Avian influenza A viruses reassort and diversify differently in mallards and mammals

Ketaki Ganti[1#], Anish Bagga[2#], Juliana DaSilva[3], Samuel S. Shepard[3], John R. Barnes[3], Susan Shriner[4], Katia Koelle[5,6*] and Anice C. Lowen[1,6*]

[1]Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, GA, USA

[2]Emory College of Arts and Sciences, Atlanta, GA, USA

[3]Influenza Division, Center for Disease Control and Prevention, Atlanta, GA, USA

[4]National Wildlife Research Center, USDA-APHIS, Fort Collins, CO, USA

[5]Department of Biology, Emory University, Atlanta, GA, USA

[6]Emory-UGA Center of Excellence for Influenza Research and Surveillance (CEIRS)

[#]These authors contributed equally

*Corresponding authors: anice.lowen@emory.edu, katia.koelle@emory.edu

Abstract

Reassortment among co-infecting influenza A viruses (IAVs) is an important source of viral diversity and can facilitate expansion into novel host species. Indeed, reassortment played a key role in the evolution of the last three pandemic IAVs. Observed patterns of reassortment within a coinfected host are likely to be shaped by several factors, including viral load, the extent of viral mixing within the host and the stringency of selection. These factors in turn are expected to vary among the diverse host species that IAV infects. To investigate host differences in IAV reassortment, here we examined reassortment of two distinct avian IAV within their natural host (mallards) and a mammalian model system (guinea pigs). Animals were co-inoculated with A/wildbird/California/187718-36/2008 (H3N8) and A/mallard/Colorado/P66F1-5/2008 (H4N6) viruses. Longitudinal samples were collected from the cloaca of mallards or the nasal tract of guinea pigs and viral genetic exchange was monitored by genotyping clonal isolates from these samples. Relative to those in guinea pigs, viral populations in mallards showed higher frequencies of reassortant genotypes and were characterized by higher genotype richness and diversity. In line with these observations, analysis of pairwise segment combinations revealed lower linkage disequilibrium in mallards as compared to guinea pigs. No clear longitudinal patterns in richness, diversity or linkage disequilibrium were present in either host. Our results reveal mallards to be a highly permissive host for IAV reassortment and suggest that reduced viral mixing limits avian IAV reassortment in a mammalian host.

Introduction

Influenza A viruses (IAVs) infect a broad range of host species. Many diverse lineages circulate in waterfowl (Anseriformes) and shorebirds (Charadriformes), with 16 hemagglutinin and 9 neuraminidase subtypes represented. This diverse viral gene pool is the ancestral source of IAV lineages circulating in poultry, swine, humans and other mammalian hosts (1, 2). Although host barriers to infection limit the range of hosts within which a given IAV lineage circulates, spillovers occur occasionally and can seed novel lineages (3, 4). When a novel IAV lineage is established in humans, the result is a pandemic of major public health consequence (5, 6).

The segmented nature of the IAV genome allows facile genetic exchange between viruses that co-infect the same host: through reassortment of intact gene segments, mixed infections frequently give rise to viral genotypes that differ from both parental strains (7). The potential for viral diversification through this mechanism is high. If co-infecting parental viruses differ in all eight gene segments, reassortment can yield 256 distinct viral genotypes from a single co-infected cell. However, the contribution of reassortment to IAV diversity can be limited by two major forms of constraint. The first is a lack of opportunity for reassortment between distinct strains. For reassortment to occur, viruses must infect the same host and the same cell within that host. These prerequisites will be met routinely only if viral spread is well-mixed and high density at the population and within-host levels, respectively. The second form of constraint is negative selection. Reassortment may be deleterious because it can break epistatic interactions among gene segments and the proteins they encode. As a result, progeny viruses with chimeric genotypes can be less fit than both parental strains (8, 9).

Despite these constraints, reassortment has repeatedly been implicated in the evolution of novel IAV lineages and is strongly associated with IAV host switching (10). Reassortment involving human seasonal strains and IAV adapted to avian and/or swine hosts led to the 1957, 1968, and most recently, 2009 influenza pandemics. The establishment in poultry of H7N9 and H5N1 subtype viruses that are highly pathogenic to humans followed from reassortment between

enzootic poultry viruses and strains introduced transiently from wild birds (11-14). More recently, reassortment of poultry H5N1 viruses with IAV in wild birds led to the spread of highly pathogenic H5N8 and H5N2 subtype viruses to North America (15). Thus, reassortment can have major consequences for IAV evolution and host range expansion.

In the work described here, we used experimental IAV coinfection to evaluate the efficiency of reassortment in mallards, a major natural host of IAV. The viral strains used for coinfection, of H3N8 and H4N6 subtypes, are typical of viruses isolated from mallards and representative of lineages co-circulating in North American waterfowl. Both the hemagglutinin (HA) and neuraminidase (NA) subtypes represented are furthermore associated with frequent mixed infection within this ecological niche (16). We compared reassortment observed in this natural host-virus pairing to that seen with the same viral strains in a guinea pig model. In this way, the impact of host species on the extent of within-host genetic diversity achieved through reassortment was evaluated. Our results revealed abundant reassortment in mallards, giving rise to high genotype richness and diversity and low linkage disequilibrium between segments. In guinea pigs, reassortment rates were lower, fewer unique genotypes were detected and relatively high prevalence of parental genotypes resulted in low diversity. These findings indicate that mallards are highly permissive for IAV reassortment, whereas mammalian hosts may present a less permissive environment for IAV reassortment. Additional analyses suggest that lower levels of reassortment in guinea pigs stem from reduced viral mixing in this host, rather than pervasive purifying selection limiting the frequency of reassortants.

Results

Robust infection in co-inoculated mallards and guinea pigs

The genomes of the two viral strains used for co-infection, influenza A/wildbird/California/187718-36/2008 (H3N8) and A/mallard/Colorado/P66F1-5/2008 (H4N6) viruses, were sequenced in full and found to be genetically distinct in all eight segments (**Table 1**). The proteins encoded by these

gene segments exhibited between 44.2% and 100% identity in their amino acid composition. The viral genome sequences were used to design eight segment-specific primer sets suitable for differentiation of the H3N8 and H4N6 gene segments by high resolution melt analysis.

To evaluate the patterns of IAV reassortment *in vivo* and its reliance on host species, groups of eight mallards and eight guinea pigs were co-inoculated the H3N8 and H4N6 viruses and shed virus was sampled daily. Viral growth in each species was robust, but differed in the kinetics observed (**Figure 1**). The average peak titer reached in mallards and guinea pigs was comparable, at 4.5 and 5.2 $\log_{10}$PFU/mL, respectively, although guinea pigs showed a much broader range of values. Kinetics were more rapid in mallards, with peak viral load seen at the earliest time point (1 day post-inoculation) and clearance observed at 4 or 5 days post-inoculation. In guinea pigs, titers generally peaked at 2 days post-inoculation and shedding ceased at 5 days post-inoculation at the earliest.

Genotypic diversity was higher in mallards than in guinea pigs

To quantify reassortment following coinfection, clonal isolates were derived from mallard cloacal swabs collected on days 1, 2 and 3 and from guinea pig nasal washes collected on days 2, 3 and 4. These time points were chosen because all individuals in the respective groups were shedding above the limit of detection on each of these days. Twenty-one clonal isolates were derived per sample. Genotyping was then performed for all eight gene segments of these isolates to evaluate whether they were of H3N8-origin or H4N6-origin. For a small minority of isolates, one or more gene segments gave an ambiguous result in our genotyping assay and the full isolate was therefore excluded from the analysis (giving a final sample size of 18-21 isolates per sample). In all individuals of both species, and at all time points examined, reassortment was detected (**Supplementary Figure 1**).

