Divergent evolutionary trajectories of influenza B viruses underlie their contemporaneous epidemic activity

Ramandeep K. Virk[a], Jayanthi Jayakumar[a], Ian H. Mendenhall[a], Mahesh Moorthy[b], Pauline Lam[a], Martin Linster[a], Julia Lim[a], Cui Lin[c], Lynette L. E. Oon[d], Hong Kai Lee[e], Evelyn S. C. Koay[e], Dhanasekaran Vijaykrishna[f,g], Gavin J. D. Smith[a,h,i,1], and Yvonne C. F. Su[a,1]

[a]Programme in Emerging Infectious Diseases, Duke-National University of Singapore (NUS) Medical School, Singapore 169857; [b]Department of Clinical Virology, Christian Medical College, Vellore, India 632004; [c]National Public Health Laboratory, Ministry of Health, Singapore 308442; [d]Department of Molecular Pathology, Singapore General Hospital, Singapore 169608; [e]Department of Laboratory Medicine, National University Hospital, Singapore 117597; [f]Department of Microbiology, Biomedicine Discovery Institute, Monash University, VIC 3800, Australia; [g]World Health Organization Collaborating Centre in Influenza Research and Surveillance, Peter Doherty Institute, VIC 3000, Australia; [h]SingHealth Duke-NUS Global Health Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore 169857; and [i]Duke Global Health Institute, Duke University, Durham, NC 27710

Edited by Peter Palese, Icahn School of Medicine at Mount Sinai, New York, NY, and approved November 18, 2019 (received for review September 25, 2019)

Influenza B viruses have circulated in humans for over 80 y, causing a significant disease burden. Two antigenically distinct lineages ("B/Victoria/2/87-like" and "B/Yamagata/16/88-like," termed Victoria and Yamagata) emerged in the 1970s and have cocirculated since 2001. Since 2015 both lineages have shown unusually high levels of epidemic activity, the reasons for which are unclear. By analyzing over 12,000 influenza B virus genomes, we describe the processes enabling the long-term success and recent resurgence of epidemics due to influenza B virus. We show that following prolonged diversification, both lineages underwent selective sweeps across the genome and have subsequently taken alternate evolutionary trajectories to exhibit epidemic dominance, with no reassortment between lineages. Hemagglutinin deletion variants emerged concomitantly in multiple Victoria virus clades and persisted through epistatic mutations and interclade reassortment—a phenomenon previously only observed in the 1970s when Victoria and Yamagata lineages emerged. For Yamagata viruses, antigenic drift of neuraminidase was a major driver of epidemic activity, indicating that neuraminidase-based vaccines and cross-reactivity assays should be employed to monitor and develop robust protection against influenza B morbidity and mortality. Overall, we show that long-term diversification and infrequent selective sweeps, coupled with the reemergence of hemagglutinin deletion variants and antigenic drift of neuraminidase, are factors that contributed to successful circulation of diverse influenza B clades. Further divergence of hemagglutinin variants with poor cross-reactivity could potentially lead to circulation of 3 or more distinct influenza B viruses, further complicating influenza vaccine formulation and highlighting the urgent need for universal influenza vaccines.

phylogeny | genetic diversity | natural selection | antigenic | vaccine

Two lineages of influenza B viruses ("B/Victoria/2/87-like" and "B/Yamagata/16/88-like," respectively, referred to as the Victoria and Yamagata viruses or lineages) cocirculate to contribute substantially to the ∼290,000–650,000 annual influenza-attributed global deaths (1). Influenza A and B viruses cause seasonal epidemics, and the clinical presentations of infection by both types are largely indistinguishable, with young children and the elderly being more susceptible to infection (1–3). Unlike influenza A viruses, which periodically emerge from animals to cause pandemics and circulate as single predominant antigenic variants, the Victoria and Yamagata lineages emerged as antigenic variants following divergence of influenza B viruses during the 1970s, and both lineages have cocirculated globally at least since 2001 (4). The recently developed quadrivalent vaccine protects against both B lineages (5), whereas the trivalent vaccines, which contains only 1 of the lineages, are frequently mismatched with respect to the circulation of the dominant influenza B lineage in a given season, resulting in low vaccine efficacy (5, 6). A universal vaccine that targets both B virus lineages remains a public health priority (7).

Influenza B viruses exhibit complex evolutionary dynamics, whose drivers are not fully understood (8–11). The hemagglutinin (HA) gene of the Victoria and Yamagata viruses evolve at a slower pace than in influenza A virus lineages, although the Victoria lineage undergoes faster rates of lineage turnover, with an average of 3.9–5.1 y, compared to the Yamagata lineage average of 6.3–7.2 y (11). Interestingly, despite slower antigenic turnover, Yamagata viruses infect a considerably older population (8, 10, 12), and diverged prior to 2000, resulting in the long-term cocirculation of 2 antigenically distinct clades (clades 2 and 3), while exhibiting an epidemiological pattern of alternating antigenic dominance between seasons (11). Since 2011, the Victoria lineage has also diversified into 2 dominant monophyletic clades (clades 1A and 1B) (11). The development of greater diversity of influenza B lineages raises a challenge to the existing quadrivalent vaccines.

Significance

Two influenza B viruses (Victoria and Yamagata) cocirculate in humans and contribute to the estimated 290,000–650,000 annual influenza-attributed deaths. Here, we analysed influenza B genomic data to understand the causes of a recent surge in human influenza B infections. We found that evolution is acting differently on Yamagata and Victoria viruses and that this has led to the cocirculation of a diverse group of influenza B viruses. If this phenomenon continues, we could potentially witness the emergence of 3 or more distinct influenza B viruses that could require their own vaccine component, thereby complicating influenza vaccine formulation and highlighting the urgency of developing universal influenza vaccines.

Author contributions: G.J.D.S. and Y.C.F.S. designed research; R.K.V., J.J., P.L., M.L., and J.L. performed research; C.L., L.L.E.O., H.K.L., and E.S.C.K. contributed new reagents; R.K.V., J.J., I.H.M., M.M., and Y.C.F.S. analyzed data; R.K.V., M.M., D.V., G.J.D.S., and Y.C.F.S. wrote the paper; and J.J. wrote in-house scripts.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences generated in this study have been deposited in GenBank (accession nos. MN588323–MN589586 [*SI Appendix*, Table S4]).

[1]To whom correspondence may be addressed. Email: gavin.smith@duke-nus.edu.sg or yvonne.su@duke-nus.edu.sg.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1916585116/-/DCSupplemental.

First published December 16, 2019.

Genome-wide analyses have suggested that, owing to the presence of greater genomic diversity, reassortment plays a crucial role in the evolution of influenza B viruses (9–11, 13). While Yamagata and Victoria viruses diverged across all 8 gene segments, frequent reassortment between the lineages has been observed. Intriguingly, the polymerase basic 1 and 2 (PB1 and PB2) and HA proteins have maintained separate Victoria and Yamagata lineages, suggesting that distinct segments were maintained through limited compatibility (9), as confirmed by in vitro experiments (14). Historically, influenza B virus epidemics have been characteristically low to sporadic in most countries during a year, with influenza A viruses predominant in most countries. Since the first reports of high influenza B virus activity in Europe in 2015, substantial epidemics have been reported globally, indicating major changes in the evolutionary and epidemiological dynamics of influenza B viruses (12, 15–19).

Here, we analyzed all available influenza B genomic data (>12,000) in combination with 158 genomes sequenced from Singapore to investigate the processes enabling the recent resurgence of epidemics due to influenza B virus (see global epidemic activity in *SI Appendix*, Fig. S1) (12, 16–19). Our results show that both lineages have undergone selective sweeps prior to resurgence that follow long periods of HA divergence; however, Victoria and Yamagata viruses showed significantly different strategies for epidemiological success. The Victoria viruses resurged through drastic changes in the HA, including nucleotide deletions, and coevolution of the neuraminidase (NA) and internal gene segments, whereas Yamagata viruses experienced stronger seasonal fluctuations that are driven by antigenic drift of the NA protein. In sum, our results show that increases in epidemic dominance by influenza B viruses is determined by a diverse set of evolutionary mechanisms, including HA nucleotide deletions, HA-NA functional balance, reassortment, epistatic interactions between segments, and antigenic drift of NA to compensate for HA stasis.

