JVI Accepted Manuscript Posted Online 26 August 2015
J. Virol. doi:10.1128/JVI.01571-15




# Positive selection in CD8+ T-cell epitopes of influenza nucleoprotein revealed by a comparative analysis of human and swine viral lineages

**Running title: Positive selection on influenza CD8+ T-cell epitopes**

Heather M. Machkovech[1,2,3], Trevor Bedford[4], Marc A. Suchard[5], and Jesse D. Bloom[1,3,#]

[1]Department of Genome Sciences, University of Washington, Seattle, WA

[2]Medical Scientist Training Program, University of Washington, Seattle, WA

[3]Division of Basic Sciences and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA

[4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA

[5]Departments of Biomathematics, Biostatistics and Human Genetics, UCLA, Los Angeles, CA

[#]To whom correspondence may be addressed: jbloom@fredhutch.org

Manuscript information: 227 words in the abstract; 9043 words in the text.

## ABSTRACT

Numerous experimental studies have demonstrated that CD8+ T-cells contribute to immunity against influenza by limiting viral replication. It is therefore surprising that rigorous statistical tests have failed to find evidence of positive selection in the epitopes targeted by CD8+ T-cells. Here we use a novel computational approach to test for selection in CD8+ T-cell epitopes. We define all epitopes in the nucleoprotein (NP) and matrix protein (M1) with experimentally identified human CD8+ T-cell responses, and then compare the evolution of these epitopes in parallel lineages of human and swine influenza that have been diverging since roughly 1918. We find a significant enrichment of substitutions that alter human CD8+ T-cell epitopes in the NP of human versus swine influenza, consistent with the idea that these epitopes are under positive selection. Furthermore, we show that epitope-altering substitutions to human influenza NP are enriched on the trunk versus the branches of the phylogenetic tree, indicating that viruses that acquire these mutations have a selective advantage. However, even in human influenza NP, sites in T-cell epitopes evolve more slowly than non-epitope sites, presumably because these epitopes are under higher inherent functional constraint. Overall, our work demonstrates that there is clear selection from CD8+ T-cells in human influenza NP, and illustrates how comparative analyses of viral lineages from different hosts can identify positive selection that is otherwise obscured by strong functional constraint.

46

47 **IMPORTANCE**

48 There is a strong interest in correlates of anti-influenza immunity that are protective
49 against diverse viral strains. CD8+ T-cells provide such broad immunity, since they target
50 conserved viral proteins. An important question is whether T-cell immunity is sufficiently
51 strong to drive influenza evolution. Although many studies have shown that T-cells limit
52 viral replication in animal models and are associated with decreased symptoms in
53 humans, no studies have proven with statistical significance that influenza evolves under
54 positive selection to escape T-cells. Here we use comparisons of human and swine
55 influenza to rigorously demonstrate that human influenza evolves under pressure to fix
56 mutations in nucleoprotein that promote escape from T-cells. We further show that
57 viruses with these mutations have a selective advantage since they are preferentially
58 located on the “trunk” of the phylogenetic tree. Overall, our results show that CD8+ T-
59 cells targeting nucleoprotein play an important role in shaping influenza evolution.

## INTRODUCTION

Both arms of the adaptive immune system help control influenza replication: antibodies neutralize virus [1] and direct the clearance of infected cells [2], while CD8+ T-cells kill infected cells that display viral peptides on their MHC class I molecules [3], [4]. While antibodies against the viral surface protein hemagglutinin (HA) provide the most potent protection when they are well matched to the viral strain [5]–[7], T-cells offer broader protection against diverse strains since they tend to recognize epitopes in more conserved internal viral proteins such as nucleoprotein (NP) and matrix protein (M1) [3], [4], [8], [9].

Studies in both mice [10]–[14] and humans [9], [15], [16] have shown that pre-existing influenza-specific CD8+ T-cells reduce the severity of disease and enhance viral clearance. For instance, pre-existing virus-specific CD8+ T-cells correlated with decreased symptoms in humans infected during the 2009 H1N1 pandemic [15]. Similarly, T-cells specific for NP were associated with decreased incidence of symptomatic infection over a multi-year study of a large human cohort [9], and CD8 T-cell responses correlate with recovery from severe H7N9 infection [16]. Therefore, experimental and epidemiological work demonstrates that CD8+ T-cells contribute to immunity against influenza.

Because humans are repeatedly infected with influenza over their lifetimes, one might expect viruses to be under evolutionary pressure to accumulate substitutions in epitopes targeted by immune memory. Indeed, there are numerous examples of the fixation of antibody-escape mutations in HA [17], [18], consistent with the notion that this protein evolves under strong selection from antibodies. Several studies have also described influenza mutations that escape recognition by CD8+ T-cells [19]. In a mouse



Journal of Virology

study, viral mutations arose that conferred T-cell escape in RAG-1 deficient mice expressing an influenza NP specific TCR [20]. Rimmelzwaan and coworkers identified the fixation of mutations in the NP of human H3N2 that mediated escape from CD8+ T-cells by altering the epitope recognized by the T-cell receptor [21]–[23] or abrogating binding of the epitope to MHC class I [24]. Valkenburg *et al* described the emergence of CD8+ T-cell escape mutations in a persistently influenza-infected infant [25]. These elegant studies demonstrate that influenza accumulates substitutions that escape CD8+ T-cells as well as antibody-mediated immunity.

But these studies do not prove that positive selection for CD8+ T-cell escape is an important driving force in influenza's evolution, since many sites in the viral genome will fix substitutions given enough time [26]–[28]. To rigorously establish the presence of positive selection, the field of molecular evolution has developed statistical tests to discern whether a subset of sites is evolving faster than expected. Most of these tests compute nonsynonymous and synonymous distances (referred to as dN and dS, respectively), and then test for sites with statistical evidence that the accumulation of nonsynonymous substitutions exceeds that of synonymous substitutions (dN/dS > 1) [29], [30]. These tests consistently find overwhelming evidence for positive selection in the antigenic sites of influenza hemagglutinin [31]–[33], but little evidence for positive selection in CD8+ T-cell epitopes [33]. One study did report that CD8+ T-cell epitopes in NP have a greater dN/dS than other sites [34]; however this study only made a pairwise comparison of two sequences and included no tests for statistical significance. Below, we have used several state-of-the-art tests to verify that CD8+ T-cell epitopes have neither an elevated frequency of sites with dN/dS > 1 nor an elevated rate of nonsynonymous

substitutions. Therefore, by standard criteria, CD8+ T-cell epitopes are not under positive selection.