To characterize the outcomes of IAV coinfection in each host, the genotypes from the clonal isolates were quantitatively analyzed. Specifically, for each sample collected from a given

individual on a given day, the frequency of unique genotypes, frequency of parental genotypes, genotype richness, and genotype diversity were determined (**Figure 2**). In mallards, the frequency of a given genotype was variable across time, reflecting infrequent carry-over of specific reassortant genotypes within a given individual (**Figure 2A** and **Supplementary Figure 2**). Parental genotypes were detected in each of the mallards on at least one sampling day, with frequencies on a given day varying between 0 and 0.6 (**Figure 2B**). Genotype richness, or the number of distinct genotypes detected in a sample, was typically high in mallards, with most samples showing richness values between 12 and the maximum value of 21 (**Figure 2C**). Diversity was measured using the Shannon-Weiner index, which considers both richness and evenness in the frequency with which genotypes are detected. Diversity in mallards tracked closely with richness and was also high, with all mallards showing diversity values greater than 2.3 on at least one sampling day (**Figure 2D**). Similar to mallards, genotype frequencies in guinea pigs were variable across time, again reflecting infrequent carry-over of specific reassortant genotypes (**Figure 2E** and **Supplementary Figure 2**). In striking contrast to mallards, however, parental genotypes were often predominant in guinea pigs, although the frequencies observed ranged widely, from 0 to 0.95 (**Figure 2F**). Genotype richness in guinea pigs was generally lower than that seen in mallards, with most samples comprising fewer than 12 unique genotypes (**Figure 2G**). Correspondingly, genotype diversity also appeared lower in guinea pigs than in mallards, with only one of the guinea pigs showing diversity above 2.3 and most values falling in the range of 0.5 to 2.0 (**Figure 2H**). Lower levels of genotype diversity in guinea pigs than in mallards were found to be statistically significant ($p = 1.3 \times 10^{-7}$; ANOVA).

In the combined data set from both species, we noted an inverse correlation between diversity and parental genotype frequency (**Figure 3A**). To test the hypothesis that low diversity is driven by high parental frequencies, we therefore calculated viral evenness in each sample with and without the inclusion of parental genotypes. Evenness, rather than diversity, was analyzed in this way because the reduction in sample size brought about by excluding parental genotypes

157 confounds effects on diversity. Since this exclusion typically has little impact on richness (reducing
158 the number of unique genotypes by 0-2), any impact on diversity of removing parental genotypes
159 would occur mainly through changes in evenness. Upon exclusion of parental genotypes,
160 evenness in guinea pigs increased, while that in mallards was less affected (**Figure 3B, C**).
161 Nevertheless, evenness was significantly lower in guinea pigs compared to mallards whether
162 parental genotypes were included or not ($p=1.47x10^{-5}$ and 0.0166, respectively, unpaired t-test).
163 Thus, high parental genotype frequency contributes, but does not fully account for reduced viral
164 diversity in guinea pigs compared to mallards.
165
166 **Parental genotypes drive higher linkage disequilibrium in guinea pigs than in mallards**
167 To assess the extent to which reassortment was shaped by genetic linkages among segments,
168 and whether this effect varied with host species, linkage disequilibrium (*D*) was evaluated for each
169 pair of viral gene segments. This analysis examines the extent to which segments assorted non-
170 randomly. For the pairing of PB2 and PB1, for example, four genotypes are possible:
171 $PB2_{H3N8}PB1_{H3N8}$, $PB2_{H3N8}PB1_{H4N6}$, $PB2_{H4N6}PB1_{H3N8}$, and $PB2_{H4N6}PB1_{H4N6}$. If these four genotypes
172 are present in a given viral population at the expected frequencies based on the prevalence of
173 the four gene segments, *D* will equal zero, indicative of random assortment. Positive associations,
174 in which specific genotypes are over-represented, will yield $D > 0$, while negative associations will
175 give $D < 0$. In both mallards and guinea pigs, pairwise *D* values tended to be positive (**Figure 4**).
176 Comparison between host species for each pairwise combination revealed that *D* was typically
177 lower in mallards than guinea pigs (Mann Whitney U test, $p<0.05$ for 16 of the 28 segment
178 pairings). While this result could suggest that reassortment in guinea pigs was more strongly
179 shaped by positive segment associations than that in mallards, high linkage disequilibrium could
180 alternatively be driven by a lack of opportunity for reassortment, leading to frequent occurrence
181 of parental segment pairings.
182

Genotype patterns are consistent with random sampling of gene segments

To gauge whether the lower levels of reassortment in guinea pigs compared to ducks is due to a lack of reassortment opportunity versus more stringent selection in guinea pigs, we assessed the types of reassortants that were observed in guinea pigs and in mallards. Reasoning that less chimeric genotypes, in which only one or two segments are derived from a second parent strain, would be less likely to suffer fitness defects due to epistasis, we tested the hypothesis that such reassortants were over-represented in our data sets relative to others, and that they would be particularly overrepresented in guinea pigs relative to mallards. The extent to which a viral genotype was chimeric was quantified by the number of gene segments stemming from the parental virus that contributed fewer gene segments to the genotype. Viral genotypes were thus categorized as having between 0 and 4 minority segments. We further quantified the total number of unique viral genotypes that belonged to each of these categories. All possible genotypes in a given category were typically not observed, with the exceptions of the 0 minority segments category (which includes only the two parental genotypes) and the 1 minority segment category in mallards (**Figure 5A**). To test whether the distribution of observed genotypes across the five categories may be due to selection, we calculated the proportion of possible genotypes that were detected in each category and compared these results to a simulated dataset (**Figure 5B**). The simulated data were generated by sampling segments at random from the set of reassortant genotypes observed in each animal. Genotypes with one minority segment tended to be more prevalent in observed compared to simulated datasets, but this trend was not statistically significant owing to wide confidence intervals in the simulated data sets (**Figure 5B**). Conversely, more highly chimeric genotypes occurred less often in observed compared to simulated datasets. This trend was statistically significant for genotypes with three or four minority gene segments in guinea pigs and for those with three minority segments in mallards. Thus, observed genotype patterns likely resulted from a combination of limited availability of certain segments in a given host and purifying selection acting more strongly on reassortant genotypes in the three and four

minority gene segment categories. Evidence for selection was, however, stronger in guinea pigs than mallards. In comparing mallards and guinea pigs, it is also clear that higher proportions of genotypes in each reassortant category were detected in mallards. This effect was recapitulated in the simulated dataset, however, indicating that this species-specific effect is a simple manifestation of increased genotype richness in mallards.

Discussion

Through experimental coinfection of mallards and guinea pigs with distinct IAVs typical of those that circulate widely in North American ducks, we examined the efficiency of IAV genetic exchange in both a natural host (mallard) and a model mammalian host (guinea pig). Robust reassortment was apparent in mallards, giving rise to highly diverse viral populations within coinfected individuals. In contrast, reassortant genotypes were less common in guinea pigs. In both hosts, temporal trends in the frequency of specific or overall reassortant genotypes were not apparent. Little recurrence of reassortant genotypes across individuals was apparent, suggesting that stochastic processes were more potent than selection in shaping within-host viral populations.

