

# Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans

YQ Shirleen Soh<sup>1,2</sup>, Louise H Moncla<sup>2,3</sup>, Rachel Eguia<sup>1</sup>, Trevor Bedford<sup>2,3</sup>,  
Jesse D Bloom<sup>1,2,4\*</sup>

<sup>1</sup>Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, United States; <sup>2</sup>Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, United States; <sup>3</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, United States; <sup>4</sup>Howard Hughes Medical Institute, Seattle, United States

**Abstract** Viruses like influenza are infamous for their ability to adapt to new hosts. Retrospective studies of natural zoonoses and passaging in the lab have identified a modest number of host-adaptive mutations. However, it is unclear if these mutations represent all ways that influenza can adapt to a new host. Here we take a prospective approach to this question by completely mapping amino-acid mutations to the avian influenza virus polymerase protein PB2 that enhance growth in human cells. We identify numerous previously uncharacterized human-adaptive mutations. These mutations cluster on PB2's surface, highlighting potential interfaces with host factors. Some previously uncharacterized adaptive mutations occur in avian-to-human transmission of H7N9 influenza, showing their importance for natural virus evolution. But other adaptive mutations do not occur in nature because they are inaccessible via single-nucleotide mutations. Overall, our work shows how selection at key molecular surfaces combines with evolutionary accessibility to shape viral host adaptation.

DOI: <https://doi.org/10.7554/eLife.45079.001>

\*For correspondence:  
jbloom@fredhutch.org

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 22

**Received:** 11 January 2019

**Accepted:** 31 March 2019

**Published:** 30 April 2019

**Reviewing editor:** Yoshihiro Kawaoka, University of Tokyo, Japan

© Copyright Soh et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Viruses are exquisitely adapted to interact with host-cell machinery to facilitate their replication. Despite significant differences in this machinery across host species, some viruses like influenza can evolve to infect divergent hosts (*Parrish et al., 2008; Webster et al., 1992*). Such zoonotic transmissions can have severe public health consequences: transmission of influenza virus from birds or pigs to humans has resulted in four pandemics over the last century (*Taubenberger and Kash, 2010*). These pandemics require the virus to adapt to the new host (*Long et al., 2019*). Delineating how viruses adapt to new hosts will aid in our ability to understand what determines if a chance zoonotic infection evolves into a human pandemic.

One critical determinant of influenza host range is the viral polymerase (*Long et al., 2019*), which transcribes and replicates the viral genome (*Eisfeld et al., 2015; te Velhuis and Fodor, 2016*). Avian influenza polymerases perform poorly in mammalian cells (*Cauldwell et al., 2014; Mänz et al., 2012; Massin et al., 2001; Naffakh et al., 2000*). This host range restriction likely arises from the need for the viral polymerase to interact with host proteins such as importin- $\alpha$  (*Resa-Infante and Gabriel, 2013*) and ANP32A (*Long et al., 2016*), which differ between avian and mammalian hosts. However, it remains unclear exactly how the molecular interfaces between the polymerase and these host proteins are altered during adaptation to humans (*Long et al., 2019*).

**eLife digest** Viruses copy themselves by hijacking the cells of an infected host, but this comes with some limitations. Cells from different species have different molecular machinery and so viruses often have to specialize to a narrow group of species. This specialization consists largely of fine-tuning the way that viral proteins interact with host proteins.

For instance, in bird flu viruses, a protein known as PB2 does not interact well with the machinery in human cells. Because PB2 proteins form part of the viral polymerase (the structure that copies the viral genome), this prevents bird flu viruses from replicating efficiently in humans. Sometimes however, changes in the PB2 protein allow bird flu viruses to better replicate in humans, potentially leading to deadly flu pandemics.

To understand exactly how this happens, researchers have previously used two approaches: examining the changes that have happened in past flu viruses, and monitoring the evolution of bird flu viruses grown in human cells in the lab. However, these approaches can only look at a small number of the many possible genetic changes to the virus. This makes it hard to anticipate the new ways that flu might adapt to human cells in the future.

To overcome this problem, Soh et al. systematically created all of the single changes to the bird flu PB2, altering every element of the protein sequence one-by-one. They then tested which of the changes to PB2 helped the virus grow better in human cells. The modifications that made the viruses thrive were on the surface of the protein, suggesting that they might improve interaction with the cell machinery of the host. Some changes have been found in bird flu viruses that have recently jumped into humans in nature, although fortunately none of these viruses have yet spread widely to cause a pandemic.

Many factors affect the evolution of viruses, and their ability to infect new species. Understanding which changes in proteins help these microbes adapt to new hosts is an important element that scientists could consider to assess future risks of pandemics.

DOI: <https://doi.org/10.7554/eLife.45079.002>

Studies of natural zoonoses and experimental passaging of viruses in the lab have identified a number of mutations that adapt avian influenza polymerases to mammalian hosts (*Bussey et al., 2010; Cauldwell et al., 2014; Cauldwell et al., 2013; Chen et al., 2006; Finkelstein et al., 2007; Gabriel et al., 2005; Hu et al., 2017; Hu et al., 2014; Kim et al., 2010; Mänz et al., 2016; Mehle and Doudna, 2009; Miotto et al., 2008; Mok et al., 2011; Naffakh et al., 2000; Reperant et al., 2012; Taft et al., 2015; Tamuri et al., 2009; Yamada et al., 2010; Zhou et al., 2011*). The best known of these mutations is E627K in the PB2 subunit of the polymerase (*Subbarao et al., 1993*). This mutation alone significantly improves avian influenza polymerase activity in mammalian cells (*Long et al., 2013; Massin et al., 2001; Mehle and Doudna, 2008; Naffakh et al., 2000*), and was considered a key step in adaptation to humans (*Taubenberger et al., 2005*). But surprisingly, the recent 2009 H1N1 pandemic lineage lacks the E627K mutation. Instead, it has acquired mutations to PB2 at sites 590 and 591 that similarly confer improved polymerase activity (*Mehle and Doudna, 2009; Yamada et al., 2010*). This fact underscores the possibility that natural evolution has explored only a small fraction of the possible host-adaptation mutations. Examining only the currently available instances of adaptation in nature or the lab may therefore overlook additional mechanisms of adaptation and evolutionary paths to future zoonoses.

Here, we map all single amino-acid mutations to an avian influenza PB2 protein that enhance growth in human cells versus avian cells. We do so by leveraging deep mutational scanning (*Boucher et al., 2014; Fowler and Fields, 2014*), which previously has only been used to measure the functional effects of mutations to several influenza proteins in mammalian cells (*Ashenberg et al., 2017; Bloom, 2014; Doud and Bloom, 2016; Du et al., 2018; Jiang et al., 2016; Lee et al., 2018; Wu et al., 2014a; Wu et al., 2014b*). We show that comparative deep mutational scanning in human versus avian cells identifies numerous human-adaptive mutations that have never before been described. These mutations cluster on the surface of the PB2 protein, highlighting potential interfaces with host factors. Some of these mutations are enriched in avian-human

transmission of H7N9 influenza, demonstrating the utility of our experiments for anticipating PB2's adaptation in nature. The human-adaptive mutations that have not been observed in nature are often inaccessible by single-nucleotide mutations. Overall, our complete map of human-adaptive mutations sheds light on how species-specific selection and evolutionary accessibility shape influenza virus's evolution to new hosts.

## Results

### Deep mutational scanning of an avian influenza PB2

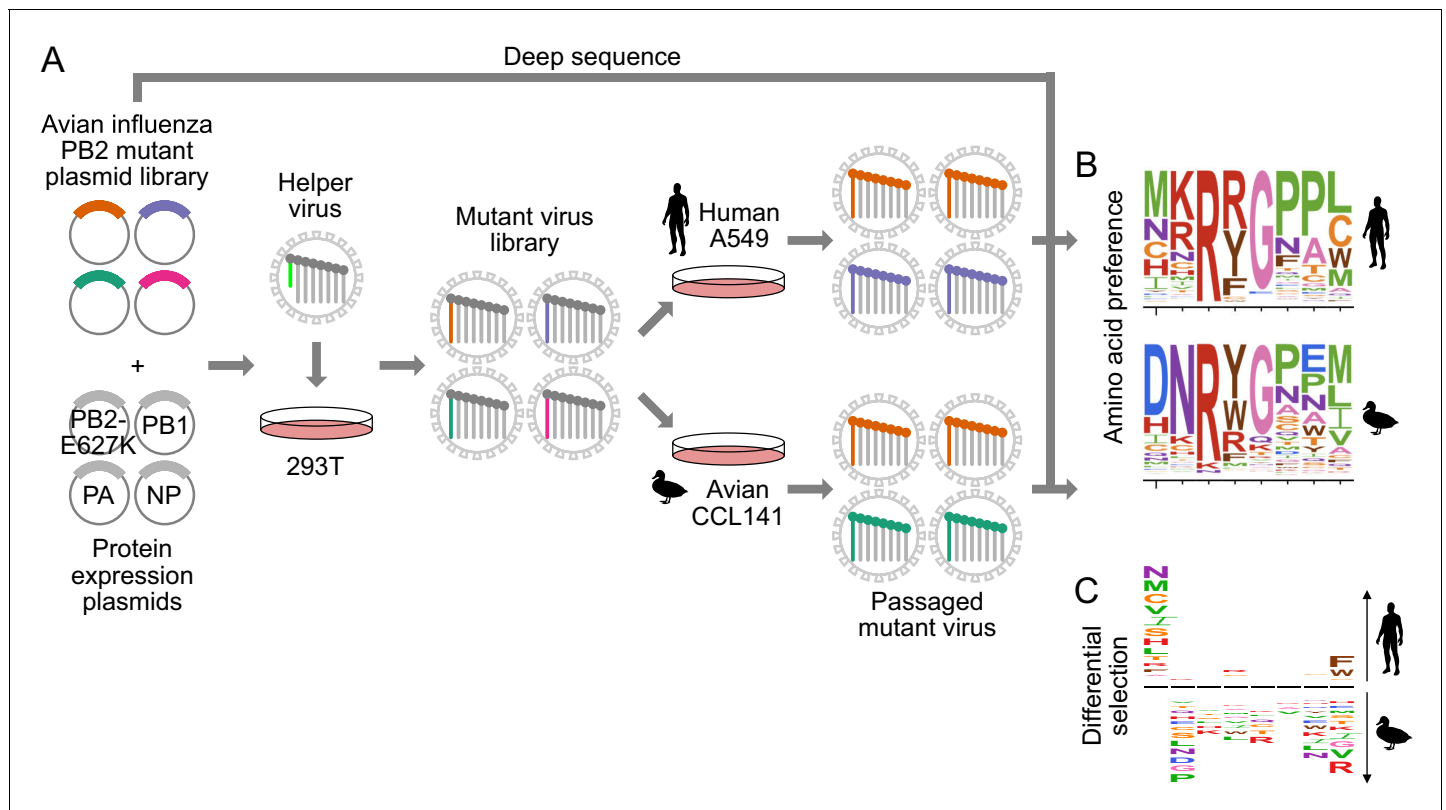
To identify host-adaptation mutations in PB2, we used deep mutational scanning to measure the effects of all amino-acid mutations to this protein in both human and avian cells. We performed these experiments using the PB2 from an avian influenza strain, A/Green-winged Teal/Ohio/175/1986 (also previously referred to as S009) (Jagger *et al.*, 2010; Mehle and Doudna, 2009). The PB2 from this strain is representative of avian influenza PB2s, most of which are highly similar (average pairwise amino-acid identity of 98.7%) (Figure 1—figure supplement 1). We mutagenized all codons in PB2 to create three replicate mutant plasmid libraries with an average of 1.4 codon substitutions per clone (Figure 1A, Figure 1—figure supplement 2A–F). Since there are 759 residues in PB2, there are  $759 \times 19 = 14,421$  amino acid mutations, virtually all of which are represented in our libraries (Figure 1—figure supplement 2G).

We generated a mutant virus library from each of the triplicate plasmid mutant libraries using a helper-virus approach, which reduces bottlenecks during generation of complex viral libraries (Doud and Bloom, 2016) (Figure 1A). For biosafety reasons, we rescued reassortant virus using polymerase (PB2, PB1, PA) and nucleoprotein (NP) genes from the avian influenza strain and the remaining viral genes (HA, NA, M, NS) from the lab-adapted A/WSN/1933(H1N1) mammalian influenza strain. We wanted to minimize selection for host-adaptive mutations during the initial library generation. Therefore, we generated the libraries in human HEK293T cells with a co-transfected protein-expression plasmid encoding the human-adapted PB2-E627K protein variant, so that all cells had a PB2 protein that could complement poorly functioning library variants.

To select for functional PB2 variants in human versus avian cells, we passaged each replicate mutant virus library at low MOI in the A549 human lung epithelial carcinoma line and CCL-141 duck embryonic fibroblasts (Figure 1A, Figure 1—figure supplement 2A). To quantify the functional selection on each mutation during viral growth, we deep sequenced the initial plasmid mutant libraries and the passaged mutant viruses to measure the frequency of mutations before and after selection (Figure 1B, Figure 1—figure supplement 3). All experiments were also performed in parallel on virus carrying wild-type PB2 as a control to quantify the rate of errors arising during sequencing, library preparation, and viral replication (Figure 1—figure supplement 2H).

To assess the efficacy of selection without the complication of errors arising from sequencing and passaging, we examined the post-selection frequency of stop and nonsynonymous mutations accessible by >1 nucleotide substitution. Stop and nonsynonymous mutations fell to 2–7% and 26–35% of their initial frequencies respectively (Figure 1—figure supplement 2H). In contrast, synonymous mutations remained at 68–87% of their initial frequency. Therefore, the experiments effectively selected for functional PB2 mutants.

We quantified selection at the amino-acid level in terms of the 'preference' of each site in the protein for each amino acid (Figure 1B) (Bloom, 2015). The preference for an amino acid is proportional to its enrichment during functional selection. We assessed the reproducibility of our experiments across biological replicates by examining the correlations of preferences for all 14,421 amino acid (Figure 1—figure supplement 2I). Biological replicates passaged in each cell type were well correlated (Pearson's *R* in human cells was 0.74 to 0.79; Pearson's *R* in avian cells was 0.76 to 0.79), and were generally better correlated within cell types than between cell types (Pearson's *R* between cell types was 0.67 to 0.78). For downstream analyses, we rescaled our preferences to match the stringency of selection in nature (see Materials and methods, Supplementary file 4, Figure 2—source data 1).



**Figure 1.** Deep mutational scanning of avian influenza PB2 in human and avian cells. (A) We mutagenized all codons of PB2 from an avian influenza strain. We generated mutant virus libraries using a helper-virus approach, and passaged libraries at low MOI in human (A549) or duck (CCL-141) cells to select for functional PB2 variants. (B) We deep sequenced PB2 mutants from the initial mutant plasmid library and the mutant virus library after passage through each cell type. We computed the ‘preference’ for each amino acid in each cell type by comparing the frequency of each mutation before and after selection. In the logo plots, the height of each letter is proportional to the preference for that amino acid at that site. (C) To identify mutations that are adaptive in one cell type versus the other, we computed the differential selection by comparing the frequency of each amino-acid mutation in human versus avian cells. Letter heights are proportional to the log enrichment of the mutation in human versus avian cells. **Figure 1—figure supplement 1** shows the phylogenetic relation of the chosen avian influenza strain to other influenza strains. **Figure 1—figure supplement 2** shows further details of deep mutation scanning experiment. **Figure 1—figure supplement 3** shows relative amplification of full-length PB2 versus PB2-GFP and PB2-deletion gene segments.

DOI: <https://doi.org/10.7554/eLife.45079.003>

The following figure supplements are available for figure 1:

**Figure supplement 1.** Phylogenetic relationship of PB2 sequence of chosen avian influenza strain to other influenza strains.