Results

Long-Term Fluctuations in Victoria and Yamagata Lineage Genetic Diversity. Estimation of changes in relative genetic diversity using a Gaussian Markov Random Fields smoothing of the Bayesian skyline plot showed that the highest levels of relative genetic diversity of the Victoria lineage occurred during the 2010–2011 and 2016–2017 seasons, with a period of decline from 2011 to 2015 (Fig. 1 *A* and *B*). This observation was consistent across the surface and internal genes (Fig. 1 *A* and *B* and *SI Appendix*, Fig. S2). We found a continuous and rapid increase in relative genetic diversity of the HA and NA of Yamagata viruses ($>10^1$) with peaks observed in late 2015 and 2018. The NA gene, and to some extent the HA, exhibited troughs in genetic diversity in mid-2013 and mid-2016, although the overall level of genetic diversity in 2016 remained higher than the diversity observed prior to 2015 (10), suggesting a greater global diversity of Yamagata viruses than during the previous decade. All other genes showed similar patterns of diversity as observed for the HA gene (Fig. 1 *C* and *D* and *SI Appendix*, Fig. S3). These patterns show that the evolutionary dynamics of the Victoria and Yamagata lineages have considerably changed, with both lineages exhibiting overall increases in global genetic diversity. These results also suggest that the recent Yamagata viruses are evolving at a faster rate than in the previous decades.

The HA gene phylogeny shows that both the Victoria and Yamagata lineages experienced a series of selective sweeps since the early to mid-2010s, resulting in single dominant clades emerging in both lineages (Fig. 1 *A* and *C* and *SI Appendix*, Fig. S4), although the timing and outcomes have been different for these emergent clades. Following a long period of diversification of the Yamagata lineage (Fig. 1*C*), clade 3 (B/Wisconsin/1/2010- and B/Phuket/3037/2013-like) viruses emerged in 2016 as a single dominant lineage that have been subject to repeated selective sweeps (see below). Clade 2 (B/Massachusetts/2/2012-like) viruses have not been detected since 2015. Similarly, for Victoria, clade 1A viruses emerged as a predominant clade since 2015 when clade 1B ceased to circulate. However, since then clade 1A viruses have undergone extensive diversification in which clades with and without multiple amino acid HA deletions emerged to cocirculate globally. While the most recently recommended Northern hemisphere vaccine strain, B/Colorado/06/2017, protects against Victoria clade 1A.1 viruses (15), owing to the periods of long divergence followed by staggered selective sweeps among the lineages, the composition of World Health Organization recommended vaccines have been updated in 7 of the last 12 y to combat epidemics of both lineages.

Victoria HA Amino Acid Deletions Coincide with Mutations in the Remaining Genes. Ancestral amino acid reconstruction of the HA of Victoria viruses shows that following the period 2011–2015, when the relative genetic diversity was the lowest, 3 amino acid substitutions (D129G, I180V, and R498K) occurred in quick succession in 1 of the subclades of clade 1A, both in the stem and antigenically dominant head domain of HA (Fig. 1*A*). However, since this period, Victoria clade 1A has diverged into multiple cocirculating clades (1A.1–1A.4). Notably, HA variants with 2-amino acid (2-aa) or 3-amino acid (3-aa) deletions have emerged independently on 4 occasions in each of the cocirculating clades and cocirculate with those that do not have a deletion. This has led to a marked increase in genetic diversity of Victoria viruses, resulting in multiple cocirculating clades that have rapidly spread worldwide (see below).

Clade 1A.1 viruses that bear a 2-aa deletion at HA 162–163 and D129G and I180V substitutions in the head of the HA emerged between June 2016 and January 2017 (mean time of most recent common ancestor [TMRCA] 2016.68, 95% highest posterior density [HPD]: 2016.44–2017.06) (Table 1). These HA changes were preceded by an R498K substitution in the HA stalk (Figs. 1 *A* and *E* and 2*A*). Interestingly, the majority of clade 1A.1 viruses also possessed 2 NA surface mutations K371Q and V401I/I401V (Figs. 1 *B* and *F* and 2*A*) that emerged during 2016 (mean TMRCA 2016.64; 95% HPD: 2016.27–2017.00), corresponding to the timing of the occurrence of HA mutations (D129G, I180V, and R498K). This suggests that compensatory mutations that maintain virus fitness may have been acquired either within or between genes during recent evolution of Victoria viruses. Noticeably, clade 1A.1 is only monophyletic in the HA and polymerase acidic (PA) genes, and the NA phylogeny shows a small number of clade 1A.1 viruses located in a different lineage that has acquired a T72K NA mutation (Fig. 1*B*).

Three separate clades of Victoria viruses harboring 3-aa deletions at positions 162–164 in the HA (clades 1A.2–1A.4) also emerged (mean TMRCA 2017.64, 2016.95, and 2016.99, respectively; see Table 1 for confidence intervals, *SI Appendix*, Fig. S5), during the same period of emergence of the 2-aa deletion variants (clade 1A.1). Clade 1A.2 viruses (denoted by green circles in Fig. 1*A*) appear more prevalent than 1A.3 and 1A.4 viruses, with viruses isolated from North America, but widely detected throughout Africa, East Asia, Europe, Middle East, Southeast Asia, and South America (*SI Appendix*, Fig. S6). Clade 1A.2 viruses possess cooccurring mutations on the HA (at K136E and G133R) both in proximity to the 162–164 deletion at the HA antigenic site and at NA A395T (Fig. 1). Clade 1A.3 viruses (denoted by magenta circles in Fig. 1*A*) have been isolated from East Asia (China), Southeast Asia (including Bangkok, Singapore, and Vietnam), and North America, whereas to date, clade 1A.4 viruses (blue circles in Fig. 1*A*) have only been detected in Laos. Clade 1A.3 viruses have mutations on the HA head (I180T and K209N, Fig. 1*A*), whereas clade 1A.4 has a single mutation (G378E) in the NA (Fig. 1*B*). Recent Victoria viruses also exhibited amino acid substitutions in the internal gene segments:

