Deep mutational scanning of an H3 hemagglutinin can inform evolutionary forecasting of human H3N2 influenza virus

Juhye M. Lee[1,4,5,†] John Huddleston[2,6,†] Michael B. Doud[1,4,5]
Kathryn A. Hooper[1,6] Trevor Bedford,[2,3]
Jesse D. Bloom[1,3,4*]

[1]Basic Sciences Division, [2]Vaccine and Infectious Diseases Division, and [3]Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[4]Department of Genome Sciences, [5]Medical Scientist Training Program, and [6]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA

[†]These authors contributed equally

[*]Correspondence: jbloom@fredhutch.org

Abstract

Abstract text.

INTRODUCTION

RESULTS

Strategy for deep mutational scanning of an H3 hemagglutinin

Previously, we measured the effect of all possible single amino-acid mutations to a highly lab-adapted H1 hemagglutinin from the A/WSN/1933 (H1N1) strain.

H3 site-specific amino-acid preferences

Estimating mutational effects from an H3N2 phylogeny

Comparing H1 and H3 preferences

DISCUSSION

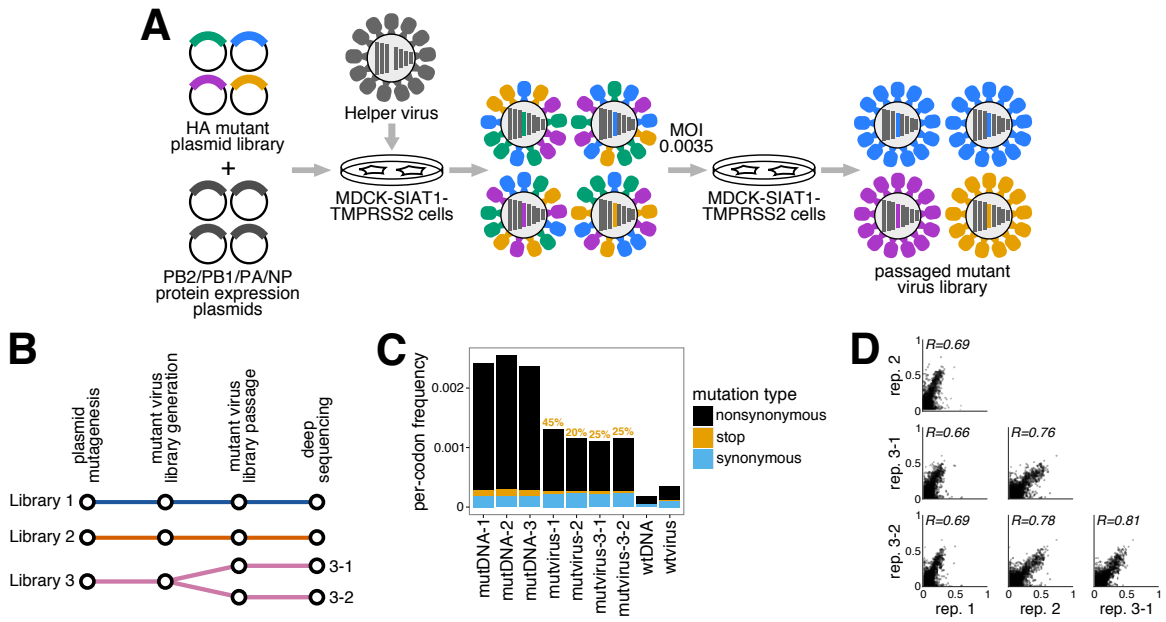

Figure 1: Overview of deep mutational scanning experiments of H3 hemagglutinin. (A) We used a helper virus approach previously described in Doud (2016) to generate the mutant virus libraries. We transfected MDCK-SIAT1-TMPRSS2 cells with the mutant plasmid library carrying all possible amino-acid mutations to the A/Perth/16/2009 (H3N2) HA, in addition to protein expression plasmids encoding the viral ribonucleoprotein complex. After transfection, we infected the cells with an HA-deficient helper virus carrying all of the WSN/1933 influenza genes. We then passage the initially generated pool of mutant viruses at low MOI to establish a genotype-phenotype linkage and select for functional HA variants. (B) All of the experiments were completed in biological triplicate, starting from independent preps of the wildtype HA genes to create the mutant plasmids. In addition, we passaged and deep sequenced library 3 in technical replicate, denoted 3-1 and 