DOI: <https://doi.org/10.7554/eLife.45079.004>

**Figure supplement 2.** Details of deep mutational scanning experiment.

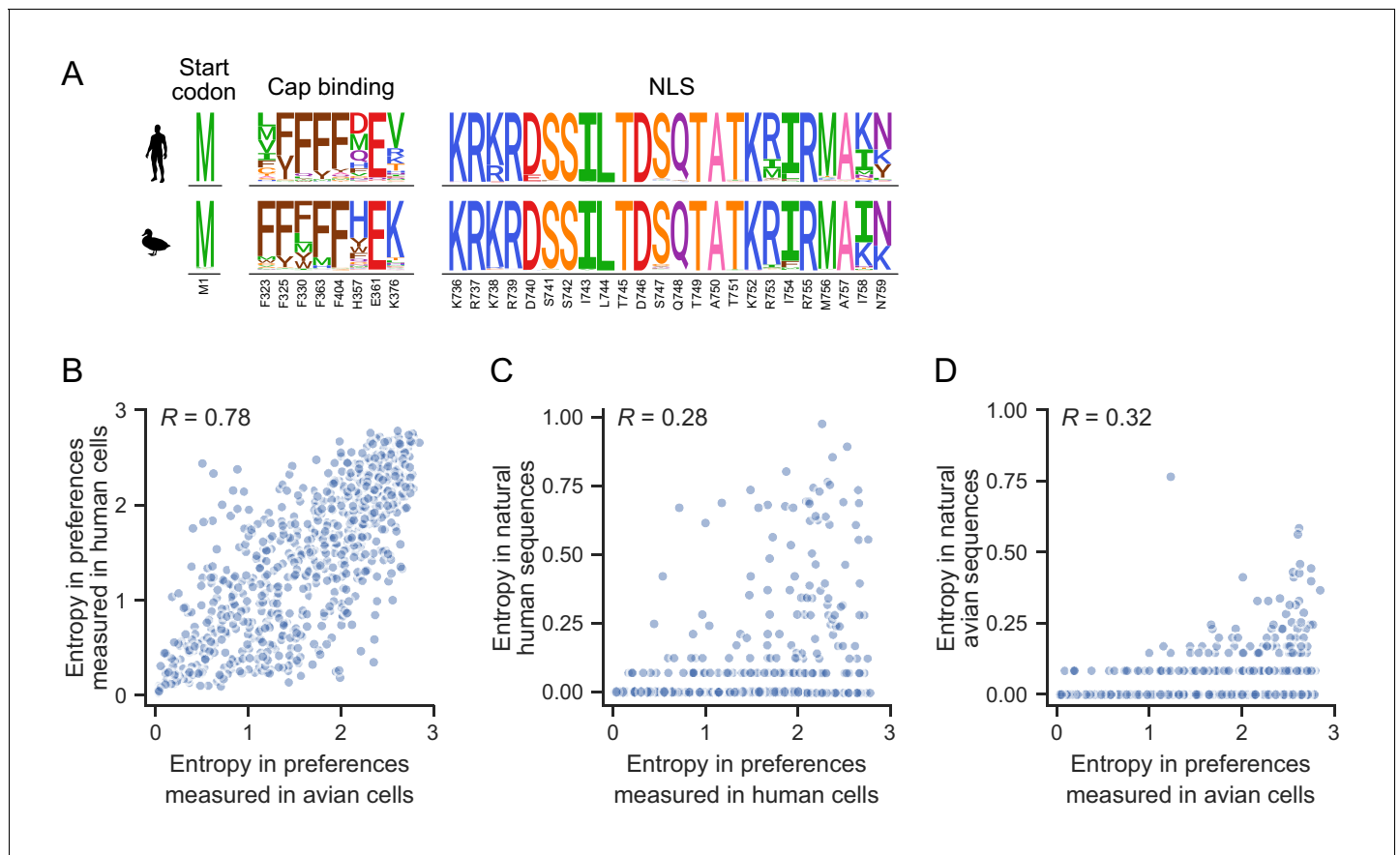
DOI: <https://doi.org/10.7554/eLife.45079.005>

**Figure supplement 3.** Relative amplification of full-length PB2 versus PB2-GFP and PB2-deletion gene segments.

DOI: <https://doi.org/10.7554/eLife.45079.006>

## Experimental measurements are consistent with natural selection and known functional constraints on PB2

Our experiments reflect known functional constraints on PB2 (**Figure 2A**, **Figure 2—figure supplement 1**). As expected, the start codon shows a strong preference for methionine in both human and avian cells. PB2’s cap-binding function is mediated by a hydrophobic cluster of five phenylalanines (F404, F323, F325, F330, F363), H357, E361, and K376 (**Guilligay et al., 2008**). Phenylalanines are strongly preferred in the hydrophobic cluster in both host cell types, with the exception of site 323, which also tolerates aliphatic hydrophobic residues in human cells (**Figure 2A**). E361 is also strongly preferred in both cell types, as is K376 in the duck cells. A number of other amino acids are tolerated at site 376 in human cells, and at site 357 in both cell types. At site 357, aromatic residues tyrosine,



**Figure 2.** Functional constraints on PB2. (A) The amino acid preferences measured in human and avian cells for key regions of PB2: the start codon, sites involved in cap-binding, and sites comprising the nuclear localization sequence (NLS). The height of each letter is proportional to the preference for that amino acid at that site. Known critical amino acids are generally strongly preferred in both cell types. (B) Correlation of the site entropy of the amino-acid preferences measured in each cell type. (C) Sites of high variability (as measured by entropy) in natural human influenza sequences occur at sites of high entropy as experimentally measured in human cells. (D) Sites with high variability in natural avian influenza sequences occur at sites of high entropy as experimentally measured in duck cells. **Figure 2—figure supplement 1** shows the complete map of amino acid preferences as measured in human and avian cells. Preferences (as well as mutation effect and differential selection for all mutations as calculated for **Figure 3**) are in **Figure 2—source data 1**.

DOI: <https://doi.org/10.7554/eLife.45079.007>

The following source data and figure supplement are available for figure 2:

**Source data 1.** Preference, mutation effect, and differential selection results for all mutations.

DOI: <https://doi.org/10.7554/eLife.45079.009>

**Figure supplement 1.** Complete map of amino acid preferences measured in human and avian cells.

DOI: <https://doi.org/10.7554/eLife.45079.008>

tryptophan, and phenylalanine are preferred in addition to histidine, consistent with previous observations that the H357W substitution enhances binding to the  $m^7$ GTP base (Guilligay et al., 2008). Finally, the two motifs comprising the C-terminal bipartite nuclear import signal, 736-KRKR-739 and 752-KRIR-755 (Tarendeau et al., 2007), are strongly and similarly preferred in both host cell types. Thus, our experimentally measured preferences largely agree with what is known about PB2 structure and function, and further suggest that functional constraints at these critical sites are similar in both human and avian cells.

To more broadly investigate whether functional constraints are similar between both cell types across the entire PB2 protein, we computed the entropy of the amino acid preferences at each site. A larger site entropy indicates a higher tolerance for mutations at that site. Site entropies are well correlated between cell types (**Figure 2B**,  $R = 0.78$ ), indicating that sites are usually under similar

functional constraint in both cell types. These protein-wide measures of mutational tolerance are also consistent with natural sequence variation. Natural variation is generally low in natural avian sequences, probably because influenza A virus is highly adapted to avian hosts so there is little pressure for additional adaptation. Natural variation is generally higher in natural human sequences, likely because of increased genetic diversity generated as a result of directional selection to adapt to the human host (*dos Reis et al., 2011*), and diversifying selection to escape immune selection on PB2-derived T-cell epitopes (*Assarsson et al., 2008*). Sites that are highly variable among publicly available natural influenza sequences tend to also be ones that we experimentally measured to be mutationally tolerant (*Figure 2C,D*).

## Identification of human-adaptive mutations

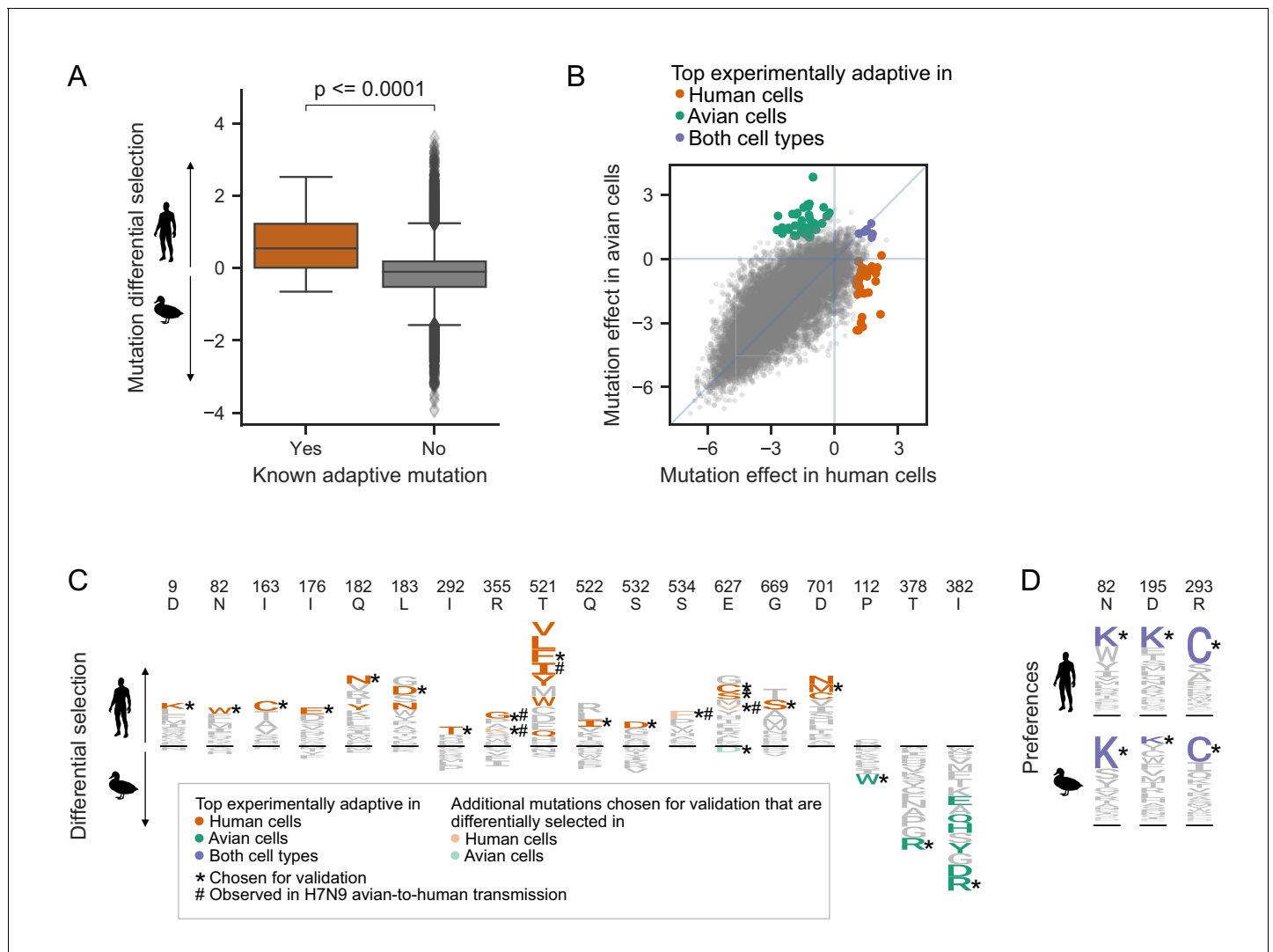
To identify mutations that are adaptive in human versus avian cells, we quantified the host-specific effect of each mutation using two different metrics. The first metric, differential selection, quantifies how much a mutation is selected in one condition versus another (*Doud et al., 2017*). Differential selection is computed by taking the logarithm of the relative enrichment of the mutation relative to the wild-type residue in human versus avian cells (*Figure 1C, Figure 3—figure supplement 1A, Figure 2—source data 1*). Differential selection greater than zero indicates that a mutation is relatively more favorable in human than avian cells.

To test if differential selection accurately identifies host-specific mutations, we asked if a set of 25 previously experimentally verified human- or mammalian-adaptive mutations (*Figure 3—source data 1*) have differential selection values greater than zero. Indeed, most of these previously characterized mutations had positive differential selection values, as expected for human-adaptive mutations (*Figure 3A*). In contrast, all other mutations have a distribution of differential selection values centered around zero.

However, there are many previously uncharacterized mutations that have differential selection values similar to or greater than those of known human-adaptive mutations (*Figure 3A*). Of course, differential selection only quantifies the extent to which a mutation is more beneficial in human than avian cells. But importantly, for a mutation to be truly adaptive, it must also be more beneficial than the wild-type amino acid in human cells. To quantify each mutation's effect relative to wild type in each cell type, we computed the logarithm of the ratio of preferences of the mutant versus wild-type amino acid (*Figure 3B, Figure 2—source data 1*). Mutation effect values greater than zero indicate that a mutation is more preferred than the wild-type residue.

We identified top experimentally adaptive mutations using both differential selection and mutation effect metrics (*Figure 3B, Figure 3—figure supplement 1B*). We focused on the 34 mutations most adaptive in our human cell selection (differential selection  $>1.5$  and mutational effect in human cells  $>1$ ). Among these 34 mutations, only one (D701N) has already been described as human adaptive. The E627K mutation is favored in human cells in our experiments, though it is not in this set of top 34 mutations. However, two other mutations at this site (E627C and E627S) are among the top 34 mutations (*Figure 3C*). S627 is naturally encoded by bat influenza and, in the context of bat influenza polymerase, supports high polymerase activity in human cells (*Tong et al., 2013; Tong et al., 2012*). In fact, it appears that many mutations at site 627 are human adaptive, with the exception of E627D. These observations are consistent with prior findings that a wide range of amino acid residues can be accommodated at site 627, and in fact improve polymerase activity in human HEK293T cells over the consensus avian wild-type glutamic acid (*Chin et al., 2014*). We additionally identified 42 mutations as adaptive in avian cells (differential selection  $<-2$  and mutational effect in avian cells  $>1$ ), and seven mutations that are more favorable than the wild-type amino acid in both cell types (mutational effect  $>1$  in both human and avian cells).

From these top adaptive mutations identified in our deep mutational scanning, we chose 26 for experimental validation. Specifically, we chose 18, four, and three mutations adaptive in human, avian, or both cell types, respectively (*Figure 3C,D*). We prioritized mutations that had consistent measurements across biological replicates. When there were multiple strongly adaptive mutations at a site, we chose just one mutation at that site to test mutations across more sites. Finally, we also validated additional mutations of particular interest, such as the only mutation at site 627 (E627D) that appeared to be favored in avian over human cells (negative differential selection), and mutations observed in avian-to-human transmission of H7N9 influenza (see below for more details).



**Figure 3.** Deep mutational scanning identifies known and novel host-adaptive mutations. (A) Distribution of experimentally measured differential selection for previously characterized human adaptive mutations and all other possible mutations to PB2. Positive differential selection means a mutation is favored in human versus avian cells. (B) Scatterplot of each mutation’s effect in human versus avian cells, showing the top adaptive mutations identified in the deep mutational scanning. (C) Logoplots showing the differential selection at the sites of mutations that we chose for functional validation. The height of each letter above the line indicates how strongly it was selected in human versus avian cells. Top adaptive mutations are colored in orange (human-adaptive) or green (avian-adaptive). Mutations chosen for functional validation are indicated by an asterisk(\*). Additional mutations chosen for validation are colored light orange (differentially selected in human) or light green (differentially selected in bird). Mutations observed in H7N9 avian-to-human transmission are indicated by #. Note that not all mutations with high differential selection in human versus avian cells are classified as top adaptive mutations because we also filtered for mutations that are substantially beneficial relative to wildtype. (D) Logoplots showing amino acid preferences at sites we chose for functional validation. Top mutations beneficial in both human and avian cells are colored purple. Mutations chosen for validation are indicated by \*. **Figure 3—figure supplement 1** shows the complete map of differential selection in human versus avian cells. Catalog of previously described human/mammalian adaptive mutations are in **Figure 3—source data 1**.

