Published in final edited form as:
Vaccine. 2008 July 18; 26(Suppl 3): C8–14.

Vaccination and antigenic drift in influenza

Maciej F. Boni
Resources for the Future, Washington, DC 20036; Department of Ecology and Evolutionary Biosylogy, Princeton University, Princeton, NJ 08540

Abstract

The relationship between influenza antigenic drift and vaccination lies at the intersection of evolutionary biology and public health, and it must be viewed and analyzed in both contexts simultaneously. In this paper, I review what is known about the effects of antigenic drift on vaccination and the effects of vaccination on antigenic drift, and I suggest some simple ways to detect the presence of antigenic drift in seasonal influenza data. If antigenic drift occurs on the time scale of a single influenza season, it may be associated with the presence of herd immunity at the beginning of the season and may indicate a need to monitor for vaccine updates at the end of the season. The relationship between antigenic drift and vaccination must also be viewed in the context of the global circulation of influenza strains and the seeding of local and regional epidemics. In the data sets I consider — from New Zealand, New York, and France — antigenic drift can be statistically detected during some seasons, and seeding of epidemics appears to be endogenous sometimes and exogenous at other times. Improved detection of short-term antigenic drift and epidemic seeding would significantly benefit influenza monitoring efforts and vaccine selection.

Keywords

influenza; antigenic drift; vaccination; evolution; haemagglutinin

Introduction

Influenza infection in human populations in characterized by seasonal epidemics in the temperate zones of both hemispheres and endemicity in tropical and subtropical regions [**? ?**]. Vaccination for influenza epidemics occurs biannually — once in March for southern hemisphere populations and a second time in September for northern hemisphere populations — with the goal of timing vaccination to precede each hemisphere's annual winter epidemic [?]. During the epidemic season, surveillance is carried out to monitor for the appearance of novel strains that elude human immunity and would indicate a need to change the current vaccine composition [? ?]. These novel strains, sometimes called immune-escape variants, are the influenza virus's evolutionary adaptations to a strong population-wide immune response. The adaptations usually involve several amino acid changes to influenza's haemagglutinin (HA) protein, the antigen responsible for entry into host epithelial cells, and the detection of which is a primary stimulus of host immune response [?]. By accumulating amino acid changes, the HA protein is said to "drift" from one form, recognizable by host antibodies, to another form which is less recognizable and more successful at infecting vaccinated and unvaccinated

hosts alike; this process is called antigenic drift [?]. Because antigenic drift can decrease a vaccine's efficacy, influenza strains are reviewed in February (for the northern hemisphere) and September (for the southern hemisphere) to detect new immune-escape variants and determine if an update to the vaccine composition is necessary [?]. Antigenic drift is responsible for the periodic need to change global vaccine composition and, since 1968, has forced vaccine updates approximately every two to five years [? ?].

For several reasons, an understanding of the short-term evolution of the influenza virus is critical for formulating vaccine policy. First, as already mentioned, short-term influenza evolution gives rise to immune-escape variants that force updates in vaccine composition. Although these novel strains appear only every two to five years, surveillance occurs on the level of regional and local epidemics — which last between three and six months [? ?] — since it is these epidemics that harbor the majority of viral reproduction and are most likely to be the source of novel immunity-evading strains. Second, vaccination may have quite a strong effect on short-term influenza evolution, as has already been documented in an avian H5N2 lineage in vaccinated chickens [?]. Standard population-genetic theory suggests that a similar process would occur in human influenza; if vaccination confers partial or imperfect immunity, a highly immune or vaccinated population can selectively pressure the virus population to evolve more quickly than usual [?]. Finally, understanding influenza evolution can help us determine the evolutionary history of the many different strains that are isolated during local influenza epidemics [? ? ? ? ?], and knowing the recent evolutionary history of influenza will help us determine how multiple lineages of influenza coexist, coevolve, and compete during an influenza season and across multiple seasons. Understanding influenza dynamics at this scale may help us determine where and when surveillance will be most effective.

