VIRUS EVOLUTION

Virus Evolution, 2023, **9(1)**, 1–14

DOI: https://doi.org/10.1093/ve/vead004

Advance access publication date: 19 January 2023

Research Article

Avian H7N9 influenza viruses are evolutionarily constrained by stochastic processes during replication and transmission in mammals

Katarina M. Braun,[1,†] Luis A. Haddock III,[1,†,‡] Chelsea M. Crooks,[1] Gabrielle L. Barry,[1] Joseph Lalli,[2] Gabriele Neumann,[3] Tokiko Watanabe,[4,5,6] Masaki Imai,[4,7] Seiya Yamayoshi,[4,7,§] Mutsumi Ito,[4] Louise H. Moncla,[8,**] Katia Koelle,[9,††] Yoshihiro Kawaoka,[3,4,7] and Thomas C. Friedrich[1,*]

[1]AIDS Vaccine Research Institute, Department of Pathobiological Sciences, University of Wisconsin-Madison, 585 Science Dr. Madison, WI 53711, USA, [2]Department of Genetics, University of Wisconsin-Madison, 425 Henry Mall Madison, WI 53706, US, [3]Influenza Research Institute, Department of Pathobiological Sciences, University of Wisconsin-Madison, 575 Science Dr. Madison, WI 53711, USA, [4]Division of Virology, Institute of Medical Science, University of Tokyo, 4 Chome-6-1 Shirokanedai Minato City, Tokyo 108-0071, Japan, [5]Department of Molecular Virology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka Suita City, Osaka 565-0871, Japan, [6]Center for Infectious Disease Education and Research (CiDER), Osaka University, 2-8 Yamadaoka Suita City, Osaka 565-0871, Japan, [7]The Research Center for Global Viral Diseases, National Center for Global Health and Medicine Research Institute, 1 Chome-21-1 Toyama Shinjuku City, Tokyo 162-8655, Japan, [8]Department of Pathobiology, University of Pennsylvania, 3800 Spruce Street Philadelphia, PA 19104, USA and [9]Department of Biology, Emory University, 1510 Clifton Road NE Atlanta, GA 30322, USA

[†]These authors contributed equally.

[‡]https://0000-0001-8916-5693

[§]https://orcid.org/0000-0001-7768-5157

[**]https://orcid.org/0000-0001-5722-1988

[††]https://orcid.org/0000-0002-0254-6141

*Corresponding author: E-mail: tfriedri@wisc.edu

Abstract

H7N9 avian influenza viruses (AIVs) have caused over 1,500 documented human infections since emerging in 2013. Although wild-type H7N9 AIVs can be transmitted by respiratory droplets in ferrets, they have not yet caused widespread outbreaks in humans. Previous studies have revealed molecular determinants of H7N9 AIV host switching, but little is known about potential evolutionary constraints on this process. Here, we compare patterns of sequence evolution for H7N9 AIV and mammalian H1N1 viruses during replication and transmission in ferrets. We show that three main factors—purifying selection, stochasticity, and very narrow transmission bottlenecks—combine to severely constrain the ability of H7N9 AIV to effectively adapt to mammalian hosts in isolated, acute spillover events. We find rare evidence of natural selection favoring new, potentially mammal-adapting mutations within ferrets but no evidence of natural selection acting during transmission. We conclude that human-adapted H7N9 viruses are unlikely to emerge during typical spillover infections. Our findings are instead consistent with a model in which the emergence of a human-transmissible virus would be a rare and unpredictable, though highly consequential, 'jackpot' event. Strategies to control the total number of spillover infections will limit opportunities for the virus to win this evolutionary lottery.

Key words: avian influenza; host adaptation; within-host evolution; transmission.

Introduction

The potential emergence of a novel avian influenza virus (AIV) in humans remains a significant public health threat (Taubenberger and Kash 2010; Lipsitch et al. 2016; Neumann and Kawaoka 2019). Despite recent advances in influenza surveillance and forecasting (Neher and Bedford 2015; Du et al. 2017; Morris et al. 2018), we still do not understand the evolutionary processes underlying the emergence of pandemic influenza viruses (Lipsitch et al. 2016; Neumann and Kawaoka 2019). H7N9 AIVs naturally circulate in aquatic birds and have been endemic in poultry since the virus's emergence in China in February 2013 (Gao et al. 2013). Since then, H7N9 viruses have spilled over into human populations, causing 1,568 confirmed infections with a case-fatality rate approaching 40 per cent across five epidemic waves (Su et al. 2017) with no recorded instances of human-to-human transmission. During the fifth and largest epidemic wave, some low-pathogenicity avian influenza (LPAI) H7N9 viruses acquired a novel motif in hemagglutinin (HA) that both facilitates systemic virus replication in chickens and enhances pathogenicity in mammals (Ke et al. 2017; Shen and Lu 2017; Wang et al. 2017; Zhang et al. 2017; Zhou et al. 2017); these viruses are designated as highly pathogenic avian influenza (HPAI) H7N9 viruses. Since natural selection can only act on the available genetic variation in a population, and limited H7N9 diversity in mammalian hosts could impede the efficiency of

mammalian adaptation, we sought to understand the evolutionary dynamics of both high- and low-pathogenicity H7N9 viruses replicating in mammals.

Both H7N9 and H5Nx AIVs are currently assigned high pandemic potential (Russell et al. 2012; Taft et al. 2015; Imai et al. 2017; Kiso et al. 2017; Qi et al. 2018; Sutton 2018; Yu et al. 2019; King et al. 2020). H7N9 viruses appear particularly threatening because, unlike H5N1 viruses, wild-type H7N9 viruses can be transmitted among ferrets via respiratory droplets (Watanabe et al. 2013; Zhang et al. 2013; Imai et al. 2017). In addition, H7N9 viruses are capable of binding human-type sialic acid receptors, in which sialic acids are linked to galactose in an α(2,6) configuration (Xiong et al. 2013; Zaraket et al. 2015; Imai et al. 2017). It is therefore unclear why there have been no documented cases of human-to-human transmission of H7N9 viruses (Zhou et al. 2018). Several factors may contribute to poor H7N9 transmissibility in humans, including preferential binding to avian-type α(2,3) sialic acid receptors, reduced polymerase activity at 33°C, which approximates the human upper respiratory tract temperature, and suboptimal HA stability, impacting successful membrane fusion (Imai et al. 2012; Galloway et al. 2013; Richard et al. 2013; Xiong et al. 2013; Linster et al. 2014; Zaraket et al. 2015; Xu et al. 2019). Nonetheless, ongoing isolated human spillover infections remain concerning because they provide an opportunity for adaptation of H7N9 viruses to human hosts, laying the groundwork for future AIV outbreaks. To our knowledge, only one previous study has examined H7N9 genetic diversity within hosts and reported lower levels of diversity in ferrets than in chickens (Zaraket et al. 2015).

In 2017, Imai et al. characterized the replication and pathogenicity of H7N9 AIV in ferrets (Imai et al. 2017). Using time series samples originally collected in that study, we performed whole-genome deep sequencing in technical duplicate and evaluated H7N9 evolutionary dynamics in seven ferret transmission events and in an additional nine infections not resulting in transmission. We compared the viral genetic diversity of these AIVs in a mammalian system to the diversity exhibited by a 2009 pandemic H1N1 virus in four ferret transmission events and one additional non-transmitting infection also observed in previously conducted studies (Imai et al. 2017, 2020). While stochastic forces played a significant role in viral evolution, we found little evidence for H7N9 mammalian adaptation in ferrets. These observations suggest that there is a high evolutionary barrier to the emergence of an H7N9 AIV capable of sustained spread in humans.

Results

H1N1 viruses transmit more frequently than H7N9 viruses in ferrets

We isolated and sequenced viral RNA (vRNA) from nasal washes collected from two previously published studies (Imai et al. 2017, 2020). Four of five donor ferrets infected with H1N1 virus (CA04) transmitted infection to a naive recipient ferret (80 per cent). By comparison, seven of sixteen ferrets infected with H7N9 AIV transmitted to their recipients (43.7 per cent) (Fig. 1). These group sizes are small, and while the rate of transmission from H1N1-infected ferrets exceeded that of H7N9-infected ferrets, the difference was not significant using a z-score test for two population proportions ($P = 0.1556$; 95 per cent confidence interval [CI] [−0.064, 0.79]) or a Mann–Whitney U test ($P = 0.17$).

Both the frequency and timing of transmission varied across H7N9 virus subgroups (Supplementary Table S1). Transmission occurred in one of four ferret pairs whose donors were infected with either the LPAI isolate (A/Anhui/1/2013; 'Anhui/1') or the HPAI isolate (A/Guangdong/17SF003/2016; 'GD/3'). The human GD/3 isolate contained neuraminidase (NA) inhibitor-sensitive (NA-289R) and -resistant (NA-289K) subpopulations; to separate these populations, each was rescued separately using reverse genetics and tested individually in the original 2017 study (Imai et al. 2017). Two of four ferrets infected with the drug-resistant variant, rGD/3-NA289K, transmitted to recipient animals (50 per cent), and three of four donors infected with the wild-type recombinant variant, rGD3/NA289R, transmitted to their recipients (75 per cent; Fig. 1).

H7N9 within-host diversity is dominated by low-frequency iSNVs

We mapped sequencing reads to the inoculating virus consensus sequence and called within-host variants found in both technical replicates (i.e. intersection intra-host single nucleotide [nt] variants [iSNVs]) in ≥1 per cent of reads. iSNV frequencies from 1 per cent to 99 per cent were highly concordant between technical replicates ($R^2 = 0.993$, Supplementary Fig. S1). All coding region changes are reported using H7 numbering for the H7N9 viruses and H1 numbering for the H1N1 virus, consistent with the numbering schemes used in Nextstrain (Hadfield et al. 2018). Combining all the data from donors and recipients, we detected iSNVs in one or more viruses at 879 different genome sites, of which 490 were synonymous, 386 were non-synonymous, and three were stop mutations. These stop mutations were found at low frequencies and were located near the ends of coding regions (nucleoprotein [NP] E292*, non-structural protein 1 [NS1] W203*, and nuclear export protein [NEP] Q119*).