Low diversity in guinea pigs may result from relatively little opportunity for reassortment, negative selection of less fit variants, or a combination of both. Our analyses suggest that a combination of these two factors is at play in guinea pigs. If opportunity for reassortment is low, owing to low rates of coinfection, this would not only favor maintenance of parental genotypes, but would also lead to out-growth of reassortants that form and then are propagated without further coinfection. Low evenness among reassortant genotypes detected in guinea pigs is therefore consistent with limited opportunity for reassortment. Furthermore, stochastic outgrowth of reassortants would lead to over-representation of differing reassortant genotypes in unrelated individuals. In line with this expectation, recurrent detection of reassortant genotypes across different individuals was not a major feature of the dataset. One exception to this generalization

was seen: a reassortant combining the NP segment of the H4N6 parent with seven segments from the H3N8 parent and was detected repeatedly in six of eight guinea pigs. More broadly, however, the reassortant genotypes detected in guinea pigs suggest that their propagation was largely a stochastic process and not the result of positive selection. Evidence of negative selection was, however, seen with the observation that highly chimeric genotypes were under represented in guinea pigs. Together, the data support roles for both stochastic and selective forces shaping reassortment patterns in guinea pigs.

In contrast to guinea pigs, the viral genotypic diversity observed in coinfected in mallards is consistent with abundant opportunity for reassortment and minimal within host selection. Several lines of evidence point to this conclusion. While parental genotypes were routinely detected, they did not predominate. Both high richness and high evenness contributed to diverse within-host viral populations. Pairwise linkage disequilibrium was low in mallards, suggesting that heterologous gene segment combinations were not strongly deleterious to viral fitness. Similarly, the reassortant gene constellations detected were largely consistent with random sampling of gene segments. The data clearly reveal mallards to be highly permissive hosts for influenza viral genetic exchange.

The evolutionary implications of robust genetic exchange within the mallard host are dependent on viral dynamics at larger scales of biological organization. For example, the extent of viral mixing within host populations is a major determinant of the frequency with which coinfections occur. Field studies indicate that this potential constraint is not strong in mallards. Within flyways, the migratory patterns of mallards and other waterfowl allow diverse IAV to be brought together at common migratory stopover points and over-wintering sites (16-19). One study revealed that, among 167 wild bird samples analyzed, 26% showed evidence of more than one subtype, indicative of mixed infection (20). Recent examination of a much broader dataset revealed mixed infection within North American waterfowl to be more common in winter, with

multiple HA/NA subtypes detected in 3% of all isolates and 13% of winter isolates (16). Thus, within flyways, IAV mixing at the population level occurs readily in wild birds.

A second feature of viral dynamics that determines the potential for reassortants to impact larger scale evolutionary trends is the tightness of the bottleneck governing transmission between hosts. While IAV transmission bottlenecks are thought to be stringent in mammalian hosts (21, 22), formal analysis of the bottleneck in waterfowl has not been undertaken to date. Since IAV transmission in these birds occurs through a fecal-oral route, rather than a respiratory route as seen in mammals, the existence of a substantially wider bottleneck is plausible.

A third, and critical, factor in considering the potential for abundant within-host reassortment to impact population-level viral evolution is the distribution of fitness effects that results from reassortment. The fitness effects of both inter- and intra-subtype reassortment of human IAV is generally deleterious (8, 9). These negative fitness effects are readily explained by the disruption of inter-segment interactions: as constellations of viral genes co-evolve, the genetic under-pinnings of epistatic interactions become lineage-specific and, as a result, reassortment breaks this epistasis (23). Interestingly, the consequences of genetic exchange may be very different for IAV circulating in wild waterfowl. In this ecological niche, IAV reassortment is frequently detected at the population level (20, 24-28). In particular, incongruence among the phylogenies of different IAV gene segments indicates that their evolution does not occur in concert; rather, gene constellations are routinely disrupted through reassortment. In the absence of stable gene constellations, there may be little opportunity for divergent epistatic interactions to evolve. Instead, the gene segments of avian IAV appear to exist as functionally-equivalent and readily interchangeable units (20). In stark contrast to IAV circulating in humans, reassortants that form in wild birds are not subjected to strong negative selection and circulate widely.

The observation herein of robust IAV reassortment within individual mallards strongly suggests that free-mixing and prevalent coinfection within mallard hosts contributes to the abundant IAV reassortment observed at the population level. Over multiple scales of biological

286 organization, IAV dynamics in mallards (and likely other members of the Anseriformes) appear to
287 overcome both spatial and selective constraints to support production, onward transmission and
288 broad circulation of reassortant viruses.
289
290 **Materials and Methods**
291 **Virus isolation and propagation**
292 The viruses used in this study were originally isolated from wild bird environmental samples as
293 part of US surveillance for avian IAV in wild birds conducted between 2006-2009 (29). Fresh wild
294 bird feces were collected using Dacron swabs and stored in BA-1 viral transport medium. Samples
295 were tested by RT PCR and positive samples were inoculated into the allantoic cavity of 11-day
296 old embryonated hen eggs at 37°C for virus isolation (30). Virus subtypes were confirmed by
297 hemagglutination inhibition and neuraminidase inhibition tests at the National Veterinary Services
298 Laboratory (Ames, Iowa).
299 The H3N8 virus stock, A/wildbird/CA/187718-36/08, was propagated in hen eggs from the
300 original environmental sample as described above. The H4N6 virus, A/mallard/CO/P66F1-5/08,
301 was also propagated in hen eggs from the original sample but was then passaged through a
302 mallard. A fecal sample from that mallard was propagated in hen eggs. Allantoic fluid was
303 harvested and pooled for each virus, and aliquots were stored at -80°C prior to use. Egg Infectious
304 Dose 50 (EID_{50}) titers were determined using the Reed & Muench (1938) method.
305
306 **Whole genome sequencing and primer design**
307 The complete viral genomes for both the H3N8 and H4N6 viruses were sequenced at the CDC
308 Influenza Division using an Illumina sequencing platform. Briefly, viral RNA was extracted from
309 virus stocks using the Qiagen Viral RNA Mini Kit (Qiagen) and was used as a template for multi-
310 segment RT PCR as described previously (31). The sequence reads were assembled using the

IRMA v 0.9.1 pipeline (32). The complete sequences were submitted to GenBank (MT982372-79 (H4N6); MT982380-87 (H3N8)).

Primers to distinguish H3N8 and H4N6 gene segments were designed based on these sequences (**Table 2**). The six non-HA, non-NA segments were typed using high resolution melt analysis. Primers anneal to conserved sequences of each segment and direct amplification of an approximately 100 bp region, which contains one to five nucleotide differences between the two viruses. These nucleotide differences confer distinct melting properties on the cDNA. Because the HA and NA segments are highly divergent, high resolution melt analysis is not feasible. For these segments, virus-specific primers were generated to enable genotyping by standard RT qPCR.

Cells

Madin Darby canine kidney (MDCK) cells from Dr. Daniel Perez at University of Georgia were used for plaque assay. The cells were maintained in minimal essential medium (Gibco) supplemented with 10% fetal bovine serum (FBS; Atlanta Biologicals), penicillin (100 IU), and streptomycin (100 µg ml^{-1}; PS; Corning). All cells were cultured at 37°C and 5% CO_2 in a humidified incubator. Cells were tested monthly for mycoplasma contamination while in use.

