Asynchrony between virus diversity and antibody selection limits influenza virus evolution

Dylan H. Morris[1*], Velislava N. Petrova[2], Fernando W. Rossine[1], Edyth Parker[3,4], Bryan T. Grenfell[1,5], Richard A. Neher[6], Simon A. Levin[1], and Colin A. Russell[4*]

[1]Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[2]Department of Human Genetics, Wellcome Trust Sanger Institute, Cambridge, UK
[3]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK
[4]Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, Amsterdam, NL
[5]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA
[6]Biozentrum, University of Basel, Basel, CH
[*]Correspondence to dhmorris@princeton.edu, c.a.russell@amsterdamumc.nl.

October 28, 2020

Abstract

Seasonal influenza viruses create a persistent global disease burden by evolving to escape immunity induced by prior infections and vaccinations. New antigenic variants have a substantial selective advantage at the population level, but these variants are rarely selected within-host, even in previously immune individuals. Using a mathematical model, we show that the temporal asynchrony between within-host virus exponential growth and antibody-mediated selection could limit within-host antigenic evolution. If selection for new antigenic variants acts principally at the point of initial virus inoculation, where small virus populations encounter well-matched mucosal antibodies in previously infected individuals, there can exist protection against reinfection that does not regularly produce observable new antigenic variants within individual infected hosts. Our results provide a theoretical explanation for how virus antigenic evolution can be highly selective at the global level but nearly neutral within host. They also suggest new avenues for improving influenza control.

Introduction

Antibody-mediated immunity exerts evolutionary selection pressure on the antigenic phenotype of seasonal influenza viruses (Archetti & Horsfall, 1950; Hensley et al., 2009). Influenza virus infections and vaccinations induce neutralizing antibodies that can prevent re-infection with previously encountered virus antigenic variants, but such re-infections nonetheless occur (Clements et al., 1986; Javaid et al., 2020; Memoli et al., 2019). At the human population level, accumulation of antibody-mediated immunity creates selection pressure favoring antigenic novelty. Circulating antigenic variants

typically go rapidly extinct following the population-level emergence of a new antigenic variant, at least for A/H3N2 viruses (Smith et al., 2004).

New antigenic variants like those that result in antigenic cluster transitions (Smith et al., 2004) and warrant updating the composition of seasonal influenza virus vaccines are likely to be produced in every infected host. Seasonal influenza viruses have high polymerase error rates (on the order of 10^{-5} mutations/nucleotide/replication (Nobusawa & Sato, 2006)), reach large within-host virus population sizes (as many as 10^{10} virions (Perelson et al., 2012)), and can be altered antigenically by single amino acid substitutions in the hemagglutinin (HA) protein (Koel et al., 2013; Linderman et al., 2014).

In the absence of antibody-mediated selection pressure, *de novo* generated antigenic variants should constitute a tiny minority of the total within-host virus population. Such minority variants are unlikely to be transmitted onward or detected with current next-generation sequencing (NGS) methods. But selection pressure imposed by the antibody-mediated immune response in previously exposed individuals could promote these variants to sufficiently high frequencies to make them easily transmissible and NGS detectable. The potential for antibody-mediated antigenic selection can be readily observed in infections of vaccinated mice (Hensley et al., 2009) and in virus passage in eggs in the presence of immune sera (Davis et al., 2018).

Surprisingly, new antigenic variants are rarely observed in human seasonal influenza virus infections, even in recently infected or vaccinated hosts (Debbink et al., 2017; Dinis et al., 2016; Han et al., 2019; Javaid et al., 2020; Leonard et al., 2016; McCrone et al., 2018; Valesano et al., 2019) (Fig. 1A,B). These observations contradict existing models of within-host influenza virus evolution (Luo et al., 2012; Volkov et al., 2010) and pathogen immune escape generally (Kennedy & Read, 2017), which model strong within-host antibody selection from the beginning of infection and therefore predict that new antigenic variants will be at consensus or fixation in detectable reinfections of previously immune hosts. This raises a fundamental dilemma. If within-host antibody selection is strong, why do new antigenic variants appear so rarely? If this selection is weak, how can there be protection against reinfection and resulting strong population-level selection?

We hypothesized that influenza virus antigenic evolution is limited by the asynchrony between virus diversity and antibody-mediated selection pressure. Antibody immunity at the point of transmission in previously infected or vaccinated individuals should reduce the initial probability of reinfection (Le Sage et al., 2020); secretory IgA antibodies on mucosal surfaces (sIgA) are likely to play a large role ((Wang et al., 2017), see Appendix section 2). But if viruses are not blocked at the point of transmission and successfully infect host cells, an antibody-mediated recall response takes multiple days to mount (Coro et al., 2006; Lam & Baumgarth, 2019) (see detailed review in the Appendix, section 2). Virus titer—and virus shedding (Lau et al., 2010)—may peak before the production of new antibodies has even begun, leaving limited opportunity for within-host immune selection. If antigenic selection pressure is strong at the point of transmission but weak during virus exponential growth, new antigenic variants could spread rapidly at the population level without being readily selected during the timecourse of a single infection. Moreover, prior work has established tight population bottlenecks at the point of influenza transmission (McCrone et al., 2018; Xue & Bloom,

2019). With a tight transmission bottleneck and weak selection during virus exponential growth, antigenic diversity generated during any particular infection will most likely be lost, slowing the accumulation of population-level antigenic diversity.

We used a mathematical model to investigate hypothesis that realistically timed antibody-mediated immune dynamics slow within-host antigenic evolution. We found three modeling results: 1. antibody neutralization at the point of inoculation can protect experienced hosts against reinfection and explain new antigenic variants' population level selective advantage. 2. If successful reinfection occurs, the delay between virus replication and the mounting of a recall antibody response renders within-host antigenic evolution nearly neutral, even in experienced hosts. 3. It is therefore reasonable that substantial population immunity may need to accumulate before new antigenic variants are likely to observed in large numbers at the population level, whereas effective within-host selection would predict that they should be readily observable even before they proliferate.

Our modeling results suggest a plausible mechanism that can explain otherwise poorly reconciled empirical patterns, and should motivate further experimental investigation of the mechanisms of immune protection and natural selection on influenza virus antigenic phenotypes at the point of transmission.

Model overview

Our model reflects the following biological realities: 1. Seasonal influenza virus infections of otherwise healthy individuals typically last 5–7 days (Suess et al., 2012); 2. In influenza virus-naive individuals, it can take up to 7 days for anti-influenza virus antibodies to start being produced (Wrammert et al., 2008), effectively resulting in no selection (Fig. 1A); 3. In previously infected ("experienced") individuals, sIgA antibodies can neutralize inoculated virions before they can infect host cells; (Wang et al., 2017) 4. However, if an inoculated virion manages to cause an infection in an experienced individual, it takes 2–5 days for the infected host to mount a recall adaptive immune response, including producing new antibodies (Coro et al., 2006; Zuccarino-Catania et al., 2014) (see Appendix section 2 for further discussion of motivating immunology). Importantly, this contrasts with previous within-host models of virus evolution, which have assumed that antibody-mediated neutralization of virions during virus replication is strong from the point of inoculation onward and is the mechanism of protection against reinfection (Luo et al., 2012; Volkov et al., 2010), and reflects new animal model evidence of sterilizing antibody immunity (Le Sage et al., 2020). We discuss existing models and hypotheses for the rarity of population level influenza antigenic variation in section 7 of the Appendix.

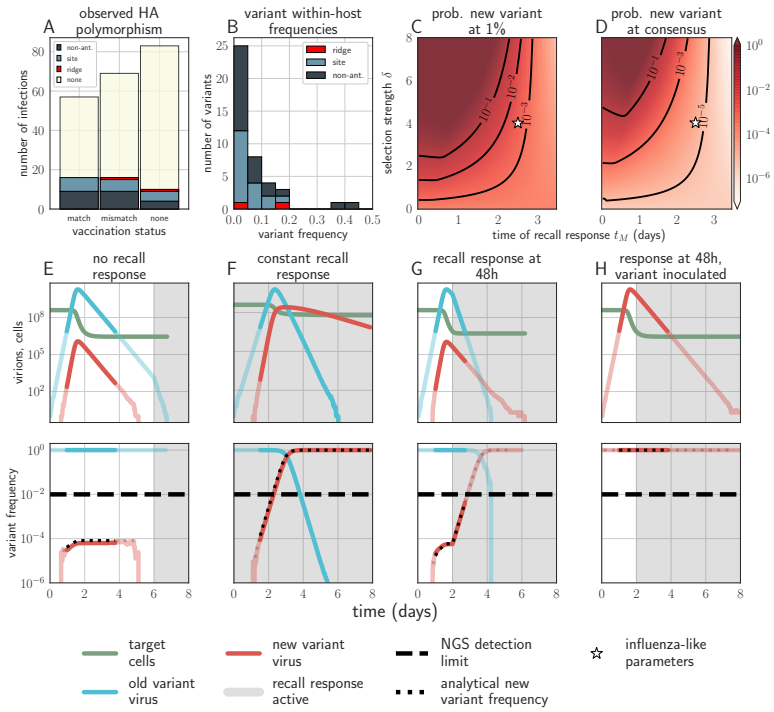

Fig. 1. Empirical within-host influenza virus variant frequencies and model within-host evolutionary dynamics. (**A, B**) meta-analysis of A/H3N2 viruses from next-generation sequencing studies of naturally infected individuals (Debbink et al., 2017; McCrone et al., 2018). (**A**) Fraction of infections with one or more observed amino acid polymorphisms in the hemagglutinin (HA) protein, stratified by likelihood of affecting antigenicity: infections with a substitution in the "antigenic ridge" of 7 key amino acid positions found by (Koel et al., 2013) in red, infections with a substitution in a classically defined "antigenic site", (Wiley et al., 1981) in blue, infections with HA substitutions only in non-antigenic regions in grey, infections with no HA substitutions in cream. Infections grouped by whether individuals had been (left) vaccinated in a year that the vaccine matched the circulating strain, (center) vaccinated in a year that the vaccine did not match the circulating strain, or (right) not vaccinated. (**B**) Distribution of plotted polymorphic sites from (**A**) by within-host frequency of the minor variant. (**C, D**) heatmaps showing model probability new antigenic variant selection to the NGS detection threshold of 1% (**C**) and to 50% (**D**) by 3 days post infection given the strength of immune selection δ, the antibody response time t_M and an old variant founding population; calculated from equation 27 in the Methods. Calculated with $c_w = 1, c_m = 0$, but for $t_M > 1$, replication selection probabilities are approximately equal for all c_w, c_m, k trios that yield a given δ (see Methods). Star denotes a plausible influenza-like parameter regime: 25% escape from sterilizing-strength immunity ($c_w = 1, c_m = 0.75, k = 20$) with a recall response at 2.5 days post infection. Black lines are probability contours. (**E–H**) example model trajectories. Upper row: absolute counts of virions and target cells. Lower row: variant frequencies for old antigenic variant (blue) and new variant (red). Dashed line shows 1% frequency, the detection limit of NGS. Dotted line shows an analytical prediction for new variant frequency according to equations 15 and 16 (see Methods). Model scenarios: (**E**) naive; (**F**) experienced with $t_M = 0$; (**G**) experienced with $t_M = 2$; (**H**) experienced with $t_M = 2$ and new antigenic variant virion incoulated. Lines faded when infection is below 5% transmission probability—approximately 10^7 virions with default parameters. All parameters as in Table 1 unless otherwise stated.

We constructed a model of within-host influenza virus evolution that can be parameterized to reflect different hypothesized immune mechanisms, different host immune statuses, and different durations of infection. In the model, virions V_i of antigenic type i infect target cells C, replicate, mutate to a new antigenic type j at a rate μ_{ij}, and decay at a rate d_v. We model the innate immune response implicitly as depletion of infectible cells. We model the antibody-mediated immune response as an increase k in the virion decay rate in the presence of well-matched antibodies. To model partial antibody cross reactivity, we scale k by a parameter $c_i \in [0, 1]$; c_i reflects the binding strength of the host's best-matched antibodies to antigenic type i. So in the presence of an antibody-mediated response, virions V_i of type i decay at a rate $d_v + c_i k$.

The model can accommodate N_v antigenic variants $i = 1, 2, ...N_v$ linked by an arbitrary network of possible substitutions and corresponding mutation probabilities μ_{ij}, but in practice we typically consider two, the new variant m and the old variant w, and neglect back-mutation from new variant to old variant ($\mu_{wm} > 0$, but $\mu_{mw} = 0$).

To assess the importance of transmission bottlenecks, initial virus diversity, and sIgA antibody neutralization in virus evolution, we model the point of transmission as a series of stochastic events which may ultimately lead to one of more virions invading cells and initiating an infection. The recipient host is inoculated with a random sample of within-host virus diversity from a transmitting host. In experienced hosts, this inoculum is probabilistically thinned by host antibodies. The founding population that initiates the infection is then randomly sampled from among any remaining virions.

Mathematically, we model the number of inoculated virions as Poisson-distributed with a mean v, so if variant i has within-host frequency f_i, the number of variant i virions inoculated is Poisson-distributed with mean vf_i. The virions then encounter antibodies, which we interpret as sIgA but can be understood to be any antigen-specific antibody-mediated protection that precedes cell infection; each virion of variant i is independently neutralized with a probability κ_i. This probability depends upon the strength of protection against homotypic reinfection κ and the sIgA cross immunity between variants σ_{ij}, $(0 \leq \sigma_{ij} \leq 1)$. So if a host with antibodies to variant j is challenged with variant i, those virions will be neutralized with a probability $\kappa_i = \kappa\sigma_{ij}$. For simplicity, we assume the same homotypic protection level across all variants and hosts, though in practice there may be variation in the immunogenicity of individual variants and in the strength of responses generated by individual hosts. We typically fix host immune histories to test the effect of host immune history on selective dynamics. When necessary, we can model a novel (non-recall) antibody response to a strain i by designating the host as experienced to i at some time t_N post-infection (see Methods).

The model is continuous-time and stochastic: cell infection, virion production, virus mutation, and virion decay are stochastic events that occur at rates governed by the current state of the system, with exponentially distributed (memoryless) waiting times. The system is approximately well-mixed: we track counts of virions and cells without an explicit account of space within the upper respiratory tract. We treat infected and dead or removed cells implicitly.

Parameterized as in Table 1, the model captures key features of influenza infections: rapid peaks approximately 2 days post infection and slower declines with clearance approximately a week post infection, with faster clearance in experienced hosts.

We give a full mathematical description of the model in the Methods.

In addition to analyzing this within-host model, we explored the between-host and population-level implications of these within-host dynamics using a simple transmission chain model and an SIR-like population-level model, which we describe in the Methods.

Results

Realistically-timed immune kinetics limit otherwise rapid adaptation during exponential growth

In our model, sufficiently strong antibody neutralization during the virus exponential growth period can potentially stop replication and block onward transmission, but this mechanism of protection results in detectable new antigenic variants in each homotypic reinfection, since the infection is terminated rapidly unless it generates a new antigenic variant that substantially escapes antibody neutralization (Fig. 2).

If there is antibody neutralization throughout virus exponential growth and it is not sufficiently strong to control the infection, this facilitates the establishment of new antigenic variants: variants can be generated *de novo* and then selected to detectable and easily transmissible frequencies (Fig. 1C,D,F, sensitivity analysis in Fig. A3). We term selection on a replicating within-host virus population "replication selection". Virus phenotypes that directly affect fitness independent of immune system interactions are likely to be subject to replication selection.

Adding a realistic delay to antibody production of two days post-infection (Lam & Baumgarth, 2019) curtails antigenic replication selection. There is no focused antibody-mediated response during the virus exponential growth phase, and so the infection is dominated by old antigenic variant virus (Fig. 1C,D,G, sensitivity analysis in Fig. A3). Antigenic variant viruses begin to be selected to higher frequencies late in infection, once a memory response produces high concentrations of cross-reactive antibodies. But by the time this happens in typical infections, both new antigenic variant and old antigenic variant populations have peaked and begun to decline due to innate immunity and depletion of infectible cells, so new antigenic variants remain too rare to be detectable with NGS (Fig. 1G).

We find that replication selection of antigenic novelty to detectable levels becomes likely only if infections are prolonged, and virus antigenic diversity and antibody selection pressure therefore coincide (Fig. 1G, see also Fig. 7). This can explain existing observations: within-host adaptive antigenic evolution can be seen in prolonged infections of immune-compromised hosts (Xue et al., 2017), and prolonged influenza infections show large within-host effective population sizes (Lumby et al., 2020).

Neutralization of virions at the point of transmission provides host protection and population-level selection without rapid within-host adaptation

Adding antibody neutralization of virions (e.g. by mucosal sIgA) at the point of inoculation to our model produces realistic levels of protection against reinfection, and

when reinfections do occur, they are overwhelmingly old antigenic variant reinfections. New antigenic variants that arise during these infections remain undetectably rare, reproducing observations from natural human infections (Fig. 1A–D, G–H, Fig. 3B–C)(Debbink et al., 2017; Dinis et al., 2016; Leonard et al., 2016; McCrone et al., 2018). The combination of mucosal sIgA protection and a realistically timed antibody recall response explains how there can exist immune protection against reinfection—and thus a population level selective advantage for new antigenic variants—without observable within-host antigenic selection in typical infections of experienced hosts.

