Journal of the Formosan Medical Association (2019) 118, 116–124

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

Original Article

Inferring the global phylodynamics of influenza A/H3N2 viruses in Taiwan

Yu-Nong Gong [a], Kuo-Chien Tsao [a,b,c], Guang-Wu Chen [a,b,d,*]

[a] *Research Center for Emerging Viral Infections, College of Medicine, Chang Gung University, Taoyuan, Taiwan*
[b] *Department of Laboratory Medicine, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan*
[c] *Department of Medical Biotechnology and Laboratory Science, College of Medicine, Chang Gung University, Taoyuan, Taiwan*
[d] *Department of Computer Science and Information Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan*

Received 7 August 2017; received in revised form 24 January 2018; accepted 29 January 2018

KEYWORDS
Influenza A/H3N2 viruses;
Phylodynamics;
Phylogeography;
Summer epidemics;
Viral quasispecies

Background/Purpose: Influenza A/H3N2 viruses are characterized by highly mutated RNA genomes. In this study, we focused on tracing the phylodynamics of Taiwanese strains over the past four decades.
Methods: All Taiwanese H3N2 HA1 sequences and references were downloaded from public database. A Bayesian skyline plot (BSP) and phylogenetic tree were used to analyze the evolutionary history, and Bayesian phylogeographic analysis was applied to predict the spatiotemporal migrations of influenza outbreaks.
Results: Genetic diversity was found to have peaked near the summer of 2009 in BSP, in addition to the two earlier reported ones in summer of 2005 and 2007. We predicted their spatiotemporal migrations and found the summer epidemic of 2005 from Korea, and 2007 and 2009 from the Western United States. BSP also predicted an elevated genetic diversity in 2015–2017. Quasispecies were found over approximately 20% of the strains included in this time span. In addition, a first-time seen N31S mutation was noted in Taiwan in 2016–2017.
Conclusion: We comprehensively investigated the evolutionary history of Taiwanese strains in 1979–2017. An epidemic caution could thus be raised if genetic diversity was found to have peaked. An example showed a newly-discovered cluster in 2016–2017 strains featuring a mutation N31S together with HA-160 quasispecies. Phylogeographic analysis, moreover, provided useful insights in tracing the possible source and migrations of these epidemics around the world. We

* Corresponding author. Department of Computer Science and Information Engineering, College of Engineering, Chang Gung University, Taoyuan 33302, Taiwan. Fax: +886 3 2118700.
E-mail address: gwchen@mail.cgu.edu.tw (G.-W. Chen).

https://doi.org/10.1016/j.jfma.2018.01.019

demonstrated that Asian destinations including Taiwan were the immediate followers, while U.S. continent was predicted the origin of two summer epidemics in 2007 and 2009.

Introduction

Influenza A/H3N2 virus has a highly-mutated RNA genome that is caused by several molecular mechanisms, including mutation, reassortment, and quasispecies.[1–3] In particular, variants in the HA1 gene further result in phylogenetic clade changes. Since 2009, seven genetic groups of H3N2 viruses have been defined by phylogeny inference.[4] Clade 3C, including three subdivisions 3C1, 3C2, and 3C3, was dominant after 2011. More new subclades, including 3C2.a from subdivision 3C2, and 3C3.a and 3C3.b from 3C3, emerged in 2014. One recent study further discovered a novel subclade 3C2.a2 (including two clusters I and II) co-circulated with 3C2.a1 (including three clusters I, II, and III) in London, United Kingdom,[5] which possessed four unique amino acid substitutions in the HA gene, including I58V and S219Y in Cluster I, and N122D (an N-linked glycosylation site) and S262N in Cluster II. These novel genetic variants potentially changed the antigenicity of H3N2, leading to outbreaks.

Previous studies indicated that influenza epidemics in Taiwan, Hong Kong, Singapore, and Japan occurred almost all seasons.[6,7] More studies have suggested that summer influenza was one of the main factors causing acute respiratory infections in children and adults, particularly in subtropical areas.[8–10] More importantly, not only were two different genetic clades detected, but viral reassortment from Taiwanese strains also circulated in the summer of 2005 and 2007.[2] For years, Asia has been recognized as a major origin of new subtypes and strains and is sitting in a pivotal geographical location due to the ease of spreading viruses via public transit.[11,12] Recently, a research topic called viral phylodynamics was raised and defined as a study of how epidemiological, immunological, and evolutionary processes act and potentially interact to shape viral phylogenies.[13] Studies reported that Eastern, Southern, and Southeast Asia contributed to the evolution of seasonal H3N2, and that most of the evolved viruses had spread worldwide.[11,14,15] Although Southeast Asia and Hong Kong may not be the sustained sources for annual H3N2 influenza epidemics, a global persistence through time in tropical Asian regions may be influenced by viral input from temperate regions.[16]

This study aims at reconstructing the phylodynamics of the Taiwanese H3N2 strains overall, and performing Bayesian phylogeographic analysis for inferring globally spatiotemporal and epidemiological dynamics, with particular attention on summer influenza epidemics. Phylodynamic and phylogeographic analysis could assist not only in detecting novel genetic diversity, but also in tracing the origins of newly emerging viruses.

Methods

Nucleotide sequences of influenza A/H3N2 viruses

We downloaded all 883 H3N2 Taiwanese strains from the Global Initiative for Sharing All Influenza Data (GISAID) database as of June 2017 and removed sequences with redundant strain names and of insufficient year/month/day information. This resulted in 543 sequences with more than 95% of the 987-nt HA1 domain from 1979 to 2017, which were used for Bayesian phylodynamic analysis encompassing these four decades in Taiwan. Since 2009, seven genetic groups of H3N2 viruses have been defined. Therefore, 24 reference strains[5,17] with clade information after 2009 including two vaccine strains (A/Victoria/361/2011 as VI11, and A/Texas/50/2012 as TX12), and another 14 vaccine strains[18,19] after 1979 were additionally downloaded to facilitate annotating the transitions and circulations of phylogenetic clades for those 543 Taiwanese strains.

