Collective interactions augment influenza A virus replication in a host-dependent manner

Kara L. Phipps[a], Ketaki Ganti[a‡], Nathan T. Jacobs[a‡], Chung-Young Lee[a‡], Silvia Carnaccini[b], Maria C. White[a], Miglena Manandhar[c], Brett E. Pickett[c†], Gene S. Tan[c,d,f], Lucas M. Ferreri[b], Daniel R. Perez[b,e] and Anice C. Lowen[a,d]*

[a]Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia, United States of America

[b]Department of Population Health, Poultry Diagnostic and Research Center, University of Georgia, Athens, Georgia, United States of America.

[c]J. Craig Venter Institute, La Jolla, California, United States of America

[d]Emory-UGA Center of Excellence for Influenza Research and Surveillance (CEIRS), Atlanta, Georgia, United States of America

[e]Center of Research on Influenza Pathogenesis (CRIP) CEIRS, New York, New York, United States of America.

[f]Division of Infectious Disease, Department of Medicine, University of California San Diego, La Jolla, CA, United States of America

[‡]These individuals contributed equally to the research presented.

[†]Present address: Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah 84602 USA

*Address correspondence: anice.lowen@emory.edu

Abstract

Infection with a single influenza A virus (IAV) is only rarely sufficient to initiate productive infection. Here, we exploit both single-cell approaches and whole-animal systems to show that the extent of IAV reliance on multiple infection varies with virus strain and host species. Influenza A/guinea fowl/HK/WF10/99 (H9N2) [GFHK99] virus exhibits much stronger dependence on collective interactions in mammalian compared to avian systems, indicating a role for host factors. Quantification of incomplete viral genomes showed that their complementation accounts for the more moderate reliance on coinfection seen in avian cells, but not the added reliance seen in mammalian cells. This finding suggests an additional form of virus-virus interaction is needed to support GFHK99 virus infection in mammalian cells. Genetic mapping implicated the PA gene segment as a major driver of this phenotype and quantification of viral RNA synthesis indicated that both replication and transcription were affected. These findings indicate that multiple distinct mechanisms underlie IAV reliance on coinfection and underscore the importance of virus-virus interactions in IAV infection, evolution and emergence.

Introduction

Classically, an infectious unit has been defined as a single virus particle which delivers its genome to a cell, initiates the viral reproductive program, and yields progeny viruses. The importance to infection of collective interactions among viruses is being increasingly recognized, however[1-3]. The delivery of multiple viral genomes to a cell allows antagonistic and beneficial interactions to occur, and these interactions have the potential to shape transmission, pathogenicity, and viral evolution.

Recent work has revealed several distinct mechanisms by which multiple viral genomes are co-delivered to a target cell[4-8]. Various mechanisms of direct cell-to-cell spread yield a similar outcome[9-11]. The implications of multiple infection in these diverse systems are still being explored. In a number of cases, however, collective delivery was demonstrated to increase the efficiency of infection relative to free virus particles[4,5], or to increase the rate of genetic exchange through recombination[6].

When distinct variants coinfect, beneficial effects such as compensation of deleterious mutations can increase overall fitness[12-14]. In the case of IAV, several lines of evidence point to a major role in infection for multiplicity reactivation, the process by which segmented genomes lacking one or more functional segments complement each other[15-19]. Conversely, negative interactions can also arise in which deleteriously mutated genes act in a dominant negative fashion. Defective interfering particles, which often potently interfere with the production of infectious progeny from a coinfected cell, are the most extreme example of such antagonism[20-22].

Importantly, multiple infection with identical viral genomes can also alter infection outcomes. Such cooperation was documented for VSV and HIV, where rates of transcription and replication were enhanced with increasing multiplicity of infection (MOI)[23,24]. Similarly, faster kinetics of virus production were seen at high MOI for poliovirus and IAV[19,25]. In these instances, it is thought that increased copy number of infecting viral genomes provides a kinetic benefit

important in the race to establish infection before innate antiviral responses take hold. Indeed, it has been suggested that multiple infection may be particularly relevant for facilitating viral growth under adverse conditions, such as antiviral drug treatment[3,26].

For IAV, an important adverse condition to consider is that of a novel host environment. IAVs occupy a broad host range, including multiple species of wild waterfowl, poultry, swine, humans, and other mammals[27,28]. Host barriers to infection typically confine a given lineage to circulation in one species or a small number of related species[29,30]. Spillovers occur occasionally, however, and can seed novel lineages. When a novel IAV lineage is established in humans, the result is a pandemic of major public health consequence[31,32]. The likelihood of successful cross-species transfer of IAV is determined largely by the presence, absence, and compatibility of host factors on which the virus relies to complete its life cycle, and on the viruses' ability to overcome antiviral defenses in the novel host[33-35].

Our objective herein was to assess the degree to which IAV relies on the delivery of multiple viral genomes to a cell to ensure production of progeny. In particular, we sought to determine whether this phenotype varies with host species and with virus strain. We therefore examined the multiplicity dependence of influenza A/Netherlands/602/2009 (H1N1) [NL09] virus and two highly divergent avian-origin viruses, influenza A/mallard/Minnesota/199106/99 (H3N8) [MaMN99] and influenza A/guinea fowl/Hong Kong/WF10/99 (H9N2) [GFHK99] in multiple host systems. Results from all virus/cell combinations tested confirm prior reports that cells multiply-infected with IAV produce more viral progeny than singly-infected cells. Importantly, however, the proportion of viral progeny that emerge from coinfected cells varies greatly with virus-host context. The GFHK99 strain exhibits an acute dependence on multiple infection in mammalian cells that is not seen for NL09 or MaMN99 viruses in mammalian cells or for GFHK99 virus in avian cells. The PA gene segment of the GFHK99 virus drives its host-specific dependence on multiple infection. In line with this finding, both bulk and single-cell measurements of viral RNA showed that polymerase activity in mammalian cells is enhanced with multiple infection. A need

for complementation of incomplete viral genomes partially accounts for this cooperative effect. Importantly, however, single-cell data indicate that additional multiplicity-dependent mechanisms support RNA synthesis by the GFHK99 virus in mammalian cells. Thus, our data point to an important role for multiple infection in determining the potential for IAV replication in diverse hosts.

Results

Degree of multiplicity dependence varies with strain and cell type

To evaluate the extent to which IAV relies on multiple infection for productive infection, we initially used reassortment and coinfection as readouts. While singly infected cells yield only parental progeny, reassortment is highly efficient within coinfected cells[36]. Monitoring reassortment therefore indicates the proportion of progeny in a bulk virus population that were produced from multiply infected cells. To ensure accurate quantification of reassortment, coinfections were performed under single-cycle conditions with homologous viruses that differ only by a silent mutation in each segment and the presence of either an HA or HIS epitope tag fused to the HA protein. The silent mutations allow identification of reassortant viruses. The epitope tags enable quantification of cellular coinfection across a range of MOIs. Homologous coinfection partners, termed WT and VAR, were generated in NL09, MaMN99 and GFHK99 strain backgrounds.

Coinfection and reassortment were examined in canine MDCK and chicken DF-1 cells (**Figure 1**). In MDCK cells, GFHK99 viruses showed a near linear relationship between total HA^+ cells and dual-HA^+ cells, suggesting a strict dependence on multiple infection for HA expression (**Figure 1A**). By contrast, HA expression resulting from infection with a single strain was more common for GFHK99 in DF-1 cells, MaMN99 in MDCK cells and NL09 in MDCK cells (**Figure 1A**). To give a quantitative measure of the differences among these virus/cell pairings, we determined the degree of linearity from the regression models of each dataset (**Figure 1B**).

The more dependent expression of *any* HA protein is on coinfection, the more linear the relationship between dual-HA^+ cells and HA^+ cells becomes. Conversely, a relationship closer to the Poisson expectation indicates less dependence on coinfection. We found that only GFHK99 virus in MDCK cells exhibits strong linearity in the relationship between HA^+ and dual-HA^+ cells (**Figure 1B**).

In line with observed levels of coinfection, GFHK99 virus exhibits high levels of reassortment in MDCK cells, indicating that nearly all progeny virus is produced from WT-VAR_1 coinfected cells (**Figure 1C**). MaMN99 or NL09 viruses infecting MDCK cells show relatively lower levels of reassortment. Moreover, reassortment of GFHK99 viruses is markedly reduced in DF-1 cells compared to MDCK cells (**Figure 1C**). These results clearly reveal differing degrees of multiplicity dependence for different virus/cell pairings and therefore indicate that multiple infection dependence is determined through virus-host interactions.

That all virus-cell pairings tested show evidence of multiplicity dependence is highlighted by comparison of the experimental reassortment data to a theoretical prediction (**Figure 1C**). This theoretical prediction was published previously[18] and is derived from a computational model in which the number of viral progeny produced by an infected cell is constant, irrespective of the number of input viral genomes. Because singly- and multiply-infected cells make equivalent numbers of progeny in this model, reassortment is predicted to increase only gradually at low levels of infection where coinfection is relatively rare. The experimental observation that reassortment reaches high levels much more rapidly than expected indicates that viral progeny production is enriched in the proportion of the infected cell population that is multiply infected.

Frequent reassortment and the linear relationship between overall HA expression and HA co-expression provide two measures of multiplicity dependence for a given virus-cell pairing. To summarize these two measures and analyze the relationships between the virus-cell pairings tested, we conducted a hierarchical clustering analysis using the parameters of the regression

models shown in **Figure 1A** and **1C**. MaMN99 and NL09 viruses in MDCK cells, as well as GFHK99 virus in DF-1 cells, represent one cluster, while GFHK99 virus in MDCK cells partitioned into its own cluster (**Figure 1D**). Coinfection dependence, therefore, is evident in all virus-cell pairings, but particularly strong for GFHK99 virus in MDCK cells.

***Strain- and host-specific phenotypes are also evident* in vivo**

To determine whether host-dependent reliance on multiple infection extended to *in vivo* models, we performed coinfections in guinea pigs and quail. GFHK99, MaMN99, and NL09 viruses were tested in guinea pigs and GFHK99 virus was examined in quail. To ensure use of comparable effective doses for each virus/host pairing, the 50% infectious dose (ID_{50}) of each virus mixture was first determined experimentally in the animal models used. Guinea pigs were then infected intranasally with 10^2 $GPID_{50}$ of each WT/VAR mixture and nasal washes were collected daily. Quail were infected with 10^2 QID_{50} of the GFHK99 virus mixture via an oculo-naso-tracheal route and tracheal swabs were collected daily. To evaluate the frequency of reassortment, plaque isolates from these upper respiratory samples were genotyped for each animal on each day. Because multicycle replication *in vivo* allows the propagation of reassortants, analysis of genotypic diversity rather than percent reassortment is more informative for these experiments. Thus, the effective diversity (Hill's N_2) was calculated for each dataset and plotted as a function of time post-inoculation (**Figure 1E, 1F**). The viruses collected from GFHK99-infected guinea pigs show much higher genotypic diversity throughout the course of infection than viruses isolated from MaMN99- or NL09-infected guinea pigs (**Figure 1E**) or GFHK99-infected quail (**Figure 1F**). These data indicate that the virus-host interactions which determine dependence on multiple infection in cell culture extend to *in vivo* infection.

Multiple infection enhances viral growth

The abundant reassortment observed with GFHK99 viruses in mammalian systems suggests that multiple infection plays a major role in determining the productivity of an infected cell. We therefore hypothesized that increasing MOI would augment the viral output of infected cells and that the magnitude of this effect would be greater for GFHK99 virus in MDCK cells than for GFHK99 virus in DF-1 cells or MaMN99 and NL09 viruses in MDCK cells. To test this prediction, we infected over a range of MOIs and then measured PFU produced per cell under single-cycle conditions. The input MOIs for this series of experiments were based on PFU as determined in MDCK cells. Which MOI conditions were saturating was determined empirically by flow cytometry in MDCK and DF-1 cells (**Supplementary Figure 1**). MOIs < 1 PFU/cell were found to be non-saturating. Under these non-saturating conditions, increasing MOI resulted in accelerated viral growth and higher burst size for all virus/cell pairings (**Figure 2A–C**). As predicted, however, increasing the MOI of GFHK99 virus in MDCK cells resulted in a further enhancement of viral amplification (**Figure 2D**). Thus, the benefit conferred by multiple infection was greater for GFHK99 virus in MDCK cells compared to the other virus-cell pairings tested.

We reasoned that the cooperative effect observed might result from i) complementation of incomplete viral genomes or ii) a benefit of increased viral genome copy number per cell. In an effort to differentiate between these possibilities, we measured growth of GFHK99 virus in MDCK and DF-1 cells infected over range of saturating MOIs. Note that plaque assays in MDCK cells underestimate the infectious potential of GFHK99 virus, and increasing viral MOI beyond 1 PFU/cell did not lead to increased levels of HA+ cells, indicating saturation (**Supplementary Figure 1**). Under these conditions, incomplete viral genomes are unlikely to be prevalent and any benefit of increasing MOI would be attributable to increasing genome copy numbers per cell. In both cell types, MOIs between 1 and 20 PFU per cell result in similar peak viral titers (**Figure 2E, F, H**). This saturation of cooperation at higher MOIs suggests diminishing returns from additional genome copies above a certain threshold. Whether this threshold is imposed by a need for complementation or another mechanism sensitive to saturation remained unclear.

The viral PA segment is a major determinant of multiple infection dependence

To identify viral genetic determinants of multiple infection dependence, we mapped segments responsible for the high reassortment phenotype of GFHK99 virus in MDCK cells. One or more genes from GFHK99 virus was placed into a MaMN99 virus background. For each chimeric genotype, WT and VAR strains were generated to allow tracking of homologous reassortment. Coinfections were performed in MDCK cells and HA expression and reassortment were measured as in **Figure 1**. Two viruses are similar to the parental GFHK99 virus in the levels of dual HA positivity observed: MaMN99:GFHK99 3PNP and MaMN99:GFHK99 PA (**Figure 3A–B**). Quantification of reassortment revealed that all chimeric viruses reassort at a higher frequency than MaMN99 parental strains, but that the MaMN99:GFHK99 3PNP and MaMN99:GFHK99 PA viruses show the highest reassortment, comparable to that seen for the parental GFHK99 genotype (**Figure 3C-D**). A clustering analysis similar to the one shown in **Figure 1D** revealed that only these two genotypes cluster with GFHK99 parental virus, while each of the other genotypes clusters with MaMN99 virus (**Figure 3E**). Thus, while other viral genes may contribute, PA is the primary genetic determinant of the high reassortment exhibited by GFHK99 virus in MDCK cells and defines a need for cooperation between coinfecting viruses.