The results of these statistical tests for positive selection seem at odds with the extensive body of experimental work described above. We hypothesized that the discrepancy arises because known CD8+ T-cell epitopes are under strong functional constraint [34]–[37]. If epitopes are highly constrained, then even strong positive selection might fail to elevate the rate of nonsynonymous substitutions in epitopes above that at less constrained non-epitope sites. To address this possibility, we developed new statistical tests that take advantage of the fact that some lineages of human influenza are paralleled by lineages of swine influenza that are not under selection from human CD8+ T-cells. Using these tests, we show that CD8+ T-cell epitopes in NP evolve significantly faster in human influenza than in swine influenza. Furthermore, we show that substitutions in these epitopes are enriched on the trunk of the phylogenetic tree, indicating that viruses that acquire them have a selective advantage that promotes their evolutionary spread. Overall, our work provides clear statistical evidence that complements prior experimental studies showing that CD8+ T-cell epitopes are under selection in human influenza [22], and suggests that the failure of conventional tests to identify this selection is due to high levels of functional constraint in epitopes.

## METHODS AND MATERIALS

### Inference of phylogenetic trees and mutation counts.

M1 and NP protein-coding sequences were downloaded from the Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html) [38].

For human influenza, we assembled sequence sets by taking sequences for H1N1 (1918 to 1957), H2N2 (1957 to 1968), and H3N2 (1968 to 2013) – if there were less than three sequences per year then we retained them all; when there were more than three for a year then we randomly selected three to retain. For swine influenza, we similarly assembled sequence sets containing up to 3 sequences per subtype per year for H1N1 (1918 to 2013), H1N2 (1999 to 2013), and H3N2 (1998 to 2013). For swine influenza, the first available sequence is from 1933. We excluded sequences previously classified as misannotated [39] or that were strong outliers based on a molecular clock analysis using RAxML [40] and Path-O-Gen (http://tree.bio.ed.ac.uk/software/pathogen/). The sequence sets are in Supplemental Files 1-4.

There are gaps in sequence availability in earlier years (most prominently, there are no sequences from between 1918 and the early 1930s). Therefore, we have reduced power to identify substitutions in these early years. However, since our comparisons are between human and swine influenza, and since both lineages have similarly sparse sequences in these early years, these gaps seem unlikely to systematically bias our study, although they may reduce its power.

We translated the sequences and inferred separate human and swine influenza phylogenies for each protein using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) [41] with a strict molecular clock, a Jones-Taylor-Thornton (JTT) [42] model of substitution, and a constant population size demographic model. **Fig. 1** shows the

maximum clade credibility trees rendered with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The trunk of each tree (dark lines in **Fig. 1**) was defined by tracing from the most recent sequence back to the oldest sequence. We used a stochastic mapping technique [43]–[45] implemented via the "MarkovJumps" feature in BEAST to estimate the posterior mean number of substitutions at each site for each phylogenetic tree and along the trunk of each tree. The times to most recent common ancestor referred to in the **Fig. 1** legend were estimated by a BEAST analysis of the joint swine and human influenza lineages. The dates of fixation of the CD8+ T-cell escape substitutions characterized by Rimmelzwaan and colleagues [21]–[24] refer to estimates obtained from Figure 2 - Supplement 2 of [37]. Supplemental File 9 lists all substitutions that are present along the trunk of at least 90% of the trees sampled from the posterior for each viral protein and lineage, along the with posterior-mean estimate of the date at which the substitution fixed on the trunk.

**Identification of CD8+ T-cell epitope sites.**

We downloaded all epitopes with an experimentally identified CD8+ T-cell response (source organism *Influenza A virus* and host *Homo sapiens*) from the Immune Epitope Database (http://www.iedb.org) [46]. We identified unique epitopes as described in the RESULTS section using a previously described software package (https://github.com/jbloom/epitopefinder/) [47], and then determined the number of unique epitopes ($E_r$) to which each site $r$ contributes (see **Fig. 2**). The epitopes and the counts of epitopes per site are in Supplemental Files 5 to 8.

**Conventional dN/dS tests for positive selection.**

We used DataMonkey (http://www.datamonkey.org/) [48] to perform two types of dN/dS analyses, the hierarchical Bayes method FUBAR (Fast Unconstrained Bayesian AppRoximation) [30] and the maximum-likelihood method FEL (Fixed Effects Likelihood) [29]. The maximum clade credibility tree from BEAST was used as the input phylogeny for FUBAR and FEL, and a REV (general reversible model) codon substitution model was specified for FEL. For both methods, we calculated the percentage of sites for which the estimated dN/dS ratio was greater than one, and the percentage of sites for which there was strong statistical support for this ratio being greater than one (posterior probability > 0.95 for FUBAR; P-value < 0.05 for FEL).

**Statistics on substitutions at each site in human and swine influenza lineages.** The posterior mean estimate of the number of nonsynonymous substitutions $S_r$ at each site $r$ was extracted from the BEAST trees. These estimates were used to compute the average substitution rates across all epitope sites (sites that fell in at least one epitope) and across all non-epitope sites, both for the entire tree and for the trunk alone. We also defined a statistic, $F$, which represents the average number of epitopes changed per substitution. This statistic is defined as

$$F = \frac{\sum_r E_r \times S_r}{\sum_r S_r}$$

where $E_r$ is the number of unique epitopes to which site $r$ contributes.

We performed statistical tests of whether we could reject the null hypothesis that there was no difference between the $F$ statistics for human versus swine and for the trunk versus the tree. To do this, we calculated the ratio of these statistics for human versus

swine or trunk versus tree, and then created a null distribution by repeatedly recalculating the statistics after randomizing the epitope counts $E_r$ among sites. The P-values represent the fraction of time the randomized statistic is greater than the actual statistic in $10^4$ randomizations.

**Availability of data and computer code.**

Data and computer code are available at https://github.com/hmmachko/TcellEpitopeComparisons

## RESULTS

### Parallel human and swine influenza lineages can reveal selection by CD8+ T-cells.

Our goal is to determine whether epitopes targeted by human CD8+ T-cells are under selection in influenza viruses that circulate in human hosts. The two most highly expressed influenza proteins are NP and M1 [49], and epitopes in these proteins are major targets of CD8+ T-cells [3], [4], [8], [9]. The NP and M1 in contemporary human H3N2 influenza have circulated in humans since at least 1918 [50], [51]. For both genes, this unbroken lineage consists of H1N1 viruses from 1918 to 1957, H2N2 viruses from 1957 to 1968, and H3N2 viruses from 1968 to the present. The red lines in **Fig. 1** show phylogenetic trees of NP and M1 from this human influenza lineage.

This human influenza lineage is closely paralleled by a swine influenza lineage descended from the common ancestor of the virus that caused concurrent pandemics in humans and swine in 1918 [50], [52]. NP and M1 of this lineage have circulated exclusively in swine since 1918 [50], [52]. The blue lines in **Fig. 1** show phylogenetic trees of NP and M1 from this swine influenza lineage. The phylogenetic trees show that both human and swine influenza undergo substantial genetic evolution in NP and M1; however, this fact alone does not reveal what forces drive this evolution. Influenza genetic evolution can be driven by positive selection, but it can also be driven by stochastic forces such as genetic hitchhiking or drift [26]–[28].

