Viral expansion after transfer is a primary driver of influenza A virus transmission bottlenecks

Short Title: Influenza viral transmission bottlenecks and evolution

Katie E. Holmes[1], David VanInsberghe[1], Lucas M. Ferreri[1], Baptiste Elie[2], Ketaki Ganti[1], Chung-Young Lee[3], Anice C. Lowen[1,4*]

[1]Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, GA, USA
[2]MIVEGEC, CNRS, IRD, Université de Montpellier, Montpellier, France
[3]Department of Microbiology, School of Medicine, Kyungpook National University, Jung-gu, Daegu, Republic of Korea
[4]Emory Center of Excellence for Influenza Research and Response (CEIRR), Atlanta, GA, USA

*anice.lowen@emory.edu

ABSTRACT

For many viruses, narrow bottlenecks acting during transmission sharply reduce genetic diversity in a recipient host relative to the donor. Since genetic diversity represents adaptive potential, such losses of diversity are thought to limit the opportunity for viral populations to undergo antigenic change and other adaptive processes. Thus, a detailed picture of evolutionary dynamics during transmission is critical to understanding the forces driving viral evolution at an epidemiologic scale. To advance this understanding, we used a barcoded virus library and a guinea pig model of transmission to decipher where in the transmission process influenza A virus populations lose diversity. In inoculated guinea pigs, we show that a high level of viral barcode diversity is maintained. Within-host continuity in the barcodes detected across time furthermore indicates that stochastic effects are not pronounced within the inoculated hosts. Importantly, in both aerosol-exposed and direct contact animals, we observed many barcodes at the earliest time point(s) positive for infectious virus, indicating robust transfer of diversity through the environment. This high viral diversity is short-lived, however, with a sharp decline seen 1-2 days after initiation of infection. Although major losses of diversity at transmission are well described for influenza A virus, our data indicate that events that occur following viral transfer and during the earliest stages of natural infection have a central role in this process. This finding suggests that host factors, such as immune effectors, may have greater opportunity to impose selection during influenza A virus transmission than previously recognized.

INTRODUCTION

The high mutation rate characteristic of RNA viruses [1, 2], coupled with genetic recombination [3, 4] and/or reassortment [5] of segmented genomes, enables constant production of viral variants [6-8]. In turn, this variation provides the substrate for viral evolution [9-11]. For some

viruses, including influenza A virus and SARS-CoV-2, viral evolution and viral spread through host populations occur on a similar timescale, such that each shapes the other [12]. Understanding viral evolution is therefore crucial in our efforts to control outbreaks.

Selection and genetic drift are two major drivers of evolutionary change in viral populations [13]. Selection is a deterministic process whereby differences in fitness result in changes in variant frequencies over time. In contrast, genetic drift is a stochastic process whereby changes in variant frequencies result from chance. In general, when population sizes are large, selection predominates over drift and changes in variant frequencies result from differences in the variants' reproductive output. When population sizes are small, drift instead predominates and changes in variant frequencies result simply from random chance associated with reproductive output. Both selection and drift can reduce genetic diversity, but by different means. Under drift, the loss of a variant comes about as a chance event. Under selection, the loss of a variant comes about from the variant having lower fitness. As such, these processes have different impacts on a population's fitness trajectory: if driven by selection, the loss of diversity increases viral population-level fitness and as such results in viral adaptation. If driven by drift, the loss of viral diversity could (by chance) either increase or decrease population-level fitness. For influenza viruses, the dynamic interplay between selection and genetic drift is still poorly understood [13-15].

A reduction in viral population size during transmission has been documented for many viruses [16-20]. Termed a transmission bottleneck, this effect is characterized by the establishment of a new infection by few viruses derived from a large source population. The reduction in population size is typically accompanied by a loss of population diversity, known as a genetic bottleneck. Classically, bottlenecks are defined as random events and are therefore a source of genetic drift [21]. When a random bottleneck is active, even highly adaptive variants may go extinct. In the case of HIV, however, the sharp reduction in population size at transmission is a selective process favoring variants that use CCR5 co-receptors [22-24]. The term selective bottleneck is often used to describe this deterministic process.

As for other viruses, loss of diversity during influenza A virus transmission has been documented. In humans with naturally occurring infections, as few as one to two viral genomes from an infected individual established in an uninfected individual [13]. Studies using influenza A virus populations with inserted genetic barcodes similarly revealed narrow transmission bottlenecks in animal models [25, 26]. More relaxed bottlenecks have been reported in natural hosts aside from humans, suggesting roles for host biology and/or modes of transmission in shaping species-specific between-host dynamics [27, 28]. Finally, avian influenza A virus transmission in mammals has been shown to involve a tight bottleneck with reductions in viral diversity attributed to either stochastic or selective forces [29-31].

While it is known that diversity is lost during influenza A virus transmission, it is unclear at which point – in the donor host, the environment, or the recipient host – this takes place. The specific stage at which diversity is lost is likely to define the potential for selection to act during transmission. We therefore made an influenza A virus library with 4096 potential barcodes to

monitor the fate of many unique viral lineages within infected guinea pigs and through the course of transmission to contacts. Our data reveal that viral diversity remains high in inoculated animals throughout infection, and many viral genomes are transferred to animals exposed by aerosols or direct contact. However, a severe bottleneck occurs in exposed animals 1-2 days after the initiation of infection, such that few lineages are sustained in the context of population expansion. Thus, in our system, losses of diversity are primarily driven by events that occur during the expansion of infection, not during the process of donor-to-recipient transfer itself. Importantly, this phase of expansion could be a previously unrecognized opportunity for selection to operate.

RESULTS

Generation of a Barcoded Influenza A Virus Library

We designed a genetic barcode for influenza A/Panama/2007/99 (H3N2) virus (Pan/99) with the goals of producing a diverse viral population while avoiding attenuation of viral replication and fitness differences among barcode variants. To achieve high diversity, twelve nucleotide sites within a 50-nucleotide region of the NA segment were made bi-allelic. Since one of two bases was possible at each of 12 positions, 2^{12} = 4096 unique barcodes were introduced (**Figure 1A**). To avoid attenuation, we introduced polymorphisms within the native sequence of Pan/99 rather than inserting a foreign sequence. To avoid fitness differences among the barcode variants, the polymorphisms introduced were synonymous and chosen based on their natural occurrence in H3N2-subtype influenza A viruses circulating in humans from 1994 to 2004. We reasoned that naturally occurring variants detected in circulation would be likely to have minimal fitness effects. We named the gene segment modified in this way Pan/99 NA-BC.

To evaluate barcode diversity in the reverse genetics plasmid and the passage 1 (P1) viral stock carrying Pan/99 NA-BC, we subjected the barcode region to next-generation sequencing. No barcode was found to be dominant in the plasmid library or the virus stock, and no nucleotide was dominant at any site within the barcode (**Figure 1B, D**). An aberrant nucleotide was detected at the twelfth barcode site in the plasmid preparation but was not carried through to the virus stock. To quantify diversity, we applied the Shannon Diversity Index (*H*), which considers the richness (i.e., the number of species present) and the evenness (i.e., the abundance of the present species) in a community [32]. To calculate *H*, each unique barcode was taken to represent a species. The measured diversity in the plasmid and virus stocks was 8.07 and 7.90, respectively (**Figure 1C**), near to the theoretical maximum (H_{max}) for this system of 8.32 and revealing little loss of barcode diversity during viral recovery from cDNA. To determine whether the barcode in the NA segment altered fitness of the Pan/99 virus, multi-cycle replication was evaluated in MDCK cells. No significant differences between wild type and Pan/99 NA-BC viruses were detected (**Figure 1E**).

Modeling Transmission of Pan/99 NA-BC Virus in Guinea Pigs

To evaluate viral dynamics within and between hosts, we inoculated guinea pigs with 5×10^4 plaque-forming units (PFU) of Pan/99 NA-BC virus intranasally [33]. After 24 hours, we placed a naïve animal with each inoculated animal in either direct or aerosol contact (**Supplemental**

130 **Figure 1A**). Three independent experiments were performed, each including four direct contact
131 and four aerosol transmission pairs. Viral shedding, assessed by plaque assay of daily nasal
132 lavage samples, was similar for all inoculated animals (**Supplemental Figure 1B**). Transmission
133 efficiency varied from 75% to 100% for pairs in direct contact and from 50% to 100% for pairs in
134 aerosol contact. The daily nasal lavages collected from both inoculated and exposed animals
135 furnished samples with which to investigate viral population dynamics and the drivers of the
136 influenza A virus transmission bottleneck.
137
138 **Stochastic Effects are Not Pronounced in Inoculated Animals**
139 To determine how barcode diversity changed during infection in inoculated animals, we
140 evaluated the barcodes present in nasal lavage samples by next-generation sequencing. In all
141 directly inoculated animals, many barcodes were detected throughout the course of infection
142 (**Figure 2** and **Supplemental Figure 2**). Using the Shannon Diversity index (*H*), we found that
143 barcode diversity either remains consistent or declines slightly over time in inoculated animals
144 (**Figure 3** and **Supplemental Figure 3**). Changes in diversity are furthermore mirrored by
145 changes in both barcode richness and evenness, indicating that declines in diversity are driven
146 both by loss of barcode species from the within-host population and skewing of the relative
147 frequencies of individual barcodes that comprise the population (**Figure 3** and **Supplemental**
148 **Figure 3**).
149
150 **High Initial Diversity in Exposed Animals Precedes a Sharp Decline**
151 To determine how barcode diversity changed between donor and recipient animals, barcodes
152 present in nasal lavage samples of recipients were analyzed. Strikingly, at the earliest
153 timepoint(s) positive for infectious virus in recipients, many barcodes are present, signifying
154 transfer of an appreciably large and diverse viral population through the environment (**Figure 2**
155 and **Supplemental Figure 2**). This observation is true of both aerosol-exposed and direct
156 contact animals. Populations present early in exposed animals are characterized by high
157 barcode diversity, richness, and evenness (**Figure 3** and **Supplemental Figure 3**).
158
159 However, the observed high barcode diversity in exposed animals is short-lived, with a sharp
160 decline seen within 1-2 days of the initiation of infection (**Figure 2, Supplemental Figure 2, and**
161 **Supplemental Figure 3**). Both richness and evenness contribute to the precipitous drop in
162 diversity seen within exposed animals (**Figure 3A** and **Supplemental Figure 3**). The stark change
163 in population composition is unique to early times after transmission: pairwise comparisons
164 between successive time points show high dissimilarity of barcode composition between the
165 first and second positive days but low dissimilarity between successive days thereafter (**Figure**
166 **3B**). While the few viral lineages that penetrate the bottleneck persist through the remainder of
167 the acute infection, those that persist differ across animals (**Figure 2** and **Supplemental Figure**
168 **2**), excluding the possibility that some barcodes carry a fitness advantage. Notably, barcode
169 diversity in some exposed animals rebounds late in the infection.
170
171 To assess the potential for the observed population dynamics to be driven by selection acting
172 on *de novo* mutations outside the barcoded region, we performed whole-genome sequencing
173 of a subset of nasal lavage samples. Specifically, viral genomes collected on the first or second

day of positivity in exposed animals from experimental replicates 1 and 2 were sequenced in full. Among these samples, a single variant was found above a 1% frequency, which was a nonsynonymous change present at 4.9% in the NP segment of guinea pig 31 on day 5 post-inoculation. This variant resulted in a methionine to valine substitution at amino acid 238. Thus, reductions in viral diversity observed in exposed animals were not driven by selective sweeps of *de novo* variants.