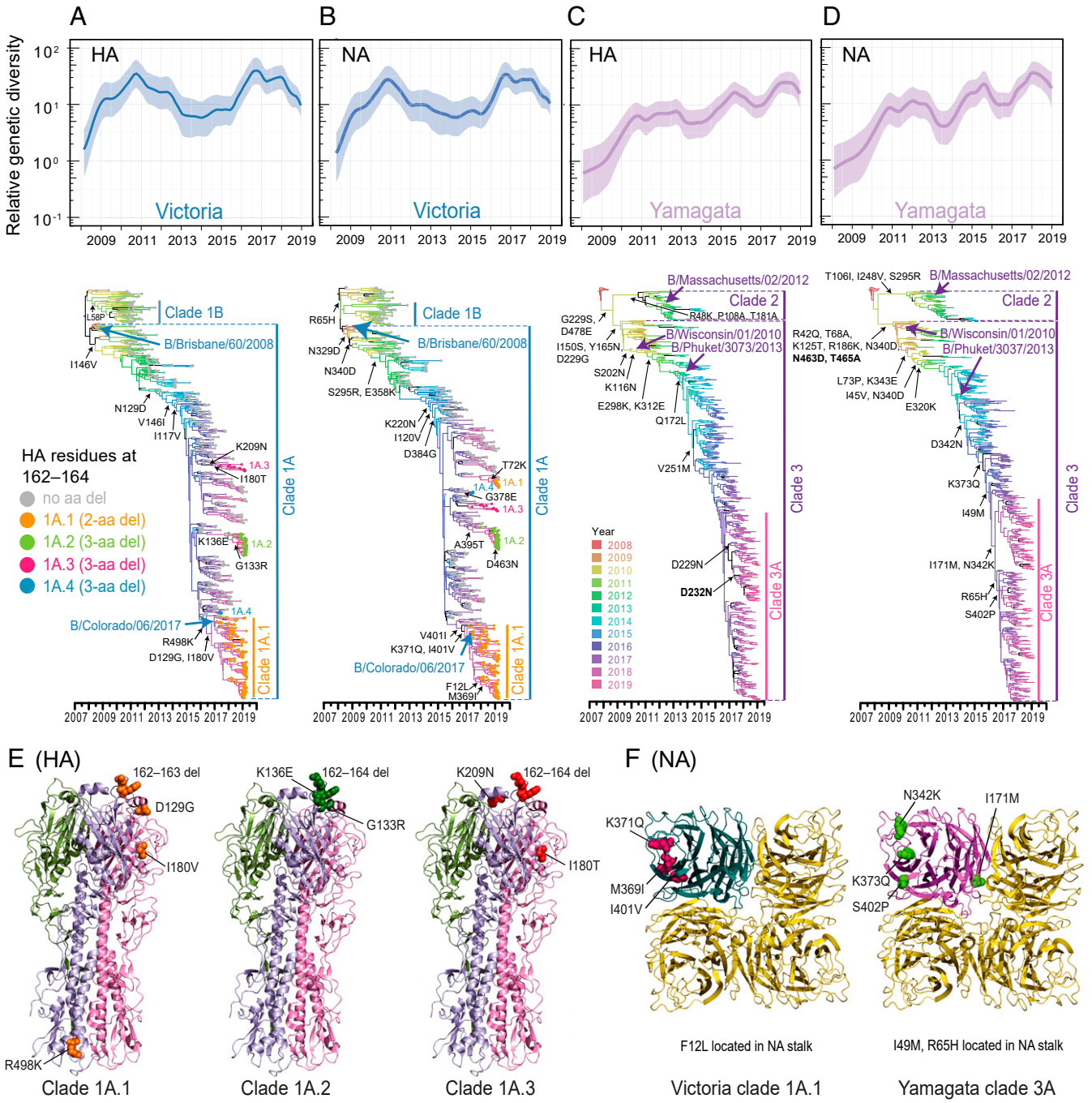

Fig. 1. Phylodynamics of global influenza B virus. Relative genetic diversities and temporal phylogenies of the HA and NA gene segments of the Victoria (*A* and *B*) and Yamagata lineages (*C* and *D*). Among the Victoria lineages, gray tip circles represent viruses in absence of HA deletions. Orange tip circles denote clade 1A.1 viruses with 2-aa deletions at positions 162–163 of the HA protein. Green, pink, and blue tip circles denote different clades (1A.2–1A.4) harboring a 3-aa deletion at positions 162–164 of the HA protein. For each gene phylogeny, trunk amino acid substitutions are mapped at the backbone of each tree. Amino acid substitutions in bold indicate potential N-linked glycosylation sites. (*E*) Three-dimensional structure of the trimeric HA (PDB ID code: 4FQM) and (*F*) tetrameric NA (PDB ID code: 3K36) proteins of B/Brisbane/60/2008 (Victoria lineage) and B/Perth/211/2001(Yamagata lineage) as reference, respectively. Trunk amino acid substitutions that occurred in the Victoria clade 1A.1–1A.3 and the Yamagata clade 3A viruses are mapped.

PB2 (T103A), PB1 (V208I), PA (I686V), and nonstructural protein (NS) (L241F) gene segments (*SI Appendix*, Figs. S7–S14). These results show that Victoria viruses possess a variety of deletion variants with additional HA mutations mostly occurring on the 120- and 160-loop regions of the HA globular head, but also in the NA and internal segments, demonstrating alternative genomic targets for optimizing virus fitness.

Resurgence of Yamagata Virus Due to HA and NA Antigenic Mutations. Between 2010 and 2014, a number of amino acid mutations accumulated sequentially on the backbone of the Yamagata clade 3 virus HA phylogeny (Fig. 1*C*). These HA mutations occurred at the antigenic sites (I150S, Y165N, S202N, and K116N) and surrounding regions (G229S, D229G, E298K, Q172L). Similarly, from 2008 to 2013, 10 sequential amino acid substitutions were

Table 1. TMRCA estimates for the HA and NA gene segments among Victoria clade 1A HA variants harboring 2-aa or 3-aa deletions

Segment	Clade	Mean TMRCA	Lower 95% HPD	Upper 95% HPD	PP
HA	1A.1 (2-aa del)	2016.68 (06 Sep 2016)	2016.44 (11 Jun 2016)	2017.06 (23 Jan 2017)	1.00
	1A.2 (3-aa del)	2017.64 (23 Aug 2017)	2017.40 (27 May 2017)	2018.01 (05 Jan 2018)	1.00
	1A.3 (3-aa del)	2016.95 (14 Dec 2016)	2016.75 (02 Oct 2016)	2017.29 (17 Apr 2017)	0.99
	1A.4 (3-aa del)	2016.99 (28 Dec 2016)	2016.74 (28 Sep 2016)	2017.31 (24 Apr 2017)	1.00
NA	1A.1 (2-aa del)	2016.64 (23 Aug 2016)	2016.27 (10 Apr 2016)	2017.00 (01 Jan 2017)	<0.95
	1A.2 (3-aa del)	2017.45 (14 Jun 2017)	2016.99 (28 Dec 2016)	2017.81 (24 Oct 2017)	0.97
	1A.3 (3-aa del)	2017.07 (27 Jan 2017)	2016.84 (05 Nov 2016)	2017.23 (26 Mar 2017)	1.00
	1A.4 (3-aa del)	2017.18 (08 Mar 2017)	2016.98 (25 Dec 2016)	2017.36 (12 May 2017)	1.00

PP, posterior probability.

observed in the NA (Fig. 1*D*). These HA and NA mutations in clade 3 viruses correspond with previous observations (11). However, since 2015, only a single HA mutation (V251M) was observed at the trunk of the Yamagata HA (Fig. 1*C*), whereas in contrast, the NA has accumulated at least 7 significant amino acid substitutions (Fig. 1*D*), indicating that immune selection on the HA has played a limited role in the recent evolution of Yamagata viruses. The first of these NA mutations (D342N) is estimated to have emerged in April 2014 (mean TMRCA: 2014.31; 95% HPD: 2013.92–2014.63), ∼1 y before the global epidemics in late 2014 to mid-2015. Five of these substitutions occurred at the surface of the NA protein, whereas 2 substitutions (I49M and R65H) are located in the NA stalk (Fig. 1*F*).

Yamagata viruses circulating in 2017–2019 formed a monophyletic lineage (clade 3A in Fig. 1 and *SI Appendix*, Fig. S15) in all segments. This clade is characterized by mutations in the NA at positions I171M and N342K and mutations in the PB1 (S591A) and PA (V326M) proteins (Fig. 2*B* and *SI Appendix*, Figs. S15–S23), while no trunk amino acid substitutions were observed in the HA, PB2, nucleoprotein (NP), matrix protein (MP), and NS. More recently, a subgroup of clade 3A viruses have acquired 2 additional HA mutations at D229N and D232N, with the latter a potential N-glycosylation site. The mean TMRCAs of NA, PB1, and PA genes of Yamagata clade 3A defining mutations were estimated around May–June 2016 (mean TMRCA 2016.44 [95% HPD, 2016.22–2016.64], 2016.46 [2016.24–2016.67], and 2016.33 [2016.15–2016.63], respectively) (Table 2 and *SI Appendix*, Figs. S15–S23), suggesting these mutations may have contributed to the severe epidemics observed in 2018 (16–19).

Patterns of Reassortment of Influenza B Viruses. We observed long-term changes in reassortment patterns, with frequent interlineage reassortment between Victoria and Yamagata viruses prior to 2016, wherein the NA and NP genes of some Victoria isolates were acquired from Yamagata viruses, but interlineage reassortment has not been detected since (*SI Appendix*, Fig. S24). This suggests that interlineage reassortments has led to the selection of 2 distinct Victoria and Yamagata genomes that are likely fit and presently outcompete emerging novel interlineage reassortants. The recent lack of reassortment between the Victoria and Yamagata lineages is consistent with observations in seasonal influenza A viruses, where emergence of between-subtype reassortants (e.g., between A/H3N2 and A/H1N1pdm09) is rare.

During periods of greater within-lineage genetic diversity (i.e., during periods when multiple clades of Victoria or Yamagata viruses cocirculated), we observed a greater rate of interclade reassortment. For example, Victoria viruses isolated during 2016–2019 have undergone reassortment at a greater frequency than in other years, when multiple clades with amino acid deletions emerged and cocirculated (Fig. 3). Victoria clade 1A.1–4 viruses are monophyletic in the HA phylogeny (Fig. 1*A*); however, phylogenies of the NA and internal genes indicate that there has been a degree of interclade reassortment (Figs. 1*B* and 3 and *SI Appendix*, Fig. S7). Clade 1A.1 viruses (shown by orange circles in Fig. 1*A*) are monophyletic in the PA, HA, and NP phylogenies but segregate into different lineages in the PB2, PB1, NA, MP, and NS trees with segments repeatedly deriving from viruses without an HA deletion; clade 1A.2 viruses (green circles) are monophyletic in the PB2, PB1, HA, and NA phylogenies but are non-monophyletic in PA, NP, MP, and NS trees; clade 1A.3 clade (magenta circles) are monophyletic in most gene segments except the PA and NS genes; and clade 1A.4 viruses (shown by blue circles) are monophyletic in most gene segments except the NS.

In comparison, among Yamagata viruses (*SI Appendix*, Fig. S25), older viruses appeared to have undergone fewer interclade reassortment events than viruses circulating from 2015 onwards. Taken together, comparison of the evolutionary histories of each gene segment showed variation in reassortment patterns within both lineages, with greater rates observed during periods of greater genetic diversity, showing a correlation between reassortment and relative genetic diversity of influenza B viruses.

Selection Pressure of Victoria and Yamagata Lineages. The mean nucleotide substitution rates (evolutionary rate) was 1.93×10^{-3} (95% HPD $1.76–2.1 \times 10^{-3}$) and 2.41×10^{-3} ($2.18–2.66 \times 10^{-3}$) substitutions per site/year, respectively, for the HA and NA genes of Victoria viruses (Table 3), comparable with previous estimates (10). By comparison, Yamagata viruses exhibited significantly higher evolutionary rates for the surface genes (HA 2.75 [2.51–2.99] $\times 10^{-3}$; NA 2.82 [2.54–3.11] $\times 10^{-3}$ substitutions per site/year). The estimates for the Yamagata lineage are significantly higher than those estimated using a dataset spanning 2002–2012 (10), suggesting that recent Yamagata viruses were evolving at a greater rate than observed previously.

