

# Evolution and rapid spread of a reassortant A(H3N2) virus that predominated the 2017–2018 influenza season

Barney I. Potter,<sup>1,\*</sup> Rebecca Kondor,<sup>2</sup> James Hadfield,<sup>1</sup> John Huddleston,<sup>1,3</sup> John Barnes,<sup>2</sup> Thomas Rowe,<sup>2</sup> Lizheng Guo,<sup>2</sup> Xiyan Xu,<sup>2</sup> Richard A. Neher,<sup>4,5,‡</sup> Trevor Bedford,<sup>1</sup> and David E. Wentworth<sup>2</sup>

<sup>1</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, USA, <sup>2</sup>Virology Surveillance and Diagnosis Branch, Influenza Division, National Center for Immunization and Respiratory Diseases (NCIRD), Centers for Disease Control and Prevention (CDC), 1600 Clifton Road, Atlanta, GA 30333, USA, <sup>3</sup>Molecular and Cellular Biology Program, University of Washington, 4109 E Stevens Way NE, Seattle, WA 98105, USA, <sup>4</sup>Biozentrum, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland and <sup>5</sup>SIB Swiss Institute of Bioinformatics, Klingelbergstrasse 70, 4056 Basel, Switzerland

\*Corresponding author: E-mail: barney.potter.24@gmail.com

†<http://orcid.org/0000-0001-9476-149X>

‡<http://orcid.org/0000-0003-2525-1407>

## Abstract

The 2017–2018 North American influenza season caused more hospitalizations and deaths than any year since the 2009 H1N1 pandemic. The majority of recorded influenza infections were caused by A(H3N2) viruses, with most of the virus's North American diversity falling into the A2 clade. Within A2, we observe a subclade which we call A2/re that rose to comprise almost 70 per cent of A(H3N2) viruses circulating in North America by early 2018. Unlike most fast-growing clades, however, A2/re contains no amino acid substitutions in the hemagglutinin (HA) segment. Moreover, hemagglutination inhibition assays did not suggest substantial antigenic differences between A2/re viruses and viruses sampled during the 2016–2017 season. Rather, we observe that the A2/re clade was the result of a reassortment event that occurred in late 2016 or early 2017 and involved the combination of the HA and PB1 segments of an A2 virus with neuraminidase (NA) and other segments a virus from the clade A1b. The success of this clade shows the need for antigenic analysis that targets NA in addition to HA. Our results illustrate the potential for non-HA drivers of viral success and necessitate the need for more thorough tracking of full viral genomes to better understand the dynamics of influenza epidemics.

**Key words:** influenza; phylogenetics; reassortment; adaptation.

## 1. Introduction

Influenza virus epidemics cause significant morbidity and mortality with about 10 per cent of the global population infected annually ([World Health Organization 2018](#)) and most recent mortality estimates indicate that there are

291,000–645,000 influenza virus-associated death each year ([Iuliano et al. 2018](#)). Seasonality of influenza epidemics can vary by geographic location. In temperate regions, influenza activity typically peaks in mid-winter, while seasonal patterns are less stereotypical in tropical or subtropical regions ([Petrova and Russell 2018](#)). In addition, the severity of

influenza epidemics can also vary by season and geography. The 2017–2018 influenza season in the USA was the most severe epidemic since the 2003–2004 season, resulting in 79,000 deaths and more than 330,000 hospitalizations in the USA alone (Garten et al. 2018).

Throughout life humans are repeatedly infected by influenza viruses, primarily due to antigenic drift wherein amino acid substitutions in the hemagglutinin (HA) and/or neuraminidase (NA) surface glycoproteins generate viruses that escape from immunity induced by prior infection or vaccination. Out of the four co-circulating lineages of epidemic influenza viruses, A(H3N2) evolves the most rapidly, with generally higher circulation levels, necessitating more frequent updates to the specific strain used in the seasonal influenza vaccine (Hay et al. 2001; Bedford et al. 2014). Due to ramifications in vaccine strain selection, improvements to tracking and predicting spread of A(H3N2) variants presents an opportunity for scientific insights to improve public health outcomes (Morris et al. 2017).

