Explaining the geographic origins of seasonal influenza A (H3N2)

Frank Wen[*1], Trevor Bedford[2], and Sarah Cobey[1]

[1]Department of Ecology and Evolution, University of Chicago,
1101 E. 57th St., Chicago, IL 60637
[2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer
Research Center

March 31, 2016

1   Abstract

Most antigenically novel and evolutionarily successful strains of seasonal influenza A (H3N2) originate in East, South, and Southeast Asia. To understand this pattern, we simulated the ecological and evolutionary dynamics of influenza in a host metapopulation representing the temperate north, tropics, and temperate south. Although seasonality and air traffic are frequently used to explain global migratory patterns of influenza, we find that other factors may have a comparable or greater impact. Notably, a region's basic reproductive number (R_0) strongly affects the antigenic evolution of its viral population and the probability that its strains will spread and fix globally: a 17-28% higher R_0 in one region can explain the observed patterns. Seasonality, in contrast, increases the probability that a tropical (less seasonal) population will export evolutionarily successful strains but alone does not predict that these strains will be antigenically advanced. The relative sizes of different host populations, their birth and death rates, and the region in which H3N2 first appears affect influenza's phylogeography in different but relatively minor ways. These results suggest general principles that dictate the spatial dynamics of antigenically evolving pathogens and offer predictions for how changes in human ecology might affect influenza evolution.

[*]frankwen@uchicago.edu

2 Introduction

Antigenic variants of seasonal influenza continuously emerge and escape human immunity in a process known as antigenic drift. These drifted strains are less easily recognized by host immunity and therefore have a transmission advantage. More antigenically advanced strains are also more likely to spread globally and successfully perpetuate the evolutionary lineage of subsequent variants.

Asia has long been recognized as a major source of not only new influenza subtypes but also new strains of seasonal influenza [1–4]. Phylogeographic analyses have shown that East, South, and Southeast Asia contribute disproportionately to the evolution of seasonal H3N2, exporting most of the evolutionarily successful strains that eventually spread globally [5–9]. The trunk of H3N2's phylogeny traces the evolutionary path of the most successful lineage and is estimated to be located in Asia 87% of the time [9]. Additionally, strains of H3N2 isolated in E-SE Asia appear to be more antigenically advanced, with new antigenic variants emerging earlier in E-SE Asia than in the rest of the world [6,10]. These observations suggest that ecological differences between regions, such as climate and human demography, affect the local antigenic evolution of H3N2, which in turn shapes its global migratory patterns. Here we ask what ecological factors might explain the disproportionate contributions of particular host populations to the evolution of an influenza-like pathogen. This information may be immediately useful for viral forecasting. Over the long term, it could help predict changes in influenza's phylogeography and identify effective global vaccination strategies.

The conspicuous role of tropical regions in H3N2 evolution has been attributed to the seasonal nature of influenza, which experiences pronounced winter epidemics in temperate regions [2, 5–8, 11]. Bottlenecks from late spring through autumn necessarily limit local evolution and reduce opportunities for strain emigration [12,13]. By contrast, in tropical and subtropical Asia, semiconnected host populations support asynchronous epidemics that enable regional persistence year-round [6,11,14]. Smaller contributions from other tropical and subtropical regions might arise from the weaker connectivity of their host populations [8,15,16].

Although seasonality clearly affects temporal patterns of viral migration [7], a robust explanation for differences in regions' long-term contributions to the evolution of H3N2 would consider the effects of seasonal variation in transmission in light of other potentially influential differences among host populations, including:

Host population size. E-S-SE Asia alone contains more than half of the global population [17]. Larger host populations should sustain larger viral populations, and in the absence of other effects, they should contribute a proportionally larger fraction of strains that happen to spread globally. Additionally, if rare mutations limit the generation of antigenic variants, larger populations could contribute a disproportionate number of antigenically novel strains with high fitness.

Host population turnover. Birth rates have historically been higher in E-S-SE Asia than in most temperate populations [17]. Demographic rates influ-

ence the replenishment of susceptibles and loss of immune individuals, thereby modulating selection for antigenic change. Faster replenishment of susceptibles increases prevalence, and thus viral abundance and diversity, but weakens the fitness advantage of antigenic variants. A more immune population imposes greater selection for antigenic change but supports a smaller, less diverse viral population. Thus, the rate of antigenic evolution may vary in a complex way with the rate of host population turnover [18].

Initial conditions. H3N2 first emerged in or near Hong Kong in 1968. The region in which a subtype emerges may effectively give the viral population a head start on evolution. The first epidemic will almost certainly occur in this region, and viruses here will be the first to experience selective pressure for antigenic change. If host migration rates are low and the founding viral population persists, this antigenic lead could be maintained or even grow in time.

Transmission rates. Differences in human behavior can affect transmission rates. The transmission rate affects a strain's intrinsic reproductive number (R_0), the expected number of secondary cases caused by a single infection in an otherwise susceptible population. Differences in regional R_0 could affect evolution in at least two ways. Higher R_0 increases the equilibrium prevalence, increasing the probability that rare beneficial mutations will appear. In addition, the rate of antigenic drift increases with R_0 in models that include mutation as a diffusion-like process [9, 19–21]. A higher intrinsic reproductive number in one population could thus accelerate the emergence of novel mutants in that area.

To determine the potential effects of these five factors on the evolution of H3N2 in space, we simulated an influenza-like pathogen in three regions, representing the temperate north, tropics, and temperate south, and analyzed the effects of these factors on two key metrics of influenza's spatial evolutionary and antigenic dynamics. The first metric measures the proportion of the trunk of the phylogeny present in the tropics (Fig. 1A). The phylogenetic trunk represents the most evolutionarily successful lineage that goes on to seed all future outbreaks. The second metric measures the degree to which tropical strains are antigenically advanced (Fig. 1B). Phenotypically, antigenic dissimilarities can be quantified as distances in antigenic space using pairwise measures of cross reactivity [10, 22]. Our model uses an analogous measure of antigenic distances, allowing us to determine the relative antigenic advancement of strains from each region.

3 Results

3.1 Influenza-like patterns

We simulated an individual-based model that included ecological and evolutionary dynamics in a metapopulation with three demes [23]. In one deme, transmission rates are constant throughout the year, and in the two others, transmission rates vary sinusoidally with opposing phases. Viral phenotypes

occur as points in 2D Euclidean space, and mutation displaces phenotypes in this 2D space according to a fixed kernel [23]. This space is analogous to an antigenic map constructed from pairwise measurements of cross-reactivity between influenza strains using a hemagglutination inhibition (HI) assay [10, 22]. Susceptibility to infection is proportional to the distance in antigenic space between the challenging strain and the nearest strain in the host's infection history, giving distant or antigenically advanced strains greater transmissive advantage.

The model reproduces characteristic ecological and evolutionary features of H3N2, except the antigenic lead in the tropics (Table 1), under the default parameters (Table 2). We restricted our analyses to simulations where the virus remained endemic and where the time to the most recent common ancestor (TMRCA) never exceeded 10 years during the 40 years of simulation. Of 100 replicate simulations, the viral population went extinct in 18 cases and exceeded the TMRCA threshold 29 times, leaving 53 simulations for analysis. The model tracks the ancestry of individual strains, allowing us to explicitly reconstruct the phylogeny of the virus and the geographic location of lineages. The phylogeny has the characteristically well-defined trunk with short branches of the H3N2 hemagglutinin (Fig. 1). This shape arises due repeated selective sweeps of antigenic variants, which reduces standing diversity; the average TMRCA across replicates was 3.72 years (SD = 0.26), comparable to empirical estimates of 3.89 years [9]. The antigenic distance from the founder increased linearly with time (Fig. 1), characteristic of H3N2's canalized antigenic evolution [22, 23]. The mean antigenic drift across replicate simulations was 0.97 antigenic units per year (SD = 0.11), comparable to observed rates of 1.01 antigenic units per year [10]. The mean annual incidence was 9.1% (SD = 0.8%), in agreement with estimates of 9-15% for all seasonal influenza reported by the World Health Organization [24].

Although all three host populations were the same size, the tropical strains were on average more evolutionarily successful. The phylogenetic trunk traces the lineage of strains that survive to the end of the simulation—in other words, the most successful lineage—and was located in the tropics 77% (SD = 13%) of the time, comparable to the observed 87% of the H3N2 phylogenetic trunk located in E-S-SE Asia [9]. Phenotypically, simulated strains were not more antigenically advanced in any particular population. However, antigenic cartography shows that while H3N2 drifts on average at 1.01 antigenic units per year globally [10], Asian strains tend to be farther drifted at any given time, and the region is thus considered to lead antigenically [6, 10]. Thus, the default parameterization does not capture the observed antigenic lead of Asian strains (Table 1).