The NA and NS genes of both the Victoria and Yamagata lineages were under a greater selection pressure (mean estimated ratios of nonsynonymous to synonymous substitutions [d_N/d_S] for Victoria and Yamagata viruses: NA [d_N/d_S: 0.28 and 0.23] and NS genes [d_N/d_S: 0.34 and 0.37] than all remaining gene segments, with d_N/d_S values ranging between 0.04–0.11 [Fig. 2*C* and Table 3]). The d_N/d_S ratios in the internal versus external branches of the HA gene was higher for the Victoria lineage (ratio = 1.1) compared with the Yamagata lineage (ratio = 0.73), indicating the Victoria HA was subject to stronger immune pressure. Conversely, in the NA, the Yamagata lineage showed a higher d_N/d_S (ratio = 1.20) in the internal versus external branches compared to the Victoria lineage (ratio = 0.96). In addition, more sites are positively selected in the NA of both Victoria and Yamagata viruses than the HA gene, showing that the NA protein is under greater pressure and potentially exhibiting a greater rate of antigenic drift. These results suggest that the NA is contributing significantly to the evolution of influenza B virus.

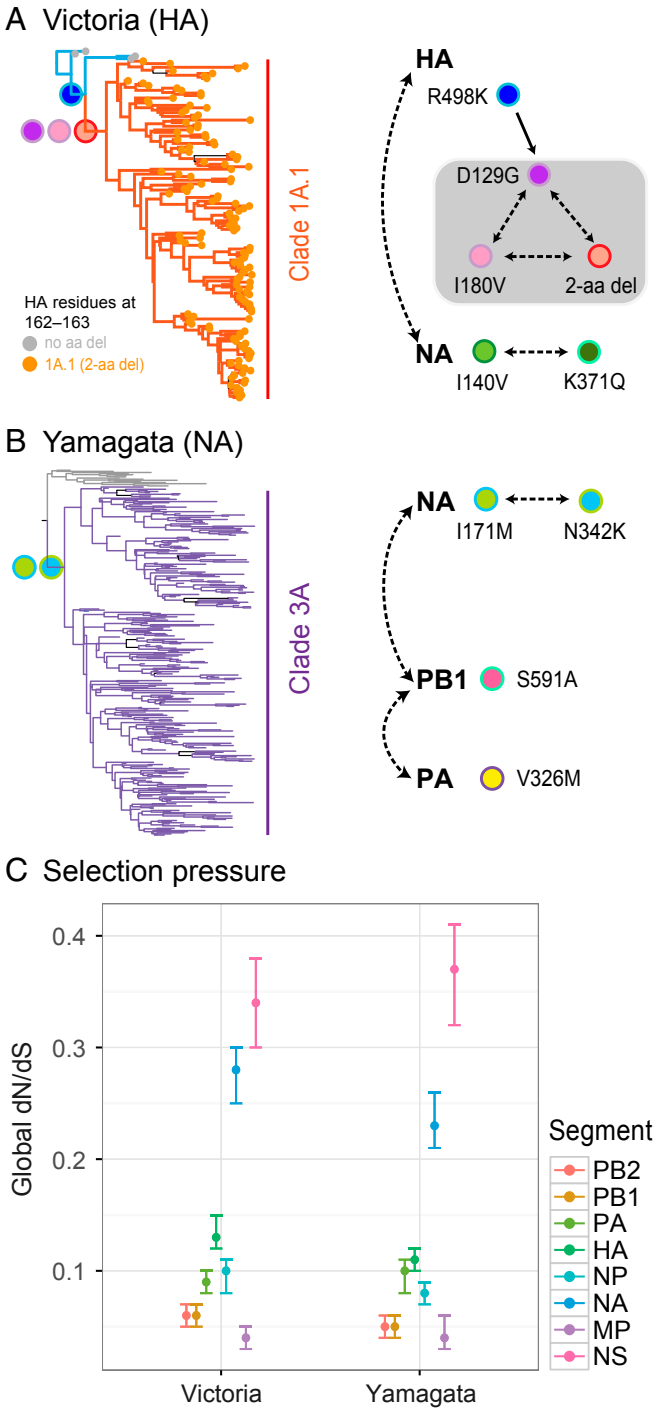

Fig. 2. Schematic diagrams showing the putative cooccurring mutations within and between gene segments, and selection pressure of influenza B viruses. (*A*) HA phylogeny of the Victoria lineage. Orange branches denote clade 1A.1 viruses with 2-aa deletions (2-aa), and blue branches denote viruses with no deletion. (*B*) NA phylogeny of the Yamagata lineage. Purple branches denote clade 3A viruses. Colored circles represent trunk amino acid mutations on gene segments. Dotted arrows indicate putative cooccurring mutations. (*C*) Estimates of global d_N/d_S ratios for each gene segment inferred using SLAC method, with the significance level at $P < 0.05$. The means of d_N/d_S ratios are represented by circle-shaped symbols with error bars indicating 95% confidence intervals. Colored bars denote different gene segments.

Viral Migration Dynamics. We reconstructed the global phylogeography of influenza B viruses using the Markov fields model (20). We found that the geographical areas occupied by trunk branches of the Victoria lineage varied through time (Fig. 4*A* and *SI Appendix*, Fig. S26 and Table S1). Multiple sources contributed to the 2011–2013 epidemics; however, East Asia constituted >50% of trunk proportions in 2013, while South America, Southeast Asia, and North America appeared to be the major source populations in subsequent years. Consistent with trunk proportions, we observed significant diffusion rates from Southeast Asia (0.70–2.21) into all other geographical areas (Fig. 4*B* and Table 4), and from North America (0.74–1.59) into most regions including East Asia, Europe, Oceania, and South America. Geographic state transition counts and migration pathways further corroborate these results (Table 4 and *SI Appendix*, Figs. S27 and S28 and Table S2). As such, this suggests that the seeding of Victoria lineage epidemics is driven by different geographic locations, where Southeast Asia and North America are major transmission sources. Notably, North America is the probable ancestral location for seeding HA deletion variants (including Victoria clade 1A.1–1A.3 viruses) (*SI Appendix*, Fig. S26).

The trunk proportions for the Yamagata lineage displayed more continuous virus circulation among multiple geographical regions through time (Fig. 4*A* and *SI Appendix*, Table S1), with no single region acting as source. Remarkably, Southeast Asia occupied a considerable amount of trunk proportions throughout 2011–2019, indicative of prolonged persistence of Yamagata viruses in human populations. Strongly supported diffusion rates have also occurred from Southeast Asia (0.70–2.72) into all other geographical regions (Fig. 4*B* and Table 4). While Oceania was less likely a significant source in any given year, interestingly the dynamics of Southeast Asia matched Oceania in the Southern hemisphere rather than a region in the Northern hemisphere. Furthermore, strong diffusion rates from Europe into most other regions are similarly observed. A greater global movement (~13 decisive pathways) was observed for Yamagata viruses compared to Victoria viruses (~8 decisive pathways) (*SI Appendix*, Fig. S28 and Table S2), consistent with the hypothesis that infecting an older age group would correlate with greater global transmission.

Global Host Age Distribution of Victoria Versus Yamagata Lineages. Demographic data from 2008 to 2019 showed that the Victoria lineage was positively skewed toward a younger population (with a median age of 13 y), whereas the Yamagata lineage displayed bimodal distribution (with a median age of 32.5 y, $P < 0.0001$) (Fig. 4*C* and *SI Appendix*, Table S3). The kernel density indicates a higher probability of Victoria virus infections under the age of 15, with peak density in 5-y-old children, whereas Yamagata viruses had a higher probability of infection between 0 and 20 y, but with a broader density distribution beyond 30 y of age. The significant difference in age distribution between the Victoria and Yamagata lineages was consistent across all geographical localities (*SI Appendix*, Fig. S29 and Table S3). Age differences between Victoria and Yamagata corroborate previous observations in region-specific studies (10, 21, 22). Based on the global HA tree, the Victoria lineages showed more sustained transmission within a given region (e.g., North America) than observed for the Yamagata lineage (*SI Appendix*, Fig. S4).