DOI: <https://doi.org/10.7554/eLife.45079.010>

The following source data and figure supplement are available for figure 3:

**Source data 1.** Catalog of previously described human/mammalian adaptive mutations.

DOI: <https://doi.org/10.7554/eLife.45079.012>

**Figure supplement 1.** Complete map of differential selection in human versus avian cells.

DOI: <https://doi.org/10.7554/eLife.45079.011>

## Human-adaptive mutations identified in deep mutational scanning improve polymerase activity and viral growth in human cells

The main function of the influenza polymerase is to transcribe and replicate the viral genome. We quantified the effect of mutations on polymerase activity using a minigenome assay which measures transcription of an engineered viral RNA encoding GFP by reconstituted influenza polymerase. To test whether the results of our deep mutational scanning are generalizable to human cells beyond the A549 cell line used in the scanning, we performed the minigenome polymerase activity assay in HEK293T as well as A549 cells.

Almost all the putative human-adaptive mutations identified in the deep mutational scanning improved transcriptional activity in human cells relative to the wild type or a synonymous mutant (**Figure 4A,B, Figure 4—source data 1**). Mutations that did not improve transcriptional activity retained at least wild-type activity. The effect of mutations in both human cell lines were remarkably consistent, suggesting that our deep mutational scanning yielded results that generalize across human cells.

In contrast, all but one of the putative avian-adaptive mutations decreased transcriptional activity in human cells compared to wild type, as expected (**Figure 4A,B**). The one mutation that did not decrease transcriptional activity, I382D, had comparable activity to wild type. Finally, mutations that are putatively adaptive in both human and avian cells had modestly improved or comparable transcriptional activity in human cells compared to wild type.

We also tested the effect of some of the mutations on viral growth to capture any effects of mutations beyond transcriptional activity. We selected some mutations that increased transcriptional activity, and others that had no effect of transcriptional activity. Our rationale in these choices was to determine if mutations that were identified in our screen but did not increase transcriptional activity still resulted in improved viral growth in human over avian cells. For each competition, we infected human (A549) and avian (CCL-141) cells with mutant and wild-type viruses mixed at a 1:1 ratio of transcriptionally active particles as determined by flow cytometry for HA expression in infected cells. We then measured the frequencies of mutant to wild-type virus by deep sequencing. To measure kinetics of viral genome replication from a single round of infection, we infected cells at MOI of 0.1, collected samples at 10 hr post infection, and sequenced vRNA from cellular extract. To measure multi-cycle replication kinetics, we infected cells at MOI of 0.01, collected samples at 48 hr post infection, and sequenced vRNA from the supernatant. At the end of the infection, we calculated the ratio of mutant to wild-type virus in human cells divided by the same ratio in avian cells. Because this quantity is a ratio of ratios, it corrects for any possible deviations from a 1:1 ratio of infectious particles in the initial inoculum. Therefore, ratios greater than one indicates that the mutation confers a relative benefit for viral growth in avian versus human cells.

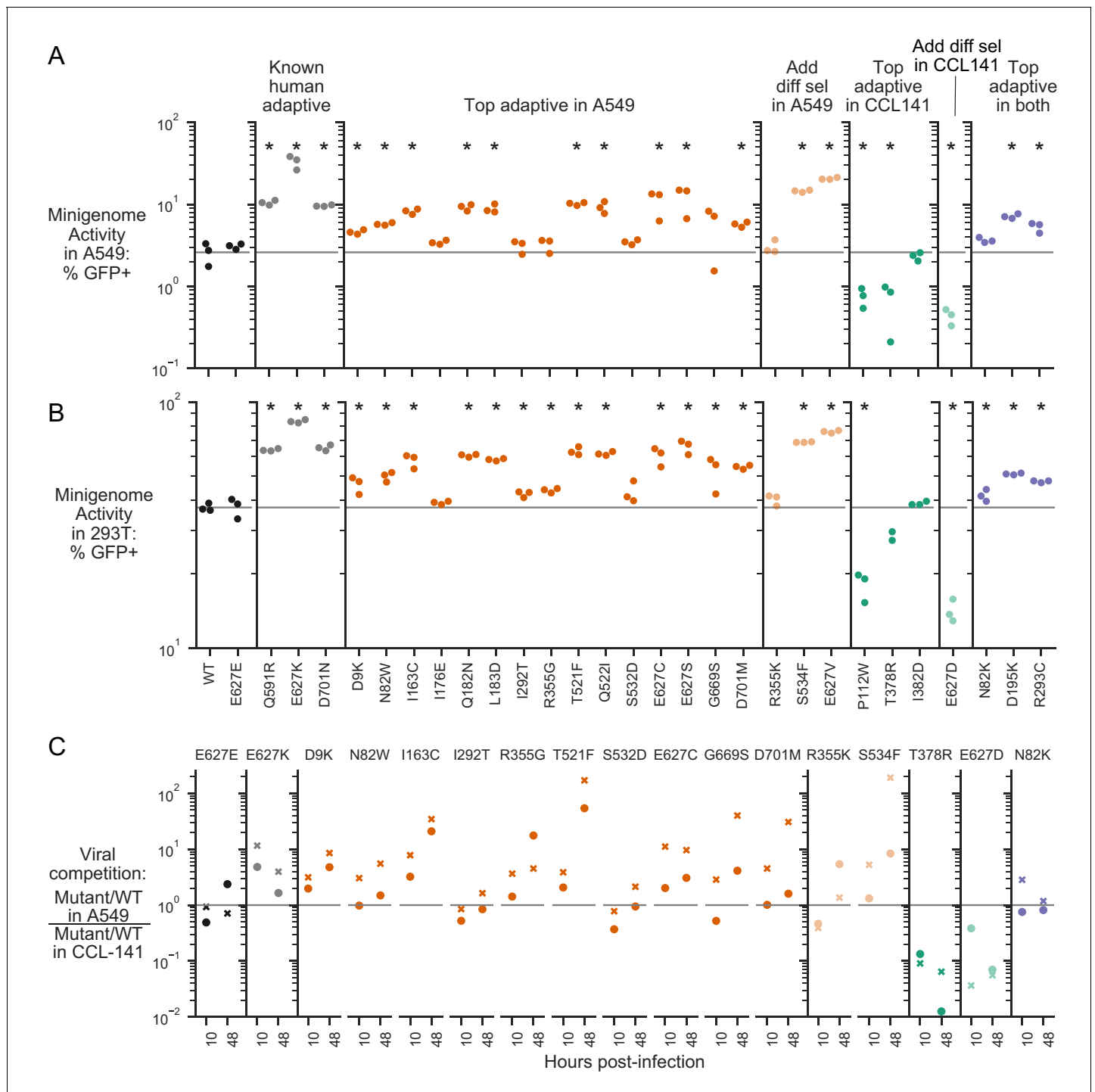
Almost all putative human-adaptive mutations identified in the deep mutational scanning improved growth in human over avian cells, as reflected by an increase in the ratio of mutant to wild type in human cells versus avian cells over the time course of the competition (**Figure 4C, Figure 4—source data 2**). Of the putative human-adaptive mutations that did not improve polymerase activity, one mutation, R355G, improved growth in human over avian cells at both 10 and 48 hr post-infection. An additional mutation, R355K, slightly improved growth in human cells by 48 hr post-infection. These two mutations may therefore confer a human-specific growth benefit due to some mechanism other than polymerase activity. As expected, both putative avian-adaptive mutations resulted in poorer growth in human versus avian cells. Finally, the N82K mutation that is putatively adaptive in both human and avian cells resulted in comparable growth in both human and avian cells, as expected.

Thus, our deep mutational scanning identified numerous previously undescribed PB2 mutations that improve polymerase activity or viral growth in human cells. In addition, it also identified an intriguing small set of mutations that enhance viral growth but not polymerase activity in human cells.

### Human-adaptive mutations cluster in regions of PB2 that are potentially important for host adaptation

The additional human-adaptive mutations we identified may improve PB2's ability to interact with important human cell factors. To identify potential interfaces for such interactions, we mapped the





**Figure 4.** Validation of top experimentally adaptive mutations. The polymerase activity of selected PB2 mutants as measured using minigenome assays in A549 (A) and HEK293T (B) human cells. The mutations chosen for characterization include previously known human adaptive mutations, top adaptive mutations identified by our deep mutational scanning (orange = human adaptive, green = avian adaptive), and additional mutations differentially selected in human (light orange) or avian (light green) cells. E627E is a synonymous mutation at site 627 used as a negative control. Minigenome activity is represented as percent of transfected cells that expressed a viral GFP reporter. The gray horizontal line indicates the mean value measured for the wild type avian PB2. Minigenome assays were performed in biological triplicate. Mutations that have significantly different minigenome activity from wild type are indicated by asterisks (unpaired t-test,  $p < 0.05$ ). (C) Competition of virus bearing the indicated mutant PB2 against virus with wild-type PB2. For each competition, human A549 and avian CCL-141 cells were infected with mutant and wild-type viruses mixed at a 1:1 ratio of transcriptionally active particles, and the frequency of each variant after viral replication was measured by deep sequencing viral RNA. For samples collected at 10 hr post infection, we infected cells at MOI of 0.1, and sequenced vRNA from cellular extract. For samples collected at 48 hr post

*Figure 4 continued on next page*

Figure 4 continued

infection, we infected cells at MOI of 0.01, and sequenced vRNA from the supernatant. The plots show the ratio of the mutant over wild-type variant in A549, divided by the same ratio in CCL-141 cells. A ratio >1 indicates that a viral mutant grows better in human than avian cells. Competition assays were performed in biological duplicate; circle and cross represent replicate experiments. Flow data for minigenome activity and mutation counts for viral competition are provided in **Figure 4—source datas 1** and **2**.

DOI: <https://doi.org/10.7554/eLife.45079.013>

The following source data is available for figure 4:

**Source data 1.** Flow cytometry data for minigenome assays.

DOI: <https://doi.org/10.7554/eLife.45079.014>

**Source data 2.** Mutant frequency data for competition assay.

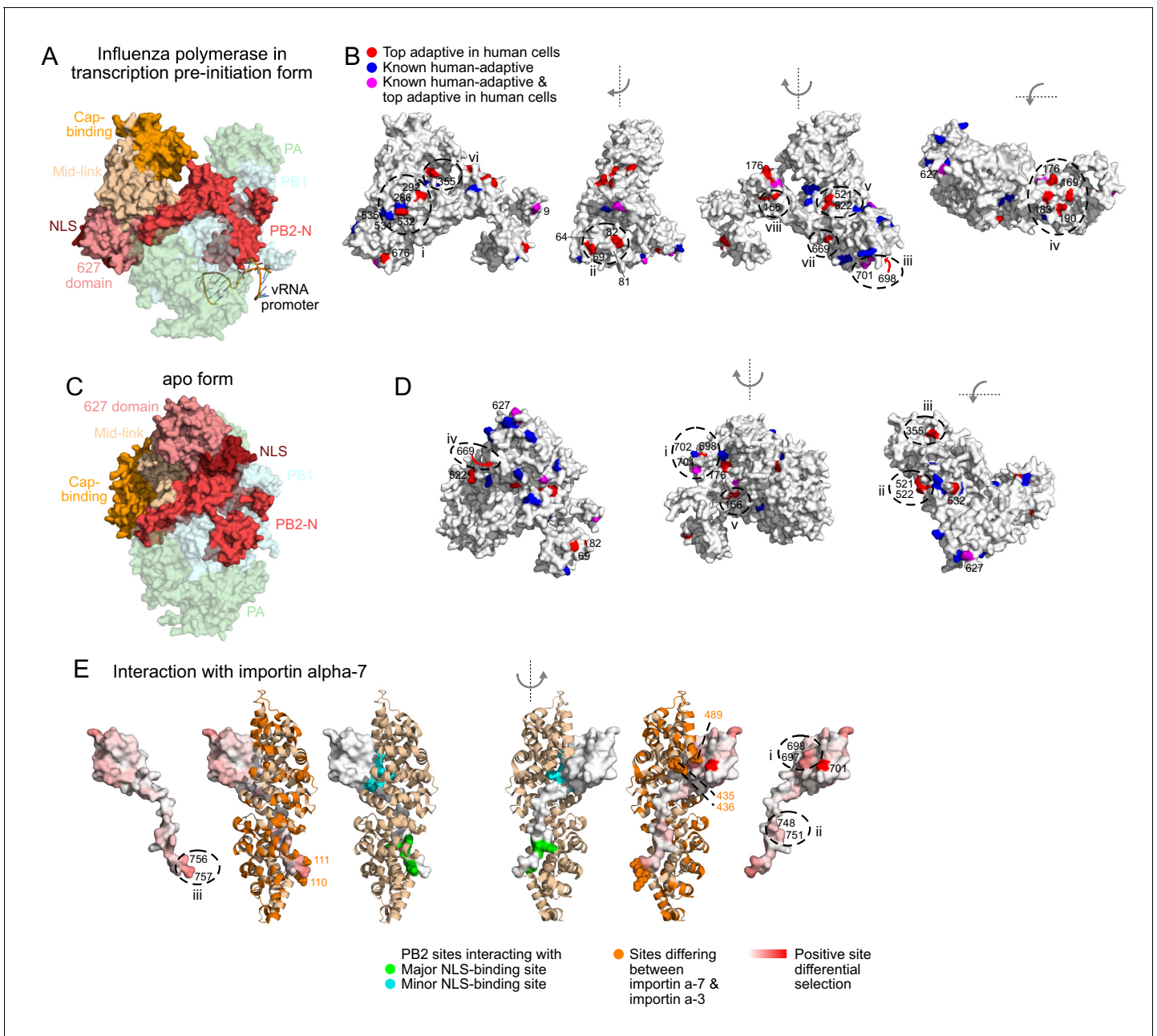
DOI: <https://doi.org/10.7554/eLife.45079.015>

sites of top human-adaptive mutations identified in the deep mutational scanning onto the structure of PB2. Many of the sites cluster in regions of PB2 that may play a role in host adaptation (**Figure 5, Figure 5—figure supplement 1**).

Most adaptive mutations are surface exposed in at least one of the two conformations of the polymerase we examined (**Figure 5—figure supplement 1A**, relative solvent accessibility >0.2). In the transcription pre-initiation form of the polymerase (PDB: 4WSB) (**Reich et al., 2014**) (**Figure 5A, B**), the experimentally identified sites 532 and 292 are located on the surface near sites of known human-adaptive sites 286, 534, and 535 (**Cauldwell et al., 2013; Mänz et al., 2016**) (**Figure 5B: i**). Experimentally identified sites 69 and 82 are located on the surface close to sites 64 and 81 (**Figure 5B: ii**), which are located near the template exit channel and were recently shown to modulate generation of mini viral RNAs that act as innate-immune agonists (**Te Velthuis et al., 2018**). Experimentally identified site 698 is located near known sites 701 and 702 (**Gabriel et al., 2005**) (**Figure 5B: iii, Figure 5D: i**). Experimentally identified mutations are also located in regions not yet identified to be important for host-adaptation: we find a cluster of experimentally identified sites on the surface of the PB2-N terminal domain (169, 176, 183, 190, **Figure 5B: iv**), partially occluded by the flexible PA endonuclease domain in the transcription pre-initiation structure. Experimentally identified sites 163 and 182 are not surface exposed, but are buried right underneath the cluster of four sites.