Tight bottlenecks lead to loss of generated diversity and mean new variants reach consensus through founder effects

Regardless of host immune status, an antigenic variant that has been generated *de novo* within a host must survive a series of population bottlenecks if it is to infect other individuals. To found a new infection, virions must be expelled from a currently infected host (excretion bottleneck), must enter another host (inter-host bottleneck), must escape mucus on the surface of the airway epithelium (mucus bottleneck), must avoid neutralization by sIgA antibodies on mucosal surfaces (sIgA bottleneck), and must infect a cell early enough to form a detectable fraction of the resultant infection (cell infection bottleneck) (Fig. 3A). The sum of all of these effects is the net bottleneck and typically results in infections being initiated by a single genomic variant (Ghafari et al., 2020; McCrone et al., 2018; Xue & Bloom, 2019). That said, bottlenecks resulting from direct contact transmission may be substantially wider than those associated with respiratory droplet or aerosol transmission (Varble et al., 2014) and more human studies are required to quantify these differences.

We find that because antigenic variants appear at very low within-host frequencies when generated *de novo* and undergo minimal or no replication selection, new antigenic variants most commonly reach detectable levels within hosts through founder effects at the point of inter-host transmission: a low-frequency antigenic variant generated in one host survives the net bottleneck to found the infection of a second host (Fig. 3D).

Given that influenza bottlenecks are thought to be on the order of a single virion (McCrone et al., 2018), any replication-competent mutant that founds an infection should occur at NGS-detectable levels, and likely at consensus. But for the same reason, these founder effects are rare events. These founder effects could be a purely neutral sampling process. It is likely quite close to neutral in a truly naive host who does not possess well-matched antibodies to the infecting old variant: all inoculated virions, regardless of antigenic phenotype, have an equal chance of becoming part of the new infection's founding population. If there are v virions that compete to found the infection and $b = 1$, then each virion founds the infection with probability $1/v$.

In our model, new antigenic variants survive the transmission bottleneck upon inoculation into a naive host with a probability approximately equal to the donor host variant frequency f_{mt} times the bottleneck size b (see Methods). New antigenic variant infections of naive hosts should therefore occur on the order of 1 in 10^5 or 1 in 10^4 of such infections given biologically plausible parameters (Fig. 3E–H).

But if different virions have different chances of being neutralized at the point of transmission, the founding process may be selective. Among the virions that encounter

antibodies, those that are less likely to be neutralized have a higher than average chance of undergoing stochastic promotion to consensus while those that are more likely to be neutralized have a lower than average chance. A new antigenic variant is then disproportionately likely to survive the net bottleneck (Figs. 1, 3, Appendix section 4.2). We term this potential selection on inoculated diversity "inoculation selection". Neutralization at the point of transmission thus not only gives new antigenic variant infections their transmission advantage (population-level selection) but may also increase the rate at which these new antigenic variant infections arise (inoculation selection).

There is some suggestive evidence of differential survival of particular (not necessarily antigenic) influenza genetic variants at the point of transmission from experiments in ferrets (Moncla et al., 2016; Wilker et al., 2013). But as Lumby and colleagues (2018) point out in a reanalysis of those experiments, it difficult empirically to distinguish selection that occurs at the point of transmission from selection that occurs during replication in the donor host because of the challenges associated with sampling the small virus populations present at the earliest stages of infection. Here we define inoculation selection as selection on the bottlenecked virus population that establishes infection in the recipient host before any virus replication has taken place in that host.

Inoculation selection depends on degree of founding competition and degree of immune escape

The strength of inoculation selection depends on the ratio of the number of virions that compete to found an infection in the absence of well-matched sIgA antibodies (the mucus bottleneck size v) to the number of virions that actually found an infection (the final cell infection bottleneck size b). The larger this v/b ratio is, the more inoculation selection in experienced hosts facilitates the survival of new antigenic variants (Fig. 3F, Fig. A1).

When new antigenic variant immune escape is incomplete due to partial cross-reactivity with previous antigenic variants, increased antibody neutralization is a double-edged sword for new antigenic variant virions. Competition to found the infection from old antigenic variant virions is reduced, but the new antigenic variant is itself at greater risk of being neutralized. The impact of inoculation selection therefore depends on the degree of similarity between previously encountered viruses and the new antigenic variant. An experienced recipient host could facilitate the survival of large-effect antigenic variants (like those seen at antigenic cluster transitions (Smith et al., 2004)) while retarding the survival of variants that provide less substantial immune escape (Fig. 3E, Fig. A1).

Inoculation selection is limited by the low frequency f_{mt} of new antigenic variants in transmitting donor hosts (due to weak replication selection), the potentially small mucus bottleneck size v, and the fact that some hosts previously infected with the same or similar antigenic variants might not possess well-matched antibodies due to original antigenic sin (Davenport & Hennessy, 1956), antigenic seniority (Lessler et al., 2012), immune backboosting (Fonville et al., 2014), or other sources of individual-specific variation in antibody production (Lee et al., 2019). These factors combined make selection and onward transmission of new variants rare.

Immune hosts can facilitate the appearance of new variants without producing rapid diversification

Onward transmission of new variants can be facilitated by natural selection—replication selection, inoculation selection or both. The degree of facilitation depends principally on four quantities: 1. $\delta\tau$, the product of the replication fitness difference $\delta = k(c_w - c_m)$ and the time under replication selection τ. This determines the degree to which the new variant is promoted by replication selection prior to transmission (increasing f_{mt}). 2. κ_w, the neutralization probability for the old variant, which must be large enough to reduce competition for the final bottleneck. 3. v/b, the ratio of the sIgA bottleneck size v to the cell infection (final) bottleneck size b. This determines the degree of competition to found the infection, and thus sets maximum potential strength of inoculation selection when κ_w is large: a $\frac{v}{b}$-fold improvement over drift for small b. 4. $1 - \kappa_m$, how likely the new variant is to avoid neutralization at the point of transmission; this scales down the maximum replication selective strength set by v/b (to zero if the new variant is neutralized with certainty; $\kappa_m = 1$). When $\delta\tau$ is small and κ_w, v/b, and $1 - \kappa_m$ are large, new variant survival is most facilitated by inoculation selection. When the opposite is true, replication selection is most important. And there are parameter regimes in which both replication and and inoculation selection provide a substantial improvement over drift (Fig. 4, see Appendix section 4.6 for mathematical intuition for these results).

At realistic parameter values and assuming all individuals develop well-matched antibodies to previously encountered antigenic variants, only $\sim$1 to 2 in 10^4 inoculations of an experienced host results in a new antigenic variant surviving the bottleneck (Fig. 3E,H, Fig. 4). This rate is likely to be an overestimate due to the factors mentioned above. Moreover, it is only about 2- to 3-fold higher than the rate of bottleneck survival in naive hosts, where new antigenic variant infections should occasionally occur via neutral stochastic founder effects. In short, even in the presence of experienced hosts, antigenic selection is inefficient and most generated antigenic diversity is lost at the point of transmission. Because of these inefficiencies, new antigenic variants can be generated in every infected host without producing explosive antigenic diversification at the population level.

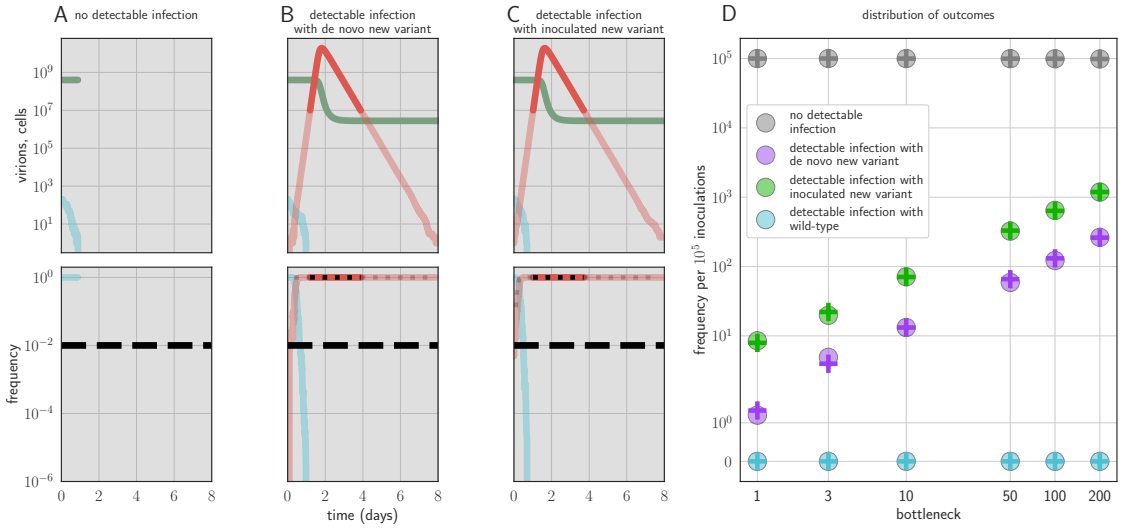

Fig. 2. Example timecourses and distribution of outcomes when antibody immunity is active from the start of infection and sufficient to prevent detectable reinfection. $t_M = 0$, $k = 20$, yielding $\mathcal{R}^w(t = 0) < 1$ for the old antigenic variant but $\mathcal{R}^m(t = 0) > 1$ for the new antigenic variant, where $\mathcal{R}^i(t)$ is the within-host effective reproduction number for variant i at time t (see Methods). No mucosal antibody neutralization ($z_w = z_m = 0$); protection is only via neutralization during replication Example timecourses from simulations with founding population (bottleneck) $b = 200$. Since neutralization during replication takes the place of mucosal sIgA neutralization, b should be understood as comparable to the sIgA bottleneck size v in a model with sIgA neutralization. (**A**–**C**) Top panels: absolute abundances of target cells (green), old antigenic variant virions (blue), and new antigenic variant virions (red). Bottom panels: frequencies of virion types. Black dotted line is an analytical prediction for the new antigenic variant frequency given the time of first appearance. Black dashed line is the threshold for NGS detection. (**D**) Probabilities of no infection, *de novo* new antigenic variant infection, inoculated new antigenic variant infection, and old antigenic variant infection per 10^5 inoculations of an immune host by a naive host. Circles are frequencies from simulation runs (10^6 runs for bottlenecks 1–10, 10^5 runs for bottlenecks 50–200). Plus-signs are analytical model predictions for probabilities (see Appendix section 3.6), with f_{mt} set equal to the average from donor host stochastic simulations for the given bottleneck. Parameters as Table 1 unless otherwise stated.

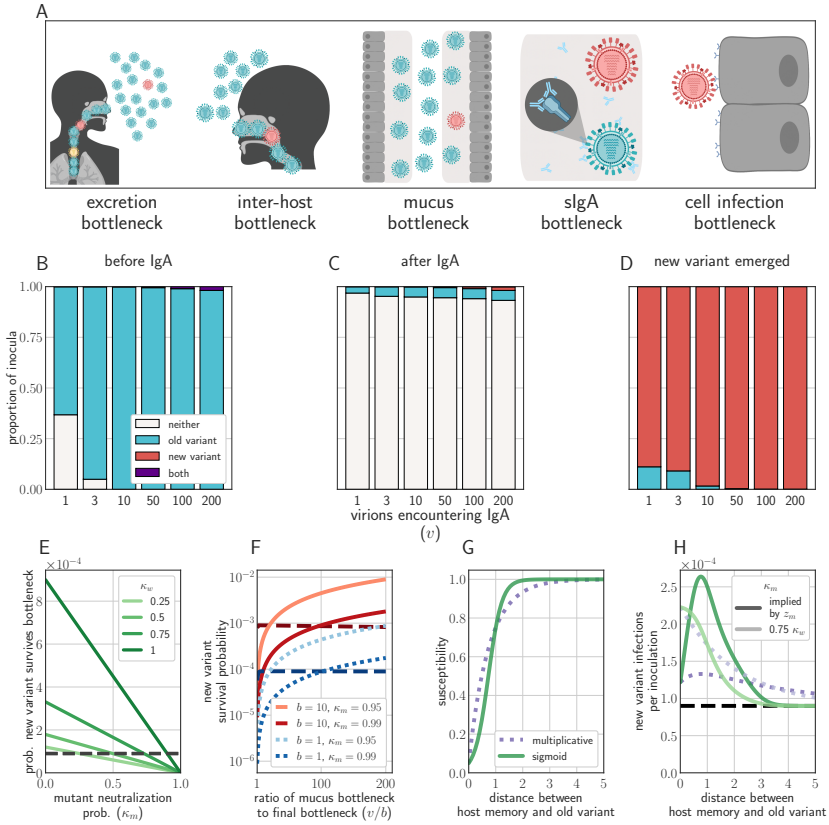

Fig. 3. Selection for antigenic variants at the point of transmission (inoculation selection). (**A**) Schematic of bottlenecks faced by both old antigenic variant (blue) and new antigenic variant (other colors) virions at the point of virus transmission. Key parameters for inoculation selection are the mucus bottleneck v—the mean number of virions that encounter sIgA—and the cell infection bottleneck size b. (**B–D**) Effect of sIgA selection at the point of inoculation with $b = 1$. (**B, C**) Analytical model distribution of virions inoculated into an immune host immediately before (**B**) and after (**C**) mucosal neutralization / the sIgA bottleneck. f_{mt} set to mean of stochastic simulations. (**D**) Distribution of founding virion populations (after the cell infection bottleneck) among individuals who developed detectable new antigenic variant infections in stochastic simulations. (**E–F**) Analytical model probability that a variant survives the final bottleneck. Dashed horizontal lines indicate probability in naive hosts. $f_{\mathrm{mt}} = 9 \times 10^{-5}$ (the approximate mean in stochastic simulations for $b = 1$). (**E**) Variant probability of surviving all bottlenecks, as a function of old antigenic variant neutralization probability κ_w and new antigenic variant mucosal neutralization probability κ_m. (**F**) New antigenic variant survival probability as a function of the ratio of v to b. (**G, H**) Effect of host susceptibility model on the appearance of antigenic novelty. Per inoculation rates of new variants surviving the bottleneck (**H**) depend on host immune status and on the relationship between virus antigenic phenotype and host susceptibility $(1 - z_w)$ (**G**). Plotted with $f_{\mathrm{mt}} = 9 \times 10^{-5}$, and 25% susceptibility to (75% protection against, $z_1 = 0.75$) a variant one antigenic cluster away from host memory. Unless noted, parameters for all plots as in Table 1.

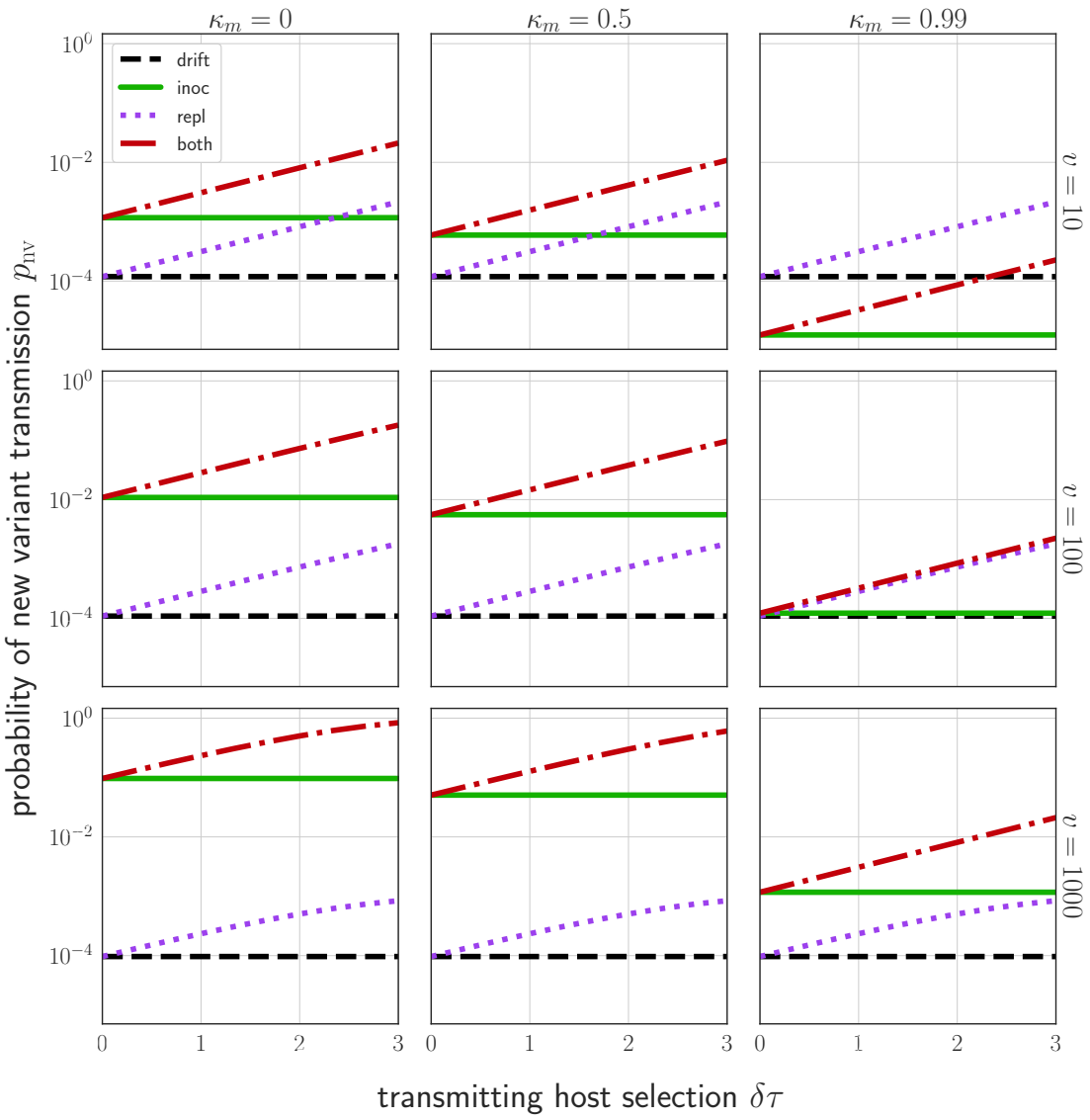

Fig. 4. Probability $p_{\mathbf{nv}}$ of a new variant surviving the transmission bottleneck as a function of donor host replication selection and recipient host inoculation selection. Calculated according to equation 43, and plotted as a function of degree of replication selection in the donor host $\delta\tau$, the product of the selection strength δ and the time duration $\tau = \max 0, t_t - t_M$ between the onset of the antibody response at t_M and the transmission event at t_t. Black dashed line: neutral (drift) expectation, where $\delta = 0$ in the donor host and the recipient host does not neutralize either the old or the new variant at the point of transmission ($\kappa_w = \kappa_m = 0$). Purple dotted line: replication selection only: $\delta\tau$ as given in the donor host, but a naive recipient host. Green solid line: inoculation selection only: $\delta = 0$ in the donor host, but a recipient host with well-matched antibodies to the old variant ($\kappa_w = 1$), with varying degrees of immune escape (κ_m as given in the columns). Red dot-dashed line: combination of both replication selection in the donor host as before and inoculation selection in the recipient host as before. Plotted with a final bottleneck of $b = 1$ and an sIgA bottleneck v as given in the rows. Parameters as in Table 1 unless otherwise noted.