Individuals infected with the Victoria lineage 2-aa deletion viruses (Fig. 4*D*) were similar in age to those without the deletion, however, infections with the 3-aa deletion viruses (median age = 15) were found to occur in marginally older individuals than those caused by viruses lacking deletions (median age = 11, $P = 0.013$), but were not significantly different to those due to 2-aa deletion viruses. The causes of differences in age are difficult to explain but are unlikely to be due to antigenic differences of the viruses as hemagglutinin inhibition (HI) assays show that the antisera raised to a no deletion vaccine strain (B/Brisbane/60/2008) did not

Table 2. TMRCA estimates for all gene segments of Yamagata clade 3A viruses

Segment	Mean TMRCA	Lower 95% HPD	Upper 95% HPD	PP
PB2	2016.17 (04 Mar 2016)	2016.16 (01 Mar 2016)	2016.65 (26 Aug 2016)	0.95
PB1	2016.46 (18 Jun 2016)	2016.24 (30 Mar 2016)	2016.67 (03 Sep 2016)	0.97
PA	2016.33 (02 May 2016)	2016.15 (25 Feb 2016)	2016.63 (19 Aug 2016)	0.95
HA	2016.56 (25 Jul 2016)	2016.39 (23 May 2016)	2016.72 (21 Sep 2016)	0.97
NP	2016.32 (28 Apr 2016)	2016.17 (04 Mar 2016)	2016.62 (15 Aug 2016)	<0.95
NA	2016.44 (11 Jun 2016)	2016.22 (23 Mar 2016)	2016.64 (23 Aug 2016)	1.00
MP	2016.55 (21 Jul 2016)	2016.75 (02 Oct 2016)	2016.35 (09 May 2016)	<0.95
NS	2016.35 (09 May 2016)	2016.71 (17 Sep 2016)	2016.27 (10 Apr 2016)	<0.95

PP, posterior probability.

discriminate strongly between 3-aa and 2-aa deletion viruses (15). The presence of different NA substitutions between Victoria clade 1A.1–1A.4 viruses further complicates explanations for the observed age differences between deletions variants.

Discussion

Influenza B virus genomes collected over the last 2 decades reveal a dynamic pattern of evolution of seasonal influenza viruses that were not observed previously using shorter time periods of data. We identified several significant changes in the evolutionary landscapes of the 2 cocirculating influenza B lineages, Victoria and Yamagata. For the Victoria lineage, evolution was predominantly acting on the antigenically dominant HA gene, with cooccurring mutations in the NA and polymerase complex that we hypothesize are required to maintain virus fitness. The HA of recent Victoria clade 1A viruses was subject to diversifying selection, with numerous amino acid mutations being fixed in antigenic sites, and the independent emergence of 2-aa or 3-aa HA deletion variants. Substitutions in the HA and NA are mostly located in the antigenic head regions, which is indicative of directional selection in escaping the host immune response, thus providing an evolutionary mechanism for the emergence and cocirculation of multiple different Victoria virus clades since 2016.

The observation of HA deletion variants is not unprecedented in influenza B viruses; the influenza B prototype strain B/Lee/1940 possessed an amino acid deletion at the same HA position, i.e., a 1-aa deletion at 163 residue (^{160}IPK*DNN^{166}) (23). Subsequently, ancestral Yamagata viruses that circulated before 1988 also harbored double (^{160}VPK*D*N^{166}) or triple (^{160}VP***KN^{166}) deletions, while no deletions were observed in ancestral Victoria viruses (23). As such, the presence of HA deletions in recent Victoria viruses (^{160}VP**DKN^{166} or ^{160}VP***KN^{166}) is not surprising. However, as these deletions occur on the 160 antigenic loop of the HA protein, they may be crucial for antigenic properties of influenza B viruses (24). At the time of writing, the current influenza vaccine strain selected for the Victoria lineage was the B/Colorado/06/2017 strain, which was nested within clade 1A.1 with a 2-aa deletion. Antigenic characterization of Victoria viruses has shown that some 3-aa deletion strains reacted poorly with antisera raised against the Victoria clade 1A.1 vaccine strain (with a 2-aa deletion), but generally reacted better to sera raised against viruses with no deletion, suggesting that Victoria viruses are undergoing antigenic diversification (15, 25). This is reflected in the subsequent 2020 vaccine composition recommendation for the Southern hemisphere where a 3-aa deletion strain in clade 1A.2 (B/Washington/02/2019-like) was selected to replace the previous 2-aa clade 1.A1 (B/Colorado/06/2017-like) vaccine strain (15). With continued cocirculation of these HA deletion variants, this reduced cross-reactivity could potentially lead to the divergence of Victoria viruses into 2 antigenically distinct variants. The circulation of 3 or more distinct influenza B viruses would drastically complicate influenza vaccine formulation.

We found that the NA of Yamagata viruses was evolving at a greater rate, with few mutations occurring in the HA, and cooccurring substitutions in the polymerase genes (PB2-PB1-PA). We also observed a marked change in population dynamics of Yamagata viruses that, from 2002 to 2013, were previously subject to weaker natural selection pressure with relatively low and constant

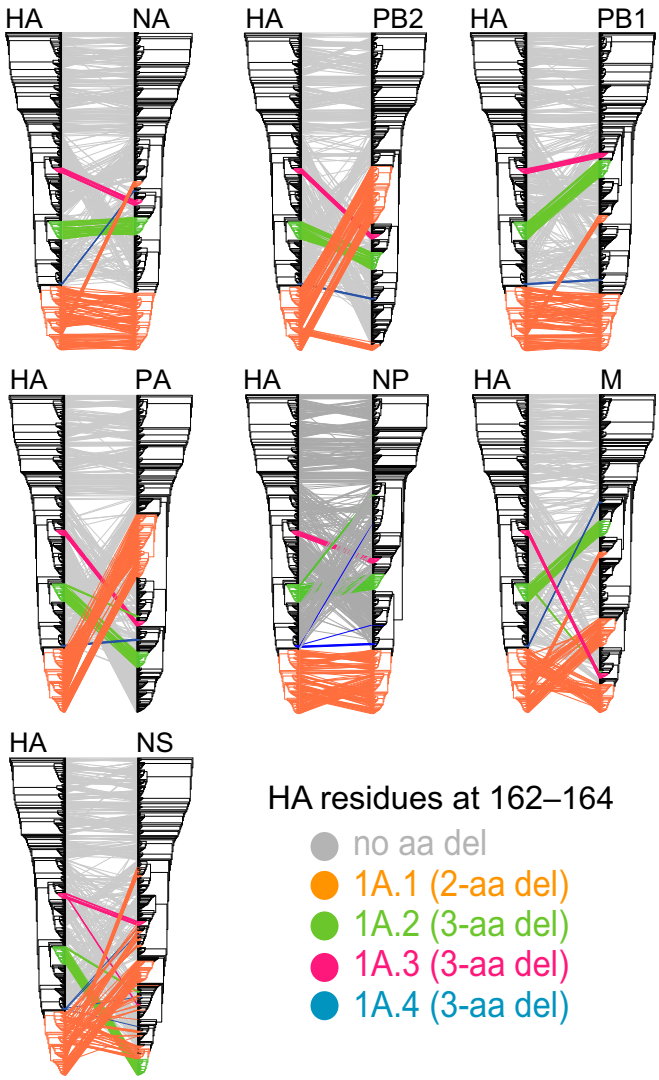

Fig. 3. Evolutionary relationships of each gene segment of the Victoria lineage, displaying interclade reassortment within the lineage, highlighting the deletion variants. Reassortment events of Yamagata viruses are shown in *SI Appendix*, Fig. S25.

Table 3. Nucleotide substitution rates and selection pressures of globally sampled Victoria and Yamagata viruses in 2008–2019

Segment*	Mean substitution rates (95% HPD)	Global d_N/d_S (95% CI)	Branch d_N/d_S			No. of positive sites (amino acid position)
			Internal	External	Internal/External	
Victoria						
PB2	1.79 (1.63–1.96)	0.06 (0.05–0.07)	0.004	0.018	0.22	1 (630)
PB1	1.77 (1.61–1.94)	0.06 (0.05–0.07)	0.025	0.037	0.68	1 (655)
PA	1.81 (1.66–1.97)	0.09 (0.08–0.10)	0.034	0.045	0.76	1 (619)
HA	1.93 (1.76–2.13)	0.13 (0.12–0.15)	0.11	0.10	1.10	0
NP	1.76 (1.58–1.95)	0.10 (0.08–0.11)	0.05	0.06	0.83	4 (242, 340, 357, 453)
NA	2.41 (2.18–2.66)	0.28 (0.25–0.30)	0.23	0.24	0.96	5 (73, 106, 320, 395, 455)
MP	2.16 (1.83–2.53)	0.04 (0.03–0.05)	0.01	0.02	0.50	0
NS	2.28 (1.96–2.65)	0.34 (0.30–0.38)	0.16	0.25	0.64	2 (205, 253)
Yamagata						
PB2	2.03 (1.86–2.21)	0.05 (0.04–0.06)	0.018	0.017	1.06	1 (479)
PB1	2.17 (1.98–2.35)	0.05 (0.04–0.06)	0.02	0.03	0.67	0
PA	2.40 (2.21–2.60)	0.10 (0.08–0.11)	0.035	0.059	0.59	1 (573)
HA	2.75 (2.51–2.99)	0.11 (0.10–0.12)	0.046	0.063	0.73	2 (229, 234)
NP	2.37 (2.13–2.64)	0.08 (0.07–0.09)	0.038	0.052	0.73	1 (451)
NA	2.82 (2.54–3.11)	0.23 (0.21–0.26)	0.26	0.22	1.20	4 (271, 358, 395, 436)
MP	2.39 (2.03–2.78)	0.04 (0.03–0.06)	0.0048	0.012	0.40	0
NS	2.85 (2.51–3.24)	0.37 (0.32–0.41)	0.21	0.28	0.75	2 (124, 200)

CI, confidence interval; d_N/d_S, the ratio of nonsynonymous to synonymous substitution rates.
*Analyses were restricted to M1 and NS1 coding regions for the MP and NS segments, respectively.

levels of genetic diversity (10). However, from 2015 onwards, we observed more frequent seasonal oscillations in genetic diversity that coincide with global epidemics of Yamagata viruses (16–19). The influenza vaccine strain for the Yamagata lineage (B/Phuket/3073/2013-like) has remain unchanged since 2015, and antigenic characterization of recent clade 3 viruses shows good cross-reactivity with the vaccine strain (25), consistent with the lack of amino acid substitutions in the HA of these viruses. Our results suggest that the NA of Yamagata viruses is under stronger immune pressure than the HA, and that NA antigenic drift may be responsible for the recent increase in epidemic activity.