We first characterized the within-host diversity in the H1N1 and H7N9 virus groups (Anhui/1, GD/3, rGD/3). The average number of unique iSNVs per ferret across all available time points varied significantly across virus groups ($P = 5.41 \times 10^{-10}$; one-way analysis of variance (ANOVA); Fig. 2A; Supplementary Fig. S2A). We found the fewest iSNVs in the H1N1 group ($n = 9$ ferrets), with an average of 24 iSNVs per ferret, ranging from 3 to 35 across all time points. This is similar to the number of seasonal influenza virus iSNVs reported in humans (Han et al. 2021). The number of iSNVs in animals infected with recombinant GD/3 viruses ($n = 13$ ferrets, grouping both rGD/3 viruses) was also low, with an average of 13 per ferret, ranging from 1 to 43 across all time points. We found more iSNVs in the ferrets infected with H7N9 biological isolates. Anhui/1 ($n = 5$ ferrets) had an average of 152 iSNVs per animal, ranging from 85 to 195 across all time points, while GD/3 ($n = 5$ ferrets) had an average of 109 iSNVs per ferret, ranging from 27 to 142 across all time points. This level of within-host diversity is not unexpected in animals directly inoculated with a high-dose viral isolate (Wilker et al. 2013; Leyson et al. 2022). The number of iSNVs found in each animal fluctuated over time and often trended downward in GD/3 and Anhui/1 donor ferrets (Supplementary Fig. S2).

Most iSNVs were detected at <10 per cent frequency (Fig. 2B). The number of iSNVs across all groups was compared to the proportion of iSNVs at various frequencies expected under a model of neutral evolution, as we have done previously (Braun et al. 2021), using a Kolmogorov–Smirnov test to compare the shape of the neutral distribution to the iSNV frequency distribution within each virus group. Looking at all iSNVs within a virus group, we found that the Anhui/1 distribution ($P = 0.006$) and the H1N1 distribution ($P = 0.006$), and to a lesser extent, rGD/3

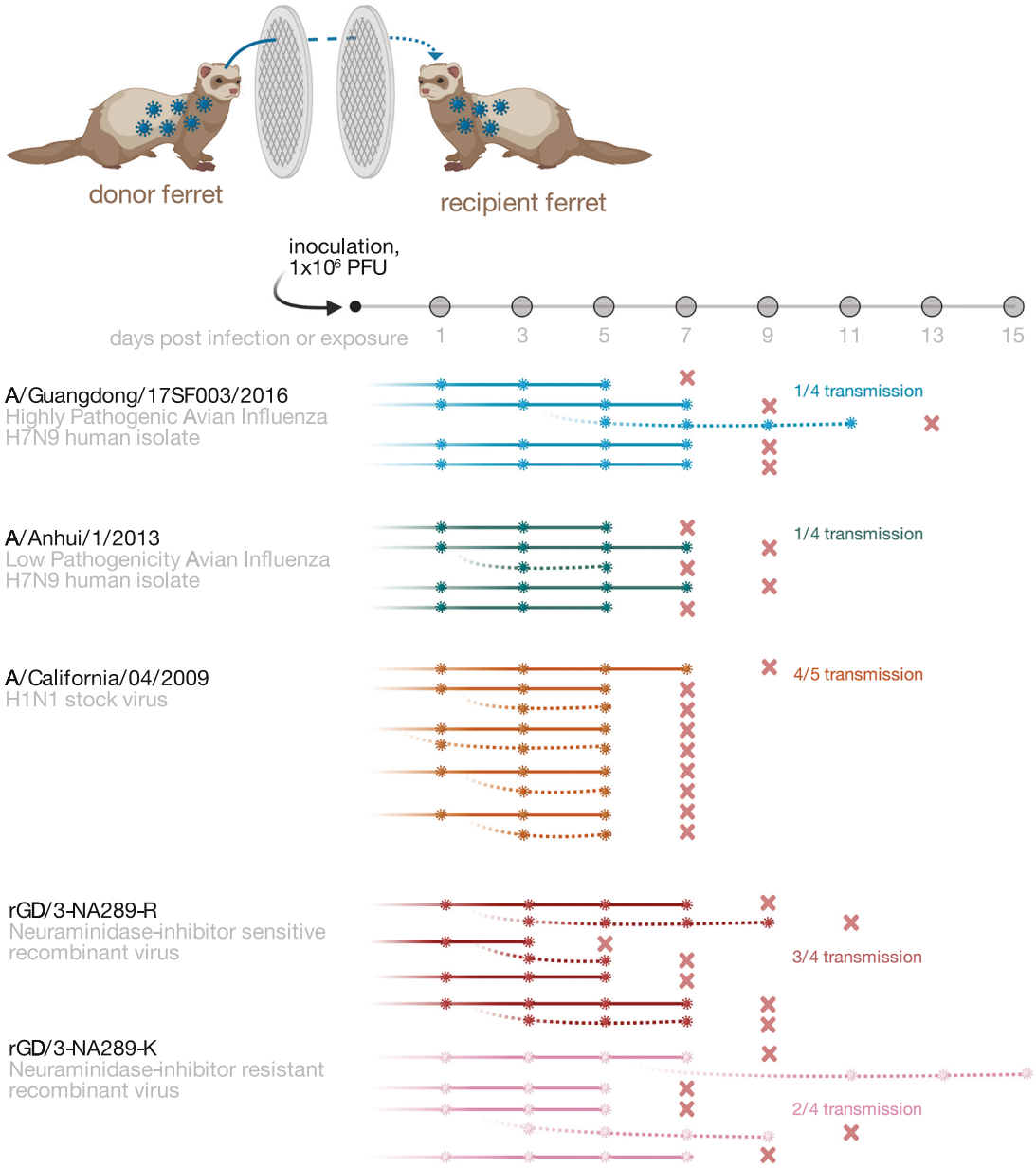

Figure 1. Overview of the experimental system and sampling timeline. Ferrets were inoculated intranasally with 10^6 PFU of a HPAI H7N9 isolate (A/Guangdong/17SF003/2016; blue), a LPAI H7N9 isolate (A/Anhui/1/2013; green), a H1N1pdm isolate (A/California/04/2009; orange), or recombinant HPAI H7N9 GD/3 viruses (rGD3-NA289R; red, rGD3-NA289K; pink). One day after inoculation, one naive recipient ferret was paired with each donor ferret. Nasal washes were collected from donor (solid lines) and recipient (dotted lines) ferrets every other day up to 15 DPI or until the ferret was euthanized. Cartoon virions denote days on which live virus was detected by plaque assay and viral RNA was extracted for whole-genome sequencing. The 'X' on donor and recipient timelines denotes the first negative sample as detected by quantitative RT-PCR (RT-qPCR). Figure was created using BioRender. Nasal wash titers can be found in the GitHub repository accompanying this manuscript.

($P = 0.031$), differed significantly from the neutral expectation, with low-frequency iSNVs at significantly higher proportions than expected under neutrality (Supplementary Fig. S3; Supplementary Table S2). The exception to this trend was GD/3 viruses, whose iSNV frequency distributions did not differ significantly from the neutral expectation ($P = 0.675$). This predominance of low-frequency iSNVs is consistent with viral population expansion and/or purifying selection acting within hosts. The frequency, genome location, and impact on amino acid sequence (synonymous vs. non-synonymous) for iSNVs detected longitudinally in HA across all ferrets are shown in Fig. 2C. iSNVs in all other gene segments are plotted in Supplementary Fig. S4.

H7N9 viral populations are subject to purifying selection, and genetic diversity is reduced following transmission

We used a common measure of nt diversity, π, to roughly assess signals of H7N9 viruses adapting to or diversifying within ferrets. This summary statistic quantifies the average number of pairwise differences per nt site among a set of viral sequences. In particular, we compared the nt diversity at non-synonymous sites (πN) to the nt diversity at synonymous sites (πS). In general, $\pi N/\pi S < 1$ indicates that purifying selection is acting to remove deleterious mutations from the viral population, and $\pi N/\pi S > 1$ indicates that diversifying selection is favoring new mutations,

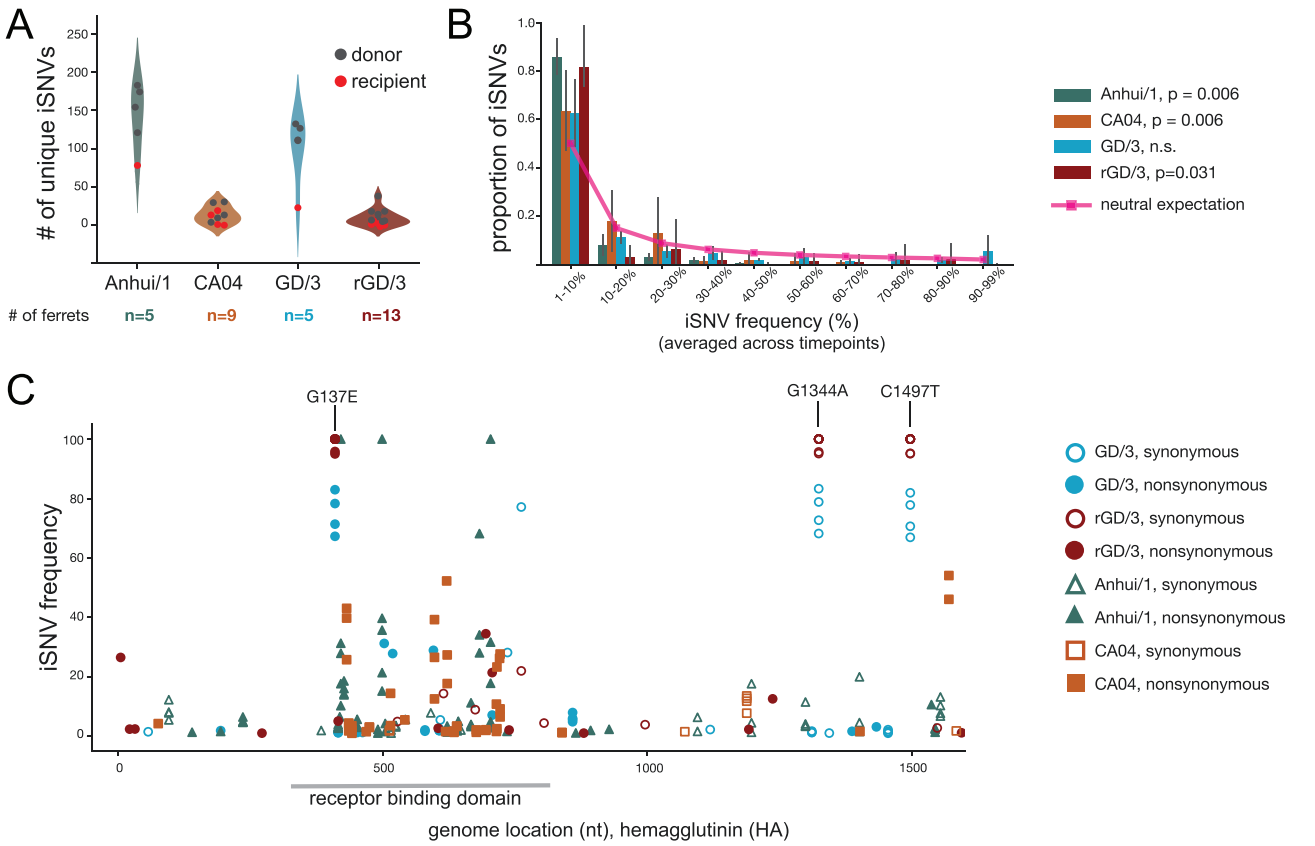

Figure 2. Frequency and location of iSNVs. (A) A violin plot showing the total number of unique iSNVs detected across all available time points plotted per virus group ($P = 6.83 \times 10^{-10}$; one-way ANOVA). The individual data points denote the number of iSNVs per ferret. Black and red dots indicate donor or recipient ferrets, respectively. (B) An iSNV frequency distribution showing the proportion of iSNVs detected per frequency bin. Error bars display variance (standard deviation) in the proportion of within-host iSNVs across ferrets. The solid magenta line denotes the expected distribution of variants in each frequency bin following a 1/x distribution in a neutral model of evolution. Virus group distributions were compared to the neutral distribution using the Kolmogorov–Smirnov test. (C) All iSNVs detected longitudinally in HA are plotted for GD/3 and rGD/3 iSNVs (circles), Anhui/1 (triangles), and H1N1 (squares). Synonymous iSNVs are denoted with open symbols and non-synonymous iSNVs with closed symbols. Three iSNVs found in multiple HPAI samples at high frequencies are labeled; G137E and two synonymous mutations at nucleotides 1,344 and 1,497. iSNVs in all other gene segments can be found in Supplementary Fig. S4.

which might be expected in the case of an AIV adapting to a mammalian host. When πN approximates πS, this suggests that allele frequencies are primarily determined by non-selective forces, such as population size changes or genetic drift, i.e. stochastic shifts in allele frequencies related to population size (Zhao and Illingworth 2019).