3.2   Seasonality

Seasonality by itself in the two temperate populations could not cause the tropics to produce more antigenically advanced strains; however, seasonality did cause the tropics to contribute a greater fraction of evolutionarily successful strains (Fig. 2). The trunk spent 87% of its time in the tropics (the same fraction that
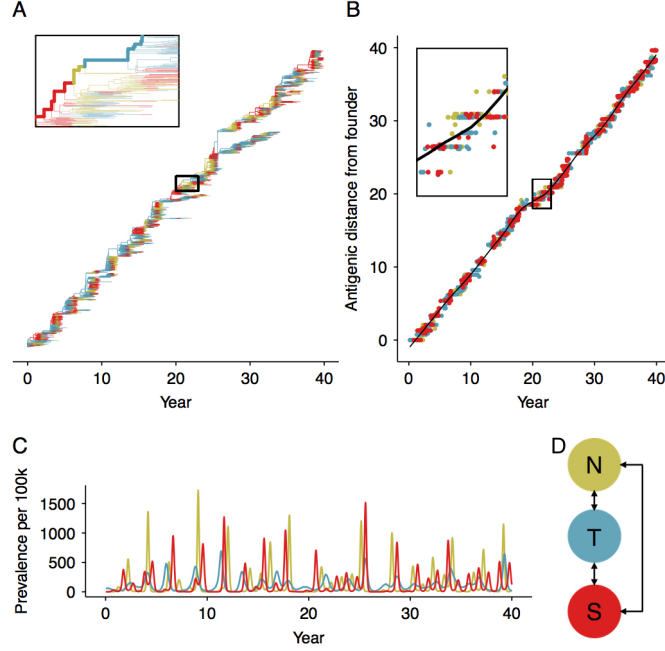
4

Figure 1: Representative output showing influenza-like behavior from a sample simulation using the default parameters (Table 2). Statistics reported here are based on 53 replicate simulations. (A) The phylogeny of the pathogen is reconstructed explicitly from the recorded ancestry of simulated strains. Branches are colored by region indicated in panel D. The trunk is determined by tracing the recorded ancestry of surviving strains at the end of the simulation. Side branches show lineages that go extinct. (B) Viruses evolve antigenically away from the founding strain in a canalized fashion. On average, the antigenic distance from the founding strain follows the trajectory indicated by the black LOESS spline fitted to viruses from all three regions. At any given point in time, strains above this line have drifted farther from the founder compared to average, and are thus considered antigenically leading. Conversely, strains below this line are considered antigenically lagging. Antigenic lead is calculated as the distance to the spline in antigenic units. (C) Prevalence of infection over time for each region. (D) Depiction of the totally connected model population, composed of the temperate north, tropics, and temperate south.

is observed in Asia [9]) with a seasonal transmission amplitude (ϵ) of 0.19 (95% CI: 18%, 20%). In multivariate sensitivity analysis, the amplitude of seasonal transmission accounted for 33% of the variation in the tropical fraction of the trunk (*SI appendix*, Fig. S2, Table S2). This result suggests that seasonal bottlenecks in temperate populations discourage seasonal strains from fixing globally, in agreement with other models [12]. However, seasonality alone could

Table 1: Properties of the default model

Statistic	Model mean ± SD	Observed (Ref)
Annual incidence	0.091±0.0077	0.09 - 0.15 [24]
Antigenic drift rate (a.u. yr^{-1})	0.97±0.11	1.01 [10]
TMRCA (years)	3.7±0.26	3.89 [9]
Frac. of trunk in the tropics	0.61±0.13	0.87 [9]
Tropics antigenic lead (a.u.)	0.0025±0.036	0.25 [6, 10]

Table 2: Default parameters

Parameter	Value	Reference
Intrinsic reproductive number (R_0)	1.8	[25, 26]
Duration of infection ν	5 days	[27]
Population size N	45 million	(see text)
Birth/death (turnover) rate γ	1/30 $year^{-1}$	[17]
Mutation rate μ	10^{-4} day^{-1}	(see text)
Mean mutation step size δ_{mean}	0.6 antigenic units	(see text)
SD mutation step size δ_{sd}	0.3 antigenic units	(see text)
Infection risk conversion s	0.07	[23, 28, 29]
Migration rate m	10^{-3} day^{-1}	(see text)
Seasonal amplitude ϵ	0.10	[30]

not explain any variation in the tropic's antigenic lead (*SI appendix*, Fig. S2, Table S3). We therefore hypothesized that ecological factors besides seasonality must contribute to regional differences in relative antigenic fitness.

3.3 Transmission rate in the tropics

Increasing R_0 in the tropics relative to the temperate populations caused the tropics to produce strains that led antigenically, while also preserving the tropics' contribution to the trunk (Fig. 3). Our results predict that a 28% increase in R_0 (95% CI: 25%, 30%) in the tropics causes the tropics to produce strains that are, on average, 0.25 antigenic units ahead of global mean, in agreement with the observed antigenic lead in Asia [6, 10]. We also estimate that a 17% increase in R_0 (95% CI: 15%, 19%) causes the phylogenetic trunk to be located in the tropics 87% of the time, reproducing the observed fraction of the H3N2 trunk in Asia [9].

The effects of R_0 on antigenic lead were robust to changes in other ecological variables as well. When we varied the other parameters (Table 2), relative R_0 in the tropics accounted for 77% of the variance in antigenic lead, making it the best predictor of antigenic lead in the tropics (*SI appendix*, Fig. S2, Table

S3). The fraction of the trunk in the tropics also increased with the relative R_0, although R_0 explained less of the variation in trunk proportion (41%), due to the effect of seasonality (*SI appendix*, Fig. S2, Table S2). In the presence of strong seasonality in the temperate populations, even antigenically lagging strains from the tropics could form part of the trunk (*SI appendix*, Fig. S2) because temperate strains sometimes failed to spread.
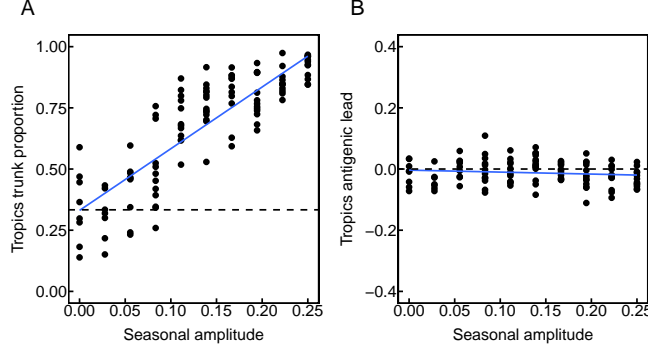

Figure 2: Seasonal amplitude ϵ in the temperate populations increases the tropics' contribution to the most evolutionarily successful lineage but alone does not affect regional differences in antigenic advancement. Transmission rates β in the temperate north and south oscillate sinusoidally in opposite phase, with amplitude ϵ. (A) Effects of seasonality on the fraction of the trunk in the tropics (Pearson's $r = 0.85$, $p < 0.001$; $R^2 = 0.72$). Each point shows the fraction of time that the phylogenetic trunk was located in the tropics during the course of one simulation. The dashed line represents the null hypothesis where tropical strains comprise one third of the phylogenetic trunk. (B) Effects on seasonality on the antigenic lead of the tropics (Pearson's $r = -0.12$, $p = 0.20$, $R^2 = 0.01$). Each point shows the average antigenic lead of tropical strains over time from one simulation. The dashed line represents the null hypothesis where tropical strains are neither antigenically ahead or behind. Blue lines represent linear least squares regression.

To better understand why increasing tropical R_0 causes the tropics to produce more antigenically advanced strains, we examined the effect of R_0 on antigenic evolution in a single deme. Simulations showed that increasing R_0 increases the rate of antigenic drift (*SI appendix*, Fig. S3). To investigate further, we derived an analytic expression for the invasion fitness of a novel mutant in a population at the endemic equilibrium (*SI appendix*, Eq. S1). When the resident and mutant strains have the same intrinsic fitness (R_0), the growth rate of an antigenically distinct, invading mutant increases linearly with R_0 (*SI appendix*, Fig. S3). The strength of this relationship depends on the antigenic distance between the mutant phenotype and the phenotype of the resident strain, such that increases in R_0 have a greater effect on distant mutants compared to closer mutants.
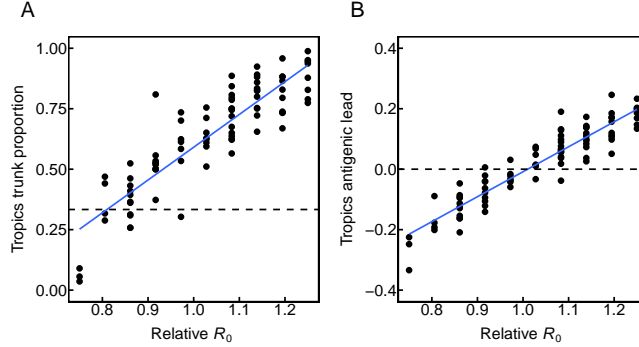
Figure 3: Increased R_0 in the tropics increases the tropics' contribution to the most evolutionarily successful lineage and the antigenic advancement of tropical strains. Relative R_0 is calculated as R_0 in the tropics divided by R_0 in the temperate regions. (A) Effect of R_0 in the tropics on the fraction of the trunk in the tropics (Pearson's $r = 0.88$, $p < 0.001$; $R^2 = 0.78$). Each point shows the fraction of phylogenetic trunk located in the tropics during one simulation. The dashed line represents the null hypothesis where tropical strains comprise one third of the phylogenetic trunk. (B) Effect of R_0 in the tropics on the antigenic lead in the tropics (Pearson's $r = 0.93$, $p < 0.001$; $R^2 = 0.87$). Each point shows the average antigenic lead of tropical strains over time from one simulation. The dashed line represents the null hypothesis where tropical strains are neither antigenically ahead or behind. Blue lines represent linear least squares regression.