Animal models

All animal experiments were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The studies were conducted under animal biosafety level 2 containment and were approved by the Institutional Animal Care and Use Committee (IACUC) of the US Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center (NWRC), Fort Collins, CO, USA (approval NWRC QA-1621) for the mallards (*Anas platyrhynchos*) and the IACUC of Emory University (protocol PROTO201700595) for the guinea pigs (*Cavia porcellus*) studies. The

337 animals were humanely euthanized following guidelines approved by the American Veterinary
338 Medical Association.
339
340 **Experimental infection of mallards**
341 Mallards were purchased from Field Trial Gamebirds in Fort Collins, CO and were approximately
342 5-6 months old at the time of testing. Prior to inoculation, all birds were confirmed to be negative
343 for influenza A virus antibodies by bELISA (33) and for viral RNA by RT-PCR. During testing, all
344 birds (n=8) were housed at the NWRC in an indoor aviary equipped with ten 2.1 m x 2.1 m x 2.4
345 m pens separated by 1.27 cm x 7.62 cm PVC-coated wire mesh. Birds were housed in groups of
346 three or four per pen and could roam freely within the pen. Each pen included a food bowl, a water
347 bowl, a small bowl of grit, and a water-filled 375 L oval stock tank for swimming and preening.
348 The H3N8 and H4N6 inoculums were prepared to 10^5 EID_{50}/mL by diluting stocks with negative
349 allantoic fluid. Birds were inoculated oro-choanally using a P1000 pipet with 1 ml H3N8 and 1 ml
350 H4N6 virus preparations. Cloacal swabs were collected daily from all birds and placed in 1 ml BA-
351 1 viral transport medium and stored at -80°C until laboratory testing.
352
353 **Experimental infection of guinea pigs:** Female Hartley strain guinea pigs weighing 250–350 g
354 were obtained from Charles River Laboratories and housed by Emory University Department of
355 Animal Resources. Before intranasal inoculation and nasal washing, the guinea pigs were
356 anaesthetized with 30 mg kg^{-1} ketamine and 4 mg kg^{-1} xylazine by intramuscular injection. The
357 guinea pigs (n=8) were inoculated intranasally with the H3N8 and H4N6 virus mixture in PBS at
358 a dose of 10^5 pfu of each virus in a total inoculation volume of 300 µl. Animals were singly housed
359 in filter top covered rat cages with food and water provided ad libitum. Daily nasal washes were
360 collected in 1 ml PBS and stored at -80°C until laboratory testing.
361
362

363 **Determination of viral loads**

364 Viral loads in mallard cloacal swab samples and guinea pig nasal wash samples were determined
365 by standard plaque assay on MDCK cells. Briefly, one aliquot was thawed, mixed well and
366 subjected to 10-fold serial dilution in PBS. Dilutions 10^{-1} to 10^{-6} were then used to inoculate
367 confluent MDCK cells in 6-well plates. Following a one-hour attachment period, inoculum was
368 removed, monolayers were washed with PBS and serum-free culture medium containing 0.6%
369 Oxoid agar was overlaid onto the cells. Cultures were incubated for two days at 37ºC, at which
370 time plaques were counted and titer determined by taking into account the initial dilution of the
371 sample. The limit of detection of the plaque assay is 50 PFU/mL, equivalent to one plaque from
372 the 10^{-1} dilution.

373

374 **Quantification of reassortment**

375 Reassortment levels were evaluated by genotyping 21 virus isolates per sample from mallard
376 cloacal swabs and guinea pig nasal washes, as described previously (7). Based on positivity for
377 infectious IAV in all eight animals, samples from days 1, 2 and 3 were chosen for evaluation of
378 reassortment in mallards, while samples from days 2, 3 and 4 were chosen for guinea pigs.
379 Briefly, plaque assays were performed on MDCK cells in 10 cm diameter dishes to isolate virus
380 clones. Serological pipettes (1 ml) were used to collect agar plugs from well-separated plaques
381 into 160 µl PBS. Using a ZR-96 viral RNA kit (Zymo), RNA was extracted from the agar plugs and
382 eluted in 40 µl nuclease-free water (Invitrogen). Reverse transcription was performed with a
383 universal IAV primer (31) using Maxima reverse transcriptase (RT; Thermofisher) according to
384 the manufacturer's protocol. The resulting cDNA was diluted 1:4 in nuclease-free water, each
385 cDNA was combined with segment-specific primers (**Table 2**) and Precision Melt Supermix (Bio-
386 Rad) and analyzed by qPCR using a CFX384 Touch real-time PCR detection system (Bio-Rad).
387 The qPCR was followed by high-resolution melt analysis to differentiate the H3N8 and H4N6
388 amplicons of the six non-HA, non-NA segments. Precision Melt Analysis software (Bio-Rad) was

used to determine the parental virus origin of each gene segment based on the melting properties of the cDNA fragments and comparison to H3N8 and H4N6 virus controls. For the HA and NA gene segments, Ct values were used to determine the origin of each segment. Ct values >= 32 were considered negative. Each plaque was assigned a genotype based on the combination of H3N8 and H4N6 genome segments, with two variants of each of the eight segments allowing for 256 potential genotypes. In a minority of samples, the origin of one or more gene segments could not be determined and the genotype from that clonal isolate was removed from the analysis.

Software

All calculations, plotting, and machine learning was done through Python 3 (34). Packages used included matplotlib (35), NumPy (36), pandas (37), SciPy (38) and statsmodels (39).

Genotype Frequencies, Richness, Diversity, and Evenness

Here, a virus genotype is defined as a unique combination of the eight IAV segments, where each segment is derived from either the H3N8 or H4N6 parental virus. In total, there are 2^8 (=256) possible unique genotypes. Genotype frequencies were calculated for each sample by dividing the number of appearances of each genotype by the total number of clonal isolates available for that sample.

Genotype richness (*S*) is given by the number of unique virus genotypes in a sample. Genotype diversity was quantified using the Shannon-Wiener Index:

$$H' = -\sum_{i=1}^{S} (p_i \ln p_i)$$

where *S* is genotype richnessand p_i is the frequency of genotype *i* in the sample (40). With a range between 0 and 1, genotype evenness is defined as the extent to which genotypes are evenly present in a sample, with 1 corresponding to maximum evenness (41). Evenness is given by:

$$\frac{H'}{H'_{max}}$$

where maximum genotype diversity H'_{max} evaluates to $\ln S$ in a sample with genotype richness *S*. Evenness values were calculated both with and without the parental genotypes included in the samples. In the evenness calculations that excluded parental genotypes, genotype frequencies, richness *S*, diversity H' and maximum diversity H'_{max} values were all recalculated using the subset of clonal isolates that were not parental genotypes.

Linkage Disequilibrium

Linkage disequilibrium *D* is a population genetic measure that quantifies the extent of non-random association of alleles at two or more loci in a population (42). Here we use *D* to characterize non-random associations of gene segments. *D* was calculated for all 8-choose-2 (= 28) possible two gene segment combinations using the formula:

$$D = p\{i, H3N8\}, \{j, H3N8\} - p\{i, H3N8\}p\{j, H3N8\}$$

where $p\{i, H3N8\}$ is the proportion of clonal isolates that have gene segment *i* derived from parental virus H3N8 $p\{j, H3N8\}$ is the proportion of clonal isolates that havegene segment *j* derived from parental virus H3N8, and $p\{i, H3N8\}, \{j, H3N8\}$ is the proportion of the clonal isolates have both gene segments *i* and *j* derived from parental virus H3N8. Calculations of *D* using H4N6 as the reference virus instead of H3N8 yield equivalent results.

Minority Segments, Unique Genotypes, and Simulation

To identify potential purifying selection in mallards and guinea pigs, an analysis of the genotypes in terms of the number of minority segments was conducted. We define a minority segment as a segment derived from the virus that contributes fewer gene segments to a genotype. Each observed genotype can thus be classified as falling into a category of containing between 0 and

4 minority gene segments. Figure 5A shows the number of virus genotypes that fall into each of these minority gene segment categories.