3-2, to gauge the amount of experimental noise within a single biological replicate. (C) Mutation frequencies of nonsynonymous, stop, and synonymous mutations for the mutant DNA, mutant virus, wildtype DNA, and wildtype virus samples. There is selection against nonsynonymous and stop codons in the mutant viruses. The percentages signify the frequency of stop codons remaining in the passaged mutant viruses relative to their starting frequency in the mutant plasmid libraries after correcting for the stop codon frequencies in the wildtype DNA and viruses. (D) The correlations and the Pearson correlation coefficient for the amino-acid preferences between each pair of replicates are shown. The biological replicates are fairly well-correlated, as are the technical replicates, indicating some degree of bottlenecking of variants during the viral passage which can contribute to experimental noise.

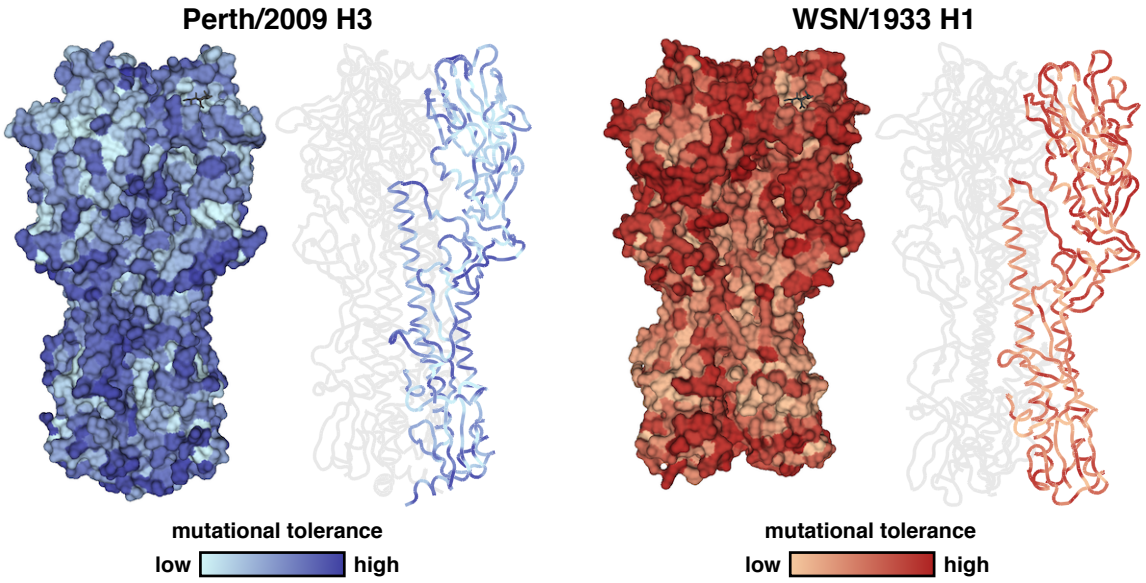

Figure 2: Different regions of HA exhibit varying degrees of mutational tolerance. Mutational tolerance as calculated by the Shannon entropy of a given site's amino-acid preferences are mapped onto the structure of the H3 trimer (PDB 4O5N, [*cite*]) and the H1 trimer (PDB 1RVX, [*cite*]), with both trimers in approximately the same orientation. The site entropies were calculated from the preferences measured in the Perth/2009 H3 (left panel) or the preferences measured in the WSN/1933 H1 (right panel). Lighter shades of blue or red signify low mutational tolerance, while darker shades of blue or red signify high mutational tolerance. For each HA, the structure on the left side colors the full HA trimer, while the structure on the right side colors only one of the monomers. The sialic acid receptor is shown as black sticks. The Perth/2009 H3 shows relatively high mutational tolerance in the stalk region compared to the head region. The head region of the WSN/1933 H1 is mutationally tolerant compared to the relatively intolerant stalk region.