Vaccine effectiveness against Yamagata viruses for the 2017–2018 season was lower than that for Victoria viruses in the United States (48% versus 76%) (26), indicating NA as a target for future vaccine development. However, less attention is paid to the antigenicity of the NA, and it is not known if human anti-NA antibodies arising from novel Yamagata strains broadly cross-react with older viruses. Increasing awareness of the importance of influenza virus NA immunity has prompted recent efforts to understand the protective effect of anti-NA immunity in animal models; studies have shown that the anti-NA antibodies can significantly reduce virus titers and shedding of influenza B viruses (27) and lower the duration of illness in influenza A/H1N1 virus (28). Crucially, anti-NA antibodies can broadly cross-react against H1N1 and H3N2 subtype viruses, but current influenza vaccines cannot induce an efficient NA response (29). Furthermore, characterization of NA antigenicity has shown that antigenic drift also occurs at sites on the lateral surface of the NA that did not inhibit NA activity (30).

Using phylodynamic analyses we have revealed key molecular processes affecting the evolution of influenza B virus that have contributed to immune escape and the resulting strong epidemic waves observed in recent years. Future research into immune responses to the NA and the effects of epistatic gene interactions are needed to understand their effect on the antigenic and genetic evolution and global circulation of influenza B virus. Addressing these aspects can potentially aid modeling to predict future circulating lineages, assist in forecasting the timescale of epidemic outbreaks, and also inform the development of universal influenza vaccines that are the cornerstone of current research efforts aimed at the prevention and control of influenza.

Materials and Methods

Ethics Statement. This study was approved by the ethics committees at the National University of Singapore (NUS)-Institutional Review Board (12–439E) and Domain Specific Review Board (2016/01279). The clinical samples were deidentified before transporting to the research laboratory at Duke-NUS.

Virus Samples and Sequencing. Viral RNA was extracted from 2,147 human influenza-positive nasopharyngeal swab samples or cultured isolates collected from 2 hospitals (Singapore General Hospital and National University Hospital) and National Public Health Laboratory during 2011–2019, using the QIAamp Viral RNA Mini Kit (Qiagen) and subtyping was performed by real-time RT-PCR. Influenza A virus was detected in 1,493 samples and influenza B in 654 samples. A subset of influenza B samples was selected for full genome sequencing using either Sanger or next generation sequencing (NGS) methods. For Sanger sequencing, the viruses were passaged in Madin-Darby Canine Kidney SIAT–1 cells and sequenced using previously published primers (31). NGS sequencing was performed on libraries of cDNA derived from viral RNA extracted directly from the swabs and followed by RT-PCR amplification of whole genome of influenza B (32, 33). The PCR products were quantified using the Qubit dsDNA HS Assay kits (Thermo Fisher Scientific) and subsequently diluted to 1 ng/μL for library preparation using the Nextera XT DNA library preparation kits (Illumina). The libraries were pooled and run on an Illumina MiSeq sequencing platform with a read length of 2 × 250 bp at the Duke-NUS Genome Biology Facility. The NGS reads were quality checked with FastQC as implemented in UGENE v1.28 (34), and the adaptors were removed in Trimmomatic v0.36 (35). Short reads were then mapped to a known reference genome using the BWA algorithm in UGENE, and consensus sequences were obtained. A total of 158 influenza B genomes (76 Victoria and 82 Yamagata) were generated.

Evolutionary Genomic Analysis of Influenza B Viruses. We downloaded >12,000 global influenza B genomes (as of May 30, 2019) from National Center for Biotechnology Information GenBank and Global Initiative on Sharing All Influenza Data databases from 2008 to 2019. To investigate the evolutionary dynamics of recent influenza B outbreaks, we first reconstructed the global evolution using the HA gene (including new sequences from Singapore) in FastTree (36). This tree was then used to classify viruses into Victoria and Yamagata lineages and prepare downsampled genome datasets of 2,078 viruses. Our final gene datasets consisted of an even sampling of Victoria (n = 1,037) and Yamagata (n = 1,041) lineages, with sample numbers varying in each year depending on data availability. Maximum likelihood phylogenies of individual gene segments (PB2, PB1, PA, HA, NP, NA, MP, and NS) were reconstructed using RAxML v8.0.14 (37) with the generalized time-reversible nucleotide substitution model and gamma rate heterogeneity.