The parallel lineages of human and swine influenza enable us to perform an internally controlled analysis of whether CD8+ T-cells represent an important selective force in driving influenza evolution, since human CD8+ T-cells target epitopes in human but not swine influenza. There are two reasons that we can be confident that swine influenza is not under selection from human CD8+ T-cells. First, the MHC class I

molecules that restrict CD8+ T-cell epitopes are highly variable among species; therefore epitopes displayed to human CD8+ T-cells will differ from those displayed to swine CD8+ T-cells [53], [54] (note that our approach does not require the human and swine epitopes to be completely non-overlapping; it simply assumes that the MHC alleles are sufficiently diverged that not all epitopes targeted by humans are also targeted by swine). Second, swine influenza is under weaker selection from immune memory than human influenza because pigs are infected only once or a few times during their short lives [55]–[59]. Therefore, swine influenza is probably under less pressure from CD8+ T-cells in general, and whatever pressure does exist will generally focus on different epitopes than those targeted by human T-cells.

**Experimentally identified human CD8+ T-cell epitopes.**

We aimed to identify CD8+ T-cell epitopes targeted by individuals in the human population. There are two plausible ways to do this: computationally predict peptides that bind to MHC class I, or collate epitopes that have been experimentally identified as eliciting responses from CD8+ T-cells isolated from humans. We chose to use experimentally identified epitopes since computational predictions are imperfect [60], and only a fraction of peptides that bind MHC class I are targets of cytolytic CD8+ T-cells [61], [62]. We extracted all influenza epitopes from the Immune Epitope Database [46] between 8 and 12 amino acids in length with an experimentally identified human CD8+ T-cell response. We retained all epitopes that aligned to at least one strain from our human and swine influenza lineages with no more than one amino-acid mismatch. We classified epitopes as redundant if they shared 8 or more amino acids and were in the same MHC class I group [63] (or supertype [64] if the group was not specified). We identified 133

unique epitopes in the seven proteins that did not reassort in the human influenza pandemics of 1957 or 1968 (NS1, NS2, PB2, PA, M1, M2, and NP). Of the 133 epitopes, 62 were in NP (47%) and 29 were in M1 (22%), consistent with reports that these two proteins are major targets of CD8+ T-cells [9]. **Fig. 2** shows the number of epitopes to which each site in NP and M1 contributes; individual sites are involved in anywhere between zero and nine epitopes.

These experimentally identified epitopes probably do not represent an exhaustive list of all sites targeted by human T-cells. In particular, some epitopes in historical strains may be overlooked since most studies use recent viral strains. However, since our analyses are internally controlled (we compare either human to swine influenza, or the trunk of the tree to side branches), missing some epitopes should not systematically bias our results.

Our approach identifies sites that contribute to epitopes in any of the influenza strains under consideration. Mutations to an epitope can mediate escape by abrogating peptide binding to MHC class I or by changing the sequence of the bound peptide such that it is no longer recognized by memory T-cells. Both types of escape have been experimentally demonstrated in the NP of human H3N2. An example of escape by abrogation of MHC-binding is the R384G mutation that fixed in 1993 [24]. Three examples of escaping T-cell recognition but not MHC-binding are the D421E/I425V mutations that fixed in 1979 [22], [23], the K103R mutation that fixed in 1980 [21], and the S259L mutation that fixed in 1990 [21]. Additionally, mutations outside an epitope can affect its processing [65], although we are unaware of documented examples of extra-epitopic escape mutations that have fixed in human influenza. Our analysis cannot distinguish among these types of escape, since most experiments identify epitopes

without characterizing how prior or subsequent mutations affect their processing, MHC binding, and recognition by T-cells.

We therefore classify sites according to the number of epitopes to which they contribute in *any* of our viral strains without attempting to determine whether the epitopes are present across all the viral strains. This approach is usually valid when mutations alter T-cell recognition without affecting processing or MHC binding, since epitopes that escape existing T-cells via such mutations will often soon be targeted by new T-cells [21]. But our approach is imperfect for mutations that abrogate binding to MHC and so eliminate the epitope from all subsequent strains. However, since the NP and M1 homologs are closely related (the maximal protein-sequence divergence among the sequences in **Fig. 1** is just 12%), even if a site fixes just a single mutation that eliminates an epitope, this would represent a substantial elevation in its rate of evolution over the timeframe of interest. Therefore, classifying sites by the number of epitopes as in **Fig. 2** should identify positions in NP and M1 that will exhibit an increased rate of substitution if there is T-cell selection.

**Conventional dN/dS tests fail to find positive selection in CD8+ T-cell epitopes.** One might hypothesize that immune selection from CD8+ T-cells would lead to a greater proportion of epitope than non-epitope sites with dN/dS > 1. We therefore used two state-of-the-art methods (a hierarchical Bayes approach [29] and a maximum-likelihood approach [30]) to identify sites in human and swine influenza with dN/dS > 1, and partitioned the sites based on whether or not they were involved in at least one experimentally identified epitope. As shown in **Fig. 3A,B**, the proportion of sites with dN/dS > 1 was actually lower for epitope than non-epitope sites in human influenza.

Additionally, we calculated the proportion of sites with strong statistical evidence of dN/dS >1 (**Fig. 3C,D)**. In no instance is there a greater proportion of epitope versus nonepitope sites with dN/dS >1. Our findings are consistent with previous work that has failed to find evidence for positively selected sites in NP or M1 [33]. So overall, state-of-the-art dN/dS tests fail to identify enhanced positive selection in T-cell epitopes in NP.

**Epitope sites do not evolve faster than non-epitope sites, although the rate is higher on the "trunk" of the human influenza NP lineage.**

We next estimated the number of nonsynonymous substitutions at each site in NP and M1 for both the full swine and influenza trees in **Fig. 1**, and for the "trunks" of these trees (dark lines in **Fig. 1**). The rationale for examining the trunk separately is that we expect beneficial substitutions to be enriched on the trunk since they will confer a selective advantage that favors the propagation of sequences that contain them. Consistent with this idea, studies of human H3N2 influenza HA have found that presumably beneficial substitutions that alter antigenicity are enriched on the trunk versus the entire tree [66], [67].

**Fig. 4** shows the ratio of the substitution rate at epitope versus non-epitope sites. In none of the cases (NP or M1, swine or human influenza, trunk or tree) is the ratio substantially greater than one when taken over the whole tree – so in no case are the epitope sites evolving faster than the non-epitope ones. This result helps explains why the dN/dS tests fail to find evidence for positive selection in the epitopes – for NP and M1 (unlike for HA), epitope sites simply don't evolve faster than their non-epitope counterparts.