Although seasonality alone did not affect antigenic lead, the effects of R_0 on antigenic lead were influenced by seasonality (Fig. 4). Introducing seasonality in the temperate populations reduced differences in antigenic phenotype between regions. When tropical strains were antigenically ahead of temperate strains (due to higher tropical R_0), introducing seasonality reduced the tropics' antigenic lead. When tropical strains were antigenically behind temperate strains (due to lower tropical R_0), introducing seasonality reduced the antigenic lag. Two factors explain the equalizing effect of seasonality on antigenic phenotype. First, high contact rates during temperate winters increase the rate of strain immigration from the tropics. Second, seasonal troughs in prevalence allow tropical strains to invade more easily due to reduced competition with local strains.

3.4 Demographic rates, population size, and initial conditions

Other ecological factors affected regional contributions to evolution but could not reproduce the observed patterns as well as differences in R_0 (*SI appendix*, Figs. S1, S2). Notably, strains were slightly more antigenically advanced in

older populations (Fig. S1). When the rate of population turnover in the tropics was half that in the temperate regions, the tropics led by 0.04 antigenic units (SD = 0.03). Larger populations generally contributed more to the trunk, although there was much variation that population size alone did not explain (*SI appendix*, Figs. S1, S2 and Tables S3, S2). Initial conditions did not have a lasting effect (*SI appendix*, Fig. S5).

3.5 Implications for other influenza subtypes

Influenza A/H1N1 and influenza B both evolve slowly compared to H3N2 and are suspected to have lower R_0 [9, 10]. Specifically, H1N1 drifts at a rate of 0.62 antigenic units per year, and the B/Victoria and Yamagata strains drift at 0.42 and 0.32 antigenic units per year respectively [10]. H1N1 and B viruses are also less apt to have Asian origins than H3N2 [9]. When we simulate with lower baseline R_0, we find that differences in R_0 between regions have a weaker influence on spatial patterns of evolution (*SI appendix*, Fig. S8). Based on the relationship between mean R_0 and antigenic drift (Fig. S3), we would expect seasonal H1N1, for example, to have an R_0 of 1.6. For this R_0, a 17% increase in R_0 causes the tropics to occupy only 79% (versus 87% for H3N2-like R_0 of 1.8) of the trunk, and a 28% increase in R_0 causes the tropics to lead by 0.20 (versus 0.25 for H3N2) antigenic units.
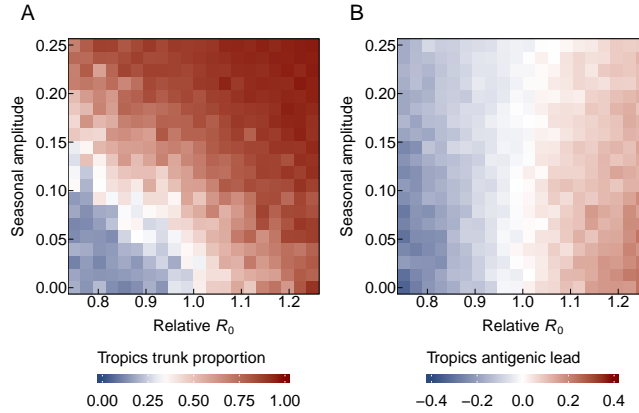

Figure 4: Seasonality in temperate populations has an equalizing effect on antigenic differences. Relative R_0 is calculated as R_0 in the tropics divided by R_0 in the temperate regions. (A) Effects of seasonality and R_0 on the fraction of the trunk in the tropics. Blue indicates that the phylogenetic trunk is located in the tropics less than 1/3 of the time, and red indicates that the trunk is the tropics more than 1/3 of the time. (B) Effects of seasonality and R_0 on antigenic lead in the tropics. Blue indicates that tropical strains are on average ahead antigenically relative to other global strains and red indicates that tropical strains are behind antigenically. Each square averages 1 to 17 replicate simulations.

4 Discussion

We find that the simplest explanation for why a host population produces more antigenically novel and evolutionarily successful strains than other populations is that its strains have a higher intrinsic fitness, or R_0. The strong effect of regional R_0 on spatial patterns of viral evolution is caused by the effect of R_0 on antigenic drift. Higher regional R_0 facilitates invasion of antigenically novel strains, resulting in faster antigenic drift. Seasonality reduces the rate at which temperate populations export strains that are evolutionarily successful, but seasonality alone cannot explain regional differences in the production of strains that are antigenically novel. Size and age can influence global patterns too, but to a lesser extent: larger populations export more strains that fix, and populations with slower replenishment of susceptibles increase the rate of antigenic evolution. These last two effects are sensitive to changes in seasonality and R_0.

Our model uses a simple host population structure and minimal evolutionary landscape. Although these simplifying assumptions limit our ability to replicate influenza's phylogeographic dynamics precisely, the model nonetheless reveals how ecological differences in populations create spatial patterns in the evolution of an influenza-like pathogen.

These results immediately raise the question of whether there is evidence of regional variation in R_0. Low reporting rates and antigenic evolution make the R_0 of influenza difficult to measure with traditional methods, but we can conjecture from several lines of evidence. Low absolute humidity favors transmission in experimental settings [31] and also influences the timing of the influenza season in the United States [32]. Based on absolute humidity alone, these results suggest that R_0 of tropical and subtropical Asia would be lower than in temperate latitudes. However, in Vietnam the onset of influenza-like illness is associated with periods of high humidity [33]. This observation suggests that humidity is not the dominant driver of influenza transmission, at least in this region.

Contact rates also influence transmission [34]. Multiple studies have detected a significant effect of school closure on influenza spread [35–37], although this trend is not without exception [38]. Households also influence risk: after one household member is infected, the average risk of secondary infection in a household contact is 10% [39]. Differences in classroom and household sizes may thus influence local transmission, and both are higher in, for instance, China and India than in Europe and the U.S. [40, 41]. Contact surveys report higher contact rates in Guangdong, China, than in European communities, whereas those in Vietnam are lower, although differences may arise from differences in survey design [42–44]. These surveys notably miss non-social, casual contacts (e.g., shared cafeterias and elevators) that might be important for influenza transmission.

Differences in local transmission rates may not scale: high rates of local transmission may be offset or attenuated by the structure of contact networks over larger areas. At the regional level, commuter and air passenger flows affect the spread of influenza epidemics, suggesting that adults are important to the

long-range dispersal of the virus [11, 16]. The frequency of long-distance contacts differs between communities [42]. Although sensitivity of R_0 to network topology is well known theoretically [45, 46], there is a need to integrate the features of local and regional empirical transmission networks to infer large-scale differences R_0.

Empirical estimates of R_0 are in theory attainable from seroprevalence. Under a simplistic, single-strain SIR model, which assumes random mixing and no maternal immunity, differences in R_0 should appear in differences in seropositivity by age. For instance, if R_0=1.8, approximately 5.1% of 2 year-olds would be seropositive, whereas 7.4% would be seropositive if R_0 were 20% higher. R_0 variation in this range could be detected by sampling as few as 1500 2 year-olds in each population. Detailed surveys of H3N2 seropositivity by age cohort exist for some European countries [47, 48] but show much faster increases in seropositivity with age than expected under the SIR model: 100% of tested children are seropositive to H3N2 by age 7 in the Netherlands and by age 12 in Germany. This discrepancy may be due to antigenic drift resulting in higher attack rates [9]. The difference may also reflect greater contact rates among school-aged children [43] and highlights the possibility that differences in exposure rates at young ages do not reflect mean differences in the populations. Such effects may be reduced by examining seroprevalence at older ages, but these estimates must balance a tradeoff between minimizing age-related correlations in transmission rates and increasing sample sizes required to detect asymptotically small differences in seropositivity. Another potential approach to measuring R_0 is to refine estimates of annual incidence in different populations. Estimates of R_0 based on annual incidence would have to incorporate the histories of recent circulating strains, survey timing and titer dynamics, and vaccination in each population.

A greatly reduced birth rate confers a slight antigenic lead, but actual differences in birth rates between regions appear too small to explain Asia's observed lead. Current birth rates across most of Europe, China, and the United States are within 10% of each other [17]. Birth rates are almost twice as high in some SE Asian countries, including Cambodia, Laos, and the Philippines. The highest birth rates are found in Africa and the Middle East, and are three to four times higher than birth rates in the United States and China. Our model suggests that these regions should contribute relatively less to influenza's antigenic evolution, assuming the differences in population structure are not associated with higher R_0, and ignoring other differences.