In analyzing short-term influenza evolution, one can look at individual nucleotide changes in the viral haemagglutinin and ask at what time scale antigenic drift occurs. Strain differences are observable during a single epidemic season, even in a single week, but the cause of this variation is usually unknown. A plausible hypothesis is that the observed variation is a transient phase of antigenic drift (a process driven by mutation and positive selection or by mutation alone) in which the viruses are mutating away from a particular ancestral strain that seeded this season's epidemic. Antigenic drift may be diffcult to detect because there are a number of other processes — reassortment, purifying selection, non-random mixing of hosts, stochastic effects during the transmission bottleneck, and immigration of strains from other populations — that generate and restrict variation during the course of an epidemic. Antigenic drift is easily observable across multiple seasons since viruses isolated many years apart are usually more different than viruses isolated a few years apart. As a short-term process, however, we do not yet know whether antigenic drift builds up gradually during annual epidemics, or whether it is a more punctuated and stochastic process which occurs at the whim of the arrival of beneficial immune-escape mutations. If antigenic drift is indeed observable during a single epidemic, surveillance efforts will need to focus on isolating the latest possible strains to determine the closest possible vaccine match. If antigenic drift is not observable on the time scale of a single epidemic, then choosing vaccine strains from the general variation present in the virus population can be done equally well in December, February, or April, since isolates from any time during the epidemic can be considered potential candidates for updating the vaccine composition. Note that a complete analysis of vaccine selection and antigenic drift would include detail on the numbers and locations of amino acid changes, especially in antigenic sites; we would need to be precise in drawing conclusions about antigenic drift on the nucleotide level versus antigenic drift on the amino acid level.

This article presents several data sets that shed light on simple questions about short-term influenza evolution on a time scale of one to two years and reviews what is known about the relationship between antigenic drift and vaccination in influenza.

Antigenic drift on a seasonal scale

Antigenic drift can have one simple effect on vaccination policy: if antigenic drift is observable on a local or regional geographic scale and on a season-long temporal scale, then monitoring for immune-escape variants must take into account that strains isolated later in a season will have a higher probability of immune escape. Testing if antigenic drift is observable on a short time scale requires dozens of sequences isolated in a single region during a single epidemic with the date of isolation included for each strain. Fortunately, such data are available thanks to the efforts of the NIAID-funded Influenza Genome Sequencing Project (http://www.niaid.nih.gov/dmid/genomes/mscs/influenza.htm). The data pooled for this study were 151 sequences from three different influenza seasons in New York state (a subset of the 156 sequences analyzed by ?]); 415 sequences isolated over six seasons in New Zealand; and 92 sequences from the 1999—2000 season in France (the data set analyzed by ?]). The IGSP-sequences, from New York and New Zealand, include the entire haemagglutinin segment which is about 1700nt long, while the French sequences are of the haemagglutinin's HA1 domain which is 987nt long.

For each influenza season and location, we would like to determine if antigenic drift is occurring. This is somewhat diffcult since we do not know which strains in the data evolved from other strains in the data, and which strains were originally introduced to begin the epidemic. Some of these relationships can be inferred with phylogenetic analysis or with knowledge of surrounding epidemics and the immigration of geographically proximate strains (e.g., see the analysis in [? ?]). Without this extra information, we can ask a more basic question: when sampling strains in a single epidemic, are pairs of strains that are more distant in time also likely to be genetically more distinct? Put more simply, is there a positive correlation between temporal distance and genetic distance for pairs of strains isolated in a single influenza epidemic? This question can be answered statistically with a Mantel test, and the answer is that this correlation is significant in some epidemics but not in others. Table 1 summarizes the results for the different data sets considered here.

Two of the seasons in particular, the 1999-2000 influenza season in New York and the 2002 season in New Zealand, appear to exhibit a strong correlation between temporal and genetic distance indicating that short-term antigenic drift was observable during these seasons. In these two cases, monitoring late in the season would have been more likely to uncover divergent strains, which may have had a higher likelihood of immune escape. It is not clear whether strains isolated late in the season are more likely to seed the next epidemic season in the same location; if this were the case, then isolating strains late in an epidemic in which antigenic drift is observable would provide a picture of the strains that will be circulating during the next epidemic season.

Five of the epidemics studied included temporal outliers; these are summertime isolates that do not clearly belong to the following or preceding epidemic season. The data from the 2000 H1N1 season in New Zealand comprise 38 isolates spanning 255 days, with a weak correlation between temporal and genetic distance ($p = .081$). However, one of these strains, outlier #1 from Table 2, is a summertime isolate. This outlier is temporally the most distant to the remaining 37 strains in that season and turns out to be genetically quite distinct as well (only two other strains during the 2000 H1N1 season have a higher mean pairwise distance to the remaining 37 strains of that season). When this outlier is removed, the epidemic season spans only 126 days and there is no observable temporal-genetic correlation among the H1N1 strains during the 2000 season in New Zealand.