To investigate the spatiotemporal dynamics of H3N2 viruses within the specific time span of interest, we downloaded all the H3N2 sequences worldwide in 2015 (from January to December, n = 936), 2007 (n = 1057), and 2009 (n = 1425), respectively. Same data trimming was applied as it was in obtaining the 543 Taiwanese H3N2 sequences earlier. To ensure that the sequences participating in the subsequent phylogeographic analysis featured the same evolutionary characteristics, phylogenetic trees were computed to sample only sequences belonging to the clades where the Taiwanese summer strains emerged. All selected sequences were then grouped by country, except the Eastern and Western United States (USA), as well as the Northern, Southern, and Hong Kong SAR in China. Countries or regions with less than 12 sequences were further removed to reduce computational complexity, leaving 361, 589, and 574 strains each in 2005, 2007, and 2009, respectively for subsequent Bayesian phylogeographic analysis. The locations of the collected sequences are provided in Supplementary Table S1.

Phylogenetic analysis

Bayesian phylogeny analysis was performed using BEAST software (version 1.8.4)[20] along with the BEAGLE library,[21] assuming a strict molecular clock. A Markov Chain Monte Carlo was used to sample evolutionary trees with parameters of 50 million generations, but discarded the initial 10% for burn-in. TRACER (version 1.6) was used for analyzing log files from BEAST, and generating Bayesian skyline plot (BSP).[22] A maximum clade credibility (MCC) tree was summarized using the TreeAnnotator program within the BEAST

package. FigTree (version 1.4.2) was used to visualize and annotate trees, respectively. Furthermore, the spatiotemporal migrations were estimated using a discrete Bayesian phylogeographic method, also performed using BEAST with the same parameters as above-mentioned, e.g., generations and burn-in value. The discrete trait substitution model was applied in the symmetric substitution model with Bayesian stochastic search variable selection.[23] Google Earth (available online: https://www.google.com/earth/) was used to display viral migration patterns.

MEGA7[24] in combination with the neighbor-joining (NJ) method[25] was used to infer the evolutionary history for clade transitions and circulations. The percentage of the replicated trees (where the associated taxa clustered together) were used in the bootstrap test with 500 replicates.[26] An NJ tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary distances were computed in the units of the number of base substitutions per site using the Jukes-Cantor method.[27] All positions with less than 95% site coverage were eliminated. In other words, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position.

Results

Phylogenetic and evolutionary dynamics of Taiwanese H3N2 strains

We applied Bayesian phylogenetic analysis to reconstruct the evolutionary dynamics of Taiwanese H3N2 strains over time. All available H3N2 sequences isolated in Taiwan were downloaded, and trimmed to include 543 HA1 domain

