Parallel evolution between genomic segments of seasonal human influenza viruses reveals RNA-RNA relationships

Jennifer E. Jones[1,2,3], Valerie Le Sage[1,3], Gabriella H. Padovani[1,3], Michael Calderon[4], Erik S. Wright[2,5*], Seema S. Lakdawala[1,2,3*]

[1]Department of Microbiology & Molecular Genetics, University of Pittsburgh, Pittsburgh, PA

[2]Center for Evolutionary Biology and Medicine, University of Pittsburgh, Pittsburgh, PA

[3]Center for Vaccine Research, University of Pittsburgh School of Medicine, Pittsburgh, PA

[4]Department of Cell Biology, Center for Biologic Imaging, University of Pittsburgh, Pittsburgh, PA

[5]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA

* Send all correspondence to lakdawala@pitt.edu and eswright@pitt.edu

Abstract

The influenza A virus (IAV) genome consists of eight negative-sense viral RNA (vRNA) segments that are selectively assembled into progeny virus particles through RNA-RNA interactions. To identify relationships between vRNA segments, we examined parallel evolution between vRNA segments of seasonal human IAV, finding that evolutionary relationships between vRNA segments differ between subtypes and antigenically-shifted strains. Intersegmental relationships were distinct between H3N2 and H1N1 viruses, but largely conserved over time in H3N2 viruses. However, parallel evolution of vRNA segments diverged between H1N1 strains isolated before and after the 2009 pandemic. Surprisingly, intersegmental relationships were not driven solely by protein sequence, which is potentially indicative of RNA-RNA driven coevolution. Colocalization of highly coevolved vRNA segments was enriched over other pairs at the nuclear

periphery during a productive viral infection. This study illustrates how phylogenetics can be applied to interrogate putative RNA interactions underlying selective assembly of IAV.

Introduction

Genetic variation is ubiquitious in RNA viruses. The rapid evolution underlying this variation can occur as a result of mutation, recombination, or reassortment, with major consequences for human disease (Andino & Domingo, 2015). In the case of influenza virus, these consequences include poor vaccine efficacy rates, immune escape, antiviral resistance, and the emergence of novel strains (Lyons & Lauring, 2018). Within the past century, influenza A virus (IAV) pandemics occurred in 1918 (H1N1), 1957 (H2N2), 1968 (H3N2), and 2009 (H1N1) (Neumann, Noda, & Kawaoka, 2009; Paules & Subbarao, 2017; Short, Kedzierska, & van de Sandt, 2018). Each of the last three influenza pandemics have been attributed to a reassortant strain composed of a novel combination of the eight viral RNA (vRNA) segments of the influenza virus genome (Neumann et al., 2009). Thus, the emergence of pandemic strains is marked by a concommittent alteration in the influenza virus genome triggered by new genetic diversity.

Public health measures to limit the impact of influenza virus outbreaks would benefit from the ability to predict reassortment between circulating influenza viruses. Genetic mutation is driven by stochastic processes and is therefore difficult to predict (Andino & Domingo, 2015). In contrast, reassortment is restricted by a number of factors. As reassortment must occur between two strains coinfecting the same cell, the spatiotemporal dynamics of coinfection as well as compatibility of RNA packaging signals can impede reassortment (Lowen, 2017; Richard, Herfst, Tao, Jacobs, & Lowen, 2018). One method for identifying potential reassortant viruses is to examine intracellular assembly of vRNA segments through biochemical techniques. Intersegmental RNA-RNA interactions have been proposed to facilitate selective assembly and packaging of eight unique vRNA segments into a progeny virion and could pose a significant hurdle to reassortment (Gavazzi et al., 2013; Noda et al., 2006). It is consequently imperative to

identify the evolutionary constraints imposed by intersegmental vRNA interactions, as this may help predict future influenza pandemics.

Epistasis within and between genes imposes evolutionary constraints that can be shaped by a number of factors, including the function, stability, or interactions between individual RNA or protein (Sardi & Gasch, 2018). Probabilistic models have revealed several destabilizing mutations in the influenza virus nucleoprotein (NP) that became fixed as a result of counterbalancing epistasis that improves NP protein stability (Gong, Suchard, & Bloom, 2013). These destabilizing mutations occur within T cell epitopes of NP that may be important for immune escape (Gong et al., 2013). Stabilizing epistasis was similarly instrumental to the emergence of oseltamivir-resistance mutations in the influenza neuraminidase (NA) (Bloom, Gong, & Baltimore, 2010). The rise of oseltamivir-resistance mutations in NA spurred investigation of parallel evolution between NA and hemagglutinin (HA), demonstrating that mutations in HA may have facilitated acquisition of oseltamivir-resistance mutations in NA (Jang & Bae, 2018; Kryazhimskiy, Dushoff, Bazykin, & Plotkin, 2011; Neverov, Kryazhimskiy, Plotkin, & Bazykin, 2015). These approaches have great potential, yet the current focus surrounds constraints on protein interactions (Escalera-Zamudio et al., 2020). Such methodologies could be further employed to investigate epistasis arising from RNA-RNA interactions. Mounting evidence from our group and others suggests that direct intermolecular interactions between vRNA segments coordinate selective assembly (Dadonaite et al., 2019; Le Sage et al., 2020). Genomic assembly contributes to heterogeneity in progeny viruses and could determine the fitness of reassortant strains after coinfection (Brooke, 2017; Lowen, 2017). Therefore, parallel evolution between vRNA segments arising from RNA interactions could reveal epistatic constraints on genetic reassortment.

In this study, we set out to combine phylogenetics and molecular biology to examine parallel evolution across vRNA segments genome-wide in seasonal human influenza viruses and identify potential epistatic relationships between vRNA segments. Unlike previous studies, our objective was to identify vRNA segments that might play key roles in genomic assembly. To evaluate

77 phylogenetic relationships among vRNA segments we relied upon the Robinson-Foulds distance
78 (*d*), a measure of topological distance between trees (Robinson, 1981). This method determines
79 the number of branch partitions that are not shared between two trees (Robinson, 1981) and is
80 therefore a quantitative measure of the topological distance between phylogenies. Higher values
81 of *d* correspond to greater topological distance, with a *d* value of 0 indicating that two trees are
82 topologically equivalent. We hypothesized that *d* would vary in accordance with the degree of
83 parallel evolution between genome segments, and sought to determine whether any observed
84 parallel evolution arose from RNA interactions.
85
86 **Results**
87 **Phylogenetic relationships between vRNA segments are not uniform in H3N2 viruses.**
88 Influenza A virus H1N1 and H3N2 subtypes have cocirculated in human populations since
89 1977 (Neumann et al., 2009). In order to explore vRNA segment relationships in seasonal human
90 IAV strains over time, we examined parallel evolution between vRNA segments in viruses
91 representative of each subtype. We used four sets of seasonal IAV strains: two sets of H3N2
92 viruses from 1995-2004 and 2005-2014, and two sets of H1N1 viruses from 2000-2008 and 2010-
93 2018 (**Table 1**). Bracketing the H3N2 viruses into two time intervals permitted investigation of the
94 conservation of vRNA relationships over time in antigenically drifted H3N2 viruses. We took a
95 similar approach with H1N1 viruses, bracketing instead on the antigenic shift event in 2009 with
96 the emergence of the pandemic swine-origin H1N1 virus. Comparison of vRNA relationships in
97 pre-pandemic (2000-2008) and post-pandemic (2010-2018) H1N1 viruses could reveal vRNA
98 relationships from viruses of two distinct lineages or, alternatively, uncover vRNA relationships
99 that remain conserved despite the potential for swapping of vRNA segments across multiple host
100 species.

Subtype	Time Period	Total Strains	Clusters with >97% identity	Final Clusters
H3N2	1995-2004	1,026	16	12
	2005-2014	3,879	17	12
H1N1	2000-2008	821	11	9
	2010-2018	4,072	14	9

Table 1. Influenza A virus strain datasets. Human H1N1 or H3N2 virus sequences for which full-length sequences are available (Influenza Research Database). Representative sequences were selected for further analysis by clustering. 'Final Clusters' indicates the number of clusters after small clusters were collapsed or omitted.

Our approach, outlined in **Figure 1**, examines evolutionary relationships between vRNA segments. We began our investigation with all H3N2 viruses for which full-length sequence information was available in the Influenza Research Database, yielding 1,026 H3N2 viruses from 1995-2004 and 3,879 H3N2 viruses from 2005-2014 (**Table 1**). However, reconstructing phylogenetic trees from all available sequences was disadvantageous, as a preliminary analysis of three-hundred sequences suggested that a great deal of phylogenetic variation could not be statistically supported by bootstrapping (branch support < 70). Instead, we used a clustering approach to select representative strains that would produce more statistically robust trees. We first concatenated sequences from all strains into full-length genomes from which we built alignments (**Figure 1A**) and clustered into operational taxonomic units on a neighbor-joining species tree (**Figure 1B**). Despite the fact that fewer full-length influenza virus genomic sequences were available prior to the 2000s, our approach resulted in a similar number of clusters within a subtype (**Table 1**), consistent with the notion that increased sequencing has led to more closely related sequences in public databases.

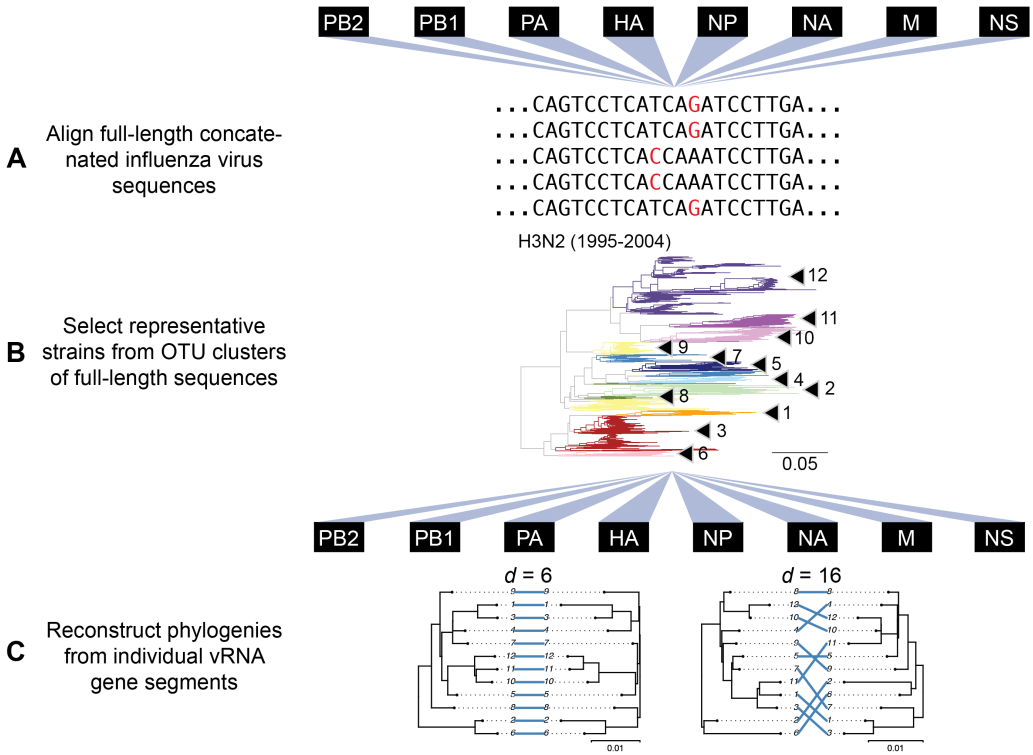


Figure 1. Experimental overview. A, Human H3N2 or H1N1 virus sequences were downloaded from the Influenza Research Database and subset into two time periods each: 1995-2004 and 2005-2014 (H3N2 viruses); 2000-2008 and 2010-2018 (H1N1 viruses). The H3N2 virus dataset (1995-2004) is illustrated here. All eight vRNA segments from each strain were concatenated into a full-length genome from which alignments were made. **B**, A species tree was built grouping strains into operational taxonomic units (OTU clusters) with at least 97% sequence identity. Arrowheads denote clusters 1-12. Seven replicate strains were randomly selected from these clusters for further analysis. **C**, Full-length genomic sequences were partitioned into individual gene sequence alignments and maximum-likelihood phylogenetic trees were reconstructed from each vRNA gene segment in each replicate. The Robinson-Foulds distances (*d*) and tanglegrams were examined from each pair of phylogenies. Left, a pair of highly congruent trees with a low *d* value. Right, a pair of discordant trees with a high *d* value. Scale bars indicate percent divergence.

The primary objective behind clustering was to reduce variation between trees that was not statistically supported by bootstrapping. The cutoff for sequence identity during clustering of the species tree was therefore an important consideration because it controlled how much variation remained in our trees. As expected, higher cutoffs (98-99% sequence identity) yielded species trees with more clusters containing fewer members while lower cutoffs (95-96% sequence identity) contained increasingly fewer clusters with more members grouped in each cluster. We selected a cutoff of 97% sequence identity based on the observation that it produced vRNA trees with an intermediate number of clusters (16-17 clusters in each species tree) but enough members in each cluster to provide multiple representative strains for comparison. This process resulted in seven replicate trees for each of the eight vRNA segments, for a total of 56 trees analyzed from each set of H3N2 viruses (**Figure 1C** and **Supplemental Figure 1**).

Concordance between phylogenetic trees built from different RNA segments is expected to be highest when there are strong positive epistatic interactions between encoded protein or RNA complexes (Kryazhimskiy et al., 2011; Neverov et al., 2015). The PB1 and PA proteins are subunits of the heterotrimeric polymerase complex and would be expected to exhibit positive epistasis (Fodor, 2013), whereas PB1 and HA do not share any known protein function. Trees built from one replicate of the PB1 and PA phylogenies from H3N2 (2005-2014) had a low *d* value of 6, suggesting parallel evolution occurs between these two genes (**Figure 2A**). By comparison, the PB1 and HA phylogenies from the same replicate had a *d* value of 14 (**Figure 2B**), suggesting that parallel evolution between PB1 and PA is stronger than between PB1 and HA. Thus, tree concordance recapitulates anticipated protein-driven parallel evolution between two influenza proteins.