As a means of validating results obtained from targeted sequencing of the barcoded region of the viral genome, we analyzed the whole-genome sequencing reads that spanned this region. This analysis relied on reads spanning the entirety of the barcode, which were relatively rare. Nevertheless, high-level observations could be confirmed: first-positive time points showed multiple barcodes present and subsequent time points revealed the same predominant barcodes in both sequencing datasets (**Supplemental Figure 4A**). Furthermore, results of targeted barcode sequencing were validated by sequencing a subset of samples multiple times to ensure reproducibility (**Supplemental Figure 4B, C**).

To evaluate whether barcodes detected on the first day of positivity in recipient animals were carried by viable viruses, we performed plaque assays with nasal lavage samples from the first and second positive days from a subset of eight recipient animals. The barcodes present in plaque isolates were then identified by sequencing. In three of the eight animals, plaques from the first day of positivity yielded a diverse set of barcodes, indicating that many independent viral lineages present early in these three animals were represented in the infectious viral population (**Figure 4**). In the remaining five animals, the plaque isolates from the first positive day were low diversity and largely matched those derived from the second positive day (**Figure 4**).

To test whether the differing population dynamics between donor and recipient animals relate to viral dose, we evaluated dynamics of barcode composition in animals inoculated with escalating doses of virus (**Supplemental Figure 5A**). Three animals were inoculated intranasally with 1×10^2 PFU, 3.3×10^2 PFU, 1×10^3 PFU, 3.3×10^3 PFU, 1×10^4 PFU, or 5×10^4 PFU of Pan/99 NA-BC virus. We saw that, at each inoculation dose, measured values for diversity, richness, and evenness are below what would be expected if all inoculated barcodes were recovered in nasal lavage samples (**Supplemental Figure 5B**). Thus, a genetic bottleneck acts within animals infected through inoculation. This bottleneck is of comparable stringency across the dose range tested. A bottleneck acting during the establishment of infection was not, however, observed in inoculated animals as it was transmission recipients.

DISCUSSION

We used a barcoded influenza A virus system to generate a high-resolution view of changes in the composition and diversity of viral populations as they expand, contract, transmit, and establish infection in a new host. Our data suggest that viral evolutionary dynamics through a transmission event comprise two distinct stages: inter-host transfer and intra-host establishment. The bottleneck acting during the first stage can be loose, allowing many

217 genotypes to pass to the new host. Conversely, the bottleneck acting at the second stage is
218 tight, such that few genotypes dominate the established within-host population. The
219 evolutionary implications of this two-stage process are potentially great. The existence of a
220 loose bottleneck between hosts could enable selection to act efficiently on the transferred viral
221 population *prior to* a stochastic contraction of diversity imposed by the process of viral
222 establishment. Indeed our data offer a potential solution to a long-standing conundrum in
223 evolutionary virology, that of disparate evolutionary dynamics at the individual and global
224 population scales [34].
225
226 The dynamics we observe here for influenza A virus transmission are analogous to those seen in
227 mice inoculated orally with a barcoded coxsackievirus, wherein high initial barcode diversity
228 was followed by a sharp decline during enteric infection [35]. The observed reduction in
229 diversity between influenza virus populations in donor animals and those that become
230 established in recipients is furthermore consistent with prior studies on influenza in human
231 cohorts and experimental animals [13, 25, 26, 29]. Our conclusion that early viral dynamics in
232 the recipient make a major contribution to the loss of diversity during transmission is, however,
233 novel. This pattern was likely not apparent in prior work due to the time points analyzed. In
234 human cohort studies, sample collection is typically triggered by the onset of symptoms, such
235 that the very early stages of infection are often not sampled [13]. Similarly, previous studies of
236 the influenza A virus transmission bottleneck in experimental models did not characterize viral
237 populations present early after transmission [25, 29].
238
239 The very transient nature of early, diverse, viral populations is also apparent within our own
240 data sets. Direct sequencing of viral genomes sampled from the nasal cavity of recipient hosts
241 revealed a collapse of barcode diversity over a period of approximately 24 hours. Conversely,
242 sequencing of infectious isolates derived from the sampled populations suggested a more rapid
243 extinction of viral lineages. Among eight recipient animals that each showed a high diversity of
244 genomes in early nasal lavages, only three showed a high diversity of infectious viruses in these
245 sampled populations. In the remaining five, the infectious viral populations detected early
246 matched the low diversity populations that persisted through the course of infection. We
247 postulate that the discrepancy seen in these five samples relates to the detection of viral
248 genomes within infected cells when sequencing bulk nasal lavage. In line with the
249 heterogeneity of viral burst size at a single-cell level [36-38], many infected cells will not
250 produce progeny viruses that can be detected by plaque assay. Thus, these five samples appear
251 to have captured a time point in which many unique barcodes were being replicated within
252 infected cells, but the bottleneck had already occurred at the level of the released viral
253 population. Conversely, the three samples that showed concordance between bulk nasal lavage
254 and plaque isolates captured an earlier phase of the transition from high to low diversity.
255
256 Based on our results, we propose a model in which many viruses are delivered to the recipient
257 and initiate infection in the nasal cavity, such that their genomes are replicated to a level
258 detectable by sequencing. Among this first generation of infected cells, only a subset produces
259 large bursts of progeny viruses early [36-39]. The genotypes replicated in those cells initiate a
260 second round of infection, which is subject to similar heterogeneity. Within these very first

generations of viral amplification, the few genotypes replicated rapidly and to high levels quickly dominate both infected-cell and released-virus populations. The eventual death of producer cells precipitates loss of barcode diversity in the infected-cell population. Among released viruses, competition for new target cells leads to the extinction of many minor lineages [40, 41]. Innate antiviral immunity likely hastens the contraction of barcode diversity both in infected cells and in released viruses [42, 43]. Within these contractions, some minor lineages may persist owing to long-lived cells [44, 45] or spatial isolation [46-48]. Our data suggest that such persistent lineages can expand late in infection as dominant lineages are cleared, yielding a transient rebound in barcode diversity.

In contrast to recipient animals infected through transmission, we found only modest reductions in diversity in donor animals during infection. This finding extended to animals inoculated with low doses, where a bottleneck acting at inoculation was apparent, but a growth-induced bottleneck acting during the establishment of infection was not. Since the kinetics of viral expansion may differ with infection route, this difference may be attributable to the timing of sampling. Alternatively, viral population dynamics may differ between animals infected through inoculation and transmission owing to differences in the mode of viral deposition. For example, viral aggregation state may differ, which would modify the potential for coordinated infection and, in turn, alter the extent of heterogeneity in viral burst size [38, 49]. Alternatively, or in addition, the size of the epithelial area across which viruses are delivered may differ. If this area is greater following intranasal instillation, the ensuing spatial separation of initial sites of infection may allow more lineages to establish within the host [46-48].

Seasonal influenza virus evolution at large geographic and temporal scales is characterized by a clear pattern of positive selection: on a recurring basis, antigenically distinct variants sweep the global viral population, driving epidemic spread [50-52]. In contrast, the within-host evolutionary dynamics of influenza viruses show strong genetic drift and purifying selection; positive selection has rarely been observed [13, 14, 53, 54]. A similar dichotomy is true for SARS-CoV-2 at within-host and population scales [55-58]. Our data suggest that these scales may be linked by a transmission process in which selective forces have an opportunity to act before potent stochastic forces within hosts. For example, antibodies present at mucosal surfaces could act on an antigenically diverse incoming viral population, mediating antigenic selection. Only positively selected variants would then be available to pass through the stochastic bottleneck associated with population expansion. Experiments performed in pre-immune hosts are needed to formally test this model.

In summary, our data reveal a two-stage transmission process in which transfer between hosts can include a large and diverse viral population but early events in the recipient host impose a stringent and stochastic bottleneck. Transmission may therefore represent an opportunity for selection to operate in the earliest phases of infection before any single genotype sweeps the population. This effect is expected to have a strong influence on viral evolution and is likely relevant across diverse viral families.

MATERIALS & METHODS

Ethical considerations

All the 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 approved by the IACUC of Emory University (PROTO201700595) for the guinea pig (*Cavia porcellus*). The animals were humanely euthanized following guidelines approved by the American Veterinary Medical Association.

Cells

Madin–Darby canine kidney (MDCK) cells were a gift from Dr. Daniel Perez, University of Georgia, Athens, GA. A seed stock of MDCK cells at passage 23 was amplified and maintained in Minimal Essential Medium (Gibco) supplemented with 10% fetal bovine serum (FBS; Atlanta Biologicals) and Normocin (Invivogen). 293T cells (ATCC, CRL-3216) were maintained in Dulbecco's Minimal Essential Medium (Gibco) supplemented with 10% FBS and Normocin. All cells were cultured at 37 °C and 5% CO_2 in a humidified incubator. The cell lines were not authenticated. All cell lines were tested monthly for *Mycoplasma* contamination while in use. The medium for the culture of influenza A virus in MDCK cells (virus medium) was prepared by supplementing Minimal Essential Medium with 4.3% bovine serum albumin (BSA; Sigma) and Normocin.

Generation of Pan/99 NA-BC plasmid

The region of Pan/99 into which the barcode was inserted was first identified by aligning 593 sequences of H3N2-subtype influenza A viruses isolated between 1994 and 2004. A region with many nucleotide substitutions was identified in the NA segment from nucleotide position 484 to 532. Twelve synonymous mutations were identified within this region. Double-stranded Ultramers (IDT) were designed that contained degenerate bases with two possible nucleotides at each of the twelve chosen barcode sites (cacagtacatgataggacccccttaycgraccytattgatgaatgarttrggtgtyccattycayytrggracyaagcaagtgtgtatagcatggtcc).

Site-directed mutagenesis was used to insert an Xho1 restriction site into the wild-type reverse genetics plasmid, pDP Pan/99 NA, prior to barcode insertion. This was done to give a means of destroying the parental template following barcode insertion. To generate a linearized template for barcode insertion, the following steps were performed. Xho1 digestion of the plasmid stock was followed by phosphatase treatment using rSAP (NEB) to dephosphorylate the cut ends of the plasmid. The plasmid was then amplified by PCR using primers that extend outward from the barcode region: P99_NA_536F 5'-caagtgtgtatagcatggtcc-3' and P99_NA_479R 5'-gggtcctatcatgtactgtg-3'. PCR purification (Qiagen QIAquick PCR Purification Kit) was used to isolate the linearized PCR product, followed by a dual digestion with Dpn1 and Xho1 to remove residual WT plasmid. PCR purification was repeated, and then an assembly reaction using the NEBuilder HiFi DNA Assembly Kit (NEB) was performed to insert the Ultramers into the linearized vector and re-circularize. The product was then transformed into DH5-α cells (NEB).

