.org/10.1016/0025-5564(87)90074-5
28. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates; 2002.
29. Boni MF, de Jong MD, van Doorn HR, Holmes EC. Guidelines for identifying homologous recombination events in influenza A virus. PLoS ONE. 2010;5:e10434. http://dx.doi.org/10.1371/journal.pone.0010434
30. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Third International AAAI Conference on Weblogs and Social Media; 2009 May 17– 20; San Jose, California [cited 2013 Mar 4]. http://www.aaai.org/Library/ICWSM/icwsm09contents.php
31. Boni MF, Gog JR, Andreasen V, Feldman MW. Epidemic dynamics and antigenic evolution in a single season of influenza A. Proc Biol Sci. 2006;273:1307–16. http://dx.doi.org/10.1098/rspb.2006.3466
32. Adams B, McHardy AC. The impact of seasonal and year-round transmission regimes on the evolution of influenza A virus. Proc Biol Sci. 2011;278:2249–56. http://dx.doi.org/10.1098/rspb.2010.2191
33. Harvey PH, Pagel MD. The comparative method in evolutionary biology. Oxford (UK): Oxford University Press; 1991.
34. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A. 2012;109:15066–71. http://dx.doi.org/10.1073/pnas.1206598109
35. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLOS Comput Biol. 2009;5:e1000520. http://dx.doi.org/10.1371/journal.pcbi.1000520
36. Rabaa MA, Hang VTT, Wills B, Farrar J, Simmons CP, Holmes EC. Phylogeography of recently emerged DENV-2 in southern Viet Nam. PLoS Negl Trop Dis. 2010;4:e766. http://dx.doi.org/10.1371/journal.pntd.0000766
37. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, Farrar J, et al. Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. PLoS Pathog. 2011;7:e1002064. http://dx.doi.org/10.1371/journal.ppat.1002064

Address for correspondence: Maciej F. Boni, Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 764 Vo Van Kiet St, Q5, Ho Chi Minh City, Vietnam; email: mboni@oucru.org

Article DOI: http://dx.doi.org/10.3201/eid1911.130349

Migration and Persistence of Human Influenza A viruses, Vietnam, 2001–2008

Technical Appendix

1. Supplementary methods

Primary Data Sets. For phylogeographic analysis, influenza virus sequences of both H1N1 and H3N2 subtypes were compiled into two datasets: a "regional data set" of whole-genome sequences from Asia and Australia/New Zealand (ANZ), and a "global data set" of geographically sub-sampled sequences (50 replicates) of the hemagglutinin (HA) segment. For the regional dataset, we downloaded from GenBank all influenza virus sequence data that met the following criteria: (a) the virus was isolated in one of Australia, New Zealand, or an Asian country, (b) all eight segments were sequenced, (c) the date of isolation was available, and (d) the virus was sampled between 1999 and 2008. Because Australian and New Zealand strains were very closely related to each other, with ancestry difficult to determine in many cases, we grouped these two countries into a single region. This regional dataset yielded 787 whole-genome sequences of subtype H3N2 and 300 whole-genome sequences of subtype H1N1 (including the new Vietnamese sequences). The regional dataset was assembled to have a dataset of manageable size on which we could perform phylogeographic analysis for the geographic region surrounding Vietnam. Australia and New Zealand were included in this analysis because of the known high frequency of travel between ANZ and Southeast Asia.

For the global HA datasets, we downloaded HA sequences from all countries, including only time-stamped sequences sampled between 2001 and 2008, with a minimum sequence length of 900nt. This yielded 3,934 H3N2 sequences and 1,970 H1N1 sequences. We generated 50 geographic subsamples of this dataset allowing for a maximum of 12 sequences per country per year. Viruses from the United States were subsampled with a maximum of three viruses per year from each of four geographic regions: Northeast, Southeast, Midwest, and West. Subsampled datasets comprised 1,140 H3N2 sequences and 554 H1N1 sequences.

Two additional datasets were assembled: one for validation of the regional dataset as representative of migration events for Vietnam (section 2 below), and one comprising all Asian and ANZ sequences from 2006 to 2008 to validate analyses on local persistence of influenza during this time (section 4.2 below).

Additional Data Sets. Two additional datasets were assembled: one for validation of the regional dataset as representative of migration events for Vietnam (section 2 below), and one comprising all Asian and ANZ sequences from 2006 to 2008 to validate analyses on local persistence of influenza during this time (section 4.2 below).

PAUP Migration Analysis. Migration events using the ACCTRAN parsimony criterion in PAUP* are reported as `unambiguous' or 'ambiguous'. Only unambiguous events where Vietnam was either the origin or the destination country were counted in the regional analysis, and all unambiguous events were counted in the global analysis to create a complete migration matrix among our pre-defined regions. In addition, PAUP* reports directionality for migration events, but we ignored this because it was not consistently repeatable across our analyses; all migration events described in this analysis should be interpreted as migratory connections, without certainty as to the direction of migration.

2. Validation of the regional dataset as a useful tool for inspecting migration events relating to Vietnam

As we were initially uncertain about the validity of using a 'regional tree' to infer migratory patterns for Vietnam, we constructed a partially sub-sampled global tree (PSGT), different from the 'global tree' mentioned in the main text, to determine what fraction of Vietnam-related global migration events were observable in the regional tree. The dataset for the PSGT was not restricted to Asia and ANZ and it was not restricted to include only whole genomes; it is partially subsampled because the Vietnam sequences are not subsampled but all other sequences are subsampled. The PSGT thus contains the exact same Vietnamese sequences as the regional tree and subsampled global sequences, from the same original sample set as in the `global tree' in the main paper. Hence, we can determine how many more migration events are associated with these VN sequences when we construct a global rather than regional phylogeny.

A total of 28 Vietnam-related parsimony-unambiguous migration events were inferred from the PSGT: five with China, four each with the United States and Hong Kong, three with Taiwan, and two each with Cambodia and South Korea. Single migration events were identified with Thailand, Myanmar, Japan, Iran, Qatar, France, Honduras, and Chile. Nineteen of these 28 migration events occurred between Vietnam and another country in East or Southeast Asia, indicating that roughly 70% of migration events for Vietnam can be observed by building a regional phylogenetic tree. Some of the long-range events present in the global tree (Honduras, Chile) must be exceptions as it is doubtful that there is a common or frequent migration route of influenza viruses between Vietnam and either of these countries.