However, further examination of **Fig. 4** reveals an interesting trend. Although there are no instances where epitope sites are evolving substantially faster than non-epitope ones, the ratio of epitope to non-epitope substitution rate is greater for the NP of human influenza than for the NP of swine influenza. This trend is particularly pronounced along the trunk of the tree, which is exactly where we would expect to see the largest increase in rate if epitope substitutions confer a selective advantage. To test if this trend was indicative of statistically significant CD8+ T-cell selection in NP, we undertook the more nuanced analysis in the next section.

**Substitutions alter more NP epitopes in human than swine influenza, especially on the trunk.**

The foregoing analyses simply subdivided sites as epitope or non-epitope based on whether they fall in at least one experimentally identified epitope. But as **Fig. 2** shows, some sites are involved in far more unique epitopes than others. To account for this, we defined a new statistic (which we denote as *F*) that gives the average number of unique epitopes that are altered per substitution. **Fig. 5A** shows this *F* statistic for the trunk and entire tree for NP and M1 for both human and swine influenza. There appears to be little difference in this *F* statistic for M1 between the trunk and the entire tree, and between human and swine influenza. But for NP, *F* is greater for human influenza than swine influenza, with the increase much larger for the trunk than for the entire tree. This result indicates that the average substitution alters more NP epitopes in human influenza than swine influenza, and that this trend is especially pronounced on the trunk of the tree – exactly as we would expect under selection for epitope-altering substitutions.

To test if the trend is statistically significant, we computed the ratio of *F* for human versus swine influenza, and then calculated a null distribution by repeatedly randomizing the epitopes among sites. For M1, the actual ratio falls near the center of the null distribution (**Fig. 5B**), confirming that there is no enhancement of the rate of fixation of epitope-altering substitutions in the M1 of human influenza. But for NP, the actual ratio falls near the top the null distribution for both the trunk and the tree (**Fig. 5B**). In particular, for the trunk there is strong statistical support (P = 0.0002) for rejecting the null hypothesis that epitope-altering substitutions are equally likely to fix in human and swine influenza. This result indicates that there is selection for substitutions that alter CD8+ T-cell epitopes in the NP of human influenza.

To corroborate this statistical finding of an enhanced rate of epitope-altering substitutions along the trunk of human versus swine influenza, we undertook a further sub-analysis of the small subset of epitopes for which specific T-cell escape mutations have been identified. The vast majority of epitopes that comprise the analysis in **Fig. 5** were identified by relatively high-throughput studies that characterize peptides eliciting T-cell responses without identifying escape mutations. However, a series of meticulous studies by Rimmelzwaan and coworkers [21], [22], [24] have identified the sites of specific escape mutations for a small number of epitopes. **Table 1** shows that substitutions are fixed at all of these sites along the trunk of human influenza, but that swine influenza has not fixed substitutions at any of these sites. This finding lends further support to the idea that human T-cells exert positive selection on human but not swine influenza.

We next tested whether selection for epitope-altering substitutions in human influenza NP was stronger on the trunk than the rest of the tree, as would be expected if

such substitutions confer a selective advantage. We computed the ratio of the *F* statistic for the trunk versus the entire tree, and generated null distributions for this statistic by randomizing the epitopes among sites (**Fig. 6**). For M1, the actual ratios are near the center of the null distributions. But for NP, the actual ratio is near the top of the null distribution (P = 0.003) for human influenza, and near the bottom for swine influenza (P = 0.01). This result demonstrates that epitope-altering substitutions to NP are significantly enriched on the trunk of the human influenza lineage, suggesting viruses that fix these mutations are more fit than other strains. The depletion of epitope-altering substitutions on the swine lineage can also be given a clear explanation: If the known CD8+ T-cell epitopes in NP are in functionally constrained regions of the protein (as has been suggested by a variety of experimental studies [34]–[37]), then substitutions in these epitopes will often be deleterious in swine influenza lineages that experience no human CD8+ T-cell selection, and so will be relatively depleted on the trunk. So overall, the results in this section provide strong statistical evidence of positive selection in the CD8+ T-cell epitopes of human influenza NP.

## DISCUSSION

We describe the first rigorous statistical evidence that CD8+ T-cell epitopes are under positive selection in human influenza. Our work adds to a growing body of evidence suggesting an important role for T-cell immunity in shaping influenza evolution. Prior studies have shown that T-cells help protect against human influenza [9], [15] and have detailed specific instances of T-cell escape [21]–[25]. Our work shows that T-cell selection increases the rate at which mutations fix in epitopes of NP, and indicates that viruses with these substitutions have a selective advantage that makes them more likely to fall along the trunk of the phylogenetic tree.

Our results also explain why conventional dN/dS tests fail to detect positive selection in CD8+ T-cell epitopes. Known human CD8+ T-cell epitopes tend to be under high functional constraint [34]–[37]. It is unclear whether this is because T-cells inherently target conserved epitopes, because repeated infections preferentially boost T-cells that target conserved epitopes, or because there is a bias towards experimentally identifying conserved epitopes. But in any case, the fact that known epitopes are under high constraint means that even fairly strong positive selection may not enhance the nonsynonymous substitution rate to a level detectable by conventional dN/dS tests. This contrasts with antibody epitopes in HA, where the ability of dN/dS tests to detect antibody-mediated positive selection is probably augmented by the fact that antigenic sites are disproportionately tolerant of point mutations [68].

The novel approach that we developed ameliorates this problem by comparing the evolution of epitopes between human and swine influenza, or between the entire phylogenetic tree and only its trunk. These comparisons should better control for site-to-site variation in functional constraint, since the comparisons are always between

homologous sites that should be subject to similar functional constraints. Admittedly, there may also be other differences in functional constraints between human and swine influenza beyond T-cells, but unless these differential constraints are systematically biased towards occurring at T-cell epitopes they should not alter the fundamental validity of our approach. By making comparisons in this way, we demonstrated clear positive selection in CD8+ T-cell epitopes in NP, both in human versus swine influenza and in the trunk versus the entire phylogenetic tree.

One interesting aspect of our study is that we found positive selection in NP but not M1. This finding is consistent with a recent large-scale study that found that NP was the only protein for which the presence of pre-existing memory T-cells correlated with decreased symptomatic infections [9]. However, our study does not preclude an important role for T-cells against M1, which contains an immunodominant HLA-A2 epitope spanning residues 58-66 [69], [70]. One study has argued that T-cells targeting this epitope are ineffectual [70], although this interpretation is disputed [71], [72]. But experiments have also shown that this epitope is under high constraint [34]. If an epitope is completely intolerant of mutations, it will of course be unable to accumulate substitutions regardless of the strength of selection. It remains unclear if our failure to detect positive selection in M1 reflects a lack of effective immunity targeting this protein or strong constraints that simply prevent the fixation of escape mutations.