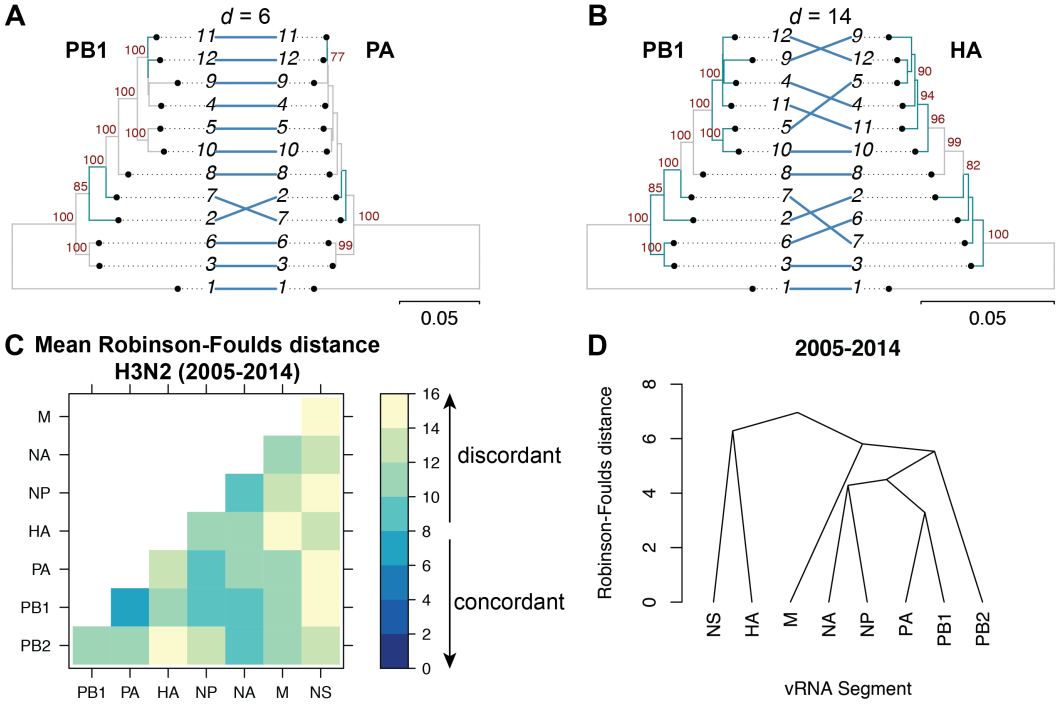


Figure 2. Coevolution between vRNA segments is hierarchical in H3N2 viruses from 2005-2014. Seven replicate maximum-likelihood phylogenies were reconstructed for each vRNA gene segment from human H3N2 virus sequences (2005-2014) as described in Figure 1. **A-B,** Tanglegrams representative of concordant (PB1 and PA gene segments) **(A)** or discordant (PB1 and HA gene segments) phylogenies **(B)** are shown from replicate 1 with discordant branches highlighted in turquoise. Robinson-Foulds distances (*d*) are shown above the tanglegram. Intersecting lines map leaves on the left tree to the corresponding leaves on the right. Strains are coded by cluster number; strain identities can be found in Supplemental Table 2. Bootstrap values greater than 70 are shown in red. Scale bars indicate percent divergence. **C**, Pairwise *d* values were calculated between each pair of phylogenies in each replicate and the mean *d* values were visualized in a heatmap. Refer to Supplemental Figure 5B for the standard error of the mean of each pair. **D**, Higher order relationships between vRNA segments were assessed with UPGMA dendrograms derived from mean *d* values. The point at which edges merge is equivalent to 1/2 *d*.

Genome-wide phylogenetic inference can distinguish parallel evolution of vRNA segments that interact during genomic assembly through RNA-RNA interactions in addition to that of the proteins encoded. To examine this, we compared all pairwise *d* values in H3N2 viruses from 2005-2014. **Figure 2C** shows the mean *d* values of all seven replicate trees for each pair of vRNA segments (refer to **Supplemental Figure 2B** for the standard error). To establish a threshold for significance in parallel evolution between vRNA trees, we determined a 95% confidence interval for *d* using a null dataset of randomly generated trees with an equivalent number of leaves (12 in this case) (**Supplemental Figure 3A & D**). Using this methodology, we found that low *d* values rarely occur by chance, with the vast majority of *d* values being greater than 15 in null trees. By comparison, mean *d* values ranged from 6.5 for PB1 & PA to 15 for PA & NS. Intriguingly, PB1, PA, NP, and NA were most highly coevolved with each other (all mean *d* values below 10), suggesting that parallel evolution between these four vRNA segments is enriched over other vRNA segments. Surprisingly, the PB2 tree was most congruent with the NA tree rather than the PB1 or PA trees, suggesting that the relationship between these segments may supercede the essential role of the PB2 protein in the polymerase complex. In contrast, the mean *d* values of the NS phylogenies with the four core vRNA segments were 14 to 15, approaching the 95% confidence threshold of 15.3. However, higher values of *d* may have resulted from a lack of statistical significance as neither the M or NS trees were strongly supported by bootstrapping (**Supplemental Figure 1B**). To further visualize higher order relationships between vRNA segments, we assembled a matrix of the pairwise *d* values between vRNA trees and constructed a dendrogram of the *d* values (**Figure 2D**). This dendrogram highlights our observation that parallel evolution is most robust between PB1, PA, NP and NA.

Parallel evolution is largely conserved over time within H3N2 viruses.

Recent studies have identified a highly plastic and redundant network of interactions between vRNA segments in influenza virions produced during productive infection, many of which may be transient (Dadonaite et al., 2019; Le Sage et al., 2020). Based on these observations, it is

plausible that vRNA relationships defined by parallel evolution change over time. To examine the conservation of evolutionary relationships in H3N2 viruses, we estimated the mean and standard error of *d* for all pairs of vRNA trees within a set of H3N2 viruses from an earlier time period (1995-2004) (**Figure 3A** and **Supplemental Figure 2A**, respectively). As was seen in the H3N2 viruses from 2005-2014, the mean *d* values ranged widely from 4.5 to 14. PB1, PA, NP, and NA remained highly coevolved in this time period, with mean *d* values from 4.5 to 8. Statistical differences between *d* values from each time period were only found for the NS segment (p-adj < 0.05, **Figure 3C**). However, NS trees had consistently low bootstrap support (**Supplemental Figure 1**), so these differences may be attributable to insufficient resolution in the underlying trees. We constructed a dendrogram of the mean *d* values of the 1995-2004 trees to examine whether the hierarchical coevolution between vRNA segments observed in the 2005-2014 viruses was also apparent in the 1995-2004 viruses (**Figure 3B**). Comparison of these dendrograms (**Figures 2D** and **3B**) using the Robinson-Foulds distance revealed a *d* value of 6, which lies within the 95% confidence interval for a tree with eight leaves (**Supplemental Figure 3A-B**). Based on these data, we conclude that there was minimal difference in the phylogenetic relationships between vRNA segments in H3N2 viruses from these two time periods.

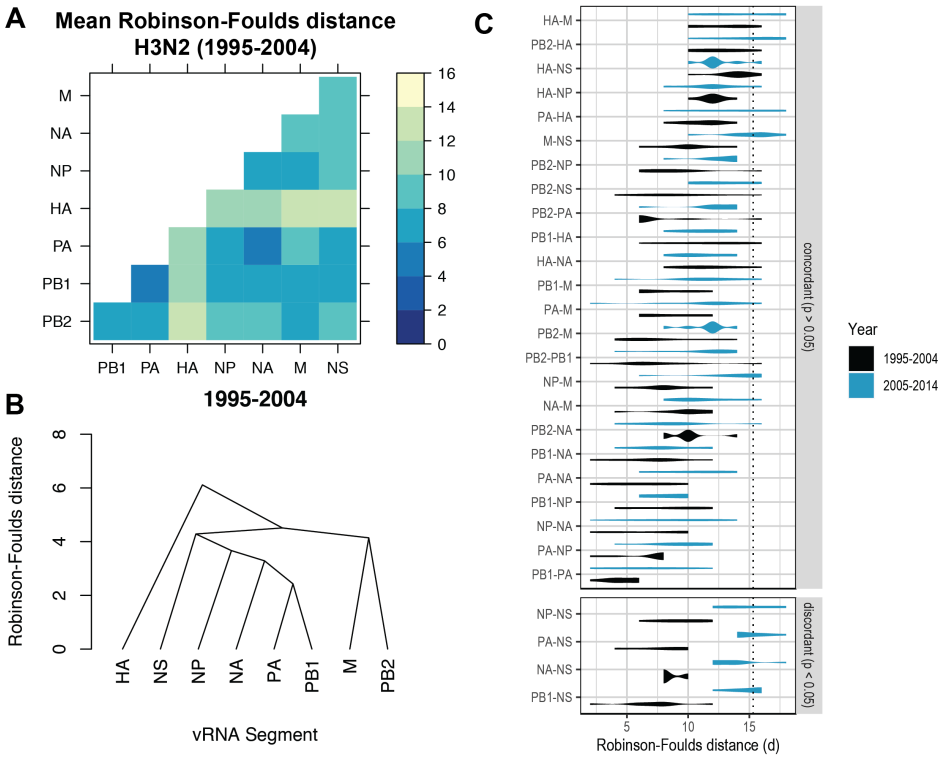


Figure 3. H3N2 virus vRNA coevolution is conserved through antigenic drift. A, Seven replicate maximum-likelihood phylogenies were reconstructed for each vRNA gene segment from human H3N2 virus sequences (1995-2004) as described in Figure 1. The Robinson-Foulds distance (*d*) between each pair of phylogenies was calculated for each replicate. The mean *d* values were visualized in a heatmap. Refer to Supplemental Figure 5A for the standard error of the mean of each pair. **B**, Higher order relationships between vRNA segments were assessed with UPGMA dendrograms derived from the mean *d* values. Scale bar corresponds to *d*. **C**, All seven replicate *d* values for each pair of trees were plotted comparing H3N2 viruses from 1995-2004 (black) to H3N2 viruses from 2005-2014 (turquoise). 'Discordant' pairs are clustered where $p < 0.05$ (Mann-Whitney *U* test with Benjamini-Hochberg correction). Dashed line, 95% confidence interval for phylogenetic concordance (determined by a null dataset; refer to Supplemental Figure 4).

Evolutionary relationships between vRNA segments are dependent upon subtype and lineage.

Our results suggest that vRNA relationships are remarkably consistent across H3N2 viruses from a period spanning two decades. To examine whether our approach captures anticipated changes in vRNA relationships, we assessed these relationships in H1N1 viruses from 2000-2008 and 2010-2018. H1N1 viruses from these time periods represent distinct lineages before and after the 2009 pandemic. This pandemic was caused by an antigenically shifted H1N1 virus that emerged from reassortment of two swine-origin viruses (Garten et al., 2009). Therefore, these two time periods represent distinct H1N1 virus lineages, and different evolutionary relationships between vRNA segments would be expected for each lineage.

193 Species trees comprising full-length concatenated H1N1 virus genomes from 2000-2008 or
194 2010-2018 were constructed and clusters were defined using the same approach described for
195 H3N2 viruses. While this method produced a similar number of clusters for both sets of H1N1
196 viruses (**Table 1**), there were fewer clusters than in H3N2 viruses, owing to the higher rate of
197 evolution observed in H3N2 viruses (Bedford et al., 2015). Seven replicate strains were selected
198 from each cluster (**Supplemental Tables 3-4**) and vRNA trees were built (**Supplemental Figure**
199 **4**). **Figure 4A-B** shows the mean *d* values of all seven replicate trees for each pair of vRNA
200 segments (refer to **Supplemental Figure 2C-D** for the standard error). The 95% confidence
201 interval cutoff for Robinson-Foulds distances corresponding to trees with 9 leaves was 8.6
202 (**Supplemental Figure 3C**) and is the threshold used for statistical comparison of parallel
203 evolution in vRNA segments from H1N1 strains.

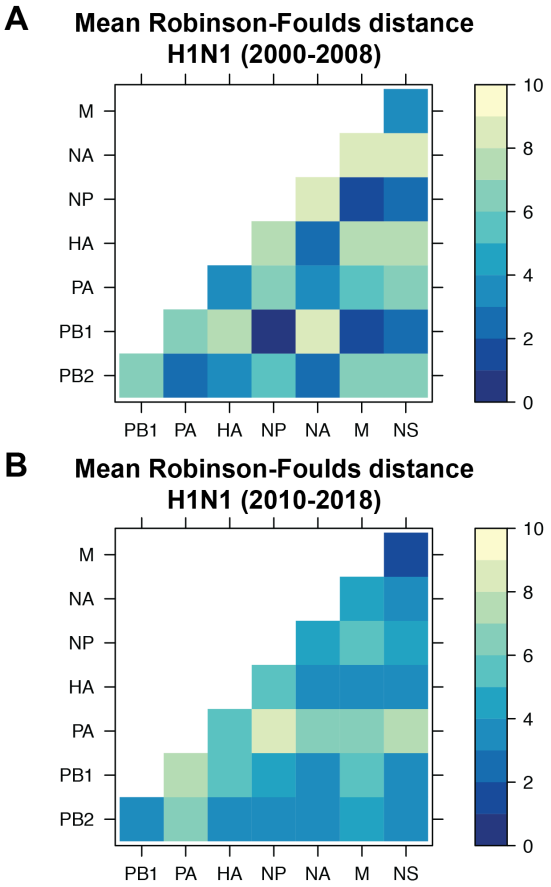


Figure 4. vRNA coevolution is dependent upon subtype and lineage. Seven replicate maximum-likelihood phylogenies were reconstructed for each vRNA gene segment from human H1N1 virus sequences from 2000-2008 **(A)** or 2010-2018 **(B)** as described in Figure 1. The pairwise Robinson-Foulds distance (*d*) between phylogenies in each replicate was calculated. The mean *d* values were visualized in a heatmap. Refer to Supplemental Figure 5C-D for the standard error of the mean of each pair.