Of the 20 regionally-inferred migration events (see main text), six could not be evaluated in the global tree, as the relevant sequences were not sub-sampled. Of the 14 remaining regionally-inferred migration events, all were supported in the global tree either directly, through an intermediary sequence, or through a missing ancestor in China; two VN-HK migrations and one VN-TW migration showed common ancestry in China with emigration (when sampling dates were taken into account) from China to both locations, indicating that the restricted sampling for the regional tree possibly missed the true migratory pattern in these situations. It is clear that some migration events will be missed by restricting sampling to a particular part of the world and to whole-genome sequences. Hence, in general, an individual migration event from a geographically regional tree should be viewed as a valid single observation — with probable, but not definite, phylogenetic support for the same migration event in a more complete global analysis — but the sum of these events does not represent a systematic migration analysis of influenza viruses among different countries. Rather, this type of regional migration analysis must be viewed as an investigation of influenza viruses in Vietnam and potential migration routes from/to New Zealand, Australia, Hong Kong, Singapore, and Taiwan — the only other countries with sufficient sampling in this dataset.

The migration events identified in the PSGT are listed in Table S1 below. Number of sequences and number of migration events were significantly correlated by Spearman ($p = 0.011$) and a Kendall ($p = 0.015$) tests.

Technical Appendix Table 1. Numbers of migration events between Vietnam and other countries/regions in the partially-subsampled global H3N2 tree.

	Number of Parsimony-Unambiguous Migration Events	Number of Sequences
China	5	84
USA	4	96
Hong Kong	4	72
Taiwan	3	53
Cambodia	2	26
South Korea	2	79
Thailand	1	51
Myanmar	1	17
Japan	1	89
Kyrgyzstan	1	5
Qatar	1	8
France	1	24
Central America	1	24
South America	1	48

3. Reassortment

Strong statistical evidence for reassortment was present in the datasets where whole genomes were available, i.e., the regional H3N2 and H1N1 datasets. The concatenated whole-genome datasets showed strong signals of mosaicism ($p < 10^{-11}$ for both, 3SEQ [1]), as did concatenated HA-NA datasets ($p < 10^{-9}$ for both, 3SEQ). Considering HA and NA for H3N2, a strong phylogenetic signal of reassortment was present (Technical Appendix Figure 2). As it is already known that human influenza viruses reassort frequently [2,3] this analysis was not pursued further.

4. Robustness of Persistence Result for Vietnamese sequences, 2007–2008

4.1 Bayesian skyride demographic model for H3N2 regional sequences

To determine if the viral persistence result described in the main text is robust to changes in viral effective population size, we performed phylogenetic inference with a Bayesian skyride model allowing for variable population size during the years 2001 to 2008. The main purpose of this analysis was to observe the presence/absence of population bottlenecks during 2007 and 2008 when viral lineages appeared to be persisting in Vietnam.

The MCMC was run for 150 million iterations, with the first 18 million designated as burn-in. Bayesian skyride was set as the demographic model, and other settings remained the same as for the constant population model. The effective sample size (ESS) values were 189 for the posterior, 34 for the prior, 93 for the likelihood, and 35 for the root height, which was the

best result that we observed after several runs and given the available computational time. The Maximum Clade Credibility (MCC) tree is shown as Technical Appendix Figure 3, with the bottom panel showing the Bayesian skyride plot with 95% posterior credible intervals. The natural-log Bayes factor for the constant population model over the skyride model was 23.7, but this unexpected result may be due to the incomplete convergence of the chain for the skyride analysis. The Bayesian skyride plot shows a drop in relative genetic diversity in the early part of 2008. This may be due to a real population bottleneck or an exogenous/endogenous strain replacement during this time.

In other analyses that we attempted for this dataset, we noted that inferred BSRs were not robust to various types of sub-analyses such as (*a*) sub-sampled regions, (*b*) individual datasets for separate regions, and (*c*) segments other than the hemagglutinin. For this reason, we were cautious in interpreting the BSRs as true reconstructions of past effective population size or relative genetic diversity, and we decided to present results in the main text based on the assumption of constant population size or constant genetic diversity.

4.2 Global sampling of sequences from 2006 through 2008

To determine if the viral persistence result described in the main text is robust to sampling frequencies of Asian and ANZ viruses during this time, we assembled a dataset with all Asian and ANZ hemagglutinin sequences (GenBank only) sampled between 2006 and 2008. In contrast to the migration analyses, it was not obvious how to devise a sub-sampling strategy for this type of analysis, where the goal is to find/reject evidence of lineage persistence in a focal region based on sampling in other regions. Hence, all sequences were included in the analysis. The final dataset included 672 H3N2 sequences, 88 of which were from Vietnam.

Two Bayesian phylogenetic trees were inferred with BEAST. The first was run under a constant population model, with a chain length of 100 million iterations and 10% burn-in (Technical Appendix Figure 4). The second was run under a Bayesian Skyride model, with a chain length of 100 million iterations and 45% burn-in.

Within the available computational time, the chain for the constant population model achieved near convergence in all parameters of interest (i.e., ESS >100; the ESS for the posterior was 183), while the BSR model did not converge. Therefore, we based our results on the maximum clade credibility tree of the constant population size model (Technical Appendix

Figure 4). This phylogeny shows that the 2007–2008 lineages from Figure 5 and Technical Appendix Figure 3 form more than a dozen sub-lineages when viewed in the regional context of all Asian and ANZ sequences. This was consistent in the incompletely converged BSR phylogeny. Two magnifications of the tree in Technical Appendix Figure 4 are shown in Technical Appendix Figure 1, suggesting lineage persistence in Vietnam for 15 months (panel A) and 11 months (panel B). These conclusions are sensitive to the absence of samples during certain periods, and to phylogenetic uncertainty that could affect the topology of these two subclades. Vietnam sentinel influenza data indicate that most weeks during this time period had positive influenza A/H3N2 cases; however, the last 2 months of 2007 saw an influenza B epidemic at which time there may have been no or low circulation of H3N2 viruses [4].