Research on influenza antigenic evolution focuses primarily on HA, whose antigenic properties are historically analyzed with hemagglutination inhibition (HI) assays (Hirst 1942). Amino acid substitutions in epitope regions, primarily in the globular head of the HA protein, alter antigenic phenotype of the virus as measured by HI titers (Smith et al. 2004; Koel et al. 2013; Bedford et al. 2014; Neher et al. 2016). However, NA evolves at nearly the same rate as HA and has similar signatures of adaptive evolution (Bhatt, Holmes, and Pybus 2011). In addition, recent work has shown the importance of anti-NA antibodies to protective immunity through natural infection and vaccination (Monto et al. 2015; Huang et al. 2018). This suggests that NA evolution is a substantial contributor to the changing antigenic properties and dynamics of influenza viruses. While most research is devoted to the study of antigenic drift via specific amino acid substitutions, separate genomic segments coding for HA and NA can reassort, allowing novel genomic constellations to arise, a process that occurs frequently in nature (Nelson et al. 2008; Marshall et al. 2013). Reassortment can sometimes create novel adaptive genotypes (Neverov et al. 2014; Dudas et al. 2015), but generally results in deleterious, incompatible genotypes (Rabadan, Levine, and Krasnitz 2008; Villa and Lässig 2017). In one notable example, the spread of the Fujian/2002 antigenic variant was attributable in part to reassortment between HA and other genomic segments, including NA (Holmes et al. 2005). The rapid development of sequencing technology and specific strategies for influenza virus genomics in the past decade enables sequencing of thousands of ‘full’ viral genomes annually, where each genome consists of the coding sequences on each of the eight influenza genome segments minus segment termini. Now, nearly all influenza-positive surveillance specimens received by the US Centers for Disease Control and Prevention are sequenced directly using a next-generation sequencing strategy and submitted to the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database (Elbe and Buckland-Merrett 2017). Sequencing direct clinical samples avoids issues of virus evolution during *in vitro* propagation. From this abundance of genomic data, the dynamics of reassortment can be analyzed at high temporal and genotypic resolution.

Here, we use tools from the Nextstrain project (Hadfield et al. 2018) to show that a reassortant A(H3N2) genome constellation dominated the 2017–2018 North American influenza season. We go on to show that viruses with this genome constellation had the HA and PB1 segments of one parental virus with the other six segments of another virus,

and that the progenitor virus likely emerged in late 2016 or early 2017.

## 2. Methods

### 2.1 Sequence and antigenic data

We built phylogenetic trees and frequency estimates for each segment of the influenza virus genome containing data from a 2-year window between April 2016 and April 2018, as well as a small set of older reference viruses. This window gives us enough context to see global influenza diversity while maintaining high resolution on the scale of months within individual clades. A total of 18,844 A(H3N2) sequences were downloaded from GISAID EpiFlu database with submission dates ranging between 15 April 2012 and 20 April 2018 (see [Supplementary Text](#)). From these, sequences were subsampled to achieve equal distribution of samples across geographic regions and over a 2-year window (see [Supplementary Text](#) TSV for full list of samples used in analysis). In addition to sampling evenly across countries and time, subsampling prioritized virus strains for which antigenic data were available and when possible there were sequence data available for all genomic segments. In total, 1,862 sequences were used to construct phylogenetic trees. In this analysis, we use antigenic data from focus reduction assays provided by collaborating centers from the WHO Global Influenza Surveillance and Response System.

### 2.2 Phylogenetic analysis

For each segment of the influenza virus genome, alignments were created using MAFFT (Katoh and Standley 2013). Aligned sequences were then cleaned of insertions relative to the reference to maintain consistent numbering of sites, and viruses that deviate too much from the expected molecular clock were removed. For each tree we included a set of reference viruses that are well characterized (i.e. past vaccine strains) with sampling dates before the 2-year window of interest. We then constructed phylogenetic trees using FastTree (Price, Dehal, and Arkin 2009) according to the Jukes–Cantor model of nucleotide evolution and a CAT model of site heterogeneity. After trees were constructed, we refined them using RAXML (Stamatakis 2014). We inferred dates and ancestral sequences of each internal node of the trees (Sagulenko, Puller, and Neher 2018). Finally, we annotated trees with attributes (i.e. geographic location), and inferred these traits for ancestral nodes using TreeTime. TreeTime (Sagulenko, Puller, and Neher 2018) implements a discrete trait phylogeographic analysis following (Lemey et al. 2009). Viruses are assigned to ten broad geographic regions and transition rates between these regions are estimated by maximizing likelihood of a CTMC model. Ancestral locations are estimated as the marginal distribution on ancestral node states following this CTMC model.

### 2.3 Reassortment analysis

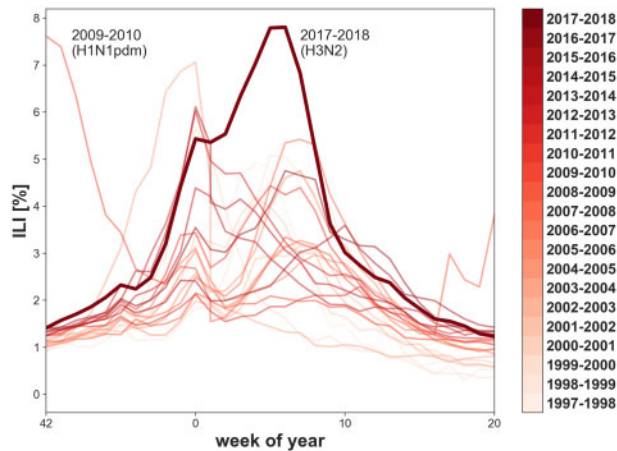
To detect reassortment, we constructed tangle trees—sets of two trees representing different segments, with tips from matching viruses connected by lines. We then compared topologies of tangle trees that matched the phylogeny of HA segment with that of other segments. When incongruent tree topologies were found, we compared whether the viruses labeled as A2 and A2/re in the HA phylogeny were adjacent in the other segment’s phylogeny, and if it appeared that A2/re was a subclade of A2. If not, we inferred that a reassortment had taken place

that combined the HA segment of an A2-like background with the other segment from a separate background.