Figure 3: The site-specific amino-acid preferences of H3 hemagglutinin. This logoplot shows the site-specific amino-acid preferences for the averaged replicates rescaled by the stringency parameter (Table 1) estimated by phydms. The height of each letter is proportional to its preference at that site, and the preferences for all sites are normalized to sum to 1. The sites are in H3 numbering. The top overlay bar shows the relative solvent accessibility. The bottom overlay bar is colored by the HA domain (sig. pep. = signal peptide, HA1 ecto. = HA1 ectodomain, HA2 ecto. = HA2 ectodomain, TM = transmembrane domain, cyto. tail. = cytoplasmic tail). The letters directly above each logo indicate the wildtype amino acid at that site.

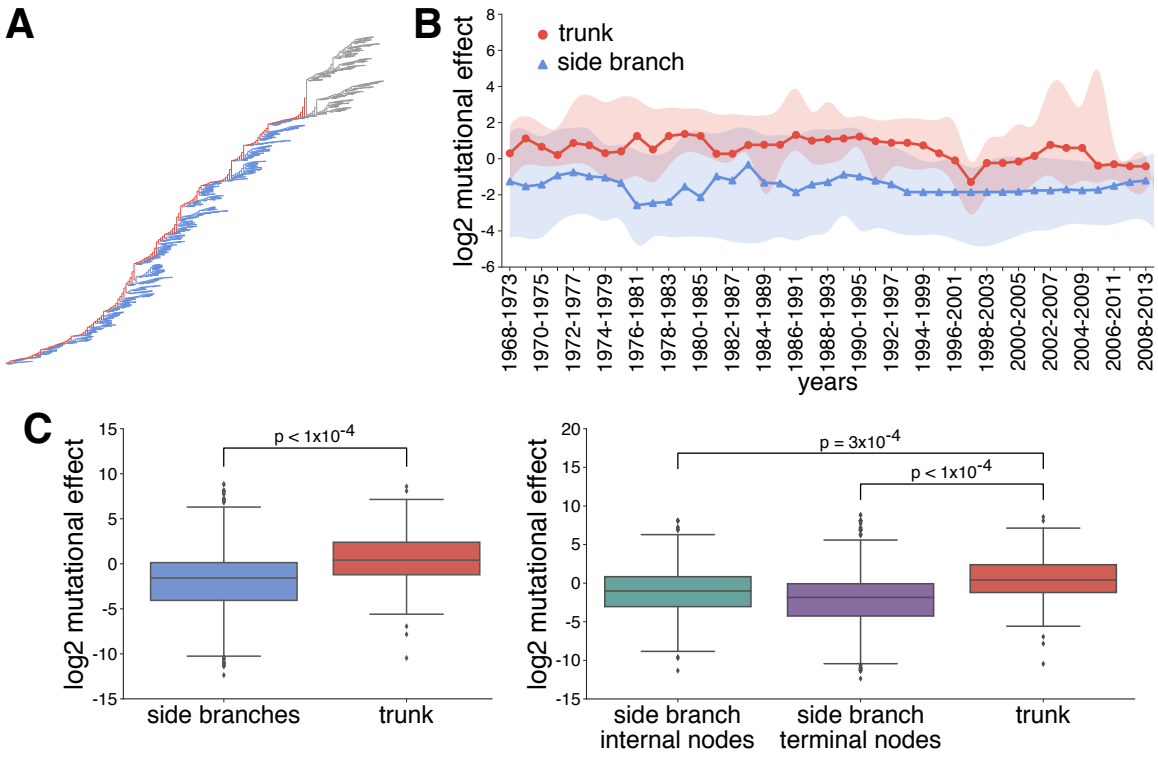

Figure 4: The trunk of a human H3N2 phylogeny has higher mutational effects than those of side branches. (A) Phylogenetic tree of human H3N2 influenza virus from 1968-present. [*We downloaded X sequences from the Influenza Virus Resource ?.... etc. inferred the tree, ancestral state reconstruction, visualized the tree. Mark Perth/2009 on the tree*] To parse out trunk mutations from side branch mutations, we first defined a set of recent nodes sampled on or after Jan. 1, 2017, and traced these nodes back to their most recent common ancestor. All branches ancestral to the MRCA of the recent nodes