References

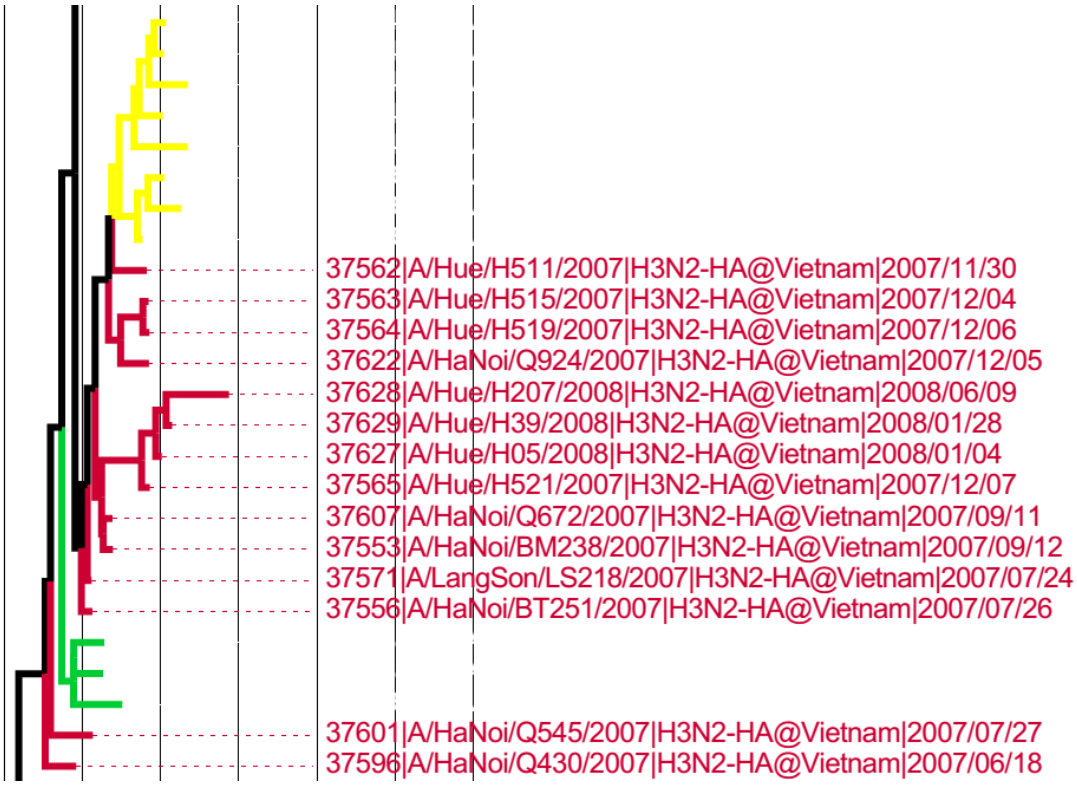
1. Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176: 1035–47.

2. Holmes EC, Ghedin E, Miller N, Taylor J, Bao Y, et al. (2005) Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol 3: e300.

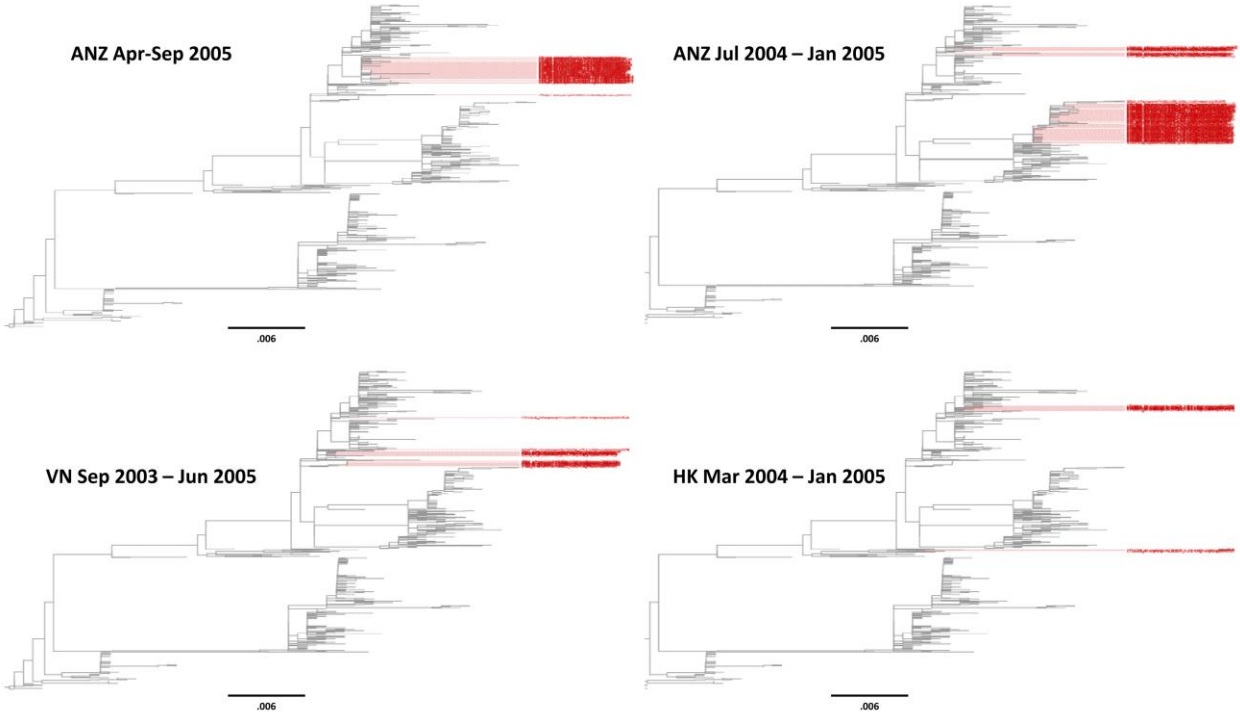
3. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, et al. (2006) Stochastic processes are key determinants of short-term evolution in influenza A virus. PLoS Pathog 2: e125.

4. Nguyen HT, Dharan NJ, Le MTQ, Nguyen NB, Nguyen CT, et al. (2009) National influenza surveillance in Vietnam, 2006–2007. Vaccine 28(2):398–402.