It is well established that antibodies are strong drivers of repeated selective sweeps in the evolution of human influenza [66], [73]. The fact that we can detect positive selection by CD8+ T-cells even in the presence of these antibody-driven selective sweeps demonstrates the importance of T-cell immunity in driving viral evolution. The

existence of such selection is consistent with modeling studies showing that T-cell immunity that reduces the infectious period can strongly favor viral escape [74].

There is considerable interest in developing vaccines that elicit stronger T-cell immunity to better protect against diverse influenza strains [3]. Our demonstration of the evolutionary importance of T-cell selection suggests that this interest is well founded. In addition, our results suggest that attempts to forecast the seasonal evolution of influenza [75], [76] could benefit from examining changes in T-cell as well as antibody epitopes.

## ACKNOWLEDGMENTS

The research reported here was supported by the NIGMS of the National Institutes of Health under grants R01GM102198 (to J.D.B) and MIDAS U01GM110721 (to T.B. and J.D.B). H.M.M. was supported in part by training grant T32AI083203 from the NIAID of the National Institutes of Health.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES


[1] M. Knossow, M. Gaudier, A. Douglas, B. Barrère, T. Bizebard, C. Barbey, B. Gigant, and J. J. Skehel, "Mechanism of neutralization of influenza virus infectivity by antibodies.," *Virology*, vol. 302, no. 2, pp. 294–298, 2002.

[2] S. Jegaskanda, E. R. Job, M. Kramski, K. Laurie, G. Isitman, R. de Rose, W. R. Winnall, I. Stratov, A. G. Brooks, P. C. Reading, and S. J. Kent, "Cross-reactive influenza-specific antibody-dependent cellular cytotoxicity antibodies in the absence of neutralizing antibodies.," *J. Immunol.*, vol. 190, no. 4, pp. 1837–48, 2013.

[3] S. A. Valkenburg, J. A. Rutigliano, A. H. Ellebedy, P. C. Doherty, P. G. Thomas, and K. Kedzierska, "Immunity to seasonal and pandemic influenza A viruses," *Microbes and Infection*, vol. 13, no. 5. pp. 489–501, 2011.

[4] G. F. Rimmelzwaan, J. H. C. M. Kreijtz, R. Bodewes, R. A. M. Fouchier, and A. D. M. E. Osterhaus, "Influenza virus CTL epitopes, remarkably conserved and remarkably variable," *Vaccine*, vol. 27, no. 45, pp. 6363–6365, 2009.

[5] J. L. Virelizier, "Host defenses against influenza virus: the role of anti-hemagglutinin antibody.," *J. Immunol.*, vol. 115, no. 2, pp. 434–439, 1975.

[6] M. L. Clements, R. F. Betts, E. L. Tierney, and B. R. Murphy, "Serum and nasal wash antibodies associated with resistance to experimental challenge with influenza A wild-type virus," *J. Clin. Microbiol.*, vol. 24, no. 1, pp. 157–160, 1986.

[7] R. B. Couch and J. A. Kasel, "Immunity to influenza in man.," *Annu. Rev. Microbiol.*, vol. 37, pp. 529–549, 1983.

[8] J. W. Yewdell, J. R. Bennink, G. L. Smith, and B. Moss, "Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 6, pp. 1785–1789, 1985.

[9] A. C. Hayward, L. Wang, N. Goonetilleke, E. B. Fragaszy, A. Bermingham, C. Andrew, D. Oliver, E. R. Millett, I. Nazareth, J. S. Nguyen-Van-Tam, J. M. Watson, M. Zambon, Flu Watch Group, A. M. Johnson, and A. J. McMichael, "Natural T cell mediated protection against seasonal and pandemic influenza," *Am. J. Respir. Crit. Care Med.*, pp. 1–53, 2015.

[10] H. Guo, F. Santiago, K. Lambert, T. Takimoto, and D. J. Topham, "T cell-mediated protection against lethal 2009 pandemic H1N1 influenza virus infection in a mouse model.," *J. Virol.*, vol. 85, no. 1, pp. 448–455, 2011.

[11] B. J. Laidlaw, V. Decman, M. A. A. Ali, M. C. Abt, A. I. Wolf, L. A. Monticelli, K. Mozdzanowska, J. M. Angelosanto, D. Artis, J. Erikson, and E. J. Wherry, "Cooperativity Between CD8+ T Cells, Non-Neutralizing Antibodies, and Alveolar Macrophages Is Important for Heterosubtypic Influenza Virus Immunity," *PLoS Pathog.*, vol. 9, no. 3, 2013.

[12] J. H. C. M. Kreijtz, R. Bodewes, G. van Amerongen, T. Kuiken, R. A. M. Fouchier, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan, "Primary influenza A virus infection induces cross-

protective immunity against a lethal infection with a heterosubtypic virus strain in mice," *Vaccine*, vol. 25, no. 4, pp. 612–620, 2007.

[13] P. M. Taylor and B. A. Askonas, "Influenza nucleoprotein-specific cytotoxic T-cell clones are protective in vivo.," *Immunology*, vol. 58, no. 3, pp. 417–20, Jul. 1986.

[14] S. L. Epstein, C. Y. Lo, J. A. Misplon, and J. R. Bennink, "Mechanism of protective immunity against influenza virus infection in mice without antibodies.," *J. Immunol.*, vol. 160, no. 1, pp. 322–7, Jan. 1998.

[15] S. Sridhar, S. Begom, A. Bermingham, K. Hoschler, W. Adamson, W. Carman, T. Bean, W. Barclay, J. J. Deeks, and A. Lalvani, "Cellular immune correlates of protection against symptomatic pandemic influenza," *Nat. Med.*, vol. 19, no. 10, pp. 1305–12, 2013.

[16] Z. Wang, Y. Wan, C. Qiu, S. Quiñones-Parra, Z. Zhu, L. Loh, D. Tian, Y. Ren, Y. Hu, X. Zhang, P. G. Thomas, M. Inouye, P. C. Doherty, K. Kedzierska, and J. Xu, "Recovery from severe H7N9 disease is associated with diverse response mechanisms dominated by CD8+ T cells," *Nat. Commun.*, vol. 6, no. May, p. 6833, 2015.

[17] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, "Mapping the antigenic and genetic evolution of influenza virus.," *Science*, vol. 305, no. 5682, pp. 371–376, 2004.

[18] G. W. Both, M. J. Sleigh, N. J. Cox, and A. P. Kendal, "Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites.," *J. Virol.*, vol. 48, no. 1, pp. 52–60, 1983.

[19] G. F. Rimmelzwaan, A. C. Boon, J. T. Voeten, E. G. Berkhoff, R. A. Fouchier, and A. D. Osterhaus, "Sequence variation in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes," *Virus Res*, vol. 103, no. 1–2, pp. 97–100, 2004.