First, we analyzed viral diversity in all infected animals in the study. In most ferrets infected with H1N1 viruses, πS exceeded or was equal to πN (Fig. 3A, orange; Supplementary Fig. S5). πS was significantly greater than πN in polymerase basic protein 2 (PB2), PA, and NA, and πN never significantly exceeded πS in any gene segment at any time point. Together, these findings provide little evidence that H1N1 virus populations in ferrets are subject to strong selection and give only weak signals of purifying selection at some time points. Weak purifying selection might be expected for a mammalian virus replicating in a mammalian host, where most mutations away from a fit consensus are likely to be deleterious. Somewhat surprisingly, πN and πS comparisons gave similar results for H7N9 viruses. πS significantly exceeded πN in all genes apart from NA in the GD/3 group and all genes apart from NA and HA in Anhui/1 group (Fig. 3A, blue and green), providing some evidence for weak purifying selection. πN was moderately elevated relative to πS (πN/πS) in the HA globular head in H7N9 viruses (GD/3—1.13 [95 per cent CI, 1.067–1.203], Anhui/1—3.59 [95 per cent CI, 3.506–3.678]; Supplementary Table S4). However, these signs of diversifying selection in the globular head of H7N9 were low when compared to the same region in H1N1 viruses (πN/πS—8.37 [95 per cent CI, 8.310–8.432]; Supplementary Table S4). The rGD/3 group had fewer iSNVs contributing to diversity measurements, but nonetheless we found no evidence of diversifying selection as πN never significantly exceeded πS (Fig. 3A, pink). Combining these analyses, we find little evidence for strong selection on HPAI or LPAI H7N9 viruses replicating in ferrets on the whole-gene level. It is important to note that these whole-gene or whole-domain analyses could mask the action of selection on individual sites.

We also compared nt diversity in donor–recipient pairs before and after transmission. We found that genome-wide nt diversity (π) did not significantly differ between donor and recipient ferrets in the H1N1 group (Supplementary Figs S6A and S7, $P = 0.125$, paired *t*-test). However, in the H7N9 group, π in the donor ferrets was significantly greater than the recipient ferrets (Fig. 3B, $P = 0.005$; paired *t*-test), showing that overall H7N9 genetic diversity is lost during transmission. As we have done previously (Wilker et al. 2013; Moncla et al. 2016), we looked for selective sweeps by comparing the change in πN and πS in each gene segment for paired donor and recipient ferrets. πN and πS at the gene level did not differ significantly between

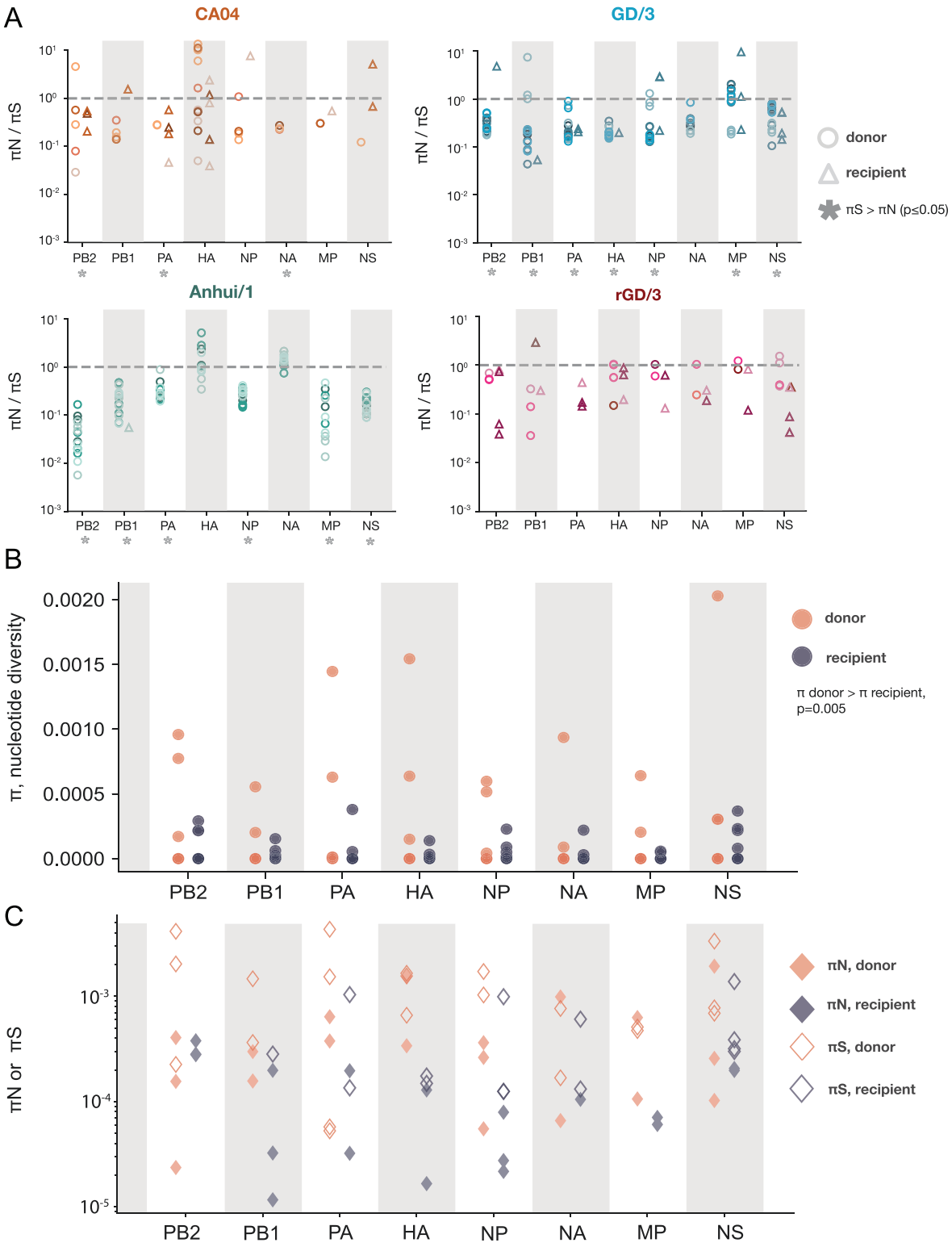

Figure 3. Patterns of viral genetic diversity within ferret hosts. (A) πN/πS nt diversity is plotted for each gene segment. Each data point represents a single ferret. Circles denote donor ferrets and triangles denote recipient ferrets. πN equal to πS ($y = 1$) is represented with a dashed gray line. Stars denote instances when πS is significantly greater than πN. (B) Genewise nt diversity is plotted for all H7N9 transmission pairs (Anhui/1, GD/3, and rGD/3). The donor ferrets are shown in orange and the recipient ferrets are shown in gray. Nucleotide diversity did not significantly differ between donor and recipient ferrets in any single gene segment but is significantly lower following transmission in recipient ferrets when assessing the entire genome ($P = 0.005$; paired *t*-test). (C) πN (solid symbols) and πS (open symbols) in the H7N9 donors and recipients are plotted for each gene segment. πN and πS in the donor ferrets are denoted by the orange diamonds. πN and πS in the recipient ferrets are denoted by the gray diamonds. Similar data as (B) and (C) are plotted for H1N1.

donor and recipient ferrets. This was true across all H1N1 transmission pairs (Supplementary Figs S6B and S7) and all H7N9 transmission pairs (Fig. 3C; Supplementary Fig. S7). These findings suggest that during transmission of these viruses, genetic diversity was purged equally across gene segments with no evidence for a selective reduction in diversity of any particular segment.

Airborne transmission results in a dramatic shift of iSNV frequencies

Strength of animal studies is that we are able to conduct frequent longitudinal sampling from each infected host, allowing us to follow individual iSNV frequencies within donors and following transmission into recipients. In general, in the absence of selection, the likelihood that an iSNV is transmitted may be proportional to its frequency in the donor at the time of transmission. However, there are exceptions to this trend. For example, one iSNV that encodes a glycine-to-glutamic acid substitution at HA position 137 (HA G137E) in the GD/3 transmission pair was present at 81 per cent at 1 day post-inoculation (DPI) in the donor ferret and decreased to a sub-consensus frequency (39.3 per cent) by 7 DPI. Despite this marked downward trend in the donor animal, G137E was transmitted to the recipient at 5 DPI and remained at ≥99 per cent from this time point onward (Fig. 4). Conversely, a mutation in the matrix gene encoding an arginine-to-lysine substitution at position 210 in matrix protein 1 (M1) (R210K) was never detected in the donor ferret above 1 per cent yet was nearly fixed (97.5 per cent) at the first time point post-infection in the recipient. Interestingly, M1 R210K then decreased in frequency in the recipient and was found at 54.5 per cent at 9 DPI. We observed similar patterns in synonymous variants. For example, a synonymous A-to-G change at nt 2,037 in the polymerase basic protein 1 (PB1) gene segment was never found above 6 per cent frequency in the donor ferret, but was nearly fixed immediately following transmission, and again decreased in frequency to 57.57 per cent by 9 DPI in the recipient ferret. It is important to note that down-trending iSNV frequencies may be influenced by stochastic fluctuations or genetic drift within the population and are not necessarily a function of fitness.