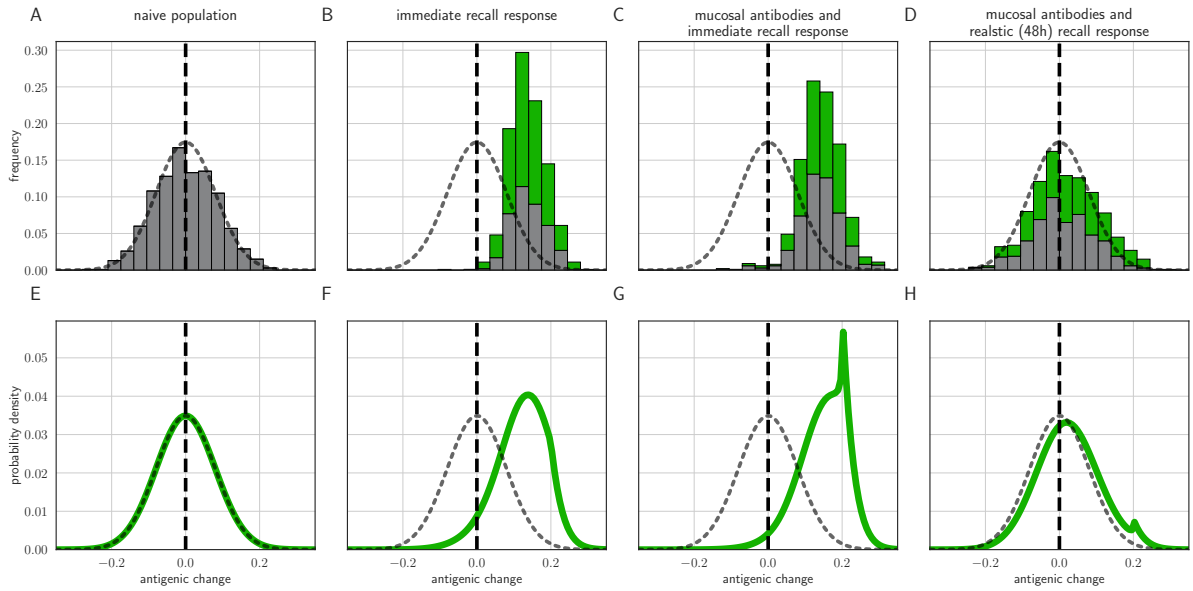

Fig. 5. Distribution of mutant effects given replication and inoculation selection. Distribution of antigenic changes along 1000 simulated transmission chains (**A**–**D**) and from an analytical model (**E**–**H**). In (**A**,**E**) all naive hosts, in other panels a mix of naive hosts and experienced hosts. Antigenic phenotypes are numbers in a 1-dimensional antigenic space and govern both sIgA cross immunity σ and replication cross immunity c. A distance of ≥ 1 corresponds to no cross immunity between phenotypes and a distance of 0 to complete cross immunity. Gray line gives the shape of Gaussian within-host mutation kernel. Histograms show frequency distribution of observed antigenic change events and indicate whether the change took place in a naive (grey) or experienced (green) host. In (**B**–**D**) distribution of host immune histories is 20% of individuals previously exposed to phenotype -0.8, 20% to phenotype -0.5, 20% to phenotype 0 and the remaining 40% of hosts naive. In (**E**), a naive hosts inoculate naive hosts. In (**F**–**H**) a host with history -0.8 inoculate hosts with history -0.8. Initial variant has phenotype 0 in all sub-panels. Model parameters as in Table 1, except $k = 25$. Spikes in densities occur at 0.2 as this is the point of full escape in a host previously exposed to phenotype -0.8.

Inoculation selection produces realistically noisy between-host evolution

To investigate the between-host consequences of adaptation given weak replication selection, tight bottlenecks, and possible inoculation selection, we simulated transmission chains according to our within-host model, allowing the virus to evolve in a 1-dimensional antigenic space (Bedford et al., 2012; Smith et al., 2004) until a generated antigenic mutant becomes the majority within-host variant. When all hosts in a model transmission chain are naive, antigenic evolution is non-directional and recapitulates the distribution of within-host mutations (Fig. 5A). When antigenic selection is constant throughout infection and even a moderate fraction of hosts are experienced, antigenic evolution is unrealistically strong: the virus evolves directly away from existing immunity and large-effect antigenic changes are observed frequently (Fig. 5B,C). When the model incorporates both mucosal antibodies and realistically timed recall responses, major antigenic variants appear only rarely and the overall distribution of emerged variants better mimics empirical observations (Fig. 5D); most notably, the phenomenon of quasi-neutral diversification within an antigenic cluster seen in Figures 1 and 2 of Smith et al., 2004. A simple analytical model (see Methods) in which generated antigenic mutants fix according to their replication and inoculation selective advantages also displays this behavior (Fig. 5B–H).

In particular, we note that whereas an immediate recall response would predict strong near-constant directed evolution of virus antigenic phenotypes away from existing immunity (Fig. 5B,C), a realistically-timed recall response predicts that small-effect, drift-like, antigenic substitutions will be observed. Even substitutions that move a virus "backward" in antigenic space and that should be more readily neutralized by existing antibodies than the ancestral variant should be observed thanks to the large role of stochasticity at the point of transmission. That said, there is a slight bias favoring forward substitutions, especially those of sufficiently large effect (Fig. 5D) to create a substantial selective advantage over the ancestral variant. Coupled with the plausible assumption that large effect substitutions are rarer than small effect substitutions (here captured qualitatively by the Gaussian mutation kernel), this predicts the pattern of quasi-neutral diversification within antigenic clusters followed by rarer directional "jumps" in phenotype. Exact rates of antigenic evolution will depend upon how these emergence processes intersect with population-level epidemic dynamics and competition among variants.

Epidemic dynamics can alter rates of inoculation selection

We used an epidemic-level model to study the consequences of individual-level inoculation selection for population-level antigenic selection. If inoculation selection is efficient, an intermediate initial fraction of immune hosts maximizes the probability that a new antigenic variant infection is created during an epidemic (Fig. 6.) This is due to a trade-off between the frequency of naive or weakly immune "generator" hosts who can propagate the epidemic and produce new antigenic variants through *de novo* mutation, and the frequency of strongly immune "selector" hosts who, if inoculated, are unlikely to be infected, but can facilitate the survival of these new antigenic variants at the point of transmission. As selector host frequency increases, epidemics become rarer and smaller, eventually decreasing opportunities for evolution, but moderate numbers of

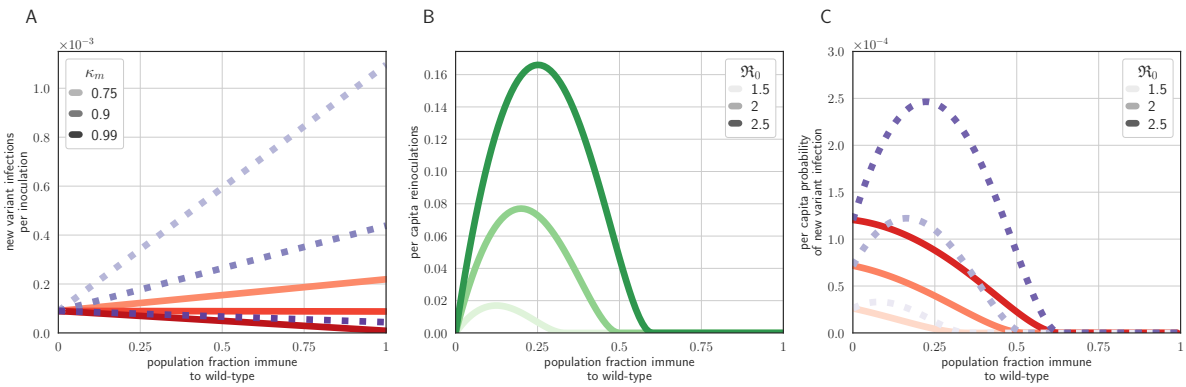

Fig. 6. Population level antigenic dynamics resulting from inoculation selection. Analytical model results (see Methods) for population-level inoculation selection, using parameters in Table 1 and $f_{\mathrm{mt}} = 9 \times 10^{-5}$ unless otherwise stated. (**A**) Probability per inoculation of a new antigenic variant founding an infection, as a function of fraction of hosts previously exposed to the infecting old antigenic variant virus, mucus bottleneck size v and sIgA cross immunity $\sigma = \kappa_m/\kappa_w$. Red solid lines: $v = 10$. Purple dotted lines: $v = 50$. (**B**) Expected per capita reinoculations of previously exposed hosts during an epidemic, given the fraction of previously exposed hosts in the population, if all hosts that were previously exposed to the circulating old antigenic variant virus are fully immune to that variant, for varying population level basic reproduction number $\mathfrak{R}_0$. (**C**) Probability per individual host that a new antigenic variant founds an infection in that host during an epidemic, as a function of the fraction of hosts previously exposed to the old antigenic variant. Other hosts naive. $\sigma = 0.75$. Red solid lines: $v = 10$; purple dotted lines: $v = 50$ (as in **A**).

efficient selectors can substantially increase the rate at which new antigenic variants reach within-host consensus (Fig. 6B,C).

Discussion

Any explanation of influenza virus antigenic evolution—and why it is not even faster—must explain why population level antigenic selection is strong, as evidenced by the typically rapid sequential population-level replacement of old antigenic variants upon the emergence of a new major antigenic variant, but within-host antigenic evolution is rarely observed.

We hypothesized that antibodies present in the repiratory tract mucosa at the time of virus exposure can effectively block transmission, but have only a small effect on viral replication once cells become productively infected. Antigenic selection after successful infection therefore begins with the mounting of a recall response 48–72 hours post infection. In this case, selection pressure can be strong at the point of transmission, but subsequently weak until after the period of virus exponential growth. This mechanistic paradigm reconciles strong but not perfect sterilizing homotypic immunity with rare observations of new antigenic variants in successfully reinfected experienced hosts.

Alternative explanations for rare new antigenic variants

We consider several possible explanations for the observed phenomenon that new antigenic variants are rare within experienced infected hosts and at the population level prior to cluster transitions. But among these candidate hypotheses, only the mechanism of small inocula, transmission-blocking mucosal antibodies, but a slow-to-mount recall adaptive immune response can explain all the aforementioned empirical observations simultaneously.

Alternative possibilities include strong immune protection through antibody neutralization early during early viral replication (an immediate recall response), heterogeneous neutralization rates during early viral replication, new antigenic variants that are deleterious in the absence of antibody selection, and the need for new variants to emerge against a favorable genetic background.

As shown above, protection through antibody neutralization early in replication can result in rare within-host observation of new antigenic variants, but it contradicts understanding of antibody kinetics and makes other empirical predictions that are unrealistic. It implies that homotypic challenge has a binary outcome: either it results in an undetectable infection that is rapidly cleared or it results in a visible infection dominated by an escape mutant (Fig. 2). This is how influenza viruses have been hypothesized to behave in previous models of immune escape (Luo et al., 2012). But empirical work (Debbink et al., 2017; Javaid et al., 2020; McCrone et al., 2018) and human challenge studies (Clements et al., 1986; Memoli et al., 2019) have shown that detectable reinfection of experienced hosts can occur without observable immune escape. Another empirical prediction of such a model is that intermediately immune hosts should be efficient selectors, since they neutralize the old variant virus poorly enough to allow it to grow, but strongly enough to impose antigenic selection upon that growing population (Volkov et al., 2010). While such intermediately immune hosts should be present from the beginning of a new antigenic cluster's circulation (Fonville et al., 2014), new variants are rarely observed until a new variant has circulated for multiple years (Smith et al., 2004).

Antibody neutralization during early replication could avoid binary outcomes if individual hosts are sufficiently heterogeneous in the strength of their neutralizing response, so that some individuals clear the infection rapidly while others barely exert antibody selection upon it. But while heterogeneity in immunity exists (Lee et al., 2019), this explanation requires extreme, bimodal hetegoneity to avoid the intermediately immunity regime in which replication selection is efficient (Volkov et al., 2010) and again requires an unrealistically early antibody response.

New antigenic variants could be replication-competent, but weakly deleterious within-host in the absence of immune selection and/or compensatory substitutions. Two studies (Gog, 2008; Kucharski et al., 2015) have invoked this hypothesis to explain population-level antigenic dynamics. However, a realistically strong antibody response during virus replication could still promote new variants during infections of experienced hosts, even in the absence of compensatory mutations (see Appendix section 7.4 for a model analysis). So a hypothesis to explain the relative weakness of selection during replication is still required, especially for weakly deleterious mutants that offer substantial immune escape. That said, under our hypothesis of founder effects

and possible inoculation selection, weak new variant deleteriousness could further limit the rate of antigenic evolution by reducing the probability that new variants are inoculated into hosts (see Appendix section 7.4).

Finally, it has been hypothesized that antigenic mutants can only proliferate at the population level if they arise against a favorable genetic background (Koelle & Rasmussen, 2015)—in the absence of deleterious substitutions elsewhere in the genome. But antigenic cluster transitions are frequently polyphyletic (see Appendix section 6) where a new variant emerges quasi-simultaneously in multiple virus lineages. Since these lineages should have different genetic backgrounds, this suggests that favorable backgrounds are readily available, and that emergence is limited instead by the presence or absence of selection pressure.

We discuss these alternative explanations in more depth in the Appendix (sections 7 and 7.4).

Relationship to prior influenza virus transmission bottleneck literature

Previous literature on influenza virus transmission mentions a "selective bottleneck" (Leonard et al., 2016; Moncla et al., 2016; Wilker et al., 2013), but these studies do not refer to antigenic inoculation selection. Rather, a "selective bottleneck" has been typically used to mean simply a tight neutral bottleneck (that leads to stochastic loss of diversity) or to non-antigenic factors that lead to preferential transmission of certain variants (Moncla et al., 2016). An important exception is (Lumby et al., 2018). The authors studied ferret transmission experiments and partitioned selection for adaptive mutants (not necessarily antigenic) into selection for transmissibility (acting at a potentially tight bottleneck) and selection during exponential growth. Subsequently, several studies have hypothesized that influenza virus antigenic selection might be weak in short-lived infections of individual experienced hosts and might occur at the point of transmission (Han et al., 2019; Lumby et al., 2020; Petrova & Russell, 2018).

To our knowledge, however, ours is the first study to undertake a mechanistic, model-based comparison between the role of selection at the point of transmission versus selection during replication in the promotion of antigenic mutants, to show that immunologically plausible mechanisms could make the former more salient than the latter, and to connect that finding to the rarity of observable new antigenic variants in homotypically reinfected human hosts. We discuss the relationship between this paradigm and those put forward in previous modeling studies at further length in section of the Appendix.

Limitations and remaining uncertainties

The study presented here nonetheless has important limitations that suggest opportunities for future investigation. This is a modeling study, and a mainly theoretical one. This is out of necessity. Quality experiments of the kind that are necessary to observe selection at the point of transmission directly and measure its strength have not been published, and we were unable to find any experimental measurements of within-host competition between known antigenic variants. For

additional discussion of unmodeled biological realities and mechanistic uncertainties, see sections 2 and 5 of the Appendix.