were defined as the trunk (shown in red), and all other branches were defined as side branches (shown in blue). (B) Using the Perth/2009 H3 preferences, we calculated the $\log_2$ mutational effect for all side branch and all trunk mutations (left panel), in addition to all mutations in internal nodes and terminal nodes on the side branches (right panel). To estimate significance, we performed 10,000 randomizations of the preferences to calculate the median difference in trunk vs side branch mutational effects, and counted how many of the randomizations exceeded the true difference in median trunk vs side branch mutational effects. The effects of trunk mutations are higher than side branch, internal and terminal node, mutations.

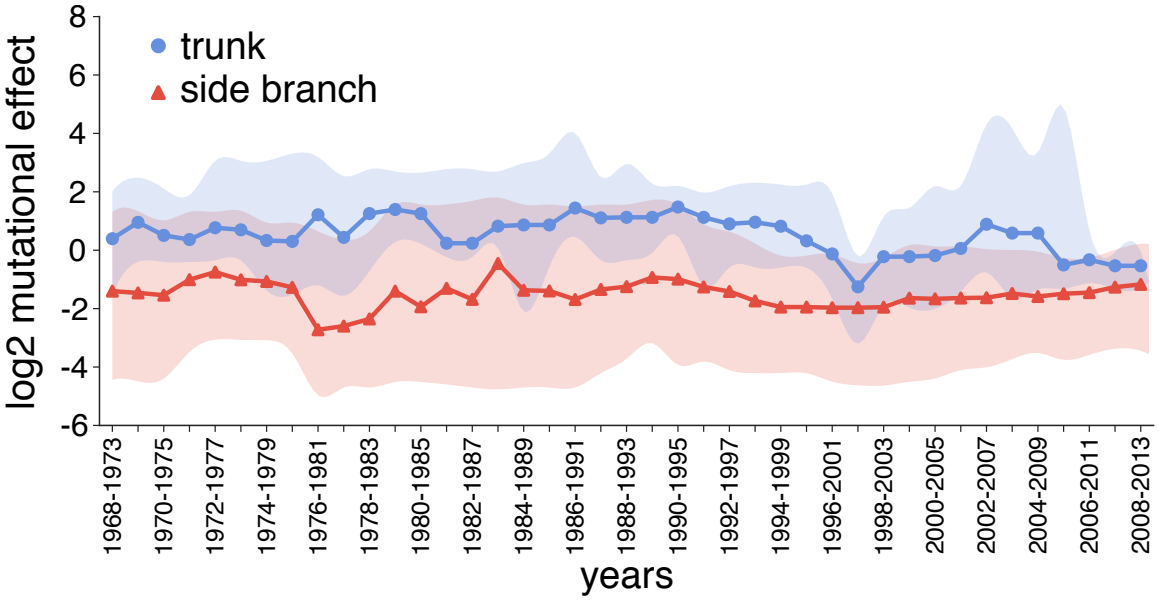

Figure 5: Trunk mutational effects are consistently higher than side branch mutational effects across time. We calculated the $\log_2$ mutational effect for trunk and side branch mutations in windows of 5 years for every year from 1968-2013. The median $\log_2$ mutational effect in a given window is shown as circles for trunk mutations and triangles for side branch mutations. The shaded region demarcates the interquartile range of trunk and side branch mutational effects. The median trunk mutational effects are consistently higher than the median side branch mutational effects for all windows.