Almost all sites that are not highly surface exposed in the transcription pre-initiation conformation of the polymerase become exposed upon conformational rearrangement of the polymerase from the transcription pre-initiation form to the apo form (PDB: 5D98) (**Hengrung et al., 2015**). The experimentally identified sites 521 and 522 which face the product exit channel, site 355 which faces the core of the polymerase, and site 669 which faces PB1 in the transcription pre-initiation structure (**Figure 5B: v – vii**), are more fully surface-exposed in the apo structure (**Figure 5C,D: ii – iv**). However, some sites remain inaccessible: site 156 faces the internal core of the polymerase in both transcription pre-initiation and apo forms (**Figure 5B: viii, Figure 5D: v**). Therefore, with only a few exceptions, the human-adaptive mutations cluster in patches on the surface of PB2. Taking a comprehensive approach has therefore allowed us to map surfaces of PB2 that might mediate host-interactions.

Next, we asked if host-adaptation mutations occur at known interaction interfaces with host proteins. One known interacting host protein is importin- $\alpha$ , which mediates nuclear import of PB2 and has been proposed to have a role in viral transcription and replication (**Resa-Infante et al., 2008; Tarendeau et al., 2007**). PB2 of avian viruses uses importin- $\alpha 3$  in human cells, whereas PB2 of mammalian-adapted viruses uses importin- $\alpha 7$  (**Gabriel et al., 2011; Gabriel et al., 2008**). We mapped total positive differential selection on each site of PB2 (see legend for **Figure 5E**), and asked how this selection on PB2 relates to its interaction with importin- $\alpha$ . Sites on PB2 that interact with the major and minor NLS-binding surfaces of importin- $\alpha$  (**Pumroy et al., 2015; Tarendeau et al., 2007**) generally have low differential selection, indicating that host-adaptation mutations do not occur at these sites (**Figure 5E; PDB: 4UAD**) (**Pumroy et al., 2015**). This is expected, since all importin- $\alpha$  isoforms share an invariant NLS-binding surface. However, adjacent PB2 sites have higher differential selection (**Figure 5E: i-iii**). Some of these PB2 sites are in close proximity to regions of importin- $\alpha$



**Figure 5.** Locations of top human-adaptive mutations on the structure of the influenza polymerase. Overall structure of the influenza polymerase complex comprising PB2, PB1 and PA in (A, B) the transcription pre-initiation form (PDB: 4WSB) and (C, D) the apo form (PDB: 5D98). PB2 domains defined as in *Pflug et al. (2017)*. (B, D) Sites of top human-adaptive mutations identified by deep mutational scanning are shown in red on the PB2 subunit of the structure. Sites of previously experimentally verified human-adaptive mutations are in blue (25 sites as listed in *Figure 3—source data 1*). Sites identified by deep mutational scanning and which were also previously known are in purple. A subset of sites are labeled and/or circled for referencing in the main text, to indicate surfaces that might mediate host-interactions. Similar results are obtained if we instead analyze the structures in terms of a continuous variable representing the extent of human-specific adaptation at each site (*Figure 5—figure supplement 1B, C*). (E) Structure of PB2 C-terminal fragment co-crystallized with importin- $\alpha 7$  (PDB: 4UAD). Sites on PB2 interacting with major and minor NLS binding surfaces of importin- $\alpha 7$  are in green and cyan respectively. Importin- $\alpha 7$  is depicted in ribbon form in tan. We used the deep mutational scanning to define a continuous variable indicating the extent of host-specific adaptation at each site of PB2. Specifically, for each site, we computed the positive site differential selection by summing all positive mutation differential selection values at that site (i.e., the total height of the letter stack in the positive direction in logoplots such as in *Figure 3D*). We mapped this differential selection onto the PB2 C-terminal fragment in red; PB2 sites with high differential selection are numbered. Regions of importin- $\alpha 7$  that differ from importin- $\alpha 3$  are colored in orange, those near PB2 sites with high differential selection are shown as spheres. For all structures, the avian influenza (S009) PB2 amino acid sequence was mapped onto the PB2 chain by one-2-one threading using Phyre2 (*Kelley et al., 2015*) (Confidence in models for 4WSB, 5D98, and 4UAD are 100%, 100%, and 99% respectively). Sites are numbered

*Figure 5 continued on next page*

Figure 5 continued

according to the S009 PB2 sequence. **Figure 5—figure supplement 1** shows relative solvent accessibility of human-adaptive mutations, as well as positive site differential selection mapped onto structures of influenza polymerase.

DOI: <https://doi.org/10.7554/eLife.45079.016>

The following figure supplement is available for figure 5:

**Figure supplement 1.** Solvent accessibility of sites of human-adaptive mutations, and positive site differential selection mapped onto structures of influenza polymerase.

DOI: <https://doi.org/10.7554/eLife.45079.017>

that differ between the  $\alpha$ -7 and  $\alpha$ -3 isoforms (**Figure 5D**: i, iii), suggesting that adaptation at these PB2 sites affects importin- $\alpha$  usage.

PB2 also interacts with the C-terminal domain of RNA polymerase II, and this interaction is proposed to stabilize the polymerase in the transcription-competent conformation (PDB: 6F5O) (**Serna Martin et al., 2018**). Similar to what we observe with importin- $\alpha$ , PB2 sites thought to directly interact with RNA polymerase II tend to have low differential selection, whereas adjacent PB2 sites have higher differential selection (**Figure 5—figure supplement 1D**). Thus, it appears that host adaptation may involve mutations at sites adjacent to the core residues that directly interact with host proteins.

### Experimentally defined human-adaptive mutations are enriched in avian-to-human transmission of H7N9 influenza

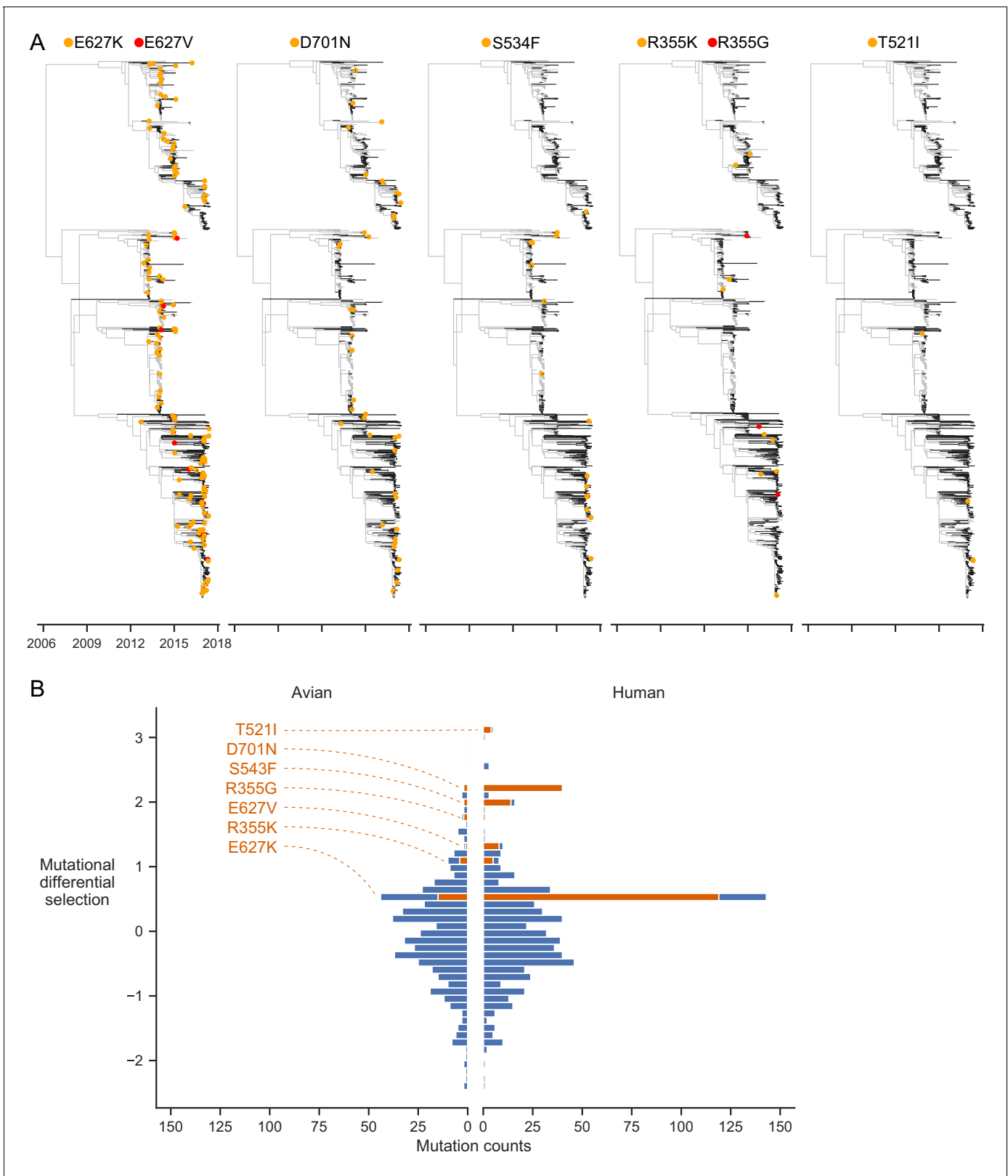
A challenge in the surveillance of non-human influenza and assessment of pandemic risk is determining which of the many mutations that occur during viral evolution are human-adaptive (**Lipsitch et al., 2016; Russell et al., 2014**). We investigated whether our experimental measurements can identify host-adaptation mutations that occur during the actual transmission of avian influenza to humans.

Avian H7N9 influenza viruses have recently caused a large number of sporadic human infections (**Su et al., 2017**). We examined mutations occurring during the evolution of H7N9 viruses that have jumped from avian to human hosts to determine whether they were enriched for changes predicted by our deep mutational scanning to be human adaptive. First, we constructed a phylogeny of H7N9 PB2 sequences, inferred ancestral sequences for all internal nodes, and assigned mutations to specific branches of the phylogenetic tree (**Figure 6A, Figure 6—figure supplement 1–5**). We then classified each mutation on the phylogenetic tree as ‘avian’ or ‘human’ based on whether it occurred on a branch connecting two avian isolates, or on a branch leading to a human isolate respectively (**Figure 6—source data 1**). As human infections by H7N9 are evolutionary dead-ends, mutations occurring during human infections should appear immediately proximal to human isolates in the phylogeny, while mutations occurring during bird infections will be ancestral.

We next asked whether mutations occurring in human hosts had higher differential selection values in our deep mutational scanning than mutations in avian hosts. Indeed, human mutations more often had high differential selection (a value  $>0.5$ ) than avian mutations (Fisher’s exact test,  $p=2.69e-7$ ) (**Figure 6B**). The H7N9 human mutations with differential selection  $>0.5$  include the well-studied human-adaptive mutations 627K and 701N. Indeed, these two mutations make up the majority of H7N9 human mutations with high differential selection. But we also identified a number of other mutations with high differential selection that occurred at least four independent times in jumps of H7N9 influenza into humans: 627V, 534F, 355K, and 521I (**Figure 6A, Figure 6—figure supplement 1–5, 6B, 3D**), only one of which has been previously characterized (E627V, **Taft et al., 2015**). A second mutation at site 355 (355G) with high differential selection also occurs during jumps of H7N9 influenza into humans. Thus, our deep mutational scanning identifies both previously characterized and novel mutations that occur in natural avian-to-human transmission of influenza.

### Most human-adaptive mutations are not accessible by single nucleotide substitutions

Our deep mutational scanning identifies many human-adaptive mutations. Why do we not observe all of them in nature? One possible explanation is that some of these mutations are inaccessible by



**Figure 6.** Experimentally identified human-adaptive mutations are enriched in avian-human transmission of H7N9 influenza. (A) Phylogeny of H7N9 influenza PB2 sequences. Branches in human and avian hosts are colored black and gray respectively. Orange or red dots indicate where a mutation was inferred to have occurred. Branch lengths are scaled by annotated and inferred dates of origin of each sequence. (B) Distribution of experimentally measured differential selection values for all mutations occurring during H7N9 evolution in human and avian hosts. A positive differential selection value indicates a mutation that is more frequent in human hosts than in avian hosts. *Figure 6 continued on next page*

Figure 6 continued

means that our experiments measured the mutation to be beneficial in human versus avian cells. A subset of top differentially selected mutations that occur frequently are labeled and plotted in orange. Enlarged phylogenetic trees are in **Figure 6—figure supplement 1–5**. Counts of mutations identified in phylogenetic analysis are in **Figure 6—source data 1**. Mutations plotted in each bin of the histogram are in **Figure 6—source data 2**.

DOI: <https://doi.org/10.7554/eLife.45079.018>

The following source data and figure supplements are available for figure 6:

**Source data 1.** H7N9 human and avian mutation counts.

DOI: <https://doi.org/10.7554/eLife.45079.024>

**Source data 2.** H7N9 human and avian mutation differential selection values and counts in each histogram bin.

DOI: <https://doi.org/10.7554/eLife.45079.025>

**Figure supplement 1.** Phylogeny of H7N9 influenza PB2 sequences showing where mutations at site 627 were inferred to have occurred.

DOI: <https://doi.org/10.7554/eLife.45079.019>

**Figure supplement 2.** Phylogeny of H7N9 influenza PB2 sequences showing where mutations at site 701 were inferred to have occurred.

DOI: <https://doi.org/10.7554/eLife.45079.020>

**Figure supplement 3.** Phylogeny of H7N9 influenza PB2 sequences showing where mutations at site 534 were inferred to have occurred.

DOI: <https://doi.org/10.7554/eLife.45079.021>

**Figure supplement 4.** Phylogeny of H7N9 influenza PB2 sequences showing where mutations at site 355 were inferred to have occurred.

DOI: <https://doi.org/10.7554/eLife.45079.022>

**Figure supplement 5.** Phylogeny of H7N9 influenza PB2 sequences showing where mutations at site 521 were inferred to have occurred.

DOI: <https://doi.org/10.7554/eLife.45079.023>

---

single nucleotide substitution from existing sequences, and are therefore less likely to arise during natural evolution (*Fragata et al., 2018*).