Multiple infection enhances viral RNA replication

Because mapping of the GFHK99 high reassortment phenotype implicated the viral PA, we tested the effects of multiple infection on viral genome replication. We measured WT viral RNA synthesis in the absence and presence of increasing amounts of a homologous coinfecting virus. To evaluate host specificity, we did this analysis in both MDCK cells and cells appropriate for each virus strain: DF-1 cells for GFHK99 virus, duck embryo cells for MaMN99 virus, and human trachea-bronchial epithelial (HTBE) cells for NL09 virus. GFHK99 virus was also tested

in HTBE cells. Cells were infected with low MOI of WT virus to ensure receipt of a single copy of the WT virus genome. Concurrently, cells were infected with increasing doses of a VAR virus. Digital droplet PCR (ddPCR) was then used to quantify replication of WT genomes using primer/probe sets that do not detect VAR templates.

The results for GFHK99 virus in both cell types show that coinfection with low to moderate doses of VAR virus increases levels of WT vRNA, while a suppressive effect is observed at high doses (**Figure 4A**). Importantly, the amount of coinfection required to reach maximal vRNA production is 10-fold greater in MDCK cells (0.5–1 PFU per cell) than in DF-1 cells (0.05–0.1 PFU per cell). The maximal impact of the VAR virus on WT vRNA production is also greater in MDCK cells: a ~60-fold enhancement is seen, compared to only ~2-fold in DF-1 cells. To verify these observations in a more physiologically relevant system, we repeated the experiment in HTBE cells differentiated at an air-liquid interface. Similar to MDCK cells, these primary human cells exhibit maximal GFHK99 WT RNA production with addition of 0.5–1 PFU per cell VAR virus, and peak RNA replication is ~20 fold higher than that seen without coinfecting virus (**Figure 4B**).

Results for NL09 virus in MDCK and HTBE cells, as well as MaMN99 virus in duck embryo and MDCK cells, revealed little to no detectable benefit of homologous virus coinfection (**Figure 4C–E**). Addition of VAR virus was seen to suppress WT vRNA production in all virus/cell pairings, but the dose of VAR at which this antagonistic effect was noted varied.

Thus, for GFHK99 virus, the introduction of coinfecting VAR virus reveals a cooperative effect acting at the level of RNA synthesis, which is particularly potent in mammalian cells. That the replication of NL09 and MaMN99 viruses was not detectably enhanced by the addition of VAR virus suggests a reduced reliance on multiple infection, consistent with reassortment and dual-HA positivity data presented above.

Multiple infection accelerates viral replication and transcription

To more finely assess the effects of multiplicity on polymerase function, we measured vRNA and mRNA over time following low or high MOI infection (**Figure 5**). GFHK99 virus was examined in DF-1 and MDCK cells and MaMN99 virus was tested in MDCK cells. To ensure viral genomes were supplied as low or single copies, a dose of 0.5 RNA copies per cell was used for low MOI infection. Under these low MOI conditions in MDCK cells, GFHK99 genomic RNA levels remain low throughout the time course. GFHK99 mRNA increases early, perhaps reflecting robust primary transcription, then accumulates at a relatively slow rate from 2 h post-infection onward (**Figure 5A**). Both viral RNA species accumulate at a significantly higher rate for GFHK99 virus in DF-1 cells and MaMN99 virus in MDCK cells (**Figure 5B, C, G**). For the high MOI dose, we used 3.0 HA-expressing units per cell, as determined by flow cytometry (**Supplementary Figure 2**). This measure of infectivity gives a functional readout for polymerase activity, and therefore ensures the vast majority of cells carry an active viral polymerase. At this high MOI, accumulation of GFHK99 mRNA and vRNA in MDCK cells occurs at a similar rate to that seen for GFHK99 virus in DF-1 cells or MaMN99 virus in MDCK cells (**Figure 5D–F, H**). Thus, a host-specific defect in GFHK99 polymerase activity that affects both replication and transcription is seen at low MOI. This defect is, however, resolved under conditions where multiple infection is prevalent.

Single-cell mRNA sequencing reveals a need for cooperation

To evaluate the heterogeneity of viral RNA synthesis at the single-cell level, we infected DF-1 or MDCK cells with GFHK99 virus under single-cycle conditions and collected cells at 8 h post-infection for mRNA barcoding on the 10x Genomics Chromium platform prior to sequencing. Cells in which no viral mRNA was detected were excluded from further analysis. In addition, we used a thresholding approach to exclude cells that were unlikely to be truly infected, but rather may carry small amounts of viral mRNA owing to lysis of neighboring cells (**Supplementary Figure 5**). Following these two exclusion steps, the number of cells remaining ranged between

110 and 440 per infection condition. Next, to allow comparison between cell types and between experiments, each cell's total transcript abundance was normalized to the median number of transcripts per cell in the relevant sample. The normalized and $\log_{10}$ transformed counts of viral transcripts per cell were then compared across conditions. We found that the amount of detected GFHK99 viral mRNA varies widely between individual DF-1 cells, which is consistent with previous observations[37-39]. In contrast, GFHK99 viral mRNA levels are uniformly low in MDCK cells, especially at the lower MOIs tested (**Figure 6A**, left facet).

To evaluate whether low transcript abundance corresponded with the failure to detect one or more polymerase-encoding segments, we stratified the data based on detection of all four segments necessary to support transcription (PB2, PB1, PA and NP). Consistent with prior reports[37], viral transcript levels are higher in cells that contained PB2, PB1, PA, and NP transcripts compared to those in which one or more of these transcripts was not detected (**Supplementary Figure 6**). We noted that the PB2, PB1, and PA mRNAs were markedly less abundant than the other viral transcripts, however. Thus, the correspondence between failure to detect polymerase transcripts and low viral mRNA abundance may simply be due to the higher likelihood of these mRNAs falling below the limit of detection in cells that have low overall viral transcription.

To measure the impact of coinfecting viruses in individual cells, we repeated the single-cell sequencing experiment with the addition of two variants of GFHK99 virus, GFHK99 $mVAR_1$ and GFHK99 $mVAR_2$. Cells were inoculated with GFHK99 WT and GFHK99 $mVAR_1$ viruses in a 1:1 ratio. Coinfection with GFHK99 $mVAR_2$ virus was performed simultaneously and this virus was used at the concentration found to give optimal support of WT viral RNA replication in **Figure 4A**: 0.1 PFU per cell in DF-1 cells and 1.0 PFU per cell in MDCK cells. After mRNA sequencing, only cells in which transcripts from all eight $mVAR_2$ virus segments were detected were analyzed further. Cells that likely obtained WT or $mVAR_1$ viral RNA through lysis were excluded (**Supplementary Figure 5**). Between 64 and 217 cells per infection condition met

both criteria to be included in further analysis. As expected, no significant difference was detected between WT and $mVar_1$ transcript levels ($p > 0.05$, linear mixed effects model) (**Supplementary Figure 6**). No differentiation is therefore made between strains in further analysis. In cells infected with both WT and $mVAR_1$ viruses, viral transcript abundance is calculated separately for each, so that coinfected cells contribute two distinct measurements to the dataset. The viral transcript levels per cell detected in this second experiment are shown in **Figure 6A** alongside data from the first experiment. In comparing the two infections, we observe that total viral transcript abundance is markedly lower in MDCK cells compared to DF-1 cells in the first infection, but this effect is almost entirely mitigated by the presence of $mVAR_2$ virus in the second infection. This reduction in the disparity between DF-1 and MDCK cells resulted from the fact that $mVAR_2$ virus increased transcript abundance by 149% in DF-1 cells, but 258% in MDCK cells ($p < 10^{-4}$, linear mixed effects model) (**Figure 6A)**. These data underscore the significance of viral collective interaction to ensure productive infection in diverse hosts.

Because only a subset of a cell's transcripts is captured and therefore reliably detected[40], the 10x platform does not allow a robust determination of segment presence or absence in a cell. The addition of $mVAR_2$ virus improves the sensitivity with which WT transcripts can be detected, however, and – as noted above – levels the playing field between MDCK and DF-1 cells. We therefore evaluated the number of distinct viral gene segments detected per cell, as indicated by the presence of the corresponding mRNAs. Failure to detect a subset of the eight segments in a given cell is very common in our datasets, as with others[37,39]. The addition of $mVAR_2$ virus increases the number of segments per cell detected, however, suggesting that failure to detect is often a result of transcript levels falling below the limit of detection of the assay (**Figure 6B**). Importantly, in the presence of $mVAR_2$ virus, the number of segments detected per cell was comparable between MDCK and DF-1 cells across all MOIs tested (**Figure 6B**). This result suggests that the frequency with which incomplete GFHK99 viral genomes are expressed is comparable between these mammalian and avian cell lines.

Frequency of incomplete GFHK99 genomes in MDCK cells is moderate

Given the limitations of single-cell mRNA sequencing for the detection of relatively rare transcripts, we employed a second single-cell approach to quantify the frequency with which fewer than eight vRNAs are replicated in GFHK99 infected MDCK cells[19]. MDCK cells were coinfected with a low MOI of GFHK99 WT virus and a high MOI of GFHK99 VAR_2 virus. The GFHK99 VAR_2 virus acts to ensure propagation of the WT virus gene segments, even when less than the full WT viral genome is available for transcription and replication. Following inoculation, single cells were sorted into wells which contain a naïve cell monolayer and multicycle replication was allowed for 48 h. To determine which viral gene segments were present in the initially sorted cell, RT-qPCR with primers that differentiate WT and VAR_2 gene segments was applied. As detailed in the Methods, the results were used to calculate the probability that a cell infected with a single WT virus contains a given segment. We termed the resultant parameter Probability Present (P_P). The experimentally determined P_P values vary among the segments, with a range of 0.56 to 0.8 (**Figure 7A**). The product of the eight P_P values gives an estimate of the proportion of singular infections in which all eight segments are available for replication. This estimate is 3.2% for GFHK99 virus in MDCK cells.

To evaluate whether incomplete viral genomes account for the focusing of GFHK99 virus production within multiply infected cells, we used our previously published computational model of IAV coinfection and reassortment[18]. In this model, the frequency of successful segment replication is governed by eight P_P parameters and an infected cell only produces virus if at least one copy of all eight segments are replicated. Importantly, in this model the amount of virus produced from productively infected cells is constant – there is no additional benefit to multiple infection. When the eight experimentally determined P_P values for GFHK99 virus in MDCK cells are used to parameterize the model, the theoretical prediction of reassortment frequency is much lower than that observed experimentally for GFHK99 viruses in MDCK cells,

but a good match with the reassortment seen in DF-1 cells (**Figure 7B**). Clearly, the frequency with which GFHK99 segments fail to be replicated cannot fully account for the high reassortment seen in the mammalian MDCK cell system, but can fully account for the more moderate reliance on multiple infection seen in DF-1 cells. Thus, in mammalian cells, a full GFHK99 viral genome is necessary but not sufficient to support robust replication.

Discussion

Our data reveal that both viral and host features dictate the degree to which productive IAV infection relies on cooperation. Differences between virus strains and host systems in multiple infection dependence lead to differences in the amount of reassortment that occurs upon coinfection, pointing to the relevance of viral collective interactions for IAV evolution and emergence. Mechanistically, our data indicate that multiple infection is needed in part for complementation of incomplete viral genomes, but that such complementation is not sufficient to ensure infection in all virus-host systems. Rather, we see that the GFHK99 virus requires a second form of cooperation to support efficient RNA synthesis in mammalian cells – a phenotype that is driven largely by the PA gene segment. For this reason, robust infection of GFHK99 virus in mammalian systems is achieved only in the context of high MOI infection. Thus, the data presented reveal that infection efficiency and need for coinfection varies with virus-host context.

An important consequence of viral genome segmentation is the potential for replication of incomplete genomes[2]. Complementation is therefore a major class of collective interaction for viruses with segmented genomes. The relevance of complementation for a given virus species likely depends on the extent to which packaging, delivery and replication of genome segments is coordinated[41]. For IAV, we and others have demonstrated that a subset of segments fails to be replicated or expressed in infected cells with high frequency[16,19,37,39]. Specifically, for influenza A/Panama/2007/99 (H3N2) virus in MDCK cells, we found that delivery of a single viral

genome results in replication of all eight segments only 1.2% of the time[19]. Data reported herein for GFHK99 virus indicate that a somewhat higher proportion of replicated viral genomes are complete – namely, 3.2%. Thus, GFHK99 virus is partially dependent on complementation for productive infection. However, the high levels of reassortment seen between GFHK99 WT and VAR viruses in mammalian cells indicate that additional cooperative interactions are at play. This is clear from the discrepancy between observed GFHK99 virus reassortment in MDCK cells and the reassortment levels expected if complementation is the only cooperative effect considered. These data point to a model in which the presence of not just complete genomes, but rather multiple copies of the viral genome, are needed to overcome host-specific barriers to GFHK99 infection in mammalian systems (**Figure 8**).

Insight into the nature of this second cooperative interaction is gleaned from the observation that the amount of viral RNA produced from GFHK99 WT viral genomes is significantly increased with the addition of a homologous virus. From these data, we can conclude that the coinfecting virus functions in *trans* to support WT virus RNA replication in mammalian cells. In addition, our genetic data implicate the viral PA in defining an acute reliance on cooperation. The PA gene segment encodes two proteins: PA and PA-X[42]. Together with PB2 and PB1, PA is an essential component of the trimeric polymerase complex[43]. PA carries the endonuclease activity required for cap-snatching from nascent Pol II transcripts, which is essential for viral transcription[44]. It also comprises much of the interface for viral polymerase dimerization, which is needed for viral replication[45]. PA-X shares its N-terminus with PA but has a unique 41 or 61 amino acid C-terminus[42,46]. PA-X contributes to the shut-off of host protein synthesis by targeting host Pol II transcripts for destruction[47,48]. Alignment of PA and PA-X proteins encoded by GFHK99 and MaMN99 viruses reveals 29 differences in PA and 9 differences in PA-X (**Supplementary Figure 7**). Three-dimensional modeling of the GFHK99 and MaMN99 PA proteins based on the recently reported structure of the A/NT/60/1968 (H3N2) virus polymerase[45] did not predict marked differences. Nevertheless, the polymorphisms noted

in and near the dimerization loop suggest a model in which the reliance of GFHK99 virus on multiple infection may result from an inefficiency of polymerase dimerization and consequent inefficiency of vRNA synthesis[45]. This model is attractive based on the results of Fan et al., who reported that aberrant *in vitro* initiation of vRNA synthesis by a dimerization mutant was ameliorated with increased polymerase concentration[45]. We speculate that multiple infection could lead to a similar effect within the infected cell. If this speculation is correct, it would suggest a role for host factors in viral polymerase dimerization, since the observed GFHK99 reliance on multiple infection is specific to mammalian cells. Finer mapping of the phenotype within PA and targeted functional studies are needed to test this model and explore other possible mechanisms.