### 3. Results

#### 3.1 Severe 2017–2018 North American influenza season

In many aspects the 2017–2018 North American influenza season was the most severe influenza epidemic in the past twelve seasons, resulting in a large increase in visits for outpatient influenza-like illness (ILI) (Fig. 1) and resulted in an estimated

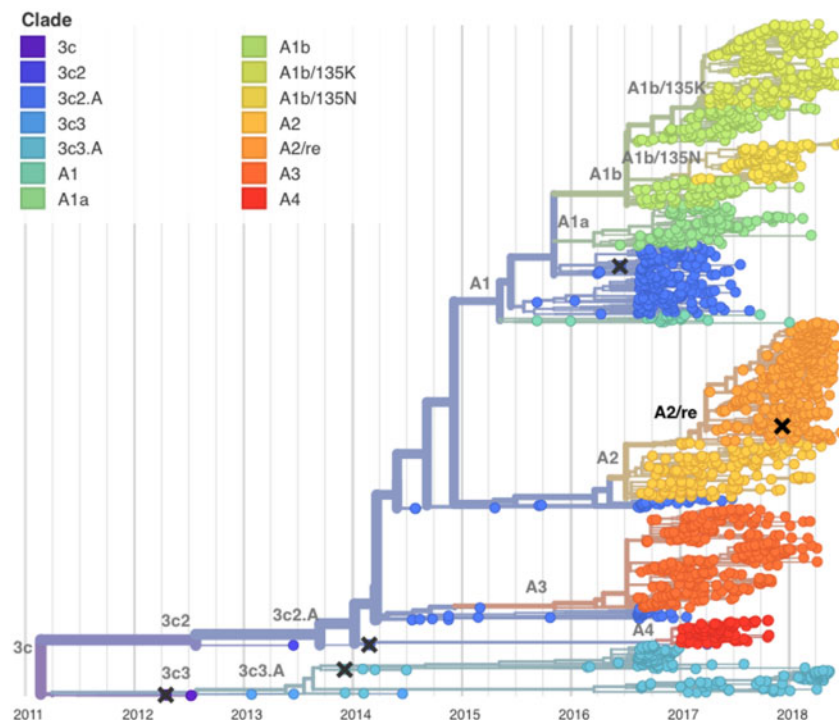


**Figure 1.** Percentage of health care visits for ILI in the US from the 1997–1998 to the 2017–2018 season with season indicated by color. The 2017–2018 season is highlighted as a thick line; it showed the highest ILI rate since the 2009 H1N1 pandemic virus. Data from [Centers for Disease Control and Prevention, \(2018\)](#).

960,000 hospitalizations and 79,000 influenza-associated deaths, particularly among those sixty-five and older ([Centers for Disease Control and Prevention, 2018](#)). In the USA, there were over 23,000 confirmed cases of influenza A viruses between October 2017 and February 2018; of these 89.9 per cent were A(H3N2) viruses. At the season's peak, influenza A viruses were responsible for 7.7 per cent of all hospitalizations and 10.1 per cent of deaths in the USA ([Budd et al. 2018](#)). The fraction of health care visits for ILI surpassed those of the 2009 influenza epidemic season (Fig. 1)—a season marked by the emergence of the A(H1N1)pdm09 pandemic virus, that continues to cause annual epidemics in humans ([Garten et al. 2009](#)). In addition to rising to high levels, the rate of hospital visits of symptoms of ILI stayed above 4 per cent for 10 weeks during the 2017–2018 North American influenza season. Unusually, the North American influenza season was disproportionately intense compared to other global regions.

#### 3.2 Genetic diversity of recent A(H3N2) viruses

Here, we analyze viral clades, representing clonal descendants of a single common ancestor, on the HA segment of A(H3N2). Rapid clade turnover results in an A(H3N2) HA phylogeny in which contemporaneous viruses typically coalesce to a single common ancestor within 3 years ([Bedford et al. 2015](#)). However, the current most recent common ancestor (MRCA) of the global A(H3N2) virus population, existed more than 6 years ago in early 2012 (Fig. 2). Correspondingly, there is significant genetic diversity among the HA segments of circulating viruses, with the competing clades 3c3.A and 3c2.A comprising the deepest split in the phylogeny. Descendants of the 3c3.A viruses circulated at low frequencies in the global population between 2016 and 2018. Several HA subclades of 3c2.A, comprising



**Figure 2.** Temporally resolved phylogenetic tree of recent A(H3N2) HA sequences. Shown are influenza viruses sampled between April 2016 and April 2018 colored by clade designation (in addition, the tree shows a small number of earlier reference viruses). Viruses are subsampled to give comparable counts across time and geography. HA segments of the majority of co-circulating A(H3N2) viruses in of the 2017–2018 Northern Hemisphere influenza season fell within clades A1b (top, yellow) or A2 (center, orange). Vaccine viruses are indicated by cross marks.