We expect these results to apply to other antigenically varying, fast-evolving pathogens, including other types of influenza. Enterovirus-71 circulates globally, and its VP1 capsid protein experiences continuous lineage replacement through time, similarly to H3N2 hemagglutinin [49]. Norovirus also demonstrates rapid antigenic evolution by amino acid replacements in its capsid protein [50]. We might expect that areas with high transmission contribute disproportionately to the antigenic evolution and global spread of these pathogens. In addition, when we simulate with lower R_0, we find that differences in R_0 between regions influence spatial patterns of antigenic variation less (*SI appendix*, Fig. S8). This

may explain why influenza A H1N1 and influenza B, which are suspected to have lower R_0 [9,10], are less apt to have Asian origins than H3N2 [9].

5   Acknowledgements

6   Author contributions

TB and SC conceived the study. FW performed the analysis and wrote the first draft of the paper. All of the authors contributed to and approve the final version.

7   Materials and methods

7.1   Description of the model

We implemented an individual-based SIR compartmental model of an influenza-like pathogen, originally described by Bedford et al. [23]. In this model, a global metapopulation is composed of three connected populations, representing tropics and temperate north and south. Individuals' compartments are updated using a τ-leaping algorithm. Within a region i, the force of infection is given by

$$F_i(t) = \beta_i(t)\frac{I_i}{N_i}$$

where I is the number of infected hosts. Between regions i and j, the force of infection is given by

$$F_{ij}(t) = m\beta_j(t)\frac{I_i}{N_j}$$

where region i is where the infection originates and region j is the destination. Here, m is a scaling factor for interregional transmission, and β_j is the transmission rate of the destination region. Transmission rates in the seasonal north and south oscillate sinusoidally in opposite phase with amplitude ϵ. After recovery from infection, a host acquires complete immunity to viruses with that specific antigenic phenotype. Hosts that clear infection accumulate an infection history that defines their immunity. In a contact event, the distances between the infecting viral phenotype and each phenotype in the susceptible host's immune history are calculated. The probability of infection after contact is proportional

12

to the distance d to the closest phenotype in the host's immune history. An individual's risk of infection by such a strain is

$$\text{Risk} = \min \{1, sd\}$$

where the proportionality constant for converting antigenic distance to a risk of infection $s = 0.07$ [23]; in other words, one unit of antigenic distance corresponds to 7% reduction in immunity. The linear relationship s between antigenic distance and susceptibility derives from studies of vaccine efficacy [23, 28, 29].

Antigenic phenotypes are represented by points in a two-dimensional Euclidean antigenic space. One unit of antigenic distance in this space corresponds to a twofold dilution of antiserum in an HI assay [22]. The model is initialized at the endemic equilibrium with antigenically identical viruses. By default, all of the initial infections occur in the tropics. Mutational events occur at a rate μ mutations per day. When a virus mutates, it moves in a random radial direction with a gamma-distributed step size. This mutation rate, along with the mutation size parameters $(\delta_{\mathrm{mean}}, \delta_{\mathrm{sd}})$ determine the accessibility of more distant mutations in antigenic space. The radial direction of mutation is chosen from a uniform distribution.

7.2   Selection of parameters

Parameter values were selected to be consistent with influenza's biology and to reproduce its major epidemiological and evolutionary patterns 2. The population size N was chosen to minimize extinctions while also making efficient use of computational resources. The population birth/death rate $\gamma = 1/30 \text{ year}^{-1}$ reflects the global crude birth rate estimates of 34 births per 1000 [17].

The proportionality constant m for calculating the between-region contact rate was calculated from the number of international air travel passengers reported by the International Civil Aviation Organization divided by the global population [51].

We chose a baseline $R_0 = 1.8$. Estimates of R_0 from the first pandemic wave H3N2 in 1968 range from 1.06-2.06 [25], and estimates of R_0 for seasonal influenza range from 1.16 to 2.5, averaging approximately 1.8 [26].

The five-day duration of infection, $1/\nu$, is based on estimates from viral shedding [27]. The transmission rate, β, is calculated using the definition of R_0:

$$R_0 = \frac{\beta}{\nu + \gamma}$$

We chose the seasonal amplitude to ensure consistent troughs during the off-season in temperate populations while remaining within reasonable estimates of seasonal transmission rates [30].

Mutations occur at a rate of 10^{-4} mutations per day. The distance of each mutation is sampled from a gamma distribution with parameters chosen to yield a mean step size of 0.6 and a standard deviation of 0.3 antigenic units. These values correspond to a reduction in immunity of 4.2% for an average mutation

(SD = 2.1%). These mutation effect parameters give the gamma distribution an exponential-like shape, so that most mutations yield small differences in antigenic fitness, while occasionally mutations will yield greater differences. We chose μ, δ_{mean}, and δ_{sd} so that the simulation would exhibit influenza-like behavior as consistently as possible (Fig. S6).

7.3   Calculation of antigenic lead and trunk proportion

We examined two metrics that describe influenza's evolutionary dynamics. For computational tractibility, these metrics were calculated using a subset of strains sampled over course of the simulation. Strains were sampled proportionally to prevalence.

To calculate antigenic lead, we first calculated the antigenic distance of each sampled strain from the founding strain (Fig. 1A). We then fit a LOESS spline to these distances over time. The spline describes the expected antigenic drift of circulating lineages at any point in time. Strains above the spline have drifted farther than average and are considered antigenically leading. Strains below the spline have drifted less than average and are considered antigenically lagging. The antigenic lead in the tropics is calculated as the average antigenic distance to this spline for all sampled tropical strains.

To calculate the fraction of the trunk in each population, we first identified strains that comprise the trunk by tracing the lineage of strains that survive to the end of the simulation (Fig. 1B). Because multiple lineages may coexist at the end of the simulation, we excluded the last 5 years of strains from trunk calculations. The fraction of the trunk in the tropics is calculated as the fraction of the time the trunk was composed of tropical strains.

7.4   Multivariate sensitivity analysis

To test the robustness of the effects of individual parameters on the antigenic lead and the phylogenetic trunk, we simulated 500 points from a Latin hypercube with dimensions representing relative R_0, seasonality, relative population size, relative population turnover rate, and the fraction of initial infecteds in the tropics (Fig. S2). The ranges for each parameter (Table S1) were chosen to remain within reasonable estimates. We simulated twenty replicates for each of the 500 unique parameter combination and discarded simulations in which the virus went extinct or the TMRCA exceeded 10 years at any time in the 40-year simulation. We performed an ANOVA on the remaining 4119 influenza-like simulations to determine each parameter's contribution to the variance in antigenic lead and the fraction of the trunk in the tropics (Table S3, S2).

7.5   Data accessibility

Code implementing the model is available at `https://github.com/cobeylab/antigen-phylogeography.git`. The complete code for reproducing these re-

14

sults is available at `https://github.com/cobeylab/influenza_phylogeography_manuscript.git`.

References

[1] K F Shortridge and C H Stuart-Harris. An influenza epicentre? *Lancet (London, England)*, 2(8302):812–813, October 1982.

[2] R G Webster, W J Bean, O T Gorman, T M Chambers, and Y Kawaoka. Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56(1):152–179, March 1992.

[3] N J Cox, T L Brammer, and H L Regnery. Influenza: global surveillance for epidemic and pandemic variants. *European journal of epidemiology*, 10(4):467–470, August 1994.

[4] N J Cox and K Subbarao. Global epidemiology of influenza: past and present. *Annual review of medicine*, 51:407–421, 2000.

[5] Andrew Rambaut, Oliver G Pybus, Martha I Nelson, Cecile Viboud, Jeffery K Taubenberger, and Edward C Holmes. The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195):615–619, May 2008.

[6] C A Russell, T C Jones, I G Barr, N J Cox, R J Garten, V Gregory, I D Gust, A W Hampson, A J Hay, A C Hurt, J C de Jong, A Kelso, A I Klimov, T Kageyama, N Komadina, A S Lapedes, Y P Lin, A Mosterin, M Obuchi, T Odagiri, A D M E Osterhaus, G F Rimmelzwaan, M W Shaw, E Skepner, K Stohr, M Tashiro, R A M Fouchier, and D J Smith. The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science (New York, N.Y.)*, 320(5874):340–346, April 2008.

[7] Justin Bahl, Martha I Nelson, Kwok H Chan, Rubing Chen, Dhanasekaran Vijaykrishna, Rebecca A Halpin, Timothy B Stockwell, Xudong Lin, David E Wentworth, Elodie Ghedin, Yi Guan, J S Malik Peiris, Steven Riley, Andrew Rambaut, Edward C Holmes, and Gavin J D Smith. Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 108(48):19359–19364, November 2011.

[8] Trevor Bedford, Sarah Cobey, Peter Beerli, and Mercedes Pascual. Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathogens*, 6(5):e1000918, May 2010.

[9] Trevor Bedford, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, Alexander Klimov, Nicola S Lewis, Xiyan Li, John W McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu,

Eugene Skepner, Derek J Smith, Marc A Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin A Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, July 2015.