Of note, the reliance of GFHK99 virus on multiple infection in mammalian cells does not extend to all avian IAVs, as is clear from the characterization herein of MaMN99 virus. A need for coinfection may therefore be a potential, but not ubiquitous, consequence of IAV mal-adaptation to host cells. H9N2 subtype viruses are, however, highly relevant in the context of zoonotic infection due to their prevalence at the poultry-human interface. Sporadic human infections with poultry-adapted H9N2 viruses have been reported, and these viruses share several related genes with H5N1 and H7N9 subtype viruses that have caused hundreds of severe human infections[49-54]. The G1 lineage to which the GFHK99 virus belongs circulates widely in the poultry of Southeast Asia, the Middle East, and North Africa and reassorts frequently with other poultry adapted IAVs[55-58]. The prevalence of reassortment suggests that the internal gene segments – which comprise the six non-HA, non-NA segments – are compatible with other genotypes. Our comparison of reassortment in guinea pigs and quail furthermore indicates that reassortment could be particularly prevalent in the context of zoonotic infection of mammals. This phenotype of high reassortment in mammals is expected to extend to the H5N1 and H7N9 subtype viruses of public health concern, which carry polymerase genes related to those of the GFHK99 virus[52-54]. While reassortment is typically deleterious[59,60], high

frequencies increase opportunity for fit genotypes to arise and adapt, and should therefore be considered in assessing the risk of emergence posed by non-human adapted IAVs.

Our data reveal novel concepts in both virus-host and virus-virus interactions: the extent of IAV reliance on cooperation is determined in part by interactions with the host cell, and interactions among homologous coinfecting viruses can be highly biologically significant. Coinfection dependence, in turn, is expected to have a strong impact on viral fitness and evolution by modulating the efficiency of spread within and between hosts and by changing how a virus population samples sequence space[61-64]. Thus, host and strain specificity in multiple infection dependence are likely to play an important role in determining the outcomes of IAV infection and evolution in diverse hosts.

Acknowledgements

We thank David Stallknecht (University of Georgia) for providing the A/mallard/MN/199106/99 (H3N8) biological isolate. We thank Hui Tao, Shamika Danzy, and Ginger Geiger for technical assistance. This work was funded in part by NIH R01 AI127799 (to ACL and DRP); NIH/NIAID Centers of Excellence in Influenza Research and Surveillance (CEIRS), contract numbers HHSN272201400004C (to ACL and GST) and HHSN272201400008C (to DRP). Additional funds were provided by the Georgia Research Alliance and the Georgia Poultry Federation (to DRP) and NIH/NIAID Genomic Centers for Infectious Diseases (GCID), Award Number U19AI110819.

Author Contributions

KLP contributed to the conception of work, experimental design, data acquisition and analysis, interpretation of data; KG, NTJ and C-YL contributed to experimental design, data acquisition and analysis and interpretation of data; SC contributed to data acquisition; MM and MCW

contributed to data acquisition and analysis; BEP contributed to data analysis and interpretation; GST contributed to conception of the work, data analysis and interpretation; LMF contributed to data acquisition; DRP contributed to experimental design, data analysis and interpretation; ACL contributed to conception of work, experimental design and data analysis and interpretation. All authors contributed to the writing of the manuscript.

Conflict of interest statement

The authors declare no conflicts of interest.

Data availability

10x Genomics single-cell sequencing data is available on the GEO database with the accession number XXXXXXXX. Other data are available from the corresponding author upon reasonable request.

Code availability

Code used for 10x Genomics analysis is available at: https://github.com/njacobs627/GFHK99_Multiplicity. Code used to run the agent-based model on influenza A virus reassortment was reported previously[18,19] and is also available at https://github.com/njacobs627/GFHK99_Multiplicity.

Methods

Cells and cell culture media

Madin-Darby canine kidney (MDCK) cells, a gift from Peter Palese, Icahn School of Medicine at Mount Sinai, were used in all experiments. MDCK cells from Daniel Perez at University of Georgia were used for plaque assays as this variant of the MDCK line was found to yield more

distinct plaques for the GFHK99 strain. Both MDCK cell lines were maintained in minimal essential medium (MEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Atlanta Biologicals), penicillin (100 IU), and streptomycin (100 µg per mL) (PS; Corning). 293T cells (ATCC CRL-3216) and DF-1 cells (ATCC CRL-12203) were maintained in Dulbecco's minimal essential medium (DMEM; Gibco) supplemented with 10% FBS and PS. Duck embryo cells (ATCC CCL-141) were maintained in Eagle's minimum essential medium (EMEM; ATCC) supplemented with 10% FBS and PS. Human tracheobronchial epithelial (HTBE) cells were acquired from Lonza and were amplified and differentiated into air-liquid interface cultures as recommended by Lonza and described by Danzy et al.[65]. All cells were cultured at 37°C and 5% CO_2 in a humidified incubator.

Medium for culture of IAV in each cell line (virus medium) was prepared by supplementing the appropriate media with 4.3% bovine serum albumin and penicillin (100 IU), and streptomycin (100 µg per mL). Ammonium chloride-containing virus medium was prepared by the addition of HEPES buffer and NH_4Cl at final concentrations of 50 mM and 20 mM, respectively.

Viruses

All viruses were generated by reverse genetics[66]. For avian viruses, 293T cells transfected with reverse genetics plasmids 16–24 h prior were injected into the allantoic cavity of 9–11 day old embryonated chicken eggs and incubated at 37°C for 40–48 h. The resultant egg passage 1 stocks were used in experiments. For NL09-based viruses, 293T cells transfected with reverse genetics plasmids 16–24 h prior were co-cultured with MDCK cells at 37°C for 40–48 h. Supernatants were then propagated in MDCK cells from low MOI to generate NL09 working stocks. Defective interfering segment content of PB2, PB1, and PA segments was confirmed to be minimal for each virus stock, as described previously[67] (**Supplementary Figure 3)**. All plaque assays were performed in MDCK cells; viruses were also titered by flow cytometry and/or RT qPCR-based methods where indicated. The reverse genetics system for influenza

A/guinea fowl/Hong Kong/WF10/99 (H9N2) virus was reported previously[68,69]. This strain has been referred to as WF10 in previous publications[68-70]. For consistency with other strains used in the present manuscript, it is referred to herein as GFHK99. A low passage isolate of influenza A/mallard/Minnesota/199106/99 (H3N8) virus, referred to herein as MaMN99, was obtained from David Stallknecht at the University of Georgia[71]. The virus was passaged once in eggs and then the eight cDNAs were generated and cloned into the pDP2002 vector[72]. To increase the recovery efficiency of viruses containing polymerase components from the MaMN99 virus, pCAGGS support plasmids encoding PB2, PB1, PA, and NP proteins of the A/WSN/33 (H1N1) strain were supplied.

GFHK99 WT, MaMN99 VAR and NL09 VAR viruses were engineered to contain a 6XHis epitope tag plus GGGS linker at the N-terminus of the HA protein following the signal peptide. GFHK99 VAR_1, MaMN99 WT and NL09 VAR viruses contain similarly modified HA genes, with an HA epitope tag plus a GGGS linker inserted at the N-terminus of the HA protein[36]. All assays except the single-cell mRNA sequencing used viruses carrying these epitope tags.

Silent mutations were introduced into VAR viruses by site-directed mutagenesis to allow genotyping of WT and VAR segment origin. The specific changes introduced into all VAR viruses used here are listed in **Supplementary Table 1.** Mutations introduced into GFHK99 VAR_1, MaMN99 VAR, and NL09 VAR viruses enable detection by high-resolution melt analysis or probe-based droplet digital PCR. Mutations introduced into the GFHK99 VAR_2 strain were designed to confer unique primer binding sites relative to GFHK99 WT virus. Viruses used for single-cell mRNA sequencing, GFHK99 $mVAR_1$ and GFHK99 $mVAR_2$, carry mutations near the 3' end of each transcript to allow detection by sequencing following oligo-dT priming[37].

Synchronized, single-cycle infection conditions

Conditions designed to synchronize viral entry and prevent propagation of progeny viruses were used for all cell culture-based infections with the exception of those performed for measurement

of P_P values. Synchronized, single-cycle infections were performed as follows: Cell monolayers were washed three times with PBS and placed on ice. Chilled virus inoculum was added to each well and incubated at 4°C for 45 minutes with occasional rocking to allow attachment without entry[73]. Inoculum was aspirated and cell monolayer was washed three times with cold PBS before addition of warm virus medium lacking trypsin. Cultures were incubated at 37°C. At 3 h post-infection, virus medium was replaced with ammonium chloride-containing virus medium. Addition of ammonium chloride to the medium prevents acidification of endosomes, thereby blocking further IAV infection[74]. Cultures were returned to 37°C for the remainder of the incubation time.

Infection of cultured cells for quantification of coinfection and reassortment

MDCK or DF-1 cells were seeded at a density of 4 x10^{5} cells per well in 6-well dishes 24 h before inoculation. Virus inoculum was prepared by combining WT and VAR viruses at high titer in a 1:1 ratio based on PFU titers, and then diluting in PBS to achieve MOIs ranging from 10 to 0.01 PFU per cell. Synchronized, single cycle infection conditions were used. In addition, owing to the low yield of IAV in DF-1 cells, acid inactivation of inoculum virus was performed at 1 h post-infection for this cell type. This procedure is needed if residual inoculum virus would otherwise comprise an appreciable proportion of the virus sampled at 12 or 16 h post-inoculation. For acid inactivation, media was aspirated and replaced with 500 µL of PBS-HCl, pH 3.00 and incubated 5 min at 37°C. Cells were then washed once with PBS before the addition of virus medium. GFHK99 virus-infected cells were harvested at 12 h post-infection due to high amounts of cytopathic effects at later time points. Cells infected with MaMN99 virus and MaMN99:GFHK99 chimeric viruses were harvested at 16 h post-infection. NL09 virus reassortment data shown in Figure 1C were reported previously[75] and are included here to allow comparison to the avian strains used.

Determination of infection and coinfection levels based on HA surface expression. To enumerate infected cells, surface expression of HIS and HA epitope tags was detected by flow cytometry (**Supplementary Figure 4**). This method was previously described in detail[36]. Cells were fixed following staining by resuspending in FACS buffer and 1% paraformaldehyde. The percentage of cells positive for either or both epitope tags is expressed as % cells HA^+. The percentage of cells positive for both epitope tags is expressed as % cells dual-HA^+. The relationship between these two parameters was evaluated by plotting % cells dual-HA^+ against % cells HA^+ and regressing the resultant curve as a polynomial:

$$\%\ cells\ dual\text{-}HA^+ = \beta_2 * \text{Expectation}(\%\ cells\ HA^+) + \beta_1 * (\%\ cells\ HA^+),$$

where β_2 and β_1 are genotype-specific. For a given fraction of cells expressing HA, the expected fraction of cells expressing both HA proteins is derived from the Poisson distribution with $\lambda = -ln(1 - \%\ cells\ HA^+)$:

$$\sum_{v=1}^{\infty} \frac{e^{-\lambda}\lambda^{v}}{v!}(1 - 2 * 0.5^{v})$$

From the regression models, we then quantified the degree of linearity using the equation:

$$\%\ linearity = \frac{|\beta_1|}{|\beta_1| + |\beta_2|}.$$

***Animal models and reassortment* in vivo**

Ethics Statement. All animal experiments were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Studies were conducted under animal biosafety level 2 (ABSL-2) containment and approved by the IACUC of the University of Georgia (protocols A201506-026-Y3-A2 and A201506-026-Y3-A5) for quail studies or the IACUC of Emory University (protocol PROTO201700595) for guinea pig studies.

Animals were humanely euthanized following guidelines approved by the American Veterinary Medical Association (AVMA).

Quail eggs obtained from the College of Veterinary Medicine, University of Georgia, were hatched at the Poultry Diagnostic and Research Center, University of Georgia. Two days before virus inoculation, quail sera were confirmed to be seronegative for IAV exposure by NP ELISA (IDEXX, Westbrook, ME). At 3 weeks of age, birds were moved into a HEPA in/out BSL2 facility and each group divided into individual isolator units.

Groups (n=6) of 3 week old Japanese quail (Coturnix Japonica) were used to determine the 50% quail infectious dose of the 1:1 GFHK99 WT and GFHK99 VAR_1 virus mixture. Each quail was inoculated with 500 µl by oculo-naso-tracheal route of virus mixture in PBS, at increasing concentrations of 10^0 to 10^6 $TCID_{50}$ per 500 µL. Tracheal and cloacal swab specimens were collected daily from each bird in brain heart infusion media (BHI). Swab samples were analyzed by $TCID_{50}$ assay and titers of tracheal swabs collected at 4 d post-inoculation were used to determine the QID_{50} by the Reed and Muench method[76]. Virus was not detected in cloacal swabs. QID_{50} was found to be equivalent to 1 $TCID_{50}$.

To quantify reassortment in quail, samples collected from quail (n=6) infected with the 10^2 $TCID_{50}$ dose of the 1:1 GFHK99 WT and GFHK99 VAR_1 virus mixture were used. These were the same birds as used to determine QID_{50}. Virus shedding kinetics were determined by plaque assay of tracheal swab samples and samples from days 1, 3, and 5 were chosen for genotyping of virus isolates.

Female Hartley strain guinea pigs weighing 250–350 g were obtained from Charles River Laboratories (Wilmington, MA) and housed by Emory University Department of Animal Resources. Prior to intranasal inoculation and nasal washing, guinea pigs were sedated by intramuscular injection with 30 mg per kg ketamine/ 4 mg per kg xylazine. The $GPID_{50}$ of GFHK99 WT/VAR_1 and MaMN99 WT/VAR virus mixtures were determined as follows. Groups of four guinea pigs were inoculated intranasally with virus mixture in PBS at doses of 10^0 to 10^5

PFU per 300 µL inoculum. Daily nasal washes were collected in 1 mL PBS and titered by plaque assay. Results from day 2 nasal washes were used to determine the $GPID_{50}$ by the Reed and Muench method[76]. The $GPID_{50}$ of GFHK99 virus was found to be 2.1 x 10^3 PFU; that of MaMN99 virus was 2.1 x 10^1 PFU. The GPID50 of NL09 virus was previously determined to be 1 x 10^1 PFU (ref[77] and data not shown).

To evaluate reassortment kinetics in guinea pigs, groups of six animals were infected with 10^2 x $GPID_{50}$ of the aforementioned GFHK99 WT / VAR_1, MaMN99 WT / VAR or NL09 WT / VAR virus mixtures. Virus inoculum was given intrasnasally in a 300 µL volume of PBS. Nasal washes were performed on days 1–6 post-inoculation and titered for viral shedding by plaque assay. HRM genotyping was performed on samples collected on day 1, 3, and 5 for each guinea pig.