To determine if selection acted against more chimeric genotypes, the proportion of genotypes in each minority gene segment category was first calculated by dividing the observed number of genotypes in that category by the number of theoretically possible genotypes in that category (Figure 5B). To determine whether these calculated proportions deviated from what would be expected by chance, we generated simulated data for both mallards and guinea pigs that reflected chance expectations. Specifically, during a round of simulation, all clonal isolates containing parental genotypes were first excluded from each of the samples. The remaining clonal isolates were then randomly permuted by gene segment to generate a simulated data set. If the simulated data set contained any parental genotypes, the data set would be tossed and another one generated. 1000 simulated data sets were generated in total. Each of these resultant simulated data sets were analyzed similarly to the observed data sets.

Statistical Measures

All confidence intervals were calculated via the proportion_confint method from the statsmodels package (39). Unpaired t-tests were conducted via the ttest_rel and the ttest_ind methods respectively from the SciPy package (38). Mann-Whitney U tests were conducted with the mannwhitneyu method and ANOVA tests were conducted with the f_oneway method, both of which are found in the SciPy package (38).

Acknowledgements

This work was funded by the National Institute of Allergy and Infectious Disease through R01 AI127799 and the Centers of Excellence for Influenza Research and Surveillance (CEIRS) contract no. HHSN272201400004C.

Figure Legends

Figure 1: Efficient viral replication was observed in mallards and guinea pigs. Infectious viral titers in A) cloacal swab samples collected from mallards and B) nasal lavage samples collected from guinea pigs are plotted against day post-inoculation. Each colored line represents an individual animal. Horizontal dashed line shows the limit of detection.

Figure 2: Viral populations in mallards showed lower frequencies of parental genotypes, higher genotype richness and higher diversity than those in guinea pigs. Results from mallards are shown in panels A-D and from guinea pigs in panels E-H. A, E) Stacked plot showing frequencies of unique genotypes detected in one representative animal over time. The lowermost two sections, in blue and orange, represent the H3N8 and H4N6 parental genotypes, respectively. B, F) Frequency of parental genotypes over time. The total frequency of H3N8 and H4N6 parental genotypes is plotted. C, G) Genotype richness over time. D, H) Diversity over time, as measured by the Shannon-Weiner index.

Figure 3: Lower levels of viral diversity in guinea pigs is only partially accounted for by high frequency of parental genotypes. A) Scatter plot of Shannon-Weiner diversity against parental genotype frequency. B) Histograms showing evenness of virus samples in mallards. C) Histograms showing evenness of virus samples in mallards, following removal of parental genotypes. D) Histograms showing evenness of virus samples in guinea pigs. E) Histograms showing evenness of virus samples in guinea pigs, following removal of parental genotypes.

Figure 4: Pairwise linkage disequilibrium was typically lower in mallards than in guinea pigs. Linkage disequilibrium for each pairwise combination of IAV gene segments is shown. Each

489 data point represents a mallard or guinea pig, with samples from all three time points aggregated
490 for a given individual. P values indicate results of a Mann Whitney U test.
491
492 **Figure 5: Genotype patterns detected in mallards and guinea pigs are consistent with**
493 **random sampling of gene segments.** A) Number of theoretically possible unique genotypes in
494 each category (brown) and the number detected in mallards (olive) and guinea pigs (grey). Data
495 from all time points and all individuals of a given host species are combined. Number of unique
496 genotypes is displayed above each bar. B) Proportion of unique genotypes in each category that
497 were detected. Observed results in mallards (solid, olive line) and guinea pigs (solid, grey line)
498 are compared to simulated data sets generated by randomly sampling segments from reassortant
499 viruses detected in each species (dashed lines). Error bars on the simulated data show the 95%
500 range of calculated proportions from the 1000 simulated data sets.
501
502 **Supplementary Figure Legends**
503 **Supplementary Figure 1: Viral genotypes detected in mallards and guinea pigs.**
504 Viral genotypes detected in mallards (A) and guinea pigs (B) are depicted with H4N6 parental
505 origin in blue and H3N8 parental origin in orange. Each table corresponds to a sample collected
506 from an individual animal on a given day, with columns representing the eight viral gene segments
507 ordered 1-8 from left to right and rows showing clonal isolates. Tables are sorted such that any
508 parental genotypes appear at the top (H4N6) or bottom (H3N8).
509
510 **Supplementary Figure 2: Viral genotype frequencies detected in mallards and guinea pigs.**
511 Stacked plots showing frequencies of unique genotypes detected in each animal over time. The
512 lowermost two sections, in blue and orange, represent the H3N8 and H4N6 parental genotypes,
513 respectively.
514

References

1. Webster RG, Hinshaw VS, Bean WJ, Jr., Turner B, Shortridge KF. 1977. Influenza viruses from avian and porcine sources and their possible role in the origin of human pandemic strains. Dev Biol Stand 39:461-8.
2. Wright PF, Neumann G, Kawaoka Y. 2006. Orthomyxoviruses, p 1691-1740. *In* Knipe DMH, P. M. (ed), Fields Virology, vol 1. Lippincott-Raven, Philidelphia.
3. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. Microbiol Rev 56:152-79.
4. Webster RG, Shortridge KF, Kawaoka Y. 1997. Influenza: interspecies transmission and emergence of new pandemics. FEMS Immunol Med Microbiol 18:275-9.
5. Taubenberger JK, Morens DM. 2006. 1918 Influenza: the mother of all pandemics. Emerg Infect Dis 12:15-22.
6. Viboud C, Miller M, Olson D, Osterholm M, Simonsen L. 2010. 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 doi:k/-/-/35hpbywfdwl4n/8 [pii]:RRN1153.
7. Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. 2013. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. PLoS Pathog 9:e1003421.
8. Phipps KL, Marshall N, Tao H, Danzy S, Onuoha N, Steel J, Lowen AC. 2017. Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny. J Virol 91.
9. Villa M, Lassig M. 2017. Fitness cost of reassortment in human influenza. PLoS Pathog 13:e1006685.
10. Ma EJ, Hill NJ, Zabilansky J, Yuan K, Runstadler JA. 2016. Reticulate evolution is favored in influenza niche switching. Proc Natl Acad Sci U S A 113:5335-9.
11. Lam TT, Wang J, Shen Y, Zhou B, Duan L, Cheung CL, Ma C, Lycett SJ, Leung CY, Chen X, Li L, Hong W, Chai Y, Zhou L, Liang H, Ou Z, Liu Y, Farooqui A, Kelvin DJ, Poon LL, Smith DK, Pybus OG, Leung GM, Shu Y, Webster RG, Webby RJ, Peiris JS, Rambaut A, Zhu H, Guan Y. 2013. The genesis and source of the H7N9 influenza viruses causing human infections in China. Nature 502:241-4.
12. Wu A, Su C, Wang D, Peng Y, Liu M, Hua S, Li T, Gao GF, Tang H, Chen J, Liu X, Shu Y, Peng D, Jiang T. 2013. Sequential Reassortments Underlie Diverse Influenza H7N9 Genotypes in China. Cell Host Microbe 4:446-52.
13. Li KS, Guan Y, Wang J, Smith GJ, Xu KM, Duan L, Rahardjo AP, Puthavathana P, Buranathai C, Nguyen TD, Estoepangestie AT, Chaisingh A, Auewarakul P, Long HT, Hanh NT, Webby RJ, Poon LL, Chen H, Shortridge KF, Yuen KY, Webster RG, Peiris JS. 2004. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature 430:209-13.
14. Bui C, Bethmont A, Chughtai AA, Gardner L, Sarkar S, Hassan S, Seale H, MacIntyre CR. 2016. A Systematic Review of the Comparative Epidemiology of Avian and Human Influenza A H5N1 and H7N9 - Lessons and Unanswered Questions. Transbound Emerg Dis 63:602-620.
15. Hall JS, Dusek RJ, Spackman E. 2015. Rapidly Expanding Range of Highly Pathogenic Avian Influenza Viruses. Emerg Infect Dis 21:1251-2.
16. Diskin ER, Friedman K, Krauss S, Nolting JM, Poulson RL, Slemons RD, Stallknecht DE, Webster RG, Bowman AS. 2020. Subtype Diversity of Influenza A Virus in North American Waterfowl: a Multidecade Study. J Virol 94.
17. Lam TT, Ip HS, Ghedin E, Wentworth DE, Halpin RA, Stockwell TB, Spiro DJ, Dusek RJ, Bortner JB, Hoskins J, Bales BD, Yparraguirre DR, Holmes EC. 2012. Migratory flyway