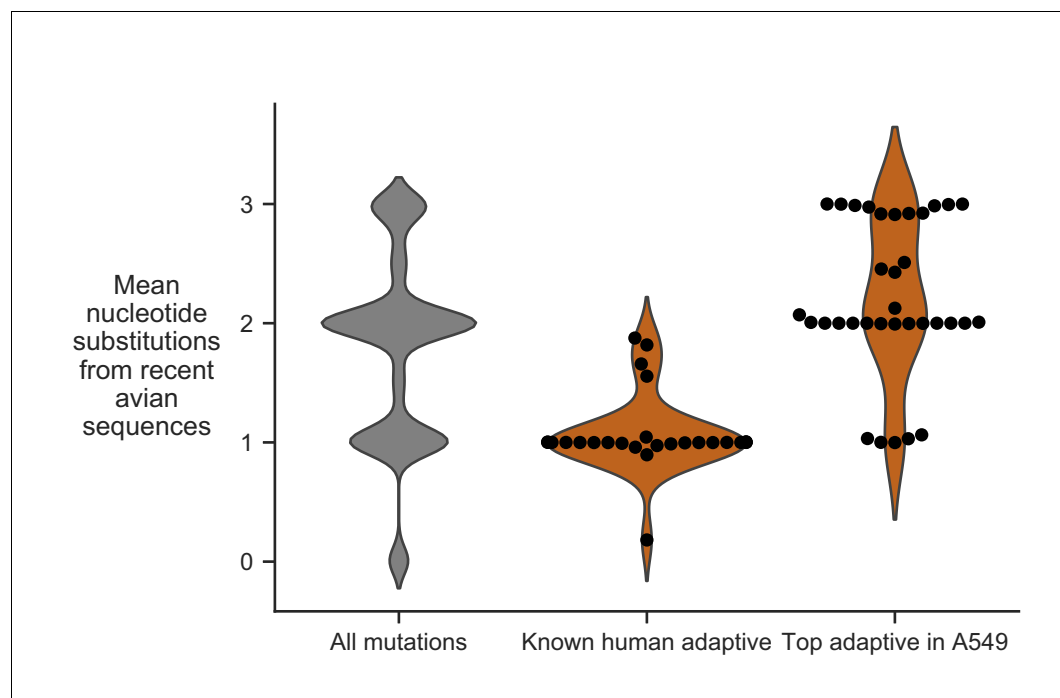
We examined if accessibility by single nucleotide substitution imposes constraint on which human-adaptive mutations arise in nature. To do so, we calculated the mean nucleotide substitutions required to access known human-adaptive mutations from all avian influenza PB2 sequences collected in the past three years. This mean number of nucleotide substitutions can range from less than one (if the mutation is already present in some avian PB2 sequences) to three (if the mutation requires three nucleotide changes from all avian PB2 sequences). The majority of previously characterized human-adaptive mutations are accessible by single nucleotide substitutions from avian PB2 sequences (**Figure 7**), suggesting that these mutations have already been characterized because they readily occur in the context of current avian influenza viruses. In contrast, most of the top human-adaptive mutations identified in our deep mutational scanning require multiple nucleotide substitutions from current avian PB2 sequences (**Figure 7**). Therefore, many of the novel human-adaptive mutations uncovered by our experiment have probably not been previously identified because they are evolutionarily inaccessible from current avian influenza sequences.

The importance of evolutionary accessibility is especially obvious if we examine the adaptive mutations that actually occur in natural influenza virus evolution. Of the five top human adaptive mutations accessible by single nucleotide substitution (**Figure 7**), three have occurred repeatedly during recent transmissions of H7N9 influenza to humans (355G, 521I, 701N; see **Figure 6A**). Thus, it may be possible to combine deep mutational scanning measurements of phenotypes with analyses of evolutionary accessibility to anticipate which mutations are most likely to arise in the context of any given starting sequence (*Fragata et al., 2018*).

## Discussion

We have measured how all single amino-acid mutations to an avian influenza PB2 affect viral growth in both human and avian cells. Our results separate the constraints on PB2 into those that are maintained across cells from diverse species versus those that are specific to human or avian cells. The vast majority of sites are under extremely similar constraint in human and avian cells, including at residues already known to be critical for PB2 function. Layered upon this conserved constraint are mutations with host-specific effects. Our approach therefore represents a powerful strategy for mapping viral determinants of cross-species transmission.

The viral determinants of influenza host specificity in the PB2 protein have been intensely studied for decades. Earlier studies addressing this question focused on mutations that have fixed during



**Figure 7.** Evolutionary accessibility of mutations from current avian influenza PB2 sequences. Distribution of mean nucleotide substitutions required to access all amino-acid mutations, previously characterized human-adaptive mutations, and top human-adaptive mutations identified in our deep mutational scanning. Mean nucleotide substitution is calculated by averaging over all avian influenza PB2 sequences collected from 2015 to 2018. Most previously characterized human-adaptive mutations are accessible by single nucleotide substitution, whereas many of the new adaptive mutations that we identified require multiple nucleotide substitutions. Mean nucleotide substitutions for each mutation are in **Figure 7—source data 1**.

DOI: <https://doi.org/10.7554/eLife.45079.026>

The following source data is available for figure 7:

**Source data 1.** Mean nucleotide substitutions from avian sequences of all mutations.

DOI: <https://doi.org/10.7554/eLife.45079.027>

viral adaptation in nature or in the lab (Bussey et al., 2010; Cauldwell et al., 2013; Chen et al., 2006; Finkelstein et al., 2007; Gabriel et al., 2005; Hu et al., 2017; Hu et al., 2014; Kim et al., 2010; Mänz et al., 2016; Mehle and Doudna, 2009; Miotto et al., 2008; Mok et al., 2011; Naffakh et al., 2000; Tamuri et al., 2009; Yamada et al., 2010; Zhou et al., 2011). Recent studies have sampled more mutations by random mutagenesis at key sites such as positions 627, 701, and 702 of PB2 (Chin et al., 2014; Chin et al., 2017), or error prone PCR across the genome (Taft et al., 2015). We can now place these previously characterized mutations in the context of a complete map, thus revealing how selection at molecular interfaces and evolutionary accessibility shapes influenza's adaptation to humans.

Examination of our maps of differential selection at known molecular interfaces, such as with importin- $\alpha$  (Pumroy et al., 2015; Tarendeau et al., 2007) and RNA polymerase II (Serna Martin et al., 2018) suggest an interesting hypothesis: host-adaptive mutations are located adjacent to but not at core residues that directly interact at host proteins, suggesting that host adaptation may involve mutations at sites at the periphery of core interactions. Of course, it remains unclear precisely how any of these new human-adaptive mutations act, though initial validations show that most improve polymerase activity in human cells. However, an intriguing set of mutations such as at site 355 improve viral growth but not polymerase activity in human cells. We speculate that the effects of these mutations on viral growth are mediated by other effects of PB2, such as in modulating the innate-immune response (Graef et al., 2010; Te Velthuis et al., 2018).

Even though we have focused our analyses mostly on host-adaptive mutations, our measurements are also useful to studies of PB2 function independent of host adaptation. Our complete maps of

amino acid preferences in each cell type provide context for the handful of residues well-known to be critical for PB2 functions. For example, although our data recapitulate most of the known functional constraints on sites previously defined to be critical for mRNA cap-binding (*Guilligay et al., 2008*), they also reveal that some of these sites can tolerate alternate residues. Thus, our complete survey of the mutational constraints at each site complements other functional and structural knowledge.

We also show that our comprehensive experimental measurements can identify human-adaptive mutations that occur during avian-to-human transmission of H7N9 influenza. These measurements therefore help address a fundamental challenge in assessing the risk of potential pandemic influenza virus strains: determining which of the many mutations observed during viral surveillance affect whether a virus will be successful in human hosts (*Lipsitch et al., 2016; Russell et al., 2014*). Our high-throughput approach therefore enables phenotypic measurements to keep pace with the challenge of interpreting the many viral mutations that are observed during genotypic surveillance of virus evolution (*Grubaugh et al., 2019*).

However, although some amino-acid mutations we experimentally identified as human-adaptive occur frequently in nature, others have never been observed. In addition, the strengths of differential selection measured in our experiments do not necessarily correspond to observed frequencies of mutations in nature. These apparent inconsistencies highlight the important role of evolutionary accessibility in shaping influenza's host adaptation. First, most human-adaptive mutations observed in natural influenza evolution are accessible by single nucleotide substitution from current avian genotypes, demonstrating how the architecture of the genetic code impacts viral adaptation (*Fragata et al., 2018*). Second, even among single-nucleotide substitutions, transition mutations are about ten-fold more frequent than transversions during influenza replication (*Bloom, 2014*). This might explain why E627K (which requires a G→A transition) is much more frequently observed than E627V (which requires a A→T transversion), despite the latter being more strongly selected in human cells in our experiments. Of course, many additional factors not captured by our experiments can impact which adaptive mutations fix in nature. Integrating our complete maps of the effects of amino-acid mutations with other data that sheds light on evolutionary opportunity, such as nucleotide accessibility, mutation rates, transmission bottlenecks, and environmental and epidemiological factors (*Geoghegan et al., 2016; Geoghegan and Holmes, 2017; Moncla et al., 2016; Olival et al., 2017; Peck and Lauring, 2018; Varble et al., 2014*) will help us understand how viruses cross species barriers in nature.

## Materials and methods

### Key resources table

| Reagent type (species) or resource        | Designation                                  | Source or reference | Identifiers                 | Additional information   |
|---|--|---------------------|-----------------------------|--|
| Cell line (Homo sapiens)                  | A549   | ATCC                | CCL-185;<br>RRID:CVCL_0023  |  |
| Cell line (Homo sapiens)                  | HEK293T                                      | ATCC                | CRL-3216;<br>RRID:CVCL_0063 |  |
| Cell line (Canis familiaris)              | MDCK-SIAT1                                   | Sigma-Aldrich       | 5071502;<br>RRID:CVCL_Z936  |  |
| Cell line (Anas platyrhynchos domesticus) | CCL-141                                      | ATCC                | CCL-141;<br>RRID:CVCL_T281  |  |
| Cell line (Canis familiaris)              | MDCK-SIAT1-tet-S009-PB2-E627K                | this paper          |                             | MDCK-SIAT1 cells expressing S009 PB2-E627K under control of a doxycycline-inducible promoter |
| Recombinant DNA reagent                   | pHW_noCMV_S009_PB2;<br>pHW_noCMVnoTerm_BsmBI | this paper          |                             | Plasmids for generating mutant plasmid library; see <b>Supplementary file 1</b>              |

*Continued on next page*



Continued

| Reagent type (species) or resource | Designation   | Source or reference   | Identifiers | Additional information  |
|------------------------------------|---|---|-------------|---|
| Recombinant DNA reagent            | pHW_S009_PB2;<br>pHW_S009_PB1;<br>pHW_S009_PA;<br>pHW_S009_NP | this paper  |             | Plasmids for generating helper virus; see <b>Supplementary file 1</b>   |
| Recombinant DNA reagent            | HDM_S009_PB2;<br>HDM_S009_PB1;<br>HDM_S009_PA;<br>HDM_S009_NP | this paper  |             | Plasmids for protein expression of S009 polymerase complex; see <b>Supplementary file 1</b>                               |
| Recombinant DNA reagent            | pHH_PB2_S009_flank_99_eGFP_100                                | this paper  |             | Plasmids for generating helper virus; see <b>Supplementary file 1</b>   |
| Recombinant DNA reagent            | pHW184_HA;<br>pHW186_NA;<br>pHW187_M;<br>pHW188_NS            | <b>Hoffmann et al. (2000)</b>   |             |   |
| Recombinant DNA reagent            | pHH-PB1-flank-eGFP  | <b>Bloom et al. (2010)</b>  |             | Reporter plasmid for minigenome assay; see <b>Supplementary file 1</b>  |
| Recombinant DNA reagent            | pcDNA3.1_mCherry  | this paper  |             | Transfection control for minigenome assay; see <b>Supplementary file 1</b>  |
| Recombinant DNA reagent            | pSBtet_RP_S009_PB2_E627K                                      | this paper  |             | Plasmid for generating PB2-expressing cell line; see <b>Supplementary file 1</b>  |
| Sequence-based reagent             | primers   | this paper  |             | See <b>Supplementary file 2</b>   |
| Commercial assay or kit            | NEBuilder HiFi DNA Assembly Master Mix                        | New England Biolabs   | E2621S      |   |
| Commercial assay or kit            | ElectroMAX DH10B competent cells                              | Invitrogen  | 18290015    |   |
| Commercial assay or kit            | Rneasy Mini Kit   | Qiagen  | 74104       |   |
| Commercial assay or kit            | Accuscript Reverse Transcriptase                              | Agilent   | 200820      |   |
| Commercial assay or kit            | KOD Hot Start Master Mix                                      | EMD Millipore   | 71842       |   |
| Commercial assay or kit            | QIAamp Viral RNA Mini Kit                                     | Qiagen  | 52904       |   |
| Commercial assay or kit            | SuperScript III   | ThermoFisher Scientific   | 18080051    |   |
| Chemical compound, drug            | BioT  | Bioland Scientific  | B01-01      |   |
| Chemical compound, drug            | Lipofectamine 3000  | ThermoFisher Scientific   | L3000015    |   |
| Antibody                           | H17-L19   | <b>Gerhard et al. (1981)</b>  |             |   |
| Software, algorithm                | dms_tools2  | <a href="https://jbloomlab.github.io/dms_tools2">https://jbloomlab.github.io/dms_tools2</a> , version 2.3.0 |             |   |
| Software, algorithm                | Jupyter notebooks that perform all steps of analyses          | this paper  |             | See <b>Supplementary file 3</b> ; <a href="https://github.com/jbloomlab/PB2-DMS">https://github.com/jbloomlab/PB2-DMS</a> |

## Cell lines and media

HEK293T, MDCK-SIAT1, and A549 (ATCC CCL-185) cells were maintained in D10 media (DMEM supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml of streptomycin). CCL-141 cells (ATCC CCL-141) were maintained in E10 media (identical to D10 except that EMEM is used in place of DMEM). Cells were grown in WSN growth media (WGM: Opti-MEM supplemented with 0.5% FBS, 0.3% BSA, 100 µg/ml CaCl<sub>2</sub>, 100 U/ml penicillin, and 100 µg/ml of streptomycin) for viral infections.

For an avian cell line, we chose to use a duck rather than a chicken cell line because ducks are natural hosts of influenza that (unlike chickens) possess RIG-I, a key innate-immune sensor of influenza (*Barber et al., 2010*).

For expansion of helper virus, we generated MDCK-SIAT1 cells expressing S009 PB2-E627K under control of a doxycycline-inducible promoter (MDCK-SIAT1-tet-S009-PB2-E627K) using a Sleeping Beauty transposon system (*Kowarz et al., 2015*). Briefly, MDCK-SIAT1 cells were transfected with pSBtet\_RP\_S009\_PB2\_E627K and pSB100X transposase vector using Lipofectamine 3000 (ThermoFisher Scientific, L3000015), and then subject to selection with 1 µg/ml puromycin. At three days post-transfection, we sorted for individual transfected cells expressing mCherry. All subsequent experiments were performed with a clonal expansion of a single transfected cell.

A549 cells were authenticated using the ATCC STR profiling service. CCL-141 cells were obtained from ATCC and used without extensive passaging. All cell lines tested negative for mycoplasma at the time they were expanded for either generating helper virus or passaging mutant plasmid libraries.

## Plasmids

Sequences for plasmids generated in this study are provided in *Supplementary file 1*.

Avian influenza polymerase plasmids: Original plasmids for PB2, PB1, PA, and NP genes from avian influenza strain A/Green-winged Teal/Ohio/175/1986 (S009) were gifts of Jeffrey Taubenberger (*Jagger et al., 2010*). For generating the mutant plasmid library, we cloned the S009 PB2 coding sequence into a pHW2000 vector (*Hoffmann et al., 2000*) from which we removed the CMV promoter (final plasmid pHW\_noCMV\_S009\_PB2). The mutant library insert was cloned into the recipient vector pHW\_noCMVnoTerm\_BsmBI, which lacks the Pol I terminator (terminator sequence is part of the insert). The reason that we generated a pHW plasmid without a CMV promoter is that we were unable to maintain a stable bacterial clone of the S009 PB2 coding sequence on the pHH21 plasmid backbone – we observed frequent deletions in the coding sequence during plasmid propagation, suggesting that the insert on the pHH21 plasmid backbone is toxic to the bacterial host. For generating helper virus and virus for viral competitions, we cloned the S009 PB2, PB1, PA, and NP coding sequences into pHW2000 (pHW\_S009\_PB2, pHW\_S009\_PB1, pHW\_S009\_PA, pHW\_S009\_NP). In all cases we used non-coding viral-RNA termini from the respective A/WSN/1933 (H1N1) gene segment. For protein expression and the minigenome assay, we cloned the PB2, PB1, PA, and NP coding sequences from S009 into a protein-expression plasmid with a CMV promoter (HDM\_S009\_PB2, HDM\_S009\_PB1, HDM\_S009\_PA, HDM\_S009\_NP). All mutants of PB2 were made by site-directed mutagenesis on the appropriate plasmid backbone.