Quantification of reassortment and effective diversity

Reassortment was quantified for *in vitro* coinfection supernatants, guinea pig nasal washes, and quail tracheal swabs as described previously[36]. Briefly, plaque assays were performed in 10 cm dishes to isolate virus clones. 1 mL serological pipettes were used to collect agar plugs into 160 µl PBS. Using a ZR-96 viral RNA kit (Zymo), RNA was extracted from the agar plugs and eluted in 40 µl nuclease free water (Invitrogen). Reverse transcription was performed using Maxima RT (Thermofisher) according to the manufacturer's protocol. The resulting cDNA was diluted 1:4 in nuclease free water and each cDNA was combined with segment specific primers (**Supplementary Table 2**) and Precision Melt Supermix (Bio-Rad) and analyzed by qPCR in a CFX384 Touch real-time PCR detection system (Bio-Rad) designed to amplify a ~100 bp region of each gene segment which contains a single nucleotide change in the VAR virus. The qPCR was followed by high-resolution melt (HRM) analysis to differentiate WT and VAR amplicons[78]. Precision Melt Analysis software (Bio-Rad) was used to determine the parental virus origin of each gene segment based on melting properties of the cDNAs and comparison to WT and VAR

controls. Each individual plaque was assigned a genotype based on the combination of WT and VAR genome segments, with 2 variants on each of 8 segments allowing for 256 potential genotypes.

Reassortment data was used to calculate viral genotypic diversity; that is, the diversity of the 256 possible WT/VAR genotypes present in a given sample. Diversity was quantified as reported previously[79], by calculating Simpson's Index, given by $D = \text{sum}(p_i^2)$, where p_i represents the proportional abundance of each genotype[80]. Simpson's Index accounts for both the raw number of species and variation in abundance of each, and is sensitive to the abundance of dominant species. These features are important to ensure that overrepresentation of a single reassortant genotype does not cause a viral population to appear more diverse than it actually is. Because Simpson's Index does not scale linearly, each sample's Simpson's Index value was converted to a corresponding Hill number to derive its effective diversity, $N_2 = 1/D$[81], which is defined as the number of equally abundant species required to generate the observed diversity in a sample community. Because it scales linearly, Hill's N_2 allows a more intuitive comparison between communities (i.e., a community with $N_2 = 10$ species is twice as diverse as one with $N_2 = 5$) and is suitable for statistical analysis by basic linear regression methods[82]. Robust linear models of N_2 vs. time were regressed using the R package *robustlmm*.

Hierarchical clustering analysis

To compare the behavior of multiple viruses during *in vitro* coinfections, a hierarchical clustering algorithm was used. In these experiments, multiplicity dependence was measured by analyzing HA co-expression and reassortment as a function of the percentage of cells expressing HA. Each of these regression models contain two parameters ($\%\ cells\ dual\text{-}HA^+ = \beta_2 *$ Expectation($\%\ cells\ HA^+$) + $\beta_1 *(\%\ cells\ HA^+)$, and $\%\ reassortment = \beta_0 + \beta_1 * ln(\%\ cells\ HA^+)$. The behavior of a given virus can therefore be expressed as a set of four parameters. These parameters were

used to calculate the distance between each pair of viruses, and Ward's method of aggolomerative hierarchial clustering was then used to organize viruses into clusters based on these distances[83]. Briefly, this algorithm combines a nearby pair of elements into a cluster with a new position that is halfway between the original individuals. This process is repeated until all elements have been incorporated into one cluster. The R package *pvclust* was used to calculate statistical support for the existence of each node by multiscale bootstrap resampling, with nodes appearing in 95% of trials being deemed statistically significant. Each dendrogram was then divided into two clusters to determine whether the behavior of GFHK99 virus in MDCK cells represented a true outgroup (**Figure 1**), or to determine whether each GFHK99:MaMN99 chimeric virus was more similar to GFHK99 or MaMN99 parent virus (**Figure 3**).

Single-cycle viral growth kinetics

DF-1 or MDCK cells were seeded at $4x10^5$ cells per well in 6 well dishes 24 h prior to infection. Virus was serially diluted using PBS. Synchronized, single cycle infection conditions with acid inactivation of inoculum virus as described above were used. At each time point, 120 µl supernatant was collected. Viral titers for each sample were assessed by plaque assay in MDCK cells. Each MOI condition was used in 5–6 wells in parallel infections. Three wells served as technical replicates for growth curve sampling while the remaining wells were harvested at 24 h post-infection to enumerate HA-expressing cells via flow cytometry. Flow cytometry data for NL09 in MDCK cells are not included owing to low sensitivity of the assay for this virus. In cases where acid inactivation was inefficient, the replicate was eliminated, and data are plotted in duplicate.

Effect of increasing multiple infection on viral RNA replication

For DF-1 and MDCK cell experiments, 12-well plates were seeded with $3x10^5$ cells per well 24 h prior to infection. For HTBE cells, cells were cultured at an air-liquid interface as previously

described[65]. Cell surfaces were washed three times with PBS prior to inoculation. Triplicate wells were then inoculated with increasing doses of VAR virus, plus 0.005 PFU per cell of WT virus in DF-1 and MDCK cells or 0.05 PFU per cell of WT virus in HTBE cells. VAR viruses used were GFHK99 VAR_2, MaMN99 VAR or NL09 VAR. After 55 minutes at 37°C, inoculum was aspirated, cells were washed three times with PBS and 500 µl per well virus medium was added. Media was exchanged for ammonium chloride-containing media 3 h later. At 12 h post-infection, virus media was removed and cells were harvested using RNAprotect Cell Reagent (Qiagen). RNA was extracted using RNeasy columns (Qiagen) and then reverse transcribed with universal influenza primers[84] and Maxima RT per protocol instructions. Digital droplet PCR (ddPCR) was performed on the resultant cDNA. For GFHK99 virus, QX200™ ddPCR™ EvaGreen Supermix (Bio-Rad) was used with a combination of PB2, M, and NS primers specific for the GFHK99 WT virus (final primer concentration of 200 nM) (**Supplementary Table 3**). For MaMN99 and NL09 viruses, QX200™ ddPCR™ Supermix for Probes (Bio-Rad) was used with NP specific primers and probes (**Supplementary Table 3**).

Strand-specific quantification of viral RNA species over time

Viruses used for this experiment were the same GFHK99 WT/VAR_1 or MaMN99 WT/VAR virus mixtures used to measure reassortment, but in this case each mixture was considered as a single virus population (i.e. the RT ddPCR assay outlined below to quantify viral m/vRNA does not differentiate between WT and VAR genotypes). Doses used were 0.5 RNA copies per cell for the low MOI and 3.0 HA-expressing units/cell for the high MOI. RNA copy numbers of virus stocks were determined by ddPCR targeting at least four segments (mean values were used). HA-expressing units per mL was determined by counting HA^+ cells by flow cytometry in the relevant cell type. Specifically, cells were infected with serial dilutions of virus under synchronized, single cycle conditions. At 24 h post-infection, cells were harvested and flow cytometry was performed targeting His and HA epitope tags. HA expressing units per mL

for each virus / cell combination was calculated based on the linear range of %HA^+ cells plotted as a function of volume of virus added to cells[85] (**Supplementary Figure 2**).

12-well plates were seeded with $2x10^5$ cells per well of MDCK or DF-1 cells and incubated at 37°C for 24 h. Synchronized, single cycle infection conditions were used, as described above. Chilled virus was added at a volume of 125 µL per well. At 0, 1, 2, 4, 6, 8, and 10 h post-infection, virus medium was aspirated and cells were harvested using 400 µL of RNAprotect Cell Reagent (Qiagen). RNA was extracted using the Qiagen RNAeasy Mini kit. Two reverse transcription reactions per sample were set up with primers targeting mRNA or vRNA of the NS segment, each containing different nucleotide barcode tags (**Supplementary Table 4**). Maxima RT was used according to the manufacturer's instructions and combined with 300 ng MDCK or 150 ng DF-1 RNA. Absolute copy number of cDNA was determined by ddPCR. Forward and reverse primers for vRNA or mRNA of NS at a total concentration of 200 nM were combined with diluted cDNA and QX200™ ddPCR™ EvaGreen Supermix (Bio-Rad). Primer sequences are given in **Supplementary Table 4**. Thermocycler protocol was 95°C for 5 min, [95°C for 30s, 57°C for 60s] repeat 40x, 4°C for 5 min, 90°C for 5 min, 4°C hold. Copy number was normalized to RNA input to give final results in units of copy number per ng RNA. We did attempt to measure cRNA in this assay but found that primers designed to be specific for cRNA cross-primed on viral mRNA.

Single-cell mRNA sequencing

For this assay, viruses were titered in DF-1 cells using flow cytometry with anti-NP antibody. DF-1 cells were used because they give more sensitive detection of GFHK99 virus infection than MDCK cells. Cells were infected with serial dilutions of virus under synchronized, single-cycle conditions. At 24 h post-infection flow cytometry was performed targeting NP: cells were processed with the BD Cytofix/Cytoperm™ Kit (catalog no. 554714) and stained with anti-NP antibody (Abcam, clone 9G8, catalog no. ab43821) followed by donkey anti-Mouse IgG Alexa

Fluor 488 (Invitrogen catalog no. A-21202). NP expression units per were calculated based on the linear range of % cells NP^+ plotted as a function of volume of virus added to cells[85] (**Supplementary Figure 2**).

To perform single-cell mRNA sequencing, MDCK and DF-1 cells were seeded into 6-well plates at $5x10^5$ cells per well. At 24 h post seeding, cells were infected with synchronized, single cycle conditions. In the first experiment, MDCK or DF-1 cells were inoculated with GFHK99 WT virus at an MOI of 0.07, 0.2, 0.6, or 1.8 NP units per cell. In the second experiment, MDCK or DF-1 cells were inoculated with a 1:1 ratio of GFHK99 WT virus and GFHK99 $mVAR_1$ virus that amounted to a MOI of 0.02, 0.07, 0.2 or 0.6 NP units per cell. GFHK99 $mVAR_2$ virus was added to MDCK and DF-1 cell infections at MOIs of 1 PFU per cell and 0.1 PFU per cell, respectively. Infected cells were incubated for 8 h at 37^oC. Culture media was then aspirated and cells washed once with 1X PBS. Cells were then trypsinized with 200 µL of 0.25% Trypsin EDTA until all cells came off the plate and were mono-dispersed. To each well, 0.5 mL of virus medium was added and replicates were pooled (2 wells per MOI). Cells for each sample were counted. Samples were spun at 150 rcf for 3 minutes and washed with 0.5 mL of 1X PBS/0.04% BSA. Washings were performed two more times. Finally, cells were resuspended with 1X PBS/0.04% BSA to give 7×10^5 cells per mL. Preparation for single-cell transcriptomic sequencing followed the protocol for 10x Genomics Chromium Single Cell platform.

Analysis of viral transcripts from single cells was performed with the sequencing data from all experiments using Cell Ranger software. Briefly, the Cell Ranger software assigns each read to individual cells and transcripts based on two sets of unique molecular identifiers (UMIs) that are ligated prior to amplification. This approach allows the quantification of amplification bias at both the cellular and transcript levels. The first step of the analytical workflow was to map the reads to concatenated transcriptomes of GFHK99 virus with the transcriptomes of dog or chicken to analyze MDCK and DF-1 cell infections, respectively. Protein coding regions for

the dog and chicken transcriptomes were identified in the GTF file associated with genome builds CanFam3.1.98 (NCBI accession number GCA_000002285.2) and GRCg6a (NCBI accession number GCA_000002315.5), respectively, while GFHK99 virus coding regions were extracted from the reverse-complement sequences of the GFHK99 strains. The aligned sequencing data is available on the GEO database with the accession number XXXXXXXX. Cell Ranger output in the form of gene and barcode counts were then analyzed in R using the package *CellrangerRkit*. To account for the issue of cellular lysis, which can allow uninfected cells to acquire viral RNA from the supernatant and thus appear infected, a preliminary analysis was conducted to exclude cells that were likely false positives. This analysis was informed by the reasoning that 1) contaminated cells were likely to contain less viral RNA than truly infected cells, and 2) the amount of viral RNA present in contaminated cells should be less consistent than the amount present in truly infected cells. Within each infection, the proportion of each cell's transcriptome that was comprised of viral RNA was calculated, and cells were ordered according to this proportion so that the marginal gain in % viral RNA from one cell to the next could be calculated. Excluding cells with no marginal gain (indicating two cells with the same % viral RNA), the marginal gain vs. % viral RNA was plotted for each infection, which showed that this marginal gain is initially a rapidly decreasing function of % viral RNA. The first local minimum of a local regression was derived for each infection, to determine the point at which marginal gain became more consistent and less affected by % viral RNA, and cells with % viral RNA below that threshold were excluded from further analysis (**Supplementary Figure 5**). To enable comparisons between samples, the median number of UMIs per cell was calculated for each infection, and the UMI counts of each cell within that infection were normalized to this median value (e.g. if the total number of UMIs in a cell was 50% of that detected in the median cell, all of its UMI counts were multiplied by 2). The number of viral RNA transcripts per cell was then calculated and $\log_{10}$-transformed.

Single-cell sorting assay for measurement of P_P values

Segment-specific P_P values were determined as previously described for influenza A/Panama/2007/99 (H3N2) virus[19], and as follows. $4x10^5$ MDCK cells were seeded into each well of a 6-well dish. 24 h later, cells were washed 3x with PBS and inoculated with 0.018 PFU per cell of GFHK99 WT virus and 1 PFU per cell of GFHK99 VAR_2 virus in 250 µL of PBS. Virus was allowed to attach at 37°C for 1 h. Inoculum was then removed and cells were rinsed 3x with PBS and 2 mL of virus medium was added to the well. After 1 h at 37°C, medium was removed and cells were washed 3x with PBS and harvested by addition of Cell Dissociation Buffer (Corning). Cells were resuspended in complete medium and washed 3x with 2 mL FACS buffer (2% FBS in PBS). A final resuspension step was performed in PBS containing 1% FBS, 10 mM HEPES, and 0.1% EDTA. Cells were strained through a cell strainer cap (Falcon) and sorted on a BD Aria II cell sorter. Gating was performed to remove debris and multiplets and one event per well was sorted into each well of a 96-well plate containing MDCK monolayers at 30% confluency in 50 µl virus medium supplemented with 1 ug per mL TPCK-treated trypsin. Following the sort, an additional 50 µl of virus medium plus TPCK-treated trypsin was added to each well and plates were centrifuged at 1,800 rpm for 2 minutes to promote cell attachment. Plates were incubated at 37°C for 48 h to allow propagation of virus from the sorted cell.