It appears that the strain variation in the 2000 H1N1 season in New Zealand is not characterized by antigenic drift; the same can be said for the variation present during the 2000 H3N2 season

in New Zealand, the 2001-2002 and 2003-2004 seasons in New York, and the 1999-2000 season in France. Therefore, detection of highly divergent isolates would not have been any more likely in the early months of these seasons than during the late months. In general, for epidemics characterized by general non-drift variation, there is no reason to believe that isolates sampled late in the season would be better potential candidates for updates in vaccine composition.

Finally, recent theoretical work has shown that the rate of antigenic drift does not have to be constant during an epidemic and that this rate tends be higher in the early phases of the epidemic [?]. This means that for seasons in which short-term antigenic drift is observable, early sampling of isolates may occur during a period of rapid evolution and/or diversification. Any conclusions based on data collected during this time should be viewed in their proper evolutionary context.

Geographic considerations

To this point, we have ignored the fact that evolution in influenza occurs on a time scale that includes highly varying viral population sizes in the form of annual epidemics and summertime bottlenecks. Understanding how the influenza virus population evolves from one epidemic, across the virus-unfriendly summertime, and into the subsequent epidemic is a crucial step in resolving the mechanisms that drive influenza's long-term evolution. The population dynamics of host interactions and disease transmission play a central role in influenza persistence, spread, and evolution during annual epidemics. During the summer, infection numbers are very low, the viral population passes through a genetic bottleneck, and chance fixation of strains is possible through the action of random genetic drift [?]. Some of the evolutionary consequences of the summertime bottleneck have been studied by ?].

In analyzing influenza dynamics across two seasons, there are two fundamental features of influenza population biology that need to be understood. First, is the seeding of epidemics in the autumn an endogenous or exogenous event? In other words, is a New Zealand strain isolated in the fall an immigrant from a neighboring region such as Australia (exogenous seeding)? Or, is this early isolate a strain which persisted from the previous springtime epidemic tail, through the New Zealand summer, and into the initial epidemic wave in the autumn (endogenous seeding)? Second, is seeding of epidemics a unique event? If so, all strains isolated in an epidemic would be descendants of the original epidemic-causing strain. ?] have suggested that influenza epidemics in New York state are seeded multiple times. When seeding is endogenous and introductions are unique events, epidemics may exhibit the classic herald wave as described by ?].

Turning again to the New Zealand data, epidemic seeding can be investigated by considering pairs of consecutive epidemics. A Mantel-like test can be used to test for a temporal-genetic correlation in pairs of strains, where a pair consists of one strain from each epidemic. For example, for the H1N1 epidemics of 2000 and 2001, a 38 × 70 matrix was populated with nucleotide distances between the 38 strains from the 2000 season and the 70 strains from the 2001 season, and a second 38 × 70 matrix was populated with the corresponding temporal distances (in days) between the 2000-strains and the 2001-strains. A correlation r^* was computed between these two matrices; then, the rows and columns of one of the matrices were reshuffled 10,000 times and 10,000 r-values were recomputed. The fraction of recomputed r-values that were greater than the original r^* is reported in the column headed "Mantel-like p-value" of Table 3.

If seeding of epidemics were endogenous, and if antigenic drift were occurring on the nucleotide level during two consecutive seasons, we should expect to observe a temporal-genetic correlation in pairs of strains isolated across these two seasons. In other words, strains

isolated 7 months apart would be genetically more related than strains isolated 17 months apart. If seeding were exogenous, no such temporal-genetic correlation should be observed, unless the population were strongly coupled to other nearby populations from which the exogenous strain introductions occurred. Testing whether a population is closed or strongly coupled to neighboring populations is a question that would need to be addressed with a phylogoegraphic analysis of influenza strains [? ?].