[20] G. Price, R. Ou, H. Jiang, L. Huang, and D. Moskophidis, "Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo.," *J. Exp. Med.*, vol. 191, no. 11, pp. 1853–1867, 2000.

[21] E. G. M. Berkhoff, M. M. Geelhoed-Mieras, R. A. M. Fouchier, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan, "Assessment of the extent of variation in influenza A virus cytotoxic T-lymphocyte epitopes by using virus-specific CD8+ T-cell clones," *J. Gen. Virol.*, vol. 88, no. 2, pp. 530–535, 2007.

[22] A. C. M. Boon, G. de Mutsert, Y. M. F. Graus, R. A. M. Fouchier, K. Sintnicolaas, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan, "Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes.," *J. Virol.*, vol. 76, no. 5, pp. 2567–2572, 2002.

[23] A. C. Boon, G. de Mutsert, D. van Baarle, D. J. Smith, A. S. Lapedes, R. A. Fouchier, K. Sintnicolaas, A. D. Osterhaus, and G. F. Rimmelzwaan, "Recognition of homo- and heterosubtypic variants of influenza A viruses by human CD8+ T lymphocytes," *J Immunol*, vol. 172, no. 4, pp. 2453–2460, 2004.

[24] J. T. Voeten, T. M. Bestebroer, N. J. Nieuwkoop, R. A. Fouchier, A. D. Osterhaus, and G. F. Rimmelzwaan, “Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes.,” *J. Virol.*, vol. 74, no. 15, pp. 6800–6807, 2000.

[25] S. A. Valkenburg, S. Quiñones-Parra, S. Gras, N. Komadina, J. McVernon, Z. Wang, H. Halim, P. Iannello, C. Cole, K. Laurie, A. Kelso, J. Rossjohn, P. C. Doherty, S. J. Turner, and K. Kedzierska, “Acute emergence and reversion of influenza A virus quasispecies within CD8+ T cell antigenic peptides.,” *Nat. Commun.*, vol. 4, no. May, p. 2663, 2013.

[26] O. G. Pybus, A. Rambaut, R. Belshaw, R. P. Freckleton, A. J. Drummond, and E. C. Holmes, “Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution,” *Mol. Biol. Evol.*, vol. 24, no. 3, pp. 845–852, 2007.

[27] L. Simonsen, C. Viboud, B. T. Grenfell, J. Dushoff, L. Jennings, M. Smit, C. Macken, M. Hata, J. Gog, M. A. Miller, and E. C. Holmes, “The genesis and spread of reassortment human influenza A/H3N2 viruses conferring adamantane resistance,” *Mol. Biol. Evol.*, vol. 24, no. 8, pp. 1811–1820, 2007.

[28] M. I. Nelson, L. Simonsen, C. Viboud, M. A. Miller, J. Taylor, K. St. George, S. B. Griesemer, E. Ghedi, N. A. Sengamalay, D. J. Spiro, I. Volkov, B. T. Grenfell, D. J. Lipman, J. K. Taubenberger, and E. C. Holmes, “Stochastic processes are key determinants of short-term evolution in influenza A virus,” *PLoS Pathog.*, vol. 2, no. 12, pp. 1144–1151, 2006.

[29] S. L. Kosakovsky Pond and S. D. W. Frost, “Not so different after all: a comparison of methods for detecting amino acid sites under selection.,” *Mol. Biol. Evol.*, vol. 22, no. 5, pp. 1208–1222, 2005.

[30] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S. L. Kosakovsky Pond, and K. Scheffler, “FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection,” *Mol. Biol. Evol.*, vol. 30, no. 5, pp. 1196–1205, 2013.

[31] R. M. Bush, W. M. Fitch, C. A. Bender, and N. J. Cox, “Positive selection on the H3 hemagglutinin gene of human influenza virus A.,” *Mol. Biol. Evol.*, vol. 16, no. 11, pp. 1457–1465, 1999.

[32] A. C.-C. Shih, T.-C. Hsiao, M.-S. Ho, and W.-H. Li, “Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 15, pp. 6283–6288, 2007.

[33] S. L. Kosakovsky Pond, A. F. Y. Poon, A. J. Leigh Brown, and S. D. W. Frost, “A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus.,” *Mol. Biol. Evol.*, vol. 25, no. 9, pp. 1809–1824, 2008.

[34] E. G. M. Berkhoff, E. de Wit, M. M. Geelhoed-Mieras, A. C. M. Boon, J. Symons, R. A. M. Fouchier, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan, “Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes.,” *J. Virol.*, vol. 79, no. 17, pp. 11239–11246, 2005.

[35] E. Assarsson, H.-H. Bui, J. Sidney, Q. Zhang, J. Glenn, C. Oseroff, I. N. Mbawuike, J. Alexander, M. J. Newman, H. Grey, and A. Sette, "Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans.," *J. Virol.*, vol. 82, no. 24, pp. 12241–12251, 2008.

[36] G. F. Rimmelzwaan, E. G. M. Berkhoff, N. J. Nieuwkoop, R. a M. Fouchier, and a D. M. E. Osterhaus, "Functional compensation of a detrimental amino acid substitution in a cytotoxic-T-lymphocyte epitope of influenza a viruses by comutations.," *J. Virol.*, vol. 78, no. 16, pp. 8946–8949, 2004.

[37] L. I. Gong, M. A. Suchard, and J. D. Bloom, "Stability-mediated epistasis constrains the evolution of an influenza protein," *Elife*, vol. 2013, no. 2, 2013.

[38] Y. Bao, P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, and D. Lipman, "The Influenza Virus Resource at the National Center for Biotechnology Information," vol. 82, no. 2, pp. 596–601, 2008.

[39] M. Krasnitz, A. J. Levine, and R. Rabadan, "Anomalies in the influenza virus genome database: new biology or laboratory errors?," *J. Virol.*, vol. 82, no. 17, pp. 8947–8950, 2008.

[40] A. Stamatakis, "RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models," *Bioinformatics*, vol. 22, no. 21, pp. 2688–2690, 2006.

[41] A. J. Drummond, M. A. Suchard, D. Xie, and A. Rambaut, "Bayesian phylogenetics with BEAUti and the BEAST 1.7," *Mol. Biol. Evol.*, vol. 29, no. 8, pp. 1969–1973, 2012.

[42] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences.," *Comput. Appl. Biosci.*, vol. 8, no. 3, pp. 275–282, 1992.

[43] V. N. Minin and M. A. Suchard, "Counting labeled transitions in continuous-time Markov models of evolution," *J. Math. Biol.*, vol. 56, no. 3, pp. 391–412, 2008.