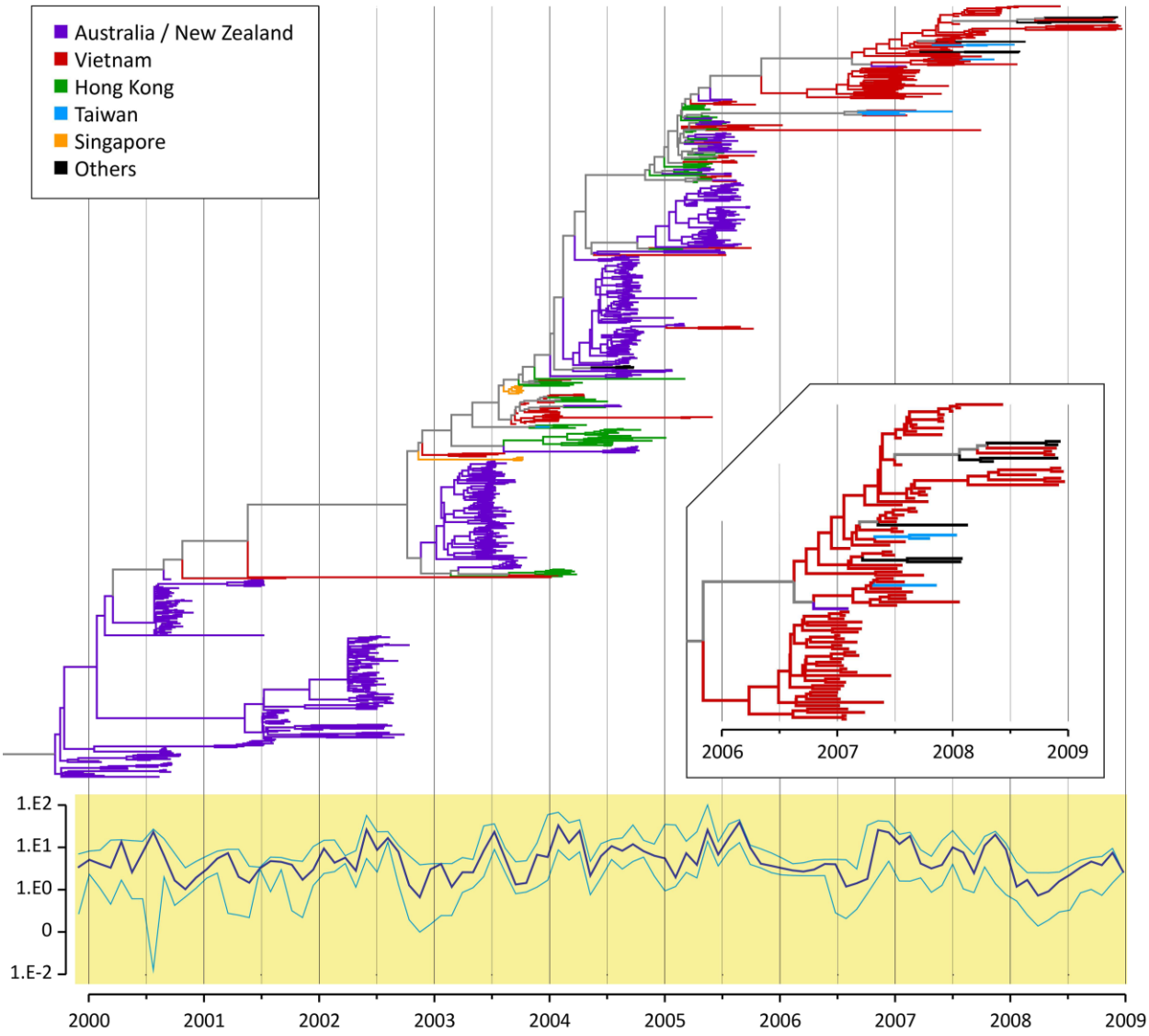
Technical Appendix Figure 1, panel A. Distance between adjacent vertical lines is 6 months. 15.5 months of persistence is seen in Vietnam, assuming intermediate sequences were also circulating in Vietnam. The three sequences at the pink tips are from Cambodia.



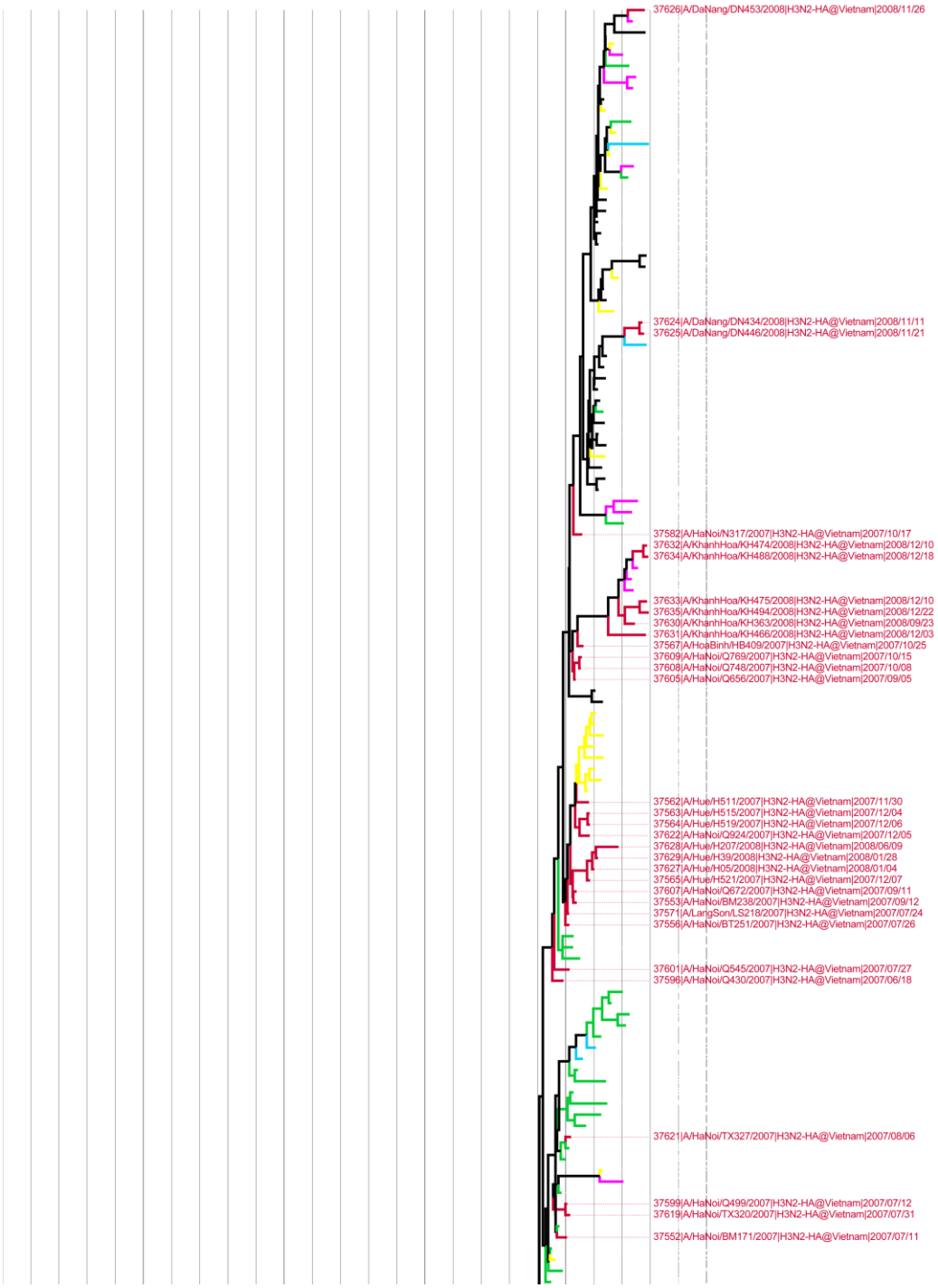
Technical Appendix Figure 1, panel B. Distance between adjacent vertical lines is 6 months. 10.5 to 11.5 months of persistence, depending on far down the phylogeny one looks. The three green tips represent sequences from China, two from Shanghai and one from Jiangxi province. The yellow clade at the top represents Japanese influenza viruses.