348 After plating onto LB-amp plates, approximately 1 x 10^4 colonies were collected and pooled into
349 LB-amp culture media and then incubated at 37°C for five hours prior to harvesting the
350 bacterial population for plasmid purification by maxiprep (Qiagen Plasmid Maxi Kit). The
351 presence of a diverse barcode in the plasmid stock was verified by next generation sequencing
352 (see below). This plasmid stock was then used to generate Pan/99 NA-BC virus by reverse
353 genetics in combination with seven plasmids encoding Pan/99 WT gene segments in a pDP2002
354 vector [59].
355
356 **Viruses**
357 Pan/99 WT and Pan/99 NA-BC viruses were derived from influenza A/Panama/2007/99 (H3N2)
358 virus (Pan/99 WT) and were generated by reverse genetics. In brief, 293T cells transfected with
359 eight reverse genetics plasmids 16–24 h previously were co-cultured with MDCK cells in virus
360 medium at 33°C for 40 h. Recovered virus was propagated in MDCK cells to generate working
361 stocks. Propagation was carried out from low MOI to avoid accumulation of defective viral
362 genomes but with a sufficient viral population size to maintain barcode diversity. Titration of
363 stocks and experimental samples was carried out by plaque assay in MDCK cells. The presence
364 of a diverse barcode in the virus stock was verified by next generation sequencing (see below).
365
366 **Growth kinetics**
367 Replication of Pan/99 WT and Pan/99 NA-BC was determined in triplicate culture wells. MDCK
368 cells in six-well dishes were inoculated at an MOI of 0.01 PFU/cell in PBS. After 1 h incubation at
369 33°C, inoculum was removed, cells were washed 3x with PBS, 2 mL virus medium was added to
370 cells, and dishes were returned to 33°C. A 120 μL volume of culture medium was sampled at
371 the indicated times points and stored at −80 °C. Viral titers were determined by plaque assay on
372 MDCK cells.
373
374 **Guinea pig infections**
375 Female Hartley strain guinea pigs weighing 250–350 grams were obtained from Charles River
376 Laboratories and housed by Emory University Department of Animal Resources. Guinea pigs
377 were sedated with ketamine (30 mg/kg) and xylazine (4 mg/kg) by intramuscular injection prior
378 to intranasal inoculation or nasal lavage. Virus inoculum was given intranasally in a 300 μL
379 volume of PBS containing 5 x 10^4 PFU of Pan/99 NA-BC. At day 1 post-inoculation, one naïve
380 animal was placed with each inoculated animal in either a cage that allowed for direct physical
381 contact (to model transmission by all modes) or a cage in which the two animals were
382 separated by a double-walled, perforated metal barrier (to model transmission via airborne
383 infectious respiratory particles). Nasal lavage was performed with 1 mL PBS per animal on days
384 1-7 post-inoculation for inoculated animals and days 2-8 post-inoculation for exposed animals.
385 Collected fluid was divided into aliquots and stored at -80 °C. Viral titers of nasal lavage samples
386 were subsequently determined by plaque assay.
387
388 For the dose escalation experiment (**Supplemental Figure 5**), guinea pigs were obtained and
389 sedated as described above. Virus inoculum was given intranasally in a 30 μL volume of PBS
390 containing 1x10^2 PFU, 3.3x10^2 PFU, 1x10^3 PFU, 3.3x10^3 PFU, 1x10^4 PFU, or 5 x 10^4 PFU of Pan/99

391 NA-BC. Nasal lavage was performed with 1 mL PBS per animal on days 1-4 post-inoculation, and
392 samples were stored as described above.
393
394 **Sample processing for next generation sequencing of barcode region**
395 The following method was used to validate plasmid and virus stocks and to evaluate
396 experimental samples. Viral RNA extraction (QiaAmp Viral RNA kit, Qiagen) was performed
397 using a 140 µL volume of each nasal lavage sample or virus stock, followed by reverse
398 transcription (Maxima RT, Thermo Fisher) with pooled Univ.F(A)+6 and Univ.F(G)+6 primers [41,
399 60]. Either cDNA or, for the purpose of plasmid validation, plasmid DNA was subjected to PCR
400 amplification with primers flanking the barcode region. Samples were sent for amplicon
401 sequencing to either Azenta Life Sciences or the Emory National Primate Research Center
402 (ENPRC) Genomics Core. For samples sent to Azenta, PCR to amplify the region containing the
403 barcode was performed using PFU Turbo AD (ThermoFisher) with primers with partial
404 sequencing adapters that generated a 155-nt product (P99_NA_adptr_428-449_F 5'-
405 acactctttccctacacgacgctcttccgatctggaacaacactaaacaacaggc-3' and P99_NA_adptr_563-582_F 5'-
406 gactggagttcagacgtgtgctcttccgatctgcttttccatcgtgacaact-3'). For samples sequenced by ENPRC,
407 amplicon PCR was performed in a similar manner but using primers with sequencing adapters
408 that generated a 100-nt product (P99NA_462F5_adptr 5'-
409 tcgtcggcagcgtcagatgtgtataagagacagggactcctcagtacatgataggacccctt-3' and P99NA_553Rev_adptr
410 5'- gtctcgtgggctcggagatgtgtataagagacagccatgctatacacacttgct-3'). Thirty PCR cycles were
411 performed for all samples. Column-based PCR purification was performed on all samples
412 followed by quantification of DNA using either NanoDrop or Qubit. Samples were normalized to
413 20 ng/uL in nuclease-free H_20.
414
415 Samples submitted to Azenta underwent Amplicon-EZ sequencing, an Illumina-based
416 sequencing service compatible with amplicons of 150-500 nt in length that does not include a
417 fragmentation step in library preparation. Amplicons were sequenced as 2 × 250 bp paired
418 reads and demultiplexed prior to delivery. At ENPRC, library preparation was performed with
419 the omission of a tagmentation step and sequencing was performed on an Illumin NovaSeq
420 6000 platform. Amplicons were sequenced as 2 x 100 bp paired reads and demultiplexed prior
421 to delivery.
422
423 **Analysis of barcode sequencing data**
424 Sequences were processed using our custom software, BarcodeID, available in the GitHub
425 repository: https://github.com/Lowen-Lab/BarcodeID. Briefly, BarcodeID uses BBTools [35] to
426 process raw reads, then uses a custom Python script to screen and identify barcode sequences
427 present in each sample, then performs an error correction step to remove spurious barcode
428 sequences, before finally calculating diversity statistics and writes summary tables. BBMerge
429 screens for adapter sequences and merges forward and reverse reads using default settings,
430 and BBDuk merges reads with a low average quality (<30). BarcodeID then screens each read to
431 verify that the nucleotides at barcode and non-barcode sites match the nucleotides expected at
432 those sites. Reads with mismatches are excluded from overall barcode sequence counts, but
433 BarcodeID collects all high-quality variant amplicons and calculates overall mismatch rates by

site to determine if any mutants with non-barcode alleles are driving any observed barcode dynamics. Samples with evidence of non-barcode driven dynamics were excluded from further analyses.

The error correction step was created to emulate the error correction software designed for bacterial 16S rRNA amplicons, DADA2 [61]. The largest difference between the two methods is how the nucleotide substitution model is generated. Where DADA2 uses a complicated statistical method, we can measure it directly from the backbone sites which should only deviate from the expected sequence due to mutation or any error accumulated through the process of library preparation and sequencing. Specifically, it measures the rate at which bases with a specific quality score match or do not match the expected base at each backbone site ($\lambda_{ji,Q}$). For example, the rate of A to T with a quality score of 25 ($\lambda_{TA,25}$) is the number of times a T with quality score of 25 was observed at a backbone site where A was expected, divided by the number of observations of all bases with a quality score of 25 at those sites. Using this error model and the observed prevalence of barcode sequences, BarcodeID estimates the probability that any barcode sequence could be derived from another through error. This probability calculated by:

$$p_{(j\rightarrow i)} = \frac{1}{1-\rho_{poison}(n_j,\lambda_{ji},0)}\sum_{a=n_i}^{\infty}\rho_{poison}(n_j,\lambda_{ji},a)$$

Where the probability of a barcode sequence *j* becoming barcode sequence *i* is Poisson distributed, and determined by the prevalence of both barcodes, n_j and n_i, and the product of probabilities of observing all nucleotide changes required to mutate barcode *j* into *i*.

BarcodeID uses the same agglomerative clustering process as DADA2 that attempts to iteratively identify barcodes that are too abundant to be explained due to error and collapse erroneous barcodes into larger barcode clusters. To initiate the agglomeration process, the most abundant barcode sequence with its associated quality score values in a sample is chosen as the initial seed barcode. BarcodeID then calculates the probability that every observed barcode sequence-quality combination could have been observed through mutation of the initial seed sequence ($p_{i,j}$ where *j* is the seed barcode and *i* is the query barcode). Barcodes are then ranked in order of increasing probability of observation due to error. If the highest ranked barcode is higher than a Bonferroni adjusted p-value threshold, then it is used to create a new barcode cluster seed. This iterative process of adding the top seed before re-calculating the probability of observation due to error of all remaining barcodes relative to the growing list of seed barcodes continues until no more new seeds are found. Then, BarcodeID enters the agglomeration phase, where query barcodes are sorted in decreasing order of probability of observation due to error, and the barcode with the highest probability of being an error derived from a seed barcode is added to that seed barcode's cluster. This process proceeds again iteratively until no barcodes can be collapsed into existing barcode clusters, at which point the remaining barcodes are included as new seeds until all observed barcodes are included. Importantly, the $P_{i>j}$ of barcodes that are observed once will always be equal to 1 and are therefore unable to be clustered. Accordingly, singleton barcodes are excluded from calculating diversity statistics.