RNA was extracted from infected cells in the 96-well plate using a ZR-96 Viral RNA Kit (Zymo Research) per manufacturer instructions. Extracted RNA was converted to cDNA using universal influenza primers[84] and Maxima RT according to manufacturer instructions. After conversion, cDNA was diluted 1:4 with nuclease-free water and used as template (4 µL per reaction) for segment-specific qPCR using SsoFast EvaGreen Supermix (Bio-Rad) in 10 µl reactions, with 200 nM final primer concentration. Primers employed targeted each segment of GFHK99 WT virus, as well as the PB2 and PB1 segments of GFHK99 VAR_2 virus. Primer sequences are listed in **Supplementary Table 3**.

Given the MOI of GFHK99 WT virus used in the experiments, an appreciable number of wells are expected to receive two or more viral genomes, and so a mathematical adjustment is needed to estimate the probability of each genome segment being delivered by a single virion. Using the relationship between MOI and the fraction of cells infected from Poisson statistics, i.e., $f = 1 - e^{-\mathrm{MOI}}$, the probability of the *i*th segment being present in a singly infected cell, or $P_{P,i}$ can be calculated from the 96-well plate using the following equation:

$$P_{P,i} = \frac{\mathrm{MOI}_i}{\mathrm{MOI}_{\mathrm{wt}}} = \frac{-\ln(1 - f_i)}{-\ln(1 - f_{\mathrm{wt}})} = \frac{\ln(1 - \frac{C_i}{A})}{\ln(1 - \frac{B}{A})}$$

where A is the number of VAR_2^+ wells, B is the number of WT^+ wells (containing any WT segment), and C_i is the number of wells positive for the WT segment in question. Wells that were negative for VAR_2 virus segments were excluded from analysis.

Figure Legends

Figure 1. Coinfection and reassortment frequencies indicate that IAV multiplicity dependence varies with virus strain and host species. A–D) MDCK or DF-1 cells were coinfected with homologous WT and VAR viruses of either GFHK99, MaMN99, or NL09 strain backgrounds at a range of MOIs. The relationship between % cells HA^+ and % cells dual-HA^+ (A) varies with strain and cell type, resulting in curves of differing % linearity (B). GFHK99, MaMN99, and NL09 viruses exhibit different reassortment levels in MDCK cells, but all show high reassortment relative to a theoretical prediction in which singly infected and multiply infected cells have equivalent burst sizes (C). GFHK99 virus reassortment levels differ in MDCK and DF-1 cells, but again reassortment in DF-1 cells remains high relative to the theoretical prediction in which multiple infection confers no advantage (C). Clustering analysis of reassortment and HA co-expression regression models determines that GFHK99 virus exhibits unique behavior in MDCK cells compared to DF-1 cells or other viruses in MDCK cells (D). Yellow circles indicate nodes with >95% bootstrap support. In guinea pigs (n=6), GFHK99 WT and VAR_1 viruses exhibit higher reassortment than MaMN99 or NL09 WT and VAR viruses, as indicated by increased genotypic diversity (E). The GFHK99 WT and VAR_1 viruses exhibit higher reassortment in guinea pigs than in quail (n=5) (F). Guinea pig data shown in panels E and F are the same. NL09 virus reassortment data shown in (C) were reported previously[75]. Shading represents 95% CI.

Figure 2. Increasing MOI increases viral productivity at sub-saturating, but not saturating MOIs. MDCK or DF-1 cells were infected under single-cycle conditions at a range of MOIs in triplicate wells for each MOI. A–F) Viral titers observed at the indicated MOIs are plotted against time post-infection for GFHK99 virus in MDCK cells (A), GFHK99 virus in DF-1 cells (B), MaMN99 virus in MDCK cells (C), NL09 virus in MDCK cells (D), GFHK99 virus in MDCK cells (E), and GFHK99 virus in DF-1 cells (F). G) Fold change in amplification (maximum PFU output

/ PFU input) relative to the MOI=0.01 PFU per cell condition is plotted for each virus-cell pairing. H) Burst size, calculated as maximum PFU output / number of HA^+ cells detected by flow cytometry, is plotted for each virus-cell pairing tested in the higher MOI range. MOIs shown are in units of PFU per cell, as determined in MDCK cells. Error bars represent mean ± standard error.

Figure 3. Coinfection and reassortment of chimeric viruses reveal a major role for the viral PA gene segment. Reverse genetics was used to place one or more genes from GFHK99 virus into a MaMN99 background. Coinfections with homologous WT and VAR strains were performed in MDCK cells as in Figure 1. The relationship between % cells HA positive and % cells dually HA positive (A, B) and reassortment levels (C, D) are shown for each genotype. Experimental results are compared to a theoretical prediction based on Poisson statistics and in which singly infected and multiply infected cells have equivalent burst sizes (Prediction). Clustering analysis of the regression models shown in (A-D) is used to partition the behavior of each chimeric genotype into one of two clusters, denoting closer similarity to GFHK99 or MaMN99 (E). Yellow circles indicate nodes with >95% bootstrap support. Data shown for GFHK99 and MaMN99 viruses are the same as those displayed in Figure 1.

Figure 4. Coinfection enhances GFHK99 vRNA synthesis in a dose and host dependent manner. Cells were coinfected with WT virus and increasing doses of VAR virus. WT virus MOI was 0.05 PFU per cell in HTBE cells and 0.005 PFU per cell in all other cell types. The fold change in WT vRNA copy number, relative to that detected in the VAR virus, is plotted for GFHK99 virus in MDCK and DF-1 cells (A), GFHK99 virus in HTBE cells (B), NL09 virus in MDCK cells (C), NL09 virus in HTBE cells (D) and MaMN99 virus in MDCK and duck embryo (DE) cells (E). n=3 cell culture dishes per condition. Error bars represent mean ± standard error.

Figure 5. High multiplicity of infection is needed for robust GFHK99 polymerase activity in MDCK cells. Dishes of MDCK or DF-1 cells (n=3) were infected with GFHK99 or MaMN99 virus at low (0.5 RNA copies per cell) or high (3 HA-expressing units per cell) MOI. NS segment vRNA and mRNA was quantified at the indicated time points (A–F). The average fold change from initial (t=0) to peak RNA copy number is plotted for low MOI infections (G) and high MOI infections (H). Error bars represent standard deviation. Significance was assessed by multiple unpaired t-tests with correction for multiple comparisons using the Holm-Sidak method, with alpha = 5.0%. Each row was analyzed individually, without assuming a consistent SD: *$p < 0.05$, **<0.01, ***<0.001. ns = not significant.

Figure 6. Homologous coinfecting virus boosts GFHK99 viral transcription in single cells and reveals comparable rates of segment detection in MDCK and DF-1 cells. DF-1 or MDCK cells were infected with GFHK99 virus (left facet) at MOIs of 0.07, 0.2, 0.6, or 1.8 NP-expressing units per cell, or a 1:1 mixture of GFHK99 WT and GFHK99 $mVAR_1$ viruses at four different total MOIs (0.02, 0.07, 0.2, 0.6 NP-expressing units per cell) and a constant amount of GFHK99 $mVAR_2$ virus (0.1 PFU per cell in DF-1 cells, 1.0 PFU per cell in MDCK cells) (right facet). Violin plots show the distribution of total viral RNA (A) or distinct viral genome segments (B) per cell in each cell-MOI infection condition. Vertical lines denote the median of each distribution. UMI = unique molecular identifier.

Figure 7. Incomplete GFHK99 virus genomes are present but not sufficiently abundant to account for observed reassortment in MDCK cells. Incomplete viral genomes were quantified experimentally by a single-cell assay which relies on the amplification of incomplete viral genomes of GFHK99 WT virus (0.018 PFU per cell) by a genetically similar coinfecting virus, GFHK99 VAR_2. Based on the rate of detection of GFHK99 WT virus segments in this assay, the probability that a given segment would be present and replicated in a singly infected

MDCK cell is reported as P_P. A) Summary of experimental P_P data. n = 4 biological replicates, distinguished by color. Horizontal bars and shading represent mean ± standard deviation. Mean $P_{P,i}$ values for each segment are indicated at the bottom of the plot area. Average P_P for each experiment was calculated as the geometric mean of the eight $P_{P,i}$ values, and the mean ± standard deviation of those four summary P_P values is shown at the top of the plot area. B) Experimentally obtained $P_{P,i}$ values in A were used to parameterize a computational model[18]. Levels of reassortment predicted using the experimentally determined parameters are shown with colors corresponding to points in (A). Levels of reassortment predicted if P_P=1.0 are shown with the dashed line. Observed reassortment of GFHK99 WT and VAR viruses in MDCK cells are shown with blue circles. Observed reassortment of GFHK99 WT and VAR viruses in DF-1 cells are shown with pink circles. Observed data are the same as those plotted in Figure 1.

Figure 8. Model: complementation of incomplete viral genomes is necessary but not sufficient for robust progeny production from mammalian cells infected with GFHK99 virus. Cells infected with single virions often replicate incomplete viral genomes and therefore do not produce viral progeny (A). In mammalian cells, complementation of incomplete viral genomes through coinfection may allow progeny production (B), but robust viral yields require the delivery of additional genomes to the cell, beyond the levels needed for complementation (C). Horizontal bars within cells represent viral gene segments successfully delivered and replicated.

References

1 Leeks, A., Sanjuan, R. & West, S. A. The evolution of collective infectious units in viruses. *Virus Res* **265**, 94-101, doi:10.1016/j.virusres.2019.03.013 (2019).

2 Brooke, C. B. Population Diversity and Collective Interactions during Influenza Virus Infection. *J Virol* **91**, doi:10.1128/JVI.01164-17 (2017).

3 Sanjuan, R. Collective Infectious Units in Viruses. *Trends Microbiol* **25**, 402-412, doi:10.1016/j.tim.2017.02.003 (2017).

924 4 Santiana, M. *et al.* Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral
925 Transmission. *Cell Host Microbe* **24**, 208-220 e208, doi:10.1016/j.chom.2018.07.006 (2018).
926 5 Chen, Y. H. *et al.* Phosphatidylserine vesicles enable efficient en bloc transmission of
927 enteroviruses. *Cell* **160**, 619-630, doi:10.1016/j.cell.2015.01.032 (2015).
928 6 Erickson, A. K. *et al.* Bacteria Facilitate Enteric Virus Co-infection of Mammalian Cells and
929 Promote Genetic Recombination. *Cell Host Microbe* **23**, 77-88 e75,
930 doi:10.1016/j.chom.2017.11.007 (2018).
931 7 Andreu-Moreno, I. & Sanjuan, R. Collective Infection of Cells by Viral Aggregates Promotes Early
932 Viral Proliferation and Reveals a Cellular-Level Allee Effect. *Current biology : CB* **28**, 3212-
933 3219.e3214, doi:10.1016/j.cub.2018.08.028 (2018).
934 8 Cuevas, J. M., Duran-Moreno, M. & Sanjuan, R. Multi-virion infectious units arise from free viral
935 particles in an enveloped virus. *Nat Microbiol* **2**, 17078, doi:10.1038/nmicrobiol.2017.78 (2017).
936 9 Cifuentes-Munoz, N., Dutch, R. E. & Cattaneo, R. Direct cell-to-cell transmission of respiratory
937 viruses: The fast lanes. *PLoS Pathog* **14**, e1007015, doi:10.1371/journal.ppat.1007015 (2018).
938 10 Lawrence, D. M. *et al.* Measles virus spread between neurons requires cell contact but not CD46
939 expression, syncytium formation, or extracellular virus production. *J Virol* **74**, 1908-1918,
940 doi:10.1128/jvi.74.4.1908-1918.2000 (2000).
941 11 McDonald, D. *et al.* Recruitment of HIV and its receptors to dendritic cell-T cell junctions. *Science*
942 **300**, 1295-1297, doi:10.1126/science.1084238 (2003).
943 12 Shirogane, Y., Watanabe, S. & Yanagi, Y. Cooperation between different RNA virus genomes
944 produces a new phenotype. *Nat Commun* **3**, 1235, doi:10.1038/ncomms2252 (2012).
945 13 Ciota, A. T., Ehrbar, D. J., Van Slyke, G. A., Willsey, G. G. & Kramer, L. D. Cooperative interactions
946 in the West Nile virus mutant swarm. *BMC Evol Biol* **12**, 58, doi:10.1186/1471-2148-12-58
947 (2012).
948 14 Xue, K. S., Hooper, K. A., Ollodart, A. R., Dingens, A. S. & Bloom, J. D. Cooperation between
949 distinct viral variants promotes growth of H3N2 influenza in cell culture. *Elife* **5**, e13974,
950 doi:10.7554/eLife.13974 (2016).
951 15 Brooke, C. B., Ince, W. L., Wei, J., Bennink, J. R. & Yewdell, J. W. Influenza A virus nucleoprotein
952 selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and
953 transmissibility. *Proc Natl Acad Sci U S A* **111**, 16854-16859, doi:10.1073/pnas.1415396111
954 (2014).
955 16 Brooke, C. B. *et al.* Most influenza a virions fail to express at least one essential viral protein.
956 *Journal of virology* **87**, 3155-3162, doi:10.1128/jvi.02284-12 (2013).
957 17 Sun, J. & Brooke, C. B. Influenza A Virus Superinfection Potential Is Regulated by Viral Genomic
958 Heterogeneity. *MBio* **9**, doi:10.1128/mBio.01761-18 (2018).
959 18 Fonville, J. M., Marshall, N., Tao, H., Steel, J. & Lowen, A. C. Influenza Virus Reassortment Is
960 Enhanced by Semi-infectious Particles but Can Be Suppressed by Defective Interfering Particles.
961 *PLoS pathogens* **11**, e1005204, doi:10.1371/journal.ppat.1005204 (2015).
962 19 Jacobs, N. T. *et al.* Incomplete influenza A virus genomes occur frequently but are readiliy
963 complemented during localized viral spread. *Nature Communications* (2019).
964 20 Nayak, D. P. Defective interfering influenza viruses. *Annu Rev Microbiol* **34**, 619-644,
965 doi:10.1146/annurev.mi.34.100180.003155 (1980).
966 21 Von Magnus, P. Incomplete forms of influenza virus. *Adv Virus Res* **2**, 59-79 (1954).
967 22 Brooke, C. B. Biological activities of 'noninfectious' influenza A virus particles. *Future Virol* **9**, 41-
968 51, doi:10.2217/fvl.13.118 (2014).
969 23 Timm, C., Gupta, A. & Yin, J. Robust kinetics of an RNA virus: Transcription rates are set by
970 genome levels. *Biotechnology and bioengineering* **112**, 1655-1662, doi:10.1002/bit.25578
971 (2015).