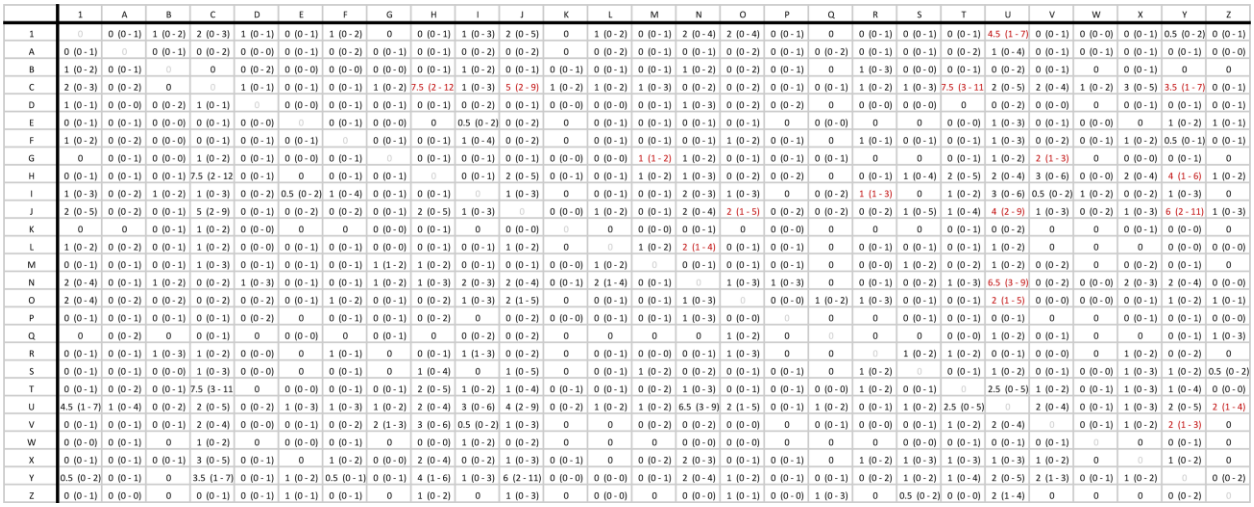
Technical Appendix Figure 2. Best ML trees (using RAxML) for the neuraminidase (NA) segment of the regional H3N2 dataset. From the HA tree of the regional dataset (Figure 1), four HA clades with strong reassortment signals were chosen and the sequences from these HA clades are shown in red here in the NA trees. The four clade labels in this figure correspond to the labels in Figure 1 of the main text. The NA sequences of the four chosen HA clades do not form monophyletic groups, indicating that the HA segments and NA segments of these viruses reassorted in the past.



Technical Appendix Figure 3. Maximum clade credibility tree for regional H3N2 HA data, generated by BEAST under a Bayesian skyride population model; these are the same sequences as Figure 1 and Figure 5 of the main text. Inset on the right shows a magnification of the tree for the 2007–2008 Vietnam sequences, to show the persistence and lineage co-circulation patterns during this time. The panel at the bottom shows the effective population size/relative genetic diversity, i.e., the Bayesian skyride plot for the regional sequences from 2000 through 2008.