476
477 To account for uneven sampling intensity, BarcodeID performs repeated subsampling and
478 linear regression to calculate diversity statistics that are more directly comparable
479 between samples. Specifically, samples were subsampled at up to 19 different read
480 depths (between 10^3 to 10^6 reads, where possible), with five iterations per subsampled
481 depth. The median value of each diversity statistic at each depth was then used to fit the
482 relationship between each diversity statistic and sampling intensity. The fit was then used
483 to calculate an adjusted value for all samples at an equivalent depth. For alpha diversity
484 statistics (richness, Shannon diversity, Simpson diversity, and evenness) the equivalent
485 depth was 10,000 reads per sample because it provided the sufficient depth to distinguish
486 samples across multiple levels of diversity, while minimizing bias due to variable sampling
487 intensity. For beta diversity statistics (Bray-Curtis and Jaccard dissimilarity), 50,000 reads
488 per sample was chosen because the added depth improved resolution in pairwise sample
489 comparisons.
490
491 **Validation of barcode sequencing data**
492 Samples were processed for barcode sequencing if infectious virus could be detected therein by
493 plaque assay (which has a limit of detection of 50 PFU/mL). We found that many reads were
494 obtained and analyzed for plaque-positive samples, irrespective of viral titer. To evaluate the
495 reproducibility of sequencing results, samples from three guinea pigs were sequenced in
496 triplicate. This set of samples includes high titer, low titer and plaque negative samples. For
497 each sample, the original RNA extract was processed independently three times to produce,
498 amplify, purify and sequence the cDNA. Results indicate that barcode composition detected in
499 plaque-positive samples is highly reproducible, while barcode composition detected in the
500 plaque-negative samples is not reproducible (**Supplemental Figure 4C**).
501
502 **Preparation of nasal lavage samples for viral whole-genome sequencing**
503 Viral RNA extraction (QiaAmp Viral RNA kit, Qiagen) was performed using a 140 μL volume of
504 each nasal lavage sample, followed by one-step reverse transcription PCR amplification of full
505 viral genomes using pooled Univ.F(A)+6, Univ.F(G)+6 primers and Univ.R primers and
506 SuperScript III Platinum kit (ThermoFisher) [41, 60]. Following PCR purification (Qiagen
507 QiaQuick PCR Purification Kit), cDNA was processed at the ENPRC core for sequencing on an
508 Illumina NovaSeq 6000 platform. Samples were sequenced as 2 x 100 bp paired reads and
509 demultiplexed prior to delivery.
510
511 **Analysis of whole-genome sequencing data**
512 Whole genome sequencing reads were merged and filtered for low average quality (≥30) using
513 BBMerge, then separated according to the segment using BLAT [62]. The reads were then
514 mapped to their corresponding reference segment using BBMap, with local alignment set to
515 false. From these alignments, we used custom Python scripts to identify iSNVs and reads that
516 map to the barcoded region of the NA segment. Cutoffs for inclusion of iSNVs were set
517 empirically. First, sites were evaluated based on their total coverage and the average quality
518 and mapping statistics. Only sites with ≥100x coverage were considered. For minor variants at
519 these sites to be included in subsequent analyses, they were required to be present at ≥1%

frequency and have an average phred score of ≥35, and the reads that contained the minor allele at any given site also had to have sufficient mapping quality to justify inclusion. Specifically, reads containing the minor allele needed an average mapping quality score of ≥40, the average location of the minor allele needed to be ≥20 bases from the nearest end of a read, and the reads overall needed to have ≤2.0 average mismatch and indel counts relative to the reference sequence.

Analysis of beta diversity

Dissimilarity between two populations can be measured using beta diversity metrics, and in this study, we used the both Jaccard index and the Bray-Curtis dissimilarity [63, 64]. Both metrics consider the species shared between two populations, and in this case, unique barcodes were considered species. In addition to presence/absence data, the Bray-Curtis dissimilarity also reflects abundance data. For either measure, a value closer to one indicates that the two populations are more dissimilar, whereas a value closer to zero indicates that the two populations are more alike in composition. Pairwise comparisons of barcode data were made between the viral populations present in plaque-positive nasal lavage samples acquired from a contact pair of guinea pigs.

Data availability

All data generated or analyzed during this study are included in the manuscript. At the time of initial submission, all raw sequencing data has been uploaded to NCBI's Sequence Read Archive under the accession number PRJNA1253752.

Acknowledgments/Funding

We are grateful to Katia Koelle for helpful discussion. This study was supported in part by R01 AI165644 and the National Institute of Allergy and Infectious Diseases (NIAID) Centers of Excellence for Influenza Research and Response (CEIRR) contract no. 75N93021C00017.

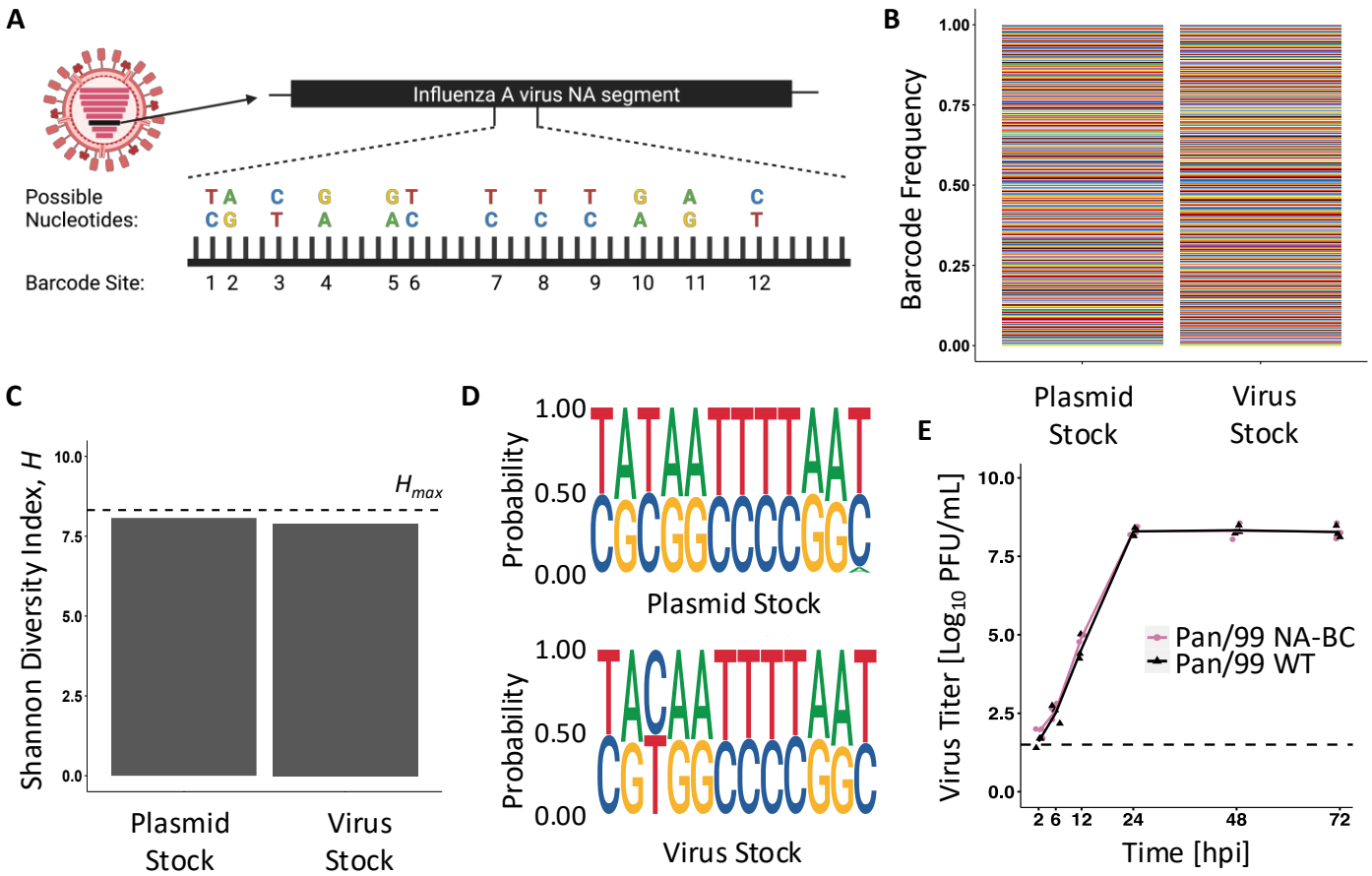


Figure 1. NA barcode diversity is maintained in both plasmid and virus stocks, and the barcode does not affect overall fitness. A) Barcode design for the NA segment of influenza A/Panama/2007/99 (H3N2) virus. At each of twelve sites, one of two nucleotides is possible, allowing for up to 4096 unique barcodes within the population. **B)** Barcodes detected in the pDP Pan/99 NA-BC plasmid preparation and passage 1 stock of Pan/99 NA-BC virus. Colors represent unique barcodes, and their frequencies within the stock are indicated by the height of the color. **C)** Shannon Diversity Index (*H*) of the stock samples is compared to the maximum possible diversity for this system (H_{max} = 8.32, shown with a horizontal dashed line). This theoretical maximum reflects a population in which all 4096 potential barcodes are equally represented. **D)** Sequence logo plots demonstrate the sequence motifs present in plasmid and passage 1 virus stocks. Each pair of nucleotides represents one of the twelve bi-allelic sites. The height of the letter indicates the corresponding nucleotide frequency in the stock. **E)** Multicycle replication of Pan/99 WT (black) and Pan/99 NA-BC (pink) viruses. Infections were performed in triplicate, and data points show results from individual cell culture dishes. Lines connect mean titers. Dashed line indicates limit of detection (50 PFU/mL). Two-way ANOVA showed no significant difference between viruses at any time point (p = 0.11). Figure 1A created in BioRender.