Within-host model

Our within-host model is a simple target cell-limited model, but the decrease in infectible cells as the infection proceeds can be qualitatively interpreted as any and all antigenicity-agnostic limiting factors that come into play as the virion and infected cell populations grow. This could include the action of innate immunity, which acts in part by killing infected cells and by rendering healthy cells difficult to infect via inflammation (see immunological review in 2 of the Appendix). The key mechanistic role played by the target cells in our model is to introduce a non-antigenic limiting factor on infections. This prevents an infection from repeatedly evolving out from under successive well-matched antibody responses, as occurs in HIV and in influenza patients with compromised immune systems (Xue et al., 2017). These factors limit the virus regardless of whether the infection remains confined to the upper respiratory tract or also infect the lower respiratory tract, thereby gaining access to additional target cells (Koel et al., 2019). The key is that non-antigenic factors prevent persistent large virus populations.

Similarly, the antibody response we introduce at 48 hours is qualitative—it is modeled simply as an increase in the virion decay rate for antibody matching virions. This response could represent IgA targeted at HA, but other antigenicity-specific modes of virus control could also be subsumed under the increased virion decay rate, for instance IgG antibodies, antibody dependent cellular cytotoxicity (ADCC), antibodies against the neuraminidase protein, or others. The key point we establish is that none of these mechanisms efficiently replication select because they all emerge once non-antigenic limiting factors have come into play. They speed clearance, but should not substantially alter virus evolution. A corollary to this point is that there are many mechanisms—and interventions—that could reduce the severity of influenza infections without substantially speeding up antigenic evolution, including universal vaccines.

Point of transmission

A key proposal of our study is that population-level antigenic selection and homotypic protection are mediated by antibody neutralization (likely sIgA) at the point of transmission. Currently, empirical evidence for antibody protection at the point of transmission is mostly indirect. Most of this evidence comes from human and animal challenge studies (Clements et al., 1986; Le Sage et al., 2020; Memoli et al., 2019). In these studies, individuals who are challenged with the same antigenic variant sometimes display apparent sterilizing immunity, but other times develop detectable infections. The study by (Memoli et al., 2019) is notable for having used very large inocula—10^6 or 10^7 $TCID_{50}$. Despite these high doses, two of the challenge subjects had neither detectable virus nor seroconversion. Similarly, ferret experiments (Le Sage et al., 2020) have shown that many ferrets develop sufficiently sterilizing immunity to prevent the virus from ever being detected, while some ferrets showed briefly detectable infections that were then cleared. While we cannot rule out a powerful immediate cellular response that was differentially evaded in the various subjects, we believe that our model, coupled with existing understanding of the timing of cellular responses and the

speed of influenza virus replication, provides a more parsimonious explanation.

Another limitation of our study is that, while we put forward mucosal sIgA as a biologically documented potential mechanism of immune protection at the point of inoculation that would not lead to strong selection during early viral replication, no modeling study can establish such a mechanism without empirical investigation. Our study reveals that such an empirical investigation would be of substantial scientific value.

We model neutralization at the point of transmission as a binomial process. Each virion is independently neutralized with a probability κ_i that depends on its antigenic phenotype and the host's immune history. As we discuss in Appendix section 5.2, this assumption of independence may be violated in practice. Careful experiments are required to develop a more realistic model of neutralization at the point of transmission.

Moreover, individual variation in immune system properties and complex effects of host immune history (Fonville et al., 2014; Lee et al., 2019) mean that even a pair of hosts who have both been previously exposed to the currently circulating variant may exert different selection pressures at the point of transmission. Modeling neutralization as a series of independent events that depend only on host history and virus phenotype is a baseline: it allows us to establish that antigenic selection at the point of transmission is possible and show what its consequences might be. But a more realistic model will be required to predict the selective pressures imposed by real hosts with real immune histories on real virions.

Finally, while we believe neutralization at the point of transmission is a crucial mechanism of protection against detectable reinfection for influenza, this may not be true for all RNA viruses. Some, such as measles and varicella, have long incubation periods even in naive hosts and induce reliable, long-lasting immune protection against detectable reinfection. This may be because they replicate slowly enough that they cannot "outrun" the adaptive response as influenza can. Neither shows influenza virus-like patterns of clocklike immune escape, suggesting that (1) escape mutants may be less available and (2) that adaptive response acts on a small population and is forceful.

Parameter uncertainties

We parameterized our models based on estimates from previous studies, but there are not good estimates for several important quantities. There are no high quality estimates of the rate of antibody-mediated neutralization in the presence of a homotypic antibody response or of how much this rate is reduced by particular antigenic substitutions, and there may be substantial inter-individual variation (Lee et al., 2019). Within-host timeseries data from antigenically heterogeneous infections are needed to estimate these quantities. Similarly, there are no good empirical data on the size of the bottlenecks that precede or follow the sIgA bottleneck (Fig. 3A of the main text). Better estimates of these bottlenecks, and of the probabilities of neutralization for individual virions encountering mucosal sIgA antibodies, would give more certainty about the strength of inoculation selection relative to neutral founder effects.

We also do not have a clear sense of exactly how neutralization probability in the respiratory tract mucosa (parameter κ in our model) and neutralization rate during

replication (parameter k in our model) relate. We expect them to be positively related, but the exact strength and shape of this relationship is unknown. Knowing whether major antigenic changes reduce both equally or reduce one more than the other could help us better quantify the potential strength of inoculation selection and replication selection. In short, better mechanistic understanding of mucosal antibody neutralization could be extremely valuable for understanding and potentially predicting influenza virus evolution.

Scaling up to the population level

How readily a particular individual host or host population helps new antigenic variants reach within-host consensus depends upon several unknown quantities: 1. how host susceptibility changes with extent of antigenic dissimilarity, 2. the ratio of virions that encounter sIgA to virions that found the infection (v/b), 3. the probability that a single new antigenic variant virion inoculated alongside old antigenic variant virions evades neutralization $1 - \kappa_m$ (Fig. 3E–H), 4. the duration τ from the onset of the antibody response to the time of transmission. Better empirical estimates of these quantities could shed light on how the distribution of host immunity shapes antigenic evolution. However, over a range of biologically plausible parameter values, our model contradicts the existing hypothesis that antigenic novelty appears when moderately immune hosts fail to block transmission and then select upon a growing virus population (Grenfell et al., 2004; Volkov et al., 2010). For influenza viruses, hosts whose mucosal immunity that regularly blocks old antigenic variant transmission may be crucial. Mucosal immunity not only produces a population-level advantage for new variants but may also play a role in their within-host emergence (Fig. 3E,H).

Implications

Our study has a number of implications for the study and control of influenza viruses.

Importance of host heterogeneity

Experienced hosts are undoubtedly heterogeneous in their immunity to a given influenza variant (Lee et al., 2019), so the overall population average protection against homotypic reinfection with variant i, z_i, is in fact an average over experienced hosts. Our model implies that the degree of neutralization difference between ancestral variant virions and new antigenic variant virions at the point of transmission strongly affects the probability of inoculation selection. Hosts with more focused immune responses—highly specific antibodies that neutralize old antigenic variant virions well and new antigenic variant virions poorly—could be especially good inoculation selectors and important sources of population-level antigenic selection. Hosts who develop less specific memory responses, such as very young children (Neuzil et al., 2006), could be less important. Similarly, immune-compromised hosts are excellent replication selectors (Lumby et al., 2020; Xue et al., 2017), and so their role in the generation of antigenic novelty and their impact on overall population level diversification rates both deserve further study.

Small-population-like evolution

Prior modeling has suggested that despite repeated tight bottlenecks at the point of transmission, evolution of influenza viruses should resemble evolution in idealized large populations (Sigal et al., 2018). In large populations, advantageous variants with small selective advantages should gradually fix and weakly deleterious variants should be purged. In these studies, diversity is rapidly generated and fit variants are selected to frequencies at which they are likely to pass through even a tight bottleneck. This is likely true of the phenotypes modeled in those studies, which include receptor binding avidity and virus budding time (Sigal et al., 2018). These phenotypes affect virus fitness throughout the timecourse of infection, so they can be efficiently replication-selected (where selection is manifest in the direct competition for infecting other cells rather than indirect competition by escaping antibodies). Indeed, next generation sequencing studies have found observable adaptative evolution of non-antigenic phenotypes in individual humans infected with avian H5N1 viruses (Welkers et al., 2019).

Seasonal influenza antigenic evolution does not resemble idealized large population evolution. Within an antigenic cluster, influenza viruses acquire substitutions that change the antigenic phenotype by small amounts. Given the large influenza virus populations within individual hosts, we might expect quasi-continuous directional pattern of evolution away from prior population immunity. Indeed, between clusters, evolution is strongly directional: only "forward" cluster transitions are observed. These are jumps—large antigenic changes. But incremental within-cluster evolution is *not* directional: the virus often evolves "backward" or "sideways" in antigenic space towards previously circulated variants (see Figs. 1 and 2 of (Smith et al., 2004)).

This noisy jump pattern is easy to explain in light of the weakness of replication selection and the importance of antigenic founder effects. Selection acts on the small sub-sample of donor-host diversity that passes through the excretion, inter-host, and mucus bottlenecks to encounter the sIgA bottleneck. Evolution via inoculation selection is therefore slower and more affected by stochasticity than evolution via replication selection (Fig. 3E–F). It resembles evolution in small populations—weakly adaptive and weakly deleterious substitutions become nearly neutral (Kimura, 1968; Ohta, 1992). If influenza virus evolution were not nearly neutral for small-effect substitutions at the within-host scale, it would be surprising to observe "backward" antigenic changes and noisy evolution at higher scales (Fig. 5A–D). In fact, there may be analogous "neutralizing" population dynamics at higher scales as well, and those may also be necessary for explaining population level noisiness. But whatever happens at higher scales, within-host replication selection would create a strongly directional bias in what population-level diversity is introduced (Fig. 5A–D), and introduces many small forward antigenic changes. Inoculation selection does not necessarily do this.

Population level neutralizing dynamics

Local influenza virus lineages rarely persist between epidemics (Bedford et al., 2015; Russell et al., 2008), and so new antigenic variants must establish chains of infections in other geographic locations in order to survive. new antigenic variant chains are most often founded when inoculations are common—that is, when existing variants are causing epidemics. Epidemics result in high levels of local competition between extant and new antigenic variant viruses for susceptible hosts (Hartfield & Alizon, 2015) as

well as metapopulation-scale competition to found epidemics in other locations. These dynamics could create tight bottlenecks between epidemics similar to those that occur between hosts, resulting in dramatic epidemic-to-epidemic diversity losses. That said, if immune hosts are present at the start of an epidemic, there will not be asynchrony between diversity and selection pressure, so new variants may pass through between-epidemic bottlenecks more readily than through between-host bottlenecks. Further work is needed to elucidate mechanisms at the population and meta-population scales.

Population immunity sets the clock of antigenic evolution

Our work suggests a simple mechanism by which accumulating immunity to an antigenic variant could produce punctuated population-level antigenic evolution. Population-level modeling has shown that influenza virus global epidemiological and phylogenetic patterns can be reproduced if new antigenic variants emerge at the population level with increasing frequency the longer an old antigenic variant circulates (Koelle et al., 2009).

If immunity to the old variant only gives new variants a population-level transmission advantage (population-level selection), we anticipate a constant rate of population-level antigenic diversification, with selective sweeps once a new variant has a sufficient population-level advantage over the old variant. If population immunity also helps new variants become the dominant within-host variant through inoculation selection, increasing population immunity to an old variant can produce increasing rates of new variant emergence (Fig. 6A). Whether this occurs in practice depends on the ecology of hosts and the relative strength of inoculation selection versus drift in partially and fully immune hosts (Fig. 4).

Potential synergy between brief antigenic replication selection late in infection and subsequent inoculation selection (Fig. 4) could further promote new variant emergence as population immunity accumulates. However, it is difficult to estimate how much this synergy matters in practice without knowing more about the kinetics of homotypic and heterotypic re-inoculation and reinfection.

One suggestive population-level pattern is that new antigenic variants frequently are observed quasi-simultaneously on multiple branches of the influenza virus phylogeny shortly prior to sweeping (see Appendix section 6). This suggests that emergence rate and population-level selection pressure do increase together. That said, alternative explanations are possible, such as reduced rates of stochastic loss of new variants at higher scales (e.g. the epidemic scale) with increased population immunity.

Regardless of the strength of these effects, however, our prediction is that new antigenic variant infection foundation events should constitute a rare but non-negligible fraction of transmission events: on the order of 1 in 10^5 or 1 in 10^4. This parsimoniously explains both why new antigenic variants lineages are hard to observe prior to undergoing positive population level selection but are readily available to be selected upon (see Appendix section 6) once that population level selection pressure has become sufficiently strong. Prior to that point they could be rendered unobservable by population-level neutralizing dynamics or by clonal interference from non-antigenic weakly adaptive mutants (Strelkowa & Lässig, 2012).

The slow rate of antigenic evolution of A/H3N2 viruses in swine lends further support to this argument. A/H3N2 viruses accumulate genetic mutations at a similar pace in swine and humans, but antigenic evolution is much slower in swine (De Jong et al., 2007; Lewis et al., 2016). Slaughter for meat means that pig population turnover is high. It follows that the frequency of experienced hosts rarely becomes sufficient to facilitate the appearance of observable new antigenic variants.

The population-level emergence of new antigenic variants, in other words, tracks the accumulation of immunity in the population, not the accumulation of genetic diversity. This suggests that A/H3N2 evolution is indeed selection-limited, not diversity-limited. But much generated antigenic diversity is invisible to surveillance: in the absence of positive selection, it is likely to be lost at bottlenecks.

Implications for other pathogens

The inoculation and replication selection paradigm has implications for the understanding and management of other pathogens. For example, HIV does not readily evolve resistance to contemporary pre-exposure prophylaxis (PrEP) antiviral drugs, but it can do so when these antivirals are taken by an individual who is already infected with HIV (Weis et al., 2016). Developing resistance at the moment of exposure is a difficult problem of inoculation selection for the virus, but developing resistance during an ongoing infection is an easier problem of replication selection. Selection on small un-diverse introduced populations may also be of interest in invasion biology and island biogeography.

Conclusion

The asynchrony between within-host virus diversity and antigenic selection pressure provides a simple mechanistic explanation for the phenomenon of weak within-host selection but strong population-level selection in seasonal influenza virus antigenic evolution. Measuring or even observing antibody selection in natural influenza virus infections is likely to be difficult because it is inefficient and consequently rare. Theoretical studies are therefore essential for understanding these phenomena and for determining which measurable quantities will facilitate influenza virus control. Our study highlights a critical need for new insights into sIgA neutralization and IgA responses to natural influenza virus infection and vaccination and shows that cross-scale dynamics can decouple selection and diversity, introducing randomness into otherwise strongly adaptive evolution.

Methods

Model notation

In all model descriptions, $X \mathrel{+}= y$ and $X \mathrel{-}= y$ denote incrementing and decrementing the state variable X by the quantity y, respectively, and $X = y$ denotes setting the variable X to the value y. $\dot{X}$ denotes the rate of event X.

Within-host model overview

The within-host model is a target cell-limited model of within-host influenza virus infection with three classes of state variables:

- C: Target **C**ells available for the virus to infect. Shared across all virus variants.
- V_i: **V**irions of virus antigenic variant i
- E_i: Binary variable indicating whether the host has previously **E**xperienced infection with antigenic variant i ($E_i = 1$ for experienced individuals; $E_i = 0$ for naive individuals).

New virions are produced through infections of target cells by existing virions, at a rate $\beta C V_i$. Infection eventually renders a cell unproductive, so target cells decline at a rate $f\beta C\bar{V}$, where $\bar{V}$ is the total number of virions of all variants. The model allows mutation: a virion of antigenic variant i has some probability μ_{ij} of producing a virion of antigenic variant j when it reproduces.

Virions have a natural per-capita decay rate d_v. A fully active specific antibody-mediated immune response to variant i increases the virion per-capita decay rate for variant i by a factor k (assumed equal for all variants). The degree of activation of the antibody response during an infection is given by a function $M(t)$, where t is the time since inoculation.

We use a parameter c_{ij} to denote the protective strength of antibodies raised against a strain j against a different strain i. $c_{ii} = 1$ by definition and $c_{ij} = 0$ indicates complete absence of cross-protection. So if host has antibodies against a strain j but not against the infecting strain i, a fully active antibody response raises the virion decay rate for strain i by $c_{ij}k$. If there are multiple candidate forms of cross protection c_{ij} and c_{ik}, we choose the strongest. We typically assume that $c_{ij} = c_{ji}$.

We assume that an antibody immune response is raised whenever the host has experienced a prior infection with a partially cross-reactive strain. For notational ease, we define the host's strongest cross-reactivity against strain i, c_i by:

$$c_i = \max c_{ij} E_j \tag{1}$$

So a recall antibody response is raised during an infection with strains $i, j, ...$ whenever one of $c_i, c_j, ... > 0$.