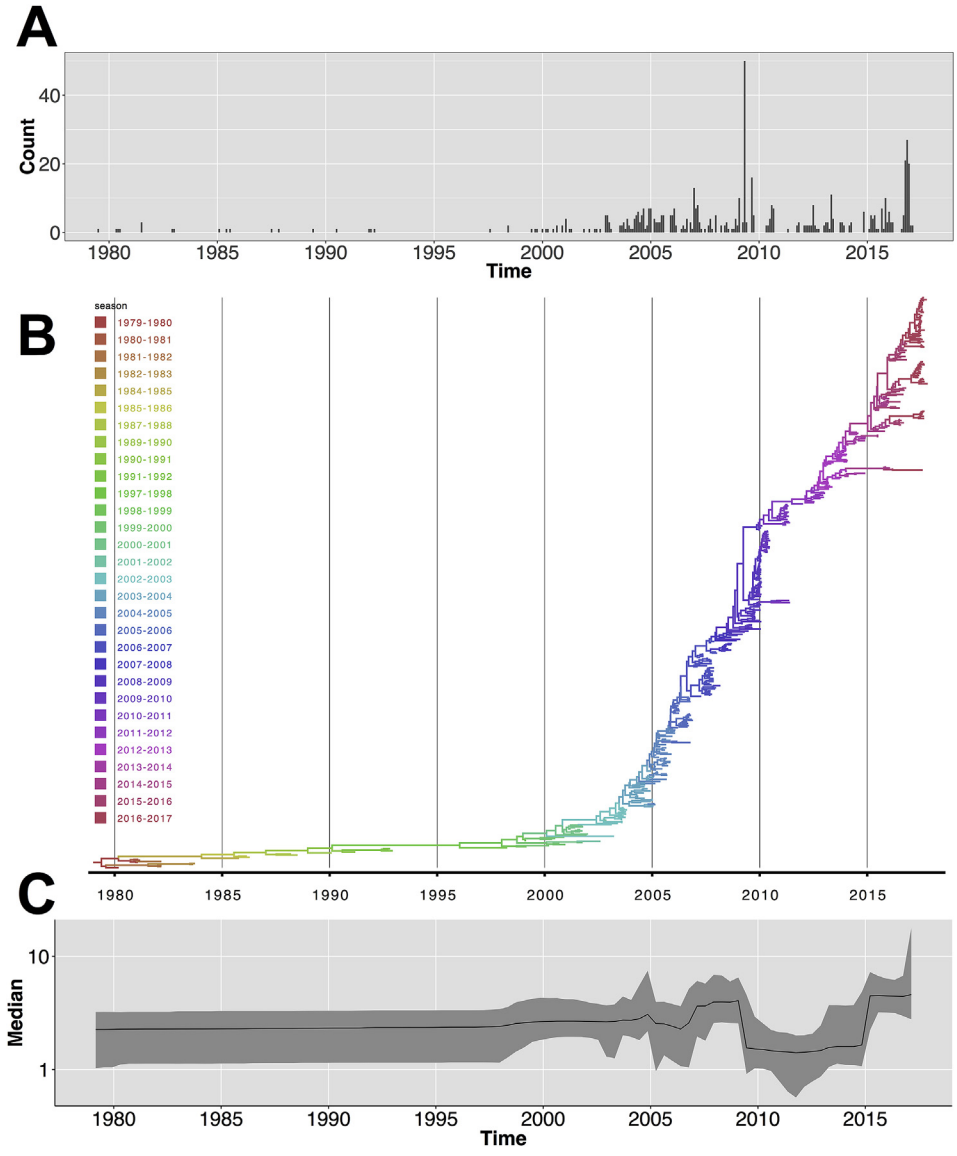


Figure 1 The evolution and phylogeny of Taiwanese H3N2 strains from 1979 to 2017: (A) monthly sequence count; (B) maximum clade credibility tree; (C) Bayesian skyline plot. In (A), a horizontal label 2010 marks the month of January in 2010. In (B), a group 2010—2011 includes strains from June 2010—May 2011. In (C), the vertical axis indicates the number of effective infections at given time t as $Ne(t)$. A solid line represents the median value with a grey area showing 95% HPD of the $Ne(t)$ estimates.

sequences for further analyses. Fig. 1A presents the monthly distribution of these H3N2 strains, which shows two distinct peaks in May 2009 (n = 50) and November 2016 (n = 27). An MCC tree of Taiwanese H3N2 strains was constructed in Fig. 1B. As stated earlier, Taiwan often experiences summer H3N2, thus blurring the beginning and end of an influenza season. For the convenience of including sequences from all months, each of the seasonal groups in this study was defined to begin from June to May in the following year. For example, strains isolated from June 1, 2005 to May 31, 2006 were annotated as 2005—2006 in Fig. 1B. The BSP in Fig. 1C represents the number of effective infections $Ne(t)$ over time to demonstrate how an effective population size contributed to new transmissions. The term effective infection is a metric based on analyzing temporal sequence variations, which is used to measure how transmissible viruses infect new host individuals. We observed that effective populations in BSPs peaked near the summer of 2009 and elevated from 2015 to 2017. The two plateaus seemed to match each of the branches of diversifying clades in Fig. 1B, respectively. The two $Ne(t)$ peaks were also found in parallel to the two large sequence counts of Fig. 1A.

Inferring global spatiotemporal dynamics by Bayesian phylogeography

As described in a previous study[2] as well as in this work, the Taiwanese strains in the summer of 2005, 2007, and 2009 all exhibited relatively higher genetic diversity. To elucidate the possible geographical origins of the viruses from each of the three summers, we downloaded all available global sequences isolated in these three years (all year round). As a result, the targeted summer strains might present different antigenic and genetic characteristics to their neighboring viruses along the timeline, so we performed NJ trees to sample only co-emerging viruses with those summer Taiwanese strains in each of the three years. To capture the global patterns of H3N2, Bayesian phylogeographic inference was applied to HA1 sequences, including 361,