Helper virus plasmids: To generate a PB2 vRNA lacking a functional PB2 protein, we cloned GFP flanked by PB2 sequence into the pHH21 vector (*Neumann et al., 1999*) (pHH\_PB2\_S009\_flank\_99\_eGFP\_100). The flanking non-coding viral-RNA termini are from WSN PB2, and the coding sequences are from S009 PB2. The length of flanking sequences, 99 and 100 bases on the 5' and 3' end of the PB2 coding sequence respectively, are based on prior experiments analyzing how much terminal sequence is needed for effective genome packaging (*Liang et al., 2005*). We mutated out start codons 5' to the GFP start site in the mRNA sense. Our helper virus rescue also required the reverse genetics plasmids encoding HA, NA, M, and NS from WSN (pHW184\_HA, pHW186\_NA, pHW187\_M, pHW188\_NS) (*Hoffmann et al., 2000*). To generate a cell line with doxycycline-inducible expression of S009 PB2 for expansion of PB2-deficient helper virus, we cloned the S009 PB2-E627K coding sequence into the pSBtet vector (pSBtet\_RP\_S009\_PB2\_E627K) (*Kowarz et al., 2015*).

Minigenome assay: In addition to protein expression plasmids described above, we used a pHH-PB1-flank-eGFP reporter (*Bloom et al., 2010*), and pcDNA3.1-mCherry as a transfection control.

## Primers

All primer sequences used in this study are provided in **Supplementary file 2**. Note that this Excel file has several worksheets giving primers for different aspects of the experiments.

## PB2 codon mutant plasmid libraries

We generated all possible codon mutations of the entire PB2 coding sequence using the PCR-based strategy described in *Bloom (2014)* with the modifications described in *Dingens et al. (2017)*. Briefly, we designed mutagenic primers tiling across the entire coding region (<https://github.com/jbloomlab/CodonTilingPrimers>; *Bloom and Dingens, 2019*; copy archived at <https://github.com/elifesciences-publications/CodonTilingPrimers>). We performed 10 cycles of fragment PCR using the mutagenic primers and end primers flanking the vRNA, followed by 20 cycles of joining PCR using only end primers (**Supplementary file 2**: Mutagenesis worksheet). We generated three independent libraries starting from mutagenesis of independent bacterial clones. The PB2 variants were cloned into the BsmBI-digested vector pHW\_noCMVnoTerm\_BsmBI using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S), and electroporated into ElectroMAX DH10B competent cells (Invitrogen, 18290015). We obtained 18–22 million transformants for each replicate library, from which we extracted plasmid by maxiprep. We randomly selected 48 clones for Sanger sequencing to evaluate the library mutation rate (<https://github.com/jbloomlab/SangerMutantLibraryAnalysis>; *Bloom et al., 2019*; copy archived at <https://github.com/elifesciences-publications/SangerMutantLibraryAnalysis>). (**Figure 1—figure supplement 2A–F**).

## Generation and passaging of mutant virus libraries

We generated mutant virus libraries using the helper-virus approach in *Doud and Bloom (2016)*, with modifications. We rescued reassortant virus using polymerase and nucleoprotein genes (PB2, PB1, PA, NP) from S009, and remaining genes (HA, NA, M, NS) from A/WSN/1933(H1N1) (WSN).

Helper virus: We plated a co-culture of  $4 \times 10^5$  HEK293T and  $0.5 \times 10^5$  MDCK-SIAT1-tet-S009-PB2-E627K in D10 media per well of a 6-well plate. At 18 hr after seeding cells, we added 1  $\mu\text{g}/\text{ml}$  doxycycline to induce PB2 expression. One hour after adding doxycycline, we transfected each well with 250 ng each of pHH\_PB2\_S009\_flank\_99\_eGFP\_100, HDM\_S009\_PB2-E627K, pHW\_S009\_PB1, pHW\_S009\_PA, pHW\_S009\_NP, pHW184\_HA, pHW186\_NA, pHW187\_M, and pHW188\_NS using BioT (Bioland Scientific, B01-01). At four hours post-transfection, we replaced D10 media with WGM supplemented with 1  $\mu\text{g}/\text{ml}$  doxycycline. We collected viral supernatant 52 hr post-transfection. To expand the helper virus, we seeded MDCK-SIAT1-tet-S009-PB2-E627K cells in D10 media 4 hr prior to infection at  $4 \times 10^6$  cells per 15 cm dish. We then infected each dish with 40  $\mu\text{l}$  of fresh viral supernatant, using WGM supplemented with 1  $\mu\text{g}/\text{ml}$  doxycycline to induce PB2 expression. At 48 hr post-infection, we collected viral supernatant containing expanded helper virus and clarified the supernatant by centrifugation at 400x g for 4 min. We measured the infectious particle (IP)/ $\mu\text{l}$  titer of the helper virus by infecting HEK293T cells with a known volume of viral supernatant, and quantifying the number of GFP+ cells by flow cytometry 18 hr post-infection.

Mutant virus library rescue: For each mutant plasmid library, we seeded 36 wells of a 6-well dish with  $1 \times 10^6$  HEK293T cells, and transfected each well 17 hr later with 375 ng each of HDM\_S009\_PB2-E627K, HDM\_S009\_PB1, HDM\_S009\_PA, HDM\_S009\_NP, and 500 ng of PB2 mutant plasmid library using BioT. For the wild-type control, we seeded six wells and used pHW\_noCMV\_S009\_PB2 in place of the mutant plasmid library. At 6 hr post-transfection, we infected cells with helper virus at MOI of 1 IP/cell in WGM. At 2 hr post-infection, we replaced the inoculum with fresh WGM. At 20 hr post-infection, we harvested viral supernatants and clarified the supernatant by centrifugation at 400x g for 4 min. Supernatants were titered by TCID<sub>50</sub> on MDCK-SIAT1 cells. The titers for the three library replicates and wild-type control were 262, 68, 100, and 1467 TCID<sub>50</sub>/ $\mu\text{l}$  respectively.

Passaging of mutant virus libraries: We aimed to passage  $1 \times 10^6$  TCID<sub>50</sub> of each mutant virus library in A549 and CCL-141 cells at MOI of 0.01 TCID<sub>50</sub>/cell as determined in MDCK-SIAT1 cells. Therefore, we aimed to have  $1 \times 10^8$  total cells each at the time of infection. The day prior to each infection, we seeded between  $8 \times 10^7$  to  $1 \times 10^8$  A549 cells in D10 in  $4 \times 5$  layered cell culture flasks (Corning, 353144), and  $8 \times 10^7$  to  $1 \times 10^8$  CCL-141 cells in E10 in  $8 \times 5$  layered cell culture flasks. To estimate the total number of each cell type at the time of each infection, we plated an

equivalent density of cells in a T225 flask. Just prior to infection, we counted the number of cells in the T225 flask, and extrapolated the total number of cells to be infected. We calculated, for an MOI of 0.01, the number of TCID<sub>50</sub>s to be passaged for each of the three library replicates and wild-type control in A549 cells to be  $9.39 \times 10^5$ ,  $1.02 \times 10^6$ ,  $1.20 \times 10^6$ , and  $9.30 \times 10^5$  respectively, and in CCL-141 cells to be  $8.46 \times 10^5$ ,  $8.92 \times 10^5$ ,  $9.54 \times 10^5$ , and  $8.58 \times 10^5$  respectively. We infected cells by removing D10 or E10 media, rinsing each flask with PBS, and then adding the calculated amount of virus diluted in WGM. At 3 hr post-infection, we replaced the inoculum with fresh WGM. The low MOI passage is expected to purge any non-replicative virus, including those containing the GFP segment. We confirmed this by our observation that there is limited spread of GFP expression over the course of infection. We harvested viral supernatant 48 hr post-infection, and clarified the supernatant by centrifugation at 400x g for 4 min.

## Barcoded subamplicon sequencing

We ultracentrifuged clarified viral supernatant at 27,000 rpm in a Beckman Coulter SW28 rotor, for 2 hr at 4°C. We resuspended the virus the residual media, then extracted RNA from 280 µl of concentrated virus using the Qiagen RNeasy Mini Kit (Qiagen, 74104). We titered the concentrated viral supernatant by TCID<sub>50</sub> on MDCK-SIAT1 cells to estimate the total TCID<sub>50</sub> from which we extracted RNA (ranged from  $2.80 \times 10^5$  to  $8.8 \times 10^6$  TCID<sub>50</sub>); we expect the TCID<sub>50</sub> titer to be a lower-bound and underestimate of the total viral variants present in the supernatant, since it measures only infectious virus, whereas we would extract RNA from both infectious and non-infectious virions.

We used a barcoded-subamplicon deep sequencing strategy that reduces the sequencing error rate (*Doud and Bloom, 2016*), [https://jbloomlab.github.io/dms\\_tools2/bcsubamp.html](https://jbloomlab.github.io/dms_tools2/bcsubamp.html)). Briefly, we reverse transcribed the full PB2 vRNA with Accuscript Reverse Transcriptase (Agilent, 200820) (primer S009-PB2-full-1F, *Supplementary file 2*: Barcoded subamplicon sequencing worksheet), and then PCR amplified the full PB2 vRNA (primers S009-PB2-full-1F and S009-PB2-full-8R, *Supplementary file 2*) using KOD Host Start Master Mix (EMD Millipore, 71842), making sure to have amplified from an estimated  $1 \times 10^7$  cDNA molecules. During this amplification, we observed that the band corresponding to full-length PB2 was generally more intense than the smaller bands likely corresponding to the PB2-GFP gene segment, as well as PB2 deletions, suggesting that full-length PB2 was the most prevalent (*Figure 1—figure supplement 3*). We then PCR amplified the PB2 gene in eight subamplicons using primers containing a random barcode to uniquely identify each template cDNA molecule (*Supplementary file 2*). Approximately  $7.5 \times 10^5$  uniquely barcoded molecules from each subamplicon library were then amplified by primers that add Illumina sequencing adaptors (*Supplementary file 2*). Finally, these libraries were deep sequenced on an Illumina HiSeq 2500 using  $2 \times 250$  bp paired-end reads to a target 3.3x coverage per barcode.

## Analysis of deep mutational scanning data

Deep mutational scanning sequence data was analyzed using *dms\_tools2* ([https://jbloomlab.github.io/dms\\_tools2](https://jbloomlab.github.io/dms_tools2), version 2.3.0). The GitHub repository <https://github.com/jbloomlab/PB2-DMS> (*Soh and Bloom, 2019*; copy archived at <https://github.com/elifesciences-publications/PB2-DMS>) contains Jupyter notebooks that perform all steps of the analyses and provide detailed step-by-step explanations and plots. The README file explains the organization of the notebooks and other files. HTML renderings of the notebooks are provided in *Supplementary file 3*. Processed results on preferences, differential selection, and mutation effect are provided in *Figure 2—source data 1*.

Rescaling of preferences: We rescaled our preferences to match the stringency of selection in nature. To do so, we first asked how well the preferences measured by the deep mutational scanning in either human or avian cells describes evolution of PB2 in both human and avian hosts. We used the preferences to generate an experimentally informed codon substitution model (ExpCM) (*Bloom, 2017*; *Hilton et al., 2017*), and asked if the ExpCMs described PB2's natural evolution better than a standard phylogenetic substitution model. The ExpCMs using amino-acid preferences vastly outperformed standard phylogenetic substitution models, suggesting that our experiments do capture some of the natural evolutionary constraint on PB2 (*Supplementary file 1*). The ExpCM stringency parameter had a value of 2.5, indicating that natural selection favors the same amino acids as our experiments, but with greater stringency (*Hilton et al., 2017*). We thus rescaled our

preferences to match the stringency of selection in nature using the ExpCM stringency parameter, and use these rescaled preferences for all subsequent analyses (**Figure 2—figure supplement 1, Supplementary file 1, Figure 2—source data 1**).

### Minigenome activity

Minigenome assays were performed in biological triplicate (starting from independent bacterial clones of each PB2 mutant) in both A549 and HEK293T cells. We seeded  $2.5 \times 10^4$  A549 or HEK293T cells per well of a 96-well plate. Cells were transfected the next day with 10 ng each of HDM\_S009\_PB2 (for the respective mutant), HDM\_S009\_PB1, HDM\_S009\_PA, HDM\_S009\_NP, 30 ng of pHH-PB1-flank-eGFP reporter, and 30 ng of pcDNA-mCherry as transfection control, using Lipofectamine 3000 (A549) or BioT (HEK293T). At 22 hr post-transfection, cells were trypsinized and analyzed by flow cytometry. We report minigenome activity as the percent of mCherry-positive cells that are GFP-positive.

### Viral competition

**Mutant virus:** We generated mutant virus by reverse genetics using pHW\_S009\_PB2 (for the respective mutant), pHW\_S009\_PB1, pHW\_S009\_PA, pHW\_S009\_NP, pHW184\_HA, pHW186\_NA, pHW187\_M, and pHW188\_NS. We seeded  $5 \times 10^5$  HEK293T cells per well of a 6-well plate, and transfected cells the next day with 250 ng each of the 8 pHW plasmids using BioT. At 2 hr post-transfection, we replaced media with WGM. At 40 hr post-infection, we collected viral supernatant. Viruses were titered by measuring the number of transcriptionally active particles as determined by flow cytometry for HA expression in infected A549 and CCL-141 cells. Briefly, we infected A549 and CCL-141 cells with a known volume of viral supernatant, and quantified the number of HA+ cells 9 hr post-infection. Cells were stained for HA using the H17-L19 antibody (**Gerhard et al., 1981**), which reacts with the WSN HA (**Doud et al., 2017**).

**Competitions:** Viral competition assays were performed in biological duplicate (starting from independent bacterial clones of each PB2 mutant). For each competition, A549 and CCL-141 cells were infected with a mixture of wild-type and PB2-mutant virus at a 1:1 ratio of transcriptionally active particles as measured for that cell type. For samples collected at 10 hr post infection, we infected a minimum of  $3.78 \times 10^5$  cells at MOI of 0.1. For samples collected at 48 hr post infection, we infected a minimum of  $7.62 \times 10^5$  cells at MOI of 0.01. At 2 hr post-infection, we replaced either D10 or E10 media with fresh WGM. At 10 hr post-infection, cells infected at MOI of 0.1 were lysed in buffer RLT and cellular RNA was extracted using the Qiagen RNeasy Mini Kit. At 48 hr post-infection, we collected viral supernatant from cells infected at MOI of 0.01, and extracted RNA using the QIAamp Viral RNA Mini Kit (Qiagen, 52904).

**Sequencing to determine mutant frequency:** We reverse transcribed the full length PB2 vRNA from the extracted RNA with SuperScript III (ThermoFisher Scientific, 18080051) (primer S009-PB2-full-1F, **Supplementary file 2: Viral competition**). For each PB2 mutant, we PCR amplified from the cDNA the region of PB2 centered around that mutated codon site (**Supplementary file 2**). This product was then subject to a second PCR using primers that add Illumina sequencing adaptors. Finally, these libraries were deep sequenced on an Illumina MiSeq using 50 bp single-end reads. Computational analyses to quantify mutant versus wild-type frequency are provided in **Supplementary file 3**, and at <https://github.com/jbloombloom/PB2-DMS>.