For this study, we consider a two-antigenic variant model with an ancestral "old antigenic variant" virions V_w and novel "new antigenic variant" virions V_m, though the model generalizes to more than two variants.

$$\begin{aligned} C^+ &: \quad C \mathrel{+}= 1 \\ C^- &: \quad C \mathrel{-}= 1 \\ V_w^+ &: \quad V_w \mathrel{+}= 1 \\ V_w^- &: \quad V_w \mathrel{-}= 1 \\ V_m^+ &: \quad V_m \mathrel{+}= 1 \\ V_m^- &: \quad V_m \mathrel{-}= 1 \end{aligned} \tag{2}$$

The events occur at the following rates:

$$
\begin{aligned}
\dot{C}^- &= f\beta C(V_w + V_m) \\
\dot{C}^+ &= p_C C\left(1 - C/C_{\max}\right) \\
\dot{V}_w^+ &= \beta C V_w r_w (1-\mu) \\
\dot{V}_w^- &= V_w\Big(d_v + c_w k M(t)\Big) \\
\dot{V}_m^+ &= \beta C\Big(V_m r_m + V_w r_w \mu\Big) \\
\dot{V}_m^- &= V_m\Big(d_v + c_m k M(t)\Big)
\end{aligned}
\tag{3}
$$

where $M(t)$ is a minimal model of a time-varying antibody response given by:

$$
M(t) = \begin{cases} 1, & t > t_M \\ 0, & \text{otherwise} \end{cases}
\tag{4}
$$

For simplicity, the equations are symmetric between old antigenic variant and new antigenic variant viruses, except that we neglect back mutation, which is expected to be rare during a single infection, particularly before mutants achieve large populations. The parameters r_w and r_m allow the two variants optionally to have distinct within-host replication fitnesses; for all results shown, we assumed no replication fitness difference ($r_w = r_m = r$), unless otherwise stated. A *de novo* antibody response raised to a not-previously-encountered variant i can be modeled by setting $E_i = 1$ at a time $t_N^i \geq t_M$ post-infection. By default, we model such a *de novo* response only for fully naive hosts, and assume that it is mounted against the majority variant at the start of infection, which is typically w.

We characterize a virus variant i by its within-host basic reproduction number $\mathcal{R}_0^i$, the mean number of new virions produced before decay by a single virion at the start of infection in a naive host:

$$
\mathcal{R}_0^i \equiv \frac{\beta C_{\max} r_i}{d_v}
\tag{5}
$$

When parametrizing our model, we fixed the within-host basic reproduction number $\mathcal{R}_0^i$, the initial target cell population $C_{\max}$, the virus reproduction rate r_i, and the shared virion decay rate d_v. We then calculated the implied β according to (5).

Another useful quantity is the within-host effective reproduction number $\mathcal{R}^i(t)$ of variant i at time t: the mean number of new virions produced before decay by a single virion of variant i at a given time t post-infection.

$$
\mathcal{R}^i(t) \equiv \frac{\beta C(t) r_i}{d_v + c_i k M(t)}
\tag{6}
$$

Note that $\mathcal{R}_0^i$ is $\mathcal{R}^i(0)$ in a naive host, and that if $\mathcal{R}^i < 1$, the virus population will usually decline.

The distribution of virions that encounter sIgA antibody neutralization depends on the mean mucosal bottleneck size v (i.e. the mean number of virions that would pass through the respiratory tract mucosa in the absence of antibodies) and on the frequency of new antigenic variant $f_{\mathrm{mt}} = f_m(t_{\mathrm{inoc}})$ in the donor host at the time of inoculation t_{inoc}. n_w old antigenic variant virions and n_m new antigenic variant virions encounter sIgA. The total number of virions $n_{\mathrm{tot}} = n_w + n_m$ is Poisson distributed with mean v and each virion is independently a new antigenic variant with probability f_{mt} and otherwise an old antigenic variant. The principle of Poisson thinning then implies:

$$\begin{aligned} n_m &\sim \mathrm{Poisson}\Big(v f_{\mathrm{mt}}\Big) \\ n_w &\sim \mathrm{Poisson}\Big(v(1 - f_{\mathrm{mt}})\Big) \end{aligned} \tag{7}$$

Note that since f_{mt} is typically small the results should also hold for a binomial model of n_w and n_m with a fixed total number of virions encountering sIgA antibodies: $v_{\mathrm{tot}} = v$.

We then model the sIgA bottleneck—neutralization of virions by mucosal sIgA antibodies. Each virion of variant i is independently neutralized with a probability κ_i. This probability depends upon the strength of protection against homotypic reinfection κ and the sIgA cross immunity between variants σ, $(0 \leq \sigma \leq 1)$:

$$\begin{aligned} \kappa_w &= \kappa \max\{E_w, \sigma E_m\} \\ \kappa_m &= \kappa \max\{E_m, \sigma E_w\} \end{aligned} \tag{8}$$

Since each virion of strain i in the inoculum is independently neutralized with probability κ_i, then given n_w and n_m, the populations that compete the pass through the cell infection bottleneck x_w and x_m are binomially distributed:

$$\begin{aligned} x_w &\sim \mathrm{Binomial}(n_w, \kappa_w) \\ x_m &\sim \mathrm{Binomial}(n_m, \kappa_m) \end{aligned} \tag{9}$$

By Poisson thinning, this is equivalent to:

$$\begin{aligned} x_w &\sim \mathrm{Poisson}\Big(v(1 - f_{\mathrm{mt}})(1 - \kappa_w)\Big) \\ x_m &\sim \mathrm{Poisson}\Big(v f_{\mathrm{mt}}(1 - \kappa_m)\Big) \end{aligned} \tag{10}$$

At this point, the remaining virions are sampled without replacement to determine what passes through the cell infection bottleneck, b, with all virions passing through if $x_w + x_m \leq b$, so the final founding population is hypergeometrically distributed given x_w and x_m.

If $\kappa_w = \kappa_m = 0$, f_{mt} is small, and v is large, this is approximated by a binomially distributed founding population of size b, in which each virion is independently a new antigenic variant with (low) probability f_{mt} and is otherwise an old antigenic variant. Alternatively, it can be approximated by a Poisson distribution with a small mean: $f_{\mathrm{mt}} b \ll 1$:

$$p_{\mathrm{drift}} \approx 1 - (1 - f_{\mathrm{mt}})^b \approx f_{\mathrm{mt}} b \tag{11}$$

When there is mucosal antibody neutralization, the variant's survival probability can be reduced below this (inoculation pruning) or promoted above it (inoculation promotion), depending upon parameters. There can be inoculation pruning even when the variant is more fit (neutralized with lower probability, positive inoculation selection) than the old antigenic variant (see Fig. 3E).

Within-host model parameters

Default parameter values for the minimal model and sources for them are given in Table 1.

Table 1: **Model parameters, default values, and sources/justifications**

Parameter	Meaning	Units	Value	Source or justification
t_M	time post-infection of antibody response in experienced hosts	days	2	literature (see review in Section 2)
t_N^w	time post infection of a novel immune response to the old antigenic variant	days	6	literature (see review in Section 2)
p_C	per capita growth rate of target cells at low density	$\frac{1}{\text{days}}$	0	ignored on the timescale of a single infection
$C_{\max}$	maximum number of target cells	cells	4×10^8	standard in the modeling literature (Baccam et al., 2006; Hadjichrysanthou et al., 2016; Luo et al., 2012)
$\mathcal{R}_0$	within-host basic reproduction number for the virus	unitless	5	empirical fits of target cell models (Hadjichrysanthou et al., 2016)
r_w	average number of virions produced by a cell infected with old antigenic variant virus	virions	100	literature (Frensing et al., 2016)
r_m	average number of virions produced by a cell infected with new antigenic variant virus	virions	100	no within-host deleteriousness for new antigenic variants
z_w	probability of no infection given homotypic re-inoculation for a host previously exposed to the old antigenic variant	unitless	0.95	challenge studies (Clements et al., 1986; Memoli et al., 2019)
β	rate of infectious contact between virions and target cells per cell per virion	$\frac{1}{\text{virions cells days}}$	calculated	from $\mathcal{R}_0$
f	number of target cells lost per infectious contact	cells	1	one cell lost per cell infection
k	additional per virion neutralization rate in the presence of an antibody response	$\frac{1}{\text{days}}$	6	varied to test hypotheses
κ_w	probability an individual inoculated old antigenic variant virion inoculated into an experienced host is neutralized in the respiratory tract mucosa	unitless	set from z_w	calculated from equation 38
κ_m	probability an individual inoculated old antigenic variant virion inoculated into an experienced host is neutralized in the respiratory tract mucosa	unitless	$\sigma\kappa_w$	reduced relative to κ_m by immune escape
c_w	percentage cross reactivity during viral replication between host antibodies and old antigenic variant	unitless	0 or 1	naive or homotypically reinfected hosts
c_m	percentage cross reactivity during viral replication between host antibodies and new antigenic variant	unitless	0	full escape variant
σ	percentage cross immunity at sIgA bottleneck between old antigenic variant and new antigenic variant	unitless	0	full escape variant
b	size of final / cell infection bottleneck	virions	1	NGS studies (McCrone et al., 2018; Xue & Bloom, 2019)
v	number of virions encountering sIgA	virions	$10 \times b$	
V_{50}	viral load at which there is a fifty-percent transmission probability	virions	10^8	chosen to give realistic transmission window (Tsang et al., 2015) and based on prior modeling studies (Russell et al., 2012)
θ	transmission threshold for threshold model	virions	10^7	chosen to be consistent with V_{50}

Selection and drift within hosts

In this section, we derive analytical expressions for the within-host frequency of the new variant over time in an infected host (Fig. 7, the probability distribution of the time of first *de novo* mutation to produce a surviving new variant lineage, and the approximate probability of replication selection to frequency x by time t given our parameters.

Within-host replicator equation

The within-host frequency of the new variant, f_m, obeys a replicator equation of the form:

$$\frac{\mathrm{d}f_m}{\mathrm{d}t} = f_m(1 - f_m)\delta(t) \tag{12}$$

where $\delta(t)$ is the fitness advantage of the new antigenic variant over the old antigenic variant at time t (see Appendix section 3 for a derivation).

If the variant is neutral in the absence of antibodies, then $\delta(t) = k(c_w - c_m)$ if $t > t_M$ and $\delta(t) = 0$ otherwise. If the new antigenic variant has emerged by t_M, then at a time $t > t_M$:

$$f_m(t) = \frac{e^{\delta(t-t_M)}}{e^{\delta(t-t_M)} + f_M^{-1} - 1} \tag{13}$$

where $\delta = k(c_w - c_m)$ and $f_M = f_m(t_M)$

When additional mutations after the first cannot be neglected, we add a correction term to $\frac{\mathrm{d}f_m}{\mathrm{d}t}$ for $t_e < t < \min\{t_M, t_{\text{peak}}\}$ (Fig. 7):

$$\frac{\mathrm{d}f_m}{\mathrm{d}t} = f_m(1 - f_m)\delta(t) + \mu\mathcal{R}_0 d_v(1 - 2f_m + f_m^2) \tag{14}$$

which for $\delta \neq 0$ yields:

$$\begin{aligned} f_m(t) \quad &\approx \frac{A\exp(\delta t) + \mu g_0}{A\exp(\delta t) + \mu g_0 - \delta} \\ A \qquad &= \frac{f_0}{1 - f_0}(\mu g_0 - \delta) - \frac{\mu g_0}{f_0} \end{aligned} \tag{15}$$

And when $\delta = 0$:

$$f_m(t) \approx \frac{\frac{1}{1-f_0} + \mu g_0 t - 1}{\frac{1}{1-f_0} + \mu g_0 t} \tag{16}$$

See Appendix section 3.2 for derivations and discussion.

Distribution of first mutation times

In our stochastic model, new variant lineages that survive stochastic extinction are produced by *de novo* mutation according to a continuous-time, variable-rate Poisson process. The cumulative distribution function for the time of the first successful mutation, t_e, depends on the mutation rate μ and the per-capita rate at which old antigenic variant virions are produced, $g_w(t) = r_w \beta C(t)$. It also depends on p_{sse}, the probability that the generated new antigenic variant survives stochastic extinction. Denoting the new variant per-capita virion production rate $g_m(t) = r_m \beta C(t)$, we calculate p_{sse} using a branching process approximation ((Ball et al., 2016)).

$$p_{\text{sse}} = \frac{g_m(t)}{g_m(t) + d_v + kM(t)\max\{E_m, cE_w\}} \tag{17}$$

Surviving mutants therefore occur at a rate $\lambda_m(t)$:

$$\lambda_m(t) = \mu g_w(t) V_w(t) p_{\text{sse}} \tag{18}$$

We define the cumulative rate $\Lambda_m(x)$:

$$\Lambda_m(x) = \int_0^x \lambda_m(x) dx \tag{19}$$

It follows that the CDF of surviving new variant mutation times is:

$$P(t_e < x) = 1 - e^{-\Lambda_m(x)} \tag{20}$$

This expression is exact for any given realization of the stochastic model if the realized values of the random variables $V_w(t)$, $C(t)$, and $g_w(t)$ are used. In practice, we mainly use it to get a closed form for the approximate new variant CDF by making the approximations that $C(t) \approx C_{\max}$ early in infection. This yields approximations for $g_w(t)$, $g_m(t)$, and $V_w(t)$:

$$g_w(t), g_m(t) \approx g_0 \equiv \mathcal{R}_0 d_v \tag{21}$$

$$V_w(t) \approx \begin{cases} b\exp\big((g_0 - d_v)t\big) & t \leq t_M \\ b\exp\big((g_0 - d_v)t_M\big)\exp\big((g0 - d_v - c_w k)(t - t_M)\big) & t > t_M \end{cases} \tag{22}$$

The resultant approximate solution for the CDF of new variant mutation times agrees well with simulations (Fig. 8).

The slightly earlier simulated mutation times in the immediate recall response case (Fig. 8A) would only make replication selection more likely in that case than our analytical approximation suggests.

Required mutation time for variant to reach a given frequency

By inverting the within-host replicator equation, we can also calculate the time $t^*(x,t)$ by which a new variant must emerge if it is to reach at least frequency x by time t. We show (see Appendix section 9.4 for derivation) that there are two candidate values for t^*, depending on whether the time that the new variant first emerges (t_e) is before or after the onset of the antibody response (t_M):

If $t_e \leq t_M$:

$$t^*_-(x,t) = \frac{\ln\left[\frac{1-x}{x}\exp\big(\delta(t-t_M)\big)+1\right] - \ln b}{g_0 - d_v} \tag{23}$$

If $t_e > t_M$:

$$t^*_+(x,t) \approx \frac{\ln\left(\frac{1-x}{x}\right) - \ln(b) + \delta t - c_w k t_M}{g_0 - d_v - c_w k + \delta} \tag{24}$$

It may be that $t^*_+(x,t) > t$ and $t^*_-(x,t) > t$. This indicates that the mutant will be at frequency x if it emerges at t itself. In that case, we therefore have $t^*(x,t) = t$. So combining:

$$t^*(x,t) = \begin{cases} t^*_-(x,t) & t^*_-(x,t) < t \text{ and } t^*_-(x,t) \leq t_M \\ t^*_+(x,t) & t^*_+(x,t) < t \text{ and } t^*_-(x,t) > t_M \\ t & \text{otherwise} \end{cases} \tag{25}$$

Finally, it is worth noting that in the case of a complete escape mutant ($c_m = 0$, $\delta = c_w k$), the approximate expression for t^*_+ is exactly equal to the equivalent approximate expression for t^*_-:

$$t^*(x,t) \approx \frac{\ln\left(\frac{1-x}{x}\right) - \ln(b) + c_w k(t - t_M)}{g_0 - d_v} \tag{26}$$

This is a linear function of t.

Probability of replication selection

Given this and the the new variant first mutation time CDF calculated in equation 20, it is straightforward to calculate the probability of replication selection to a given frequency a by time t, assuming that $\mathcal{R}_0^w > 1$ early in infection:

$$p_{\text{repl}}(a,t) = P(t_e < t^*(a,t)) = 1 - e^{-\Lambda(t^*(a,t))} \tag{27}$$

This analytical model agrees well with simulations (Fig. 9). We use it used to calculate the heatmaps shown in Fig. 1, with the $C(t) \approx C_{\max}$ early infection approximations that give us a closed form for $\Lambda_m(x)$.

When $t^* > t_M$, the integral $\int_0^{t^*} \lambda_m(x)dx$ can be evaluated piecewise, first from 0 to t_M, and then from t_M to t^*. This can also be used to find evaluate the probability density function (PDF) of first mutation times, as needed.

Finally, note that for $t_e < t_M$, the expression for p_{repl} in practice depends only on $\delta = (c_w - c_m)k$, not on c_w, c_m and k separately. In the absence of an antibody response, the mutant is generated with near certainty by 1 day post-infection ($P(t_e < 1) \approx 1$, Fig. 8B). So when $t_M \geq 1$, values for p_{repl} calculated with $c_w = 1$ $c_m = 0$, and $\delta = k$—as in (Fig. 1C,D)—will in fact hold for any c_w, c_m, and k that produce that fitness difference δ.

When $\mathcal{R}_0^w < 1$ early in infection, the probability of replication selection depends on the probability of generating an escape mutant before the infection is extinguished:

$$p_{\text{repl}} \approx \frac{\mathcal{R}^w(0)}{1 - \mathcal{R}^w(0)} b\mu p_{\text{sse}} \tag{28}$$

See Appendix section 3.6 for a derivation.

Point of transmission model

In this section, we describe our model of the point of transmission, including how sIgA neutralization may impose selection pressure.

Transmission probability

Given contact between an infected and an uninfected host, we assume that a transmission attempt occurs with probability proportional to the current virus population size V_{tot} in the infected host:

$$P(\text{transmit}) = 1 - \exp\left(\log(2)\frac{V_{\text{tot}}}{V_{50}}\right) \tag{29}$$

The parameter V_{50} sets the scaling, and reflects virus population size at which there is a 50% chance of successful transmission to a naive host. We used a default value of $V_{50} = 10^8$ virions. In addition to this *probabilistic model*, we also consider an alternative *threshold model* in which a transmission attempt occurs with certainty if V_{tot} is greater than a threshold θ and does not occur otherwise.