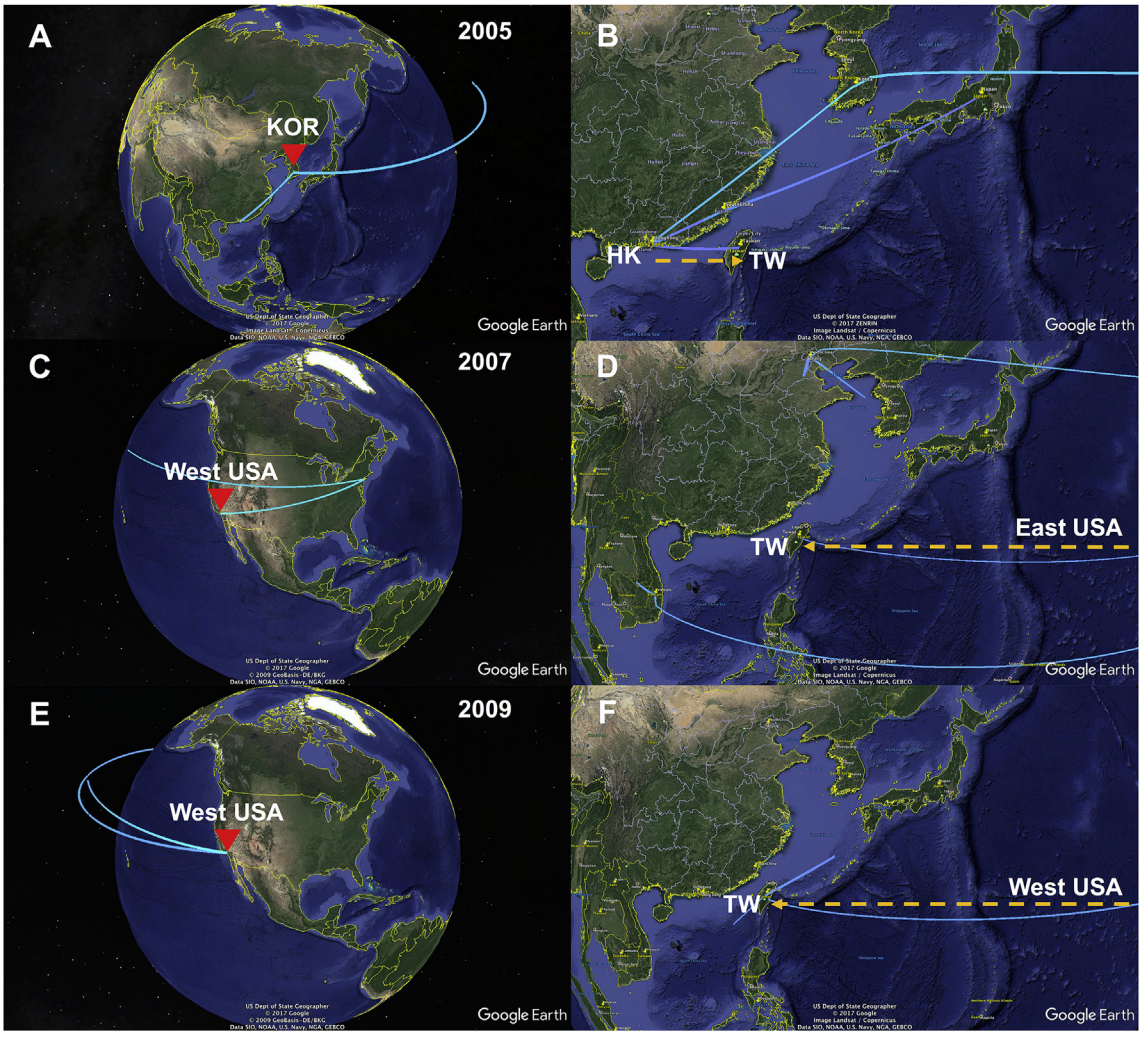
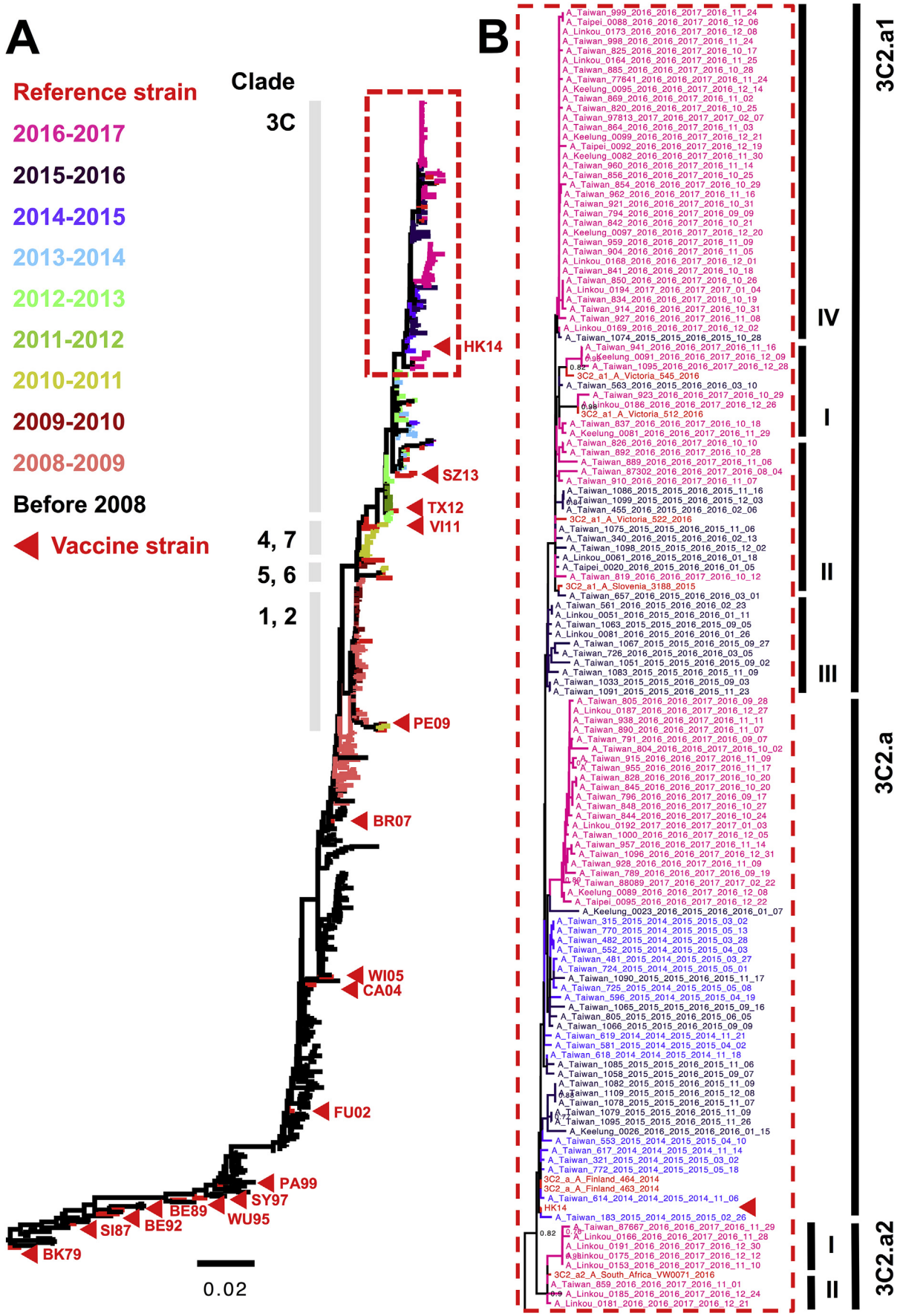


Figure 2 Global spatiotemporal inference of Taiwanese summer H3N2 virus migration. (A) 2005 summer epidemic was introduced from Korea; and (C) 2007 and (E) 2009 from the Western USA, which are marked by red triangles. (B) shows that the midpoint prior to Taiwan in the global migration of the 2005 summer flu was Hong Kong (dash line). (D) and (F) show the Taiwanese summer H3N2 viruses in 2007 and 2009 each directly related to the Eastern and Western USA, respectively.