[10] Trevor Bedford, Marc A Suchard, Philippe Lemey, Gytis Dudas, Victoria Gregory, Alan J Hay, John W McCauley, Colin A Russell, Derek J Smith, and Andrew Rambaut. Integrating influenza antigenic dynamics with molecular evolution. *eLife*, 3:e01914, 2014.

[11] Cecile Viboud, Wladimir J Alonso, and Lone Simonsen. Influenza in tropical regions. *PLoS medicine*, 3(4):e89, April 2006.

[12] Ben Adams and Alice Carolyn McHardy. The impact of seasonal and year-round transmission regimes on the evolution of influenza A virus. *Proceedings. Biological sciences / The Royal Society*, 278(1716):2249–2256, August 2011.

[13] Daniel Zinder, Trevor Bedford, Edward B Baskerville, Robert J Woods, Manojit Roy, and Mercedes Pascual. Seasonality in the migration and establishment of H3N2 Influenza lineages with epidemic growth and decline. *BMC evolutionary biology*, 14(1):272, 2014.

[14] X Cheng, Y Tan, M He, T T Y Lam, X Lu, C Viboud, J He, S Zhang, J Lu, C Wu, S Fang, X Wang, X Xie, H Ma, M I Nelson, H f Kung, E C Holmes, and J Cheng. Epidemiological dynamics and phylogeography of influenza virus in southern China. *Journal of Infectious Diseases*, 207(1):106–114, January 2013.

[15] Joseph Chan, Antony Holmes, and Raul Rabadan. Network Analysis of Global Influenza Spread. *PLoS computational biology*, 6(11):e1001005–10, November 2010.

[16] Philippe Lemey, Andrew Rambaut, Trevor Bedford, Nuno Faria, Filip Bielejec, Guy Baele, Colin A Russell, Derek J Smith, Oliver G Pybus, Dirk Brockmann, and Marc A Suchard. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens*, 10(2):e1003932, February 2014.

[17] United Nations, Department of Economic and Social Affairs, Population Division, New York. *World Population Prospects: The 2012 Revision*, 2013.

[18] Bryan T Grenfell, Oliver G Pybus, Julia R Gog, James L N Wood, Janet M Daly, Jenny A Mumford, and Edward C Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science (New York, N.Y.)*, 303(5656):327–332, January 2004.

[19] Juan Lin, Viggo Andreasen, Renato Casagrandi, and Simon A Levin. Traveling waves in a model of influenza A drift. *Journal of theoretical biology*, 222(4):437–445, June 2003.

[20] Julia R Gog and Bryan T Grenfell. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences*, 99(26):17209–17214, December 2002.

[21] Adam Kucharski and Julia R Gog. Influenza emergence in the face of evolutionary constraints. *Proceedings. Biological sciences / The Royal Society*, 279(1729):645–652, February 2012.

[22] Derek J Smith, Alan S Lapedes, Jan C de Jong, Theo M Bestebroer, Guus F Rimmelzwaan, Albert D M E Osterhaus, and Ron A M Fouchier. Mapping the antigenic and genetic evolution of influenza virus. *Science (New York, N.Y.)*, 305(5682):371–376, July 2004.

[23] Trevor Bedford, Andrew Rambaut, and Mercedes Pascual. Canalization of the evolutionary trajectory of the human influenza virus. *BMC biology*, 10(1):38, 2012.

[24] World Health Organization. *Influenza Fact Sheet*, 2014.

[25] Charlotte Jackson, Emilia Vynnycky, and Punam Mangtani. Estimates of the transmissibility of the 1968 (Hong Kong) influenza pandemic: evidence of increased transmissibility between successive waves. *American journal of epidemiology*, 171(4):465–478, February 2010.

[26] Matthew Biggerstaff, Simon Cauchemez, Carrie Reed, Manoj Gambhir, and Lyn Finelli. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infectious Diseases*, 14(1):1–20, September 2014.

[27] Fabrice Carrat, Elisabeta Vergu, Neil M Ferguson, Magali Lemaitre, Simon Cauchemez, Steve Leach, and Alain-Jacques Valleron. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American journal of epidemiology*, 167(7):775–785, April 2008.

[28] Vishal Gupta, David J Earl, and Michael W Deem. Quantifying influenza vaccine efficacy and antigenic distance. *Vaccine*, 24(18):3881–3888, May 2006.

[29] Andrew W Park, Janet M Daly, Nicola S Lewis, Derek J Smith, James L N Wood, and Bryan T Grenfell. Quantifying the impact of immune escape on transmission dynamics of influenza. *Science (New York, N.Y.)*, 326(5953):726–728, October 2009.

[30] James Truscott, Christophe Fraser, Simon Cauchemez, Aronrag Meeyai, Wes Hinsley, Christl A Donnelly, Azra Ghani, and Neil Ferguson. Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *Journal of the Royal Society, Interface / the Royal Society*, 9(67):304–312, February 2012.

[31] Jeffrey Shaman and Melvin Kohn. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3243–3248, March 2009.

[32] Jeffrey Shaman, Virginia E Pitzer, Cécile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental united states. *PLoS Biol*, 8(2):e1000316, Feb 2010.

[33] Pham Quang Thai, Marc Choisy, Tran Nhu Duong, Vu Dinh Thiem, Nguyen Thu Yen, Nguyen Tran Hien, Daniel J Weiss, Maciej F Boni, and Peter Horby. Seasonality of absolute humidity explains seasonality of influenza-like illness in Vietnam. *Epidemics*, 13:65–73, December 2015.

[34] J Wallinga, P Teunis, and M Kretzschmar. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *American journal of epidemiology*, 164(10):936–944, September 2006.

[35] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M Ferguson. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452(7188):750–754, April 2008.

[36] Anthony Heymann, Gabriel Chodick, Brian Reichman, Ehud Kokia, and Joseph Laufer. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *The Pediatric infectious disease journal*, 23(7):675–677, July 2004.

[37] Dennis L Chao, M Elizabeth Halloran, and Ira M Longini Jr. School Opening Dates Predict Pandemic Influenza A(H1N1) Outbreaks in the United States. *The Journal of Infectious Diseases*, 202(6):877–880, September 2010.

[38] Benjamin J Cowling, Eric H Y Lau, Conrad L H Lam, Calvin K Y Cheng, Jana Kovar, Kwok Hung Chan, J S Malik Peiris, and Gabriel M Leung. Effects of school closures, 2008 winter influenza season, Hong Kong. *Emerging infectious diseases*, 14(10):1660–1662, October 2008.

[39] Tim K Tsang, Lincoln L H Lau, Simon Cauchemez, and Benjamin J Cowling. Household Transmission of Influenza Virus. *Trends in Microbiology*, pages 1–11, November 2015.

[40] United Nations, Department of Economic and Social Affairs, Statistics Division, New York. *UNSD Demographic Statistics*, 2016.

[41] OECD. *Education at a Glance 2015, OECD Indicators*, 2015.

[42] J M Read, J Lessler, S Riley, S Wang, L J Tan, K O Kwok, Y Guan, C Q Jiang, and D A T Cummings. Social mixing patterns in rural and urban areas of southern China. *Proceedings of the Royal Society B: Biological Sciences*, 281(1785):20140268–20140268, April 2014.

[43] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, March 2008.

[44] Peter Horby, Quang Thai Pham, Niel Hens, Thi Thu Yen Nguyen, Quynh Mai Le, Dinh Thoang Dang, Manh Linh Nguyen, Thu Huong Nguyen, Neal Alexander, W John Edmunds, Nhu Duong Tran, Annette Fox, and Tran Hien Nguyen. Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS ONE*, 6(2):e16965, 2011.

[45] J C Miller. Spread of infectious disease through clustered populations. *Journal of The Royal Society Interface*, 6(41):1121–1134, October 2009.

[46] F R Adler. The effects of averaging on the basic reproduction ratio. *Mathematical Biosciences*, 111(1):89–98, September 1992.

[47] R Bodewes, G de Mutsert, F R M van der Klis, M Ventresca, S Wilks, D J Smith, M Koopmans, R A M Fouchier, A D M E Osterhaus, and G F Rimmelzwaan. Prevalence of Antibodies against Seasonal Influenza A and B Viruses in Children in Netherlands. *Clinical and Vaccine Immunology*, 18(3):469–476, March 2011.

[48] Andreas Sauerbrei, R Schmidt-Ott, H Hoyer, and P Wutzler. Seroprevalence of influenza A and B in German infants and adolescents. *Medical Microbiology and Immunology*, 198(2):93–101, February 2009.

[49] Kok Keng Tee, Tommy Tsan-Yuk Lam, Yoke Fun Chan, Jon M Bible, Adeeba Kamarulzaman, C Y William Tong, Yutaka Takebe, and Oliver G Pybus. Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *Journal of Virology*, 84(7):3339–3350, April 2010.

[50] Lisa C Lindesmith, Eric F Donaldson, Anna D Lobue, Jennifer L Cannon, Du-Ping Zheng, Jan Vinje, and Ralph S Baric. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS medicine*, 5(2):e31, February 2008.