Heatmaps comparing the *d* value between vRNA pairs suggest that vRNA relationships are not conserved across H1N1 viruses of different lineages (**Figure 4, A** vs **B**). We explored this further by constructing dendrograms of the *d* values for the pre-pandemic and post-pandemic H1N1 viruses (**Figure 5A-B**). The *d* value was 10 between pre-pandemic and post-pandemic H1N1 viruses, confirming that a high degree of incongruence existed between H1N1 viruses of different lineages (**Figure 5B**). To further explore individual differences between pairs of vRNA trees in H1N1 viruses of different lineages, we plotted *d* values for the pre-pandemic H1N1 viruses alongside the *d* values for the post-pandemic H1N1 viruses (**Figure 5C**). In stark contrast to the

213 relatively conserved vRNA relationships observed in H3N2 viruses over time, many relationships
214 between vRNA segments were disrupted in post-pandemic H1N1 viruses. Parallel evolution
215 between PB1 and NP (mean *d* increased from 1 to 5; *p-adj* < 0.05, Mann-Whitney test) was
216 notably displaced by stronger coevolution of PB1 with NA (mean *d* decreased from 9 to 3; *p-adj*
217 < 0.05). The M and NS trees remained highly coevolved across H1N1 lineages, but each one was
218 significantly more coevolved with the HA and NA trees in post-pandemic viruses (*p-adj* < 0.05).
219 The PB2 trees diverged significantly from the PA trees in favor of greater parallel evolution with
220 the NP and NS trees (*p-adj* < 0.05). Some of these data can be explained by weaker bootstrap
221 support in H1N1 trees, particularly those from H1N1 viruses from 2010-2018 (**Supplemental**
222 **Figure 4**). However, these data suggest that viruses from different lineages of the same subtype
223 develop distinct vRNA evolutionary relationships, which is an important consideration when
224 predicting reassortment between emerging viruses from different lineages.

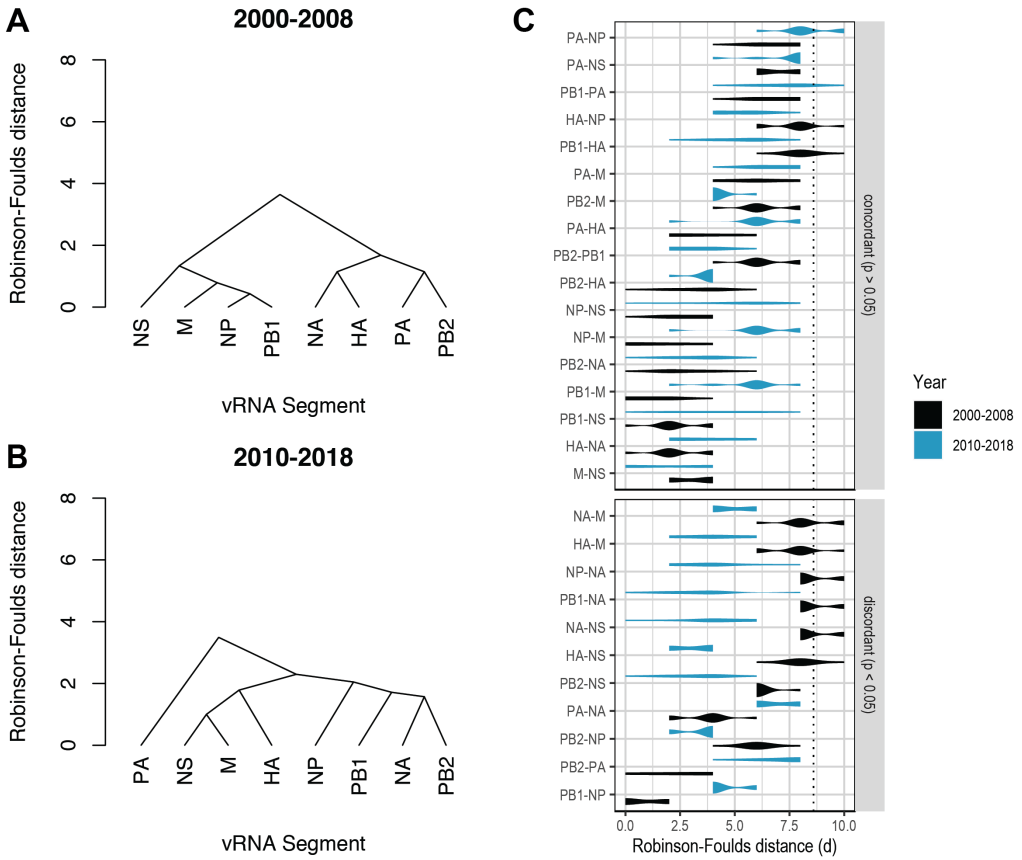


Figure 5. vRNA coevolution diverges in antigenically shifted H1N1 viruses. Higher order relationships between vRNA segments corresponding to H1N1 viruses from 2000-2008 **(A)** or 2010-2018 **(B)** were assessed with UPGMA dendrograms derived from the mean *d* values in Figure 4. Scale bar corresponds to *d*. **C**, All seven replicate *d* values for each pair of trees were plotted comparing H1N1 viruses from 2000-2008 (black) to H1N1 viruses from 2010-2018 (turquoise). 'Discordant' pairs are clustered where $p < 0.05$ (Mann-Whitney *U* test with Benjamini-Hochberg correction). Dashed line, 95% confidence interval for phylogenetic concordance (determined by a null dataset; refer to Supplemental Figure 4).

Comparison of the dendrograms from H1N1 and H3N2 viruses revealed some expected similarities as well as differences (**Fig 2D, 3D,** and **5A-B**). For example, PB1 and NP share a common evolutionary relationship across all four sets of influenza viruses examined in this study (**Figures 2D, 3B,** and **5A-B**), which may be expected based on the shared role of the encoded proteins in replication (Fodor, 2013). However, in H3N2 viruses PB1 and NP are next most closely related to NA and NP (**Figures 2D** and **3B**), while in H1N1 viruses their relationship with other segments varies (**Figure 5A-B**). We compared the overall similarity of all four dendrograms by computing the Robinson-Foulds distance. The pre-pandemic H1N1 virus dendrogram was significantly different from both of the H3N2 virus dendrograms: $d = 10$ when compared to either

1995-2004 or 2005-2014 H3N2 virus dendrograms. A *d* value of 10 is well outside of the 95% confidence interval cutoff of 6.5 determined by the null dataset for trees with 8 leaves (**Supplemental Figure 3A-B**). Likewise, *d* = 10 for the post-pandemic H1N1 virus dendrogram when compared to either of the H3N2 virus dendrograms. These results are in direct contrast to our previously determined distance for the H3N2 virus dendrograms to one another (*d* = 6). Overall, these data indicate that parallel evolution between vRNA segments is distinct between influenza subtypes isolated from humans within similar time scales.

Parallel evolution in H3N2 viruses is not driven solely by protein-coding mutations.

As discussed previously, a phylogenetics approach such as ours would encompass parallel evolution driven by either protein and RNA relationships. We have already shown that known protein relationships between PB1 and PA, two members of the polymerase complex, are identified by our approach (**Figure 2A**). However, the observation that PB2 is more coevolved with NA than with either PB1 or PA (**Figure 2C**) suggests that our method also reveals protein-independent parallel evolution, since these proteins are not known to function together during infection. Using H3N2 viruses from 2005-2014, which yielded vRNA trees with the highest overall bootstrap support (**Supplemental Figure 1B**), we explored the extent to which parallel evolution between vRNA segments is driven by protein-coding mutations. To do so, we converted the vRNA sequence alignments, which are negative-sense, into positive-sense RNA (i.e., coding sense) and translated the coding sequences into amino acid alignments. For the M and NS sequence alignments that encode two splice variants each, the M1/M2 and NS1/NS2 amino acid alignments were both translated and analyzed individually. Neighbor-joining trees were reconstructed from the amino acid alignments and the evolutionary relationships between H3N2 proteins were analyzed by the Robinson-Foulds distance. We took the resultant *d* values and constructed a dendrogram between all pairs of protein trees, as was previously done with vRNA trees (**Supplemental Figure 5**). This dendrogram appears distinct from the dendrogram built from the corresponding gene (vRNA) trees (**Figure 2D**). As might be expected, the greatest degree of

parallel evolution lying at the core of this dendrogram was between HA and NA, two viral glycoproteins with coordinated functions in attachment, motility, and entry (Bloom et al., 2010; Sakai, Nishimura, Naito, & Saito, 2017).

To compare parallel evolution between influenza proteins to that of the parent vRNA segments, the mean *d* values from the gene trees were plotted against the mean *d* values from the protein trees (**Figure 6**). In the case of the M and NS segments, mean *d* values for all M1/M2 or NS1/NS2 combinations are shown. Many vRNA pairs, such as PB2 and PB1, lie along the identity line, indicating that protein interactions are more likely to drive parallel evolution in those vRNA segments. Interestingly, HA and NA were the only pair of vRNA segments that lay significantly above the identity line, strongly supporting the observation made by others that epistatic interactions between these proteins constrains their evolution (Jang & Bae, 2018; Kryazhimskiy et al., 2011; Neverov et al., 2015). Of particular interest was that several vRNA segments, such as PB2 and NA (**Figure 6, open diamond**), lay significantly below the identity line. This could be indicative of either purifying selection, or of greater parallel coevolution between the vRNA segments than the proteins encoded. While this is not altogether unexpected, considering that the mutation rate of a protein is unlikely to be as high as the mutation rate of the corresponding gene, we would expect conserved RNA interactions to also have this effect. These observations suggest that parallel evolution may identify putative RNA interactions between vRNA segments that could facilitate selective assembly and packaging.

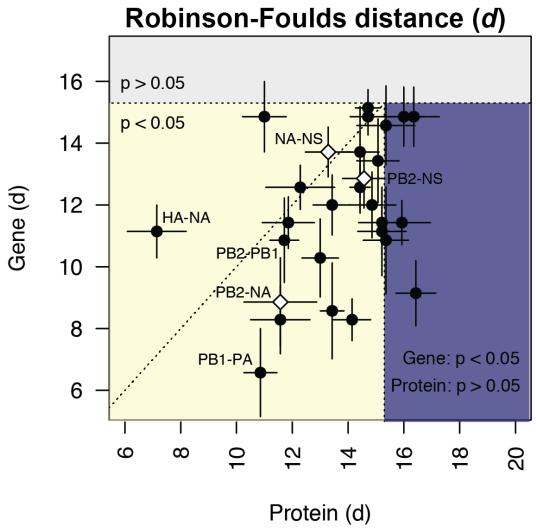


Figure 6. Protein-coding substitutions do not fully account for gene coevolution. H3N2 virus vRNA gene sequence alignments from 2005-2014 were translated into the corresponding amino acid alignments. Neighbor-joining phylogenies were reconstructed from these alignments and the Robinson-Foulds distance (*d*) was tabulated for all protein tree pairs. The mean *d* value of each pair of protein trees was plotted against the mean *d* value of the corresponding gene trees. For the M and NS gene segments, which encode multiple protein products, *d* values were calculated for each protein tree individually and the average *d* values were used. Error bars indicate the standard error of the mean (SEM) for replicate trees (n = 7). Dashed horizontal and vertical lines, 95% confidence interval (CI) for phylogenetic concordance, as determined by a null dataset (refer to Supplemental Figure 4). The region shaded yellow lies within the 95% CI for both gene and protein trees with the identity line plotted. The region shaded blue lies within the 95% CI for gene trees but not protein trees. The region shaded gray lies outside the 95% CI for both gene and protein trees.

PB2 and NA viral ribonucleoprotein complexes (vRNPs) preferentially colocalize at the nuclear periphery *in vitro*.

To address whether parallel evolution between the PB2 and NA segments corresponds with their behavior during influenza virus infection, we examined whether these vRNA segments preferentially colocalize in infected cells. During influenza virus infection, viral RNA are synthesized in the nucleus and then transported to the plasma membrane for packaging on endocytic vesicles (Lakdawala, Fodor, & Subbarao, 2016). We therefore examined intracellular colocalization of PB2, NA, and NS during infection with an H3N2 virus. These segments

encompass a pair of segments with high gene-based parallel evolution (PB2-NA) as well as pairs with little evidence of parallel evolution (PB2-NS; NA-NS) (**Figures 2C** and **6, open diamonds**).

We quantified colocalization using our established method for examining intracellular colocalization of vRNA segments by fluorescence *in situ* hybridization (FISH) and immunofluorescence (IF) in productively infected cells (Lakdawala et al., 2014; Nturibi, Bhagwat, Coburn, Myerburg, & Lakdawala, 2017). Lung epithelial A549 cells were infected with an H3N2 virus representative of the time period analyzed (A/Perth/16/2009) for 8 hours and stained for three vRNA segments, NP protein, and nuclei. The NP antibody stain was used to normalize the pairwise colocalization data to the total number of vRNP foci present in the cells. Entire cell volumes were captured to analyze the colocalization of vRNA segments specifically within the cytoplasm by masking the signal within the nucleus. A representative image of an infected cell from one of three independently performed experiments is shown after deconvolution (**Figure 7A**) and segmentation of cytoplasmic foci (**Figure 7B**).