3c2.A1 to 3c2.A4, circulated at high frequencies between 2016 and 2018. Throughout the following, we drop the 3c2 to abbreviate these names to A1a, A1b, A2, A3, and A4. By April 2018, viruses in clades A1b and A2 predominated the global population, comprising an estimated 86 per cent of A(H3N2) viruses (Fig. 2).

Clade A1 emerged in 2015 and represented the majority of influenza A(H3N2) viruses in the 2016–2017 season. Most viruses in this clade differ from the previous vaccine virus (A/Hong Kong/4801/2014) by amino acid substitutions in HA (N121K, N171K, I406V, and G484E). Clade A1 further diversified into two subclades now named A1a and A1b. These are distinguished by the amino acid substitutions HA:G479E and HA:K92R,H311Q, respectively. Further rapid evolution took place within A1b which split into two subclades defined by the HA amino acid substitutions E62G, R142G, T135K, and T135N. In the past year A1b has out-competed A1a.

Clade A2 is defined by a series of amino acid substitutions in HA: T131K, R142K, R261Q. This series of substitutions completed in mid-2016; no further amino acid changes were observed and the frequency of this clade remained fairly constant from late 2016 to mid-2017 (Fig. 3). In the 2017–2018 season in North America, however, an A2 HA subclade dominated the viruses circulating, yet lacked additional amino acid substitutions. This subclade—which we denote as A2/re (Fig. 2)—appears to have arisen as a result of a reassortment event, as discussed below. The rapid rise of subclade A2/re—coupled with an extraordinarily high incidence of that subclade in North America—stands out. By the beginning of 2018, we find the clade viruses belonging to clade A2 increasing to make up almost 70 per cent of A(H3N2) circulating viruses, overtaking A1b as the dominant HA clade (Fig. 3).

Across the entire HA phylogeny, we observe several sites that undergo repeated nucleotide transitions. One example of this is at site HA:160, which shows signatures of positive selection when analyzed by SLAC (Kosakovsky Pond and Frost 2005), but this seems to be driven by frequent, unsuccessful mutations, rather than ones that ultimately find success (Supplementary Table S2 and Fig. S1). This presents an issue as sites that undergo the same mutation throughout the tree and show signatures of positive selection are often the result of passaging adaptation, not to adaptations that arise within natural viral circulation (Meyer and Wilke 2015; McWhite, Meyer, and

Wilke 2016). To circumvent this issue, we instead focus on mutations that fix or rise to high frequency to determine adaptive origin.

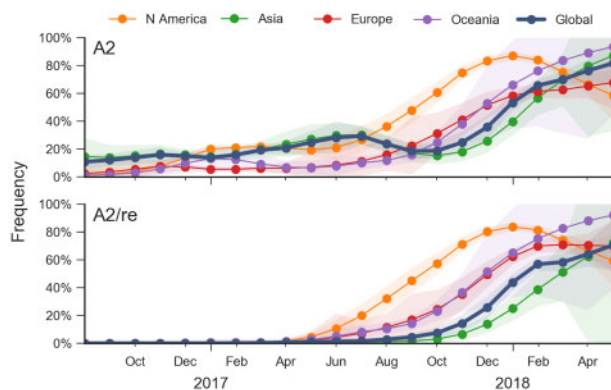
### 3.3 Antigenic analysis illustrates little antigenic drift

The impact that the amino acid substitutions of the various co-circulating HA clades have on antibody escape were analyzed using postinfection ferret antisera against cell-propagated viruses including vaccine strains (Supplementary Table S1). The recommended A(H3N2) vaccine viruses were A/Hong Kong/4801/2014-like (i.e. A/Michigan/15/2014) for both the 2016–2017 and the 2017–2018 Northern Hemisphere seasons. Focus reduction assay of reference viruses and contemporary test viruses circulating worldwide demonstrate that antisera to either HA clade 3c2.A A/Michigan/15/2014 virus or HA clade 3c2.A1 viruses A/Singapore/INFIMH-16-0019/2016 efficiently neutralize viruses expressing 3c2.A, A1, or A2 HA RNAs (Supplementary Table S1 and Fig. S2). In fact, only viruses expressing 3c3.A HAs, which represent a small proportion of viruses circulating, showed evidence of significant antigenic drift (i.e. 8-fold or greater reductions in titer as compared to titers of the homologous virus and antisera) (Supplementary Table S1). Nevertheless, antisera to a clade A2 virus (A/Nebraska/2/2017) efficiently neutralizes other A2 viruses and poorly neutralizes viruses from other HA clades (Supplementary Table S1 and Fig. S2). Collectively, antigenic analysis data indicates that while the A2 HA proteins are antigenically distinct, they have not significantly drifted from 3c2.A HA proteins and it does not explain their predominance in North America during the 2017–2018 season. Consistently, substitutions to epitope sites show that the A2 clade is slightly drifted relative to ancestral 3c2.A viruses, but show little variation within clade A2 (Supplementary Fig. S2).