[51] International Civil Aviation Organization. *Annual report of the ICAO council: 2014*, 2014.

8 Supporting Information

8.1 Invasion analysis

We assume that the host population supports a resident strain at the endemic equilibrium. We develop an expression for the fitness of an invading mutant strain to explain how the selection coefficient of the mutant changes with R_0.

Here, S, I, and R represent the fraction of susceptible, infected, and recovered individuals. The birth rate γ and the death rate are equal, so the population size is constant. All individuals are born into the susceptible class. Transmission occurs at rate β, and recovery occurs at rate ν.

$$\frac{dS}{dt} = \gamma(1 - S) - \beta SI$$
$$\frac{dI}{dt} = \beta SI - (\nu + \gamma)I$$
$$\frac{dR}{dt} = \nu I - \gamma R$$

We solve for the endemic equilibrium values of S_{eq}, I_{eq}, R_{eq}.

$$\frac{dI}{dt} = 0 = \beta S_{\text{eq}} I_{\text{eq}} - (\nu + \gamma)I_{\text{eq}}$$
$$S_{\text{eq}} = \frac{\nu + \gamma}{\beta} \equiv \frac{1}{R_0}$$

R_0, the basic reproductive number, is defined as the number of secondary infections from a single infected individual in a totally susceptible population. Continuing to solve for I_{eq} and R_{eq}, we have

$$\frac{dS}{dt} = 0 = \gamma(1 - S_{\text{eq}}) - \beta S_{\text{eq}} I_{\text{eq}}$$
$$I_{\text{eq}} = \frac{\gamma}{\beta}(R_0 - 1)$$
$$\frac{dR}{dt} = 0 = \nu I_{\text{eq}} - \gamma R_{\text{eq}}$$
$$R_{\text{eq}} = \frac{\nu}{\beta}(R_0 - 1)$$

To find the selection coefficient, we develop an expression for the effective reproductive number R_e for both the resident and mutant strains. R_e is the expected number of secondary infections from a single infected individual in a given population. We will use the relationship

$$R_e = SR_0$$

The mutant strain is d antigenic units from the resident strain. The conversion factor between antigenic units and infection risk is notated by s. Thus, the susceptibility to the mutant is given by $\min\{sd, 1\}$, and immunity to the mutant is $\max\{1 - sd, 0\}$. For ease of notation, we assume $sd \leq 1$, and use $k = 1 - sd$ The fraction of the population immune to the invading strain is denoted by R'. Note that the population is at the endemic equilibrium of the resident strain, and not the mutant.

$$R' = (1 - sd)R_{\text{eq}}$$
$$= \frac{\nu}{\beta}(R_0 - 1)k$$

We start by allowing coinfection. The fraction of susceptibles to the mutant strain is given by

$$S' = 1 - R' - \frac{1}{N}$$
$$= 1 - \frac{\nu k}{\beta}(R_0 - 1) - \frac{1}{N}$$

For large N, we have

$$S' = 1 - \frac{\nu k}{\beta}(R_0 - 1)$$

As defined by our initial set of ODEs, the growth rates of the mutant and resident strains are

$$\frac{dI'}{dt} = I'[\beta S' - (\nu + \gamma)]$$
$$\frac{dI}{dt} = I_{\text{eq}}[\beta S_{\text{eq}} - (\nu + \gamma)]$$

To get the selection coefficient, we take the difference between the growth rates:

$$s = [\beta S' - (\nu + \gamma)] - [\beta S_{\text{eq}} - (\nu + \gamma)]$$
$$= \beta - \gamma k(R_0 - 1) - \frac{\beta}{R_0}$$

Recall that $\beta = (\nu + \gamma)R_0$

$$s = (\nu + \gamma)R_0 - \nu k(R_0 - 1) - (\nu + \gamma)$$

Simplifying,

$$s = (\nu sd + \gamma)(R_0 - 1) \tag{S1}$$

Now disallowing coinfection, we have

$$S' = 1 - R' - I_{\text{eq}} - I'$$
$$= 1 - \frac{\nu}{\beta}(R_0 - 1)k - \frac{\gamma}{\beta}(R_0 - 1) - \frac{1}{N}$$

For large N,

$$S' = 1 - (R_0 - 1)[\frac{\nu k + \gamma}{\beta}]$$

Using the same arithmetic as in the case with coinfection, it follows that

$$s = \beta - (\nu k + \gamma)(R_0 - 1) - \frac{\beta}{R_0}$$

Simplifying,

$$s = (\nu sd)(R_0 - 1) \tag{S2}$$

In summary, the selection coefficient of an invading mutant strain increases linearly with the R_0, which is shared by both strains. The slope of this relationship is proportional to the distance d between the two strains in antigenic space (Fig. ??).

8.2 Detecting differences in R_0

In the SIR model, the force of infection is

$$F = \beta I$$

where β is the transmission rate and I is the fraction of infecteds. At the endemic equilibrium, the cumulative fraction of seropositive individuals at a given age a is

$$f(a) = 1 - \exp(-\beta I_{\text{eq}} a)$$
$$= 1 - \exp(-R_0(\nu + \gamma I_{\text{eq}} a)$$

where ν is the recovery rate and γ is the birth/death rate. I_{eq}, the fraction of infecteds at the endemic equilibrium, is given by

$$I_{\text{eq}} = \frac{\gamma}{\beta}(R_0 - 1)$$

Figure **??** shows the fraction of seropositive individuals by age for the baseline $R_0 = 1.8$ and a 20% higher $R_0 = 2.16$. The difference in the percentage of seropositive two-year-olds between the two groups is approximately 2.3%. The sample size in each group required to detect a difference $f_2(a) - f_1(a)$ with α confidence and $1 - \beta$ power is

$$N = \frac{f_1(1 - f_1) + f_2(1 - f_2)}{(f_1 - f_2)^2}(\Phi_{\alpha/2} + \Phi_\beta)^2$$

For legibility, $f_i(a)$ is written as f_i. To detect a 20% difference in R_0 between two populations with 0.05 significance and 0.80 power, we would require a sample of at least 1503 individuals in both groups.

9 Supplemental tables and figures

Table S1: Parameter ranges used in Latin hypercube sampling

Parameter	Range
Relative R_0	0.8−1.2
Seasonal amplitude (ϵ) in temperate populations	0.0−0.15
Relative population size (N)	0.5−2.0
Relative turnover rate (γ)	0.5−2.0
Fraction of initial infecteds (I_0) in tropics	0.0−1.0

Table S2: ANOVA of the fraction of trunk in tropics from multivariate sensitivity analysis

Parameter	Df	Sum Sq	Frac of var	Mean Sq	F value	Pr($>$F)
Relative N	1	5.04	0.017	5.04	316.48	$<$0.0001
Fraction I_0 in tropics	1	0.62	0.002	0.62	38.94	$<$0.0001
Relative R_0	1	114.49	0.406	114.49	7193.84	$<$0.0001
Relative turnover	1	2.37	0.008	2.37	148.97	$<$0.0001
Seasonal amplitude	1	94.04	0.334	94.04	5908.50	$<$0.0001
Residuals	4109	65.40	0.232	0.02		
Total	4114	281.96	1.000			

Table S3: ANOVA of the tropics' antigenic lead from multivariate sensitivity analysis