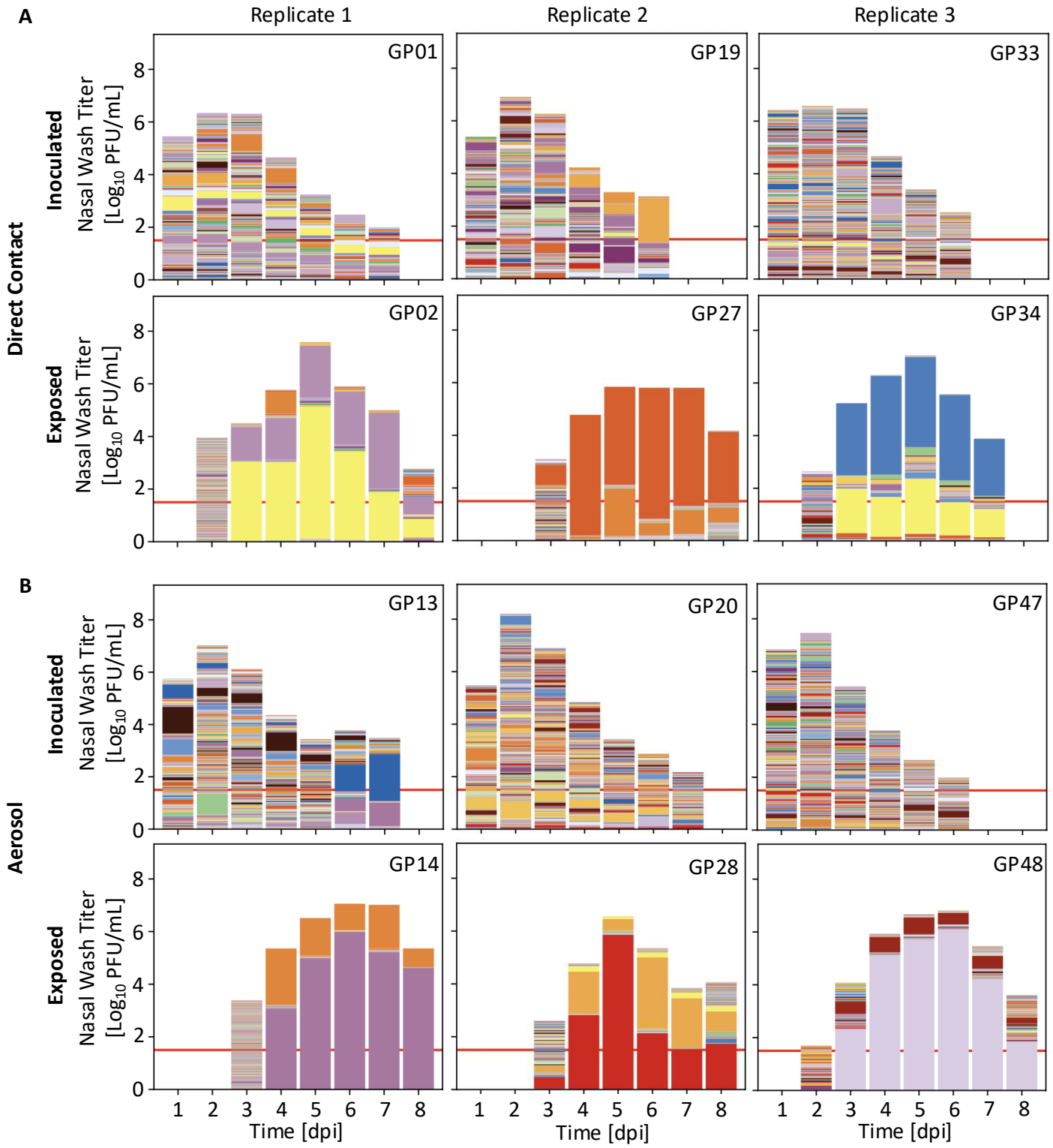


Figure 2. Population diversity declines between inoculated and exposed guinea pigs. Viral titers in nasal lavage samples are indicated by the overall height of the bar. Red lines show LOD (50 PFU/mL). Colors within the bars represent unique barcodes, and the height of each color indicates barcode frequency within the sample. Plots for individual animals are paired with

those of their cage mate. Representative pairs for direct contact (**A**) and aerosol (**B**) exposure are shown from three experimental replicates. Guinea pig ID numbers are shown in the upper right corner of each plot. Data from additional animals are shown in Supplemental Figure 2.

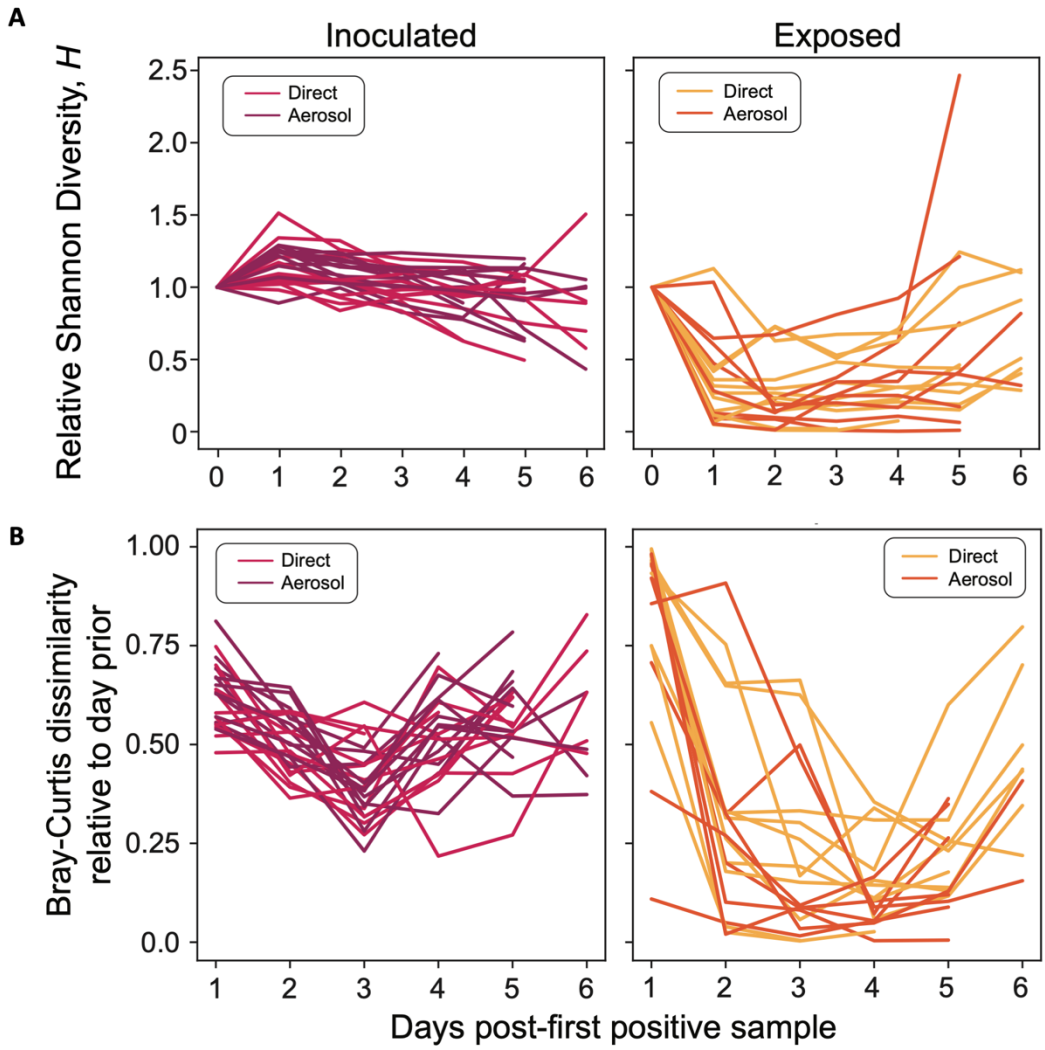
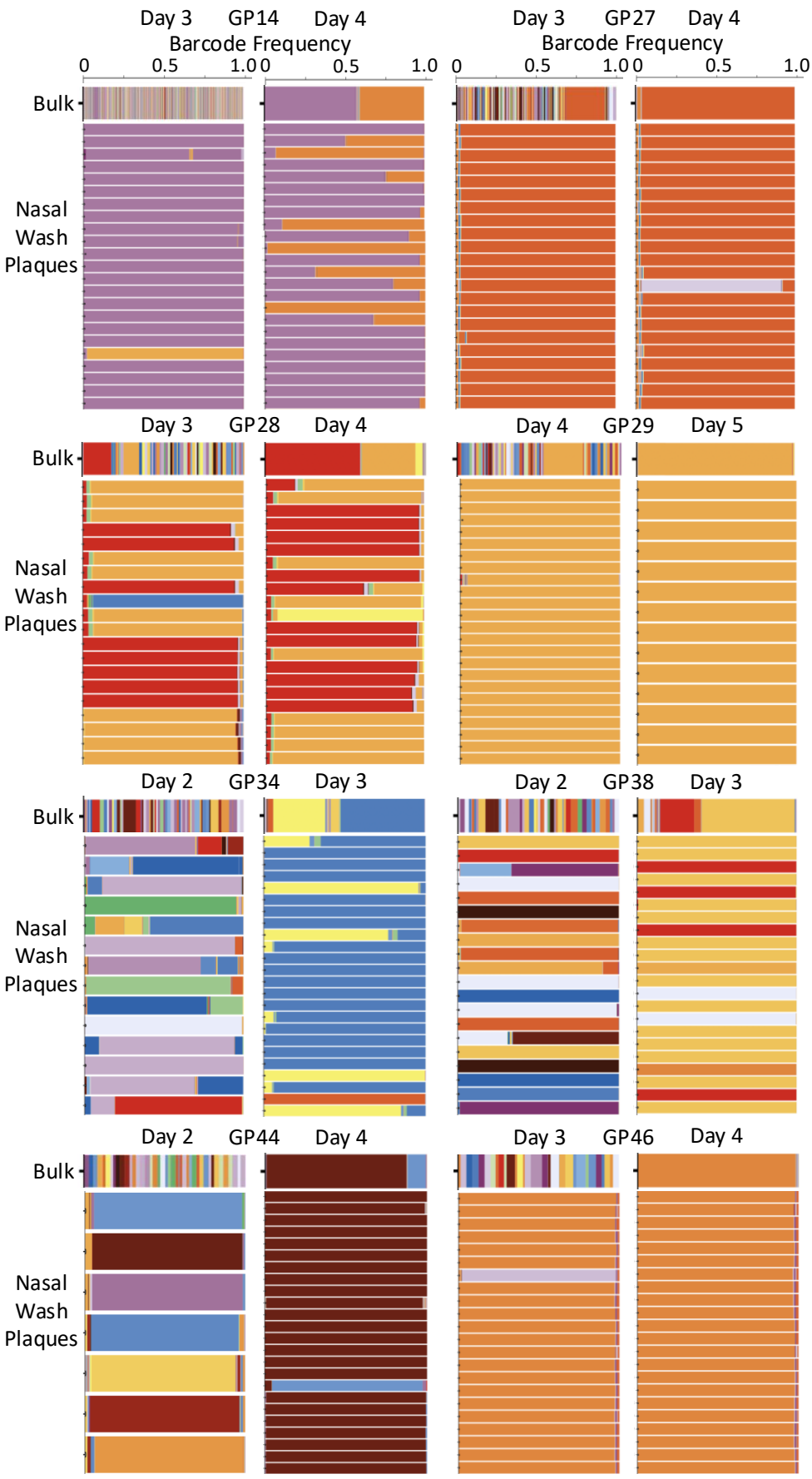
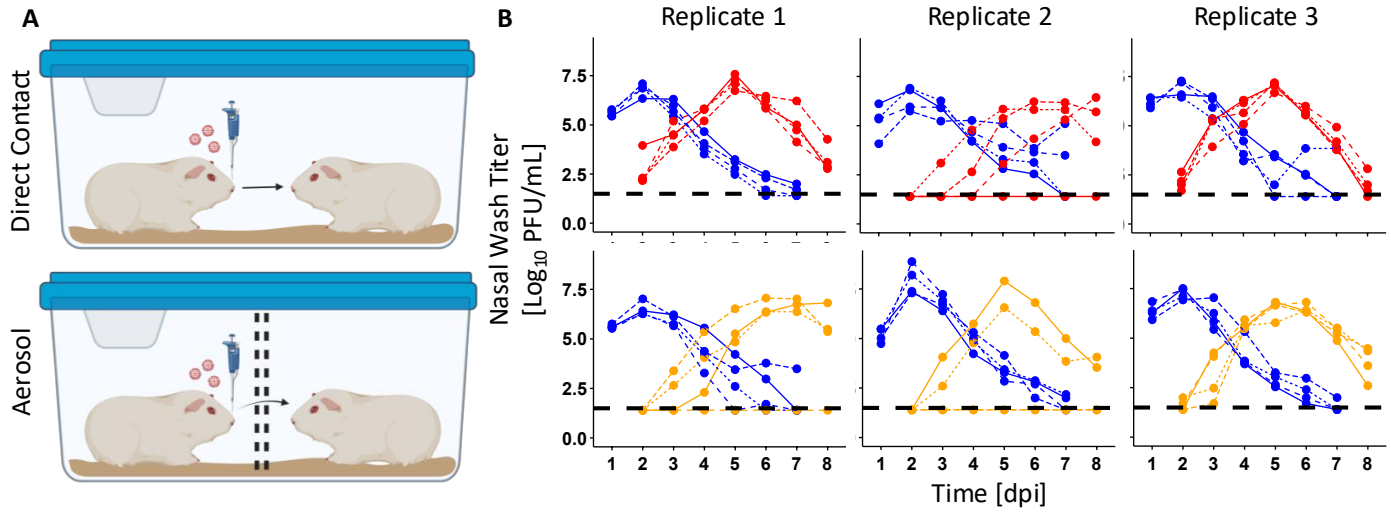


Figure 3. Initial high barcode diversity in exposed animals plummets after the first 1-2 days of viral positivity. A) For each exposed animal, the Shannon diversity index *(H)* was normalized to that on the first day of viral positivity (as determined by plaque assay). Data from a total of 24 inoculated animals (left) and 19 exposed animals (right) across three experiments are shown. The Shannon diversity in each animal relative to their first day of positivity is on average 3.7x lower in exposed animals compared to inoculated animals across all timepoints ($p<0.0001$). **B)** Changes in barcode composition between successive days was assessed using the Bray-Curtis dissimilarity index to compare barcode composition in each sample to that observed on the prior day. Lines connect data points for one animal, with inoculated (left) and exposed (right) animals shown on separate graphs.