[44] J. D. O'Brien, V. N. Minin, and M. A. Suchard, "Learning to count: Robust estimates for labeled distances between molecular sequences," *Mol. Biol. Evol.*, vol. 26, no. 4, pp. 801–814, 2009.

[45] P. Lemey, V. N. Minin, F. Bielejec, S. L. K. Pond, and M. A. Suchard, "A counting renaissance: Combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection," *Bioinformatics*, vol. 28, no. 24, pp. 3248–3256, 2012.

[46] R. Vita, J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, J. R. Cantrell, D. K. Wheeler, J. L. Gabbard, D. Hix, A. Sette, and B. Peters, "The immune epitope database (IEDB) 3.0," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D405–D412, 2015.

[47] L. I. Gong and J. D. Bloom, "Epistatically Interacting Substitutions Are Enriched during Adaptive Protein Evolution," *PLoS Genet.*, vol. 10, no. 5, 2014.

[48] W. Delport, A. F. Y. Poon, S. D. W. Frost, and S. L. Kosakovsky Pond, "Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology," *Bioinformatics*, vol. 26, no. 19, pp. 2455–2457, 2010.

[49] S. Kummer, M. Flottmann, B. Schwanhausser, C. Sieben, M. Veit, M. Selbach, E. Klipp, and A. Herrmann, "Alteration of protein levels during influenza virus H1N1 infection in host cells: A proteomic survey of host and virus reveals differential dynamics," *PLoS One*, vol. 9, no. 4, 2014.

[50] M. Dos Reis, A. J. Hay, and R. a. Goldstein, "Using non-homogeneous models of nucleotide substitution to identify host shift events: Application to the origin of the 1918 'spanish' influenza pandemic virus," *J. Mol. Evol.*, vol. 69, no. 4, pp. 333–345, 2009.

[51] N. J. Cox and K. Subbarao, "Global epidemiology of influenza: past and present.," *Annu. Rev. Med.*, vol. 51, pp. 407–421, 2000.

[52] C. Brockwell-Staats, R. G. Webster, and R. J. Webby, "Diversity of influenza viruses in swine and the emergence of a novel human pandemic influenza A (H1N1)," *Influenza Other Respi. Viruses*, vol. 3, no. 5, pp. 207–213, 2009.

[53] C. Renard, E. Hart, H. Sehra, H. Beasley, P. Coggill, K. Howe, J. Harrow, J. Gilbert, S. Sims, J. Rogers, A. Ando, A. Shigenari, T. Shiina, H. Inoko, P. Chardon, and S. Beck, "The genomic sequence and analysis of the swine major histocompatibility complex," *Genomics*, vol. 88, no. 1, pp. 96–110, 2006.

[54] K. Lienert and P. Parham, "Evolution of MHC class I genes in higher primates.," *Immunol. Cell Biol.*, vol. 74, no. 4, pp. 349–356, 1996.

[55] M. G. Sheerar, B. C. Easterday, and V. S. Hinshaw, "Antigenic conservation of H1N1 swine influenza viruses," *J. Gen. Virol.*, vol. 70, no. 12, pp. 3297–3303, 1989.

[56] S. Noble, M. S. McGregor, D. E. Wentworth, and V. S. Hinshaw, "Antigenic and genetic conservation of the haemagglutinin in H1N1 swine influenza viruses," *J. Gen. Virol.*, vol. 74, no. 6, pp. 1197–1200, 1993.

[57] R. J. Garten, C. T. Davis, C. A. Russell, B. Shu, S. Lindstrom, A. Balish, W. M. Sessions, X. Xu, E. Skepner, V. Deyde, M. Okomo-Adhiambo, L. Gubareva, J. Barnes, C. B. Smith, S. L. Emery, M. J. Hillman, P. Rivailler, J. Smagala, M. de Graaf, D. F. Burke, R. A. M. Fouchier, C. Pappas, C. M. Alpuche-Aranda, H. López-Gatell, H. Olivera, I. López, C. A. Myers, D. Faix, P. J. Blair, C. Yu, K. M. Keene, P. D. Dotson, D. Boxrud, A. R. Sambol, S. H. Abid, K. St George, T. Bannerman, A. L. Moore, D. J. Stringer, P. Blevins, G. J. Demmler-Harrison, M. Ginsberg, P. Kriner, S. Waterman, S. Smole, H. F. Guevara, E. A. Belongia, P. A. Clark, S. T. Beatrice, R. Donis, J. Katz, L. Finelli, C. B. Bridges, M. Shaw, D. B. Jernigan, T. M. Uyeki, D. J. Smith, A. I. Klimov, and N. J. Cox, "Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans.," *Science*, vol. 325, no. 5937, pp. 197–201, 2009.

[58] C.-J. Wei, J. C. Boyington, K. Dai, K. V Houser, M. B. Pearce, W.-P. Kong, Z.-Y. Yang, T. M. Tumpey, and G. J. Nabel, "Cross-neutralization of 1918 and 2009 influenza viruses: role of glycans in viral evolution and vaccine design.," *Sci. Transl. Med.*, vol. 2, no. 24, p. 24ra21, 2010.

[59] A. L. Vincent, K. M. Lager, W. Ma, P. Lekcharoensuk, M. R. Gramer, C. Loiacono, and J. a. Richt, “Evaluation of hemagglutinin subtype 1 swine influenza viruses from the United States,” *Vet. Microbiol.*, vol. 118, no. 3–4, pp. 212–222, 2006.

[60] K. Roomp, I. Antes, and T. Lengauer, “Predicting MHC class I epitopes in large datasets.,” *BMC Bioinformatics*, vol. 11, p. 90, 2010.

[61] Y. Deng, J. W. Yewdell, L. C. Eisenlohr, and J. R. Bennink, “MHC affinity, peptide liberation, T cell repertoire, and immunodominance all contribute to the paucity of MHC class I-restricted peptides recognized by antiviral CTL.,” *J. Immunol.*, vol. 158, no. 4, pp. 1507–1515, 1997.

[62] J. W. Yewdell, “Confronting Complexity: Real-World Immunodominance in Antiviral CD8+ T Cell Responses,” *Immunity*, vol. 25, no. 4. pp. 533–543, 2006.

[63] “HLA nomenclature,” 2015. [Online]. Available: http://hla.alleles.org/nomenclature/naming.html.

[64] J. Sidney, B. Peters, N. Frahm, C. Brander, and A. Sette, “HLA class I supertypes: a revised and updated classification.,” *BMC Immunol.*, vol. 9, p. 1, 2008.

[65] T. M. Allen, M. Altfeld, X. G. Yu, K. M. O’Sullivan, M. Lichterfeld, S. Le Gall, M. John, B. R. Mothe, P. K. Lee, E. T. Kalife, D. E. Cohen, K. A. Freedberg, D. A. Strick, M. N. Johnston, A. Sette, E. S. Rosenberg, S. A. Mallal, P. J. R. Goulder, C. Brander, and B. D. Walker, “Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection.,” *J. Virol.*, vol. 78, no. 13, pp. 7069–7078, 2004.