565 and geographical distance are barriers to the gene flow of influenza virus among North
566 American birds. Ecol Lett 15:24-33.
567 18. Krauss S, Walker D, Pryor SP, Niles L, Chenghong L, Hinshaw VS, Webster RG. 2004.
568 Influenza A viruses of migrating wild aquatic birds in North America. Vector Borne Zoonotic
569 Dis 4:177-89.
570 19. Hill NJ, Ma EJ, Meixell BW, Lindberg MS, Boyce WM, Runstadler JA. 2016. Transmission
571 of influenza reflects seasonality of wild birds across the annual cycle. Ecol Lett 19:915-25.
572 20. Dugan VG, Chen R, Spiro DJ, Sengamalay N, Zaborsky J, Ghedin E, Nolting J, Swayne
573 DE, Runstadler JA, Happ GM, Senne DA, Wang R, Slemons RD, Holmes EC,
574 Taubenberger JK. 2008. The evolutionary genetics and emergence of avian influenza
575 viruses in wild birds. PLoS Pathog 4:e1000076.
576 21. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic
577 processes constrain the within and between host evolution of influenza virus. Elife 7.
578 22. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, Garcia-Sastre A,
579 tenOever BR. 2014. Influenza A virus transmission bottlenecks are defined by infection
580 route and recipient host. Cell Host Microbe 16:691-700.
581 23. Lowen AC. 2017. Constraints, Drivers, and Implications of Influenza A Virus
582 Reassortment. Annu Rev Virol 4:105-121.
583 24. Wille M, Tolf C, Avril A, Latorre-Margalef N, Wallerstrom S, Olsen B, Waldenstrom J. 2013.
584 Frequency and patterns of reassortment in natural influenza A virus infection in a reservoir
585 host. Virology 443:150-60.
586 25. Deng G, Tan D, Shi J, Cui P, Jiang Y, Liu L, Tian G, Kawaoka Y, Li C, Chen H. 2013.
587 Complex reassortment of multiple subtypes of avian influenza viruses in domestic ducks
588 at the dongting lake region of china. J Virol 87:9452-62.
589 26. Hatchette TF, Walker D, Johnson C, Baker A, Pryor SP, Webster RG. 2004. Influenza A
590 viruses in feral Canadian ducks: extensive reassortment in nature. J Gen Virol 85:2327-
591 2337.
592 27. Macken CA, Webby RJ, Bruno WJ. 2006. Genotype turnover by reassortment of
593 replication complex genes from avian influenza A virus. J Gen Virol 87:2803-2815.
594 28. Venkatesh D, Poen MJ, Bestebroer TM, Scheuer RD, Vuong O, Chkhaidze M,
595 Machablishvili A, Mamuchadze J, Ninua L, Fedorova NB, Halpin RA, Lin X, Ransier A,
596 Stockwell TB, Wentworth DE, Kriti D, Dutta J, van Bakel H, Puranik A, Slomka MJ, Essen
597 S, Brown IH, Fouchier RAM, Lewis NS. 2018. Avian Influenza Viruses in Wild Birds: Virus
598 Evolution in a Multihost Ecosystem. J Virol 92.
599 29. Deliberto TJ, Swafford SR, Nolte DL, Pedersen K, Lutman MW, Schmit BB, Baroch JA,
600 Kohler DJ, Franklin A. 2009. Surveillance for highly pathogenic avian influenza in wild
601 birds in the USA. Integr Zool 4:426-39.
602 30. Szretter KJ, Balish AL, Katz JM. 2006. Influenza: propagation, quantification, and storage.
603 Curr Protoc Microbiol Chapter 15:Unit 15G 1.
604 31. Zhou B, Wentworth DE. 2012. Influenza A virus molecular virology techniques. Methods
605 Mol Biol 865:175-92.
606 32. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. 2016. Viral deep
607 sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler.
608 BMC Genomics 17:708.
609 33. Sullivan HJ, Blitvich BJ, VanDalen K, Bentler KT, Franklin AB, Root JJ. 2009. Evaluation
610 of an epitope-blocking enzyme-linked immunosorbent assay for the detection of
611 antibodies to influenza A virus in domestic and wild avian and mammalian species. J Virol
612 Methods 161:141-6.
613 34. Van Rossum G, Drake FL. 2009. Python 3 Reference Manual. CreateSpace, Scotts
614 Valley, CA.

35. Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9:90-95.
36. Oliphant TE. 2006. A guide to NumPy vol 1. Trelgol Publishing, USA.
37. McKinney W. 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference 445:51-56.
38. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy C. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261-272.
39. Seabold S, Perktold J. 2010. statsmodels: Econometric and statistical modeling with python. 9th Python in Science Conference doi:10.25080/Majora-92bf1922-011.
40. Spellerberg IF, Fedor PJ. 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Global Ecology and Biogeography 12:177-179.
41. Grunwald NJ, Goodwin SB, Milgroom MG, Fry WE. 2003. Analysis of genotypic diversity data for populations of microorganisms. Phytopathology 93:738-46.
42. Mueller JC. 2004. Linkage disequilibrium for different scales and applications. Brief Bioinform 5:355-64.

Table 1: Nucleotide and amino acid differences between A/wildbird/California/187718-36/2008 (H3N8) and A/mallard/Colorado/P66F1-5/2008 (H4N6) viruses.

Segment	% Nucleotide identity	Protein	% Amino acid identity
PB2	91.3	PB2	98.8
PB1	94.3	PB1	99.2
		PB1-F2	85.6
PA	87.7	PA	98.3
		PA-X	98.4
HA	65.0	HA	68.7
NP	93.7	NP	100
NA	55.6	NA	44.2
M	98.2	M1	100
		M2	99.0
NS	94.6	NS1	98.7
		NEP	100

Table 2: Primers used for viral genotyping

Primer name	Sequence
PB2 F	CATGCTGGGAGCAAATGTACA
PB2 R	TCTTACTATGTTCCTGGCAGC
PB1 F	GGAACAGGATACACCATGGA
PB1 R	GTTGTCCATTTCCCCTTTTCTG
PA F	GGGATTCCTTTCGTCAGTC
PA R	GCCTGCGCATGGTTC
NP F	AGGGCACTTGTGCGTACT
NP R	CCTTTCACTGCTGCTCCA
M F	GCATCGGTCTCACAGACA
M R	CCTGCCATTTGCTCCATG
NS F	GAATCCGACGAGGCACT
NS R	TGGGCATGAGCATGAACC
H3 F	AAATGGAGGGAGTGGAGCTT
H3 R	TGAACTCCCCACACGTACAA
H4 F	CTGCCCAGGAATTAGTGGAA
H4 R	TCTGGCACATCAAATGGGTA
N8 F	TGAAAGACCGGAGCCCCTAT
N8 R	AGGGCCCGTTACTCCAATTG
N6 F	TTGGGAAATGGGGCAAGCA
N6 R	GCCTTCCCTTGTACCAGACC