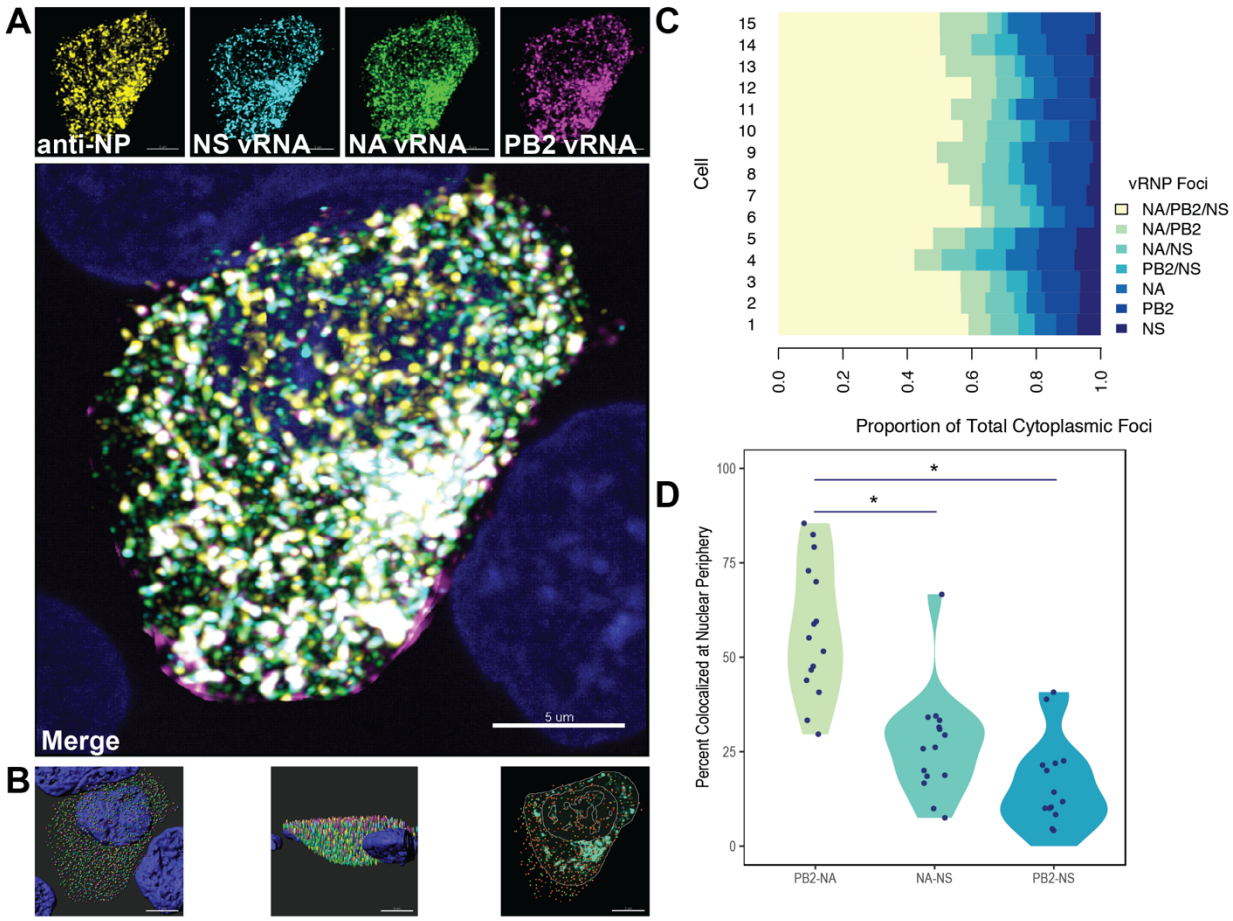


Figure 7. Colocalization of vRNA segments at the nuclear periphery correlates with evolutionary relationships during productive viral infection. A549 cells were infected with A/Perth/16/2009 (H3N2) at an MOI of 2 or mock infected. Cells were fixed at 8 hours post-infection and combination fluorescence *in situ* hybridization/immunofluoresence (FISH and IF, respectively) was performed. FISH probes targeting the NS, NA and PB2 vRNA segments were labeled with Alexa Fluor 488, Quasar 570 and Quasar 670, respectively. Antibodies targeting nucleoprotein (NP) were used with an anti-mouse Alexa Fluor 594 secondary antibody. Nuclei were labeled with DAPI. Coverslips were mounted and volumetric imaging was performed to obtain Nyquist sampling. **A,** A maximum projection image of a representative cell is shown after cell segmentation. Scale bar corresponds to 5 µm. **B,** A 3D rendering of the cell after analysis. **C,** Colocalization of vRNA segments was assessed in 15 individual infected cells. **D,** Quantification of each pair of vRNA segments within 300 nm of the nuclear border. Each point represents an individual cell (n = 15). Aggregate data from three independently performed experiments are shown. * denotes $p < 0.05$ (Mann-Whitney *U* test)

Whole cytoplasmic analysis of vRNP foci colocalization in fifteen individually analyzed cells revealed that the majority of cytoplasmic foci contained all three NA, PB2 and NS segments (**Figure 7C**). This observation may represent the complex nature of genomic assembly, where all segments come together en route to the plasma membrane. In addition, the cytoplasm includes vRNA segments colocalized at the nuclear periphery prior to packaging, where there are more complete assembly complexes (Lakdawala et al., 2014). Given that newly synthesized vRNA segments are exported from the nucleus as assembly intermediates comprising greater than two but fewer than 8 segments, these perinuclear assembly intermediates may function as nodes for

311 further assembly (Majarian, Murphy, & Lakdawala, 2018). Therefore, we assessed the potential
312 for PB2, NA and NS to colocalize at the nuclear periphery, where assembly intermediates first
313 begin to form. We defined localization at the nuclear periphery to within three-hundred
314 nanometers, the limit of resolution in this system. Examination of newly exported vRNP
315 complexes within three-hundred nanometers of the nuclear periphery revealed an enrichment of
316 PB2-NA vRNP complexes over either NA-NS or PB2-NS vRNP complexes. These data suggest
317 that PB2 and NA vRNA segments could comprise one such node from which to build the entire
318 complex of all eight segments and support the hypothesis that RNA assembly interactions, in
319 addition to protein interactions, can drive parallel evolution between vRNA segments in influenza
320 viruses.
321
322 **Discussion**
323 In this study, we used phylogenetics and molecular biology methods to investigate genome-
324 wide relationships between vRNA segments in seasonal human influenza A viruses. We found
325 that parallel evolution varies considerably between vRNA segments, with distinct relationships
326 forming in different influenza virus subtypes (H1N1 vs H3N2) and between H1N1 virus lineages
327 that arose from distinct host origins. We further demonstrate that evolutionary relatedness
328 between vRNA segments in H3N2 viruses is largely conserved over time. Importantly, our data
329 suggest that parallel evolution cannot be attributed solely to protein interactions, and we
330 successfully predicted intracellular colocalization between two coevolved vRNA segments during
331 infection with an H3N2 virus. Thus, we present a phylogenetic approach for interrogating putative
332 RNA associations that could be broadly applied toward the study of genomic assembly and
333 reassortment in segmented viruses.
334 Selective assembly of all eight genomic segments is fundamental to the production of fully
335 infectious virus particles. We and others have used a variety of biochemical approaches to
336 investigate the mechanisms that promote selective assembly (Dadonaite et al., 2019; Le Sage et

al., 2020). We previously demonstrated that binding of vRNA segments by the NP protein is non-uniform and non-random (Le Sage et al., 2018; Lee et al., 2017), supporting the model that intersegmental RNA interactions facilitate selective assembly. Biochemical approaches to define bona fide intersegmental RNA-RNA interactions demonstrated that the interaction network is highly flexible and varies between H1N1 and H3N2 viruses (Dadonaite et al., 2019; Le Sage et al., 2020). These observations are consistent with our conclusion that RNA interactions constrain parallel evolution between vRNA segments in a manner sensitive to the genetic context studied.

The approach we present here differs from other experimental approaches in that we identify a novel, conserved RNA-based relationship in H3N2 viruses. For example, we found that relationships between PB1, PA, NP and NA are enriched over other segments in H3N2 viruses and conserved over time. One might expect PB1, PA and NP to coevolve because of the functions of the proteins they encode: the polymerase subunits PB2, PB1 and PA form a supramolecular complex around each vRNA segment with NP (Fodor, 2013). However, this explanation does not account for the parallel evolution observed between vRNP protein components and NA, and our microscopy data demonstrates that the NA segment preferentially colocalizes with the vRNA of one such vRNP component, supporting the possibility that parallel evolution of NA with PB1, PA and NP could also be driven by RNA interactions. These observations suggest that RNA relationships with the NA segment may facilitate selective assembly of vRNA segments. Further work should be directed at determining the underlying nature driving the novel relationship between these segments.

Previous pandemic influenza viruses emerged through reassortment (Neumann et al., 2009). Prediction of future influenza pandemics relies on understanding assembly of vRNA segments within a cell. As we have discussed, experimental investigations of intersegmental RNA interactions indicate that the vRNA interactome is distinct among virus strains and highly plastic (Dadonaite et al., 2019; Le Sage et al., 2020). Therefore, experimental approaches are unlikely to provide the holistic view necessary to predict reassortment outcomes of two circulating

363 influenza strains. In contrast, we identified several conserved relationships between vRNA
364 segments in H3N2 viruses that could impose constraints on reassortment. Thus, investigation of
365 epistatic relationships between vRNA segments through phylogenetics could inform the
366 sequence-based prediction of barriers to reassortment in emerging influenza viruses.
367
368 **Materials and Methods**
369 **Viruses and cells**
370 Human adenocarcinoma alveolar basal epithelial cells (A549, ATCC) were maintained in high-
371 glucose Dulbecco's Modified Eagle Medium (DMEM, Sigma) supplemented with 10% fetal bovine
372 serum (FBS, HyClone), 2% L-glutamine, and 1% penicillin/streptomycin. Madin-Darby canine
373 kidney epithelial cells (MDCK, ATCC) were maintained in MEM supplemented with 10% FBS. All
374 cells were maintained at 37°C in 5% CO_2. Reverse genetics plasmids of the influenza A virus,
375 A/Perth/16/2009 (H3N2), were kindly provided by Dr. Jesse Bloom (Fred Hutchinson Cancer
376 Research Center, Seattle). Recombinant virus was rescued as previously described (Lakdawala
377 et al., 2011). Virus titers were determined by 50% tissue culture infectious dose ($TCID_{50}$) on
378 MDCK-SIAT cells using the endpoint titration method (Reed & Muench, 1938).
379 **Influenza A virus sequences**
380 FASTA files of each genomic segment of human influenza A virus sequences of H1N1 and
381 H3N2 viruses were downloaded from the Influenza Research Database (IRD,
382 http://www.fludb.org) (Zhang et al., 2016) on June 22, 2018 and July 3, 2018, respectively. Strains
383 lacking full-length genomic sequence data were excluded.
384 **Clustering and sequence selection**
385 Sequences were read into R (version 3.5.2) using the DECIPHER (version 2.18.1) package
386 (Wright, 2015) and subset into the time periods 1995-2004 and 2005-2014 (H3N2 strains) or
387 2000-2008 and 2010-2018 (H1N1 strains). Time periods were selected in part to ensure a similar
388 level of genetic diversity between strains. In each strain, all eight vRNA segments were

389 concatenated into a full-length genome from which alignments were constructed. A neighbor-
390 joining species tree was built by clustering strains into operational taxonomic units with sequence
391 identity cutoffs ranging from 95-99%. In H3N2 viruses from 1995-2004, there were 3, 7, 16, 53,
392 and 259 clusters corresponding to cutoffs of 95%, 96%, 97%, 98%, and 99% sequence identity,
393 respectively. The 95-96% sequence identity cutoffs were discarded, as these produced trees with
394 an insufficient number of branches for comparison by the Robinson-Foulds distance. However,
395 as the cutoff for sequence identity was increased from 97% to 99%, we observed a corresponding
396 decrease in bootstrap support for trees built from representative sequences. A sequence identity
397 cutoff of 97% was therefore selected to ensure the greatest degree of robustness in tree
398 topologies. Small clusters occurred infrequently and were either omitted or collapsed into a single
399 cluster. Seven replicate strains were randomly chosen from each cluster for further study and
400 visually inspected for sequencing ambiguities. A list of all strains analyzed and the corresponding
401 accession numbers can be found in **Supplemental Tables 1-4**.

402 **Phylogenetic tree reconstruction and analysis**

403 Maximum-likelihood phylogenies were reconstructed under an HKY85 model for either full-
404 length genomes or individual vRNA segments with 100 or 1,000 bootstrap replicates, as indicated,
405 using the DECIPHER package in R. The phangorn package (version 2.5.5) (Schliep, 2011) was
406 used to identify an appropriate model of evolution for phylogenetic reconstruction. Strain names
407 are coded by cluster number in all trees. Phylogenetic trees of full-length concatenated genomes
408 are shown in **Supplemental Figure 6**. Neighbor-joining protein phylogenies were built from amino
409 acid alignments after translation of the corresponding coding sequence alignments.

410 Tanglegrams were built from pairs of vRNA phylogenies within replicates using the phytools
411 package (version 0.7-70) (Revell, 2012). The Robinson-Foulds distance (*d*) was calculated for
412 each pair of phylogenies using the ape package (version 5.4-1) (Paradis & Schliep, 2019).

Dendrograms visualizing the mean *d* values between vRNA segments were built using the UPGMA method. Tanglegrams were constructed between each pair of dendrograms and a *d* value was calculated for each tanglegram.

Fluorescence *in situ* hybridization and immunofluorescence

Custom Stellaris RNA FISH oligonucleotide probes specific for the H3N2 virus NS, NA and PB2 vRNA segments were purchased from BioSearch Technologies (refer to **Supplemental Table 5** for FISH probe sequences). Each custom probe mix is comprised of 20 to 40 20-mers that span the length of the vRNA segment of interest. Probes with high complementarity against other vRNA segments or positive-sense RNA were excluded during the design process. The NS probe was purchased with a terminal amine group and manually conjugated to the Alexa Fluor 488 fluorophore using the Alexa Fluor 488 Oligonucleotide Amine Labeling Kit (Invitrogen). The NA and PB2 probes were labeled by the manufacturer with the Quasar 570 and Quasar 670 fluorophores, respectively.