A
Reference strain
2016-2017
2015-2016
2014-2015
2013-2014
2012-2013
2011-2012
2010-2011
2009-2010
2008-2009
Before 2008
Vaccine strain
Clade
3C
4, 7
5, 6
1, 2
HK14
SZ13
TX12
VI11
PE09
BR07
WI05
CA04
FU02
PA99
SY97
BE89
WU95
SI87
BE92
BK79
0.02
B
3C2.a1
IV
I
II
III
3C2.a
3C2.a2
I
II
3C2_a1_A_Victoria_545_2016
3C2_a1_A_Victoria_512_2016
3C2_a1_A_Victoria_522_2016
3C2_a1_A_Slovenia_3188_2015
3C2_a_A_Finland_464_2014
3C2_a_A_Finland_463_2014
HK14
3C2_a2_A_South_Africa_VW0071_2016

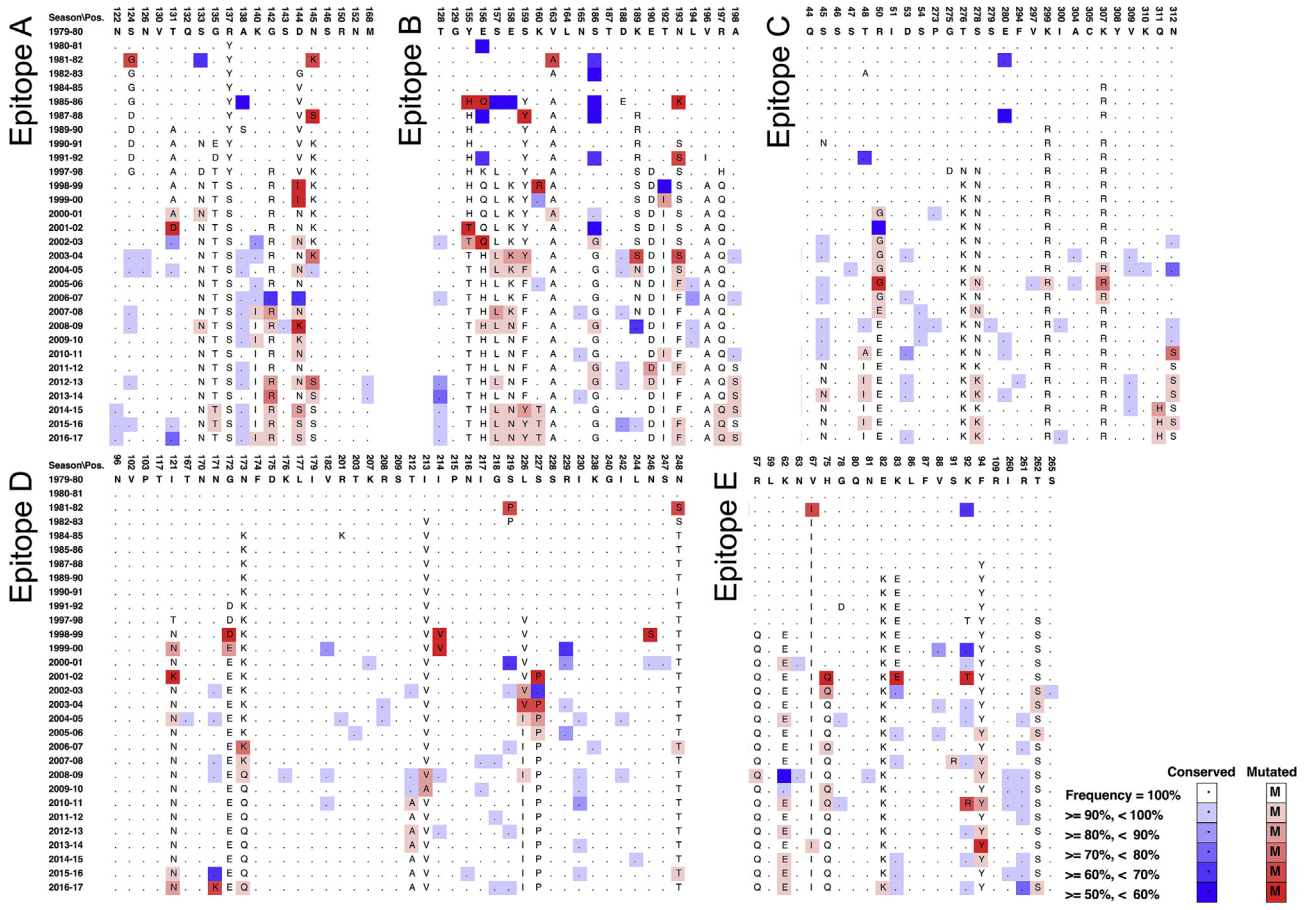


Figure 4 Amino acid transitioning in the five epitopes of Taiwanese strains over the past four decades. Using the amino acid residues in 1979-80 season as a base, a letter appearing in any subsequent seasons represents a dominant residue which is different from the base residue, and a dot represents the same base residue was maintained. Different shadings were used for the letter conservation levels, with a completely white background showing 100% dominance for either the original base (a dot) or a mutated residue.