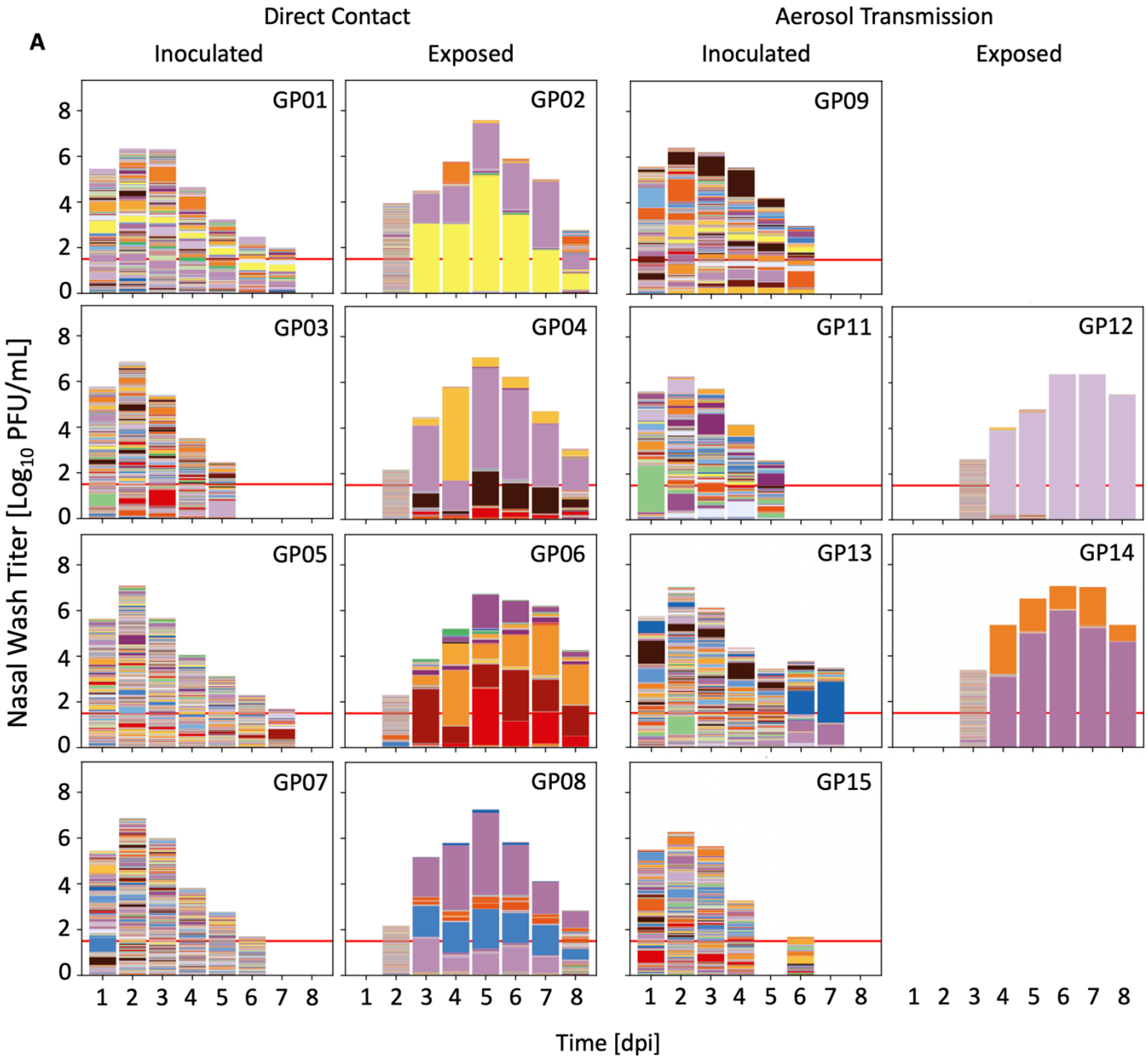
617

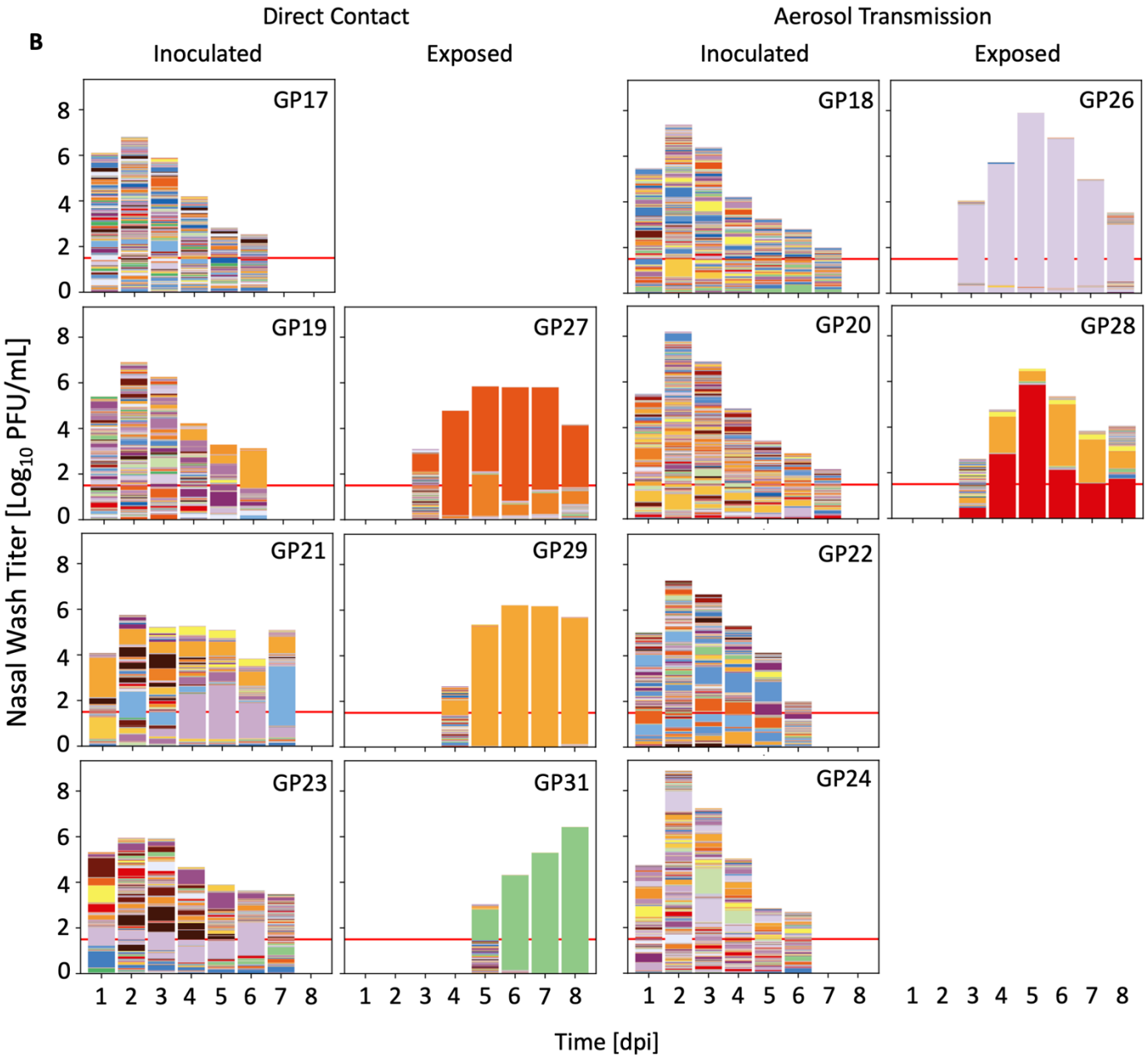

Day 3
GP14
Day 4
Barcode Frequency
0
0.5
1.0
Day 3
GP27
Day 4
Bulk
Nasal
Wash
Plaques
Day 3
GP28
Day 4
Day 4
GP29
Day 5
Day 2
GP34
Day 3
Day 2
GP38
Day 3
Day 2
GP44
Day 4
Day 3
GP46
Day 4