Assuming that New Zealand is a closed population with regard to influenza epidemics, we attempt to identify endogenous seeding by detecting a temporal-genetic correlation across two consecutive epidemic seasons. Again, we need to be aware of temporal outliers that can give a false impression of a temporal-genetic correlation. For the 2000 and 2001 H1N1 seasons, one temporal outlier (#4) has a major effect on the correlation analysis. Temporally, this outlier falls squarely between the two epidemics; genetically, it appears to cluster with the the strains circulating during the 2000 epidemic. Removing this outlier indicated that for pairs of strains isolated from the 2000 and 2001 H1N1 seasons there is no association between their temporal distance and their genetic distance (see Table 3). This precludes the possibility of strains isolated late in the 2000 season heralding the early strains of the 2001 season.

For three pairs of H3N2 seasons — 2000/2001, 2001/2002, and 2004/2005 — it appears that epidemic seeding may have been endogenous. For these seasons, depending on the persistence and evolution of the viruses in the summertime, the late springtime isolates of one season may have heralded the early autumn isolates of the following season. Heralding seems to work well from 2004 to 2005 — mean pairwise distance between isolates from the last two weeks of 2004 to the first three isolates of 2005 is 4.29nt — but not as well in the other two cases. Neither heralding nor endogenous seeding occured from 2002 to 2003 because of the arrival in early 2003 of the novel A/Fujian/411/02-like strains (from outside New Zealand). The Mantel-like correlation analysis does not reveal anything about the dynamics and evolution of influenza across the 2003—2004 seaons, nor across the 2000-2001 H1N1 seasons. Note that summertime outliers can have a quite a large effect on the correlation analysis, and that their phylogenetic relationship to the wintertime strains can be quite variable. Outlier #4 seems to cluster with strains isolated during the *preceding* winter. Outliers #5 and #6 seem to cluster with strains isolated in the *following* winter. Outlier #7 seems to bridge the 2003 epidemic with the 2004 epidemic.

Understanding the geographic spread of influenza and the seeding of local and regional epidemics is critical for vaccine policy because it can help focus monitoring efforts. For example, it is well known that careful monitoring of novel variants is of great public health value in China and Southeast Asia [? ?].

Effects of vaccination on antigenic drift

It has been known for some time that strong host immunity is a driving force of antigenic drift in influenza [? ?]. A straightforward population-genetic analysis reveals that this results from strong population-wide immunity (or herd immunity) imposing an environment on the virus where it has low fitness and where new immune-escape mutations can provide quite a significant fitness benefit [?]. Figure 1 presents a simple schematic diagram showing how theory predicts that antigenic drift should depend on host immunity. The vertical axis can be interpreted as the number of nucleotide differences between strains isolated at the beginning of the epidemic and strains isolated at the end of the epidemic. As herd immunity increases, we should expect to see more antigenic drift; however, if immunity is high enough to prevent the population-wide spread of the pathogen, the epidemic cannot take off and the virus does not evolve. Thus, an intermediate amount of population-wide immunity results in the most antigenic drift; at the individual host level, ?] predicted a similar relationship between

immunity and evolution. Note that vaccine-induced immunity and naturally-acquired immunity are not always identical on an individual level (see ?] and references therein), a difference that will affect immunity and viral evolution on a population-wide level. This fact is ignored by many studies; the discussion presented here also considers “vaccinated populations” and “immune populations” as one and the same.

Although testing the hypothesis that herd immunity drives antigenic drift can prove diffcult, there are a small number of studies whose results are consistent with the relationship presented in Figure 1. A natural experiment in Mexican chicken populations showed that population-wide immunity can drive the evolution of H5N2 avian influenza. In contrast to that found in human populations, HA in avian populations is believed to be antigenically stable [? ?] due to a lack of significant natural immunity or vaccine-induced immunity in birds. After the implementation of a vaccination program in Mexican chickens in 1995, ?] observed a subsequent pattern of amino acid changes which indicated that the virus population was mutating away from the vaccine strain with many substitutions occurring in the antigenic sites (epitopes) of the HA. The overall rate of nucleotide substitution was almost twice as high as that seen in unvaccinated populations. Such a natural experiment is somewhat harder to come by in human populations. In the 2003-2004 human influenza epidemic in Finland, ?] suggested that the observed lack of diversity among the strains circulating during the fall of 2003 resulted from the strains’ antigenic novelty and low host immune pressure. As long as sampling occurs regularly and frequently, then this observation — that antigenic novelty is associated with low antigenic diversity — is consistent with the hypothesis that immunity drives antigenic drift. A similar association was seen during the 2003 influenza season in New Zealand (Table 1); the introduction of the antigenically novel A/Fujian/411/02-like strains to New Zealand in 2003 [? ?] was accompanied by low observed viral diversity during the 2003 epidemic season.