Three independent fluorescence *in situ* hybridization and immunofluorescence (FISH-IF) experiments were performed. A549 cells were seeded directly onto 1.5 mm circular coverslips (Fisher Scientific) in tissue culture dishes. The next day, cells were infected at a multiplicity of infection (MOI) of 2 with A/Perth/16/2009 (H3N2) or mock infected in diluent. Cells were fixed at 8 hours post-infection with 4% paraformaldehyde and permeabilized overnight in ice cold 70% ethanol. Prior to hybridization, cells were rehydrated in wash buffer (10% formamide and 2X saline sodium citrate [SSC] in DEPC-treated H_2O) and then incubated at 28°C overnight in hybridization buffer (10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNA-free BSA, 1 mg/ml *E. coli* tRNA, 2X SSC, and 10% formamide in DEPC-treated H_2O) with anti-influenza A virus NP antibody (Millipore, 1:2,000) and FISH probes. After hybridization, cells were washed and incubated with Alexa Fluor 594 goat anti-mouse (Invitrogen, 1:2,000) and DAPI (Sigma, 1:5,000) in wash buffer. Coverslips were mounted on slides in ProLong Diamond antifade mountant (Thermo Fisher).

439

440 **Confocal Imaging**

441 Microscope slides were imaged on a Leica SP8 confocal microscope equipped with a pulsed
442 white light laser as an excitation source and an acousto-optical beam splitter (AOBS) and Leica
443 Hybrid Detectors. All imaging was performed with a 100X oil immersion objective with a numerical
444 aperture of 1.4. Sequential scanning with a line averaging of 3 between frames was used. To
445 obtain Nyquist sampling, z-stacks of each cell were taken with a step size of 0.17 µm to achieve
446 a pixel size of 45 nm x 45 nm x 170 nm. The following custom parameters were established using
447 single-color infected controls for sensitive detection of all five fluorophores: 405 nm excitation
448 wavelength (λ_{ex}) with 0.5% laser power and a detection range of 415 to 470 nm (PMT1; DAPI),
449 488 nm λ_{ex} with 10% laser power and a detection range of 493 to 540 nm with time gating of 1 to
450 6 nanoseconds (ns) (HyD4; Alexa Fluor 488), 582 λ_{ex} with 15% laser power and a detection range
451 of 590 to 635 nm with time gating of 1.5 to 6 ns (HyD4; Cal Fluor Red 590), 545 nm λ_{ex} with 5%
452 laser power and a detection range of 545 to 568 nm with time gating of 1.5 to 6 ns (HyD4; Quasar
453 570), 647 nm λ_{ex} with 5% laser power and a detection range of 670 to 730 nm with time gating of
454 1.5 to 6 ns (HyD5; Quasar 670). In each experiment, five volumetric z-stacks were imaged of
455 infected cells and one z-stack was imaged of mock infected cells.

456 **FISH Colocalization**

457 Background subtraction and deconvolution of confocal images were performed manually for
458 each channel using Huygens Essential software (version 19.04, Scientific Volume Imaging B.V.).
459 In each experiment, images taken of mock-infected cells were deconvolved using the same
460 parameters as those of infected cells. 3D reconstruction and colocalization analysis of the
461 resulting images were performed using Imaris software (version 8.4.2, Bitplane AG) as previously
462 described (Lakdawala et al., 2014; Nturibi et al., 2017). Briefly, the cell of interest in each image
463 was segmented using the ‘Surfaces’ and ‘Cell’ tools in Imaris software. DAPI signal was used to
464 mask nuclear signal from the remaining channels. The ‘Spots’ tool was then used to populate the

reconstructed cell with four different sets of Spots corresponding to foci from each of the remaining channels. In each experiment, the mock infected cell was analyzed in an identical manner and the fluorescence intensity for each channel of the mock-infected cell was used to establish fluorescence intensity thresholds at which 97% or more of the background signal was removed prior to Spot generation. A modified Matlab extension was then used to quantify spot colocalization using a distance threshold of 300 nm as previously described (Nturibi et al., 2017). Colocalization data was imported into the Cell and all data was exported and analyzed in R.

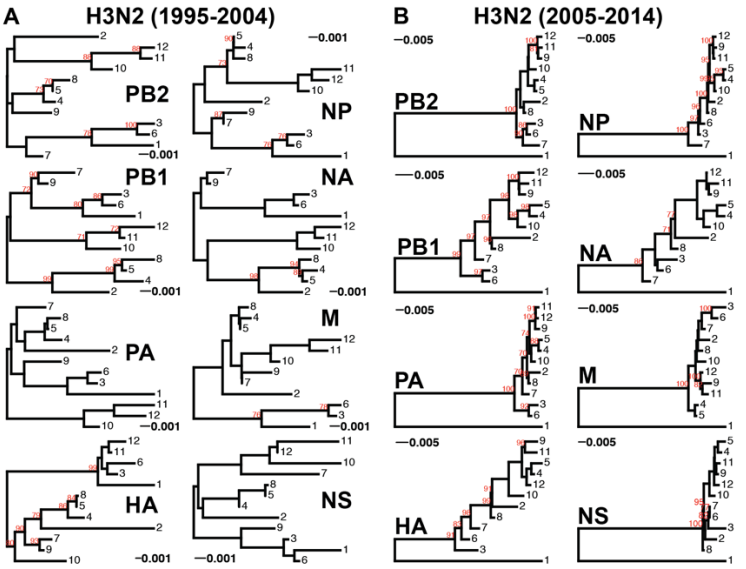
Statistics

Sets of null trees were used to determine confidence intervals for the Robinson-Foulds distance (*d*) between phylogenetic trees. A set of 1,000 randomly sampled, unrooted trees with 8, 9, or 12 tips were built using the ape package. The Robinson-Foulds distance (*d*) was calculated for all pairs of trees and these were fit to a linear regression model. Null *d* values were either log-transformed or transformed by the Yeo-Johnson method (Yeo & Johnson, 2000), as indicated. Mean *d* values calculated for pairs of vRNA trees were considered significant if they fell within the first five percentiles as compared to null *d* values from random trees with the same number of tips (i.e. 95% of null *d* values were greater than the mean *d* value for a given pair of vRNA trees). A Mann-Whitney *U* test with a Benjamini-Hochberg post-hoc correction was used to identify statistically significant differences between *d* values from two time periods. A Mann-Whitney *U* test was also used to determine statistical significance of FISH-IF colocalization data.

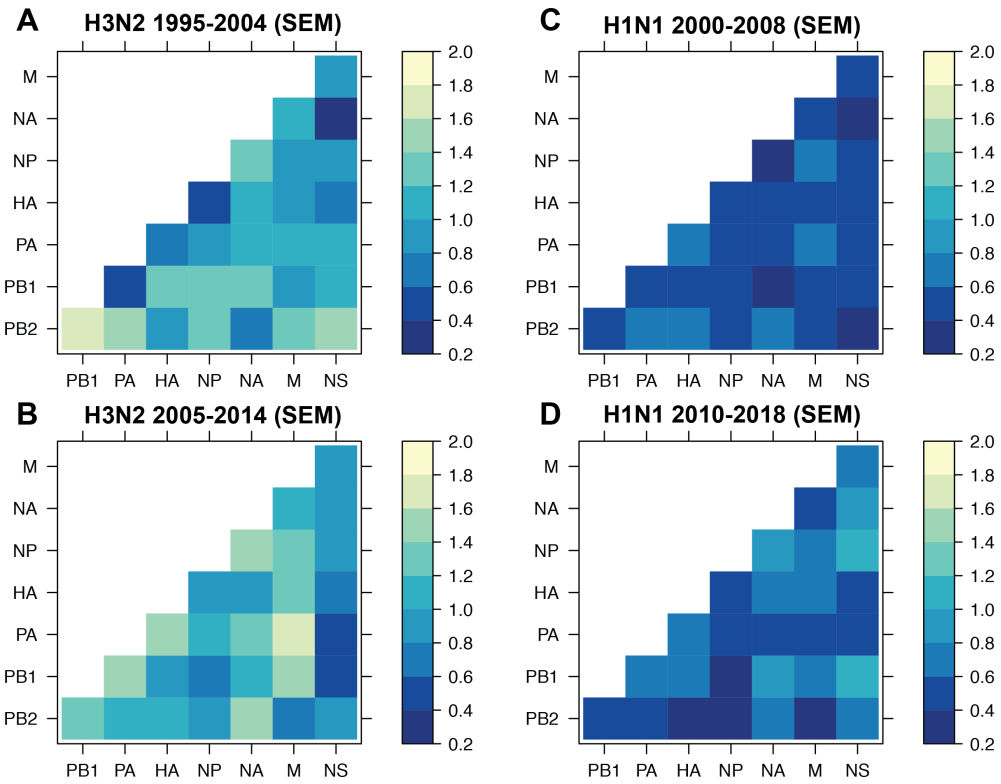
Code availability

Custom code for analysis of parallel evolution in concatenated, full-length genomic influenza virus sequences is available on GitHub (https://github.com/Lakdawala-Lab/Parallel-Evolution/).

Supplemental Material

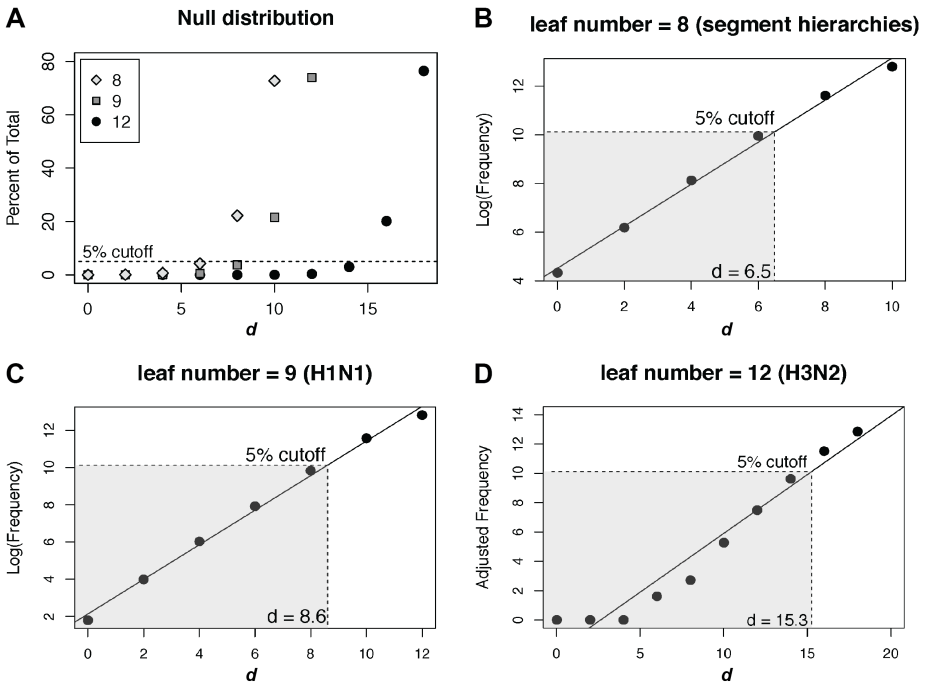


Supplemental Figure 1. Phylogenies of vRNA segments from H3N2 virus sequences. Maximum-likelihood phylogenetic reconstruction of each vRNA segment from H3N2 virus sequences shown in Supplemental Tables 1-2. **A**, Representative phylogenies from replicate 2 for the 1995-2004 viruses. **B**, Representative phylogenies from replicate 7 for the 2005-2014 viruses. Sequences are coded by OTU cluster. Bootstrapping was performed with 100 replicates (bootstrap values greater than 70 are shown in red). Scale bars indicate percent divergence.



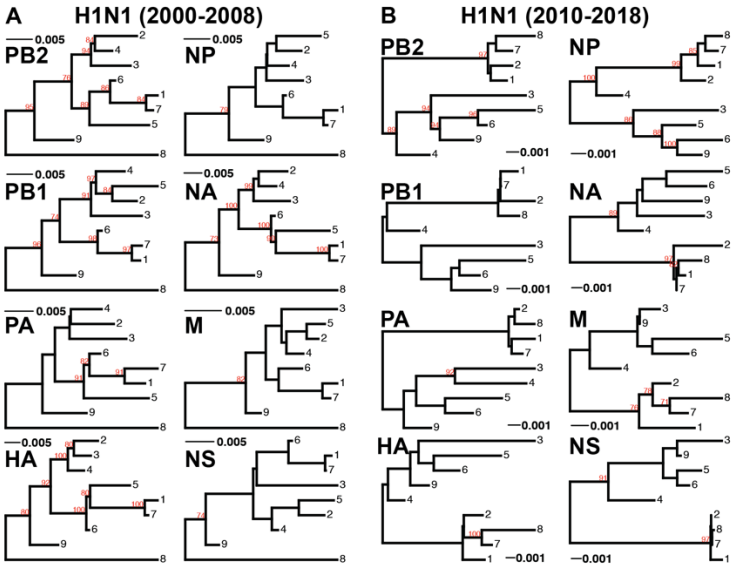
Supplemental Figure 2. The standard error of the mean (SEM) of replicate Robinson-Foulds distances. The SEM of all pairwise Robinson-Foulds distances (*d*) was determined for H3N2 viruses from 1995-2004 (corresponding mean *d* values shown in Figure 3A) **(A)**, H3N2 viruses from 2005-2014 (corresponding mean *d* values shown in Figure 2C) **(B)**, H1N1 viruses from 2000-2008 (corresponding to mean *d* values from Figure 4A) **(C)**, and H1N1 viruses from 2010-2018 (corresponding to Figure 4B) **(D)**.