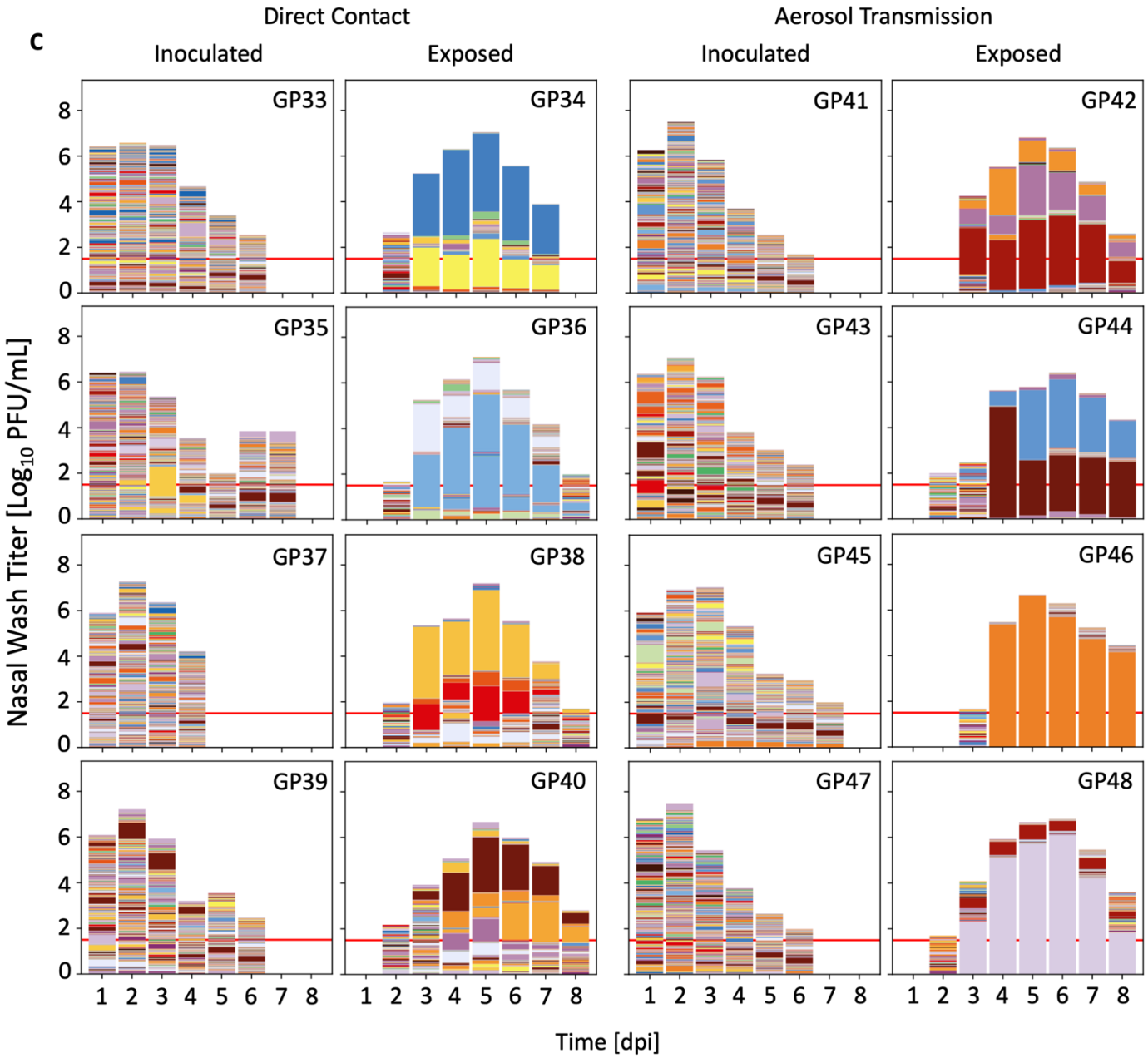
618 **Figure 4. Infectious viruses isolated early after transmission carry diverse barcodes in a subset**
619 **of animals.** For eight exposed animals, up to 24 plaques were isolated from nasal wash samples
620 collected on the first and second days of positivity. Barcodes detected in isolated plaques are
621 shown beneath those detected in the corresponding nasal lavage sample, with each row
622 representing a plaque. Each color represents a unique barcode, with frequency shown on the
623 horizontal axis.
624
625
626



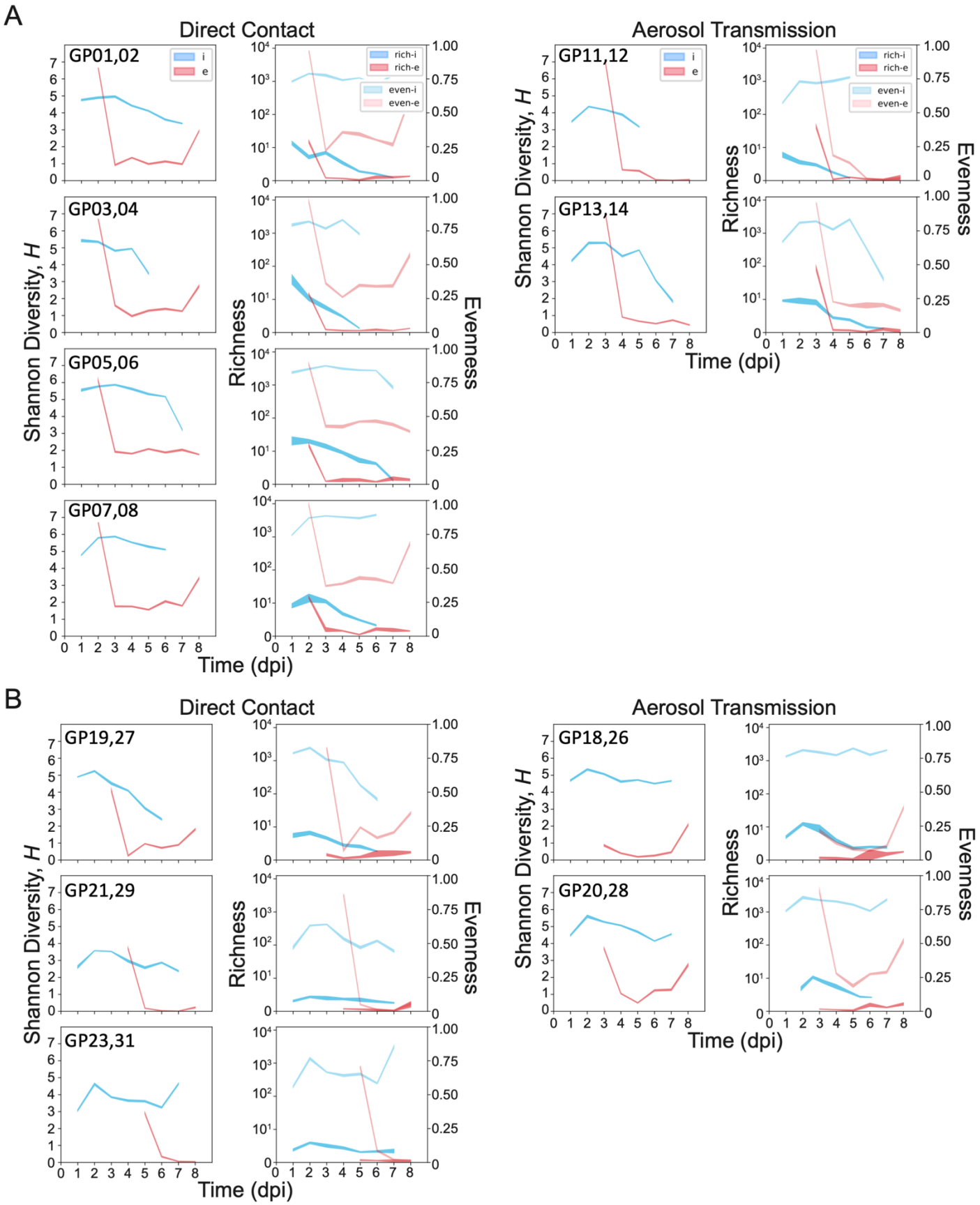
627 **Supplemental Figure 1. Pan/99 NA-BC infects inoculated animals and transmits to exposed**
628 **animals. A)** Schematic showing transmission set-up in guinea pigs. For each of the three
629 experimental replicates, eight guinea pigs were intranasally inoculated with 5 x 10^4 PFU of Pan/99
630 NA-BC in 300 **µ**L. Twenty-four hours post-inoculation, a single naïve animal was placed with each
631 inoculated animal. Cages either allowed for direct contact (n = 4) or maintained separation with
632 a double-walled, perforated metal barrier (n = 4). **B)** Viral titers in nasal lavage samples in direct
633 contact and aerosol exposure settings from replicates 1, 2, and 3. Inoculated animals are shown
634 in blue and exposed animals in red (direct contact) or yellow (aerosol contact). The dashed black
635 line represents the limit of detection (50 PFU/mL). Paired animals share the same line type.
636 Supplemental Figure 1A created in BioRender.
637
638
639
640
641
642
643
644
645
646
647

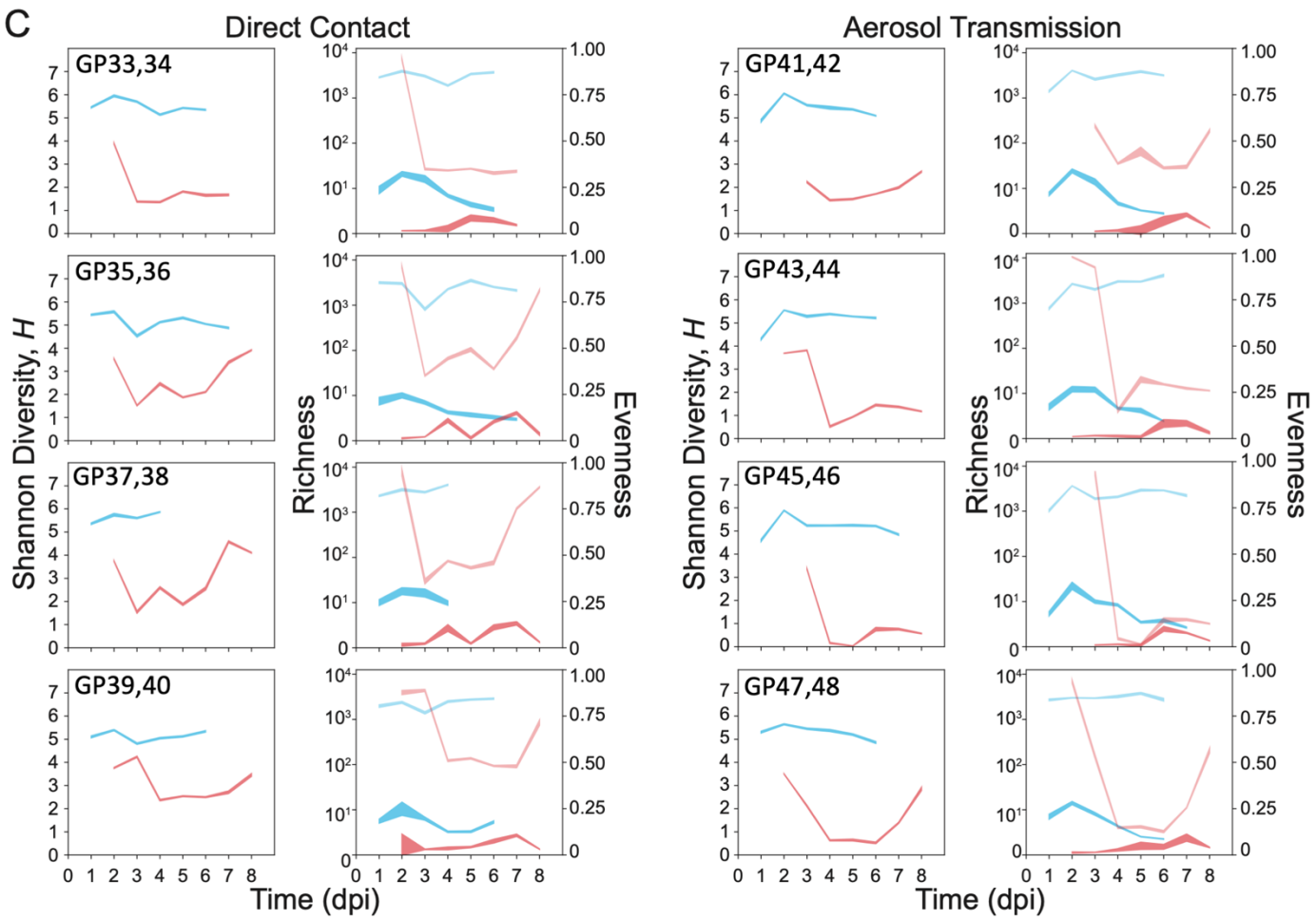

A
Direct Contact
Aerosol Transmission
Inoculated
Exposed
Inoculated
Exposed
GP01
GP02
GP09
GP03
GP04
GP11
GP12
GP05
GP06
GP13
GP14
GP07
GP08
GP15
Nasal Wash Titer [Log10 PFU/mL]
Time [dpi]