Because host immunity plays a key role in determining the rate of influenza evolution, we would like to be able to estimate the amount of population-wide immunity in host populations to give some indication of how much near-term antigenic evolution to expect. ?] used maximum-likelihood methods to fit a statistical epidemic model to 18 seasons of influenza data from France. Their results suggest that human populations may be between 63% and 95% immune at the onsets of annual epidemics. ?] estimated that, for the initial phase of the 1968 Hong Kong pandemic, the Hong Kong population was 39% immune. In general, herd immunity estimates for influenza are diffcult to find in the literature; for a historical review of theory and estimates of herd immunity, see ?].

The relationship between herd immunity and antigenic drift has a clear implication for devising vacci-nation strategies: that vaccinating too many people can have negative consequences. Because of influenza’s effect on morbidity and mortality during annual epidemics, it would be imprudent to vaccinate fewer people with the hope of reducing antigenic drift. However, in seasons when high numbers of vaccines are administered, or almost equivalently, during seasons when we suspect natural host immunity to be quite high against the circulating strains, we should expect the emergence of immune-escape variants since the evolutionary pressure favoring them is strong. The case can be made that monitoring efforts should be strengthened during these seasons.

Discussion

The data compiled for this study indicate that antigenic drift is in fact observable in some influenza seasons, though we do not yet know why some seasons exhibit a temporal-genetic correlation in influenza isolates while others do not. The implication for vaccination efforts, however, is clear. If antigenic drift is indeed occurring on the time scale of a single epidemic, monitoring for immune-escape variants ought to focus on the latter part of the epidemic with

continued monitoring late into the spring months. As some of the New Zealand data have shown, surveillance in the summertime can also be quite valuable. Understanding the short-term effects of antigenic drift on vaccination and those of vaccination on antigenic drift may yield insight into some of the mechanisms generating the observed pattern in long-term antigenic drift. Smith and colleagues [?] have shown that long-term change in antigenic properties is a punctuated process, even though the underlying genetic process is gradual. In the Smith study, antigenic change was quantified by defining a unit of antigenic distance as a twofold dilution in a haemugglutinin inhibition (HI) assay, and antigenic evolution was then easily visualized by plotting antigenic distances from H3N2 strains isolated between 1968 and 2003 to an ancestral H3N2 strain from 1968. Antigenic change occurs in what appear to be discrete jumps every few years and the strains appear to cluster into different antigenic types; nucleotide change during this time period is gradual.

A second consequence of the ?] study is that although we may be confident in our estimate of the rate at which nucleotide changes accumulate in influenza and have a rough estimate of the rate at which amino acid changes accumulate, we still cannot predict the rate at which antigenic properties change. In other words, there is no simple mapping from nucleotide/amino-acid distance to antigenic distance, even though it has been known for quite some time that the correlation between these two distances is positive [? ?]. In the analysis presented by ?], cluster transitions between antigenically distinct groups of strains can occur by a single amino acid change or by more than a dozen amino acid changes.

The reason that a simple mapping of genetic distance onto antigenic distance would be useful is that it would give us significant predictive capacity in knowing when immunity-evading strains emerge. With such a mapping in hand, it would suffice to sequence strains obtained from various influenza sentinel networks and compute their divergence from the currently employed vaccine strain. In practice, this is done (along with HI-assays), but it is far from perfect in estimating the amount of immune escape. Several studies have suggested proxies for quantifying or identifying immune escape and in most cases analyzed their predictive value. These proxies include identifying variants with more than four amino acid differences in at least two epitopic regions [?]; counting amino acid differences in the HA1 segment (329 amino acids) of the haemagglutinin [?]; counting amino acid differences in the five epitopes (comprising a total of 131 amino acids) of the HA [?]; counting amino acid changes among the 18 positively selected positions identified by ? ?]; computing the maximum amino acid divergence among the five epitopes [?]; and several others [?]. All of these methods have some predictive ability. For comparisons of the different approaches, see the more recent works of ?] and ?]. For problems and pitfalls of predictive methods, see ?].