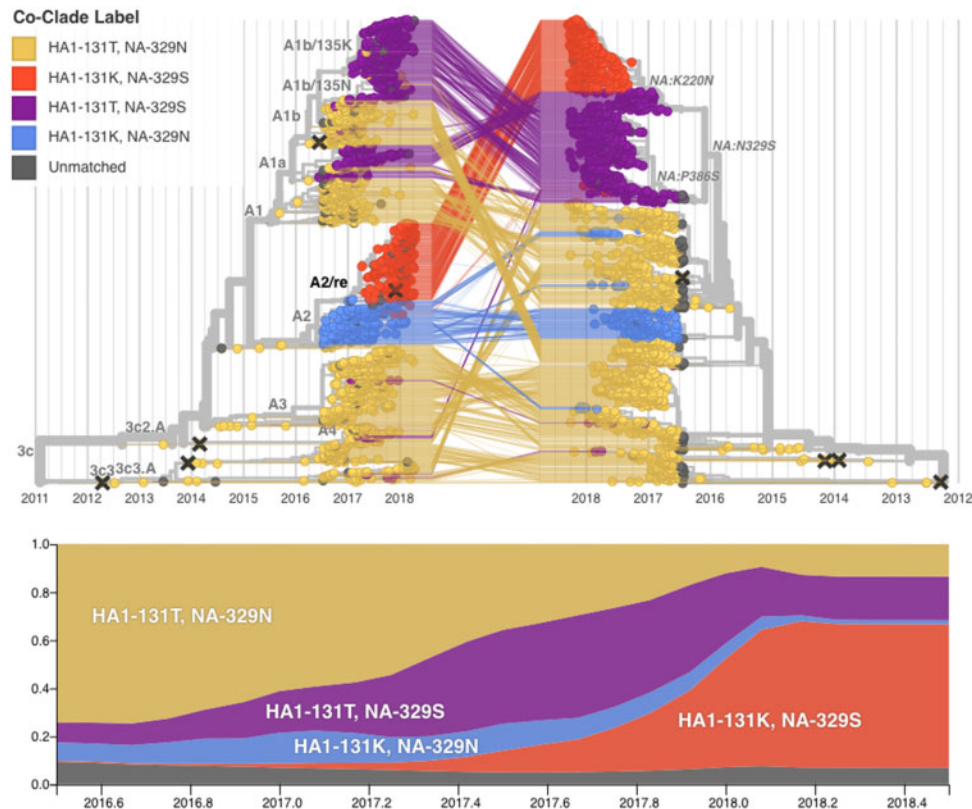
### 3.4 Genome reassortment events among circulating viruses leads to A2/re genotype

When reassortment occurs between viruses from established clades in the HA and NA RNA segments, incongruent phylogenetic relationships are created so that a reassortment event within clade A2 viruses is immediately visible (Fig. 4). In the HA phylogeny, viruses from A2 and A2/re are closely related and form a monophyletic clade. However in the NA phylogeny, the NA of A2 viruses appear in a clade distant from the NA of A2/re viruses. Viruses whose HA belongs to clade A2/re possess NA with the N329S substitution which is shared with viruses that belong to clade A1b. The NA segments of viruses belonging to HA clade A2/re differ from viruses belonging to clade A2 by NA substitutions I176M, K220N, N329S, D339N, and P386S. Importantly, the amino acid substitution N329S is present in the NA of viruses from both successful HA clades (A2/re and A1b), as indicated by the local branch index which summarizes recent clade growth (Supplementary Fig. S3) (Neher, Russell, and Shraiman 2014). This reassortment was verified statistically by use of GiRaF (Nagarajan and Kingsford 2011), which confirms the strains present in the A2/re clade and those in the matched NA clade as being part of a reassortment event (conf = 0.99, Supplementary Fig. S4).

Temporal phylogenetic analysis places the MRCA of the A2/re genotype in the HA segment between December 2016 and February 2017 (95% HPD 28 December 2016 to 11 February 2017) and in the NA segment between January 2017 and March 2017 (95% HPD 24 January 2017 to 13 March 2017), suggesting that most likely a reassortment event in January or February 2017



**Figure 3.** Frequency trajectories of A2 and A2/re clades partitioned by region. We estimate frequencies of different clades based on sample counts and collection dates. These estimates are based on all available data and global frequencies are weighted according to regional population size and relative sampling intensity. We use a Brownian motion process prior to smooth frequencies from month-to-month (Neher and Bedford 2015). Transparent bands show an estimate the 95 per cent confidence interval based on sample counts.



**Figure 4.** (Top) Reassortment combines existing genetic diversity and results in incongruent phylogenetic trees. The figure shows time-scaled phylogenetic trees of HA (left) and NA (right) colored genotype at HA site 131 and NA site 329. The HA and NA sequences from the virus isolates are connected by lines. We see that the closest relatives for the viruses belonging to A2/re genotype (red) in HA are A2 viruses (blue), while their closest relatives in NA are A1 viruses (purple). (Bottom) Global estimated frequency plots of virus clades over time. There is moderate growth of A1 viruses prior to the reassortment, but after the reassortment takes place we observe rapid growth of the reassortant clade to approximately 70 per cent frequency. Branches on the HA tree are labeled by HA clade designation; branches on the NA tree are labeled by NA amino acid substitution. Vaccine strains are indicated by cross marks.

gave rise to a virus that seeded the rapidly growing population with this genome constellation. We observe that viruses basal to the reassortant A2/re clade in both HA and NA are predominantly from North and South America with A/Minnesota/12/2017, collected in February 2017, representing an early example. A discrete trait phylogeographic model places the root of the A2/re reassortment event as 83 per cent North America/17 per cent South America (HA) and 100 per cent North America (NA).