### H7N9 phylogenetic analysis

The phylogenetic tree was generated using Nextstrain's augur pipeline (**Hadfield et al., 2018**), and ancestral state reconstruction and adjustment of branch lengths according to sequence isolation date were performed with TreeTime (**Sagulenko et al., 2018**). Ancestral state reconstruction was only performed for nucleotide states, and was not used to infer ancestral host states. Instead, we inferred host-state transitions associated with each branch of the tree in a way that leveraged the prior knowledge that most H7N9 viruses circulate in avian hosts, and that most human infections arise from direct avian-to-human transmissions (**Su et al., 2017**). Specifically, for each node in the tree, starting from the root, we gathered all tips descending from that node. If that clade included only human sequences, and its parent node also included only human sequences, then the current clade falls within a monophyletic human clade, and the branch leading to it was labeled human-to-

human. If the current clade includes only human sequences but the parent node includes non-human sequences, then the branch leading to the clade was labeled avian-to-human. If the current clade includes both human and non-human sequences, then the branch leading to the clade was labeled avian-to-avian. Mutations were classified as human if they occurred on human-to-human and avian-to-human branches, and are classified as avian if they occurred on avian-to-avian branches. Note that since H7N9 human influenza typically results from avian-to-human transmissions, the branches we label as human-to-human in reality likely arise from avian-to-human transmissions, but which we cannot accurately reconstruct due to insufficient sampling of avian sequences. For this reason, human-to-human and avian-to-human branches were grouped together. Further details are provided in Jupyter notebook at <https://github.com/jbloomlab/PB2-DMS>, or in **Supplementary file 3**.

### Accessibility of mutations

We calculated the accessibility, or mean nucleotide substitutions required to access an amino-acid mutation, from avian influenza PB2 sequences collected from 2015 through 2018. The accessibility of codon *c* to amino-acid *a* by single-nucleotide mutations is defined as the minimum number of nucleotide mutations needed to generate any codon for that amino-acid. For a collection of sequences, we calculate the accessibility as the weighted average of the accessibilities of all codons observed at that site in the collection of sequences. Accessibility was calculated using code as documented here ([https://jbloomlab.github.io/dms\\_tools2/dms\\_tools2.utils.html?highlight=accessibility#dms\\_tools2.utils.codonEvolAccessibility](https://jbloomlab.github.io/dms_tools2/dms_tools2.utils.html?highlight=accessibility#dms_tools2.utils.codonEvolAccessibility)), in Jupyter notebook at <https://github.com/jbloomlab/PB2-DMS>, and in **Supplementary file 3**.

### Quantification and statistical analysis

Quantification and statistical analysis was performed in Python and a complete description is available in main text, methods, associated figure legends, and computational Jupyter notebooks.

### Data and software availability

Deep sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA511556.

The GitHub repository <https://github.com/jbloomlab/PB2-DMS> contains Jupyter notebooks that perform all steps of computational analyses and provide detailed step-by-step explanations and plots. The README file explains the organization of the notebooks and other files. HTML renderings of the notebooks are provided in **Supplementary file 3**.

### Acknowledgements

We thank William Fowler for help with molecular cloning, Juhye Lee, Adam Dingens, and John Hudleston for advice and computer code, and Harmit Malik, Katherine Xue, Allison Greaney, and Tyler Starr for providing comments on this manuscript. We thank Jeffrey Taubenberger for providing plasmids for A/Green-winged Teal/Ohio/175/1986 (S009), and Andrew Mehle and Steven Baker for advice about growing virus containing S009 polymerase and NP genes. YQSS was supported by the Mahan Postdoctoral Fellowship (Computational Biology Program of Fred Hutchinson Cancer Research Center) and the Damon Runyon Postdoctoral Fellowship. This work was supported by NIH grant R01 AI127893 from the NIAID to TB and JDB, and NIH grant R35 GM119774-01 from the NIGMS to TB. TB is a Pew Biomedical Scholar. JDB is an Investigator of the Howard Hughes Medical Institute.

---

### Additional information

#### Funding

| Funder  | Grant reference number | Author                          |
|---|------------------------|---------------------------------|
| National Institute of Allergy and Infectious Diseases | R01 AI127893           | Trevor Bedford<br>Jesse D Bloom |
| National Institute of General Medical Sciences        | R35 GM119774-01        | Trevor Bedford                  |

---

|   |             |                  |
|---|-------------|------------------|
| Damon Runyon Cancer Research Foundation | DRG-2271-16 | Y Q Shirleen Soh |
| Pew Charitable Trusts                   |             | Trevor Bedford   |
| Howard Hughes Medical Institute         |             | Jesse D Bloom    |
| Mahan Postdoctoral Fellowship           |             | Y Q Shirleen Soh |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

YQ Shirleen Soh, Conceptualization, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Louise H Moncla, Data curation, Software, Formal analysis, Methodology, Writing—review and editing; Rachel Eguia, Validation, Investigation, Writing—review and editing; Trevor Bedford, Data curation, Supervision, Funding acquisition, Methodology, Writing—review and editing; Jesse D Bloom, Conceptualization, Software, Supervision, Funding acquisition, Methodology, Writing—review and editing

### Author ORCIDs

YQ Shirleen Soh  <https://orcid.org/0000-0002-7422-7414>

Trevor Bedford  <http://orcid.org/0000-0002-4039-5794>

Jesse D Bloom  <http://orcid.org/0000-0003-1267-3408>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.45079.036>

Author response <https://doi.org/10.7554/eLife.45079.037>

## Additional files

### Supplementary files

- Supplementary file 1. Plasmid sequences.  
DOI: <https://doi.org/10.7554/eLife.45079.028>
- Supplementary file 2. Primer sequences.  
DOI: <https://doi.org/10.7554/eLife.45079.029>
- Supplementary file 3. Jupyter notebooks documenting computational analyses.  
DOI: <https://doi.org/10.7554/eLife.45079.030>
- Supplementary file 4. Comparison of ExpCM to standard phylogenetic substitution models.  
DOI: <https://doi.org/10.7554/eLife.45079.031>
- Transparent reporting form  
DOI: <https://doi.org/10.7554/eLife.45079.032>

### Data availability

Deep sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA511556. All data generated or analyzed during this study are included in the manuscript and supporting files. Source data files have been provided for Figures 2, 3, 4, 6, and 7. The GitHub repository <https://github.com/jbloombloom/PB2-DMS> contains Jupyter notebooks that perform all steps of computational analyses and provide detailed step-by-step explanations and plots.

The following dataset was generated:

| Author(s) | Year | Dataset title   | Dataset URL   | Database and Identifier      |
|-----------|------|---|---|------------------------------|
| Soh YQS   | 2019 | Deep mutational scanning of avian influenza PB2 to identify host- | <a href="https://www.ncbi.nlm.nih.gov/bioproject/">https://www.ncbi.nlm.nih.gov/bioproject/</a> | NCBI BioProject, PRJNA511556 |

## References

- Ashenberg O, Padmakumar J, Doud MB, Bloom JD. 2017. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLOS Pathogens* **13**:e1006288. DOI: <https://doi.org/10.1371/journal.ppat.1006288>
- Assarsson E, Bui HH, Sidney J, Zhang Q, Glenn J, Oseroff C, Mbawuiké IN, Alexander J, Newman MJ, Grey H, Sette A. 2008. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. *Journal of Virology* **82**:12241–12251. DOI: <https://doi.org/10.1128/JVI.01563-08>, PMID: 18842709
- Barber MR, Aldridge JR, Webster RG, Magor KE. 2010. Association of RIG-I with innate immunity of ducks to influenza. *PNAS* **107**:5913–5918. DOI: <https://doi.org/10.1073/pnas.1001755107>, PMID: 20308570
- Bloom JD, Gong LI, Baltimore D. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**:1272–1275. DOI: <https://doi.org/10.1126/science.1187816>, PMID: 20522774
- Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular Biology and Evolution* **31**:1956–1978. DOI: <https://doi.org/10.1093/molbev/msu173>, PMID: 24859245
- Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* **16**:1–13. DOI: <https://doi.org/10.1186/s12859-015-0590-4>, PMID: 25990960
- Bloom JD. 2017. Identification of positive selection in genes is greatly improved by using experimentally informed site-specific models. *Biology Direct* **12**:1. DOI: <https://doi.org/10.1186/s13062-016-0172-z>, PMID: 28095902
- Bloom JD, Doud M, Greaney AJ, Haddox H. 2019. SangerMutantLibraryAnalysis. a97bb61. Github. <https://github.com/jbloombiolab/SangerMutantLibraryAnalysis>
- Bloom JD, Dingens A. 2019. Tiling primers for codon mutagenesis. 24dd336. Github. <https://github.com/jbloombiolab/CodonTilingPrimers>
- Boucher JI, Cote P, Flynn J, Jiang L, Laban A, Mishra P, Roscoe BP, Bolon DN. 2014. Viewing protein fitness landscapes through a next-gen lens. *Genetics* **198**:461–471. DOI: <https://doi.org/10.1534/genetics.114.168351>, PMID: 25316787
- Bussey KA, Bousse TL, Desmet EA, Kim B, Takimoto T. 2010. PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. *Journal of Virology* **84**:4395–4406. DOI: <https://doi.org/10.1128/JVI.02642-09>, PMID: 20181719
- Cauldwell AV, Moncorgé O, Barclay WS. 2013. Unstable polymerase-nucleoprotein interaction is not responsible for avian influenza virus polymerase restriction in human cells. *Journal of Virology* **87**:1278–1284. DOI: <https://doi.org/10.1128/JVI.02597-12>, PMID: 23115299
- Cauldwell AV, Long JS, Moncorgé O, Barclay WS. 2014. Viral determinants of influenza A virus host range. *Journal of General Virology* **95**:1193–1210. DOI: <https://doi.org/10.1099/vir.0.062836-0>, PMID: 24584475
- Chen GW, Chang SC, Mok CK, Lo YL, Kung YN, Huang JH, Shih YH, Wang JY, Chiang C, Chen CJ, Shih SR. 2006. Genomic signatures of human versus avian influenza A viruses. *Emerging Infectious Diseases* **12**:1353–1360. DOI: <https://doi.org/10.3201/eid1209.060276>, PMID: 17073083
- Chin AWH, Li OTW, Mok CKP, Ng MKW, Peiris M, Poon LLM. 2014. Influenza A viruses with different amino acid residues at PB2-627 display distinct replication properties in vitro and in vivo: revealing the sequence plasticity of PB2-627 position. *Virology* **468-470**:545–555. DOI: <https://doi.org/10.1016/j.virol.2014.09.008>, PMID: 25262472
- Chin AWH, Leong NKC, Nicholls JM, Poon LLM. 2017. Characterization of influenza A viruses with polymorphism in PB2 residues 701 and 702. *Scientific Reports* **7**:1–12. DOI: <https://doi.org/10.1038/s41598-017-11625-y>
- Dingens AS, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive mapping of HIV-1 escape from a broadly neutralizing antibody. *Cell Host & Microbe* **21**:777–787. DOI: <https://doi.org/10.1016/j.chom.2017.05.003>, PMID: 28579254
- dos Reis M, Tamuri AU, Hay AJ, Goldstein RA. 2011. Charting the host adaptation of influenza viruses. *Molecular Biology and Evolution* **28**:1755–1767. DOI: <https://doi.org/10.1093/molbev/msq317>, PMID: 21109586
- Doud MB, Ashenberg O, Bloom JD. 2015. Site-Specific amino acid preferences are mostly conserved in two closely related protein homologs. *Molecular Biology and Evolution* **32**:2944–2960. DOI: <https://doi.org/10.1093/molbev/msv167>, PMID: 26226986
- Doud MB, Hensley SE, Bloom JD. 2017. Complete mapping of viral escape from neutralizing antibodies. *PLOS Pathogens* **13**:e1006271. DOI: <https://doi.org/10.1371/journal.ppat.1006271>, PMID: 28288189
- Doud MB, Bloom JD. 2016. Accurate measurement of the effects of all Amino-Acid mutations on influenza hemagglutinin. *Viruses* **8**:155. DOI: <https://doi.org/10.3390/v8060155>, PMID: 27271655
- Du Y, Xin L, Shi Y, Zhang TH, Wu NC, Dai L, Gong D, Brar G, Shu S, Luo J, Reiley W, Tseng YW, Bai H, Wu TT, Wang J, Shu Y, Sun R. 2018. Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. *Science* **359**:290–296. DOI: <https://doi.org/10.1126/science.aan8806>, PMID: 29348231
- Eisfeld AJ, Neumann G, Kawaoka Y. 2015. At the centre: influenza A virus ribonucleoproteins. *Nature Reviews Microbiology* **13**:28–41. DOI: <https://doi.org/10.1038/nrmicro3367>, PMID: 25417656



- Finkelstein DB**, Mukatira S, Mehta PK, Obenauer JC, Su X, Webster RG, Naeve CW. 2007. Persistent host markers in pandemic and H5N1 influenza viruses. *Journal of Virology* **81**:10292–10299. DOI: <https://doi.org/10.1128/JVI.00921-07>, PMID: 17652405
- Fowler DM**, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nature Methods* **11**:801–807. DOI: <https://doi.org/10.1038/nmeth.3027>, PMID: 25075907
- Fragata I**, Matuszewski S, Schmitz MA, Bataillon T, Jensen JD, Bank C. 2018. The fitness landscape of the codon space across environments. *Heredity* **121**:422–437. DOI: <https://doi.org/10.1038/s41437-018-0125-7>, PMID: 30127529
- Gabriel G**, Dauber B, Wolff T, Planz O, Klenk HD, Stech J. 2005. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *PNAS* **102**:18590–18595. DOI: <https://doi.org/10.1073/pnas.0507415102>, PMID: 16339318
- Gabriel G**, Herwig A, Klenk HD. 2008. Interaction of polymerase subunit PB2 and NP with importin alpha1 is a determinant of host range of influenza A virus. *PLOS Pathogens* **4**:e11. DOI: <https://doi.org/10.1371/journal.ppat.0040011>, PMID: 18248089
- Gabriel G**, Klingel K, Otte A, Thiele S, Hudjetz B, Arman-Kalcek G, Sauter M, Schmidt T, Rother F, Baumgarte S, Keiner B, Hartmann E, Bader M, Brownlee GG, Fodor E, Klenk HD. 2011. Differential use of importin- $\alpha$  isoforms governs cell tropism and host adaptation of influenza virus. *Nature Communications* **2**:156. DOI: <https://doi.org/10.1038/ncomms1158>, PMID: 21245837
- Geoghegan JL**, Senior AM, Holmes EC. 2016. Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. *Proceedings of the Royal Society B: Biological Sciences* **283**:20160727. DOI: <https://doi.org/10.1098/rspb.2016.0727>, PMID: 27581875
- Geoghegan JL**, Holmes EC. 2017. Predicting virus emergence amid evolutionary noise. *Open Biology* **7**:170189. DOI: <https://doi.org/10.1098/rsob.170189>, PMID: 29070612
- Gerhard W**, Yewdell J, Frankel ME, Webster R. 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* **290**:713–717. DOI: <https://doi.org/10.1038/290713a0>, PMID: 6163993
- Graef KM**, Vreede FT, Lau YF, McCall AW, Carr SM, Subbarao K, Fodor E. 2010. The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *Journal of Virology* **84**:8433–8445. DOI: <https://doi.org/10.1128/JVI.00879-10>, PMID: 20538852
- Grubaugh ND**, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, Andersen KG. 2019. Tracking virus outbreaks in the twenty-first century. *Nature Microbiology* **4**:10–19. DOI: <https://doi.org/10.1038/s41564-018-0296-2>, PMID: 30546099
- Guilligay D**, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ, Cusack S. 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology* **15**:500–506. DOI: <https://doi.org/10.1038/nsmb.1421>, PMID: 18454157
- Hadfield J**, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**:4121–4123. DOI: <https://doi.org/10.1093/bioinformatics/bty407>, PMID: 29790939
- Hengrung N**, El Omari K, Serna Martin I, Vreede FT, Cusack S, Rambo RP, Vonnrhein C, Bricogne G, Stuart DI, Grimes JM, Fodor E. 2015. Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527**:114–117. DOI: <https://doi.org/10.1038/nature15525>, PMID: 26503046
- Hilton SK**, Doud MB, Bloom JD. 2017. phydms: software for phylogenetic analyses informed by deep mutational scanning. *PeerJ* **5**:e3657. DOI: <https://doi.org/10.7717/peerj.3657>, PMID: 28785526
- Hoffmann E**, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. *PNAS* **97**:6108–6113. DOI: <https://doi.org/10.1073/pnas.100133697>, PMID: 10801978
- Hu YJ**, Tu PC, Lin CS, Guo ST. 2014. Identification and chronological analysis of genomic signatures in influenza A viruses. *PLOS ONE* **9**:e84638. DOI: <https://doi.org/10.1371/journal.pone.0084638>, PMID: 24416256
- Hu M**, Yuan S, Zhang K, Singh K, Ma Q, Zhou J, Chu H, Zheng BJ. 2017. PB2 substitutions V598T/I increase the virulence of H7N9 influenza A virus in mammals. *Virology* **501**:92–101. DOI: <https://doi.org/10.1016/j.virol.2016.11.008>, PMID: 27889648
- Jagger BW**, Memoli MJ, Sheng ZM, Qi L, Hrabal RJ, Allen GL, Dugan VG, Wang R, Digard P, Kash JC, Taubenberger JK. 2010. The PB2-E627K mutation attenuates viruses containing the 2009 H1N1 influenza pandemic polymerase. *mBio* **1**:e00067-10. DOI: <https://doi.org/10.1128/mBio.00067-10>, PMID: 20689744
- Jiang L**, Liu P, Bank C, Renzette N, Prachanronarong K, Yilmaz LS, Caffrey DR, Zeldovich KB, Schiffer CA, Kowalik TF, Jensen JD, Finberg RW, Wang JP, Bolon DNA. 2016. A balance between inhibitor binding and substrate processing confers influenza drug resistance. *Journal of Molecular Biology* **428**:538–553. DOI: <https://doi.org/10.1016/j.jmb.2015.11.027>, PMID: 26656922
- Kelley LA**, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**:845–858. DOI: <https://doi.org/10.1038/nprot.2015.053>, PMID: 25950237
- Kim JH**, Hatta M, Watanabe S, Neumann G, Watanabe T, Kawaoka Y. 2010. Role of host-specific amino acids in the pathogenicity of avian H5N1 influenza viruses in mice. *Journal of General Virology* **91**:1284–1289. DOI: <https://doi.org/10.1099/vir.0.018143-0>, PMID: 20016035