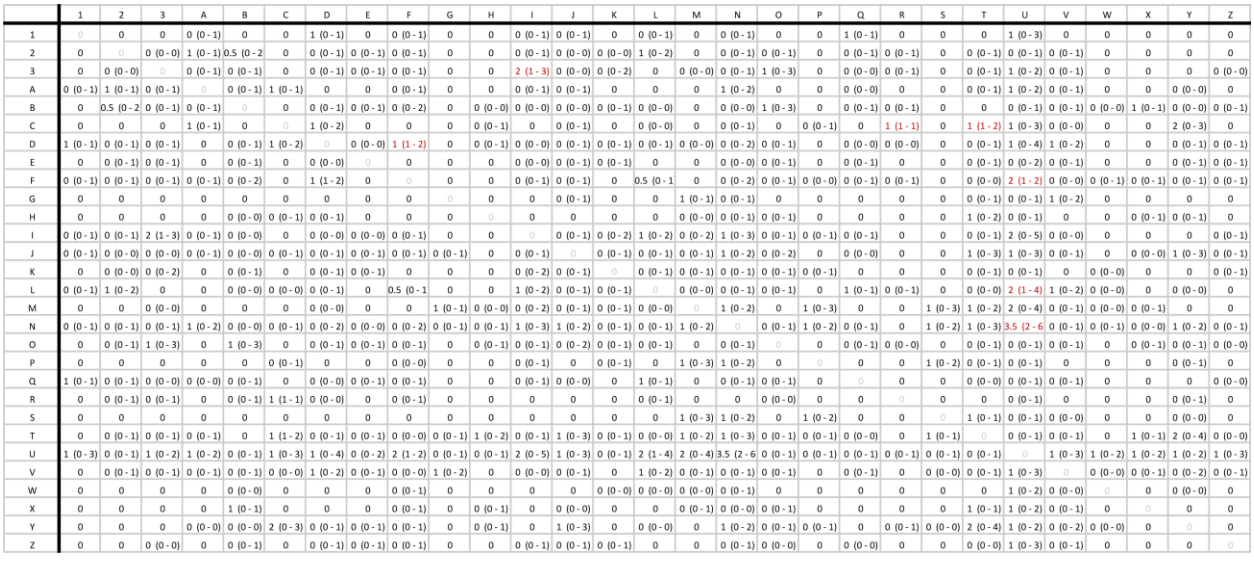


Technical Appendix Figure 4. Detail of maximum clade credibility tree for all Asian and ANZ H3N2 sequences sampled between 2006 and 2008, generated by BEAST under a constant population model. The tips of the tree are colored according to sampling location: Vietnam (red); Japan and Korea (yellow); China (green); Hong Kong (blue); Taiwan (turquoise); Singapore, Malaysia, and the Philippines (orange); Thailand, Cambodia, and Myanmar (pink); Australia, New Zealand, Solomon Islands, and New Caledonia (purple); all other countries shown in black. To view full tree, click on image.



	1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	0	0 (0 - 1)	1 (0 - 2)	2 (0 - 3)	1 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0	0 (0 - 1)	1 (0 - 3)	2 (0 - 5)	0	1 (0 - 2)	0 (0 - 1)	2 (0 - 4)	2 (0 - 4)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	4.5 (1 - 7)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0.5 (0 - 2)	0 (0 - 1)
A	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0 (0 - 2)	0	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	1 (0 - 4)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)
B	1 (0 - 2)	0 (0 - 1)	0	0	0 (0 - 2)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0 (0 - 2)	0 (0 - 1)	0	1 (0 - 3)	0 (0 - 0)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0	0 (0 - 1)	0	0
C	2 (0 - 3)	0 (0 - 2)	0	0	1 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	7.5 (2 - 12)	1 (0 - 3)	5 (2 - 9)	1 (0 - 2)	1 (0 - 2)	1 (0 - 3)	0 (0 - 2)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	1 (0 - 3)	7.5 (3 - 11)	2 (0 - 5)	2 (0 - 4)	1 (0 - 2)	3 (0 - 5)	3.5 (1 - 7)	0 (0 - 1)
D	1 (0 - 1)	0 (0 - 0)	0 (0 - 2)	1 (0 - 1)	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	1 (0 - 3)	0 (0 - 2)	0 (0 - 2)	0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 2)	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)
E	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 0)	0	0.5 (0 - 2)	0 (0 - 2)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	0	0 (0 - 0)	0	1 (0 - 3)	0 (0 - 1)	0 (0 - 0)	0	1 (0 - 2)	1 (0 - 1)
F	1 (0 - 2)	0 (0 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	1 (0 - 4)	0 (0 - 2)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0	1 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	0 (0 - 2)	0 (0 - 1)	1 (0 - 2)	0.5 (0 - 1)	0 (0 - 1)
G	0	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 0)	1 (1 - 2)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	1 (0 - 2)	2 (1 - 3)	0	0 (0 - 0)	0 (0 - 1)	0
H	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	7.5 (2 - 12)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	2 (0 - 5)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	1 (0 - 3)	0 (0 - 2)	0 (0 - 2)	0	0 (0 - 1)	1 (0 - 4)	2 (0 - 5)	2 (0 - 4)	3 (0 - 6)	0 (0 - 0)	2 (0 - 4)	4 (1 - 6)	1 (0 - 2)
I	1 (0 - 3)	0 (0 - 2)	1 (0 - 2)	1 (0 - 3)	0 (0 - 2)	0.5 (0 - 2)	1 (0 - 4)	0 (0 - 1)	0 (0 - 1)	0	1 (0 - 3)	0	0 (0 - 1)	0 (0 - 1)	2 (0 - 3)	1 (0 - 3)	0	0 (0 - 2)	1 (1 - 3)	0	1 (0 - 2)	3 (0 - 6)	0.5 (0 - 2)	1 (0 - 2)	0 (0 - 2)	1 (0 - 3)	0
J	2 (0 - 5)	0 (0 - 2)	0 (0 - 1)	5 (2 - 9)	0 (0 - 1)	0 (0 - 2)	0 (0 - 2)	0 (0 - 1)	2 (0 - 5)	1 (0 - 3)	0	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)	2 (0 - 4)	2 (1 - 5)	0 (0 - 2)	0 (0 - 2)	0 (0 - 2)	1 (0 - 5)	1 (0 - 4)	4 (2 - 9)	1 (0 - 3)	0 (0 - 2)	1 (0 - 3)	6 (2 - 11)	1 (0 - 3)
K	0	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 0)	0	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 0)	0	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 0)	0	0	0	0 (0 - 1)	0 (0 - 2)	0	0	0 (0 - 1)	0 (0 - 0)	0