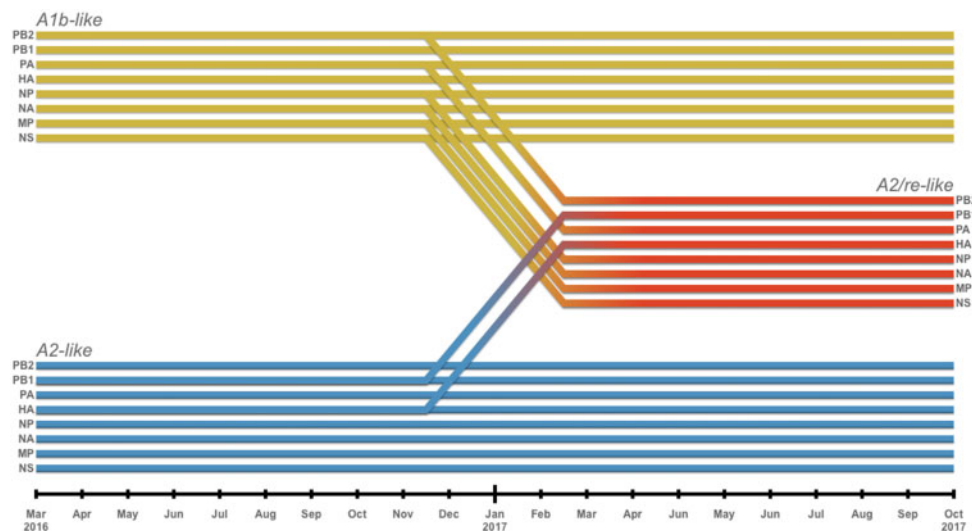
To quantify the rise of A2/re, we examined two mutations—one in HA and one in NA—that combine to identify the beginning of the reassortant constellation. The presence of all four pairwise combinations of these two mutation states indicates that a reassortment event has occurred, as only three states can arise in the absence of reassortment. In the HA phylogeny we use the mutation HA:T131K, and in NA we use NA:N329S. In our data, the three non-reassortant genotypes were observed at frequencies around 10–20 per cent in 2016, but the A2/re genotype first appeared in February 2017. By December 2017, this genotype had risen to above 70 per cent frequency in North America; this is a comparable rate to that at which the 3c3 clade rose to predominance during 2012–2013 (Fig. 4). For this analysis, we use only HA:T131K as an identifying mutation for the derived clade, though we can find the same results through using HA1 sites 142 and 261 as all three can be used to define the transition from the ancestral state of HA—in our case the split of A2 from A1.

The HA phylogeny for A2/re viruses are also incongruent compared to A2 viruses within trees for the six other RNA segments (PB2, PA, NP, NA, MP, NS). As with the NA, the six other RNA

segments of A2/re viruses are most closely related to those of HA clade A1b containing viruses (Supplementary Fig. S5). Phylogenetic analysis of the PB1 shows that viruses from the reassortant clade A2/re are adjacent to A2 viruses (Supplementary Fig. S5). Upon examining all other segments for incongruent phylogenies, we find evidence that the A2/re genotype is the result of reassortment events pairing HA and PB1 segments from A2 viruses with all other segments from A1b viruses (Fig. 5).

#### 4. Discussion

Most A(H3N2) influenza viruses that circulated during the 2017–2018 season in North America had hemagglutinin consensus sequences that were very similar to those of A2 viruses, which also circulated during the previous 2016–2017 season. Consistently, antigenic analysis did not suggest significant antigenic evolution among these viruses. Only genomic sequencing and analysis of all viral RNA segments revealed that recent A2 viruses differ substantially in six of the eight RNA segments. This new genotype which we call A2/re is likely the result of a single reassortment event that took place in late 2016 or early 2017. This reassortant genotype combines the HA and PB1 segments of A2-like viruses with all other segments from an A1b-like virus. Viruses with the A2/re genotype quickly rose to high frequency within North America, coincident with one of the most severe influenza virus epidemics in recent years. During the 2018 influenza vaccine selection for the Southern Hemisphere's 2019 season A/Switzerland/8060/2017, an A2/re



**Figure 5.** Segments from two backgrounds of A(H3N2) reassorted to form the A2/re clade (red). Segments HA and PB1 from an A2-like background (blue) combined with the six other segments from an A1b-like virus (yellow). Diagonal lines between November 2016 and February 2017 indicate the inferred timing of the reassortment.

virus, was selected as the A(H3N2) component of the vaccine (World Health Organization 2017).