Figure 1

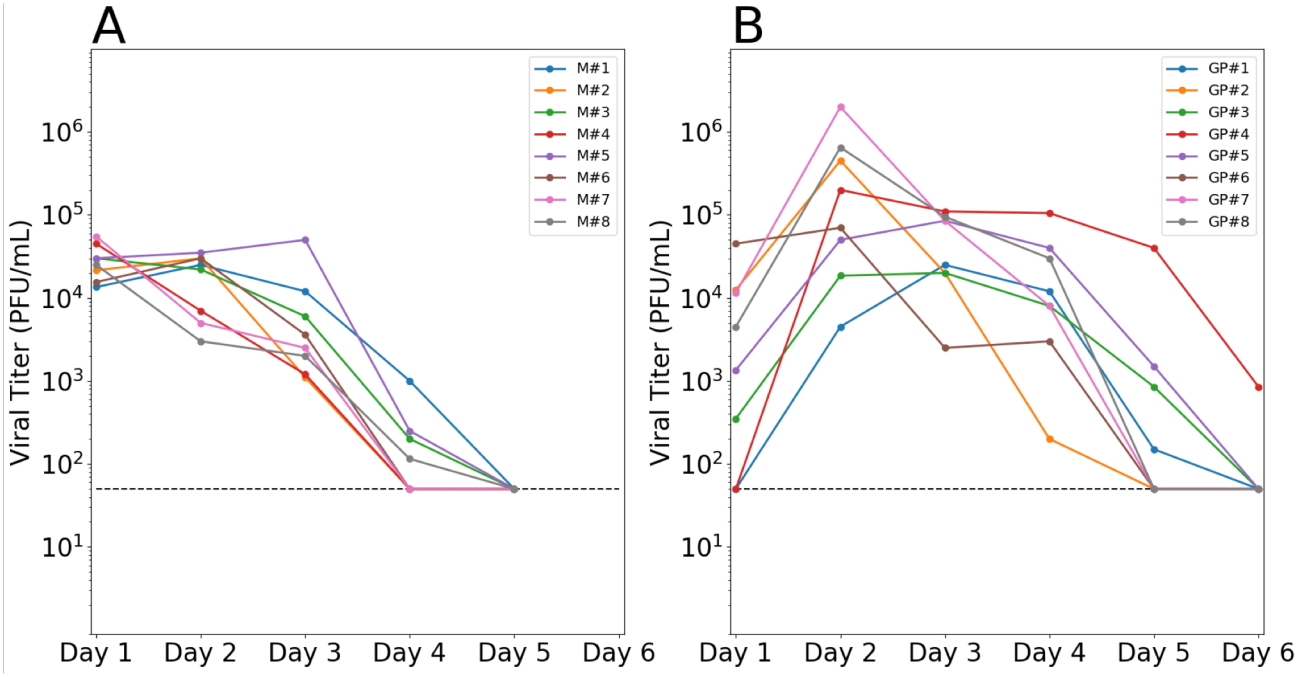

Figure 2

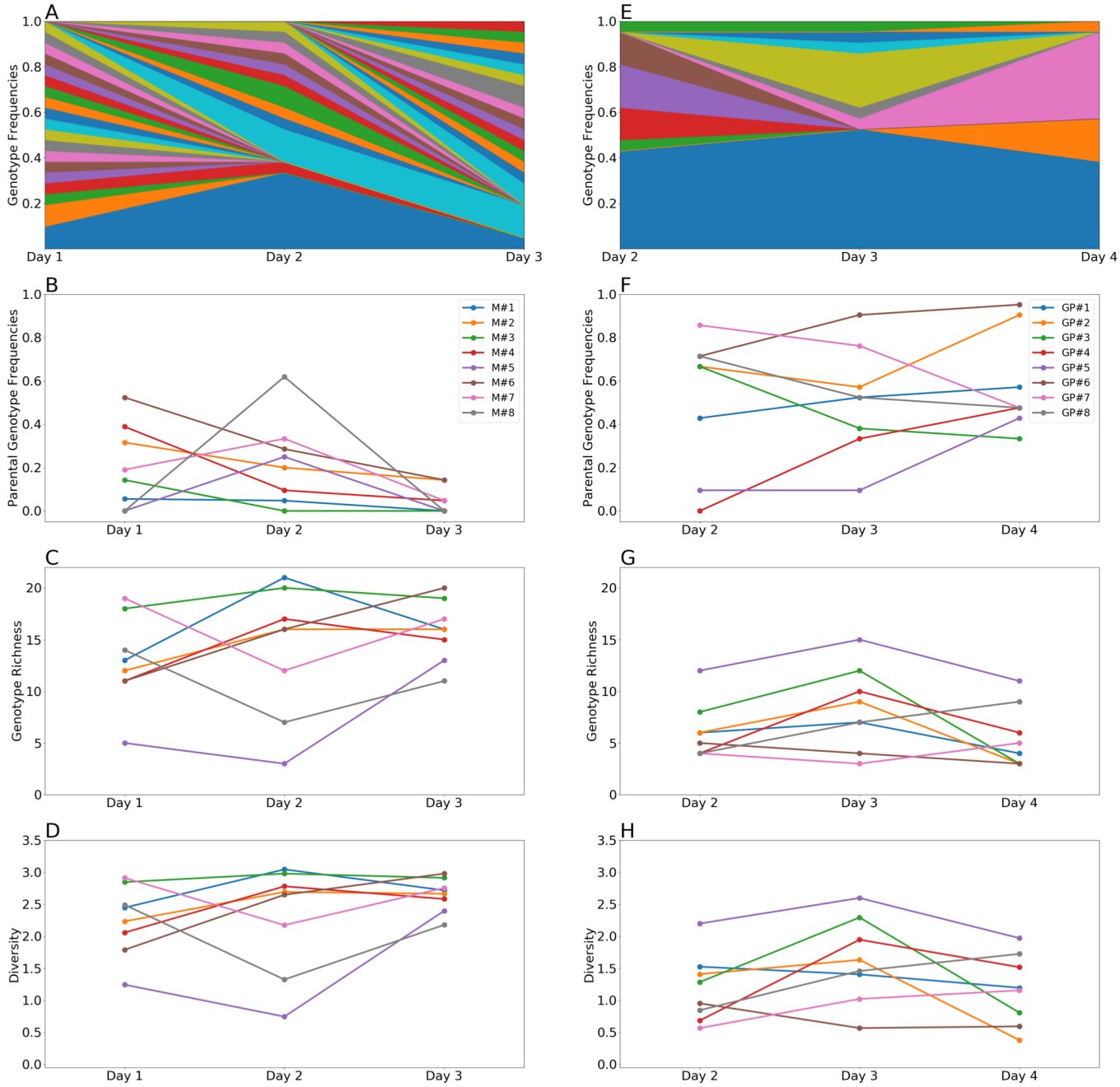