Bottleneck survival

If there are x_w old antigenic variant virions and x_m variant virions, at least one variant passes through the cell infection bottleneck with probability:

$$p_{\text{cib}}(x_w, x_m, b) = 1 - \frac{\binom{x_w}{b}}{\binom{x_w+x_m}{b}} \tag{30}$$

Summing $p_{\text{cib}}(x_w, x_m)$ over the possible values of x_w and x_m weighted by their joint probabilities yields the variant's overall probability of successful onward transmission

$p_{\text{surv}}(f_{\text{mt}}, v, b, \kappa, c)$:

$$p_{\text{surv}} = \sum_{x_m, x_w} P(x_m, x_w) p_{\text{cib}}(x_w, x_m, b) \tag{31}$$

$P(x_w, x_m)$ is the product of the probability mass functions for x_w and x_m:

$$P(x_m, x_w) = \frac{\bar{w}^{x_w}}{x_w!} \frac{\bar{m}^{x_m}}{x_m!} e^{-(x_w + x_m)} \tag{32}$$

where $\bar{w} = v(1 - f_{\text{mt}})(1 - \kappa_w)$ and $\bar{m} = v f_{\text{mt}}(1 - \kappa_m)$

At low donor host variant frequencies $f_{\text{mt}} \ll 1$ p_{surv} can be approximated using the fact that almost all probability is given to $x_m = 0$ and $x_m = 1$ (0 or 1 new antigenic variant after the sIgA bottleneck).

At least one new antigenic variant survives the sIgA bottleneck with probability:

$$p_{\text{inoc}}(\kappa_m, v, f_{\text{mt}}) = 1 - e^{-v f_{\text{mt}}(1 - \kappa_m)} \tag{33}$$

If f_{mt} is small, there will almost always be only one such virion ($x_m = 1$). That new antigenic variant's probability of surviving the cell infection bottleneck depends upon how many old antigenic variant virions are present after neutralization, x_w:

$$p_{\text{cib}}(x_w, 1, b) = 1 - \frac{\binom{x_w}{b}}{\binom{x_w+1}{b}} = 1 - \frac{x_w - b + 1}{x_w + 1} = \frac{b}{x_w + 1} \tag{34}$$

Summing over the possible x_w, we obtain a closed form for the unconditional probability $p_{\text{cib}}(\kappa_w, v, b)$ (see section 9.5 for a derivation):

$$p_{\text{cib}}(\kappa_w, v, b) = (1 - e^{-\bar{w}}) \frac{b}{\bar{w}} + e^{-\bar{w}} \sum_{j=0}^{b-1} \frac{\bar{w}^j}{j!} \left(1 - \frac{b}{j+1}\right) \tag{35}$$

where $\bar{w} = v(1 - f_{\text{mt}})(1 - \kappa_w)$ is the mean number of old antigenic variant virions present after sIgA neutralization. If $b = 1$, this reduces the just the first term, and it is approximately equal to just the first term when $\bar{w}$ is large, since $e^{-\bar{w}}$ becomes small.

It follows that there is an approximate closed form for the probability that a variant survives the final bottleneck:

$$p_{\text{surv}}(\kappa_m, \kappa_w, v, b, f_{\text{mt}}) \approx p_{\text{inoc}}(\kappa_m, v, f_{\text{mt}}) * p_{\text{cib}}(\kappa_w, v, b) \tag{36}$$

We use this expression to calculate the analytical variant survival probabilities shown in the main text, and to gain conceptual insight into the strength of inoculation selection relative to neutral processes (see Appendix section 4.5).

Neutralization probability and probability of no infection

A given per-virion sIgA neutralization probability κ_i implies a probability z_i that a transmission event involving only virions of variant i fails to result in an infected cell.

Some transmissions fail even without sIgA neutralization; this occurs with probability $\exp(-v)$. Otherwise, with probability $1 - \exp(-v)$, n_i virions must be neutralized by sIgA to prevent an infection. We define z_i as the probability of no infection given inoculated virions in need of neutralization (i.e. given $n_i > 0$).

The probability that there are no remaining virions of variant i after mucosal neutralization is $\exp\big(-v(1-\kappa_i)\big)$. So we have:

$$z_i = \frac{\exp\big(-v(1-\kappa_i)\big) - \exp(-v)}{1 - \exp(-v)} \tag{37}$$

This can be solved algebraically for κ_i in terms of z_i, but it is more illuminating first to find the overall probability of no infection given inoculation p_{no} and then solve:

$$\begin{aligned} p_{\text{no}} &= \exp\big(-v(1-\kappa_i)\big) \\ \kappa_i &= 1 + \frac{\ln(p_{\text{no}})}{v} \end{aligned} \tag{38}$$

An infection occurs with probability $(1 - \exp(-v))(1 - z_i)$ (the probability of at least one virion needing to be neutralized times the conditional probability that it is not) so:

$$p_{\text{no}} = 1 - (1 - \exp(-v))(1 - z_i) \tag{39}$$

This yields the same expression for κ_i in terms of z_i as a direct algebraic solution of 37.

For moderate to large v, $1 - \exp(-v)$ approaches 1 so p_{no} approaches z_i and κ_i approaches $1 + \frac{\ln(z_i)}{v}$. This reflects the fact that for even moderately large v, it is almost always the case that $n_i > 0$: a least one virion must be neutralized to prevent infection. In those cases, z_i can be interpreted as the (approximate) probability of no infection given a transmission event (i.e. as p_{no}).

Note also that seeded infections can also go stochastically extinct; this occurs with approximate probability $\frac{1}{\mathcal{R}_0}$. At the start of infection, if there is no antibody response, $\mathcal{R}_0$ is large (5 to 10) stochastic extinction probabilities should be low ($\frac{1}{5}$ to $\frac{1}{10}$), and equal in immune and naive hosts. We have therefore parametrized our model in terms of z_i, the probability that no cell is ever infected, as that probability determines to leading order the frequency with which immunity protects against detectable reinfection given challenge.

Susceptibility models

Translating host immune histories into old antigenic variant and new antigenic variant neutralization probabilities for the analysis in Fig. 3G,H requires a model of how susceptibility decays with antigenic distance, which we measure in terms of the typical distance between two adjacent "antigenic clusters" (Smith et al., 2004). Fig. 3 shows

results for two candidate models: a multiplicative model used in a prior modeling and empirical studies of influenza evolution (Asaduzzaman et al., 2018; Boni et al., 2006), and a sigmoid model as parametrized from data on empirical HI titer and protection against infection (Coudeville et al., 2010).

In the multiplicative model, the probability $z(i, x)$ of no infection with variant i given that the nearest variant to i in the immune history is variant x is given by:

$$z(i, x) = z_0 \left(\frac{z_1}{z_0} \right)^{d(i,x)} \tag{40}$$

where z_0 is the probability of no infection given homotypic reinfection, $d(i, x)$ is the antigenic distance in antigenic clusters between i and x, and z_1 is the probability of no infection given $d(i, x) = 1$.

In the sigmoid model:

$$z(i, x) = 1 - \frac{1}{1 + e^{b\left(\ln\left(T(i,x)\right) - a\right)}} \tag{41}$$

where $a = 2.844$ and $b = 1.299$ (α and β estimated in (Coudeville et al., 2010)) and $T(i, x)$ is the individual's HI titer against variant i (Coudeville et al., 2010). To convert this into a model in terms of z_0 and z_1 we calculate the typical homotypic titer T_0 implied by z_0 and the n-fold-drop D in titer per unit of antigenic distance implied by z_1, since units in antigenic space correspond to a n-fold reductions in HI titer for some value n (Smith et al., 2004). We calculate T_0 by plugging z_0 and T_0 into equation 41 and solving. We calculate D by plugging z_1 and $T_1 = T_0 D^{-1}$ into equation 41 and solving. We can then calculate $T(i, x)$ as:

$$T(i, x) = T_0 D^{-d(i,x)} \tag{42}$$

Probability of bottleneck survival

With these analyses in hand, it is possible to combine the within-host and the point of transmission processes to calculate an overall probability that a new variant survives the transmission bottleneck. Equation 36 gives the probability of bottleneck survival given f_{mt} and the properties of the recpient host. And given a time of transmission t_t, we can calculate the probability distribution of f_{mt} using our expressions for the CDF of successful mutation times t_e (equation 20) and for $f_m(t)$ (equations 13, 15). To calculate the overall probability p_{nv}, we average over the possible values of f_{mt}, weighted by their probability:

$$p_{\text{nv}} = \int_0^\infty p_{\text{surv}}(f_m(t_t \mid t_e = t)) p(t_e = t)\, dt \tag{43}$$

Within-host simulations (Figs. 1, 3)

To evaluate the relative probabilities of replication selection and inoculation selection for antigenic novelty and to check the validity of the analytical results, we simulated 10^6 transmissions from a naive host to a previously immune host. The transmitting host was simulated for 10 days, which given the selected model parameters is sufficient time for almost all infections to be cleared. Time of transmission was randomly drawn from the time period during which the infected host had a transmissible virus population and an inoculum of fixed size was drawn from the within-host virus population at that time. Variant counts in the inoculum were Poisson distributed with probability equal to the variant frequency within the transmitting host at the time of transmission. We simulated the recipient host until the clearance of infection and found the maximum frequency of transmissible variant: that is, the variant frequencies when the transmission probability was greater than 5×10^{-2} (probabilistic model) or when the virus population was above the transmission threshold θ (threshold model). For Fig. 3D, we defined an infection with an emerged new antigenic variant as an infection with a maximum transmissible new variant frequency of greater than 50%.

Transmission chain model (Fig. 5)

To study evolution along transmission chains with mixed host immune statuses, we modeled the virus phenotype as existing in a 1-dimensional antigenic space. Host susceptibility s_x to a given phenotype x was $s_x = \min\{1, \min\{|x - y_i|\}\}$ for all phenotypes y_i in the host's immune history.

Within-host cross immunity c_{ij} between two phenotypes y_i and y_j was equal to $\max\{0, 1 - |y_i - y_j|\}$. When mucosal antibodies were present, protection against infection z_x was equal to the strength of homotypic protection $z_{\max}$ scaled by susceptibility: $z_x = (1 - s_x)z_{\max}$. κ_x was calculated from z_x by $\exp\big(-v(1 - \kappa_x)\big) = z_x$. We used $z_{\max} = 0.95$. Note that this puts us in the regime in which intermediately immune hosts are the best inoculation selectors (Fig. 3H). We set $k = 25$ so that there would be protection against reinfection in the condition with an immediate recall response but without mucosal antibodies.

We then simulated a chain of infections as follows. For each inoculated host, we tracked two virus variants: an initial majority antigenic variant and an initial minority antigenic variant. If there were no minority antigenic variants in the inoculum, a new focal minority variant (representing the first antigenic variant to emerge *de novo*) was chosen from a Gaussian distribution with mean equal to the majority variant and a given variance, which determined the width of the mutation kernel (for results shown in Fig. 5, we used a standard deviation of 0.08). We simulated within-host dynamics in the host according to our within-host stochastic model.

We founded each chain with an individual infected with all virions of phenotype 0, representing the current old antigenic variant.

We simulated contacts at a fixed, memoryless contact rate $\rho = 1$ contacts per day. Given contact, a transmission attempt occurred with a probability proportional to donor host viral load, as described above. If a transmission attempt occurred, we chose a random immune history for our recipient host according to a pre-specified distribution of host immune histories. We then simulated an inoculation and, if applicable,

subsequent infection, according to the within-host model described above. If the recipient host developed a transmissible infection, it became a new donor host. If not, we continued to simulate contacts and possible transmissions for the donor host until recovery. If a donor host recovered without transmitting successfully, the chain was declared extinct and a new chain was founded.

We iterated this process until the first phenotypic change event—a generated or transmitted minority phenotype becoming the new majority phenotype. We simulated 1000 such events for each model and examined the observed distribution of phenotypic changes compared to the mutation kernel.

For the results shown in Fig. 5, we set the population distribution of immune histories as follows: 20% −0.8, 20% −0.5, 20% 0.0, and the remaining 40% of hosts naive. This qualitatively models the directional pressure that is thought to canalize virus evolution (Bedford et al., 2012) once a cluster has begun to circulate.

Analytical mutation kernel shift model (Fig. 5)

To assess the causes of the observed behavior in our transmission chain model, we also studied analytically how replication and inoculation selection determine the distribution of observed fixed antigenic changes given the mutation kernel when a host with one immune history inoculates another host with a different immune history.

We fixed a transmission time $t = 2$ days, roughly corresponding to peak within-host virus titers. For each possible new variant phenotype, we calculated p_{nv} according to equation 43, with parameters given by the the old variant antigenic phenotype, new variant phenotype, and host immune histories. Finally, we multiplied each phenotype's survival probability by the same Gaussian mutation kernel used in the chain simulations (with mean 0 and s.d. 0.08), and normalized the result to determine the predicted distribution of surviving new variants given the mutation kernel and the differential survival probabilities for different phenotypes.

Population-level model (Fig. 6)

To evaluate the probability of variants being selected and proliferating during a local influenza virus epidemic, we first noted that the per-inoculation rate of new antigenic variant chains for a population with n_s susceptibility classes (which can range from full susceptibility to full immunity) is:

$$\sum_{i=1}^{n_s} s_i(0) p_{\text{surv}}(\kappa_m(i), \kappa_w(i), v, b, f_{\text{mt}}) \tag{44}$$

where $\kappa_w(i)$ and $\kappa_m(i)$ are the mucosal antibody neutralization probabilities for the old antigenic variant and the new antigenic variant associated with susceptibility class i, and $s_i(0) = \frac{S_i(0)}{N}$ is the initial fraction of individuals in susceptibility class i.

We then considered a well-mixed population with frequency-dependent transmission, where infected individuals from all susceptibility classes are equally infectious if infected. Using an existing result from epidemic theory (Magal et al., 2018), we calculated R_∞, the average fraction of individuals who are infected if an epidemic

occurs in such a population. During such an epidemic, each individual will on average be inoculated (challenged) $\mathfrak{R}_0 R_\infty$ times, where $\mathfrak{R}_0$ is the population-level basic reproduction number (Miller, 2012). We can then calculate the probability that a new variant transmission chain is started in an arbitrary focal individual:

$$\mathfrak{R}_0 R_\infty \sum_{i=1}^{n_s} s_i(0) p_{\text{surv}}(\kappa_m(i), \kappa_w(i), v, b, f_{\text{mt}}) \tag{45}$$

Sensitivity analysis (Fig. A3)

We assessed the sensitivity of our results to parameter choices by re-running our simulation models with randomly generated parameter sets chosen via Latin Hypercube Sampling from across a range of biologically plausible values. Figure A3 gives a summary of the results.

We simulated 50 000 infections of experienced hosts ($c_w = 1$) according to each of 10 random parameter sets. We selected parameter sets using Latin Hypercube sampling to maximize coverage of the range of interest without needing to study all possible permutations. We did this for the following bottleneck sizes: 1, 3, 10, 50.

We analyzed two cases: one in which the immune response is unrealistically early and one in which it is realistically timed. In the unrealistically early antibody response model, t_M varied between $t_M = 0$ and $t_M = 1$. In the realistically timed antibody response model, t_M varied between $t_M = 2$ and $t_M = 4.5$. Other parameter ranges were shared between the two models (Table 2).

Table 2: **Sensitivity analysis parameter ranges shared between models**

Parameter	Minimum value	Maximum value
t_N	6	9
$C_{\max}$	10^8	10^9
$\mathcal{R}_0$	5	15
r	10	500
z_w	0.70	0.99
z_m/z_w	0.5	0.9
μ	0.33×10^{-6}	0.33×10^{-4}
d_v	2	8
k	3	16
c_m	0.5	1
θ	10^6	10^8
v/b	1	50

Discussion of sensitivity analysis results can be found in Section 8 below.

Meta-analysis (Fig. 1)

We downloaded processed variant frequencies and subject metadata from the two NGS studies of immune-competent human subjects naturally infected with A/H3N2 with known vaccination status (Debbink et al., 2017; McCrone et al., 2018) from the study

Github repositories and journal websites. We independently verified that reported antigenic site and antigenic ridge amino acid substitutions were correctly indicated, determined the number of subjects with no NGS-detectable antigenic amino acid substitutions, and produced figures from the aggregated data.

Computational methods

For within-host model stochastic simulations, we used the Poisson Random Corrections (PRC) tau-leaping algorithm (Hu & Li, 2009). We used an adaptive step size; we chose step sizes according to the algorithm of Hu and Li (Hu & Li, 2009) to ensure that estimated next-step mean values were non-negative, with a maximum step size of 0.01 days. Variables were set to zero if events performed during a timestep would have reduced the variable to a negative value. For the sterilizing immunity simulations in Figure 2, we used a smaller maximum step size of 0.001 days in recipient hosts to better handle mutation dynamics involving very small numbers of replicating virions.

We obtained numerical solutions of equations, including systems of differential equations and final size equations, in Python using solvers provided with SciPy (Jones et al., 2001).