Since the type of sequence analysis presented here may not always be possible in real time during the course of an epidemic (collecting samples and sequencing viral RNA takes time), we might like some simpler and more coarse indicators of rapid antigenic evolution during a single influenza season. Such indicators may be found by looking at epidemic sizes, lengths, or delays in onset. A robust statistical analysis determining the informativeness of such indicators has not yet been done, one of the reasons being a lack of long and accurate time series of influenza cases in most regions. In analyzing the predictive value of such indicators, we would need to keep in mind that a key population-genetic feature of influenza evolution is that immune escape occurs as a result of both (*i*) waiting for the accumulation of enough amino acid changes to alter the structure of the HA1 and (*ii*) waiting for a particular beneficial mutation (or combination of mutations) that has a significant effect on the structure of the HA1. Correlating epidemic lengths at different regional scales [? ?] with the amount of observed evolution or immune scape would reveal the importance of the first “waiting process” mentioned above, because the number of accumulated mutations is proportional to the length of the epidemic. Likewise, correlating epidemic sizes to long-term antigenic drift patterns could

help reveal whether antigenic drift is primarily driven by the stochastic arrival of beneficial mutations; the long-term dynamics and evolution of such a process have already been studied with a mathematical model [?]. Both analyses would have to be done in a region where seeding is believed to be largely endogenous.

The key ecological processes affecting influenza evolution that need to be more fully understood are the global geographic spread of influenza and the seeding of local and regional epidemics. The correlation methods presented here for analyzing epidemic seeding are simple and intuitive, but somewhat crude; a complete statistical analysis would include strains from other regions and a robust phylogenetic analysis. For local vaccination planning, if it can be determined that seeding is endogenous and if the late strains of one season appear to be phylogenetically close to the early strains of the following season, then local monitoring can be quite informative for local vaccination recommendations. If seeding is exogenous, no amount of local monitoring during an epidemic season can aid in determining the best vaccine strains for next season. The data presented in this study indicate that, in New Zealand, seeding is endogenous for some seasons and exogenous for others. ?] also obtain mixed results from Finnish influenza data

Finally, because current vaccine recommendations are made on a global scale rather than a local scale, it would be useful to determine the regions that play the largest role in affecting global antigenic drift. A complete analysis of global influenza evolution — using, for example, a detailed Ferguson-like model [?] — could indicate where most of the evolution occurs, where summertime bottlenecks play the largest role, and what determines exogenous/endogenous seeding in different regions. A better understanding of these processes will be crucial to improving global influenza monitoring and to understanding influenza dynamics and evolution on a global scale.

Acknowledgements

Thanks to E. C. Holmes for providing the New York sequences, to A. Lavenu for providing the French sequences, and to C. Viboud for pointing me to the New Zealand data. Thanks to E. C. Holmes, C. Viboud, D. L. Smith, M. W. Feldman, A. C. Campbell, and two anonymous reviewers for comments and suggestions. MFB is supported by NIH grants GM28016 and HG000205.