24 Boulle, M. *et al.* HIV Cell-to-Cell Spread Results in Earlier Onset of Viral Gene Expression by Multiple Infections per Cell. *PLoS Pathog* **12**, e1005964, doi:10.1371/journal.ppat.1005964 (2016).
25 Sanjuán, R. & Thoulouze, M.-I. Why viruses sometimes disperse in groups?(†). *Virus Evol* **5**, vez014-vez014, doi:10.1093/ve/vez014 (2019).
26 Sigal, A. *et al.* Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* **477**, 95-98, doi:10.1038/nature10347 (2011).
27 Webster, R. G., Hinshaw, V. S., Bean, W. J., Jr., Turner, B. & Shortridge, K. F. Influenza viruses from avian and porcine sources and their possible role in the origin of human pandemic strains. *Dev Biol Stand* **39**, 461-468 (1977).
28 Wright, P. F., Neumann, G. & Kawaoka, Y. in *Fields Virology* Vol. 1 (ed D. M. & Howley Knipe, P. M.) 1691-1740 (Lippincott-Raven, 2006).
29 Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* **56**, 152-179 (1992).
30 Webster, R. G., Shortridge, K. F. & Kawaoka, Y. Influenza: interspecies transmission and emergence of new pandemics. *FEMS Immunol Med Microbiol* **18**, 275-279, doi:S0928-8244(97)00058-8 [pii] (1997).
31 Taubenberger, J. K. & Morens, D. M. 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis* **12**, 15-22 (2006).
32 Viboud, C., Miller, M., Olson, D., Osterholm, M. & Simonsen, L. Preliminary Estimates of Mortality and Years of Life Lost Associated with the 2009 A/H1N1 Pandemic in the US and Comparison with Past Influenza Seasons. *PLoS Curr*, RRN1153, doi:k/-/-/35hpbywfdwl4n/8 [pii] (2010).
33 Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol* **17**, 67-81, doi:10.1038/s41579-018-0115-z (2019).
34 Rajsbaum, R. *et al.* Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein. *PLoS Pathog* **8**, e1003059, doi:10.1371/journal.ppat.1003059 (2012).
35 Manz, B. *et al.* Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathog* **9**, e1003279, doi:10.1371/journal.ppat.1003279

PPATHOGENS-D-12-02051 [pii] (2013).
36 Marshall, N., Priyamvada, L., Ende, Z., Steel, J. & Lowen, A. C. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS pathogens* **9**, e1003421, doi:10.1371/journal.ppat.1003421 (2013).
37 Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7**, doi:10.7554/eLife.32303 (2018).
38 Ramos, I. *et al.* Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1. *J Virol* **93**, doi:10.1128/JVI.00559-19 (2019).
39 Cristobal Vera, J. *et al.* A common pattern of influenza A virus single cell gene expression heterogeneity governs the innate antiviral response to infection. *bioRxiv* (2019).
40 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
41 Lowen, A. C. It's in the mix: Reassortment of segmented viral genomes. *PLoS Pathog* **14**, e1007200, doi:10.1371/journal.ppat.1007200 (2018).

1018 42 Jagger, B. W. *et al.* An overlapping protein-coding region in influenza A virus segment 3
1019 modulates the host response. *Science* **337**, 199-204, doi:10.1126/science.1222213 (2012).
1020 43 Te Velthuis, A. J. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of
1021 viral RNA synthesis. *Nat Rev Microbiol* **14**, 479-493, doi:10.1038/nrmicro.2016.87 (2016).
1022 44 Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA
1023 subunit. *Nature* **458**, 914-918, doi:nature07745 [pii]

1024 10.1038/nature07745 (2009).
1025 45 Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral genome
1026 replication. *Nature* **573**, 287-290, doi:10.1038/s41586-019-1530-7 (2019).
1027 46 Shi, M. *et al.* Evolutionary conservation of the PA-X open reading frame in segment 3 of
1028 influenza A virus. *J Virol* **86**, 12411-12413, doi:10.1128/JVI.01677-12 (2012).
1029 47 Gaucherand, L. *et al.* The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA
1030 Processing Machinery to Limit Host Gene Expression. *Cell reports* **27**, 776-792 e777,
1031 doi:10.1016/j.celrep.2019.03.063 (2019).
1032 48 Khaperskyy, D. A., Schmaling, S., Larkins-Ford, J., McCormick, C. & Gaglia, M. M. Selective
1033 Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff
1034 Protein. *PLoS Pathog* **12**, e1005427, doi:10.1371/journal.ppat.1005427 (2016).
1035 49 Butt, K. M. *et al.* Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. *J*
1036 *Clin Microbiol* **43**, 5760-5767, doi:43/11/5760 [pii]

1037 10.1128/JCM.43.11.5760-5767.2005 (2005).
1038 50 Peiris, M. *et al.* Human infection with influenza H9N2. *Lancet* **354**, 916-917,
1039 doi:S0140673699033115 [pii] (1999).
1040 51 Guan, Y. *et al.* H9N2 influenza viruses possessing H5N1-like internal genomes continue to
1041 circulate in poultry in southeastern China. *J Virol* **74**, 9372-9380 (2000).
1042 52 Guan, Y., Shortridge, K. F., Krauss, S. & Webster, R. G. Molecular characterization of H9N2
1043 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong?
1044 *Proc Natl Acad Sci U S A* **96**, 9363-9367 (1999).
1045 53 Lam, T. T. *et al.* The genesis and source of the H7N9 influenza viruses causing human infections
1046 in China. *Nature* **502**, 241-244, doi:10.1038/nature12515

1047 nature12515 [pii] (2013).
1048 54 Wu, A. *et al.* Sequential Reassortments Underlie Diverse Influenza H7N9 Genotypes in China.
1049 *Cell Host Microbe* **4**, 446-452, doi:S1931-3128(13)00298-9 [pii]

1050 10.1016/j.chom.2013.09.001 (2013).
1051 55 Gu, M., Xu, L., Wang, X. & Liu, X. Current situation of H9N2 subtype avian influenza in China. *Vet*
1052 *Res* **48**, 49-49, doi:10.1186/s13567-017-0453-2 (2017).
1053 56 Pusch, E. A. & Suarez, D. L. The Multifaceted Zoonotic Risk of H9N2 Avian Influenza. *Vet Sci* **5**, 82,
1054 doi:10.3390/vetsci5040082 (2018).
1055 57 Kayed, A. S. *et al.* Surveillance for avian influenza viruses in wild birds at live bird markets, Egypt,
1056 2014-2016. *Influenza Other Respir Viruses* **13**, 407-414, doi:10.1111/irv.12634 (2019).
1057 58 Xu, K. M. *et al.* The genesis and evolution of H9N2 influenza viruses in poultry from southern
1058 China, 2000 to 2005. *J Virol* **81**, 10389-10401, doi:10.1128/JVI.00979-07 (2007).
1059 59 White, M. C. & Lowen, A. C. Implications of segment mismatch for influenza A virus evolution. *J*
1060 *Gen Virol* **99**, 3-16, doi:10.1099/jgv.0.000989 (2018).
1061 60 Villa, M. & Lassig, M. Fitness cost of reassortment in human influenza. *PLoS Pathog* **13**,
1062 e1006685, doi:10.1371/journal.ppat.1006685 (2017).

1063 61 Chao, L., Tran, T. & Matthews, C. Muller's Ratchet and the Advantage of Sex in the Rna Virus 6.
1064 *Evolution; international journal of organic evolution* **46**, 289-299, doi:10.1111/j.1558-
1065 5646.1992.tb02038.x (1992).
1066 62 Froissart, R. *et al.* Co-infection weakens selection against epistatic mutations in RNA viruses.
1067 *Genetics* **168**, 9-19, doi:10.1534/genetics.104.030205 (2004).
1068 63 Novella, I. S., Reissig, D. D. & Wilke, C. O. Density-dependent selection in vesicular stomatitis
1069 virus. *J Virol* **78**, 5799-5804, doi:10.1128/JVI.78.11.5799-5804.2004 (2004).
1070 64 Wilke, C. O. & Novella, I. S. Phenotypic mixing and hiding may contribute to memory in viral
1071 quasispecies. *BMC Microbiol* **3**, 11, doi:10.1186/1471-2180-3-11 (2003).
1072 65 Danzy, S. *et al.* Mutations to PB2 and NP proteins of an avian influenza virus combine to confer
1073 efficient growth in primary human respiratory cells. *Journal of virology* **88**, 13436-13446,
1074 doi:10.1128/jvi.01093-14 (2014).
1075 66 Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R. G. A DNA transfection
1076 system for generation of influenza A virus from eight plasmids. **97**, 6108-6113,
1077 doi:10.1073/pnas.100133697 %J Proceedings of the National Academy of Sciences (2000).
1078 67 Schwartz, S. L. & Lowen, A. C. Droplet digital PCR: A novel method for detection of influenza
1079 virus defective interfering particles. *J Virol Methods* **237**, 159-165,
1080 doi:10.1016/j.jviromet.2016.08.023 (2016).
1081 68 Perez, D. R., Webby, R. J., Hoffmann, E. & Webster, R. G. Land-based birds as potential
1082 disseminators of avian mammalian reassortant influenza A viruses. *Avian Dis* **47**, 1114-1117
1083 (2003).
1084 69 Song, H., Nieto, G. R. & Perez, D. R. A new generation of modified live-attenuated avian
1085 influenza viruses using a two-strategy combination as potential vaccine candidates. *J Virol* **81**,
1086 9238-9248, doi:JVI.00893-07 [pii]

1087 10.1128/JVI.00893-07 (2007).
1088 70 Sorrell, E. M., Wan, H., Araya, Y., Song, H. & Perez, D. R. Minimal molecular constraints for
1089 respiratory droplet transmission of an avian-human H9N2 influenza A virus. *Proc Natl Acad Sci U*
1090 *S A* **106**, 7565-7570, doi:0900877106 [pii]

1091 10.1073/pnas.0900877106 (2009).
1092 71 Brown, J. D. *et al.* Intestinal excretion of a wild bird-origin H3N8 low pathogenic avian influenza
1093 virus in mallards (Anas Platyrhynchos). *Journal of wildlife diseases* **48**, 991-998,
1094 doi:10.7589/2011-09-280 (2012).
1095 72 Chen, H. *et al.* Partial and full PCR-based reverse genetics strategy for influenza viruses. *PLoS*
1096 *One* **7**, e46378, doi:10.1371/journal.pone.0046378

1097 PONE-D-12-21490 [pii] (2012).
1098 73 Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. Infectious entry pathway of influenza virus in a
1099 canine kidney cell line. *J Cell Biol* **91**, 601-613, doi:10.1083/jcb.91.3.601 (1981).
1100 74 Yoshimura, A. & Ohnishi, S. Uncoating of influenza virus in endosomes. *J Virol* **51**, 497-504
1101 (1984).
1102 75 Phipps, K. L. *et al.* Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort
1103 Efficiently but Produce Attenuated Progeny. *J Virol* **91**, doi:10.1128/JVI.00830-17 (2017).
1104 76 Reed, L. J. & Muench, H. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12.
1105 *American Journal of Epidemiology* **27**, 493-497, doi:10.1093/oxfordjournals.aje.a118408 (1938).
1106 77 Campbell, P. J. *et al.* The M segment of the 2009 pandemic influenza virus confers increased NA
1107 activity, filamentous morphology and efficient contact transmissibility to A/Puerto Rico/8/1934-
1108 based reassortant viruses. *J Virol*, doi:JVI.03607-13 [pii]

1109 10.1128/JVI.03607-13 (2014).
1110 78 Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. High-resolution
1111 genotyping by amplicon melting analysis using LCGreen. *Clin Chem* **49**, 853-860 (2003).
1112 79 Richard, M., Herfst, S., Tao, H., Jacobs, N. T. & Lowen, A. C. Influenza A virus reassortment is
1113 limited by anatomical compartmentalization following co-infection via distinct routes. *J Virol*,
1114 doi:10.1128/JVI.02063-17 (2017).
1115 80 Simpson, E. H. Measurement of diversity. *Nature* **163**, 688-688 (1949).
1116 81 Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427-
1117 432 (1973).
1118 82 Jost, L. Entropy and diversity. *Oikos* **113**, 363-375 (2006).
1119 83 Ward, J. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American*
1120 *Statistical Association* **48**, doi:10.2307/2282967 (1963).
1121 84 Zhou, B. *et al.* Single-reaction genomic amplification accelerates sequencing and vaccine
1122 production for classical and Swine origin human influenza a viruses. *J Virol* **83**, 10309-10313,
1123 doi:JVI.01109-09 [pii]

1124 10.1128/JVI.01109-09 (2009).
1125 85 Drayman, N., Kler, S., Ben-nun-Shaul, O. & Oppenheim, A. Rapid method for SV40 titration. *J*
1126 *Virol Methods* **164**, 145-147, doi:10.1016/j.jviromet.2009.12.003 (2010).