Even in the case of amino acid substitutions that may be adaptive in humans, variant frequencies were not maintained across the transmission event. For instance, a valine-to-isoleucine substitution at position 219 in M1 may play a role in AIV adaptation to mammals but appears to be under negative selection in avian hosts (Furuse et al. 2009). An iSNV encoding this substitution increased in frequency from 34.7 per cent to 84.3 per cent in one donor but nonetheless could not be detected on any of the sequencing reads from the recipient immediately after transmission. M1 V219I then appeared to arise *de novo* in the recipient ferret at a later time point, suggesting that positive selection may act on individual sites within individual hosts, despite a lack of evidence for positive selection at the gene level at the time of transmission. A similar trend was observed for an aspartic acid-to-asparagine substitution at position 701 in PB2, a mutation associated with enhanced replication in mammals (Subbarao, Kawaoka, and Murphy 1993; Gabriel et al. 2005; Steel et al. 2009). Note that variants that increase replicative fitness *within* a host may not necessarily enhance transmissibility *between* hosts. In addition, we cannot formally exclude the possibility that these variants were transmitted and were present at the earliest time points in recipient animals below the limit of detection. No variants consistently increased in frequency over time in the Anhui/1 or rGD/3 groups.

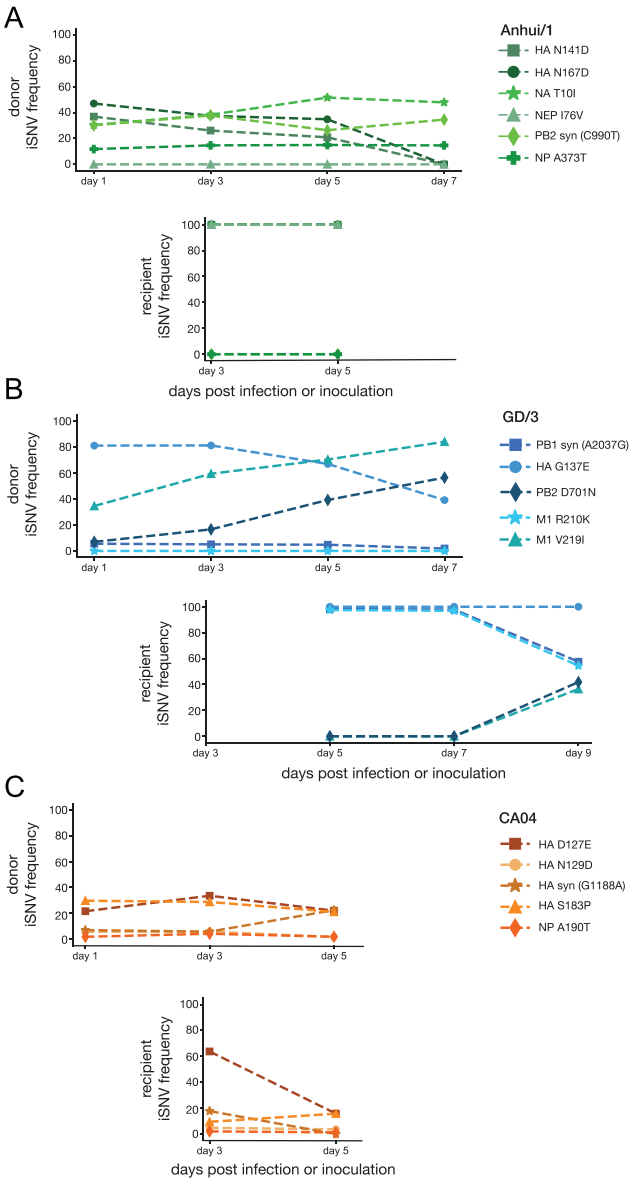

Figure 4. Frequency dynamics of iSNVs across the transmission event. The frequencies of representative iSNVs are plotted over time in donor ferrets (top plot) and following transmission into the associated recipient ferret (bottom plot) in the (A) Anhui/1 transmission pair, (B) GD/3 transmission pair, and (C) H1N1 transmission pairs. Each iSNV is plotted as y = 0 at time points when it is not detected at ≥1 per cent frequency and is absent at time points when no viral RNA was recovered for deep sequencing. We did not plot iSNV frequencies beyond Day 9 in the recipient ferret.

Unlike iSNV dynamics in the H7N9-transmission events, which resulted almost exclusively in fixation or loss of iSNVs in the recipient ferrets, eight iSNV sites in the H1N1 donor ferrets remained polymorphic at intermediate frequencies immediately following transmission (e.g. HA D127E and S183P; Fig. 4C).

These results highlight how airborne transmission can dramatically alter the frequency of influenza virus variants across hosts. While the vast majority of variants are lost at the time of transmission, we observed that potentially deleterious mutations can be transmitted and putatively adaptive ones may not be. These observations suggest that natural selection is generally weak and is outweighed by stochastic processes during H7N9

influenza virus transmission in this model. Additional work to characterize the fitness benefit or cost of individual mutations is needed to determine the full range of evolutionary forces acting within individual hosts.

Airborne transmission of H7N9 viruses in ferrets is characterized by a very narrow transmission bottleneck

Narrow transmission bottlenecks, in which a very small number of viruses found a new infection, cause a founder effect and purge most low-frequency iSNVs, regardless of their fitness (Edwards et al. 2006; McCrone and Lauring 2018). Conversely, wide transmission bottlenecks allow more viruses to initiate infection, reducing the chance that beneficial or rare variants are lost. The vast majority of iSNVs detected in all H7N9 and H1N1 donor ferrets were lost during transmission and were not found in recipients. However, a very small number of iSNVs in the Anhui/1 and GD/3 donor ferrets were transmitted and fixed (>99 per cent frequency) in the recipient ferret (Fig. 5A).

To infer transmission bottleneck sizes, we applied the beta-binomial inference method (Leonard et al. 2017) to estimate the number of transmitted viruses that could account for the pattern of iSNVs observed immediately before and after transmission for each pair. These estimates suggest that fewer than eleven viruses initiated infection in all recipient ferrets. The maximum likelihood estimate for the mean transmission bottleneck sizes for the H1N1 pairs was 5.72 ($n = 4$ pairs; 95 per cent CI: 2.57–10.00; Fig. 5B). We evaluated seven transmission events in the H7N9 group: one Anhui/1 pair, one GD/3 pair, and five rGD/3 pairs. However, two of the rGD/3 transmission events (Pairs 9 and 11) were excluded because the donor had no detectable polymorphic sites, making it impossible to estimate a bottleneck size. The maximum likelihood mean transmission bottleneck size among the remaining H7N9 pairs was 1.40 (95 per cent CI: 1.02–2.96; Fig. 5B). The mean H1N1 transmission bottleneck estimate is larger (looser) than the combined H7N9 estimate; although with only nine transmission pairs informing these estimates, and a slight overlap of the CIs between groups, this difference did not reach statistical significance.

Ferret iSNVs are not significantly enriched in human-associated H7N9 substitutions

Each H7N9 infection in a human has so far represented a unique avian spillover event. If selection is strong at a given site in the

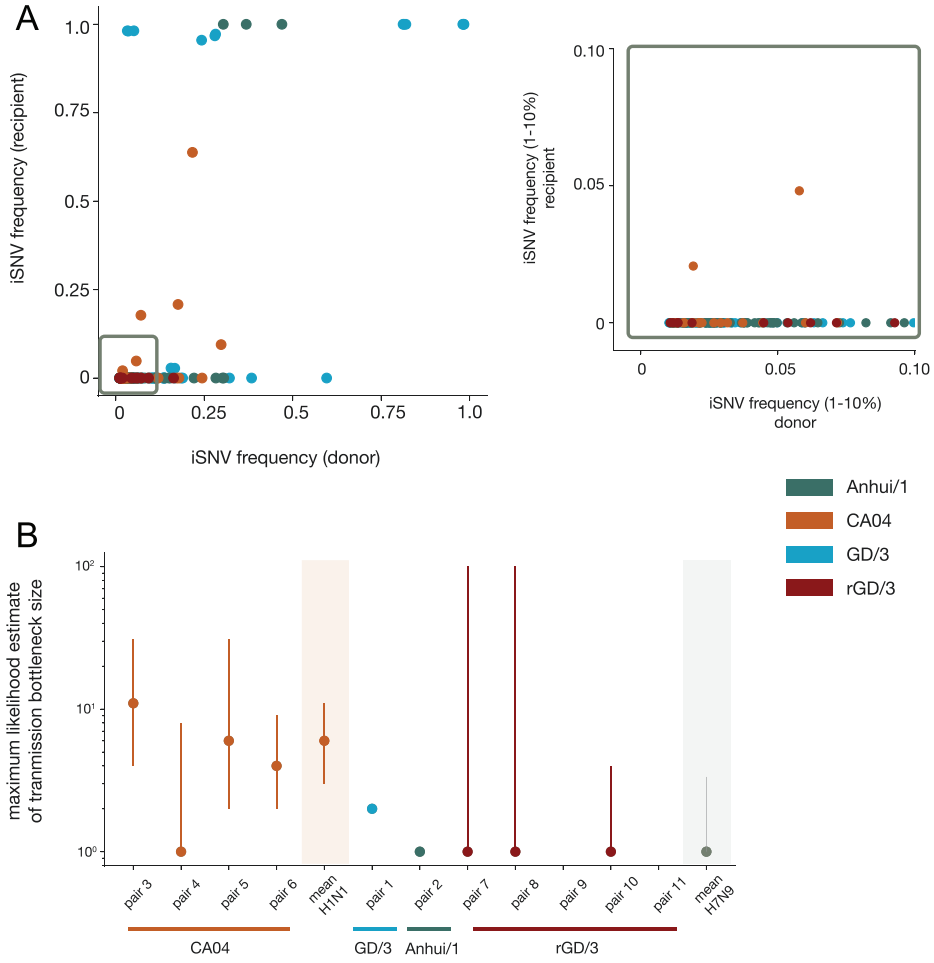

Figure 5. H1N1 and H7N9 transmission bottlenecks in ferret donor–recipient pairs. (A) 'TV plots' showing intersection iSNV frequencies in all eleven donor–recipient pairs. The inset box highlights low-frequency iSNVs (1–10 per cent). Colors denote virus groups. The plot uses the first time point at which the virus became detectable in the recipient (y axis) and the time point preceding such detection for the donor (x axis). (B) Maximum likelihood estimates for mean transmission bottleneck size with 95 per cent CI in individual donor–recipient pairs. Bottleneck sizes could not be estimated for two pairs (rGD/3 pair 9 and pair 11) because there were no polymorphic sites detected in the donor. The mean H1N1 estimate was calculated using Pairs 3, 4, 5, and 6. The mean H7N9 estimate was calculated using Pairs 1, 2, 7, 8, and 10.

genome, then we might observe mutations at that site in multiple independent infections. To look for such a signal, we compared non-synonymous iSNVs detected in this study ($n = 262$) to non-synonymous single nucleotide polymorphisms (SNPs) found at the most distal nodes in Nextstrain's H7N9 phylogenetic trees for all gene segments ($n = 2{,}071$ unique isolates) (Hadfield et al. 2018), representing more than 66 per cent of all sequences available on the Global Initiative on Sharing Avian Influenza Data (GISAID) ($n = 3{,}131$ isolates). This tree was created from publicly available H7N9 sequences collected from birds ($n = 621$) and humans ($n = 1{,}023$). Among the list of variants shared between our data and Nextstrain's, we looked for those that were proportionally overrepresented in sequences from humans. We excluded iSNVs detected in the rGD/3 samples because the inoculum was nearly clonal and few iSNVs were detected in ferrets.