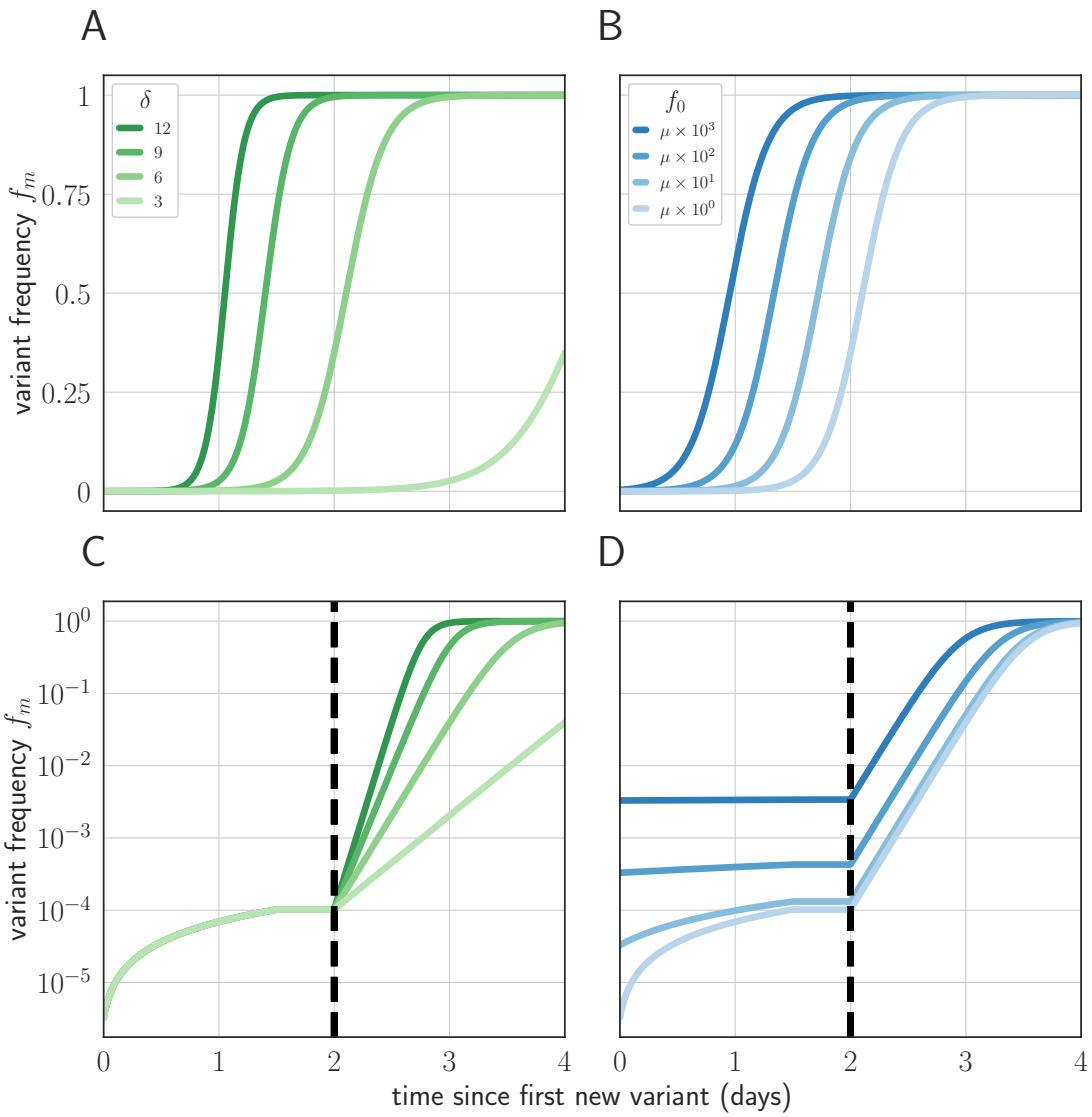

Fig. 7. Variant within-host frequency as a function of time and initial variant frequency, according to derived replicator equation (equations 13, 15). (**A, B**) Variant frequency over time for an initially present new variant. (**A**) selection strength δ varied, with initial frequency f_0 equal to the mutation rate $\mu = 0.33 \times 10^{-5}$. (**B**) initial frequency f_0 varied, with $\delta = 6$. (**C, D**) Variant frequency over time when antigenic selections begins at $t = 2$ days after first variant emergence, with ongoing mutation prior to that point. (**C**) δ varied and f_0 fixed as in (**A**); (**D**) f_0 varied and δ fixed as in (**B**). Parameters as in Table 1 unless otherwise noted.

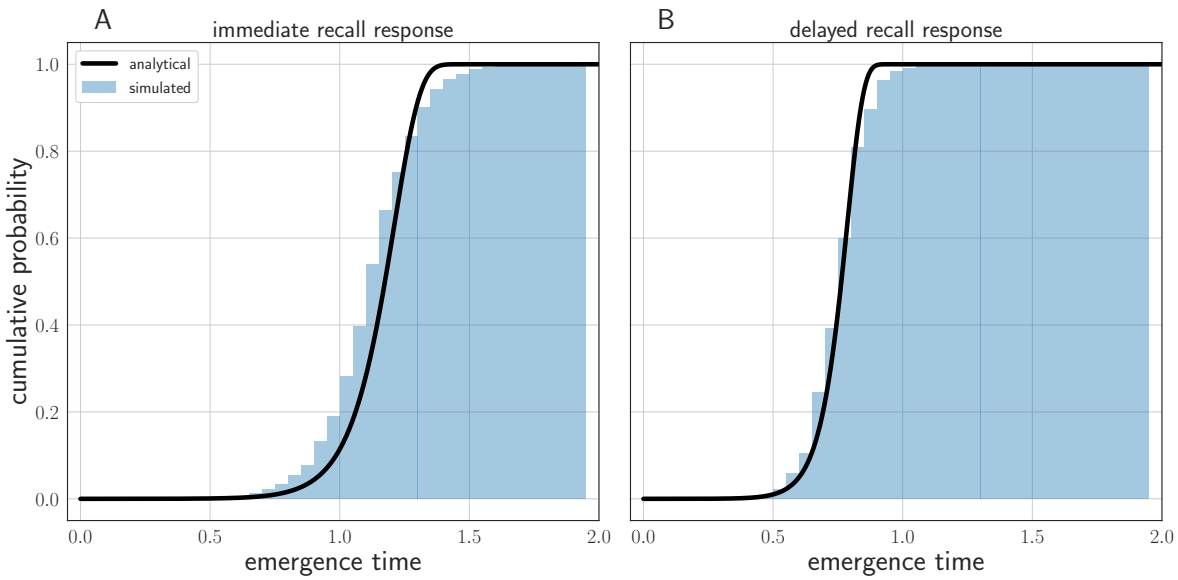

Fig. 8. Comparison of analytically calculated cumulative distribution function (CDF) for time of first successful *de novo* mutation with simulations. Black line shows analytically calculated CDF. Blue cumulative histogram shows distribution of new variant mutation times for 250 000 simulations from the stochastic within-host model with (**A**) an immediate recall response ($t_M = 0$) and (**B**) a realistic recall response at 48 hours post-infection ($t_M = 2$). Other model parameters as parameters in Table 1. Note that time of first successful mutation tends to be later with an immediate recall response than with a delayed recall response. This occurs because the cumulative number of viral replication events grows more slowly in time at the start of the infection because of the strong, immediate recall response.

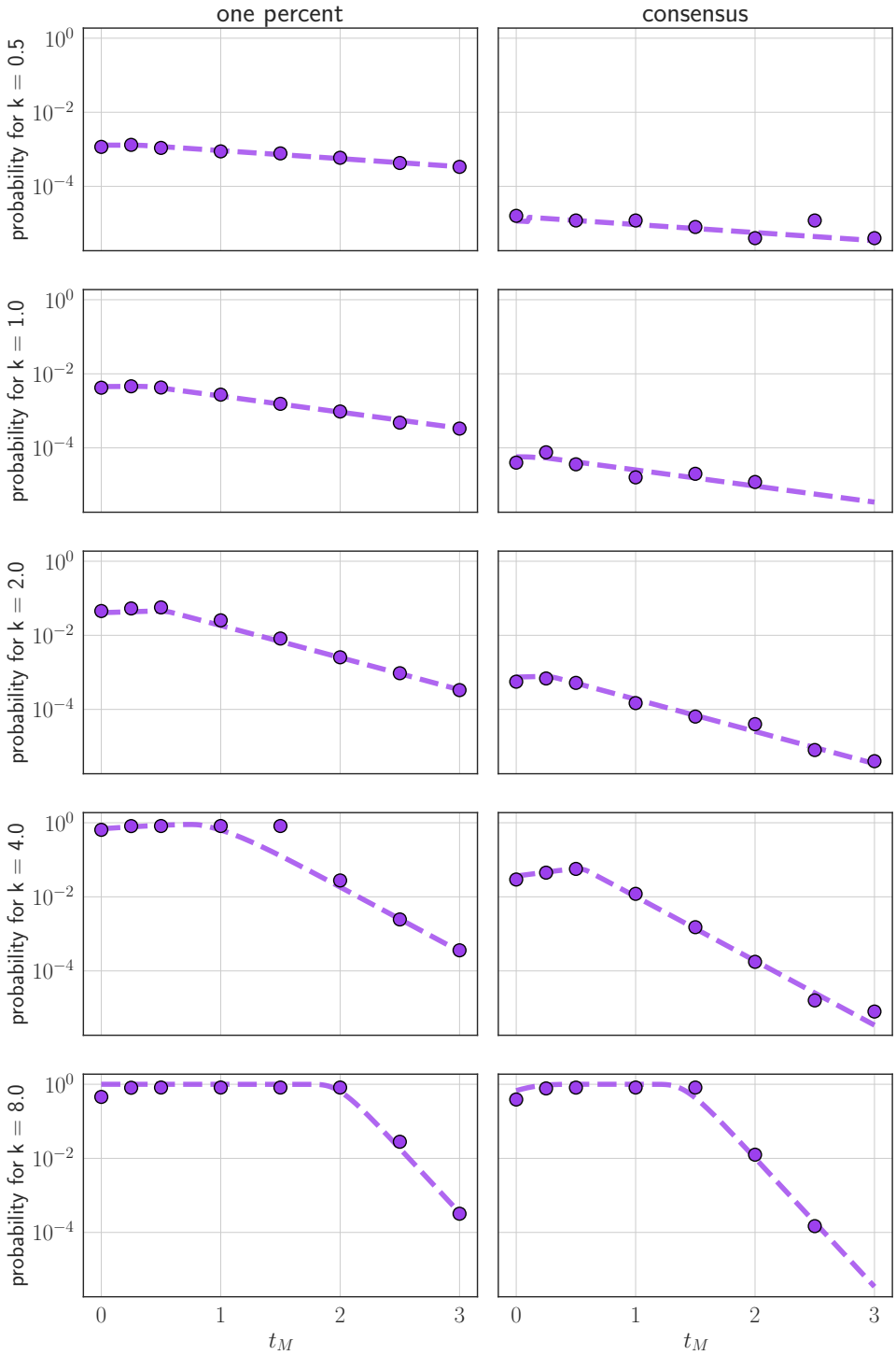

Fig. 9. Comparison of analytically calculated probability of replication selection with stochastic simulations. Probability of replication selection to one percent (left column) or consensus (right column) by $t = 3$ days post-infection as a function of k and t_M for 250 000 simulations from the stochastic within-host model. $c = 0$ (so the fitness difference $\delta = k$). Other model parameters as parameters in Table 1. Dashed lines show analytical prediction and dots show simulation outcomes.

References

Archetti, I., & Horsfall, F. L. (1950). Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *Journal of Experimental Medicine*, *92*(5), 441–462.

Asaduzzaman, S. M., Ma, J., & van den Driessche, P. (2018). Estimation of cross-immunity between drifted strains of influenza A/H3N2. *Bulletin of Mathematical Biology*, *80*(3), 657–669.

Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., & Perelson, A. S. (2006). Kinetics of influenza A virus infection in humans. *Journal of Virology*, *80*(15), 7590–7599.

Ball, F., Britton, T., & Neal, P. (2016). On expected durations of birth–death processes, with applications to branching processes and SIS epidemics. *Journal of Applied Probability*, *53*(1), 203–215.

Bazinet, A. L., Zwickl, D. J., & Cummings, M. P. (2014). A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Systematic Biology*, *63*(5), 812–8. https://doi.org/10.1093/sysbio/syu031

Bedford, T., Rambaut, A., & Pascual, M. (2012). Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biology*, *10*(1), 38.

Bedford, T., Riley, S., Barr, I. G., Broor, S., Chadha, M., Cox, N. J., Daniels, R. S., Gunasekaran, C. P., Hurt, A. C., Kelso, A., Et al. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, *523*(7559), 217.

Bonhoeffer, S., Lipsitch, M., & Levin, B. R. (1997). Evaluating treatment protocols to prevent antibiotic resistance. *Proceedings of the National Academy of Sciences*, *94*(22), 12106–12111.

Boni, M. F., Gog, J. R., Andreasen, V., & Feldman, M. W. (2006). Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1592), 1307–1316.

Burke, D. F., & Smith, D. J. (2014). A recommended numbering scheme for influenza A HA subtypes. *PLoS ONE*, *9*(11), e112302. https://doi.org/10.1371/journal.pone.0112302

Canini, L., Holzer, B., Morgan, S., Dinie Hemmink, J., Clark, B., sLoLa Dynamics Consortium, Woolhouse, M. E., Tchilian, E., & Charleston, B. (2020). Timelines of infection and transmission dynamics of H1N1pdm09 in swine. *PLoS Pathogens*, *16*(7), e1008628.

Cao, P., & McCaw, J. M. (2017). The mechanisms for within-host influenza virus control affect model-based assessment and prediction of antiviral treatment. *Viruses*, *9*(8), 197.

Caton, A. J., Brownlee, G. G., Yewdell, J. W., & Gerhard, W. (1982). The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell*, *31*(2), 417–427.

Clements, M., Betts, R., Tierney, E., & Murphy, B. (1986). Serum and nasal wash antibodies associated with resistance to experimental challenge with influenza A wild-type virus. *Journal of Clinical Microbiology*, *24*(1), 157–160.

Coro, E. S., Chang, W. W., & Baumgarth, N. (2006). Type I IFN receptor signals directly stimulate local B cells early following influenza virus infection. *Journal of Immunology*, *176*(7), 4343–4351.

Coudeville, L., Bailleux, F., Riche, B., Megas, F., Andre, P., & Ecochard, R. (2010). Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: Development and application of a Bayesian random-effects model. *BMC Medical Research Methodology*, *10*(1), 18.

Davenport, F. M., & Hennessy, A. V. (1956). A serologic recapitulation of past experiences with influenza A; antibody response to monovalent vaccine. *Journal of Experimental Medicine*, *104*(1), 85–97.

Davis, A. K., McCormick, K., Gumina, M. E., Petrie, J. G., Martin, E. T., Xue, K. S., Bloom, J. D., Monto, A. S., Bushman, F. D., & Hensley, S. E. (2018). Sera from individuals with narrowly focused influenza virus antibodies rapidly select viral escape mutations in ovo. *Journal of Virology*, *92*(19), e00859–18.

De Jong, J., Smith, D. J., Lapedes, A., Donatelli, I., Campitelli, L., Barigazzi, G., Van Reeth, K., Jones, T. C., Rimmelzwaan, G., Osterhaus, A., Et al. (2007). Antigenic and genetic evolution of swine influenza A (H3N2) viruses in europe. *Journal of Virology*, *81*(8), 4315–4322.

Debbink, K., McCrone, J. T., Petrie, J. G., Truscon, R., Johnson, E., Mantlo, E. K., Monto, A. S., & Lauring, A. S. (2017). Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathogens*, *13*(1), e1006194. https://doi.org/10.1371/journal.ppat.1006194

Dinis, J. M., Florek, N. W., Fatola, O. O., Moncla, L. H., Mutschler, J. P., Charlier, O. K., Meece, J. K., Belongia, E. A., & Friedrich, T. C. (2016). Deep sequencing reveals potential antigenic variants at low frequencies in influenza A virus-infected humans. *Journal of Virology*, *90*(7), 3355–3365.

Ferguson, N. M., Galvani, A. P., & Bush, R. M. (2003). Ecological and immunological determinants of influenza evolution. *Nature*, *422*(6930), 428.

Fonville, J. M., Wilks, S., James, S. L., Fox, A., Ventresca, M., Aban, M., Jones, T. C., Le, N., Pham, Q., Et al. (2014). Antibody landscapes after influenza virus infection or vaccination. *Science*, *346*(6212), 996–1000.

Frensing, T., Kupke, S. Y., Bachmann, M., Fritzsche, S., Gallo-Ramirez, L. E., & Reichl, U. (2016). Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in mdck cells. *Applied Microbiology and Biotechnology*, *100*(16), 7181–7192.

Ghafari, M., Lumby, C. K., Weissman, D., & Illingworth, C. J. R. (2020). Inferring transmission bottleneck size from viral sequence data using a novel haplotype reconstruction method. *bioRxiv*.

Gog, J. R. (2008). The impact of evolutionary constraints on influenza dynamics. *Vaccine*, *26*, C15–C24.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, *303*(5656), 327–332.

Hadjichrysanthou, C., Cauët, E., Lawrence, E., Vegvari, C., de Wolf, F., & Anderson, R. M. (2016). Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. *Journal of The Royal Society Interface*, *13*(119), 20160289.