| Parameter	Df	Sum Sq	Frac of var	Mean Sq	F value	Pr(>F)
Relative N	1	1.94	0.033	1.94	754.71	<0.0001
Fraction I_0 in tropics	1	0.02	<0.001	0.02	8.75	0.0031
Relative R_0	1	44.55	0.766	44.55	17344.41	<0.0001
Relative turnover	1	1.04	0.018	1.04	406.81	<0.0001
Seasonal amplitude	1	0.02	<0.001	0.02	9.53	0.0020
Residuals	4109	10.56	0.182	0.00		
Total	4114	58.140	1.000			
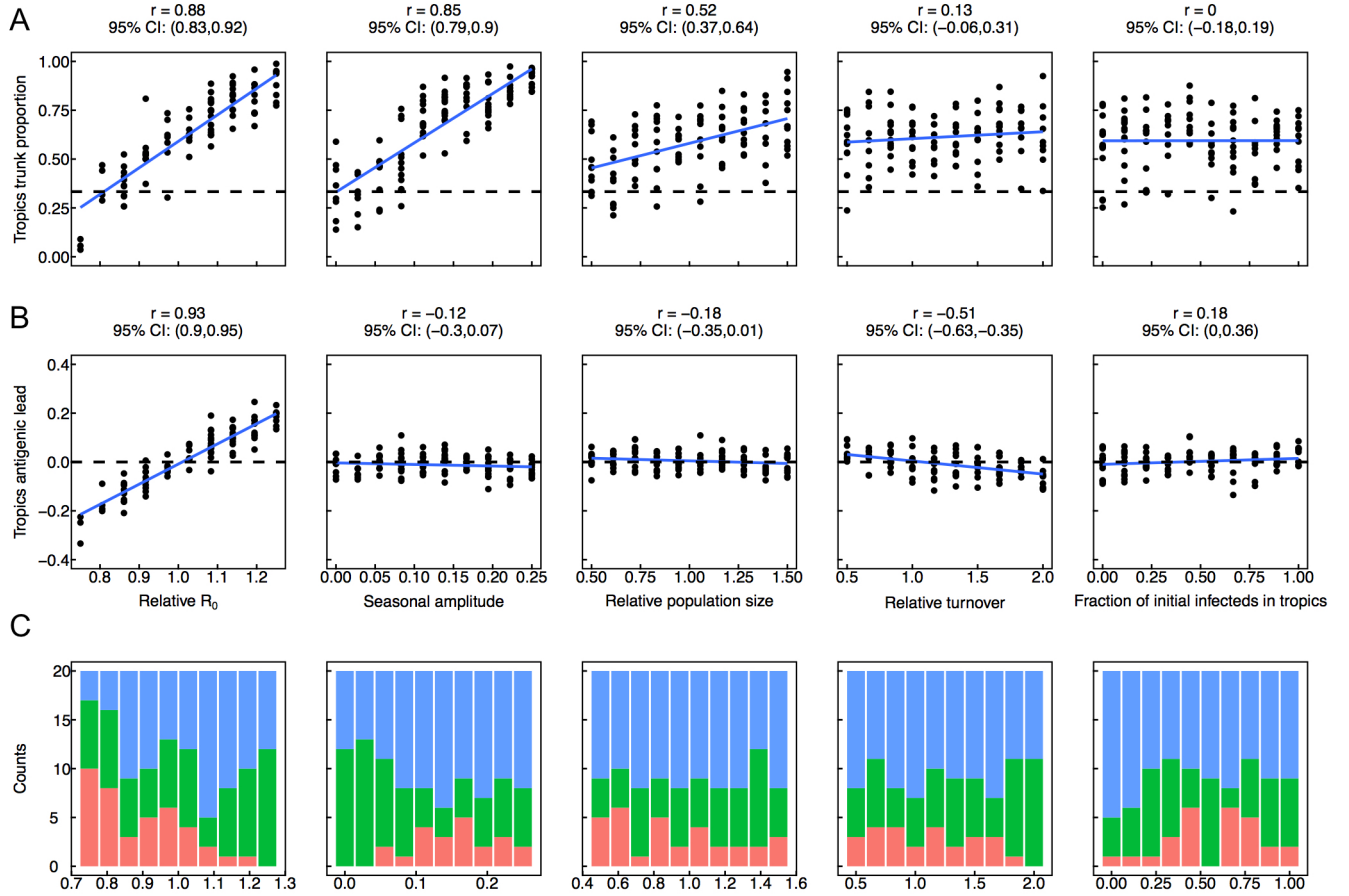
Figure S1: Univariate sensitivity analysis showing effects of individual parameters on (A) the antigenic lead and (B) the fraction of the phylogenetic trunk in the tropics. In each column of plots, only the parameter indicated on the x-axis is varying; all others are held constant at the default value. Each point represents the mean value over a single simulation. Blue lines indicate linear least squares regression. The dashed lines represent the null hypotheses where (A) the trunk is distributed equally among the three regions or (B) tropical strains are neither antigenically ahead or behind. (C) Number of simulations that went extinct (red), exceeded the TMRCA limit (green), or were suitable for analysis (blue).
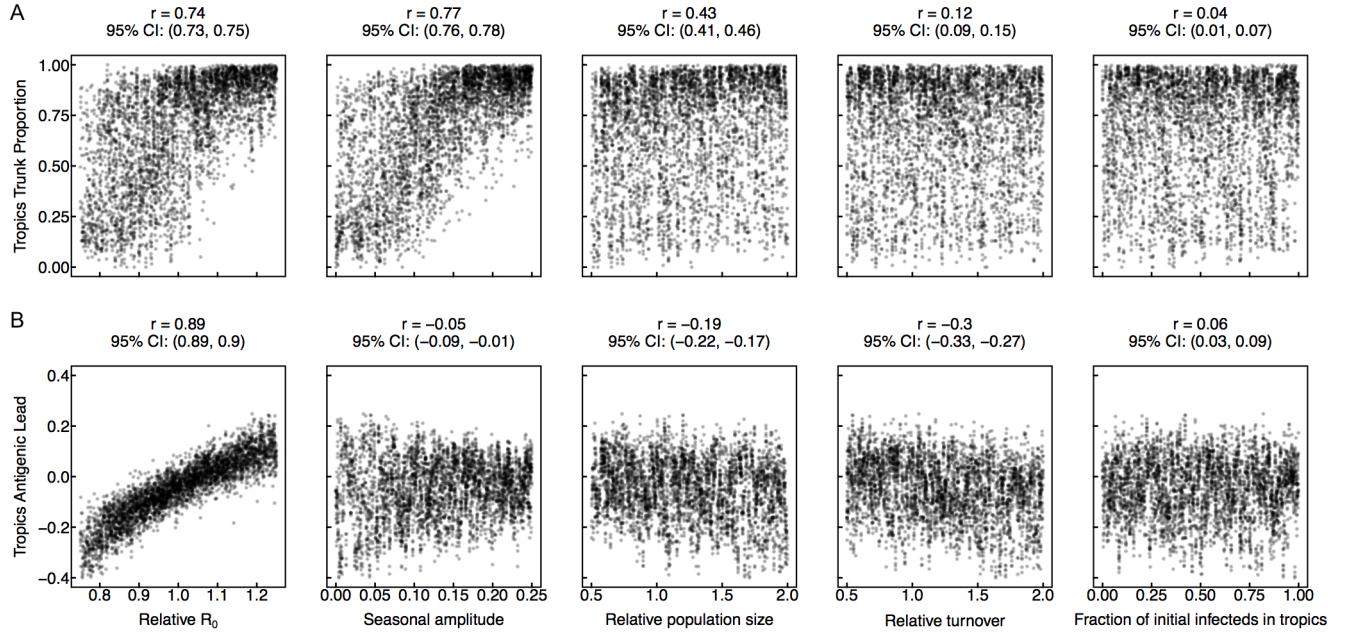
Figure S2: Multivariate sensitivity analysis showing effects of individual parameters on the (A) antigenic lead and (B) the fraction of the phylogenetic trunk in the tropics. Horizontal axes are projections of a Latin hypercube with dimensions corresponding to the five parameters indicated. Each point shows the mean value over a single simulation. The dashed lines represent the null hypotheses where (A) the trunk is distributed equally among the three regions or (B) tropical strains are neither antigenically ahead or behind. Pearson's correlation coefficients and associated 95% confidence intervals are indicated.
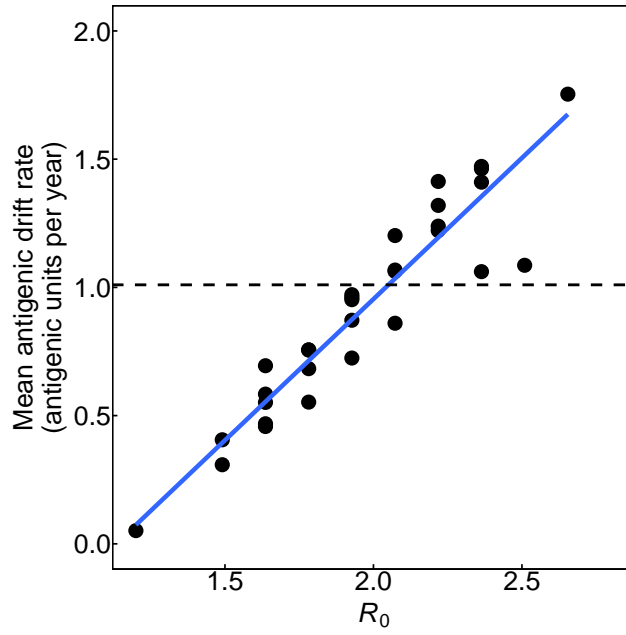
Figure S3: Effect of R_0 on antigenic drift in a single deme. Each point shows the mean antigenic drift rate from a single simulation. The blue line represents linear least squares regression, and the dashed line indicates the empirical estimate of the rate of antigenic drift for H3N2 [10]. Pearson's $r = 0.94$, $p < 0.001$; 95% CI: (0.88, 0.97).
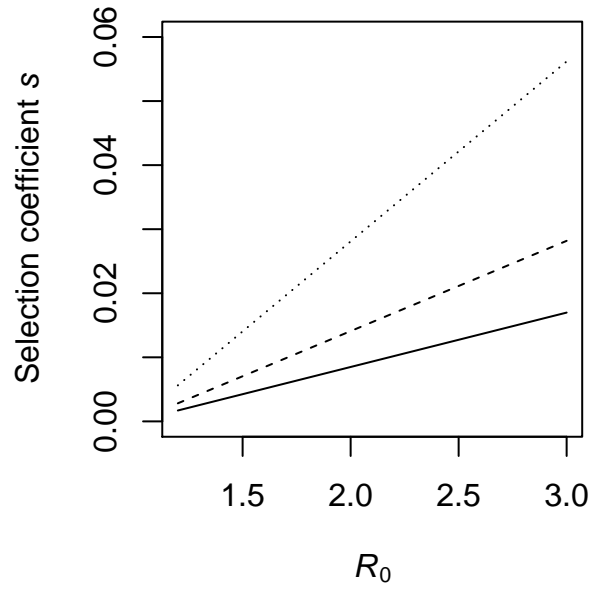