Considering all iSNVs we detected, around half (46.6 per cent, 77/165) of the Anhui/1 iSNVs and 36.1 per cent (35/97) of the GD/3 iSNVs can be found in at least one bird or human H7N9 sequence on Nextstrain. Of the seventy-seven Anhui/1 iSNVs, forty-nine were in human sequences, ninie were in bird sequences, and nineteen were found in both. Of the thirty-five GD/3 iSNVs, twenty were in human sequences, eight were in bird sequences, and seven were found in both. A complete summary of iSNVs and their respective occurrences in human- or bird-derived sequences from all gene segments is found in Supplementary Table S3. As a group, neither the Anhui/1 variants nor the GD/3 variants were significantly enriched in human sequences compared to bird sequences (GD/3 $P = 0.052$, Anhui $P = 0.049$; Fisher's exact test).

We plotted the number of occurrences of each of our iSNVs in bird and human sequences in Fig. 6. Four Anhui/1 iSNVs were significantly enriched among bird sequences: HA L235Q, HA D289N, NA V22I, and non-structural (NS) R44K. Two iSNVs were enriched among mammalian sequences: PB2 K627E in GD/3 (discussed here) and PB2 D701N in Anhui/1, which is linked to mammalian adaptation. We also identified a few putative mammalian-adapting mutations that arose sporadically in ferrets but were not enriched among human surveillance sequences. These mutations included PB2 K562R (Quan et al. 2018), HA A143T (Ning et al. 2019), and matrix protein (MP) V219I (Furuse et al. 2009).

A glutamic acid-to-lysine change at residue 627 in PB2 (E627K) is a key mutation previously shown to improve polymerase processivity in mammalian hosts (Subbarao, Kawaoka, and Murphy 1993; Gabriel et al. 2005; Steel et al. 2009). The Anhui/1 isolate's consensus sequence, which we used as our reference, already contained a lysine at this residue (Fig. 6B, red asterisks). Therefore, we report this iSNV as the glutamic acid-to-lysine change (E627K). Importantly, although the K627E iSNV met criteria for inclusion in this surveillance analysis, we only detect K627E in a single ferret, at a single time point, near 1 per cent frequency (Fig. 6B).

Discussion

The evolutionary pathways by which AIVs might adapt to cause widespread outbreaks in humans are poorly defined. Our study examined the viral dynamics of wild-type LPAI and HPAI H7N9 viruses in a ferret model, a well-studied mammalian system which closely recapitulates human respiratory physiology and transmission (Belser et al. 2020). We found that H7N9 viruses in ferrets are subject to mild purifying selection and no evidence that HPAI H7N9 iSNVs arising in ferrets were significantly more likely to appear in human rather than bird surveillance sequences. These findings are consistent with a virus that is capable of efficient replication, if not transmission, in humans and is not under strong within-host selective pressure in mammalian systems. However, our results shed light on several significant barriers to generation, selection, and, in particular, onward transmission of further mammal-adapting mutations. We speculate that short infection times, purifying selection, and narrow, non-selective transmission bottlenecks combine to limit the capacity of H7N9 viruses to adapt during typical spillover infections.

Some have theorized that the rate-limiting step in viral host switching is not the generation of adaptive variants within hosts, but the successful transmission of these variants between hosts (Russell et al. 2012). Our data support this hypothesis. We detected multiple iSNVs in donor ferrets putatively linked to mammalian adaptation (PB2 K526R, PB2 D701N, HA A143T, and M1 V219I) or enriched in human surveillance sequences (PB2 D701N) that were not transmitted onward. Indeed, the vast majority of iSNVs arising in ferrets were lost during transmission. This is the result of an extremely narrow transmission bottleneck—our estimates indicate that new infections were founded by 1–3 H7N9 viruses. Our quantitative results thus confirm a speculation previously made by Zaraket et al. in a study of LPAI H7N9 transmission in ferrets (Zaraket et al. 2015).

In general, in the absence of selection, the higher the frequency of an iSNVs in the donor, the greater its probability of being transmitted. However, very narrow transmission bottlenecks can have unpredictable effects on the virus population that found infection in a new host. Strong selection could result in a narrow bottleneck that favors the transfer of particular variants. On the other hand, very small founding population sizes could reduce the efficiency of natural selection during transmission. Mutations that increase a virus's replicative capacity in one host may not necessarily be favored for transmission or could simply be lost through stochastic effects. Transmission may therefore sometimes reduce overall viral fitness. We saw two examples of this when M1 V219I and PB2 D701N, both previously linked to mammalian adaptation (Furuse et al. 2009), appeared to be lost at the time of transmission and arose *de novo* once again in the recipient ferret. Therefore, when transmission bottlenecks are narrow and selection at the time of transmission is not sufficiently strong, common, mildly deleterious variants may become fixed and low-frequency adaptive variants are very likely to be lost, ultimately slowing the pace of viral adaptation. Here, the ferret model system allows us to analyze unambiguously linked donors and recipients and quantify bottleneck sizes.

Despite the success of mass poultry immunizations with the H5/H7 bivalent vaccine, H7N9 viruses are likely to continue to evolve and sporadically spillover into humans. H7N9 remains common in poultry, and large populations of unvaccinated ducks and wild waterfowl may serve as reservoirs for ongoing adaptation and reassortment of HPAI H7N9 viruses (Ma, Yang, and Fang 2019). Furthermore, Wu et al. characterized H7N9 viruses collected at live poultry markets and farms between 2017 and 2019 and found evidence for accelerated evolution away from the vaccine strain in the 2018–9 swabs (Wu et al. 2021). Our study suggests that H7N9 viruses are somewhat unlikely to acquire enhanced human transmissibility during a single infection. However, this risk is additive and may become non-negligible with an increasing number of human spillover infections. This emphasizes the importance of population health interventions to reduce opportunities for avian viruses to spill over into humans and, even more so, the opportunity for avian and mammalian viruses to co-infect a single host. These

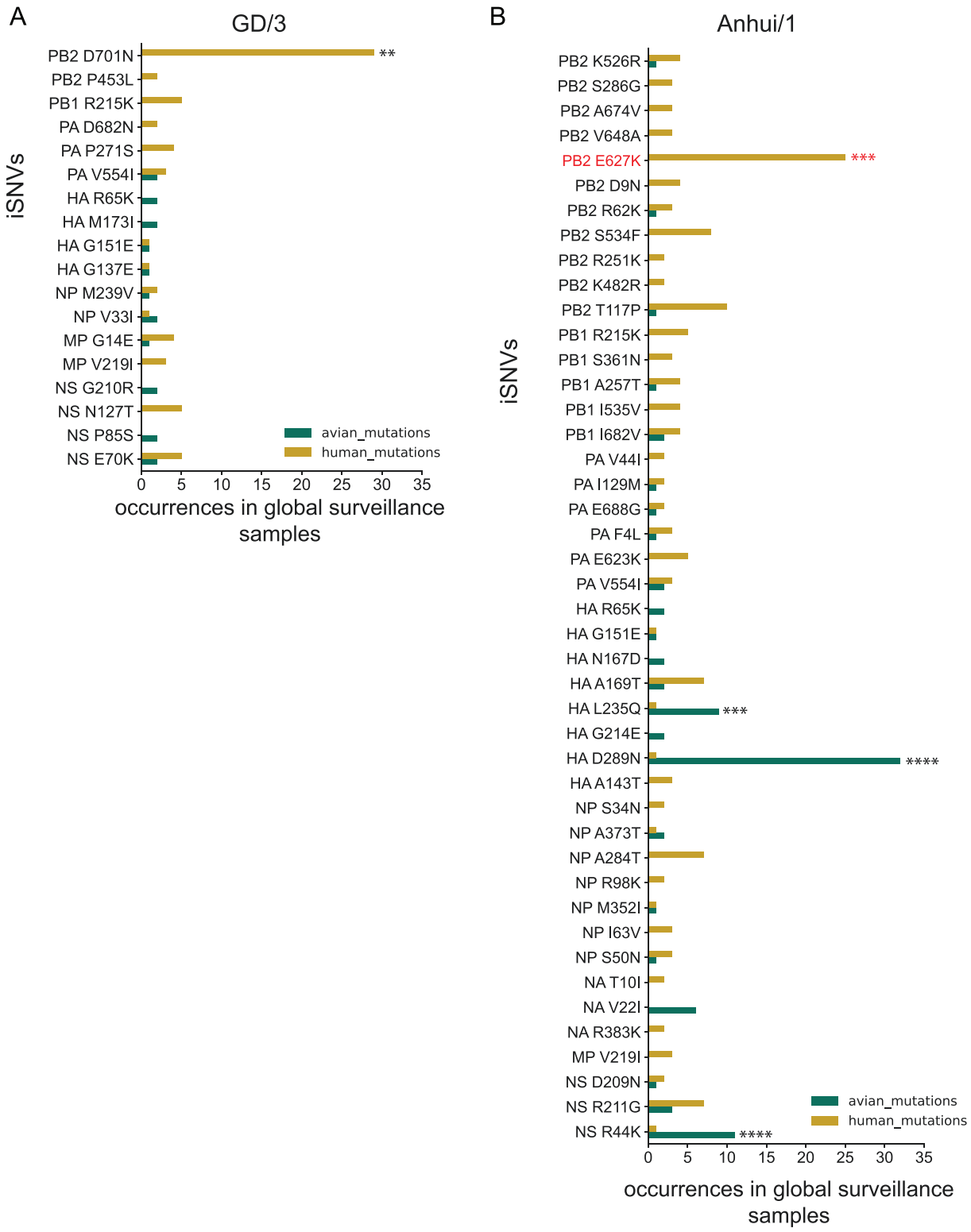

Figure 6. iSNVs found in H7N9 global surveillance sequences. Occurrences of non-synonymous iSNVs in H7N9 global surveillance sequences in the (A) GD/3 and (B) Anhui/1 datasets. This figure displays iSNVs with more than two occurrences. A plot with all non-synonymous iSNVs can be found in Supplementary Fig. S8. Fisher's exact test was used to assess for significant enrichment in human or bird sequences, and these results are shown with asterisks (* $P = 0.05$–0.01, ** $P = 0.01$–0.001, *** $P = 0.0001$–0.0000).

interventions are reviewed in full in the China–World Health Organization (WHO) Joint Mission on Human Infection with Avian Influenza A(H7N9) Virus (WHO 2013) and include, but are not limited to, continued poultry vaccination, culling, poultry movement restrictions, distancing at live animal markets, and others (Short et al. 2015).