[66] T. Bedford, M. A. Suchard, P. Lemey, G. Dudas, V. Gregory, A. J. Hay, J. W. McCauley, C. A. Russell, D. J. Smith, and A. Rambaut, “Integrating influenza antigenic dynamics with molecular evolution,” *Elife*, vol. 2014, no. 3, 2014.

[67] T. Bedford, A. Rambaut, and M. Pascual, “Canalization of the evolutionary trajectory of the human influenza virus,” *BMC Biology*, vol. 10, no. 1. p. 38, 2012.

[68] B. Thyagarajan and J. D. Bloom, “The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin,” *Elife*, vol. 2014, no. 3, pp. 1–26, 2014.

[69] A. C. M. Boon, G. de Mutsert, Y. M. F. Graus, R. A. M. Fouchier, K. Sintnicolaas, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan, “The magnitude and specificity of influenza A virus-specific cytotoxic T-lymphocyte responses in humans is related to HLA-A and -B phenotype.,” *J. Virol.*, vol. 76, no. 2, pp. 582–590, 2002.

[70] D. B. Keskin, B. B. Reinhold, G. L. Zhang, A. R. Ivanov, B. L. Karger, and E. L. Reinherz, “Physical detection of influenza A epitopes identifies a stealth subset on human lung epithelium evading natural CD8 immunity,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 7, pp. 2151–2156, 2015.

[71] C. E. van de Sandt and G. F. Rimmelzwaan, “Immunodominant responses to the influenza virus M1 58-66 epitope: Stealth or protection?,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 19, pp. E2417–E2417, 2015.

[72] D. B. Keskin, B. R. Reinhold, G. L. Zhang, A. R. Ivanov, B. L. Karger, and E. L. Reinherz, "Reply to van de Sandt and Rimmelzwaan: Matching epitope display with functional avidity," *Proc. Natl. Acad. Sci.*, vol. 112, no. 19, pp. E2418–E2418, 2015.

[73] A. Rambaut, O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C. Holmes, "The genomic and epidemiological dynamics of human influenza A virus.," *Nature*, vol. 453, no. 7195, pp. 615–619, 2008.

[74] J. R. Gog, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and B. T. Grenfell, "Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 19, pp. 11143–11147, 2003.

[75] M. Luksza and M. Lassig, "A predictive fitness model for influenza.," *Nature*, vol. 507, no. 7490, pp. 57–61, 2014.

[76] R. A. Neher, C. A. Russell, and B. I. Shraiman, "Predicting evolution from the shape of genealogical trees," *Elife*, vol. 3, p. e03568, 2014.

**Figure 1. Phylogenetic trees for human and swine influenza M1 and NP.**

Maximum clade credibility trees for NP and M1. The swine influenza lineage is in blue, and the human influenza lineage is in red. The dark blue and red lines represent the trunk. The dotted black lines indicate that human and swine influenza lineages share a recent common ancestor (estimated times to most recent common ancestor are 13 and 6 years for M1 and NP, respectively).

**Figure 2. The distribution of CD8 T-cell epitopes in M1 and NP**.

The number of unique experimentally identified CD8+ T-cell epitopes to which each site contributes for M1 and NP.

**Figure 3. Conventional dN/dS tests do not detect positive selection in T-cell epitopes.**

State-of-the-art methods for detecting positive selection fail to find any enrichment in sites with dN/dS > 1 at epitopes. Percentage of sites estimated to have dN/dS > 1 using (A) the hierarchical Bayesian approach implemented in FUBAR and (B) the maximum-likelihood approach implemented in FEL. (C) The percentage of sites for which FUBAR reports a posterior probability > 0.95 that dN/dS > 1. (D) The percentage of sites for which FEL reports dN/dS > 1 with $P < 0.05$.

**Figure 4. Epitope sites do not evolve faster than non-epitope sites.**

Shown are the ratios of the average nonsynonymous substitution rate of epitope sites versus nonepitope sites. When the substitution rate is computed across the entire tree, the epitope sites always have a lower substitution rate than the non-epitope sites.

However, along the trunk of the tree for NP from human influenza, the substitution rate in epitopes is slightly greater than the substitution rate at non-epitopes.

**Figure 5. The average substitution changes more NP epitopes in human influenza than swine influenza.**

The *F* statistic is the average number of epitopes altered per substitution. (A) The average substitution alters a similar number of epitopes in M1 in both human and swine influenza, and for both the entire tree and along the trunk alone. However, for NP the average substitution alters substantially more epitopes in human than swine influenza, particularly along the trunk of the tree. (B) The increased rate of epitope-altering substitutions to NP along the trunk of human versus swine influenza is statistically significant. The violin plots show the null distribution of the ratio of *F* for human to swine influenza, with the median shown by the red lines and the actual value shown by the red circles. The P-values (fraction of the null distribution ≥ the actual value) are: M1 trunk 0.49, M1 tree 0.65, NP trunk 0.0002, NP tree 0.098.

**Figure 6. Epitope-altering substitutions to human influenza NP are enriched on the trunk of the tree.**

The ratio of the average number of epitopes altered per substitution (*F*) for the trunk versus the entire tree. The violin plots show the actual values versus the null distributions as in **Fig. 5B**. There is no difference between the trunk and the entire tree for M1 (P-values are: M1 human influenza: 0.59 and M1 swine influenza: 0.57). But for NP, epitope-altering substitutions are significantly enriched on the trunk of the human influenza tree

738 relative to the rest of the tree (P = 0.003), and significantly depleted on the trunk of the
739 swine influenza lineage (P = 0.01 for depletion) relative to the rest of the tree.

740

741 **Table 1. Number of trunk substitutions for the small subset of NP epitopes where**
742 **specific escape substitutions have been experimentally validated.**

| site in NP | Experimental evidence | Number of trunk substitutions for human influenza | Number of trunk substitution for swine influenza |
|---|---|---|---|
| 103 | K103R escapes recognition by a HLA-B*1503-restricted T-cell [21] | 2 (K103R in 1981, R103K in 1997) | none |
| 259 | S259L escapes recognition by a HLA-B*4002-restricted T-cell [21] | 2 (L259S in 1972, S259L in 1990) | none |
| 384 | R384G abrogates MHC binding in HLA-B*2705 [24] | 1 (R384G in 1990) | none |
| 421 | D421E escapes recognition by a HLA-B*3501-restricted T-cell [22] | 1 (D421E in 1977) | none |
| 425 | I425V escapes recognition by a HLA-B*3501-restricted T-cell [22] | 2 (I425V in 1975, V425I in 1999) | none |

743