Figure S4: Relationship between R_0 and the selection coefficient for an invading strain with the resident at endemic equilibrium. The relationship for three different antigenic distances d between the invading strain and the resident strain is shown: 0.6 (solid line), 1.0 (dashed line), or 2.0 (dotted line) antigenic units.
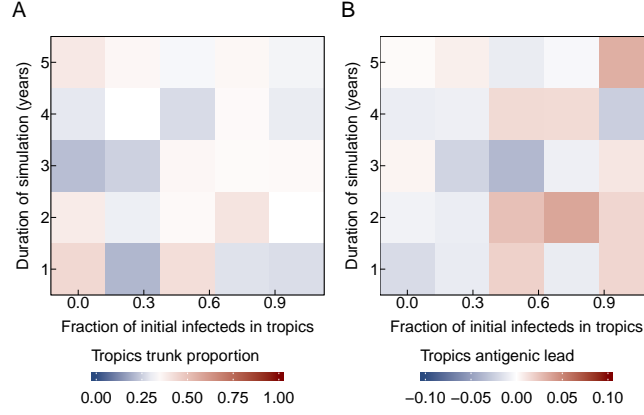
Figure S5: Effect of initial conditions on the antigenic lead and the fraction of the trunk in the tropics early in the simulation. Each square represents the average value for n=5 to 11 replicate simulations.
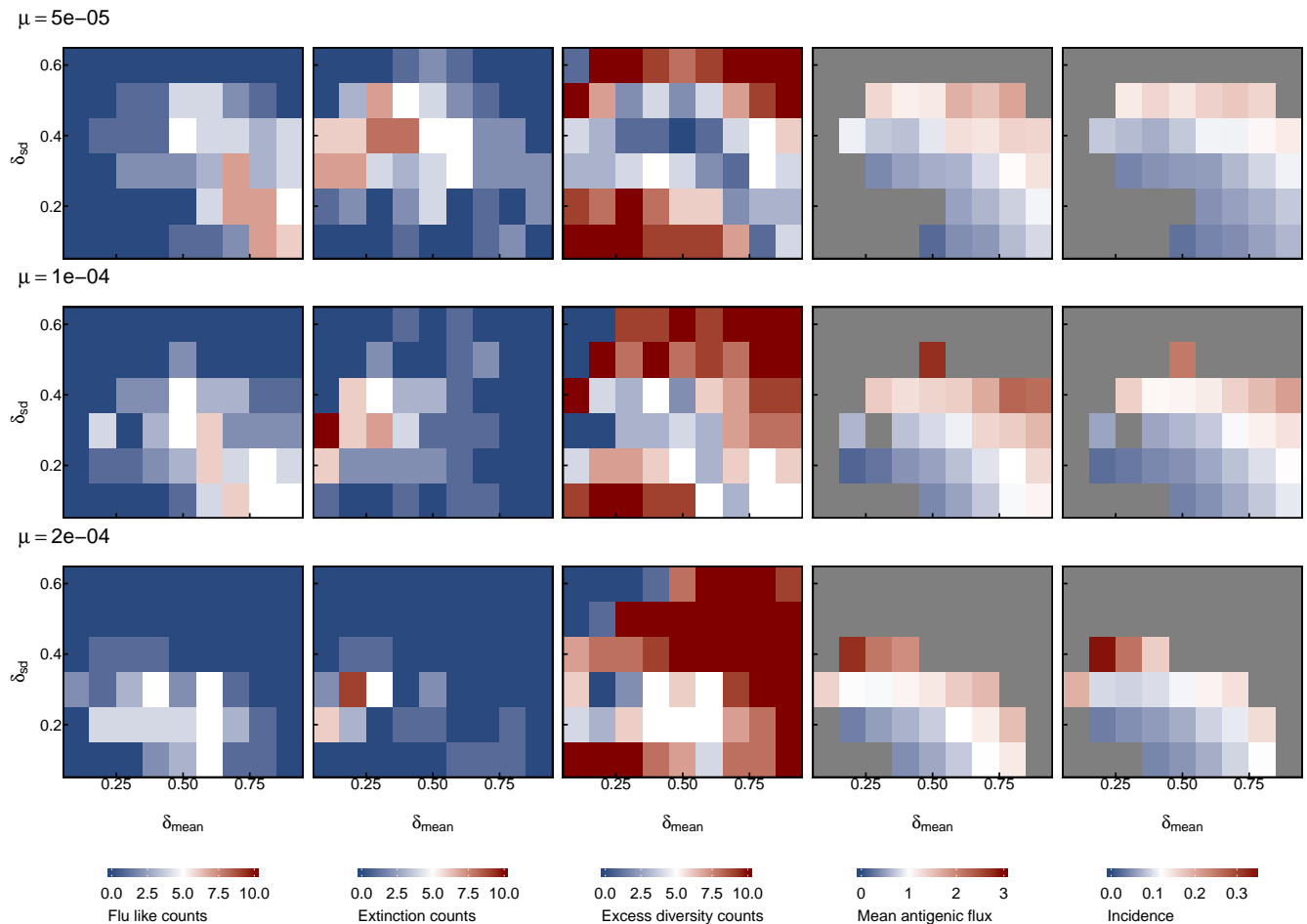
Figure S6: Sensitivity of flu-like behavior to changes in the mutational parameters, the mutation rate μ, mean mutation size δ_{mean}, and standard deviation of the mutation size δ_{sd}. Each square represents ten replicate simulations. The number of simulations where the virus went extinct is shown in the second column, and the number of simulations where the viral population exceeded a TMRCA of 10 years is shown in the third column. The remaining simulations are considered influenza-like and are shown in the first column. The reported mean antigenic drift rates and prevalences are averages over the influenza-like replicates.
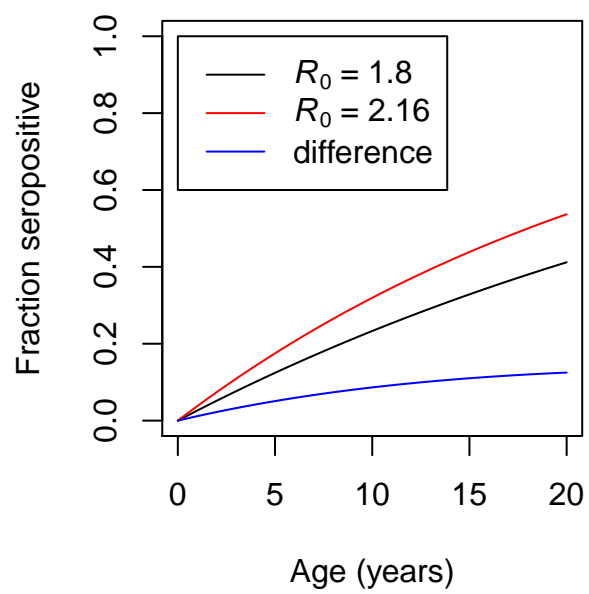
Figure S7: Theoretical increase in the fraction of seropositive individuals with age with $R_0 = 1.8$ and a 20% higher $R_0 = 2.16$.
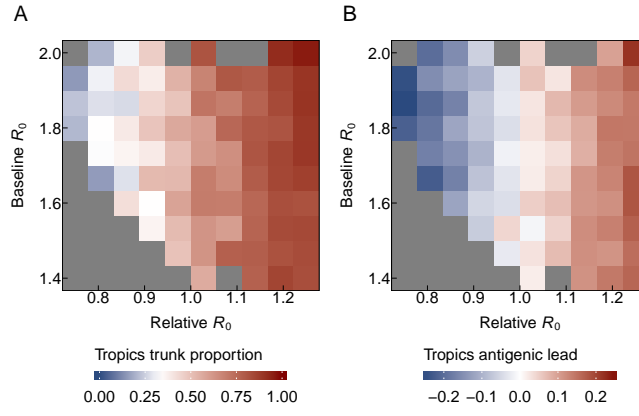
Figure S8: Lowering baseline R_0 decreases the effect of relative R_0 on the fraction of the trunk and antigenic lead in the tropics. (A) Effects of baseline and relative R_0 on the fraction of the trunk in the tropics. Blue indicates that the phylogenetic trunk is located in the tropics less than $1/3$ of the time, and red indicates that the trunk is the tropics more than $1/3$ of the time. (B) Effects of baseline R_0 and relative R_0 on antigenic lead in the tropics. Blue indicates that tropical strains are on average ahead antigenically relative to other strains, and red indicates that tropical strains are behind antigenically. Each square represents an average from 1 to 14 replicate simulations. Grey squares indicate parameter combinations where all of twenty attempted simulations either went extinct or exceeded the TMRCA threshold of 10 years.