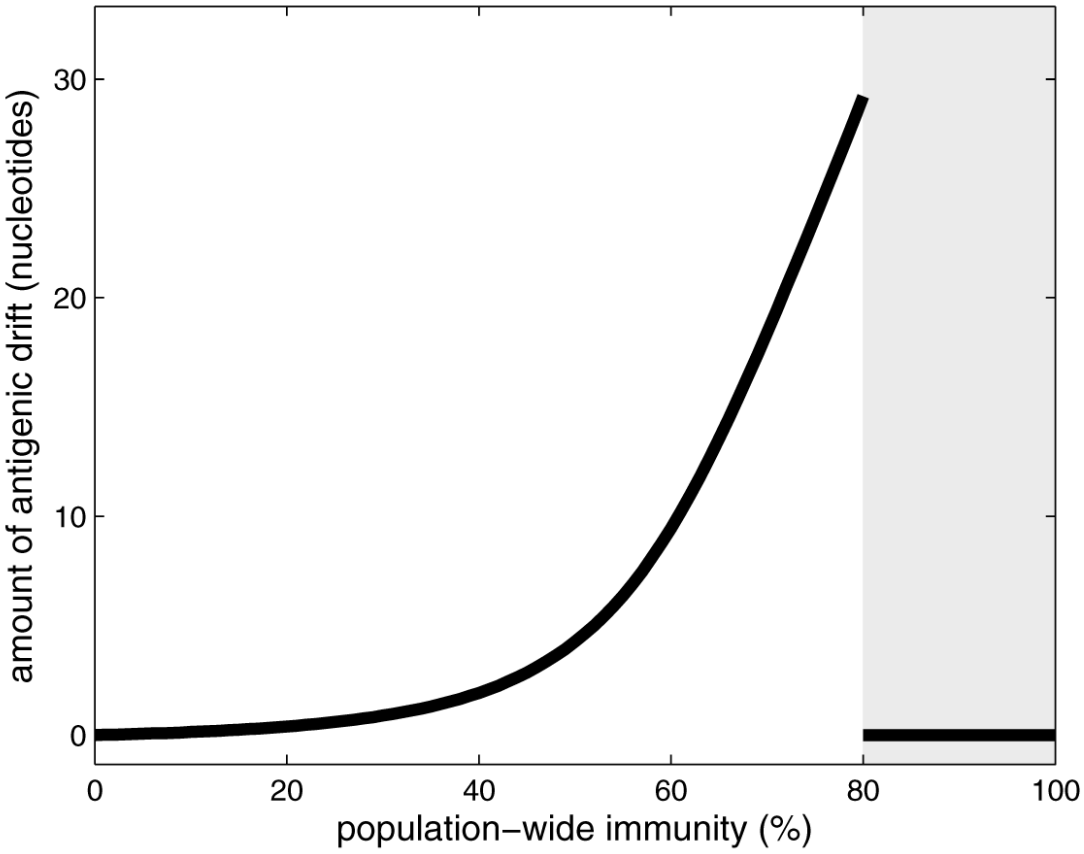

Figure 1.
Schematic diagram of antigenic drift as a function of host immunity, as predicted by the theory presented in ?]. Vertical axis represents the mean distance between strains isolated at the beginning of the epidemic and strains isolated at the end of the epidemic. The amount of observed antigenic drift increases as immunity in the host population increases and pressures the virus population to evolve. In this example, the gray-shaded area indicates that herd immunity is higher than 80%, meaning that, on average, each host is more than 80% immune; in this part of the graph, there is insuffcient potential to transmit the virus among hosts (basic reproduction ratio < 1) and there is no epidemic. The absence of an epidemic implies that there is no viral reproduction and no viral evolution. In reality, if the basic reproduction ratio were truly less than unity, a small amount of disease transmission could still occur, allowing for some viral evolution.

Table 1

Temporal-genetic correlation for 12 influenza epidemics. Note that we count the 2000 and 2001 seasons in New Zealand as having two epidemics each since strains of both subtypes were isolated in those years. 10,000 permutations were done for each Mantel *p*-value (100,000 for the first row), which determines at what level we can reject the null hypothesis that there is no correlation between the number of days separating two isolates and the number of nucleotide differences between two isolates. For five of the epidemics, a second row shows the statistics and *p*-values recomputed after removal of "temporal outliers"; these are strains isolated in the summertime, well before the onset of the usual influenza season. Temporal outliers are described in Table 2. Abbreviations: pwd = pairwise distance; sd = standard deviation; NY = New York (state); NZ = New Zealand; FR = France.

Location	Season	Subtype	Number isolates	Mean pwd (sd)	Max pwd	Mantel *p*-value	Span in days
NY	1999-2000	H3N2	37	5.98 (5.38)	22	10^{-4}	151 (10/3-3/1)
NY	2001-2002	H3N2	51	8.05 (3.57)	19	.162	121 (11/21- 3/21)
NY	2003-2004	H3N2	63	7.54 (3.50)	20	.158	102 (10/27- 2/5)
NZ	2000	H1N1	38	4.45 (5.95)	26	.081	255 (2/11-10/22)
			37	4.19 (5.96)	26	.343	126 (6/19-10/22)
NZ	2000	H3N2	38	14.72 (11.23)	32	.131	315 (1/20-11/29)
			36	14.44 (11.64)	32	.263	209 (5/5-11/29)
NZ	2001	H1N1	70	4.67 (7.18)	29	.011	221 (1/9-8/16)
			69	4.13 (6.50)	29	.116	121 (4/18-8/16)
NZ	2001	H3N2	18	19.12 (14.39)	37	.053	150 (4/17-9/13)
NZ	2002	H3N2	69	5.82 (5.63)	22	.003	227 (1/31-9/13)
			67	5.44 (5.53)	22	.028	156 (4/11-9/13)
NZ	2003	H3N2	76	1.78 (1.40)	9	.118	96 (5/19-8/22)
NZ	2004	H3N2	60	3.88 (4.28)	20	.062	273 (1/26-10/24)
			59	3.73 (4.27)	20	.103	97 (7/20-10/24)
NZ	2005	H3N2	46	12.39 (7.92)	25	.076	190 (3/4-9/9)
FR	1999-2000	H3N2	92	6.21 (6.10)	20	.83, ref [?]	41 (1/5-2/14)