589, and 574 worldwide strains of 2005, 2007, and 2009, respectively. Globally migrating patterns of these H3N2 viruses in 2005, 2007, and 2009 are presented in Fig. 2. Supplementary Files S1 to S3 show their animated spatio-temporal projections simulated using Google Earth. Fig. 2A shows that the 2005 summer epidemic originated from Korea, then spread simultaneously to Hong Kong and the Western USA. The route to Hong Kong was followed directly by Taiwan (dash line in Fig. 2B). Fig. 2C and E showed that the 2007 and 2009 summer epidemics both originated from the Western USA, while the 2007 virus went to the Eastern USA followed by multiple Asian destinations including Taiwan (Fig. 2D), and the 2009 spread across the Pacific directly to Taiwan (Fig. 2F). Note that each node in an MCMC sampling was annotated with a location in the phylogeographic reconstruction. The certainty of this tree reconstruction was calculated by using the frequencies of root state across the MCMC samplings. Frequencies of these three mentioned roots were 82.5% (Korea, 2005), 29.9% (Western USA, 2007), and 17.55% (Western U.S., 2009). Second-place root frequencies for the three cases were 10.35% (Western USA), 27.25% (Eastern USA), and 15.75% (Taiwan), respectively. While the 2005 Korean origin was well supported by a high root frequency, the source for 2007 summer viruses could go either Western (29.9%) or Eastern USA (27.2%). Similarly, Taiwan in 2009 could also be the possible origin (15.75%) comparing to the top 17.55% in the western USA root.

Phylogenetic evolution and viral quasispecies of the Taiwanese H3N2 strains

We collected all HA sequences of Taiwanese strains (n = 543) over the past four decades, and 24 reference and 14 vaccine strains from this time span. An NJ tree was reconstructed to present clade circulations and transitions of the Taiwanese H3N2 strains, as shown in Fig. 3A. To

Figure 3 Neighbor-joining tree of the Taiwanese H3N2 viruses in recent decades. (A) The construction of clade transitions and circulations used a neighbor-joining tree of 38 reference/vaccine and 543 Taiwanese strains from 1979 to 2017. Major genetic clades were annotated. (B) This subtree showed strains after 2014 with the majority belonging to Clades 3C2.a, 3C2.a1 (with four sub-clusters), and 3C2.a2 (with two). Significant bootstrap support values greater than 70% were labeled at nodes.

summarize, 2008–2009 strains belonged to Clades 1–2 (dominant) and 5; 2009–2010 strains to Clades 2 (dominant), 1, 4, and 7; 2010–2012 strains to Clades 4 (dominant), 1, 6, and 3C; and 2012–2014 strains to 3C. The vaccine strain A/Perth/16/2009 (adopted in 2010/11 and 2011/12 seasons, abbreviated as PE09) was clustered with TW 2008–2011 strains in Clade 1; VI11 (2012/13 and 2013/14 vaccine strain) and TX12 (2014/15 vaccine strain) with 2010–2013 strains in Clade 3C1; A/Switzerland/9715293/2013 (2015/16 vaccine strain, SZ13) with 2012–2014 strains in Clade 3C3; and A/Hong Kong/4801/2014 (2016/17 and 2017/18 vaccine strain, HK14) with 2014–2017 strains in Clade 3C2.a. In general these vaccine strains were found circulating in Taiwan approximately 2–4 seasons ahead before they were selected as the vaccine strain in respective season. For example, TX12 was designated as a vaccine strain in 2014/15 season. However, TW strains as early as in 2010/11 season were already found clustering with TX12 in Clade 3C1. Detailed phylogenies over the past four decades were provided in Supplementary Fig. S1. A subtree in Fig. 3B further shows the topology for recent isolates: 2014–2015 strains mostly belonged to 3C2.a (17 out of 21 strains), followed by 3C2 (n = 2, data shown in Supplementary Figs. S1) and 3C3.b (n = 2, Supplementary Fig. S1); 2015–2016 strains mostly to 3C2.a1 (21 out of 35), and some to 3C2.a (n = 13) and 3C3.b (n = 1, Supplementary Fig. S1). 2016–2017 strains belonged to 3C2.a1 (n = 34 in Cluster IV, n = 7 in Cluster I, and n = 6 in Cluster II), followed by 3C2.a (n = 22), 3C2.a2 (n = 8), and 3C3.b (n = 1, Supplementary Fig. S1). Note that Cluster IV in Clade 3C2.a1 was newly discovered in this study, and the eight 2016–2017 strains in Clade 3C2.a2 were the first Taiwanese H3N2 viruses ever made into this Clade.

This new 3C2.a1 Cluster IV included 35 Taiwanese strains (34 from the 2016–2017 season and one from 2015 to 2016). A novel substitution, N31S (all 34 from 2016 to 2017), was further identified when compared to the reference strain 3C2.a1. Moreover, 12 strains (11 from 2016 to 2017 and one from 2015 to 2016) in this new cluster contained a mixture of nucleotides at codon positions 478–480 corresponding to amino acid position 160. Further exploring HA1 sequences from other Taiwanese strains revealed a total of 33 having this quasispecies, with nucleotide compositions of AMA (n = 21), AYA (n = 8), AHA (n = 1), ASA (n = 1), RYA (n = 1), and RCA (n = 1). These resulted in a single amino acid polymorphism consisting of A, I, K, R, T, and V. The prevalence of this particular quasispecies were 28.6% (6 out of 21) in 2014–2015, 22.9% (8 out of 35) in 2015–2016, and 24.4% (19 out of 78) in 2016–2017. It was suggested that this quasispecies persistently existed from 2014. Interestingly, the remaining 23 strains having no this quasispecies in Cluster IV also showed diverse residues of T (n = 18), K (n = 4), and R (n = 1) at HA-160.