While HA antigenic properties are well characterized across A(H3N2) viruses, antigenic characteristics of NA are not as well understood. The NA protein of these viruses differs from the previous A2 NA viruses at several positions including position 329. Position 329 changed repeatedly over the past few years and substitutions at this site could result in a loss of N-linked glycosylation. The N329S clade rose rapidly in prevalence prior to the reassortment event that gave rise to A2/re, supporting the hypothesis that the genomic background including the NA segment acquired in the reassortment event aided the success of A2/re viruses. In addition, the NA acquired two substitutions—I176M and P386S—prior to the reassortment event. Further data are required to understand the full effect of these substitutions on NA antigenicity, both within the A2/re viruses and other genotypes. In addition to antigenic effects, replication capacity of the virus depends on interactions between different segments and their stoichiometry, as has been demonstrated for different HA and NA variants (Yen et al. 2011).

While it is clear that there was rapid expansion/dissemination of A2/re genotype viruses, we cannot at this time show evidence that this clade's success was solely the result of the reassortment event. Reassortment is relatively common in influenza virus evolution (Holmes et al. 2005; Nelson et al. 2008). Hence assigning significance to any specific reassortment event is difficult and the rapid growth of the A2/re genotype might also be driven by other epidemiological factors. We also cannot be sure of the geographic origin of this reassortant virus. Our subsampling method samples evenly through time and geography—thereby minimizing geographic bias. However, we cannot be certain of the exact geographic origin of the reassortment prior to its rise to prominence within North America.

The capacity of influenza viruses to spread likely depends on the specific genome constellation of all eight RNA segments. Hence it is important to systematically survey viral evolution and reassortments of influenza viruses by genomic sequencing and a comprehensive analysis of these data. Retrospective analysis of historical data should be conducted to quantify the degree to which reassortment events give rise to successful genotypes (Nelson et al. 2008; Dudas et al. 2015; Villa and Lässig

2017). Results from such analysis can then inform predictive modeling efforts to anticipate composition of future influenza virus populations (Łuksza and Lässig 2014; Neher, Russell, and Shraiman 2014; Morris et al. 2017; Klingle et al. 2018). In addition to paramount importance for public health, the wealth of genomic longitudinal data with high spatiotemporal resolution make influenza viruses an ideal system to address general questions of evolutionary biology on epistasis, reassortment, and recombination.

The inclusion of genomic predictors that draw information from all eight RNA segments into fitness models will yield more robust prediction methodologies that account for more than just HA variance. We believe that integrating these sources of information into models of influenza evolution will help to predict future influenza genotype growth. Furthermore, it would aid mitigating the effect of future influenza outbreaks that would otherwise pose great risk to global human health.

## Supplementary Data

Supplementary data are available at *Virus Evolution* online.

## Funding

This work was supported by NIH National Institute of General Medical Science (NIGMS) grant (R35 GM119774-01 to T.B.) and by NIH National Institute of Allergy and Infectious Diseases (NIAID) grant (U19 AI117891 to T.B.). T.B. is a Pew Biomedical Scholar. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

**Conflict of interest:** None declared.

## References

Bedford, T. et al. (2014) 'Integrating Influenza Antigenic Dynamics with Molecular Evolution', *eLife*, 3: e01914.