494



Supplemental Figure 3. Null distribution of *d* values. A, The null distribution of *d* values in sets of 1,000 randomly sampled, unrooted trees with either 8 (gray diamonds, corresponding to the number of tips in each hierarchical dendrogram), 9 (gray squares, the number of tips in each H1N1 phylogeny), or 12 (black circles, the number of tips in each H3N2 phylogeny) leaf tips was determined. A dashed line demarcates the threshold for the 95% confidence interval. **B-D,** The null distributions shown in **A** were log transformed **(B, C)** or Yeo-Johnson transformed (Yeo & Johnson, 2000) **(D)** and fit to a linear regression model to establish the cutoff for the first five percentiles (shaded region), which was set as the 95% confidence interval cutoff.. The value of *d* at which 95% confidence is exceeded is indicated.

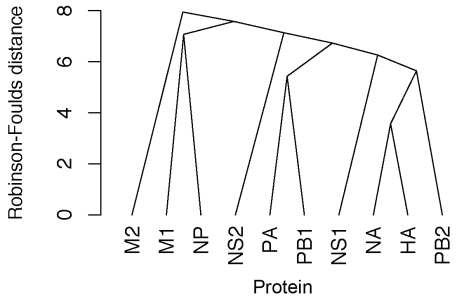
495



Supplemental Figure 4. Phylogenies of vRNA segments from H1N1 virus sequences. Maximum-likelihood phylogenetic reconstruction of each vRNA segment from H1N1 virus sequences shown in Supplemental Tables 3-4. **A**, Representative phylogenies from replicate 6 for the 2000-2008 viruses. **B**, Representative phylogenies from replicate 3 for the 2010-2018 viruses. Sequences are coded by OTU cluster. Bootstrapping was performed with 100 replicates (bootstrap values greater than 70 are shown in red). Scale bars indicate percent divergence.

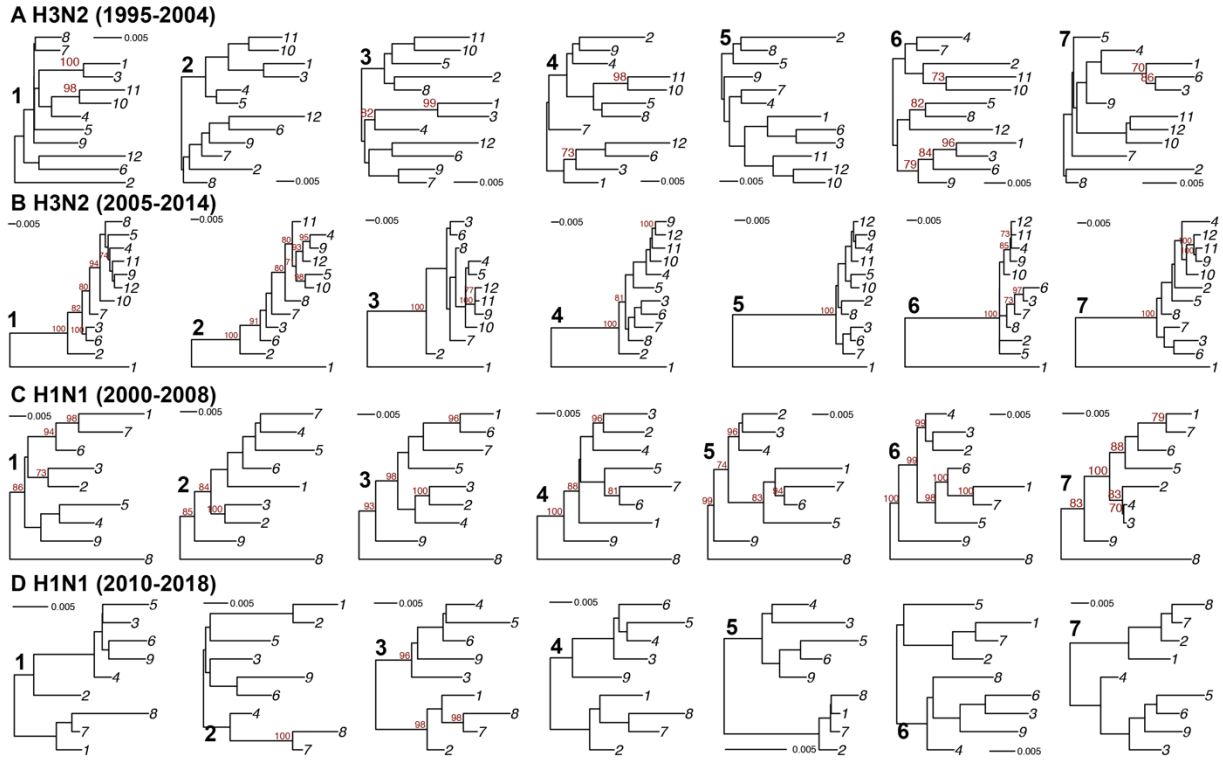
496

497



Supplemental Figure 5. Coevolution of influenza A virus proteins in H3N2 viruses from 2005-2014. H3N2 virus vRNA gene sequence alignments were translated into the corresponding amino acid alignments. Neighbor-joining trees were reconstructed from these alignments and the Robinson-Foulds distance (*d*) was tabulated for all protein tree pairs. Higher order relationships between IAV proteins were assessed in a UPGMA dendrogram. Scale bar corresponds to *d*.

498



Supplemental Figure 6. Replicate phylogenies of full-length concatenated H3N2 and H1N1 virus sequences. Maximum-likelihood phylogenetic reconstruction of full-length H3N2 virus (**A-B**) and H1N1 virus (**C-D**) genomic sequences shown in Supplemental Tables 1-4. Sequences are coded by OTU cluster. The numbers 1-7 in bold indicate replicates. Bootstrapping was performed with 1,000 replicates (bootstrap values greater than 70 are shown in red). Scale bars indicate percent divergence.

499

H3N2 1995-2004	Representative strains analyzed in each replicate and the corresponding accession numbers.						
Cluster ID	1	2	3	4	5	6	7
1	A/Christchurch/ 339/2004 CY002904, CY002977 - 83	A/New York/ 366/2004 CY002192 - 99	A/Ashburton/ 280/2004 CY002954 - 61	A/New York/ 319/2004 CY002712 - 9	A/Christchurch/ 14/2004 CY002922 - 9	A/New York/ 372/2004 CY002224 - 31	A/Christchurch/ 13/2004 CY002906 - 13
2	A/Siena/ 3/1995 CY038503 - 10	A/Netherlands/ 5/1998 CY114485 - 92	A/Hong Kong/ 49/1995 CY114205 -12	A/Johannesburg/ 10/1997 CY112853 - 60	A/Hong Kong/ 1/1997 CY112837 - 44	A/Netherlands/ 1/1995 CY114221 - 8	A/Finland/ 381/1995 CY114181 - 8
3	A/Netherlands/ 213/2003 HQ166049 - 56	A/New York/ 198/2003 CY001013 - 20	A/New York/ 32/2003 CY000025 - 32	A/Netherlands/ 217/2003 CY112981 - 88	A/New York/ 204/2003 CY002520 - 7	A/Wellington/ 01/2004 CY012104 - 11	A/Fujian/ 445/2003 CY121448 - 55
4	A/New York/ 150/2000 CY000465 - 72	A/New York/ 316/1999 CY002160 - 7	A/New York/ 430/2000 CY003448 - 55	A/New York/ 171/1999 CY000673 - 80	A/New York/ 321/1999 CY001864 - 71	A/New York/ 142/2000 CY001520 - 7	A/New York/ 180/2000 CY000737 - 44
5	A/Memphis/ 59/1999 CY002112 - 9	A/New York/ 250/1998 CY001493 - 00	A/New York/ 324/1999 CY001976 - 83	A/New York/ 456/1999 CY003608 - 15	A/New York/ 458/1999 CY003624 - 31	A/New York/ 278/1999 CY001968 - 75	A/New York/ 289/1998 CY001784 - 91
6	A/Singapore/ 1/1996 CY112829 - 36	A/Wyoming/ 03/2003 CY034108 - 15	A/Taiwan/ 9/2004 CY040082 - 9	A/Hong Kong/ HKU19/2004 CY038607 - 14	A/Yucatan/ 844/2003 CY032961 - 8	A/Hong Kong/ HKU20/2004 CY038615 - 22	A/Hong Kong/ HKU29/2004 CY038631 - 8
7	A/Malaysia/ 12974/1999 CY118562 - 9	A/South Carolina/ NHRC0001/1999 CY090901 - 8	A/Auckland/ 10/1997 CY114293 - 00	A/HaNoi/ 1766/2001 CY105190 - 7	A/California/ 32/1999 CY121424 - 31	A/Illinois/ NHRC0001/1999 CY090893 - 00	A/Georgia/ NHRC0001/1999 CY090885 - 92
8	A/New York/ 435/2000 CY006659 - 66	A/Malaysia/ 10877/1999 CY118482 - 9	A/Malaysia/ 10997/1999 CY118498 - 05	A/Panama/ 2007/1999 CY034100 - 7	A/New York/ 526/1997 CY007979 - 86	A/New York/ 502/1998 CY008180 - 7	A/New York/ 513/1997 CY006267 - 74
9	A/New York/ 247/1998 CY001504 - 11	A/New South Wales/ 15/1999 CY016539 - 46	A/New York/ 601/1996 CY015516 - 23	A/Dunedin/ 1/2000 CY012616 - 23	A/Waikato/ 6/2000 CY013048 - 55	A/Wellington/ 21/2000 CY019989 - 96	A/New South Wales/ 4/1999 CY016627 - 34
10	A/New York/ 100/2002 CY000225 - 32	A/New York/ 110/2002 CY000113 - 20	A/New York/ 192/2003 CY000753 - 60	A/New York/ 213/2003 CY001405 - 12	A/New York/ 71/2001 CY000481 - 8	A/New York/ 193/2003 CY000865 - 72	A/New York/ 99/2002 CY001080 - 7
11	A/New York/ 11/2003 CY000513 - 20	A/Mexico/ DIF2662/2003 KJ855448 - 55	A/New York/ 22/2003 CY000361 - 8	A/New York/ 46/2003 CY000785 - 92	A/New York/ 10/2004 CY000761 - 8	A/New York/ 59/2003 CY000957 - 64	A/New York/ 65/2003 CY001221 - 8
12	A/England/ 431/2003 CY087979 - 86	A/England/ 442/2003 CY088019 - 26	A/Scotland/ 76/2003 CY088123 - 30	A/England/ 567/2003 CY088347 - 54	A/Scotland/ 78/2003 CY088131 - 38	A/England/ 787/2003 CY088211 - 18	A/England/ 518/2003 CY088203 - 10

Supplemental Table 1. Human H3N2 sequences analyzed from 1995-2004 and the corresponding GenBank accession numbers. Human H3N2 sequences from 1995-2004 were downloaded from the Influenza Research Database and full-length genomes were concatenated and grouped into operational taxonomic units (numbered 1-12 under Cluster ID) with at least 97% sequence identity. Representative sequences were selected from these clusters for further analysis. Each vertical column indicates one replicate (seven replicates total).