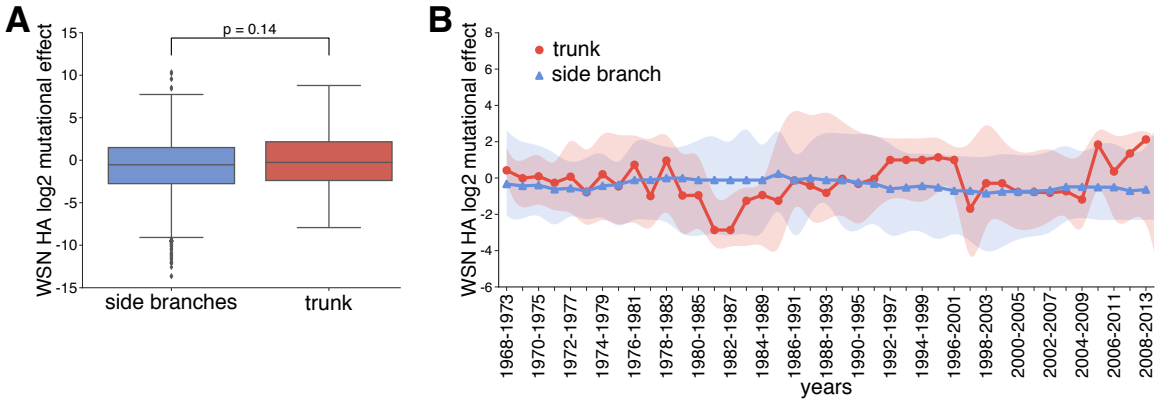

Figure 6: The WSN/1933 H1 preferences do not reveal differences in trunk vs side branch mutational effects (A) We calculated the $\log_2$ mutational effects of the same set of trunk and side branch mutations from the inferred H3N2 phylogeny in Figure **??** using the WSN/1933 H1 preferences. We again randomized the preferences for a total of 10,000 iterations and counted the number of randomizations that exceeded that true difference in median trunk vs side branch mutational effects to calculate a p-value. Using the WSN/1933 H1 preferences, there is not a significant difference in trunk vs side branch mutational effects. (B) We also performed the same sliding window analysis shown in Figure **??**, but using the WSN/1933 H1 preferences. There is not a distinct difference in trunk and side branch mutational effects.