L	1 (0 - 2)	0 (0 - 2)	0 (0 - 1)	1 (0 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0	0	1 (0 - 2)	2 (1 - 4)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0	0	0	0 (0 - 0)	0 (0 - 0)
M	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (1 - 2)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	1 (0 - 2)	0 (0 - 2)	1 (0 - 2)	0 (0 - 2)	0	0 (0 - 2)	0 (0 - 1)	0
N	2 (0 - 4)	0 (0 - 1)	1 (0 - 2)	0 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	1 (0 - 3)	2 (0 - 3)	2 (0 - 4)	0 (0 - 1)	2 (1 - 4)	0 (0 - 1)	0	1 (0 - 3)	1 (0 - 3)	0	0 (0 - 1)	0 (0 - 2)	1 (0 - 3)	6.5 (3 - 9)	0 (0 - 2)	0 (0 - 0)	2 (0 - 3)	2 (0 - 4)	0 (0 - 0)
O	2 (0 - 4)	0 (0 - 2)	0 (0 - 2)	0 (0 - 2)	0 (0 - 2)	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0 (0 - 2)	1 (0 - 3)	2 (1 - 5)	0	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	0	0 (0 - 0)	1 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	2 (1 - 5)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	1 (0 - 2)	1 (0 - 1)
P	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0	0 (0 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	0 (0 - 0)	0	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)
Q	0	0 (0 - 2)	0	0 (0 - 1)	0	0 (0 - 0)	0	0 (0 - 1)	0	0 (0 - 2)	0 (0 - 2)	0	0	0	0	1 (0 - 2)	0	0	0	0	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)	0	0	0 (0 - 1)	1 (0 - 3)
R	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	1 (0 - 2)	0 (0 - 0)	0	1 (0 - 1)	0	0 (0 - 1)	1 (1 - 3)	0 (0 - 2)	0	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	1 (0 - 3)	0	0	1 (0 - 2)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0	0 (0 - 2)	0
S	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 3)	0 (0 - 0)	0	0 (0 - 1)	0	1 (0 - 4)	0	1 (0 - 5)	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	1 (0 - 2)	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0 (0 - 0)	1 (0 - 3)	1 (0 - 2)	0.5 (0 - 2)
T	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	7.5 (3 - 11)	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	2 (0 - 5)	1 (0 - 2)	1 (0 - 4)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)	0	2.5 (0 - 5)	1 (0 - 2)	0 (0 - 1)	1 (0 - 3)	1 (0 - 4)	0 (0 - 0)
U	4.5 (1 - 7)	1 (0 - 4)	0 (0 - 2)	2 (0 - 5)	0 (0 - 2)	1 (0 - 3)	1 (0 - 3)	1 (0 - 2)	2 (0 - 4)	3 (0 - 6)	4 (2 - 9)	0 (0 - 2)	1 (0 - 2)	1 (0 - 2)	6.5 (3 - 9)	2 (1 - 5)	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	1 (0 - 2)	2.5 (0 - 5)	0	2 (0 - 4)	0 (0 - 1)	1 (0 - 3)	2 (0 - 5)	2 (1 - 4)
V	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	2 (0 - 4)	0 (0 - 0)	0 (0 - 1)	0 (0 - 2)	2 (1 - 3)	3 (0 - 6)	0.5 (0 - 2)	1 (0 - 3)	0	0	0 (0 - 2)	0 (0 - 2)	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	1 (0 - 2)	2 (0 - 4)	0	0 (0 - 1)	1 (0 - 2)	2 (1 - 3)	0
W	0 (0 - 0)	0 (0 - 1)	0	1 (0 - 2)	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 0)	1 (0 - 2)	0 (0 - 2)	0	0	0	0 (0 - 0)	0 (0 - 0)	0	0	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0
X	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	3 (0 - 5)	0 (0 - 1)	0	1 (0 - 2)	0 (0 - 0)	2 (0 - 4)	0 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0	0 (0 - 2)	2 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0	1 (0 - 2)	1 (0 - 3)	1 (0 - 3)	1 (0 - 3)	1 (0 - 2)	0	0	1 (0 - 2)	0
Y	0.5 (0 - 2)	0 (0 - 1)	0	3.5 (1 - 7)	0 (0 - 1)	1 (0 - 2)	0.5 (0 - 1)	0 (0 - 1)	4 (1 - 6)	1 (0 - 3)	6 (2 - 11)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	2 (0 - 4)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	1 (0 - 2)	1 (0 - 4)	2 (0 - 5)	2 (1 - 3)	0 (0 - 1)	1 (0 - 2)	0	0 (0 - 2)
Z	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 1)	1 (0 - 1)	0 (0 - 1)	0	1 (0 - 2)	0	1 (0 - 3)	0	0 (0 - 0)	0	0 (0 - 0)	1 (0 - 1)	0 (0 - 0)	1 (0 - 3)	0	0.5 (0 - 2)	0 (0 - 0)	2 (1 - 4)	0	0	0	0 (0 - 2)	0