Han, A. X., Maurer-Stroh, S., & Russell, C. A. (2019). Individual immune selection pressure has limited impact on seasonal influenza virus evolution. *Nature Ecology & Evolution*, *3*(2), 302.

Hartfield, M., & Alizon, S. (2014). Epidemiological feedbacks affect evolutionary emergence of pathogens. *The American Naturalist*, *183*(4), E105–E117.

Hartfield, M., & Alizon, S. (2015). Within-host stochastic emergence dynamics of immune-escape mutants. *PLoS Computational Biology*, *11*(3), e1004149.

Hensley, S. E., Das, S. R., Bailey, A. L., Schmidt, L. M., Hickman, H. D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J. R., Et al. (2009). Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science*, *326*(5953), 734–736.

Hoft, D. F., Lottenbach, K. R., Blazevic, A., Turan, A., Blevins, T. P., Pacatte, T. P., Yu, Y., Mitchell, M. C., Hoft, S. G., & Belshe, R. B. (2017). Comparisons of the humoral and cellular immune responses induced by live attenuated influenza vaccine and inactivated influenza vaccine in adults. *Clin. Vaccine Immunol.*, *24*(1), e00414–16.

Hu, Y., & Li, T. (2009). Highly accurate tau-leaping methods with random corrections. *Journal of Chemical Physics*, *130*(12), 03B618.

Iwasa, Y., Michor, F., & Nowak, M. A. (2004). Evolutionary dynamics of invasion and escape. *Journal of Theoretical Biology*, *226*(2), 205–214.

Javaid, W., Ehni, J., Gonzalez-Reiche, A. S., Carreno, J. M., Hirsch, E., Tan, J., Khan, Z., Kriti, D., Ly, T., Kranitzky, B., Barnett, B., Cera, F., Prespa, L., Moss, M., Albrecht, R. A., Mustafa, A., Herbison, I., Hernandez, M. M., Pak, T., ... van Bakel, H. (2020). Real time investigation of a large nosocomial influenza a outbreak informed by genomic epidemiology. *medRxiv*. https://doi.org/10.1101/2020.05.10.20096693

Jones, E., Oliphant, T., Peterson, P., Et al. (2001). SciPy: Open source scientific tools for Python. http://www.scipy.org/

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kennedy, D. A., & Read, A. F. (2017). Why does drug resistance readily evolve but vaccine resistance does not? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1851), 20162562.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624–626.

Koel, B. F., Burke, D. F., Bestebroer, T. M., van der Vliet, S., Zondag, G. C., Vervaet, G., Skepner, E., Lewis, N. S., Spronken, M. I., Russell, C. A., Et al. (2013). Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science*, *342*(6161), 976–979. https://doi.org/10.1126%2Fscience.1244730

Koel, B. F., van Someren Greve, F., Vigeveno, R. M., Pater, M., Russell, C. A., & de Jong, M. D. (2019). Disparate evolution of virus populations in upper and lower airways of mechanically ventilated patients. *bioRxiv*, 509901.

Koelle, K., Cobey, S., Grenfell, B. T., & Pascual, M. (2006). Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*, *314*(5807), 1898–1903. https://doi.org/10.1126/science.1132745

Koelle, K., Kamradt, M., & Pascual, M. (2009). Understanding the dynamics of rapidly evolving pathogens through modeling the tempo of antigenic change: Influenza as a case study. *Epidemics*, *1*(2), 129–137.

Koelle, K., & Rasmussen, D. A. (2015). The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. *eLife*, *4*, e07361.

Krammer, F. (2019). The human antibody response to influenza A virus infection and vaccination. *Nature Reviews Immunology*, 1.

Krammer, F., & Palese, P. (2019). Universal influenza virus vaccines that target the conserved hemagglutinin stalk and conserved sites in the head domain. *Journal of Infectious Diseases*, *219*(Supplement_1), S62–S67.

Kucharski, A. J., & Gog, J. R. (2011). Influenza emergence in the face of evolutionary constraints. *Proc. R. Soc. Lond. B: Biol. Sci.*

Kucharski, A. J., Lessler, J., Read, J. M., Zhu, H., Jiang, C. Q., Guan, Y., Cummings, D. A., & Riley, S. (2015). Estimating the life course of influenza A (H3N2) antibody responses from cross-sectional data. *PLoS Biology*, *13*(3), e1002082.

Lam, J. H., & Baumgarth, N. (2019). The multifaceted B cell response to influenza virus. *Journal of Immunology*, *202*(2), 351–359.

Lau, L. L., Cowling, B. J., Fang, V. J., Chan, K.-H., Lau, E. H., Lipsitch, M., Cheng, C. K., Houck, P. M., Uyeki, T. M., Peiris, J. M., Et al. (2010). Viral shedding and clinical illness in naturally acquired influenza virus infections. *Journal of Infectious Diseases*, *201*(10), 1509–1516.

Le Sage, V., Jones, J. E., Kormuth, K. A., Fitzsimmons, W. J., Nturibi, E., Padovani, G. H., Arevalo, C. P., French, A. J., Avery, A. J., Manivanh, R., McGrady, E. E., Bhagwat, A. R., Lauring, A. S., Hensley, S. E., & Lakdawala, S. S. (2020). Pre-existing immunity provides a barrier to airborne transmission of influenza viruses. *bioRxiv*. https://doi.org/10.1101/2020.06.15.103747

Lee, J. M., Eguia, R., Zost, S. J., Choudhary, S., Wilson, P. C., Bedford, T., Stevens-Ayers, T., Boeckh, M., Hurt, A., Lakdawala, S. S., Hensley, S. E., & Bloom, J. D. (2019). Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *bioRxiv*, 670497.

Leonard, A. S., McClain, M. T., Smith, G. J., Wentworth, D. E., Halpin, R. A., Lin, X., Ransier, A., Stockwell, T. B., Das, S. R., Gilbert, A. S., Et al. (2016). Deep sequencing of influenza A virus from a human challenge study reveals a selective bottleneck and only limited intrahost genetic diversification. *Journal of Virology*, *90*(24), 11247–11258.

Lessler, J., Riley, S., Read, J. M., Wang, S., Zhu, H., Smith, G. J., Guan, Y., Jiang, C. Q., & Cummings, D. A. (2012). Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathogens*, *8*(7), e1002802.

Lewis, N. S., Russell, C. A., Langat, P., Anderson, T. K., Berger, K., Bielejec, F., Burke, D. F., Dudas, G., Fonville, J. M., Fouchier, R. A., Et al. (2016). The global antigenic diversity of swine influenza A viruses. *eLife*, *5*, e12217.

Linderman, S. L., Chambers, B. S., Zost, S. J., Parkhouse, K., Li, Y., Herrmann, C., Ellebedy, A. H., Carter, D. M., Andrews, S. F., Zheng, N.-Y., Et al. (2014). Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013–2014 influenza season. *Proceedings of the National Academy of Sciences*, *111*(44), 15798–15803.

Lumby, C. K., Nene, N. R., & Illingworth, C. J. (2018). A novel framework for inferring parameters of transmission from viral sequence data. *PLoS Genetics*, *14*(10), e1007718.

Lumby, C. K., Zhao, L., Breuer, J., & Illingworth, C. J. R. (2020). A large effective population size for within-host influenza virus infection. *eLife*, *9*(e56915). https://doi.org/10.7554/eLife.56915

Luo, S., Reed, M., Mattingly, J. C., & Koelle, K. (2012). The impact of host immune status on the within-host and population dynamics of antigenic immune escape. *Journal of The Royal Society Interface*, *9*(75), 2603–2613.

Magal, P., Seydi, O., & Webb, G. (2018). Final size of a multi-group SIR epidemic model: Irreducible and non-irreducible modes of transmission. *Mathematical Biosciences*, *301*, 59–67.

Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., & Klenk, H.-D. (2004). Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. *Journal of Virology*, *78*(22), 12665–12667.

McCrone, J. T., Woods, R. J., Martin, E. T., Malosh, R. E., Monto, A. S., & Lauring, A. S. (2018). Stochastic processes constrain the within and between host evolution of influenza virus. *eLife*, *7*, e35962.

Memoli, M. J., Han, A., Walters, K.-A., Czajkowski, L., Reed, S., Athota, R., Angela Rosas, L., Cervantes-Medina, A., Park, J.-K., Morens, D. M., Et al. (2019). Influenza A reinfection in sequential human challenge: Implications for protective immunity and universal vaccine development. *Clinical Infectious Diseases*, ciz281.

Miller, J. C. (2012). A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 1–17.

Moncla, L. H., Zhong, G., Nelson, C. W., Dinis, J. M., Mutschler, J., Hughes, A. L., Watanabe, T., Kawaoka, Y., & Friedrich, T. C. (2016). Selective bottlenecks shape evolutionary pathways taken during mammalian adaptation of a 1918-like avian influenza virus. *Cell Host & Microbe*, *19*(2), 169–180.

Neuzil, K. M., Jackson, L. A., Nelson, J., Klimov, A., Cox, N., Bridges, C. B., Dunn, J., DeStefano, F., & Shay, D. (2006). Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naive 5–8-year-old children. *Journal of Infectious Diseases*, *194*(8), 1032–1039.

Nobusawa, E., & Sato, K. (2006). Comparison of the mutation rates of human influenza A and B viruses. *Journal of Virology*, *80*(7), 3675–3678.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, *23*(1), 263–286.

Pawelek, K. A., Huynh, G. T., Quinlivan, M., Cullinane, A., Rong, L., & Perelson, A. S. (2012). Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Computational Biology*, *8*(6), e1002588.

Perelson, A. S., Rong, L., & Hayden, F. G. (2012). Combination antiviral therapy for influenza: Predictions from modeling of human infections. *Journal of Infectious Diseases*, *205*(11), 1642–1645.

Perron, G. G., Inglis, R. F., Pennings, P. S., & Cobey, S. (2015). Fighting microbial drug resistance: A primer on the role of evolutionary biology in public health. *Evolutionary Applications*, *8*(3), 211–222.

Petrova, V. N., & Russell, C. A. (2018). The evolution of seasonal influenza viruses. *Nature Reviews Microbiology*, *16*(1), 47.

Quinlivan, M., Nelly, M., Prendergast, M., Breathnach, C., Horohov, D., Arkins, S., Chiang, Y.-W., Chu, H.-J., Ng, T., & Cullinane, A. (2007). Pro-inflammatory and antiviral cytokine expression in vaccinated and unvaccinated horses exposed to equine influenza virus. *Vaccine*, *25*(41), 7056–7064.

Recker, M., Pybus, O. G., Nee, S., & Gupta, S. (2007). The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proceedings of the National Academy of Sciences*, *104*(18), 7711–7716.

Russell, C. A., Fonville, J. M., Brown, A. E., Burke, D. F., Smith, D. L., James, S. L., Herfst, S., Van Boheemen, S., Linster, M., Schrauwen, E. J., Katzelnick, L., Mosterín, A., Kuiken, T., Maher, E., Neumann, G., Osterhaus, A. D. M. E., Kawaoka, Y., Fouchier, R. A. M., & Smith, D. J. (2012). The potential for respiratory droplet–transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science*, *336*(6088), 1541–1547.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., Et al. (2008). The global circulation of seasonal influenza A (H3N2) viruses. *Science*, *320*(5874), 340–346.

Sigal, D., Reid, J. N., & Wahl, L. M. (2018). Effects of transmission bottlenecks on the diversity of influenza a virus. *Genetics*, *210*(3), 1075–1088.

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D., & Fouchier, R. A. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science*, *305*(5682), 371–376.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–3. https://doi.org/10.1093/bioinformatics/btu033

Strelkowa, N., & Lässig, M. (2012). Clonal interference in the evolution of influenza. *Genetics*, *192*(2), 671–682. https://doi.org/10.1534/genetics.112.143396

Suess, T., Remschmidt, C., Schink, S. B., Schweiger, B., Heider, A., Milde, J., Nitsche, A., Schroeder, K., Doellinger, J., Braun, C., Et al. (2012). Comparison of shedding characteristics of seasonal influenza virus (sub) types and influenza A (H1N1) pdm09; Germany, 2007–2011. *PloS ONE*, *7*(12), e51653.

Tsang, T. K., Cowling, B. J., Fang, V. J., Chan, K.-H., Ip, D. K., Leung, G. M., Peiris, J. M., & Cauchemez, S. (2015). Influenza A virus shedding and infectivity in households. *Journal of Infectious Diseases*, *212*(9), 1420–1428.

Valesano, A. L., Fitzsimmons, W. J., McCrone, J. T., Petrie, J. G., Monto, A. S., Martin, E. T., & Lauring, A. S. (2019). Influenza b viruses exhibit lower within-host diversity than influenza a viruses in human hosts. *BioRxiv*, 791038.

Varble, A., Albrecht, R. A., Backes, S., Crumiller, M., Bouvier, N. M., Sachs, D., García-Sastre, A., Et al. (2014). Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host & Microbe*, *16*(5), 691–700.

Volkov, I., Pepin, K. M., Lloyd-Smith, J. O., Banavar, J. R., & Grenfell, B. T. (2010). Synthesizing within-host and population-level selective pressures on viral populations: The impact of adaptive immunity on viral immune escape. *Journal of The Royal Society Interface*, *7*(50), 1311–1318.

Wang, Y.-Y., Harit, D., Subramani, D. B., Arora, H., Kumar, P. A., & Lai, S. K. (2017). Influenza-binding antibodies immobilise influenza viruses in fresh human airway mucus. *European Respiratory Journal*, *49*(1), 1601709.

Weis, J. F., Baeten, J. M., McCoy, C. O., Warth, C., Donnell, D., Thomas, K. K., Hendrix, C. W., Marzinke, M. A., Mugo, N., Matsen IV, F. A., Et al. (2016). Prep-selected drug resistance decays rapidly after drug cessation. *AIDS*, *30*(1), 31.

Welkers, M. R., Pawestri, H. A., Fonville, J. M., Sampurno, O. D., Pater, M., Holwerda, M., Han, A. X., Russell, C. A., Jeeninga, R. E., Setiawaty, V., Et al. (2019). Genetic diversity and host adaptation of avian H5N1 influenza viruses during human infection. *Emerging Microbes & Infections*, *8*(1), 262–271.

Wikramaratna, P. S., Sandeman, M., Recker, M., & Gupta, S. (2013). The antigenic evolution of influenza: drift or thrift? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120200.

Wiley, D., Wilson, I., & Skehel, J. (1981). Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature*, *289*(5796), 373.

Wilker, P. R., Dinis, J. M., Starrett, G., Imai, M., Hatta, M., Nelson, C. W., OConnor, D. H., Hughes, A. L., Neumann, G., Kawaoka, Y., Et al. (2013). Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nature Communications*, *4*, 2636.

Wrammert, J., Smith, K., Miller, J., Langley, W. A., Kokko, K., Larsen, C., Zheng, N.-Y., Mays, I., Garman, L., Helms, C., Et al. (2008). Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature*, *453*(7195), 667.

Xue, K. S., & Bloom, J. D. (2019). Reconciling disparate estimates of viral genetic diversity during human influenza infections. *Nature Genetics*, *51*(9), 1298–1301. https://doi.org/https://doi.org/10.1038/s41588-019-0349-3

Xue, K. S., & Bloom, J. D. (2020). Linking influenza virus evolution within and between human hosts. *Virus evolution*, *6*(1), veaa010.

Xue, K. S., Stevens-Ayers, T., Campbell, A. P., Englund, J. A., Pergam, S. A., Boeckh, M., & Bloom, J. D. (2017). Parallel evolution of influenza across multiple spatiotemporal scales. *eLife*, *6*, e26875. https://doi.org/10.7554/eLife.26875

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zinder, D., Bedford, T., Gupta, S., & Pascual, M. (2013). The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS Pathogens*, *9*(1), e1003104.

Zuccarino-Catania, G. V., Sadanand, S., Weisel, F. J., Tomayko, M. M., Meng, H., Kleinstein, S. H., Good-Jacobson, K. L., & Shlomchik, M. J. (2014). CD80 and PD-L2 define functionally distinct memory B cell subsets that are independent of antibody isotype. *Nature Immunology*, *15*(7), 631.

Acknowledgments

We thank Daniel B. Cooney, Andrea L. Graham, Joseph Gibson, Katelyn M. Gostic, Alvin X. Han, Chadi M. Saad-Roy, and Edward C. Schrom for helpful discussions. We thank Christopher J. Illingworth, Daniel B. Weissman, and an anonymous reviewer for helpful comments on a prior version of this manuscript.

We thank the GISAID Initiative and the influenza surveillance and research groups who openly shared the genetic sequence data used in this work (a table with accession numbers and associated labs is available online in the project materials and data repositories on Github and OSF, see below for links).

Author contributions

CAR, DHM, and VNP conceived and designed the research. DHM created and analyzed the mathematical models, with support from FWR, RAN, and SAL. DHM analyzed the within-host data. EP conducted phylogenetic analyses. DHM produced the figures. All authors contributed substantially to the interpretation of results and the writing and revising of the manuscript.

Competing interests

The authors declare no competing interests.

Data and materials availability

All code, data, and other materials needed to reproduce the analysis in this paper are provided online on the project Github repository:
https://github.com/dylanhmorris/asynchrony-influenza

They are also available on OSF:
https://doi.org/10.17605/OSF.IO/jdsbp/

Output data generated by stochastic simulations, within-host NGS meta-analysis, and phylogenetic analyses are archived on OSF:
https://doi.org/10.17605/OSF.IO/jdsbp/