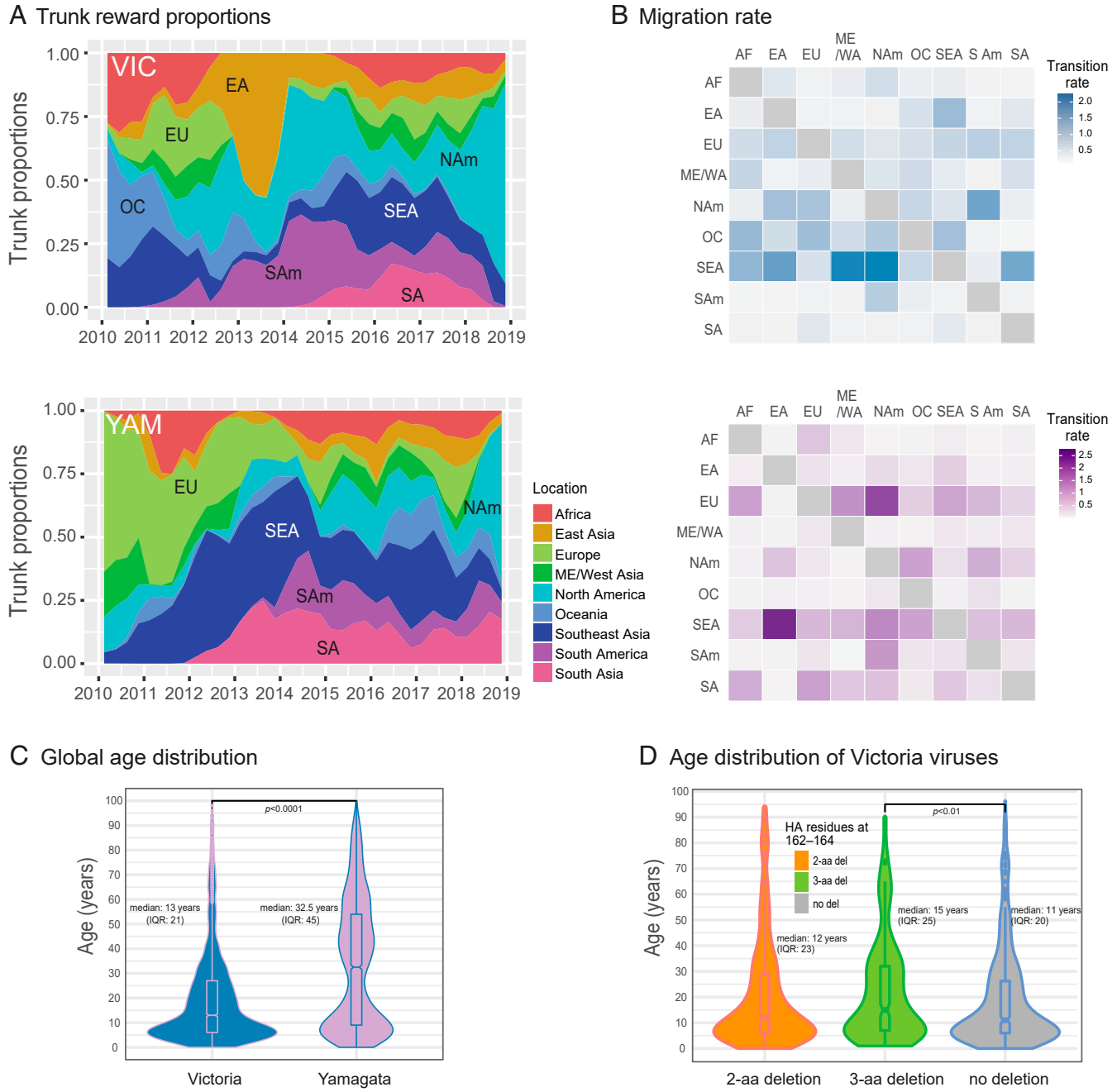

Fig. 4. Phylogeography and age distribution of influenza B viruses. (*A*) Proportion of ancestral geographical location states estimated on the phylogenetic trunk of Victoria (VIC) and Yamagata (YAM) viruses through time. Shaded areas represent estimated ancestral location state proportions in each location. (*B*) Patterns of migration rates are shown on a 2D matrix where each cell corresponds to mean migration rates between pairwise geographic locations. Blue and purple cells denote Victoria and Yamagata, respectively. Abbreviations for location states: Africa (AF), East Asia (EA), Europe (EU), Middle East/West Asia (ME/WA), North America (NAm), Oceania (OC), South America (SAm), South Asia (SA), and Southeast Asia (SEA). (*C*) Age density distribution of influenza B virus by lineage. The violin plot indicates the distribution, and the box-plot indicates the interquartile range (IQR). The median ages are represented by a notched horizontal line. The vertical line indicates the 1.5X IQR, and the outliers are represented by dots. The Mann–Whitney *U* test was used to compare the median ages, with statistical significance indicated by $P < 0.0001$. (*D*) Age density distribution of Victoria clade 1A viruses with multiple amino acid deletions on the HA protein. Orange plot represents viruses with 2-aa deletion (clade 1A.1), green plot indicates viruses with 3-aa deletion (clade 1A.2–1A.4), and gray plot indicates viruses with no deletion. Mann–Whitney *U* test *P* values indicated by $P < 0.05$.

Population dynamics were estimated for Victoria (685 sequences) and Yamagata (687 sequences) viruses using BEAST v1.10.4 (38), after checking for sufficient temporal signal using TempEst (39). Molecular clock and changes in relative genetic diversity were estimated using the uncorrelated lognormal relaxed-clock with Gaussian Markov random field smoothing of the effective population size (40). For all analyses, the SRD06 codon position model (41) for codon partitioning of the HKY85+Γ substitution model was used. For each dataset, at least 4 independent runs of 100 million generations, sampling every 10,000 generations, were conducted. The runs were checked for convergence in Tracer v1.7 (42) and combined after removing appropriate burn-in values using LogCombiner. The resulting maximum clade credibility trees were generated using TreeAnnontator. For each gene segment, ancestral trunk substitutions were reconstructed for Victoria and Yamagata lineages using the treesub program (43) as previously described (44) and mapped to maximum likelihood phylogenies.

Table 4. Asymmetric migration rates between location states inferred using the BSSVS model for the Victoria and Yamagata lineages

	Migration rates*								
	Africa	East Asia	Europe	Middle East/West Asia	North America	Oceania	Southeast Asia	South America	South Asia
Victoria									
Africa	—	**0.44**	0.28	0.33	**0.60**	0.31	0.37	0.23	0.21
East Asia	0.34	—	0.27	0.29	0.26	**0.54**	**1.20**	0.28	0.43
Europe	0.63	0.80	—	0.60	0.47	**0.59**	0.76	**0.85**	**0.80**
Middle East/West Asia	**0.75**	0.27	0.36	—	0.49	0.65	0.38	0.26	0.57
North America	0.32	**1.09**	**1.04**	**0.32**	—	**0.74**	0.51	**1.59**	**0.32**
Oceania	**1.24**	**0.62**	**1.15**	0.63	**0.93**	—	**1.13**	0.27	0.20
Southeast Asia	**1.29**	**1.71**	0.53	**2.13**	**2.21**	**0.70**	—	0.36	**1.56**
South America	0.22	0.21	0.29	0.25	**0.90**	0.32	0.29	—	0.29
South Asia	0.26	0.25	**0.43**	0.28	0.26	0.29	0.41	0.26	—
Yamagata									
Africa	—	0.21	**0.75**	0.36	0.20	0.22	0.29	0.20	0.25
East Asia	0.28	—	0.41	0.27	0.51	0.35	0.66	0.26	0.27
Europe	**1.21**	0.22	—	**1.37**	**2.27**	**0.67**	**1.15**	**0.92**	0.58
Middle East/West Asia	0.25	0.22	0.35	—	0.26	0.24	0.33	0.31	0.37
North America	0.25	**0.77**	0.34	0.23	—	**1.20**	0.44	**1.06**	**0.61**
Oceania	0.21	0.25	0.26	0.23	0.35	—	0.26	**0.44**	0.22
Southeast Asia	**0.70**	**2.72**	**0.91**	**0.96**	**1.47**	**1.20**	—	**0.82**	**0.91**
South America	0.24	0.22	0.38	0.18	**1.30**	0.38	0.41	—	0.39
South Asia	**1.07**	0.20	**1.21**	**0.74**	**0.80**	0.32	**0.72**	**0.45**	—

*Migration rates in bold indicate supported rates with Bayes factor ≥ 3.

Reassortment Analysis. Based on the global HA phylogeny of influenza B lineages, the Victoria and Yamagata split into 2 lineages and their branches were colored separately. Interlineage reassortment of gene segments between Victoria and Yamagata lineage viruses were then examined based on the HA branch coloring. To investigate the extent of interclade reassortment within Victoria and Yamagata lineages, pairwise incongruence between the HA gene and individual gene segments (PB2, PB1, PA, NP, NA, M, and NS) were constructed using the dendextend R package. The twines were colored based on the HA phylogenies. The program GiRaF (45) was used to identify the reassortment events sampling from 1,000 unrooted trees generated from MrBayes, using a burn-in of 500 and sampling every 200 iterations for 200,000 generations. Five independent runs were performed, and reassortant events with confidence levels >0.9 were considered.

Selection Analysis. Selection pressures for each gene segment were estimated for Victoria and Yamagata lineages separately, using globally representative datasets during 2008–2019 (n = 500). The ratio of the number of nonsynonymous to synonymous (d_N/d_S) mutations was estimated using the Single-Likelihood Ancestor Counting (SLAC) method (46) using DataMonkey (47). Branch-wise d_N/d_S ratios of the internal vs. external branches were estimated using the 2-model ratio in CODEML (48). The Mixed Effects Model of Evolution (49) in DataMonkey was used to identify codon sites undergoing positive selection, with the significance level of $P < 0.05$. In addition, potential N-glycosylation sites (N-X-S/T sequons, where X is not P) on the HA and NA proteins of the Victoria and Yamagata viruses were predicted using the NetNGlyc 1.0 Server (50) at a threshold of >0.5.

Structural Mapping of HA and NA Proteins. To visualize the positions of trunk substitutions on the virus surface proteins, amino acid residues were mapped onto the 3-dimensional structure of a Victoria virus HA protein (Protein Data Bank [PDB] ID code 4FQM) and a Yamagata virus NA (PDB ID code 3K36) using the PyMOL program (Molecular Graphics System, version 2.3 Schrödinger, LLC).

Phylogeographic Analyses of Influenza B Viruses. To infer the spatial-temporal dynamics of each influenza B lineage, a Bayesian Stochastic Search Variable Selection (BSSVS) model (20) was applied in BEAST to estimate the asymmetric diffusion rates between localities coded as discrete geographical locations: Africa, East Asia, Europe, Middle East/West Asia, North America, Oceania, Southeast Asia, South America, and South Asia. The uncorrelated lognormal relaxed molecular clock, the SRD06 codon position model, and exponential coalescent tree prior was used for Markov chain Monte Carlo runs. At least 4 independent runs using 100 million generations were used, with sampling every 10,000 generations. Convergence was assessed, and the runs were combined after removing appropriate burn-in using LogCombiner. The estimated viral migration rates between locations were summarized using the resulting log files. Significant diffusion rates were identified using Bayes Factors (BF): ≥1,000 was deemed as decisive support, 100 ≤ BF < 1000 as very strong support, 10 ≤ BF < 100 as strong support and 3 ≤ BF < 10 as supported. We also estimated the number of location state transitions (measure for the gene flow between localities) from Markov jump counts (51), and the total number of state counts in and out of each location were used to generate the plot in Fig. 4. Moreover, the trunk of phylogenies reconstructed from time-stamped HA sequences represents persistent lineages between epidemics (52, 53). The duration the ancestral population occupies a particular location was estimated from the Markov rewards (51), estimating the contribution of each location to the persistence of calculated trunk proportions for each discrete location. Occupation of the trunk by a single location would suggest a dominant source-sink population dynamic while multiple locations constitute metapopulation and rapid global migration.