As described in Clade 3C2.a2, four HA substitutions including I58V and S219Y in Cluster I, and N122D and S262N in Cluster II emerged in London strains from 2016 to 2017. Five out of eight Taiwanese strains in Clade 3C2.a2 had the former two substitutions, and only one strain had the latter two. In addition to these four substitutions, all eight Taiwanese 3C2.a2 strains showed an amino acid substitution Q at position 197, when compared to reference strain A/South Africa/VW0071/2016 with residue K.

In addition to amino acid changes in specific clades, we comprehensively explored amino acid transitionings of 131 antigenic sites in five epitopes to gain an insight into the evolutionary history of Taiwanese H3N2 strains. In each of these seasons, frequencies of the dominant residues were summarized in Fig. 4, in which a letter or a shaded dot signaled the presence of site mutation using 1979-80 sequences as a base. One example is the dominant residue K in HA-171 of epitope D in the 2016-17 season, which is a mutated residue from N showing a frequency of 60.3% in K (shaded red). Another example is a wild-type T in HA-131 of epitope A also in 2016-17 season, with a T frequency of 71.8% (shaded blue dot). Since only dots of completely white background represent 100% conserved to the wild-type residue, HA-131 for the 2016-17 season was considered having mutations. It was found that 14 (73.7%), 19 (86.4%), 22 (81.5%), 28 (68.3%), and 17 (77.3%) out of 19, 22, 27, 41, and 22 antigenic sites in epitope A, B, C, D, and E, respectively, presented at least one mutation over the past four decades. This amino acid transitioning map showed not only frequencies of antigenic sites, but also their changes through time. A cell of darker background indicates the residue composition is more diverse and potentially on the verge to change. An epidemic caution should be taken if amino acid transitioning emerged or occurred in consecutive seasons (e.g., HA-171 of epitope D in 2015–2017 seasons).

Discussion

Phylogenetic and evolutionary dynamics of the Taiwanese H3N2 strains over time were investigated in this study. Sequence data prevalence and BSPs clearly revealed the stepwise emergence of strains associated with novel genetic clades in a Bayesian phylogenetic tree. As reported, summer influenza surveillance provided important clues of global epidemics. In this study, Bayesian phylogeographic analysis was used to trace the origins of viral epidemics in 2005, 2007, and 2009. Moreover, clade circulations and transitions of the Taiwanese strains were completely described, and novel genetic diversities were discovered in recent strains after 2014.

In order to trace the origin of Taiwanese summer H3N2 viruses, in each sampling period we chose to group sequences primarily by country, in addition to the Eastern and Western USA, as well as the Northern, Southern, and Hong Kong SAR in China, leading to 9 to 16 geographical locations worldwide. We found that Korea was the source of summer epidemics in 2005 with excellent statistical support. While U.S. continent was predicted the origin of the 2007 and 2009 summer epidemics, Asian destinations including Taiwan were the immediate followers in these outbreaks.

Vaccination has been considered as an effective way to prevent influenza virus infection; however, people who have been vaccinated still risk infection. There are two major reasons to explain this phenomenon: the first involves waning antibodies in the host; and second, the antigenicity drift in the virus via molecular mechanisms (e.g., mutation, quasispecies). In this study, we discovered a new Cluster IV featuring the mutation N31S in Clade 3C2.a1. While there was only a small portion of worldwide H3N2

viruses showed N31S (1.7% from 2015 to 2016, increased to 7.0% in 2017, data provided by nextflu[28]), 44% of Taiwanese strains in 2016—2017 exhibited this substitution. Although HA 31 was not located in any known antigenic sites of HA gene, accompanying this substitution is a high frequency HA 160 quasispecies in this new cluster (34%) than in the other three (16%). This mixture position at 160 was important for antigenicity changes,[29] and seemed to continue to circulate and be sustained in Taiwanese strains, which may have cause infections in vaccinated patients.

The BSP in Fig. 1C showed peaks or plateaus for the effective population, that did not perfectly coincide with sequence counts of our dataset (Fig. 1A), or epidemic peaks observed in influenza surveillance report in Taiwan. This could be a consequence of limited availability and insufficient sampling diversity of sequences collected from public database for such phylodynamic analyses.[2,30] We mentioned that although 883 TW sequences were retrieved from GISAID, only 543 long-enough sequences (>95% of the 987-nt HA1) were used in producing Fig. 1. While some of the disqualified 340 sequences were removed because of redundancy, 202 were due to either missing month/day (n = 169) or day (n = 33) information for such analyses. Even all these 202 sequences were properly dated, there always exists a possible gap between the published TW sequences and the actual virus population. In future work, a comprehensive surveillance data is needed to unravel the genetic dynamics associated with population bottlenecks.

To conclude, we investigated the phylogenetic and evolutionary dynamics of HA1 sequences from Taiwanese H3N2 strains where the effective increase in virus population size was found elevated near the summer of 2009. New phylogenetic lineages that emerged in summer may have been progenitors of seasonal epidemic strains; however, the sources of summer epidemic are still unknown. We predicted that the Taiwanese 2005 summer H3N2 viruses originated from Korea; and that the ones from summer 2007 and 2009 were from the Eastern and Western USA, respectively. In addition, the effective increase in population size also rose from 2015 to 2017. To trace this change, clade circulations and transitions of Taiwanese strains from the last decade were completely described, and a new Cluster IV in Clade 3C2.a1 was discovered. We also further identified the novel substitution N31S and an HA-160 quasispecies from recent Taiwanese strains. Observing that the BSP elevation stays high at the end of 2016—2017 sampling period (February 2017), it is suggested that novel H3N2 clades may continue emerging. Phylodynamic and phylogeographic analyses allowed us not only to trace the origins of newly emerging viruses, but to also detect novel genetic diversity.