- et al. (2015) 'Global Circulation Patterns of Seasonal Influenza Viruses Vary with Antigenic Drift', *Nature*, 523: 217–20.
- Bhatt, S., Holmes, E. C., and Pybus, O. G. (2011) 'The Genomic Rate of Molecular Adaptation of the Human Influenza A Virus', *Molecular Biology and Evolution*, 28: 2443–51.
- Budd, A. P. et al. (2018) 'Update: Influenza Activity—United States, October 1, 2017–February 3, 2018', *Morbidity and Mortality Weekly Report*, 67: 169–79.
- Centers for Disease Control and Prevention. (2018) Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths in the United States—2017–2018 Influenza Season—CDC (pubd online Dec 2018). <<https://www.cdc.gov/flu/about/burden/estimates.htm>> accessed 2 Dec 2018.
- Dudas, G. et al. (2015) 'Reassortment between Influenza B Lineages and the Emergence of a Coadapted PB1-PB2-HA Gene Complex', *Molecular Biology and Evolution*, 32: 162–72.
- Elbe, S., and Buckland-Merrett, G. (2017) 'Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health', *Global Challenges*, 1: 33–46.
- Garten, R. et al. (2018) 'Update: Influenza Activity in the United States during the 2017–18 Season and Composition of the 2018–19 Influenza Vaccine', *Morbidity and Mortality Weekly Report*, 67: 634.
- Garten, R. J. et al. (2009) 'Antigenic and Genetic Characteristics of Swine-origin 2009 A(H1N1) Influenza Viruses Circulating in Humans', *Science*, 325: 197.
- Hadfield, J. et al. (2018) 'Nextstrain: Real-time Tracking of Pathogen Evolution', *Bioinformatics*, 34: 4121.
- Hay, A. J. et al. (2001) 'The Evolution of Human Influenza Viruses', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356: 1861.
- Hirst, G. K. (1942) 'The Quantitative Determination of Influenza Virus and Antibodies by Means of Red Cell Agglutination', *Journal of Experimental Medicine*, 75: 49–64.
- Holmes, E. C. et al. (2005) 'Whole-genome Analysis of Human Influenza A Virus Reveals Multiple Persistent Lineages and Reassortment among Recent H3N2 Viruses', *PLoS Biology*, 3: e300.
- Huang, Q. S. et al. (2018) 'Risk Factors and Attack Rates of Seasonal Influenza Infection: Results of the SHIVERS Seroepidemiologic Cohort Study', *The Journal of Infectious Diseases*, 219: 347–57.
- Iuliano, A. D. et al. (2018) 'Estimates of Global Seasonal Influenza-associated Respiratory Mortality: A Modelling Study', *The Lancet*, 391: 1285–300.
- Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.
- Klingen, T. R. et al. (2018) 'In Silico Vaccine Strain Prediction for Human Influenza Viruses', *Trends in Microbiology*, 26: 119–31.
- Koel, B. F. et al. (2013) 'Substitutions Near the Receptor Binding Site Determine Major Antigenic Change during Influenza Virus Evolution', *Science*, 342: 976–9.
- Kosakovsky Pond, S. L., and Frost, S. (2005) 'Not So Different after All: A Comparison of Methods for Detecting Amino Acid Sites under Selection', *Molecular Biology and Evolution*, 22: 1208–22.
- Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520–16.
- Łuksza, M., and Lässig, M. (2014) 'A Predictive Fitness Model for Influenza', *Nature*, 507: 57–61.
- Marshall, N. et al. (2013) 'Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch', *PLoS Pathogens*, 9: e1003421.
- McWhite, C. D., Meyer, A. G., and Wilke, C. O. (2016) 'Sequence Amplification via Cell Passaging Creates Spurious Signals of Positive Adaptation in Influenza Virus H3N2 Hemagglutinin', *Virus Evolution*, 2: vew026.
- Meyer, A. G., and Wilke, C. O. (2015) 'Geometric Constraints Dominate the Antigenic Evolution of Influenza H3N2 Hemagglutinin', *PLoS Pathogens*, 11: 1–20.
- Monto, A. S. et al. (2015) 'Antibody to Influenza Virus Neuraminidase: An Independent Correlate of Protection', *Journal of Infectious Diseases*, 212: 1191–9.
- Morris, D. H. et al. (2017) 'Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology', *Trends in Microbiology*, 26: 102–18.
- Nagarajan, N., and Kingsford, C. (2011) 'GiRaF: Robust, Computational Identification of Influenza Reassortments via Graph Mining', *Nucleic Acids Research*, 39: e34.
- Neher, R. A. et al. (2016) 'Prediction, Dynamics, and Visualization of Antigenic Phenotypes of Seasonal Influenza Viruses', *Proceedings of the National Academy of Sciences*, 113: E1701–9.
- , and Bedford, T. (2015) 'Nextflu: Real-time Tracking of Seasonal Influenza Virus Evolution in Humans', *Bioinformatics*, 31: 3546–8.
- , Russell, C. A., and Shraiman, B. I. (2014) 'Predicting Evolution from the Shape of Genealogical Trees', *eLife*, 3: e03568.
- Nelson, M. I. et al. (2008) 'Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus since 1918', *PLoS Pathogens*, 4: 1–12.
- Neverov, A. D. et al. (2014) 'Intrasubtype Reassortments Cause Adaptive Amino Acid Replacements in H3N2 Influenza Genes', *PLoS Genetics*, 10: e1004037.
- Petrova, V. N., and Russell, C. A. (2018) 'The Evolution of Seasonal Influenza Viruses', *Nature Reviews Microbiology*, 16: 47–60.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009) 'FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix', *Molecular Biology and Evolution*, 26: 1641–50.
- Rabadan, R., Levine, A. J., and Krasnitz, M. (2008) 'Non-random Reassortment in Human Influenza A Viruses', *Influenza and Other Respiratory Viruses*, 2: 9–22.
- Sagulenko, P., Puller, V., and Neher, R. A. (2018) 'TreeTime: Maximum-likelihood Phylodynamic Analysis', *Virus Evolution*, 4: vex042.
- Smith, D. J. et al. (2004) 'Mapping the Antigenic and Genetic Evolution of Influenza Virus', *Science*, 305: 371–6.
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies', *Bioinformatics*, 30: 1312–3.
- Villa, M., and Lässig, M. (2017) 'Fitness Cost of Reassortment in Human Influenza', *PLoS Pathogens*, 13: e1006685.
- Wolf, Y. I. et al. (2006) 'Long Intervals of Stasis Punctuated by Bursts of Positive Selection in the Seasonal Evolution of Influenza A Virus', *Biology Direct*, 1: 34.
- World Health Organization. (2018) Influenza (Seasonal) (pubd online Nov 2018). <[http://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal))> accessed 4 Aug 2019.
- (2017) 'Recommended Composition of Influenza Virus Vaccines for use in the 2017/2018 Northern Hemisphere Influenza Season', *Weekly Epidemiological Record*, 93: 553–76.
- Yen, H. L. et al. (2011) 'Hemagglutinin–Neuraminidase Balance Confers Respiratory-droplet Transmissibility of the Pandemic H1N1 Influenza Virus in Ferrets', *Proceedings of the National Academy of Sciences*, 108: 14264–9.