We speculate that the evolutionary processes and the patterns of selection acting on wild-type avian viruses in a mammalian

system are distinct from those acting on reassortant or engineered avian viruses. Ferguson et al. (2013) and Sobel Leonard et al. (2016) have previously described the concept of viral fitness landscapes, which are defined by 'fitness peaks' and 'fitness valleys' resulting from unique combinations of virus and host genotype interactions. The topography of this landscape is expected to change with shifting host immune environments, epistatic interactions, new reassortant genotypes, etc. The fitness peaks, areas of high viral fitness, on this landscape are occupied by viruses like seasonal H1N1 in a human (a wild-type mammalian virus in a mammal) and H7N9 in a chicken (a wild-type avian virus in a bird). These viruses are already well adapted to their hosts and have limited nearby evolutionary space to become fitter; that is, mutations in these viruses will tend to be deleterious, moving the virus away from a local fitness peak. Such viral populations are likely to be characterized by purifying selection and genetic drift because any new mutation is unlikely to possess a large enough selection coefficient to be positively selected in the setting of an acute infection. We saw no evidence of adaptive evolution within hosts and a regime in which genetic drift and purifying selection dominate. Altogether, this suggests that most mutations in H7N9 viruses replicating in ferrets are deleterious, as we would expect for viruses near local fitness maxima. It is important to note, however, that H7N9 viruses replicating in mammals could nonetheless occupy relatively small local fitness maxima, akin to small hills in a wide valley. That is, the existence of these local fitness maxima does not exclude the possibility that H7N9 viruses could acquire significantly greater fitness within mammals, merely that pathways to such increased fitness likely pass through areas of reduced fitness.

Interestingly, we also did not find evidence for selection during transmission of avian H7N9 viruses in mammals. This contrasts with our previous studies in which ferret transmission of genetically engineered H5N1 and '1918-like' H1N1 AIVs was associated with selective sweeps acting on HA (Watanabe et al. 2014; Moncla et al. 2016). In these sweeps, selection appeared to favor transmission and/or replication of only a subset of HA sequences in recipient ferrets, as evidenced by sharp decreases in genetic diversity in HA but not other gene segments. We posit that these engineered viruses resemble hypothetical 'transitional states' distant from local fitness maxima. For such viruses, many new mutations may confer fitness advantages and be positively selected within hosts and/or swept to fixation during transmission. We might expect such unfit viruses to be unstable and therefore likely transient in nature. However, selective sweeps between hosts or rapid diversification within a host may be 'evolutionary signatures' that indicate viruses with heightened pandemic potential. Importantly, surveillance approaches aimed at detecting evolutionary signatures of within- and/or between-host selection would be agnostic to AIV subtype, genetic background, and are less likely to be confounded by epistasis than traditional approaches that query lists of mutations of concern. Such sequence-agnostic approaches could therefore provide an important complement to traditional risk assessments for AIVs, particularly for subtypes for which there are little data on the phenotypic impact of specific mutations.

Like most ferret studies, the results of these experiments are limited by relatively small sample sizes and the biological differences between ferrets and humans. Ferrets are the most relevant animal model system for studying respiratory infections; however, there are anatomical, physiological, and immunological differences between ferrets and humans, highlighted by the fact that H7N9 AIVs are transmitted between ferrets but are not known to do so between humans (Wong et al. 2019). Accordingly, there may be fewer or different evolutionary pressures acting on the H7N9 viruses in ferrets compared to humans. We also included clonal, recombinant viruses (rGD/3), which, as stated previously, have less diversity than viral isolates and will thus be subject to different evolutionary forces. In addition, direct inoculation of donor ferrets does not fully recapitulate a human spillover infection. In particular, high-dose inoculation with a biological isolate may allow a greater number of more diverse genomes to establish infection than in natural infections. Patterns of genetic diversity might differ in the case of H7N9 transmitting directly from a bird to a human. Our results should be corroborated by further investigations, including natural spillover infections, if possible, and targeted virological and epidemiological research (Buhnerkempe et al. 2015).

Assessing zoonotic risk and adaptive potential of AIV remains a critical public health challenge. By characterizing patterns of within-host diversity, quantifying the stringency and patterns of selection acting on typical transmission bottlenecks, identifying the fate of known adaptive mutations within individuals and across transmission events, and characterizing typical and non-typical evolutionary signatures, we can continue to assemble an understanding of AIV evolution. We hope that these methods may be applicable to other zoonotic respiratory viruses, including the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in order to better assess their ongoing adaptive potential.

Materials and methods

Ferret transmission experiments, sample collection, and availability

No new transmission experiments were performed as part of this study. We used residual nasal swabs collected from ferrets as part of previously published studies (Imai et al. 2017, 2020). Animal studies were approved prior to the start of the study by the Institutional Animal Care and Use Committee and performed in accordance with the Animal Care and Use Committee guidelines at the University of Wisconsin-Madison. In this previously described study, four groups of four ferrets were directly inoculated with various H7N9 viruses (1×10^6 plaque-forming units (PFU)) and one group of two ferrets was infected with an H1N1pdm virus for comparison (inoculated or donor ferrets). Samples from three of the four total H1N1 pairs were derived from a separate and similar study by the Kawaoka group (Imai et al. 2020). The H7N9 viruses used in this study included a high-pathogenic human isolate—A/Guangdong/17SF003/2016 ('GD/3'), two recombinant viruses thaat have an arginine or lysine at position 289 (H7 numbering) to confer NA-inhibitor sensitivity or resistance, respectively, on the background of the GD/3 consensus sequence—rGD/3-NA289R and rGD/3-NA289K ('rGD/3'), and a low-pathogenic H7N9 virus—A/Anhui/1/2013 ('Anhui/1'). The H1N1 comparator group was inoculated with a representative 2009 pandemic virus—A/California/04/2009 ('CA04').

Four (GD/3, rGD/3-NA289R, rGD/3-NA289K, Anhui/1) or six (CA04) serologically confirmed naive ferrets (recipient ferrets) were placed in enclosures adjacent to the donor ferret (separated by ~5 cm) on Day 2 post-inoculation. Pairs of ferrets were individually co-housed in adjacent wireframe enclosures, which allow for the spread of virus by respiratory droplet (occurring via either small or large airborne respiratory particles), but not by direct or indirect (via fomite) contact. Nasal washes were collected from donor ferrets on Day 1 after inoculation and from recipient ferrets on Day 1 after co-housing and then every other day

(for up to 15 days) for virus titration. Virus titers in nasal washes were determined by plaque assay on Madin-Darby canine kidney (MDCK) cells. vRNA was available for isolation from nasal wash samples collected from donor ferrets on Days 1, 3, 5 and 7 post-infection and from recipient ferrets on Days 3, 5, 7, 9, 11, 13, and 15 post-infection.

Viruses

A/Guangdong/17SF003/2016 was propagated in embryonated chicken eggs to prepare a virus stock after being isolated from a fatal human case treated with oseltamivir (Zhu et al. 2017). A/Anhui/1/2013 was also propagated in embryonated chicken eggs after being isolated from an early human infection (Watanabe et al. 2013). A/California/04/2009 was propagated in MDCK cells and was originally obtained from the Centers for Disease Control (Itoh et al. 2009). Recombinant viruses, rGD3-NA289K and rGD3-NA289R, were generated by plasmid-based reverse genetics as previously described by Neumann and Kawaoka (2015). We sequenced this inoculum and plot iSNVs in Supplementary Fig. S9. Unfortunately, inoculums for A/Anhui/1, A/California/04, rGD3-NA289K and rGD3-NA289R, were not available to sequence.

Template preparation

Total nucleic acids including vRNA were extracted from nasal washes and were reverse transcribed (RT-PC) using Superscript IV VILO (Invitrogen, USA) and the Uni12 primer (AGCAAAAGCAGG) in a total reaction volume of 20 μl. The complete reverse transcription protocol can be found here: https://github.com/tcflab/protocols/blob/master/VILO_Reverse_Transcription_h7n9_GLB_2019-02-15.md.

Single-stranded complementary DNA was used as a template for PCR amplification to amplify all eight genes using segment-specific primers using high-fidelity Phusion 2X DNA polymerase (New England BioLabs, Inc., USA). Primer sequences are available in the GitHub repository accompanying this manuscript (Braun 2023). PCR was performed by incubating the reaction mixtures at 98°C for 30 s, followed by thirty-five cycles at 98°C for 10 s, 51°C–72°C depending on gene segment for 30 s, 72°C for 120 s, followed by a final extension step at 72°C for 5 min. The complete PCR protocol, including segment-specific annealing temperatures and primer sequences, can be found here: https://github.com/tcflab/protocols/blob/master/Phusion_PCR_h7n9_GLB_2019-02-21.md.

PCR products were separated by electrophoresis on a 1 per cent agarose gel (Qiagen, USA). The bands corresponding to full-length gene segments were excised, and the DNA was recovered using QIAquick gel extraction kit (Qiagen, USA). To control for RT-PCR and sequencing errors, especially in low-titer samples, all samples were prepared in complete technical replicate starting from vRNA (McCrone and Lauring 2016; Grubaugh et al. 2019). We sequenced samples with low or no coverage, typically from low-titer samples, a third time, and merged sequencing reads with only one of the first two replicates to minimize coverage gaps.

Deep sequencing

Gel-purified PCR products were quantified using Qubit dsDNA high-sensitivity kit (Invitrogen, USA) and were diluted in an elution buffer to a concentration of 1 ng/μl. All segments originating from the same samples with a non-zero concentration as determined by high-sensitivity DNA (hsDNA) Qubit (Invitrogen, USA) were pooled equimolarly, and these genome pools were again quantified by Qubit. Each equimolar genome pool was diluted to a final concentration of 0.2 ng/μl (1 ng in 5 μl volume). Each sample was made compatible for deep sequencing using the Nextera XT DNA sample preparation kit (Illumina, USA). Specifically, each sample or genome was enzymatically fragmented and tagged with short oligonucleotide adapters, followed by fifteen cycles of PCR for template indexing. Individual segments with undetectable concentrations by Qubit dsDNA were tagmented and indexed separately to maximize recovery of complete genomes.