Vaccine. Author manuscript; available in PMC 2009 July 18.

Table 2

Temporal outliers for New Zealand strains. Genetic outlier rank is computed by ranking the mean distance from one strain to the remaining strains in that season. For example, outlier #4 has a mean distance of 23.13nt to the other 69 strains of the 2001 H1N1 season. The strain A/Canterbury/119/2001 (isolated Aug 2; 5nt differences from outlier #4) has a mean distance of 26.03nt to the other 69 strains of the 2001 H1N1 season. The other 68 strains in that season have mean distances between 2.49nt and 23.07nt; therefore, outlier #4 ranks second out of 70 in being genetically the most distant to the other strains circulating during its season. Outlier #6 is genetically identical to A/Wellington/7/2002 isolated on Jun 11 2002.

Number	Season	Subtype	Name	Date of isolation	Genetic outlier rank
#1	2000	H1N1	A/Dunedin/2/2000	Feb 11 2000	3/38
#2	2000	H3N2	A/Canterbury/2/2000	Jan 20 2000	14/38
#3	2000	H3N2	A/Nelson Marlborough/1/2000	Feb 4 2000	15/38
#4	2001	H1N1	A/Canterbury/01/2001	Jan 19 2001	2/70
#5	2002	H3N2	A/Waikato/2/2002	Jan 31 2002	4/69
#6	2002	H3N2	A/Wellington/6/2002	Feb 10 2002	11/69
#7	2004	H3N2	A/Wellington/1/2004	Jan 26 2004	6/60

Table 3

Between-season distances for New Zealand strains. Table shows relationships among strains isolated in consecutive seasons; e.g. the mean distance between an H1N1 strain isolated in 2000 and an H1N1 strain isolated in 2001 is 22.01nt. The Mantel-like *p*-value indicates at what significance level we can reject the null hypothesis that given two strains from consecutive seasons there is no correlation between their temporal distance and their genetic distance. 10,000 permutations were done for each row. Abbreviations: pwd = pairwise distance; sd = standard deviation. ([a]): one outlier was isolated in the summer between the 2000 and 2001 seasons (A/Canterbury/01/2001); it has a mean distance of 10.97nt to the 2000 strains and 23.13nt to the 2001 strains. ([b]): two outliers were isolated in the summer between the 2001 and 2002 seasons (A/Waikato/2/2002 and A/Wellington/6/2002); they have a mean distance of 32.56nt to the 2001 strains and 12.02nt to the 2002 strains. ([c]): one outlier was isolated in the summer between the 2003 and 2004 seasons (A/Wellington/1/2004); it has a mean distance of 14.95nt to the 2003 strains and 8.24nt to the 2004 strains.

Seasons	Subtype	Outliers removed	Mean pwd (sd)	Min pwd	Max pwd	Mantel-like *p*-value
2000-2001	H1N1		22.01 (4.24)	7	30	.046
		#1, #4 ([a])	22.16 (4.07)	7	30	.408
2000-2001	H3N2		19.07 (11.35)	1	36	.056
		#2, #3	19.07 (11.61)	1	36	.058
2001-2002	H3N2		32.84 (8.71)	5	47	.006
		#5, #6 ([b])	32.85 (8.71)	5	47	.005
2002-2003	H3N2		57.53 (1.69)	54	64	.953
		#5, #6	57.51 (1.69)	54	64	.991
2003-2004	H3N2		20.94 (1.68)	14	29	.022
		#7 ([c])	21.04 (1.49)	19	29	.984
2004-2005	H3N2		11.86 (3.84)	2	27	.003
		#7	11.81 (3.84)	2	27	.002