Conflicts of interest

The authors have no conflicts of interest relevant to this article.

Acknowledgments

The authors would like to thank Dr. Shin-Ru Shih (Chang Gung University, Taiwan) for constructive discussions. A total of six grants supported this work, including four (Grant No. MOST 104-2221-E-182-064, 105-2221-E-182-074, 106-2221-E-182-066, and 106-2320-B-182A-013-MY3) from the Ministry of Science and Technology, Taiwan, and two (Grant No. CLRPG3B0046 and CMRPG3F1882) from Chang Gung Memorial Hospital, Taoyuan, Taiwan.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jfma.2018.01.019.

References

1. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol* 2010 Oct;**84**(19):9733—48.
2. Lin J-H, Chiu S-C, Lin Y-C, Cheng J-C, Wu H-S, Salemi M, et al. Exploring the molecular epidemiology and evolutionary dynamics of influenza A virus in Taiwan. *PLoS One* 2013;**8**(4):e61957.
3. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev* 2012 Jun;**76**(2):159—216.
4. European Centre for Disease Prevention and Control (ECDC). Influenza virus characterization, summary Europe, March 2012. Available from: http://ecdc.europa.eu/en/publications/Publications/1204-TED-CNRL-report.pdf. [Accessed 1 August 2017].
5. Harvala H, Frampton D, Grant P, Raffle J, Ferns RB, Kozlakidis Z, et al. Emergence of a novel subclade of influenza A(H3N2) virus in London, December 2016 to January 2017. *Euro Surveill* 2017 Feb 23;**22**(8).
6. Park AW, Glass K. Dynamic patterns of avian and human influenza in east and southeast Asia. *Lancet Infect Dis* 2007 Aug;**7**(8):543—8.
7. Sunagawa S, Iha Y, Taira K, Okano S, Kinjo T, Higa F, et al. An epidemiological analysis of summer influenza epidemics in okinawa. *Intern Med* 2016;**55**(24):3579—84.
8. Fu Y, Pan L, Sun Q, Zhu W, Zhu L, Ye C, et al. The clinical and etiological characteristics of influenza-like illness (ILI) in outpatients in Shanghai, China, 2011 to 2013. *PLoS One* 2015;**10**(3):e0119513.
9. Chen J-F, Sun B-C, Yuan J, Zhang R-S, Ou X-H. Surveillance of influenza virus during 2010-2012 in Changsha, China. *Southeast Asian J Trop Med Public Health* 2014 Mar;**45**(2):319—25.
10. Zhang D, He Z, Xu L, Zhu X, Wu J, Wen W, et al. Epidemiology characteristics of respiratory viruses found in children and adults with respiratory tract infections in southern China. *Int J Infect Dis* 2014 Aug;**25**:159—64.
11. 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 Apr 18;**320**(5874):340—6.
12. Uyeki TM, Zane SB, Bodnar UR, Fielding KL, Buxton JA, Miller JM, et al. Large summertime influenza A outbreak among tourists in Alaska and the Yukon Territory. *Clin Infect Dis* 2003 May 1;**36**(9):1095—102.
13. Volz EM, Koelle K, Bedford T. Viral phylodynamics. In: Wodak S, editor. *PLoS computational biology*, vol. 9(3); 2013 Mar 21. e1002947.
14. Bedford T, Cobey S, Beerli P, Pascual M. Global migration dynamics underlie evolution and persistence of human influenza a (H3N2). In: Ferguson NM, editor. *PLoS pathogens*, vol. 6(5); 2010 May 27. e1000918.
15. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* 2015 Jun 8;**523**(7559):217—20.
16. 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 USA* 2011 Nov 29;**108**(48):19359—64.
17. Haveri A, Ikonen N, Julkunen I, Kantele A, Anttila V, Ruotsalainen E, et al. Reduced cross-protection against influenza A(H3N2) subgroup 3C.2a and 3C.3a viruses among Finnish healthcare workers vaccinated with 2013/14 seasonal influenza vaccine. *Euro Surveill* 2015 Feb 5;**20**(5):21028.
18. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, et al. Mapping the antigenic and genetic evolution of influenza virus. *Science* 2004 Jul 16;**305**(5682):371—6.
19. World Health Organization. WHO recommendations on the composition of influenza virus vaccines. Available from:http://www.who.int/influenza/vaccines/virus/recommendations/en/. [Accessed 24 January 2018]
20. Drummond AJ, Rambaut ABEAST. Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007 Nov 8;**7**:214.
21. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. *Bioinformatics* 2009 Jun 1;**25**(11):1370—6.
22. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006 May;**4**(5):e88.
23. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009 Sep;**5**(9):e1000520.
24. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016 Jul;**33**(7):1870—4.
25. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987 Jul;**4**(4):406—25.
26. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985 Jul;**39**(4):783—91.
27. Jukes TH, Cantor CR. *Evolution of protein molecules*. New York: In Mammalian Protein Metabolism; 1969. p. 31—132.
28. Neher RA, Bedford T. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* 2015 Nov 1;**31**(21):3546—8.
29. Skowronski DM, Sabaiduc S, Chambers C, Eshaghi A, Gubbay JB, Krajden M, et al. Mutations acquired during cell culture isolation may affect antigenic characterisation of influenza A(H3N2) clade 3C.2a viruses. *Euro Surveill* 2016;**21**(3):30112.
30. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. *PLoS Pathog* 2010 May 6;**6**(5):e1000884.