Samples were purified using two consecutive AMPure bead cleanups (0.5× and 0.7×) and were quantified once more using Qubit dsDNA high-sensitivity kit (Invitrogen, USA). If quantifiable at this stage, independent gene segments were pooled into their corresponding genome pools. The average sample fragment length and purity were determined using Agilent hsDNA kit and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). After passing quality control measures for loading the sequencing machine, genomes were pooled into six groups of approximately thirty samples, which were sequenced on independent sequencing runs. Libraries of thirty genomes were pooled in equimolar ratios to a final concentration of 4 nM, and 5 μl of each 4 nM pool was denatured in 5 μl of freshly diluted 0.2 N NaOH for 5 min. Denatured pooled libraries were diluted to a final concentration of 16 pM, apart from the first library which was diluted to 12 pM, with a PhiX-derived control library accounting for 1 per cent of total DNA loaded onto the flow cell. A total of 600 μl of diluted, denatured library was loaded onto a 600-cycle v3 reagent cartridge (Illumina, USA). Average quality metrics were recorded, reads were demultiplexed, and FASTQ files were generated on Illumina's BaseSpace platform (Teiling 2016).

Sequence data analysis—quality filtering and variant calling

FASTQ files were processed using custom bioinformatic pipelines, available on GitHub https://github.com/tcflab/Sniffles2. Briefly, read ends were trimmed to achieve an average read quality score of Q30 and a minimum read length of 100 bases using Trimmomatic (Bolger, Lohse, and Usadel 2014). Paired-end reads were merged and mapped to a reference sequence using Bowtie2 (Langmead and Salzberg 2012). GD/3 and rGD/3 samples were mapped to the consensus sequence of the A/Guangdong/17SF006/2016 human isolate (GISAID isolate ID: EPI_ISL_249309) (Watanabe et al. 2013). Anhui/1 samples were mapped to the consensus sequence of the A/Anhui/1/2013 human isolate (GISAID isolate ID: EPI_ISL_138739) (Watanabe et al. 2013). H1N1 samples were mapped to A/California/04/2009 reference sequence (GISAID isolate ID: EPI_ISL_29618). To ensure even coverage and reduce resequencing bias, alignment files were randomly subsampled to 200,000 reads per genome using seqtk if total coverage exceeded this value (Li 2012).

The average genome sequence depth was 40,787 (±18,563) reads per genome (Supplementary Fig. S10). iSNVs were called with Varscan (Koboldt et al. 2009) using a frequency threshold of 1 per cent, a minimum coverage of 100 reads, and a base quality threshold of Q30 or higher. Variants were called independently for technical replicates, and only iSNVs called in both replicates, 'intersection iSNVs', were used for additional analyses (Robasky, Lewis, and Church 2014). If an iSNV was only found in one replicate, it was discarded. iSNV frequency is reported as the average frequency found across both replicates. iSNVs are annotated to determine the impact of each variant on the amino acid sequence. iSNVs were annotated in ten open reading frames: PB2, PB1, PA, HA, NP, NA, M1, matrix protein 2 (M2), NS1, and NEP; though for

some analyses M1 and M2 are jointly represented as MPs, an NS1 and NEP are jointly represented as NS.

Sequence data analysis—diversity statistics

nt diversity was calculated using π summary statistics. π quantifies the average number of pairwise differences per nt site among a set of sequences and was calculated using SNPGenie (Nelson, Moncla, and Hughes 2015; SNPGenie 2021). SNPGenie adapts the Nei and Gojobori method of estimating nt diversity (π) and its synonymous (π_S) and non-synonymous (π_N) partitions from next-generation sequencing data (Nei and Gojobori 1986). As most random non-synonymous mutations are likely to be disadvantageous, we expect that $\pi_N = \pi_S$ suggests neutrality and that allele frequencies are determined primarily by genetic drift. $\pi_N < \pi_S$ indicates that purifying selection is acting to remove new deleterious mutations, and $\pi_N > \pi_S$ indicates that diversifying selection is favoring new mutations and may indicate positive selection is acting to preserve multiple amino acid changes (Hughes 1999). We used paired *t*-tests to evaluate the hypothesis that $\pi_N = \pi_S$ within gene segments as well as the hypothesis that $\pi_N = \pi_S$ across samples. Code is available to replicate these analyses in the GitHub repository accompanying this manuscript (Braun 2023).

Sequence data analysis—estimating transmission bottleneck size

The beta-binomial model, explained in detail by Leonard et al. (2017), was used to infer effective transmission bottleneck sizes (Nb) for each transmission pair. Sobel-Leonard defines N_b as the number of virions that successfully establish lineages persisting to the first sampling time point in the recipient. This model does not account for virions that transiently replicate in the recipient and rapidly die out. In this model, the probability of iSNV transmission is determined by iSNV frequency in the donor at the time of sampling. The model incorporates sampling noise arising from a finite number of reads and therefore accounts for the possibility of false-negative variants that are not called in recipient animals due to conservative variant-calling thresholds (≥1 per cent in both technical replicates). Code for estimating transmission bottleneck sizes using the beta-binomial approach has been adapted from the original scripts, available here: https://github.com/koellelab/betabinomial_bottleneck (R version). H1N1 and H7N9 mean transmission bottleneck sizes were estimated using the approach outlined in a reanalysis study for SARS-CoV-2 deep sequencing data authored by Martin and Koelle (2021). Scripts have been added to our GitHub repository.

Sequence data analysis—enumerating iSNVs occurrences in surveillance samples

H7N9 phylogenies obtained from Nextstrain (Hadfield et al. 2018) in a JSON format were parsed using an open-source, custom Python script adapted from Moncla et al. (2020) to extract non-synonymous amino acid substitutions from the sequences present on a maximum likelihood phylogenetic tree. These phylogenies are generated with GISAID sequences (filtered by specific criteria) and displayed for each of the main gene segments: PB2, PB1, PA, HA, NP, NA, MP, and NS. We extracted a list of mutations from these trees and associated each mutation with the corresponding host of origin (avian host or human host). We found the intersection between iSNVs detected in our GD/3 and Anhui/1 datasets and the mutations parsed from the phylogenetic trees and counted the number of occurrences each iSNV was found in avian sequences, human sequences, or both. We tested whether occurrences of our iSNVs were enriched in human versus avian datasets using Fisher's exact test. A Bonferroni correction test was implemented to denote statistical significance with an alpha of 0.05. For readability, the iSNVs represented in Fig. 6 were filtered to remove those with less than two occurrences in human and/or avian hosts. The complete visualization of the iSNVs and their occurrences is displayed in Supplementary Fig. S8. Code to replicate these analyses is available in the GitHub repository accompanying this manuscript (Braun 2023). All figures were generated using R (ggplot2) or Python (Matplotlib) with packages including plotly, seaborn, numpy, and scipy, and were edited using Adobe Illustrator for clarity and readability.

Data availability

Primary data generated and analyzed in this study have been deposited in the Sequence Read Archive (SRA) under BioProject ID: PRJNA758865. Individual SRA identifiers can also be found in our GitHub repository.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

We would like to acknowledge and thank Dr. Kelsey Florek for her work to develop the Sniffles bioinformatic pipeline for identification of mutations in influenza populations. We thank research groups from all over the world for surveilling and sharing their viral sequence data. We also thank the teams at GISAID for maintaining a repository for influenza sequence data and to the Nextstrain team for providing open-source toolkits for analysis and visualization of these data.

Funding

This project was funded in part by National Institute of Allergy and Infectious Disease (NIAID) R01 AI125392—Mechanisms of influenza transmission bottlenecks: impact on viral evolution awarded to T.C.F. K.M.B. was supported by NIAID F30 (F30AI145182-02). L.A.H. was supported by the National Human Genome Research nstitute (NHGRI T32 (HG002760). The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Conflic t of interest: None declared.

References

Belser, J. A. et al. (2020) 'A Guide for the Use of the Ferret Model for Influenza Virus Infection', *The American Journal of Pathology*, 190: 11–24.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

Braun, K. (2023) *H7N9_evolution_in_mammals: To Accompany the Manuscript Entitled 'Stochastic Processes within Hosts Constrain Adaptation of Wildtype H7N9 Avian Influenza Viruses to Mammalian Hosts'* GitHub <https://github.com/katarinabraun/H7N9_evolution_in_mammals> accessed 12 January 2023.

Braun, K. M. et al. (2021) 'Acute SARS-CoV-2 Infections Harbor Limited Within-host Diversity and Transmit via Tight Transmission Bottlenecks', *PLoS Pathogens*, 17: e1009849.

Buhnerkempe, M. G. et al. (2015) 'Mapping Influenza Transmission in the Ferret Model to Transmission in Humans', *eLife*, 4: e07969.

Du, X. et al. (2017) 'Evolution-informed Forecasting of Seasonal Influenza A (H3N2)', *Science Translational Medicine*, 9: eaan5325.

Edwards, C. T. T. et al. (2006) 'Population Genetic Estimation of the Loss of Genetic Diversity during Horizontal Transmission of HIV-1', *BMC Evolutionary Biology*, 6: 28.

Ferguson, A. L. et al. (2013) 'Translating HIV Sequences into Quantitative Fitness Landscapes Predicts Viral Vulnerabilities for Rational Immunogen Design', *Immunity*, 38: 606–17.

Furuse, Y. et al. (2009) 'Evolution of the M Gene of the Influenza A Virus in Different Host Species: Large-scale Sequence Analysis', *Virology Journal*, 6: 67.

Gabriel, G. et al. (2005) 'The Viral Polymerase Mediates Adaptation of an Avian Influenza Virus to a Mammalian Host', *Proceedings of the National Academy of Sciences of the United States of America*, 102: 18590–5.

Galloway, S. E. et al. (2013) 'Influenza HA Subtypes Demonstrate Divergent Phenotypes for Cleavage Activation and pH of Fusion: Implications for Host Range and Adaptation', *PLoS Pathogens*, 9: e1003151.

Gao, R. et al. (2013) 'Human Infection with A Novel Avian-origin Influenza A (H7N9) Virus', *The New England Journal of Medicine*, 368: 1888–97.

Grubaugh, N. D. et al. (2019) 'An Amplicon-based Sequencing Framework for Accurately Measuring Intrahost Virus Diversity Using PrimalSeq and iVar', *Genome Biology*, 20: 8.

Hadfield, J. et al. (2018) 'Nextstrain: Real-time Tracking of Pathogen Evolution', *Bioinformatics*, 34: 4121–3.

Han, A. X. et al. (2021) 'Within-host Evolutionary Dynamics of Seasonal and Pandemic Human Influenza A Viruses in Young Children', *eLife*, 10: e68917.

Hughes, A. L. (1999) *Adaptive Evolution of Genes and Genomes*. New York: Oxford University Press.

Imai, M. et al. (2012) 'Experimental Adaptation of an Influenza H5 HA Confers Respiratory Droplet Transmission to a Reassortant H5 HA/H1N1 Virus in Ferrets', *Nature*, 486: 420–8.

——— et al. (2017) 'A Highly Pathogenic Avian H7N9 Influenza Virus Isolated from a Human Is Lethal in Some Ferrets Infected via Respiratory Droplets', *Cell Host & Microbe*, 22: 615–626.e8.