- Kowarz E**, Löscher D, Marschalek R. 2015. Optimized sleeping beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology Journal* **10**:647–653. DOI: <https://doi.org/10.1002/biot.201400821>, PMID: 25650551
- Lee JM**, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *PNAS* **115**:E8276–E8285. DOI: <https://doi.org/10.1073/pnas.1806133115>, PMID: 30104379
- Liang Y**, Hong Y, Parslow TG. 2005. Cis -Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments cis -Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. *Journal of Virology* **79**:10348–10355.
- Lipsitch M**, Barclay W, Raman R, Russell CJ, Belsler JA, Cobey S, Kasson PM, Lloyd-Smith JO, Maurer-Stroh S, Riley S, Beauchemin CA, Bedford T, Friedrich TC, Handel A, Herfst S, Murcia PR, Roche B, Wilke CO, Russell CA. 2016. Viral factors in influenza pandemic risk assessment. *eLife* **5**:e18491. DOI: <https://doi.org/10.7554/eLife.18491>, PMID: 27834632
- Long JS**, Howard WA, Núñez A, Moncorgé O, Lycett S, Banks J, Barclay WS. 2013. The effect of the PB2 mutation 627K on highly pathogenic H5N1 avian influenza virus is dependent on the virus lineage. *Journal of Virology* **87**:9983–9996. DOI: <https://doi.org/10.1128/JVI.01399-13>, PMID: 23843645
- Long JS**, Giotis ES, Moncorgé O, Frise R, Mistry B, James J, Morisson M, Iqbal M, Vignal A, Skinner MA, Barclay WS. 2016. Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* **529**: 101–104. DOI: <https://doi.org/10.1038/nature16474>, PMID: 26738596
- Long JS**, Mistry B, Haslam SM, Barclay WS. 2019. Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology* **17**:67–81. DOI: <https://doi.org/10.1038/s41579-018-0115-z>, PMID: 30487536
- Mänz B**, Brunotte L, Reuther P, Schwemmler M. 2012. Adaptive mutations in NEP compensate for defective H5N1 RNA replication in cultured human cells. *Nature Communications* **3**:802. DOI: <https://doi.org/10.1038/ncomms1804>, PMID: 22549831
- Mänz B**, de Graaf M, Mögling R, Richard M, Bestebroer TM, Rimmelzwaan GF, Fouchier RAM. 2016. Multiple natural substitutions in avian influenza A virus PB2 facilitate efficient replication in human cells. *Journal of Virology* **90**:5928–5938. DOI: <https://doi.org/10.1128/JVI.00130-16>, PMID: 27076644
- Massin P**, van der Werf S, Naffakh N. 2001. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *Journal of Virology* **75**:5398–5404. DOI: <https://doi.org/10.1128/JVI.75.11.5398-5404.2001>, PMID: 11333924
- Mehle A**, Doudna JA. 2008. An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host & Microbe* **4**:111–122. DOI: <https://doi.org/10.1016/j.chom.2008.06.007>, PMID: 18692771
- Mehle A**, Doudna JA. 2009. Adaptive strategies of the influenza virus polymerase for replication in humans. *PNAS* **106**:21312–21316. DOI: <https://doi.org/10.1073/pnas.0911915106>, PMID: 19995968
- Miotto O**, Heiny A, Tan TW, August JT, Brusica V. 2008. Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics* **9**:S18. DOI: <https://doi.org/10.1186/1471-2105-9-S1-S18>, PMID: 18315849
- Mok CK**, Yen HL, Yu MY, Yuen KM, Sia SF, Chan MC, Qin G, Tu WW, Peiris JS. 2011. Amino acid residues 253 and 591 of the PB2 protein of avian influenza virus A H9N2 contribute to mammalian pathogenesis. *Journal of Virology* **85**:9641–9645. DOI: <https://doi.org/10.1128/JVI.00702-11>, PMID: 21734052
- Moncla LH**, Zhong G, Nelson CW, Dinis JM, Mutschler J, Hughes AL, Watanabe T, Kawaoka Y, Friedrich TC. 2016. Selective bottlenecks shape evolutionary pathways taken during mammalian adaptation of a 1918-like avian influenza virus. *Cell Host & Microbe* **19**:169–180. DOI: <https://doi.org/10.1016/j.chom.2016.01.011>, PMID: 26867176
- Naffakh N**, Massin P, Escriou N, Crescenzo-Chaigne B, van der Werf S. 2000. Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *Journal of General Virology* **81**:1283–1291. DOI: <https://doi.org/10.1099/0022-1317-81-5-1283>, PMID: 10769071
- Neumann G**, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, Hughes M, Perez DR, Donis R, Hoffmann E, Hobom G, Kawaoka Y. 1999. Generation of influenza A viruses entirely from cloned cDNAs. *PNAS* **96**:9345–9350. DOI: <https://doi.org/10.1073/pnas.96.16.9345>, PMID: 10430945
- Olival KJ**, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**:646–650. DOI: <https://doi.org/10.1038/nature22975>, PMID: 28636590
- Parrish CR**, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* **72**:457–470. DOI: <https://doi.org/10.1128/MMBR.00004-08>, PMID: 18772285
- Peck KM**, Lauring AS. 2018. Complexities of viral mutation rates. *Journal of Virology* **92**:e01031-17. DOI: <https://doi.org/10.1128/JVI.01031-17>, PMID: 29720522
- Pflug A**, Lukarska M, Resa-Infante P, Reich S, Cusack S. 2017. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Research* **234**:103–117. DOI: <https://doi.org/10.1016/j.virusres.2017.01.013>, PMID: 28115197
- Pumroy RA**, Ke S, Hart DJ, Zachariae U, Cingolani G. 2015. Molecular determinants for nuclear import of influenza A PB2 by importin  $\alpha$  isoforms 3 and 7. *Structure* **23**:374–384. DOI: <https://doi.org/10.1016/j.str.2014.11.015>, PMID: 25599645

- Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**:361–366. DOI: <https://doi.org/10.1038/nature14009>, PMID: 25409151
- Reperant LA, Kuiken T, Osterhaus AD. 2012. Adaptive pathways of zoonotic influenza viruses: from exposure to establishment in humans. *Vaccine* **30**:4419–4434. DOI: <https://doi.org/10.1016/j.vaccine.2012.04.049>, PMID: 22537992
- Resa-Infante P, Jorba N, Zamarreño N, Fernández Y, Juárez S, Ortín J. 2008. The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication. *PLOS ONE* **3**:e3904. DOI: <https://doi.org/10.1371/journal.pone.0003904>, PMID: 19066626
- Resa-Infante P, Gabriel G. 2013. The nuclear import machinery is a determinant of influenza virus host adaptation. *BioEssays* **35**:23–27. DOI: <https://doi.org/10.1002/bies.201200138>, PMID: 23239226
- Russell CA, Kasson PM, Donis RO, Riley S, Dunbar J, Rambaut A, Asher J, Burke S, Davis CT, Garten RJ, Gnanakaran S, Hay SI, Herfst S, Lewis NS, Lloyd-Smith JO, Macken CA, Maurer-Stroh S, Neuhaus E, Parrish CR, Pepin KM, et al. 2014. Improving pandemic influenza risk assessment. *eLife* **3**:e03883. DOI: <https://doi.org/10.7554/eLife.03883>, PMID: 25321142
- Sagulenko P, Puller V, Neher RA. 2018. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evolution* **4**. DOI: <https://doi.org/10.1093/ve/vex042>, PMID: 29340210
- Serna Martin I, Hengrung N, Renner M, Sharps J, Martínez-Alonso M, Masiulis S, Grimes JM, Fodor E. 2018. A mechanism for the activation of the influenza virus transcriptase. *Molecular Cell* **70**:1101–1110. DOI: <https://doi.org/10.1016/j.molcel.2018.05.011>, PMID: 29910112
- Soh YQ, Bloom JD. 2019. Deep mutational scanning of avian influenza PB2 to comprehensively map adaptation to the human host. 6a9f8a6. Github. <https://github.com/jbloomlab/PB2-DMS>
- Su S, Gu M, Liu D, Cui J, Gao GF, Zhou J, Liu X. 2017. Epidemiology, evolution, and pathogenesis of H7N9 influenza viruses in five epidemic waves since 2013 in China. *Trends in Microbiology* **25**:713–728. DOI: <https://doi.org/10.1016/j.tim.2017.06.008>, PMID: 28734617
- Subbarao EK, London W, Murphy BR. 1993. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of Virology* **67**:1761–1764.
- Taft AS, Ozawa M, Fitch A, Depasse JV, Halfmann PJ, Hill-Batorski L, Hatta M, Friedrich TC, Lopes TJ, Maher EA, Ghedin E, Macken CA, Neumann G, Kawaoka Y. 2015. Identification of mammalian-adapting mutations in the polymerase complex of an avian H5N1 influenza virus. *Nature Communications* **6**:7491. DOI: <https://doi.org/10.1038/ncomms8491>, PMID: 26082035
- Tamuri AU, Dos Reis M, Hay AJ, Goldstein RA. 2009. Identifying changes in selective constraints: host shifts in influenza. *PLOS Computational Biology* **5**:e1000564. DOI: <https://doi.org/10.1371/journal.pcbi.1000564>, PMID: 19911053
- Tarandeu F, Boudet J, Guilligay D, Mas PJ, Bougault CM, Boulo S, Baudin F, Ruigrok RW, Daigle N, Ellenberg J, Cusack S, Simorre JP, Hart DJ. 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Structural & Molecular Biology* **14**:229–233. DOI: <https://doi.org/10.1038/nsmb1212>, PMID: 17310249
- Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. 2005. Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**:889–893. DOI: <https://doi.org/10.1038/nature04230>, PMID: 16208372
- Taubenberger JK, Kash JC. 2010. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host & Microbe* **7**:440–451. DOI: <https://doi.org/10.1016/j.chom.2010.05.009>, PMID: 20542248
- Te Velthuis AJW, Long JC, Bauer DLV, Fan RLY, Yen HL, Sharps J, Siegers JY, Killip MJ, French H, Oliva-Martín MJ, Randall RE, de Wit E, van Riel D, Poon LLM, Fodor E. 2018. Mini viral RNAs act as innate immune agonists during influenza virus infection. *Nature Microbiology* **3**:1234–1242. DOI: <https://doi.org/10.1038/s41564-018-0240-5>, PMID: 30224800
- te Velthuis AJ, Fodor E. 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology* **14**:479–493. DOI: <https://doi.org/10.1038/nrmicro.2016.87>, PMID: 27396566
- Tong S, Li Y, Rivaller P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA, Turmelle AS, Moran D, Rogers S, Shi M, Tao Y, Weil MR, Tang K, Rowe LA, Sammons S, Xu X, et al. 2012. A distinct lineage of influenza A virus from bats. *PNAS* **109**:4269–4274. DOI: <https://doi.org/10.1073/pnas.1116200109>, PMID: 22371588
- Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, Chen LM, Johnson A, Tao Y, Dreyfus C, Yu W, McBride R, Carney PJ, Gilbert AT, Chang J, Guo Z, et al. 2013. New world bats harbor diverse influenza A viruses. *PLOS Pathogens* **9**:e1003657. DOI: <https://doi.org/10.1371/journal.ppat.1003657>, PMID: 24130481
- Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, tenOever BR. 2014. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host & Microbe* **16**:691–700. DOI: <https://doi.org/10.1016/j.chom.2014.09.020>, PMID: 25456074
- Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiological Reviews* **56**:152–179. PMID: 1579108
- Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen SH, Lu IH, Lin CY, Chin RG, Luan HH, Nguyen N, Nelson SF, Li X, Wu TT, Sun R. 2014a. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Scientific Reports* **4**:4942. DOI: <https://doi.org/10.1038/srep04942>, PMID: 24820965

- Wu NC**, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Luan HH, Li X, Wu TT, Sun R. 2014b. High-throughput identification of loss-of-function mutations for anti-interferon activity in the influenza A virus NS segment. *Journal of Virology* **88**:10157–10164. DOI: <https://doi.org/10.1128/JVI.01494-14>, PMID: 24965464
- Xue J**, Chambers BS, Hensley SE, López CB. 2016. Propagation and characterization of influenza virus stocks that lack high levels of defective viral genomes and hemagglutinin mutations. *Frontiers in Microbiology* **7**:326. DOI: <https://doi.org/10.3389/fmicb.2016.00326>, PMID: 27047455
- Yamada S**, Hatta M, Staker BL, Watanabe S, Imai M, Shinya K, Sakai-Tagawa Y, Ito M, Ozawa M, Watanabe T, Sakabe S, Li C, Kim JH, Myler PJ, Phan I, Raymond A, Smith E, Stacy R, Nidom CA, Lank SM, et al. 2010. Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLOS Pathogens* **6**: e1001034. DOI: <https://doi.org/10.1371/journal.ppat.1001034>, PMID: 20700447
- Zhou B**, Li Y, Halpin R, Hine E, Spiro DJ, Wentworth DE. 2011. PB2 residue 158 is a pathogenic determinant of pandemic H1N1 and H5 influenza A viruses in mice. *Journal of Virology* **85**:357–365. DOI: <https://doi.org/10.1128/JVI.01694-10>, PMID: 20962098