1127

Figure 1

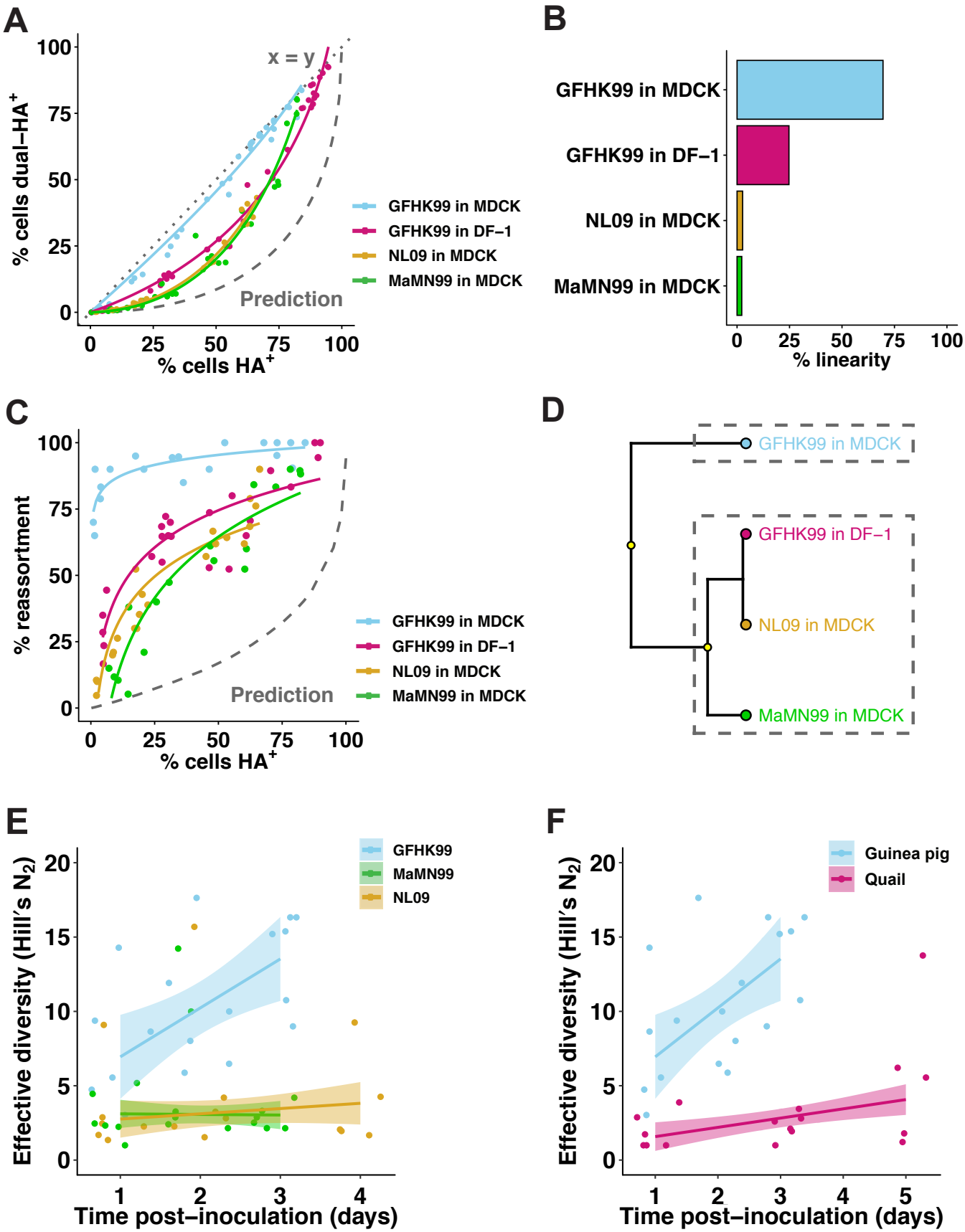

A
% cells dual–HA+
% cells HA+
x = y
Prediction
GFHK99 in MDCK
GFHK99 in DF–1
NL09 in MDCK
MaMN99 in MDCK
B
GFHK99 in MDCK
GFHK99 in DF–1
NL09 in MDCK
MaMN99 in MDCK
% linearity
C
% reassortment
% cells HA+
Prediction
GFHK99 in MDCK
GFHK99 in DF–1
NL09 in MDCK
MaMN99 in MDCK
D
GFHK99 in MDCK
GFHK99 in DF–1
NL09 in MDCK
MaMN99 in MDCK
E
Effective diversity (Hill's N2)
Time post–inoculation (days)
GFHK99
MaMN99
NL09
F
Effective diversity (Hill's N2)
Time post–inoculation (days)
Guinea pig
Quail

Figure 2

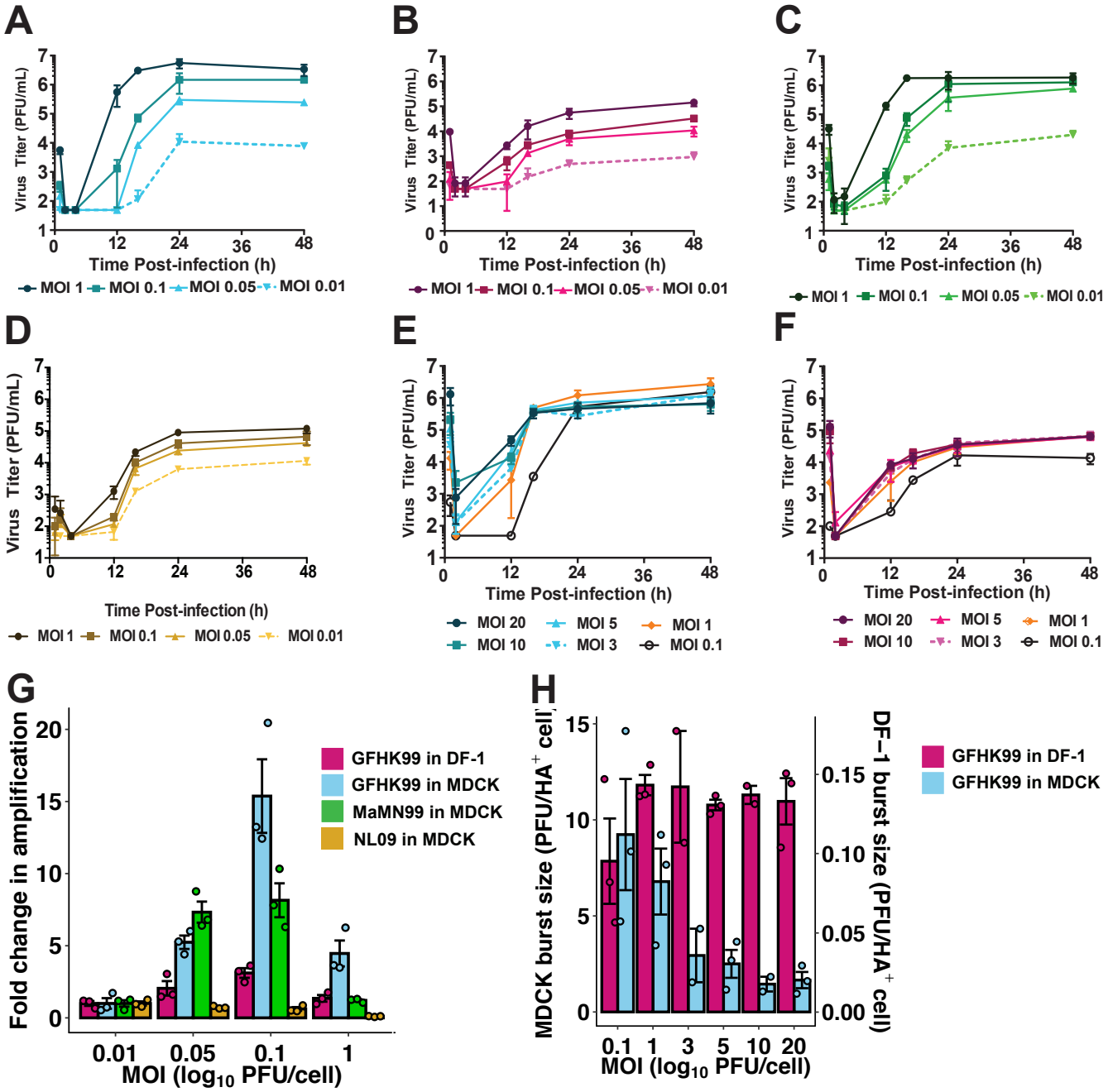

A
B
C
D
E
F
G
H
Virus Titer (PFU/mL)
Time Post-infection (h)
MOI 1
MOI 0.1
MOI 0.05
MOI 0.01
MOI 20
MOI 10
MOI 5
MOI 3
Fold change in amplification
MOI (log10 PFU/cell)
GFHK99 in DF-1
GFHK99 in MDCK
MaMN99 in MDCK
NL09 in MDCK
MDCK burst size (PFU/HA+ cell)
DF-1 burst size (PFU/HA+ cell)

Figure 3

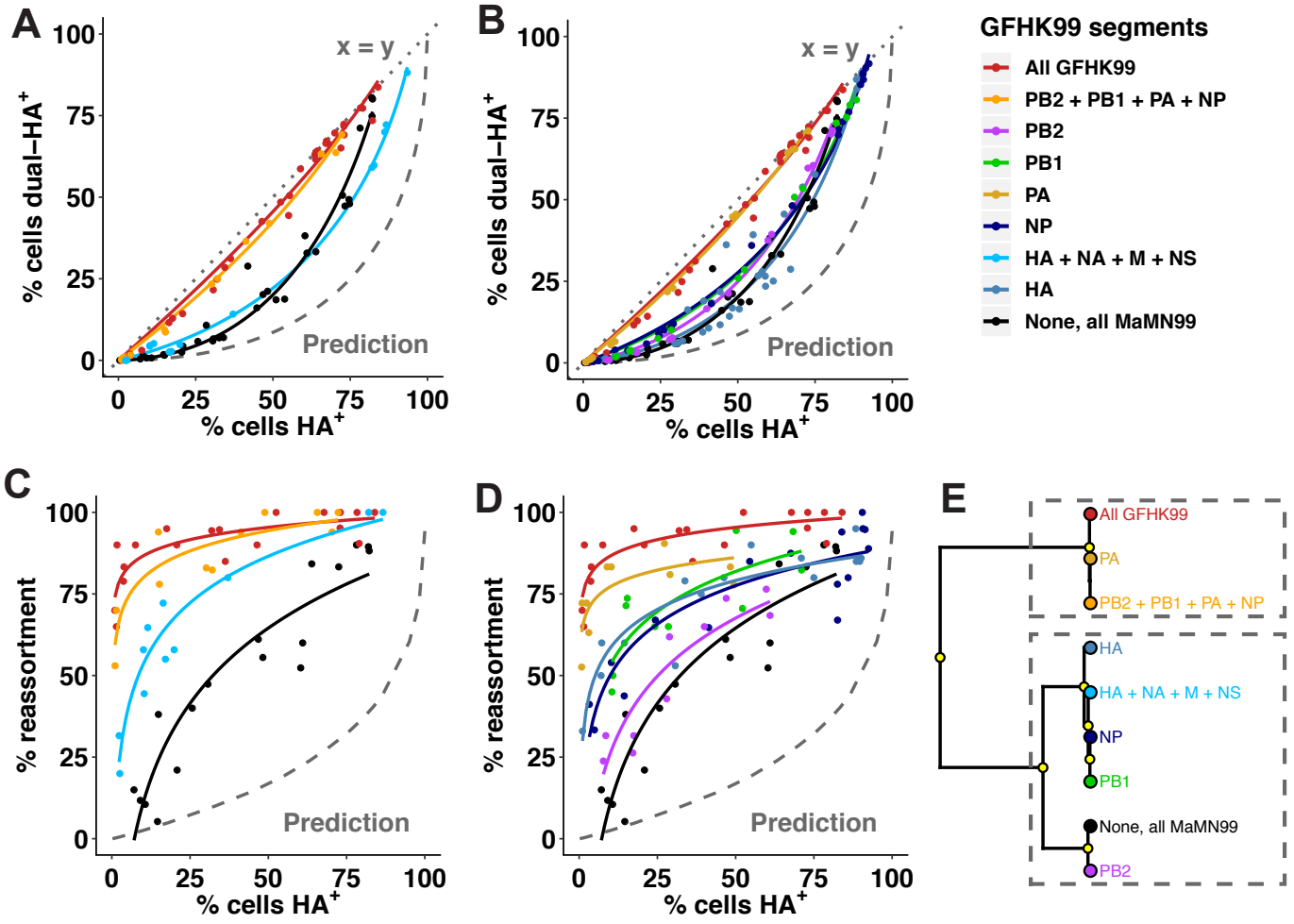

A
B
C
D
E
% cells dual–HA+
% cells HA+
% reassortment
x = y
Prediction
GFHK99 segments
All GFHK99
PB2 + PB1 + PA + NP
PB2
PB1
PA
NP
HA + NA + M + NS
HA
None, all MaMN99

Figure 4

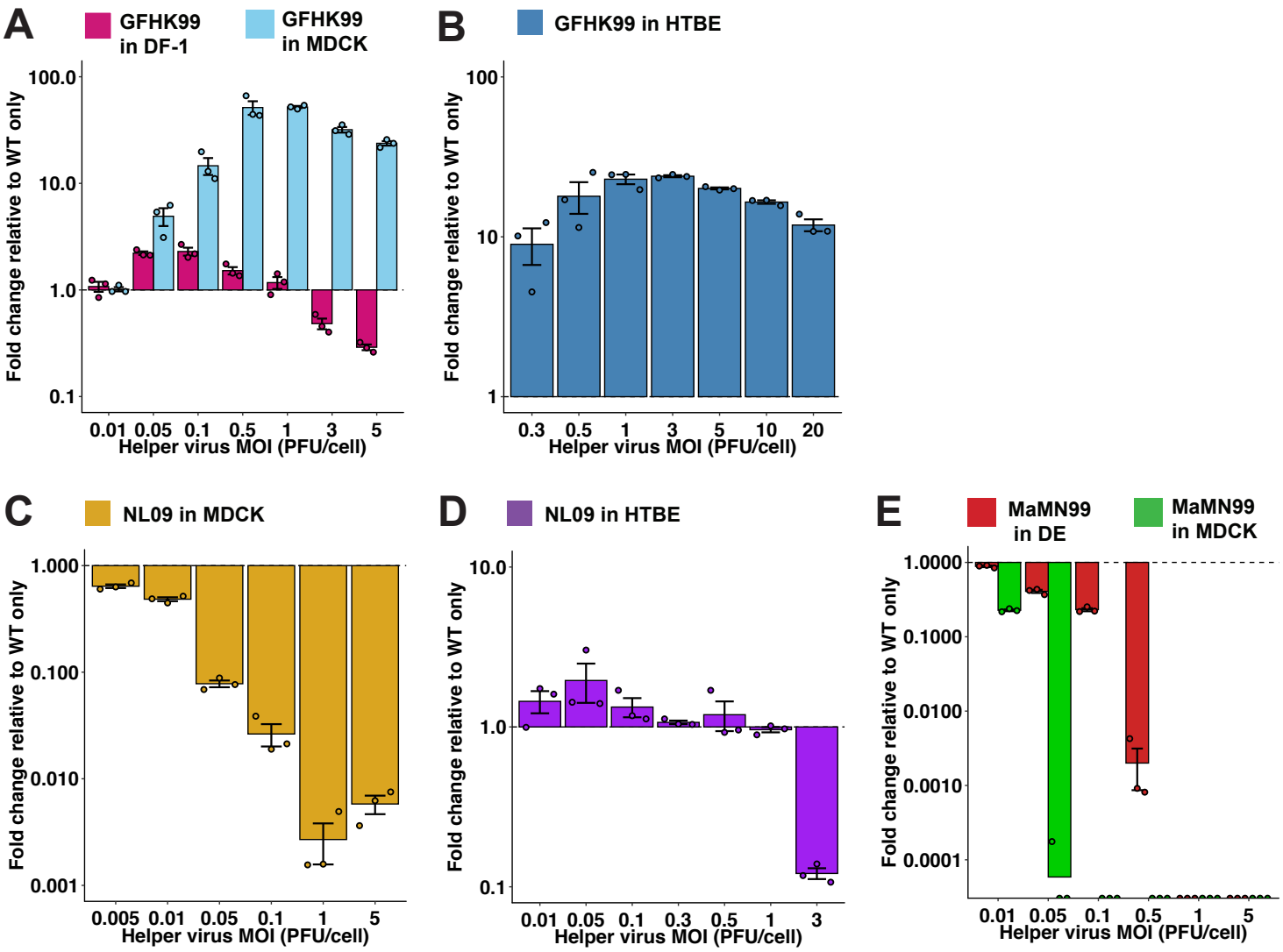

A
GFHK99 in DF-1
GFHK99 in MDCK
Fold change relative to WT only
100.0
10.0
1.0
0.1
0.01 0.05 0.1 0.5 1 3 5
Helper virus MOI (PFU/cell)
B
GFHK99 in HTBE
Fold change relative to WT only
100
10
1
0.3 0.5 1 3 5 10 20
Helper virus MOI (PFU/cell)
C
NL09 in MDCK
Fold change relative to WT only
1.000
0.100
0.010
0.001
0.005 0.01 0.05 0.1 1 5
Helper virus MOI (PFU/cell)
D
NL09 in HTBE
Fold change relative to WT only
10.0
1.0
0.1
0.01 0.05 0.1 0.3 0.5 1 3
Helper virus MOI (PFU/cell)
E
MaMN99 in DE
MaMN99 in MDCK
Fold change relative to WT only
1.0000
0.1000
0.0100
0.0010
0.0001
0.01 0.05 0.1 0.5 1 5
Helper virus MOI (PFU/cell)