500

H3N2 2005-2014	Representative strains analyzed in each replicate and the corresponding accession numbers.						
Cluster ID	1	2	3	4	5	6	7
1	*A/Indiana/ 21/2012 KJ942688 - 95	*A/Minnesota/ 11/2010 KJ942624 - 31	*A/Ontario/ 1252/2007 EU399751 - 8	*A/Ontario/ RV1273/2005 DQ469955 - 62	*A/Indiana/ 08/2011 JQ070793 - 800	*A/Indiana/ 10/2011 JN992753, JQ070778, JQ070780 - 1, JQ070792, KJ942594, KJ942598 - 9	*A/Iowa/ 07/2011 JQ290173 - 80
2	A/Singapore/ 139N/2007 KF533055, KF533061, KF836400, KF836413, KP223213, KP223232, KP223252, KP223178	A/Singapore/ 25Z/2007 KF533050, KF533056, KF836394, KF836409, KP223170, KP223205, KP223224, KP223244	A/Soledade/ LACENRS-820/2011 KY924952, KY925128, KY925167, KY925402, KY925635, KY925943, KY926297, KY926359	A/Nanjing/ 1654/2010 HQ664919, HQ664922 - 28	A/Nanjing/ 1655/2010 HQ664914 - 18, HQ664920 - 21, HQ664936	A/Nanjing/ 1663/2010 HQ664929 - 35, HQ664937	A/Qinghai/ 178/2010 KY049985 - 92
3	A/South Australia/ 22/2005 CY017131 - 8	A/Hong Kong/ HKU44/2005 CY039175 - 82	A/Quebec/ 26-281106/2006 KM438087, KM438105, KM438116, KM438129, KM438140, KM438152, KM438164, KM438176	A/New York/ 359/2005 CY002000 - 7	A/New York/ 391/2005 CY002056 - 63	A/New York/ 465/2005 CY006084 - 91	A/Quebec/ 26-031205/2005 KM438091, KM438101, KM438112, KM438125, KM438136, KM438148, KM438163, KM438175
4	A/Connecticut/ Flu086/2012 KM654605, KM654653, KM654723, KM654771, KM654793, KM654839, KM654891, KM654963	A/DC/ 1/2014 KU888850 - 7	A/Porto Alegre/ LACENRS-2376/2014 KY925119, KY925301, KY925321, KY925689, KY925806, KY926038, KY926104, KY926418	A/Santa Cruz/ 2821/2012 KF612183, KF612187, KF612190, KF612194, KF612198, KF612202, KF612206, KF612210	A/Canoas/ LACENRS-1317/2014 KY925875, KY925904, KY925907, KY924954, KY925991, KY926059, KY926164, KY926208	A/Arvorezinha/ LACENRS-2055/2013 KY925028, KY925428, KY925762, KY925949, KY925989, KY926006, KY926151, KY926211	A/Alegrete/ LACENRS-1624/2014 KY925158, KY925257, KY925311, KY925323, KY925352, KY925358, KY926147, KY926192
5	A/Saudi Arabia/ 05/2013 KY681482, KY681507, KY681532, KY681557, KY681582, KY681607, KY681632, KY681657	A/Kuwait/ F099/2014 KT888170, KT888246, KT888294, KT888454, KT888521, KT888867, KT889001, KT889268	A/Michigan/ 08/2014 KR057640 - 7	A/Nevada/ 18/2014 KT838445, KT838562, KT838565, KT838582, KT838592, KT838601, KT838697, KT838722	A/Maine/ 06/2014 KT837919, KT837922, KT837985, KT837992, KT838047, KT838108, KT838149, KT838177	A/Hawaii/ 39/2014 KR057629 - 36	A/Arkansas/ 01/2014 KT837141 - 2, KT837149, KT837166, KT837209, KT837220, KT837233, KT837259
6	A/New York/ 933/2006 CY019333, CY019336 - 40, CY019845, CY020102	A/South Carolina/ NHRC0001/2005 CY090989 - 96	A/Texas/ NHRC0001/2006 CY091037 - 44	A/Illinois/ NHRC0001/2005 CY091005 - 12	A/HaNoi/ ISBM16/2005 CY104212 - 9	A/Netherlands/ 42/2006 CY114389 - 96	A/Georgia/ NHRC0001/2005 CY091125 - 32
7	A/Queensland/ 52/2005 CY017989 - 96	A/Otago/ 1/2005 CY014039 - 46	A/New York/ 7/2006 CY014159 - 66	A/Taiwan/ 70002/2007 CY039399 - 406	A/New York/ 1009/2006 CY172231 - 8	A/Western Australia/ 69/2005 CY016979 - 86	A/Hong Kong/ HKU66/2005 CY039247 -54
8	A/Florida/ 22/2009 JX905397 - 9, JX905401 - 3, JX905407 - 8	A/New York/ 3738/2009 CY050820 - 7	A/Brisbane/ 10/2007 CY039089, CY039094, KJ609212 - 5, KJ609217 - 8	A/Thailand/ CU-1102/2008 EU625364, EU625367, FJ913003 - 8	A/Mexico/ 24025/2009 CY147723 - 30	A/Uruguay/ 716/2007 CY121635 - 9, EU716426 - 7, EU716434	A/Pennsylvania/ PIT04/2008 CY036927 - 34
9	A/Sydney/ 156/2014 CY252023 - 30	A/New York/ WC-LVD-14-073/2014 CY207355 - 62	A/Brisbane/ 341/2014 CY252143 - 50	A/Canberra/ 82/2014 CY251959 - 66	A/Nicaragua/ 7945_05/2014 CY240571 - 8	A/Houston/ JMM_170/2013 CY186179 - 86	A/Chicago/ YGA_04046/2012 CY171215 - 22
10	A/Canberra/ 35/2012 CY251175 - 82	A/Brisbane/ 21/2012 CY250865 - 72	A/Peru/ PER041/2011 CY160608 - 15	A/Parma/ 22/2012 KU322183, KU322225, KU322268, KU322311, KU322354, KU322397, KU322440, KU322483	A/South Australia/ 2/2013 CY251423 - 30	A/Victoria/ 120/2012 CY251151 - 8	A/Boston/ YGA_00046/2013 CY168167 - 74
11	A/Cambodia/ V1012345/2011 KU299790 - 7	A/Netherlands/ 034/2010 CY114509 - 16	A/Rhode Island/ 09/2012 CY264630, CY264632, CY264634 - 5, KF789834, KY116965, KY116967, KY116971	A/Chicago/ YGA_04114/2012 CY171479 - 86	A/Boston/ DOA2-097/2012 CY148300 - 7	A/Victoria/ 1003/2012 CY251375 - 82	A/South Australia/ 8/2013 CY251479 - 86
12	A/Texas/ JMM_24/2012 CY134900 - 7	A/Santa Clara/ YGA_03020/2013 CY170599 - 606	A/Nicaragua/ 6777_01_TR1/2013 CY189687 - 94	A/Nicaragua/ 30071_01_TR1/ 2013 CY189751 - 8	A/Boston/ DOA2-232/2013 CY149228 - 35	A/Texas/ JMM_21/2012 CY134876 - 83	A/Boston/ DOA2-245/2013 CY149316 - 23

Supplemental Table 2. Human H3N2 sequences analyzed from 2005-2014 and the corresponding GenBank accession numbers. Human H3N2 sequences from 2005-2014 were downloaded from the Influenza Research Database and full-length genomes were concatenated and grouped into operational taxonomic units (numbered 1-12 and labeled Cluster ID) with at least 97% sequence identity. Representative sequences were selected from these clusters for further analysis. Each vertical column indicates one replicate (seven replicates total). Asterisks denote H3N2v strains.

H1N1 2000-2008	Representative strains analyzed in each replicate and the corresponding accession numbers.						
Cluster ID	1	2	3	4	5	6	7
1	A/South Korea/ AF10/2008 CY044349 - 56	A/Taiwan/ 0045/2006 HQ291849, HQ291874, HQ291899, HQ291924, HQ291949, HQ291974, HQ291999, HQ292024	A/Taiwan/ 0586/2006 HQ291848, HQ291876, HQ291901, HQ291923, HQ291951, HQ291973, HQ291998, HQ292023	A/Taiwan/ 192/2006 HQ291869, HQ291875, HQ291900, HQ291944, HQ291950, HQ291994, HQ292019, HQ292044	A/Taiwan/ 2823/2008 HQ291853, HQ291880, HQ291907, HQ291928, HQ291957, HQ291978, HQ292003, HQ292028	A/Taiwan/ 2832/2008 HQ291862, HQ291883, HQ291908, HQ291937, HQ291958, HQ291987, HQ292012, HQ292037	A/Taiwan/ 8885/2008 HQ291856, HQ291889, HQ291914, HQ291931, HQ291964, HQ291981, HQ292006, HQ292031
2	A/New York/ 1066/2007 CY172655 - 62	A/Tennessee/ UR06-0076/2007 CY028107 - 14	A/Kentucky/ UR06-0258/2007 CY028163 - 70	A/Ohio/ UR06-0411/2007 CY027987 - 94	A/Oregon/ UR06-0609/2007 CY033465 - 72	A/Kansas/ UR06-0102/2007 CY031534 - 7, CY031551 - 4	A/Texas/ UR06-0323/2007 CY041442 - 9
3	A/Texas/ UR06-0503/2007 CY027883 - 90	A/Texas/ UR06-0542/2007 CY026339 - 46	A/Texas/ UR06-0468/2007 CY026499 - 506	A/Texas/ UR06-0467/2007 CY025445 - 52	A/Texas/ UR06-0420/2007 CY028219 - 26	A/Texas/ UR06-0026/2007 CY028323 - 30	A/Memphis/ 5/2003 CY019883 - 90
4	A/Netherlands/ 26/2007 HQ166041 - 8	A/New York/ 443/2001 CY003472 - 9	A/New York/ 220/2002 CY002528 - 35	A/New York/ 230/2003 CY002624 - 31	A/Hanoi/ ISBM31/2005 CY104822 - 9	A/TayNguyen/ TN334/2005 CY105350 - 7	A/Christchurch/ 1/2003 CY125188 - 95
5	A/Tottori/ 08T010/2008 CY043550 - 7	A/Nagasaki/ 07N035/2008 CY043486 - 93	A/Kyoto/ 07K454/2008 CY043454 - 61	A/Niigata/ 07F102/2008 CY043382 - 9	A/New York/ 08-1326/2008 JN582059 - 66	A/DaNang/ DN467/2008 CY104934 - 41	A/ThaiBinh/ TB289/2008 CY105126 - 33
6	A/Singapore/ 14/2001 CY125148 - 55	A/Malaysia/ 33166/2005 CY119186 - 93	A/Malaysia/ 34450/2006 CY119194 - 201	A/Hong Kong/ 2637/2004 CY125204 - 11	A/Canada/ 591/2004 CY125196 - 203	A/Canterbury/ 106/2004 CY007467 - 74	A/Malaysia/ 1706215/2007 CY119226 - 33
7	A/Gunma/ 07G002/2008 CY043406 - 13	A/Kyoto/ 07K303/2008 CY043438 - 45	A/Nagasaki/ 07N005/2008 CY043462 - 9	A/Niigata/ 07F191/2008 CY043398 - 405	A/Niigata/ 07F125/2008 CY043390 - 7	A/Hong Kong/ 1870/2008 CY121648 - 55	A/Hue/ H259/2008 CY104958 - 65
8	A/Chile/ 4795/2000 CY125132 - 39	A/South Australia/ 25/2000 CY017139 - 46	A/South Australia/ 26/2000 CY017379 - 86	A/South Australia/ 29/2000 CY017019 - 26	A/South Australia/ 30/2000 CY017813 - 20	A/South Australia/ 36/2000 CY017147 - 54	A/South Australia/ 64/2000 CY016731 - 8
9	A/Taiwan/ 141/2002 CY084228 - 35	A/Canterbury/ 152/2001 CY011088 - 95	A/Wellington/ 17/2001 CY012304 - 11	A/Canterbury/ 47/2001 CY010252 - 9	A/Russia/ 2187/2002 CY125164 - 71	A/New York/ 306/2001 CY003400 - 7	A/New York/ 205/2001 CY001952 - 9

Supplemental Table 3. Human H1N1 sequences analyzed from 2000-2008 and the corresponding GenBank accession numbers. Human H1N1 sequences from 2000-2008 were downloaded from the Influenza Research Database and full-length genomes were concatenated and grouped into operational taxonomic units (numbered 1-9 under Cluster ID) with at least 97% sequence identity. Representative sequences were selected from these clusters for further analysis. Each vertical column indicates one replicate (seven replicates total).

H1N1 2010-2018	Representative strains analyzed in each replicate and the corresponding accession numbers.						
Cluster ID	1	2	3	4	5	6	7
1	A/Gramado/ LACENRS-1287/2016 KY925104, KY925179, KY925188, KY925751, KY925798, KY925960, KY926084, KY926299	A/Taipei/ 0045/2016 KY487178, KY487296, KY487367, KY487426, KY487456, KY487565, KY487602, KY487714	A/Vacaria/ LACENRS-495/2016 KY925355, KY925365, KY925611, KY925636, KY925736, KY925871, KY926039, KY926287	A/Torres/ LACENRS-1102/2016 KY925032, KY925049, KY925190, KY925284, KY925584, KY926217, KY926257, KY926372	A/Baltimore/ 0107/2016 KY487084, KY487183, KY487439, KY487451, KY487472, KY487556, KY487562, KY487722	A/Taipei/ 0046/2016 KY487170, KY487436, KY487504, KY487531, KY487560, KY487635, KY487734, KY487740	A/Baltimore/ 0008/2016 KY487127, KY487343, KY487410, KY487491, KY487576, KY487604, KY487672, KY487698
2	A/Minnesota/ 35/2017 CY243771 - 8	A/Florida/ 95/2015 KX004656 - 63	A/Hawaii/ 83/2015 KX004967 - 74	A/Florida/ 83/2015 KX005095, KX005097 - 8, KX005100, KX005102, KX005472, KX005475, KX005477	A/Florida/ 12/2016 KX406328 - 35	A/Washington/ 01/2017 CY218075 - 82	A/North Carolina/ 12/2016 KX407448 - 55
3	A/Warsaw/ INS3_657/2011 CY176594 - 601	A/Khon Kaen/ INS3_649/2012 CY176538 - 45	A/Chicago/ YGA_04019/2012 CY171159 - 66	A/Tennessee/ F2068/2011 CY167660 - 7	A/Lima/ INS3_677/2012 CY176722 - 9	A/Bronx/ INS3_673/2012 CY176690 - 7	A/Georgia/ M5081/2012 CY147964, CY147998, CY148061, CY148064, CY148081, CY148159, CY148163, CY148218
4	A/Helsinki/ 848/2013 KF560110 - 7	A/Helsinki/ 804/2013 KF560054 - 61	A/Helsinki/ 911/2013 KF560158 - 65	A/Helsinki/ 100/2013 KF559390 - 7	A/Helsinki/ 1199/2012 KF560278 - 85	A/Helsinki/ 205/2013 KF559438 - 45	A/Helsinki/ 737/2013 KF559990 - 7
5	A/Parma/ 644/2010 KU322148, KU322190, KU322233, KU322276, KU322319, KU322362, KU322405, KU322448	A/Parma/ 7/2011 KU322178, KU322220, KU322263, KU322306, KU322349, KU322392, KU322435, KU322478	A/Parma/ 48/2011 KU322176, KU322218, KU322261, KU322304, KU322347, KU322390, KU322433, KU322476	A/Parma/ 126/2011 KU322173, KU322215, KU322258, KU322301, KU322344, KU322387, KU322430, KU322473	A/Singapore/ GP917/2011 CY124683 - 90	A/Singapore/ GP887/2011 CY124663 - 70	A/Singapore/ GP767/2011 CY124611 - 8
6	A/Bangkok/ INS478/2010 CY096298 - 305	A/Sydney/ DD3-37/2010 CY092638 - 45	A/Texas/ JMS413/2010 CY061187 - 94	A/Bangkok/ INS424/2010 CY071327 - 34	A/Vienna/ INS368/2010 CY071055 - 62	A/Berlin/ INS430/2010 CY071375 - 82	A/Athens/ INS396/2010 CY071167 - 74
7	A/NewYork/ A-WC-LVD-16-067/2016 CY259847 - 54	A/Alaska/ 263/2015 KU509765 - 7, KU509771, KX004187 - 8, KX004696, KX004702	A/NewYork/ A-WC-LVD-16-040/2016 CY259655 - 62	A/NewYork/ A-WC-LVD-16-018/2016 CY258913 - 20	A/NewYork/ A-WC-LVD-16-016/2016 CY259463 - 70	A/NewYork/ A-WC-LVD-16-053/2016 CY259751 - 8	A/NewYork/ A-WC-LVD-16-014/2016 CY259447 - 54
8	A/Mexico/ 8017/2017 MF593572 - 9	A/Mexico/ 4436/2016 MF593514 - 21	A/Mexico/ 8517/2017 MF593580 - 7	A/Mexico/ 4687/2017 MF593551 - 8	A/Mexico/ 4435/2016 MF593506 - 13	A/Mexico/ 4440/2016 MF593522 - 9	A/Mexico/ 17517/2017 MF593617 - 24
9	A/Thailand/ SN11783/2012 KP637595 - 602	A/Milano/ 21/2012 KU321908, KU321939, KU321971, KU322003, KU322033, KU322065, KU322097, KU322129	A/Taiwan/ 1829/2011 JQ693691, JQ693701, JQ693711, JQ693721, JQ693731 JQ693741, JQ693751, JQ693761	A/Indiana/ 04/2011 KC882315, KC882319 - 21, KC882343 - 5, KC882371	A/Michigan/ 03/2011 KC882264, KC882275 - 80, KC882296	A/Pennsylvania/ 07/2011 KC881821 - 4, KC882000, KC882022 - 3, KC882124	A/Budapest/ WRAIR3789T/2011 CY097977 - 84

Supplemental Table 4. Human H1N1 strains analyzed from 2010-2018 and the corresponding GenBank accession numbers. Human H1N1 sequences from 2010-2018 were downloaded from the Influenza Research Database and full-length genomes were concatenated and grouped into operational taxonomic units (numbered 1-9 under Cluster ID) with at least 97% sequence identity. Representative sequences were selected from these clusters for further analysis. Each vertical column indicates one replicate (seven replicates total).