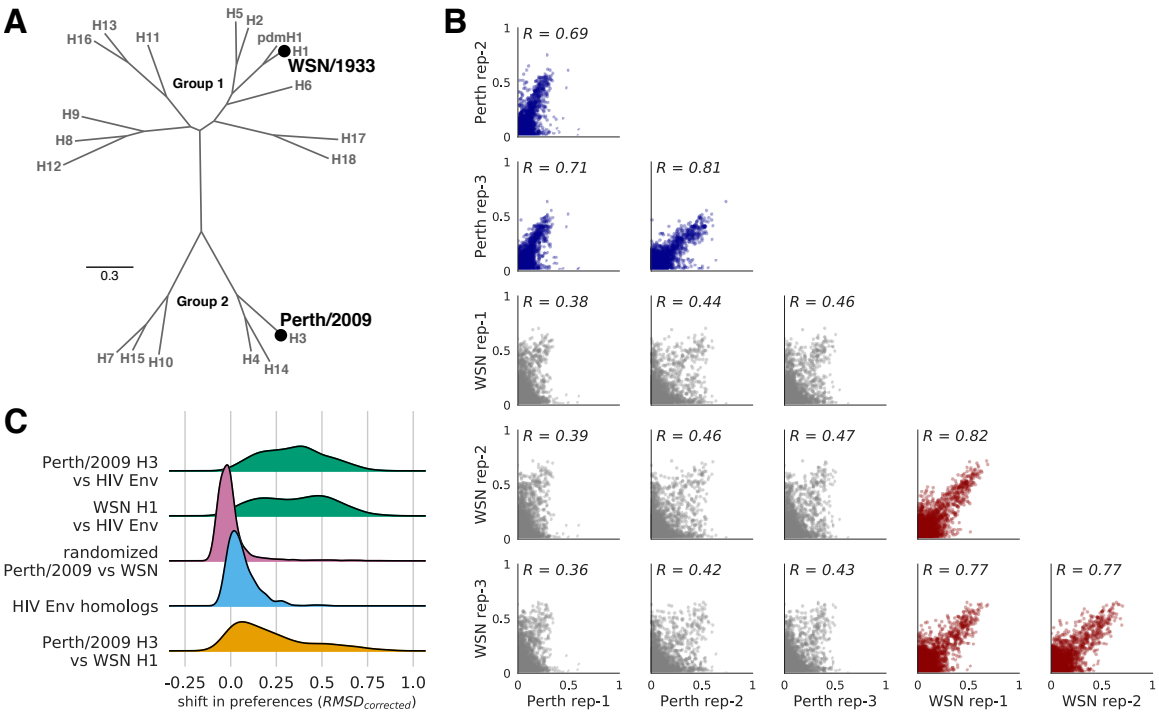

Figure 7: The HA homologs exhibit many large shifts in preference compared to shifts for other viral protein homologs (A) A phylogenetic tree of the HA subtypes, with the two HA's, WSN/1933 H1 and Perth/2009 H3, for which we have measured amino-acid preferences denoted on the tree. The WSN/1933 H1 and the Perth/2009 H3 share ~42% amino-acid identity. (B) The distribution of shifts in preference for various homolog pairs are shown.

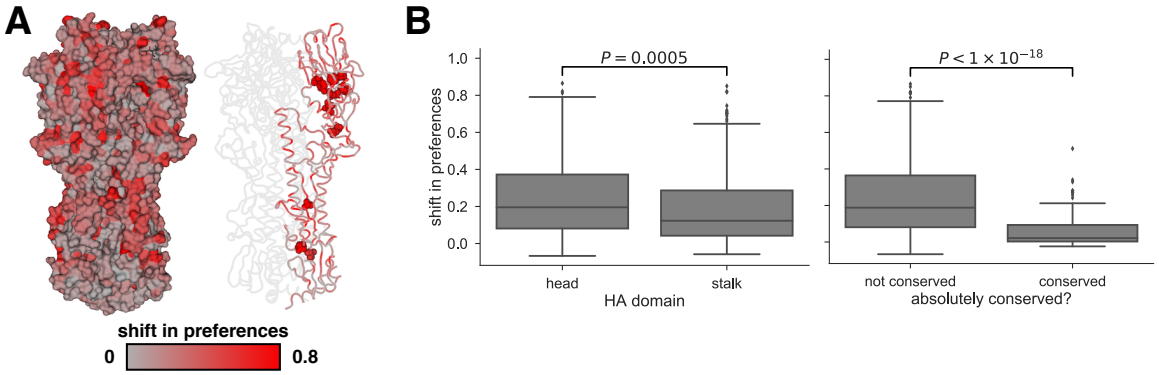

Figure 8: Shifts in preferences mapped onto the structure of HA (A) The preference shifts as calculated by $RMSD_{corrected}$ between the two HA homologs is mapped onto the structure of HA (PDB 4O5N, citation). The left structure shows the HA trimer, and the right structure colors one of the monomers. The sialic acid receptor is in black sticks. Gray indicates little shifts in preference, while red indicates large shifts in preference. The top ten most shifted sites are shown in spheres on the monomer. (B)

METHODS

HA numbering

Unless otherwise indicated, all sites are in H3 numbering.

Generation of HA codon mutant plasmid libraries

Generation and passaging of mutant viruses

Barcoded subamplicon sequencing

Analysis of deep sequencing data

Quantification of mutational effects and sequence preferences from an H3N2 phylogeny

Data availability and source code

Deep sequencing data are available from the Sequence Read Archive under BioSample accession [*add accession*].

ACKNOWLEDGMENTS

We thank Sarah Hilton, Hugh Haddox, Sidney Bell...the Fred Hutch Genomics Core... Funding...