Figure 5

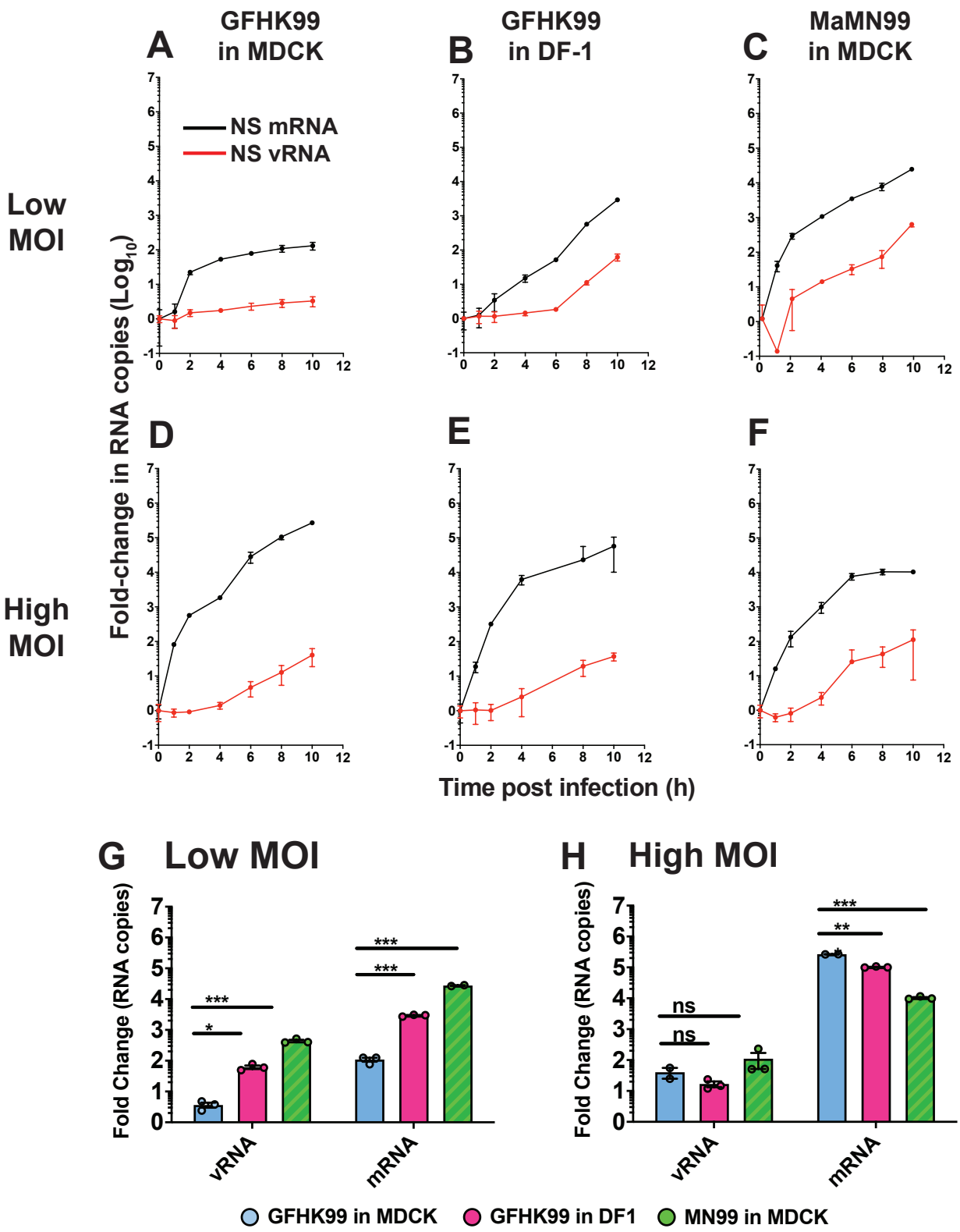

A
GFHK99 in MDCK
B
GFHK99 in DF-1
C
MaMN99 in MDCK
NS mRNA
NS vRNA
Low MOI
High MOI
D
E
F
Fold-change in RNA copies (Log10)
Time post infection (h)
G Low MOI
H High MOI
Fold Change (RNA copies)
***
*
***
***
ns
ns
***
**
vRNA
mRNA
GFHK99 in MDCK
GFHK99 in DF1
MN99 in MDCK

Figure 6

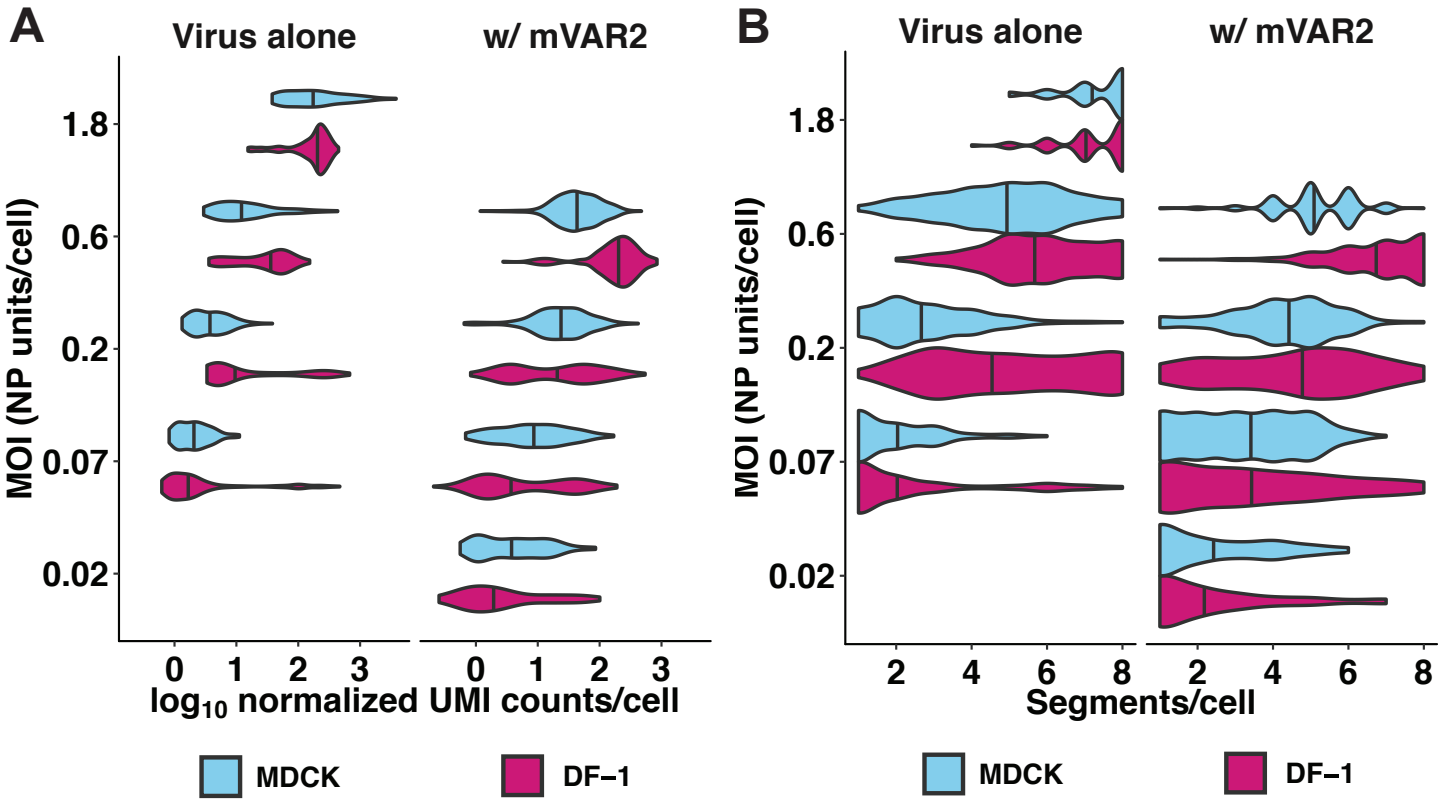
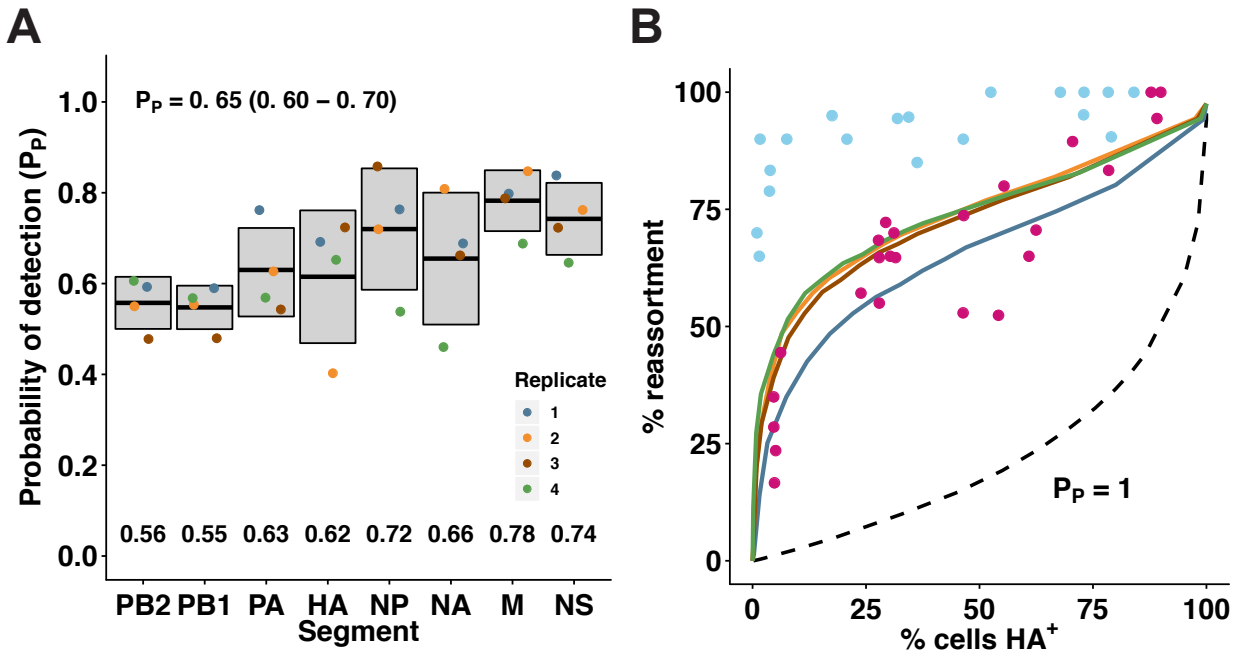

A
Virus alone
w/ mVAR2
MOI (NP units/cell)
1.8
0.6
0.2
0.07
0.02
0 1 2 3
0 1 2 3
log10 normalized UMI counts/cell
MDCK
DF–1
B
Virus alone
w/ mVAR2
MOI (NP units/cell)
1.8
0.6
0.2
0.07
0.02
2 4 6 8
2 4 6 8
Segments/cell
MDCK
DF–1

Figure 7


A
B
P_P = 0. 65 (0. 60 – 0. 70)
Probability of detection (P_P)
1.0
0.8
0.6
0.4
0.2
0.0
Replicate
1
2
3
4
0.56 0.55 0.63 0.62 0.72 0.66 0.78 0.74
PB2 PB1 PA HA NP NA M NS
Segment
% reassortment
100
75
50
25
0
P_P = 1
0 25 50 75 100
% cells HA+

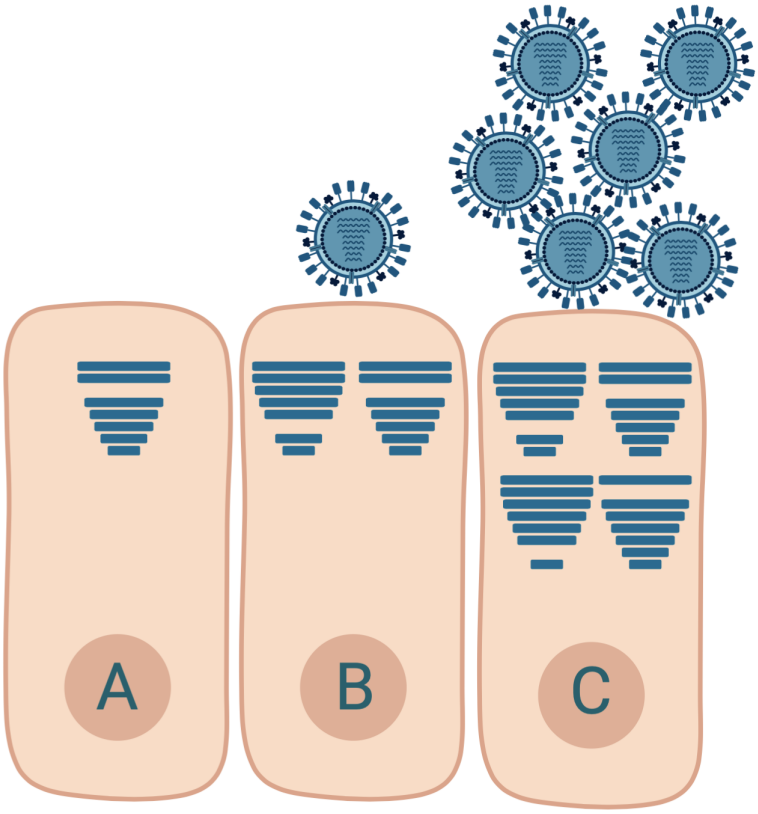



Figure 8