——— et al. (2020) 'Influenza A Variants with Reduced Susceptibility to Baloxavir Isolated from Japanese Patients Are Fit and Transmit through Respiratory Droplets', *Nature Microbiology*, 5: 27–33.

Itoh, Y. et al. (2009) '*In Vitro* and *in Vivo* Characterization of New Swine-origin H1N1 Influenza Viruses', *Nature*, 460: 1021–5.

Ke, C. et al. (2017) 'Human Infection with Highly Pathogenic Avian Influenza A(H7N9) Virus, China', *Emerging Infectious Diseases*, 23: 1332–40.

King, J. et al. (2020) 'Novel HPAIV H5N8 Reassortant (Clade 2.3.4.4b) Detected in Germany', *Viruses*, 12: 281.

Kiso, M. et al. (2017) 'Emergence of Oseltamivir-Resistant H7N9 Influenza Viruses in Immunosuppressed Cynomolgus Macaques', *The Journal of Infectious Diseases*, 216: 582–93.

Koboldt, D. C. et al. (2009) 'VarScan: Variant Detection in Massively Parallel Sequencing of Individual and Pooled Samples', *Bioinformatics*, 25: 2283–5.

Langmead, B., and Salzberg, S. L. (2012) 'Fast Gapped-read Alignment with Bowtie 2', *Nature Methods*, 9: 357–9.

Leonard, A. S. et al. (2017) 'Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus', *Journal of Virology*, 91: e00171-17.

Leyson, C. M. et al. (2022) 'Low Pathogenicity H7N3 Avian Influenza Viruses Have Higher Within-Host Genetic Diversity than a Closely Related High Pathogenicity H7N3 Virus in Infected Turkeys and Chickens', *Viruses*, 14: 554.

Li, H. (2012) 'seqtk Toolkit for Processing Sequences in FASTA/Q Formats', *GitHub*, 767: 69.

Linster, M. et al. (2014) 'Identification, Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 Virus', *Cell*, 157: 329–39.

Lipsitch, M. et al. (2016) 'Viral Factors in Influenza Pandemic Risk Assessment', *eLife*, 5: e1849.

Martin, M. A., and Koelle, K. (2021) 'Comment on "Genomic Epidemiology of Superspreading Events in Austria Reveals Mutational Dynamics and Transmission Properties of SARS-CoV-2"', *Science Translational Medicine*, 13: eabh1803.

Ma, M.-J., Yang, Y., and Fang, L.-Q. (2019) 'Highly Pathogenic Avian H7N9 Influenza Viruses: Recent Challenges', *Trends in Microbiology*, 27: 93–5.

McCrone, J. T., and Lauring, A. S. (2016) 'Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling', *Journal of Virology*, 90: 6884–95.

——— (2018) 'Genetic Bottlenecks in Intraspecies Virus Transmission', *Current Opinion in Virology*, 28: 20–5.

Moncla, L. H. et al. (2016) 'Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus', *Cell Host & Microbe*, 19: 169–80.

——— (2020) 'Quantifying Within-host Diversity of H5N1 Influenza Viruses in Humans and Poultry in Cambodia', *PLoS Pathogens*, 16: e1008191.

Morris, D. H. et al. (2018) 'Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology', *Trends in Microbiology*, 26: 102–18.

Neher, R. A., and Bedford, T. (2015) 'Nextflu: Real-time Tracking of Seasonal Influenza Virus Evolution in Humans', *Bioinformatics*, 31: 3546–8.

Nei, M., and Gojobori, T. (1986) 'Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions', *Molecular Biology and Evolution*, 3: 418–26.

Nelson, C. W., Moncla, L. H., and Hughes, A. L. (2015) 'SNPGenie: Estimating Evolutionary Parameters to Detect Natural Selection Using Pooled Next-generation Sequencing Data', *Bioinformatics (Oxford, England)*, 31: 3709–11.

Neumann, G., and Kawaoka, Y. (2015) 'Transmission of Influenza A Viruses', *Virology*, 479-480: 234–46.

——— (2019) 'Predicting the Next Influenza Pandemics', *The Journal of Infectious Diseases*, 219: S14–S20.

Ning, T. et al. (2019) 'Antigenic Drift of Influenza A (H7N9) Virus Hemagglutinin', *The Journal of Infectious Diseases*, 219: 19–25.

Qi, W. et al. (2018) 'Emergence and Adaptation of a Novel Highly Pathogenic H7N9 Influenza Virus in Birds and Humans from a 2013 Human-Infecting Low-Pathogenic Ancestor', *Journal of Virology*, 92: e00921–17.

Quan, C. et al. (2018) 'New Threats from H7N9 Influenza Virus: Spread and Evolution of High- and Low-pathogenicity Variants with High Genomic Diversity in Wave Five', *Journal of Virology*, 92: e00301–18.

Richard, M. et al. (2013) 'Limited Airborne Transmission of H7N9 Influenza A Virus between Ferrets', *Nature*, 501: 560–3.

Robasky, K., Lewis, N. E., and Church, G. M. (2014) 'The Role of Replicates for Error Mitigation in Next-generation Sequencing', *Nature Reviews Genetics*, 15: 56–62.

Russell, C. A. et al. (2012) 'The Potential for Respiratory Droplet-transmissible A/H5N1 Influenza Virus to Evolve in a Mammalian Host', *Science*, 336: 1541–7.

Shen, Y., and Lu, H. (2017) 'Global Concern regarding the Fifth Epidemic of Human Infection with Avian Influenza A (H7N9) Virus in China', *Bioscience Trends*, 11: 120–1.

Short, K. R. et al. (2015) 'One Health, Multiple Challenges: The Inter-species Transmission of Influenza A Virus', *One Health*, 1: 1–13.

SNPGenie. (2021), GitHub <https://github.com/chasewnelson/SNPGenie> accessed 12 April 2021.

Sobel Leonard, A. et al. (2016) 'Deep Sequencing of Influenza A Virus from A Human Challenge Study Reveals A Selective Bottleneck and Only Limited Intrahost Genetic Diversification', *Journal of Virology*, 90: 11247–58.

Steel, J. et al. (2009) 'Transmission of Influenza Virus in a Mammalian Host Is Increased by PB2 Amino Acids 627K or 627E/701N', *PLoS Pathogens*, 5: e1000252.

Su, S. et al. (2017) 'Epidemiology, Evolution, and Pathogenesis of H7N9 Influenza Viruses in Five Epidemic Waves since 2013 in China', *Trends in Microbiology*, 25: 713–28.

Subbarao, E. K., Kawaoka, Y., and Murphy, B. R. (1993) 'Rescue of an Influenza A Virus Wild-type PB2 Gene and A Mutant Derivative Bearing A Site-specific Temperature-sensitive and Attenuating Mutation', *Journal of Virology*, 67: 7223–8.

Sutton, T. C. (2018) 'The Pandemic Threat of Emerging H5 and H7 Avian Influenza Viruses', *Viruses*, 10: 461.

Taft, A. S. et al. (2015) 'Identification of Mammalian-adapting Mutations in the Polymerase Complex of an Avian H5N1 Influenza Virus', *Nature Communications*, 6: 7491.

Taubenberger, J. K., and Kash, J. C. (2010) 'Influenza Virus Evolution, Host Adaptation, and Pandemic Formation', *Cell Host & Microbe*, 7: 440–51.

Teiling, C. (2016) BaseSpace: Simplifying Metagenomic Analysis.

Wang, X. et al. (2017) 'Epidemiology of Avian Influenza A H7N9 Virus in Human Beings across Five Epidemics in Mainland China, 2013-17: An Epidemiological Study of Laboratory-confirmed Case Series', *The Lancet Infectious Diseases*, 17: 822–32.

Watanabe, T. et al. (2013) 'Characterization of H7N9 Influenza A Viruses Isolated from Humans', *Nature*, 501: 551–5.

——— et al. (2014) 'Circulating Avian Influenza Viruses Closely Related to the 1918 Virus Have Pandemic Potential', *Cell Host & Microbe*, 15: 692–705.

WHO. *China–WHO Joint Mission on Human Infection with Avian Influenza A(H7N9) Virus* (18–24 April 2013) <https://www.who.int/publications/m/item/china-who-joint-mission-on-human-infection-with-avian-influenza-a(h7n9)-virus-18-24-april-2013> accessed 12 April 2022.

Wilker, P. R. et al. (2013) 'Selection on Haemagglutinin Imposes a Bottleneck during Mammalian Transmission of Reassortant H5N1 Influenza Viruses', *Nature Communications*, 4: 2636.

Wong, J. et al. (2019) 'Improving Immunological Insights into the Ferret Model of Human Viral Infectious Disease', *Influenza and Other Respiratory Viruses*, 13: 535–46.

Wu, Y. et al. (2021) 'Accelerated Evolution of H7N9 Subtype Influenza Virus under Vaccination Pressure', *Virologica Sinica*, 36: 1124–32.

Xiong, X. et al. (2013) 'Receptor Binding by an H7N9 Influenza Virus from Humans', *Nature*, 499: 496–9.

Xu, Y. et al. (2019) 'Avian-to-Human Receptor-Binding Adaptation of Avian H7N9 Influenza Virus Hemagglutinin', *Cell Reports*, 29: 2217–2228.e5.

Yu, D. et al. (2019) 'The Re-emergence of Highly Pathogenic Avian Influenza H7N9 Viruses in Humans in Mainland China, 2019', *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 24: 1900273.

Zaraket, H. et al. (2015) 'Mammalian Adaptation of Influenza A(H7N9) Virus Is Limited by a Narrow Genetic Bottleneck', *Nature Communications*, 6: 6553.

Zhang, Q. et al. (2013) 'H7N9 Influenza Viruses are Transmissible in Ferrets by Respiratory Droplet', *Science*, 341: 410–4.

Zhang, F. et al. (2017) 'Human Infections with Recently-emerging Highly Pathogenic H7N9 Avian Influenza Virus in China', *The Journal of Infection*, 75: 71–5.

Zhao, L., and Illingworth, C. J. R. (2019) 'Measurements of Intrahost Viral Diversity Require an Unbiased Diversity Metric', *Virus Evolution*, 5: vey041.

Zhou, L. et al. (2017) 'Sudden Increase in Human Infection with Avian Influenza A(H7N9) Virus in China, September-December 2016', *Western Pacific Surveillance and Response Journal: WPSAR*, 8: 6–14.

——— et al. (2018) 'Clusters of Human Infection and Human-to-Human Transmission of Avian Influenza A(H7N9) Virus, 2013-2017', *Emerging Infectious Diseases*, 24: 397–400.

Zhu, W. et al. (2017) 'Biological Characterisation of the Emerged Highly Pathogenic Avian Influenza (HPAI) A(H7N9) Viruses in Humans, in Mainland China, 2016 to 2017', *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 22: 30533.