Technical Appendix Figure 5. Median number of migratory connections between regions for global H3N2 tree (95% ranges in parentheses). Based on 50 subsamples. Matrix is symmetric. A simple zero indicates that there were zero migration events between these two regions for all 50 subsamples. Region pairs where 95% range is bounded away from zero are shown in red. Region codes: 1, South America; A, Africa; B, Bulgaria, Croatia, Romania, Hungary, Greece, Slovenia; C, China; D, Denmark, Iceland, Norway, Sweden, Finland; E, El Salvador, Guatemala, Honduras, Nicaragua; F, France, Spain, Netherlands; G, Cambodia; H, Hong Kong, Macau; I, India, Bangladesh, Nepal, Iran, Afghanistan, Kyrgyzstan; J, Japan; K, Canada; L, Ireland, United Kingdom; M, Malaysia; N, New Zealand, Australia, New Caledonia; O, Austria, Germany, Italy; P, Philippines; Q, Qatar, Kuwait, Iraq; R, Russia, Ukraine, Latvia; S, Singapore; T, Taiwan; U, USA; V, Vietnam; W, Myanmar; X, Thailand; Y, South Korea, North Korea; Z, Mexico.



	1	2	3	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	0	0	0	0 (0 - 1)	0	0	1 (0 - 1)	0	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 1)	0	0	1 (0 - 1)	0	0	0	1 (0 - 3)	0	0	0	0	0
2	0	0	0 (0 - 0)	1 (0 - 1)	0.5 (0 - 2	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 0)	0 (0 - 0)	1 (0 - 2)	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0	0
3	0	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	2 (1 - 3)	0 (0 - 0)	0 (0 - 2)	0	0 (0 - 0)	0 (0 - 1)	1 (0 - 3)	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0	0	0	0 (0 - 0)
A	0 (0 - 1)	1 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	1 (0 - 1)	0	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0	0	0	1 (0 - 2)	0	0	0 (0 - 0)	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0	0 (0 - 0)	0	0
B	0	0.5 (0 - 2	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 0)	1 (0 - 3)	0	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 1)	0 (0 - 0)	0 (0 - 1)
C	0	0	0	1 (0 - 1)	0	0	1 (0 - 2)	0	0	0	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 0)	0	0 (0 - 1)	0	0 (0 - 1)	0	1 (1 - 1)	0	1 (1 - 2)	1 (0 - 3)	0 (0 - 0)	0	0	2 (0 - 3)	0
D	1 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	1 (0 - 2)	0	0 (0 - 0)	1 (1 - 2)	0	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 2)	0 (0 - 1)	0	0 (0 - 0)	0 (0 - 0)	0	0 (0 - 1)	1 (0 - 4)	1 (0 - 2)	0	0	0 (0 - 1)	0 (0 - 1)
E	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 0)	0	0	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)
F	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0	1 (1 - 2)	0	0	0	0 (0 - 1)	0 (0 - 1)	0	0.5 (0 - 1	0	0 (0 - 2)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	2 (1 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)
G	0	0	0	0	0	0	0	0	0	0	0	0 (0 - 1)	0	0	1 (0 - 1)	0 (0 - 1)	0	0	0	0	0	0	0 (0 - 1)	1 (0 - 2)	0	0	0	0	0
H	0	0	0	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0	0	0	0	0	0	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0	0	0	0	1 (0 - 2)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0
I	0 (0 - 1)	0 (0 - 1)	2 (1 - 3)	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 2)	1 (0 - 2)	0 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	2 (0 - 5)	0 (0 - 0)	0	0	0	0 (0 - 1)
J	0 (0 - 1)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0 (0 - 2)	0	0 (0 - 0)	0	0	1 (0 - 3)	1 (0 - 3)	0 (0 - 1)	0	0 (0 - 0)	1 (0 - 3)	0 (0 - 1)
K	0	0 (0 - 0)	0 (0 - 2)	0	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0	0	0	0 (0 - 2)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	0	0 (0 - 1)
L	0 (0 - 1)	1 (0 - 2)	0	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0	0.5 (0 - 1	0	0	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 1)	0 (0 - 1)	0	0 (0 - 0)	2 (1 - 4)	1 (0 - 2)	0 (0 - 0)	0	0 (0 - 0)
M	0	0	0 (0 - 0)	0	0	0	0 (0 - 0)	0	0	1 (0 - 1)	0 (0 - 0)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0	1 (0 - 2)	0	1 (0 - 3)	0	0	1 (0 - 3)	1 (0 - 2)	2 (0 - 4)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0
N	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0 (0 - 0)	0 (0 - 1)	0 (0 - 2)	0 (0 - 0)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	1 (0 - 3)	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	1 (0 - 2)	0	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0	1 (0 - 2)	1 (0 - 3)	3.5 (2 - 6	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)
O	0	0 (0 - 1)	1 (0 - 3)	0	1 (0 - 3)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)
P	0	0	0	0	0	0 (0 - 1)	0	0	0 (0 - 0)	0	0	0 (0 - 1)	0	0 (0 - 1)	0	1 (0 - 3)	1 (0 - 2)	0	1 (0 - 2)	0	0	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	0	0	0 (0 - 1)	0
Q	1 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 0)	0	1 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0	0	0	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)	0	0	0	0 (0 - 0)
R	0	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	1 (1 - 1)	0 (0 - 0)	0	0 (0 - 1)	0	0	0	0	0	0 (0 - 1)	0	0	0 (0 - 0)	0	0	0	0	0 (0 - 1)	0	0	0	0 (0 - 1)	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (0 - 3)	1 (0 - 2)	0	1 (0 - 2)	0	0	1 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0	0	0 (0 - 0)	0
T	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	1 (1 - 2)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	1 (0 - 2)	0 (0 - 1)	1 (0 - 3)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	1 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0	1 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0	1 (0 - 1)	2 (0 - 4)	0 (0 - 0)
U	1 (0 - 3)	0 (0 - 1)	1 (0 - 2)	1 (0 - 2)	0 (0 - 1)	1 (0 - 3)	1 (0 - 4)	0 (0 - 2)	2 (1 - 2)	0 (0 - 1)	0 (0 - 1)	2 (0 - 5)	1 (0 - 3)	0 (0 - 1)	2 (1 - 4)	2 (0 - 4)	3.5 (2 - 6	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	1 (0 - 3)	1 (0 - 2)	1 (0 - 2)	1 (0 - 2)	1 (0 - 3)
V	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0 (0 - 1)	0 (0 - 0)	1 (0 - 2)	0	0 (0 - 0)	0 (0 - 1)	0	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 0)	0 (0 - 1)	0	0 (0 - 1)	1 (0 - 3)	0	0 (0 - 0)	0 (0 - 1)	0 (0 - 1)
W	0	0	0	0	0 (0 - 0)	0	0	0	0 (0 - 1)	0	0	0	0	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)	0 (0 - 1)	0	0	0	0	0	1 (0 - 2)	0 (0 - 0)	0	0	0 (0 - 0)	0
X	0	0	0	0	1 (0 - 1)	0	0	0	0 (0 - 1)	0	0 (0 - 1)	0	0 (0 - 0)	0	0	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)	0	0	0	1 (0 - 1)	1 (0 - 2)	0 (0 - 1)	0	0	0	0
Y	0	0	0	0 (0 - 0)	0 (0 - 0)	2 (0 - 3)	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0	1 (0 - 3)	0	0 (0 - 0)	0	1 (0 - 2)	0 (0 - 1)	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 0)	2 (0 - 4)	1 (0 - 2)	0 (0 - 2)	0 (0 - 0)	0	0	0
Z	0	0	0 (0 - 0)	0	0 (0 - 1)	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 1)	0 (0 - 1)	0	0	0 (0 - 1)	0 (0 - 0)	0	0 (0 - 0)	0	0	0 (0 - 0)	1 (0 - 3)	0 (0 - 1)	0	0	0	0

Technical Appendix Figure 6. Median number of migratory connections between regions for global H1N1 tree (95% ranges in parentheses). Based on 50 subsamples. Matrix is symmetric. A simple zero indicates

that there were zero migration events between these two regions for all 50 subsamples. Region pairs where 95% range is bounded away from zero are shown in red. Region codes: 1, South America; A, Africa; B, Bulgaria, Croatia, Romania, Hungary, Greece, Slovenia; C, China; D, Denmark, Iceland, Norway, Sweden, Finland; E, El Salvador, Guatemala, Honduras, Nicaragua; F, France, Spain, Netherlands; G, Cambodia; H, Hong Kong, Macau; I, India, Bangladesh, Nepal, Iran, Afghanistan, Kyrgyzstan; J, Japan; K, Canada; L, Ireland, United Kingdom; M, Malaysia; N, New Zealand, Australia, New Caledonia; O, Austria, Germany, Italy; P, Philippines; Q, Qatar, Kuwait, Iraq; R, Russia, Ukraine, Latvia; S, Singapore; T, Taiwan; U, USA; V, Vietnam; W, Myanmar; X, Thailand; Y, South Korea, North Korea; Z, Mexico.