A PB2

Sequence	Location
cagaagatacggaccagcat	193
attctcggaaaagatgccgg	297
cagttctcttcattgactgt	402
catttgataccacccagata	470
acagtgggtttgtcagaact	521
tttagttcctaaggccatta	550
ctcatcgtcaatgatgtggg	703
ttattatctcctgaggaggt	762
tcgagaccaacgtgggaatg	784
tggttagcattgatcggttt	812
gtgtggatgaatactccagt	845
taccagatatgactccaagc	905
agccgaagcaataatcgtgg	1114
gtgagtggaagggacgaaca	1140
gcagaacccgactgaagaac	1399
agtatcagcagatccactag	1483
attattgcagccaggaacat	1518
gactcaaggaacgtgttggg	1594
ggagccaggatactaacatc	1779
gtcaagatacgccgaagagt	1881
aatggacccgtgacaagtac	1992
cccgtcacttaggatgaaat	2170
tacacatcagggagacagga	2196

B NA

Sequence	Location
atgcctatataagctttcgc	30
tgatggggcggacatcaatc	52
attgttgtgttttgtggcac	102
aagtctggtggacctcaaac	125
gcatcaatcggtgcttttat	179
gtattttctctgttgaaggc	206
agaggtaatatgtccggtta	231
ggtccaaacctaactccaaa	281
accttcaaggtcattgaagg	303
agaagtcacgctcaggttat	326
agggggtcatggagtgaaag	397
aaaacgacagctccagcagt	446
agcattgtttccagttatgt	498
taggcccatcgtagatataa	532
gagacaactggaaaggctcc	554
gcaaattgtcaggaagtgct	632
acttttcattgaggaggga	667
gcttcaggaagagctgatac	693
tcggaatgcgtttgtatcaa	747
aaaaatcctcaggacccagg	769
atagtattggttcatggtcc	791
gcttcatttacgatgggaga	818
ctgggcatgatgaaaatgca	845
aaagcatggctgcatgtttg	870
cagctcaagttgtcacgatg	892
tttgggaaccaagcaagtgt	925
atgaatgagttgggtgttcc	951
gcgatcctgacaagtgttat	1058
taattcaattcggctttccg	1120
aggatttgcacctttttcta	1144
tggtcaaagccgcaatgtaa	1170
tagtgtatctgaccaacacc	1235
acaaccaagtaatgctgtgt	1289
aatagccatcctggtaacta	1354
cactattgccacaatatgct	1384
taacgattggctctgtttct	1406

C NS

Sequence	Location
gctgtttgaagtggaacagg	56
ggctgattgaagaagtgaga	138
aatggcgagaacagctaggt	180
tctacagagattcgcttgga	252
ctcaccattgccttcttttc	363
agggagcagttgttggcgaa	385
tgaccgactagagaccatag	429
gacccctttgcatcagaatg	505
atgcccaagcaaaaagtgga	527
ctgcttcgcgatacataact	589
taaaatgaccatggcctcca	612
aggaagaatctgatgaggca	634
cccatgttggaaagcagatt	670
ggtctagacatcaaagcagc	692
aaggggaagaggcaacact c	714
tttccaggtagattgctttc	819
atggattccaacactgtgtc	842

Supplemental Table 5. FISH probe sequences. Custom oligonucleotide probes targeting PB2 **(A)**, NA **(B)**, and NS **(C)** vRNA were designed from A/Panama/2007/1999 (H3N2) virus sequences using the Stellaris probe designer (BioSearch Technologies). Oligos exhibiting significant complementarity against other vRNA segments and/or positive-strand complementarity were excluded.

Acknowledgments

All confocal microscopy imaging was performed at the Center for Biologic Imaging at the University of Pittsburgh. JEJ is supported by a T32 (T32 AI049820) and the Catalyst Award (University of Pittsburgh Center for Evolutionary Biology and Medicine). This work is funded by the National Institutes of Health NIAID (R01 AI139063). We thank members of the Lakdawala and Wright labs for technical support and constructive feedback on this manuscript.

Competing Interests

The authors have no competing interests to disclose.

References

Andino, R., & Domingo, E. (2015). Viral quasispecies. *Virology, 479-480*, 46-51. doi:10.1016/j.virol.2015.03.022

Bedford, T., Riley, S., Barr, I. G., Broor, S., Chadha, M., Cox, N. J., . . . Russell, C. A. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature, 523*(7559), 217-220. doi:10.1038/nature14460

Bloom, J. D., Gong, L. I., & Baltimore, D. (2010). Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science (New York, N.Y.), 328*(5983), 1272-1275. doi:10.1126/science.1187816

Brooke, C. B. (2017). Population Diversity and Collective Interactions during Influenza Virus Infection. *Journal of Virology, 91*(22), e01164-01117. doi:10.1128/JVI.01164-17

Dadonaite, B., Gilbertson, B., Knight, M. L., Trifkovic, S., Rockman, S., Laederach, A., . . . Bauer, D. L. V. (2019). The structure of the influenza A virus genome. *Nature microbiology, 4*(11), 1781-1789. doi:10.1038/s41564-019-0513-7

Escalera-Zamudio, M., Golden, M., Gutiérrez, B., Thézé, J., Keown, J. R., Carrique, L., . . . Pybus, O. G. (2020). Parallel evolution in the emergence of highly pathogenic avian influenza A viruses. *Nature Communications, 11*(1), 5511. doi:10.1038/s41467-020-19364-x

Fodor, E. (2013). The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta Virol, 57*(2), 113-122. doi:10.4149/av_2013_02_113

Garten, R. J., Davis, C. T., Russell, C. A., Shu, B., Lindstrom, S., Balish, A., . . . Cox, N. J. (2009). Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science (New York, N.Y.), 325*(5937), 197-201. doi:10.1126/science.1176225

Gavazzi, C., Yver, M., Isel, C., Smyth, R. P., Rosa-Calatrava, M., Lina, B., . . . Marquet, R. (2013). A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proceedings of the National Academy of Sciences, 110*(41), 16604. doi:10.1073/pnas.1314419110

Gong, L. I., Suchard, M. A., & Bloom, J. D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife, 2*, e00631-e00631. doi:10.7554/eLife.00631

548 Jang, J., & Bae, S.-E. (2018). Comparative Co-Evolution Analysis Between the HA and NA
549 Genes of Influenza A Virus. *Virology : research and treatment, 9*, 1178122X18788328-
550 11178122X18788328. doi:10.1177/1178122X18788328
551 Kryazhimskiy, S., Dushoff, J., Bazykin, G. A., & Plotkin, J. B. (2011). Prevalence of Epistasis in
552 the Evolution of Influenza A Surface Proteins. *PLOS Genetics, 7*(2), e1001301.
553 doi:10.1371/journal.pgen.1001301
554 Lakdawala, S. S., Fodor, E., & Subbarao, K. (2016). Moving On Out: Transport and Packaging
555 of Influenza Viral RNA into Virions. *Annual Review of Virology, 3*(1), 411-427.
556 doi:10.1146/annurev-virology-110615-042345
557 Lakdawala, S. S., Lamirande, E. W., Suguitan, A. L., Jr., Wang, W., Santos, C. P., Vogel, L., . . .
558 Subbarao, K. (2011). Eurasian-Origin Gene Segments Contribute to the Transmissibility,
559 Aerosol Release, and Morphology of the 2009 Pandemic H1N1 Influenza Virus. *PLOS*
560 *Pathogens, 7*(12), e1002443. doi:10.1371/journal.ppat.1002443
561 Lakdawala, S. S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A. J., Lamirande, E. W., . . .
562 Subbarao, K. (2014). Influenza A Virus Assembly Intermediates Fuse in the Cytoplasm.
563 *PLOS Pathogens, 10*(3), e1003971. doi:10.1371/journal.ppat.1003971
564 Le Sage, V., Kanarek, J. P., Snyder, D. J., Cooper, V. S., Lakdawala, S. S., & Lee, N. (2020).
565 Mapping of Influenza Virus RNA-RNA Interactions Reveals a Flexible Network. *Cell reports,*
566 *31*(13). doi:10.1016/j.celrep.2020.107823
567 Le Sage, V., Nanni, A. V., Bhagwat, A. R., Snyder, D. J., Cooper, V. S., Lakdawala, S. S., &
568 Lee, N. (2018). Non-Uniform and Non-Random Binding of Nucleoprotein to Influenza A and
569 B Viral RNA. *Viruses, 10*(10), 522. doi:10.3390/v10100522
570 Lee, N., Le Sage, V., Nanni, A. V., Snyder, D. J., Cooper, V. S., & Lakdawala, S. S. (2017).
571 Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids*
572 *Research, 45*(15), 8968-8977. doi:10.1093/nar/gkx584
573 Lowen, A. C. (2017). Constraints, Drivers, and Implications of Influenza A Virus Reassortment.
574 *Annual Review of Virology, 4*(1), 105-121. doi:10.1146/annurev-virology-101416-041726
575 Lyons, D. M., & Lauring, A. S. (2018). Mutation and Epistasis in Influenza Virus Evolution.
576 *Viruses, 10*(8), 407. doi:10.3390/v10080407
577 Majarian, T. D., Murphy, R. F., & Lakdawala, S. (2018). Learning the sequence of influenza A
578 genome assembly during viral replication using point process models and fluorescence in
579 situ hybridization. *bioRxiv*, 322966. doi:10.1101/322966
580 Neumann, G., Noda, T., & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-
581 origin H1N1 influenza virus. *Nature, 459*, 931. doi:10.1038/nature08157
582 Neverov, A. D., Kryazhimskiy, S., Plotkin, J. B., & Bazykin, G. A. (2015). Coordinated Evolution
583 of Influenza A Surface Proteins. *PLOS Genetics, 11*(8), e1005404.
584 doi:10.1371/journal.pgen.1005404
585 Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., & Kawaoka, Y. (2006).
586 Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature,*
587 *439*(7075), 490-492. doi:10.1038/nature04378
588 Nturibi, E., Bhagwat, A. R., Coburn, S., Myerburg, M. M., & Lakdawala, S. S. (2017).
589 Intracellular Colocalization of Influenza Viral RNA and Rab11A Is Dependent upon
590 Microtubule Filaments. *Journal of Virology, 91*(19), e01179-01117. doi:10.1128/JVI.01179-
591 17
592 Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and
593 evolutionary analyses in R. *Bioinformatics (Oxford, England), 35*(3), 526-528.
594 doi:10.1093/bioinformatics/bty633
595 Paules, C., & Subbarao, K. (2017). Influenza. *The Lancet, 390*(10095), 697-708.
596 doi:https://doi.org/10.1016/S0140-6736(17)30129-0

Reed, L. J., & Muench, H. (1938). A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12. *American Journal of Epidemiology, 27*(3), 493-497. doi:10.1093/oxfordjournals.aje.a118408

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution, 3*(2), 217-223. doi:https://doi.org/10.1111/j.2041-210X.2011.00169.x

Richard, M., Herfst, S., Tao, H., Jacobs, N. T., & Lowen, A. C. (2018). Influenza A Virus Reassortment Is Limited by Anatomical Compartmentalization following Coinfection via Distinct Routes. *Journal of Virology, 92*(5). doi:10.1128/jvi.02063-17

Robinson, D. F., Foulds, L.R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences, 53*(1-2), 131-147.

Sakai, T., Nishimura, S. I., Naito, T., & Saito, M. (2017). Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Scientific Reports, 7*, 45043-45043. doi:10.1038/srep45043

Sardi, M., & Gasch, A. P. (2018). Genetic background effects in quantitative genetics: gene-by-system interactions. *Current Genetics, 64*(6), 1173-1176. doi:10.1007/s00294-018-0835-7

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics (Oxford, England), 27*(4), 592-593. doi:10.1093/bioinformatics/btq706

Short, K. R., Kedzierska, K., & van de Sandt, C. E. (2018). Back to the Future: Lessons Learned From the 1918 Influenza Pandemic. *Frontiers in cellular and infection microbiology, 8*, 343-343. doi:10.3389/fcimb.2018.00343

Wright, E. S. (2015). DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics, 16*(1), 322. doi:10.1186/s12859-015-0749-z

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*(4), 954-959. doi:10.1093/biomet/87.4.954

Zhang, Y., Aevermann, B. D., Anderson, T. K., Burke, D. F., Dauphin, G., Gu, Z., . . . Scheuermann, R. H. (2016). Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Research, 45*(D1), D466-D474. doi:10.1093/nar/gkw857