Figure 3

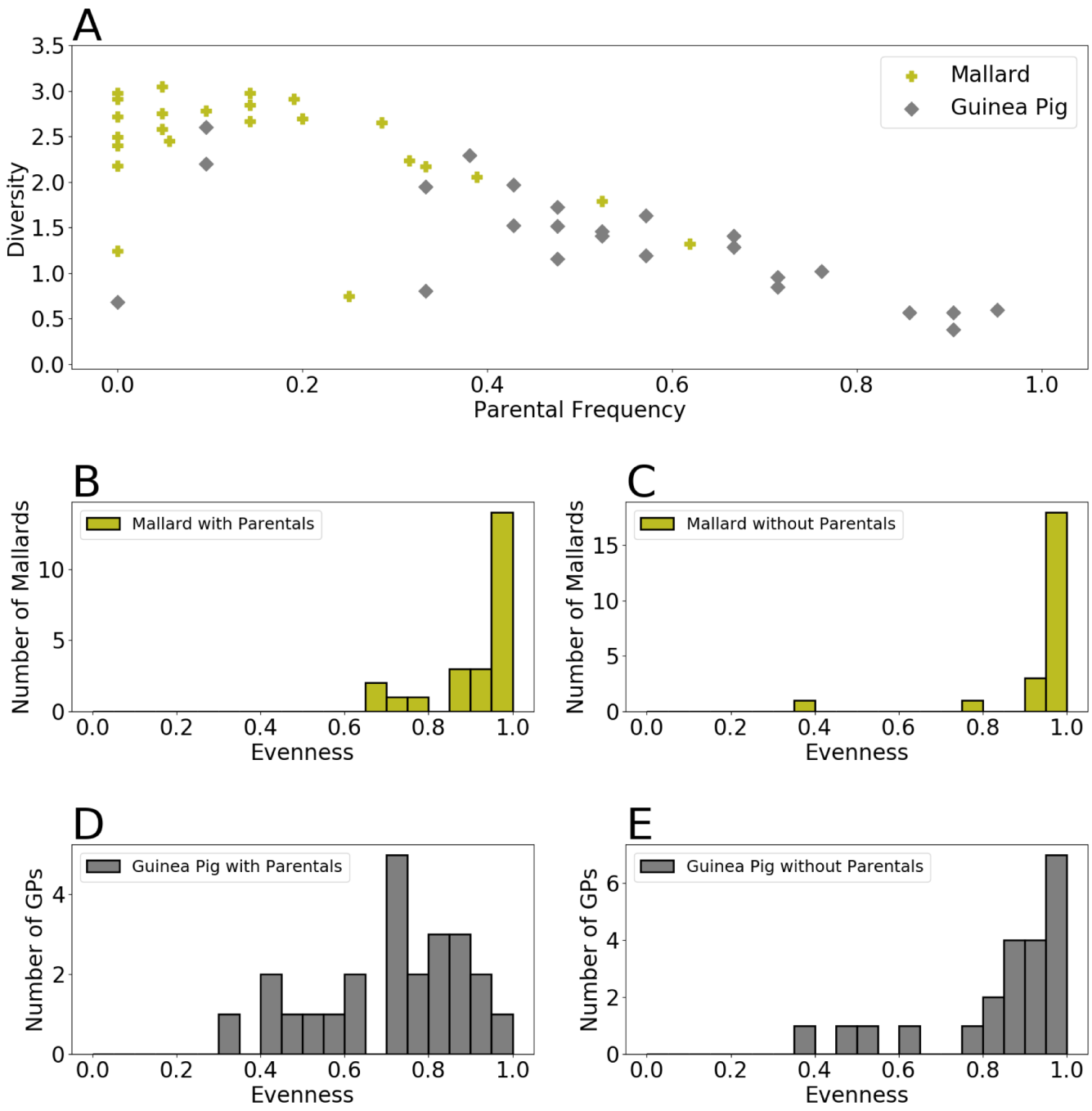

Figure 4

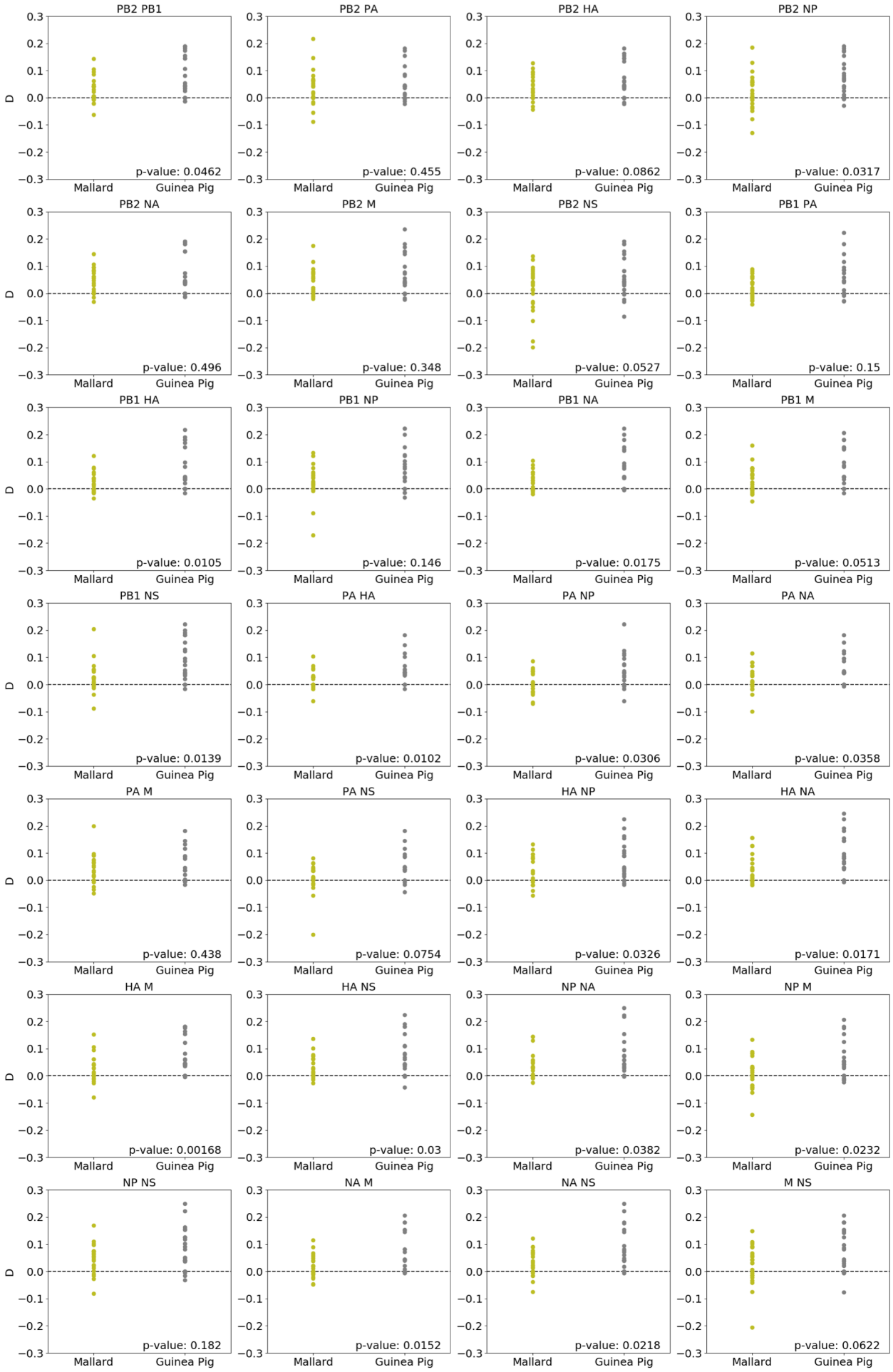

Figure 5