Global Host Age Distribution. We downloaded the global sequences of influenza B viruses (Victoria: 5,302 and Yamagata: 7,360) from 2008 to 2019 to retrieve corresponding host age data. We compared the age distribution of hosts infected with Victoria and Yamagata viruses globally and across different regions. Within the Victoria clade 1A lineage, we compared the age distribution of hosts infected by Victoria viruses exhibiting different amino acid deletion mutations. Nonparametric comparisons of age distributions between lineages and Victoria HA deletion mutants were performed using the Mann–Whitney U test. A Kruskal–Wallis test was used to examine if geographic regions differed in median age for each of Victoria and Yamagata lineages. The violin distribution plots were generated and statistical analyses performed using the *ggplot2* and *ggpubr* in R statistics packages.

Data Availability. The sequences generated in this study have been deposited in GenBank database under accession numbers MN588323 to MN589586 (*SI Appendix*, Table S4).

ACKNOWLEDGMENTS. We thank Zawiah Binte Taha and Xavier Teng for technical support. This study was supported by the Duke-NUS Signature Research Programme funded by the Ministry of Health, Singapore, by Ministry of Health Grant MOH/CDPHRG/0012/2014, National Medical Research Council Grant NMRC/OFIRG/0008/2016, and Contract HHSN272201400006C from the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services.

1. A. D. Iuliano *et al.*; Global Seasonal Influenza-associated Mortality Collaborator Network, Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *Lancet* **391**, 1285–1300 (2018).
2. K. E. Lafond *et al.*; Global Respiratory Hospitalizations—Influenza Proportion Positive (GRIPP) Working Group, Global role and burden of influenza in pediatric respiratory hospitalizations, 1982-2012: A systematic analysis. *PLoS Med.* **13**, e1001977 (2016).
3. A. Mosnier *et al.*; GROG network, Clinical characteristics are similar across type A and B influenza virus infections. *PLoS One* **10**, e0136186 (2015).
4. M. W. Shaw *et al.*, Reappearance and global spread of variants of influenza B/Victoria/2/87 lineage viruses in the 2000-2001 and 2001-2002 seasons. *Virology* **303**, 1–8 (2002).
5. V. Tisa *et al.*, Quadrivalent influenza vaccine: A new opportunity to reduce the influenza burden. *J. Prev. Med. Hyg.* **57**, E28–E33 (2016).
6. T. Heikkinen, N. Ikonen, T. Ziegler, Impact of influenza B lineage-level mismatch between trivalent seasonal influenza vaccines and circulating viruses, 1999-2012. *Clin. Infect. Dis.* **59**, 1519–1524 (2014).
7. W. Sun *et al.*, Development of influenza B universal vaccine candidates using the "Mosaic" hemagglutinin approach. *J. Virol.* **93**, e00333-19 (2019).
8. T. Bedford *et al.*, Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217–220 (2015).
9. G. Dudas, T. Bedford, S. Lycett, A. Rambaut, Reassortment between influenza B lineages and the emergence of a coadapted PB1-PB2-HA gene complex. *Mol. Biol. Evol.* **32**, 162–172 (2015).
10. D. Vijaykrishna *et al.*, The contrasting phylodynamics of human influenza B viruses. *eLife* **4**, e05055 (2015).
11. P. Langat *et al.*, Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS Pathog.* **13**, e1006749 (2017).
12. I. G. Barr, D. Vijaykrishna, S. G. Sullivan, Differential age susceptibility to influenza B/Victoria lineage viruses in the 2015 Australian influenza season. *Euro Surveill.* **21**, 30118 (2016).
13. R. Chen, E. C. Holmes, The evolutionary dynamics of human influenza B virus. *J. Mol. Evol.* **66**, 655–663 (2008).
14. J. I. Kim *et al.*, Reassortment compatibility between PB1, PB2, and HA genes of the two influenza B virus lineages in mammalian cells. *Sci. Rep.* **6**, 27480 (2016).
15. World Health Organization, WHO recommendations on the composition of influenza virus vaccines. https://www.who.int/influenza/vaccines/virus/recommendations/en/. Accessed 9 September 2019.
16. C. Adlhoch, R. Snacken, A. Melidou, S. Ionescu, P. Penttinen; The European Influenza Surveillance Network, Dominant influenza A(H3N2) and B/Yamagata virus circulation in EU/EEA, 2016/17 and 2017/18 seasons, respectively. *Euro Surveill.* **23**, 18-00146 (2018).
17. R. Garten *et al.*, Update: Influenza activity in the United States during the 2017-18 season and composition of the 2018-19 influenza vaccine. *MMWR Morb. Mortal. Wkly. Rep.* **67**, 634–642 (2018).
18. D. M. Skowronski *et al.*, Early season co-circulation of influenza A(H3N2) and B(Yamagata): Interim estimates of 2017/18 vaccine effectiveness, Canada, January 2018. *Euro Surveill.* **23**, 18-00035 (2018).
19. N. S. Korsun *et al.*, Predominance of influenza B/Yamagata lineage viruses in Bulgaria during the 2017/2018 season. *Epidemiol. Infect.* **147**, e76 (2019).
20. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
21. Y. Tan *et al.*, Differing epidemiological dynamics of influenza B virus lineages in Guangzhou, southern China, 2009-2010. *J. Virol.* **87**, 12447–12456 (2013).
22. D. M. Skowronski *et al.*, Age-related differences in influenza B infection by lineage in a community-based sentinel system, 2010-2011 to 2015-2016, Canada. *J. Infect. Dis.* **216**, 697–702 (2017).
23. R. Nerome *et al.*, Evolutionary characteristics of influenza B virus since its first isolation in 1940: Dynamic circulation of deletion and insertion mechanism. *Arch. Virol.* **143**, 1569–1583 (1998).
24. N. Nakagawa, R. Kubota, T. Nakagawa, Y. Okuno, Antigenic variants with amino acid deletions clarify a neutralizing epitope specific for influenza B virus Victoria group strains. *J. Gen. Virol.* **82**, 2169–2172 (2001).
25. European Centre for Disease Prevention and Control, Influenza virus characterization, summary Europe, September 2018. (ECDC, Stockholm). https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/influenza-virus-characterisation. Accessed 9 September 2019.
26. Centers for Disease Control and Prevention, Seasonal influenza vaccine effectiveness, 2017-2018. https://www.cdc.gov/flu/vaccines-work/2017-2018.html. Accessed 9 September 2019.
27. M. McMahon *et al.*, Mucosal immunity against neuraminidase prevents influenza B virus transmission in Guinea Pigs. *MBio* **10**, e00560-19 (2019).
28. H. E. Maier *et al.*, Pre-existing anti-neuraminidase antibodies are associated with shortened duration of influenza A (H1N1)pdm virus shedding and illness in naturally infected adults. *Clin. Infect. Dis.*, ciz639 (2019).
29. Y. Q. Chen *et al.*, Influenza infection in humans induces broadly cross-reactive and protective neuraminidase-reactive antibodies. *Cell* **173**, 417–429.e10 (2018).
30. A. Yasuhara *et al.*, Antigenic drift originating from changes to the lateral surface of the neuraminidase head of influenza A virus. *Nat. Microbiol.* **4**, 1024–1034 (2019).
31. E. Hoffmann *et al.*, Rescue of influenza B virus from eight plasmids. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11411–11416 (2002).
32. B. Zhou *et al.*, Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics. *J. Clin. Microbiol.* **52**, 1330–1337 (2014).
33. H. K. Lee, C. K. Lee, J. W. Tang, T. P. Loh, E. S. Koay, Contamination-controlled high-throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer. *Sci. Rep.* **6**, 33318 (2016).
34. O. Golosova *et al.*, Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. *PeerJ* **2**, e644 (2014).
35. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2–Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
37. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
38. M. A. Suchard *et al.*, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
39. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
40. V. N. Minin, E. W. Bloomquist, M. A. Suchard, Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471 (2008).
41. B. Shapiro, A. Rambaut, A. J. Drummond, Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23**, 7–9 (2006).
42. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
43. A. U. Tamuri, Treesub: Annotating ancestral substitution on a tree. Version 0.2. https://github.com/tamuri/treesub. Accessed 8 July 2019.
44. Y. C. F. Su *et al.*, Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nat. Commun.* **6**, 7952 (2015).
45. N. Nagarajan, C. Kingsford, GiRaF: Robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res.* **39**, e34 (2011).
46. S. L. Kosakovsky Pond, S. D. Frost, Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222 (2005).
47. W. Delport, A. F. Poon, S. D. Frost, S. L. Kosakovsky Pond, Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455–2457 (2010).
48. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
49. B. Murrell *et al.*, Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
50. R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.*, 310–322 (2002).
51. V. N. Minin, M. A. Suchard, Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* **56**, 391–412 (2008).
52. J. Bahl *et al.*, Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19359–19364 (2011).
53. T. Bedford, S. Cobey, P. Beerli, M. Pascual, Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathog.* **6**, e1000918 (2010).