B
Direct Contact
Aerosol Transmission
Inoculated
Exposed
Inoculated
Exposed
GP17
GP18
GP26
GP19
GP27
GP20
GP28
GP21
GP29
GP22
GP23
GP31
GP24
Nasal Wash Titer [Log10 PFU/mL]
Time [dpi]

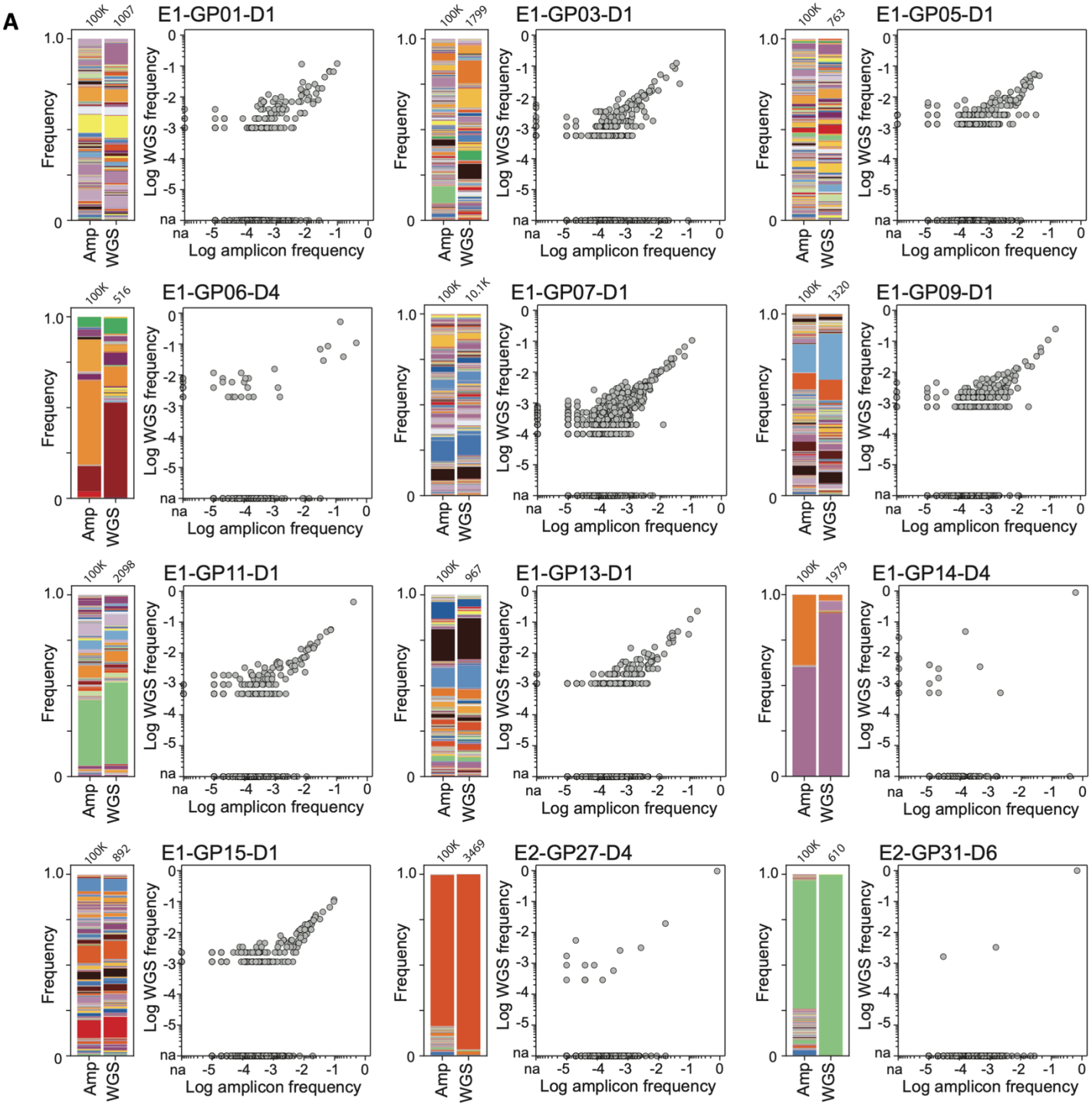


Supplemental Figure 2. Population diversity declines between inoculated and exposed guinea pigs. Data from experimental replicates 1, 2, and 3 are shown in panels **A, B,** and **C,** respectively. Guinea pig ID numbers are shown in the upper right corner of each plot. Nasal lavage titers are indicated by the total height of each bar. Colors within the bars represent unique barcodes, and the height of each color indicates the relative frequency within the sample. Only samples that were plaque-positive are shown. Red lines show LOD of 50 PFU/mL. Plots for individual animals are paired with those of their cage mate. For the exposed animal in the first aerosol transmission pair of Replicate 1 (GP10), most reads were discarded due to poor quality, and the data from this animal were excluded from further analyses.

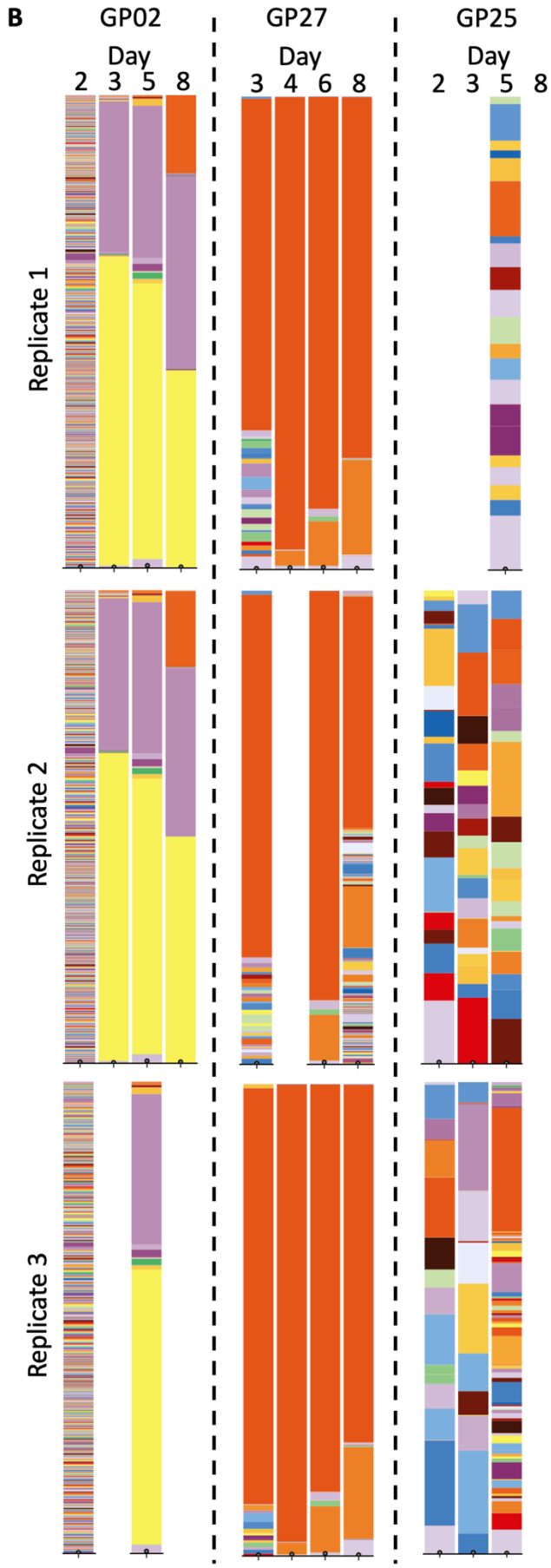

A
Direct Contact
Aerosol Transmission
GP01,02
GP03,04
GP05,06
GP07,08
GP11,12
GP13,14
i
e
rich-i
rich-e
even-i
even-e
Shannon Diversity, H
Richness
Evenness
Time (dpi)
B
Direct Contact
Aerosol Transmission
GP19,27
GP21,29
GP23,31
GP18,26
GP20,28
Shannon Diversity, H
Richness
Evenness
Time (dpi)

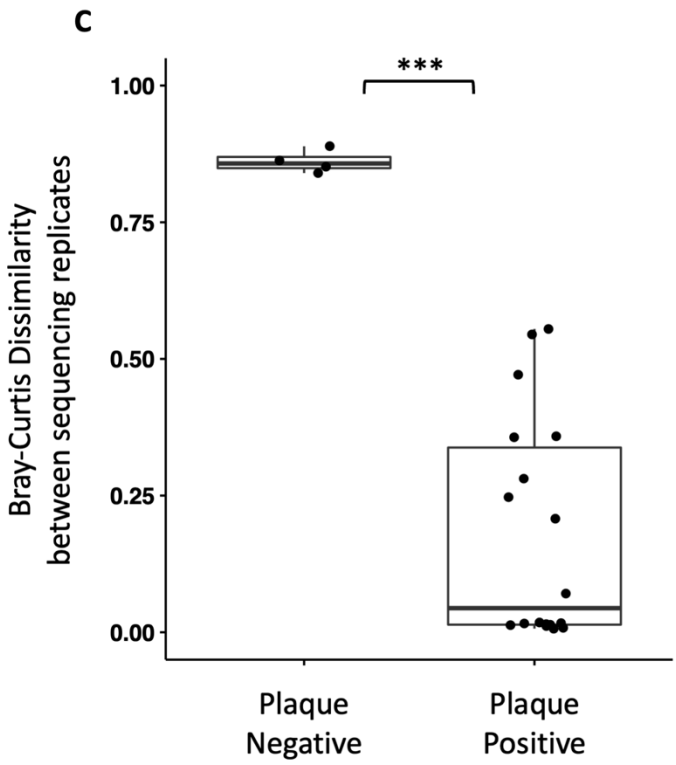


Supplemental Figure 3. Changes in richness and evenness both contribute to alterations in diversity. Shannon diversity (left), richness (right, left axis), and evenness (right, right axis) were determined for inoculated animals (i, blue) and exposed animals (e, red) in replicate experiments 1, 2, and 3 (**A, B,** and **C**, respectively). Data from cage mates are paired. In the right-hand plots, bold colors show richness and faded colors show evenness. Line widths show 95% confidence intervals based upon subsampling of barcode reads.

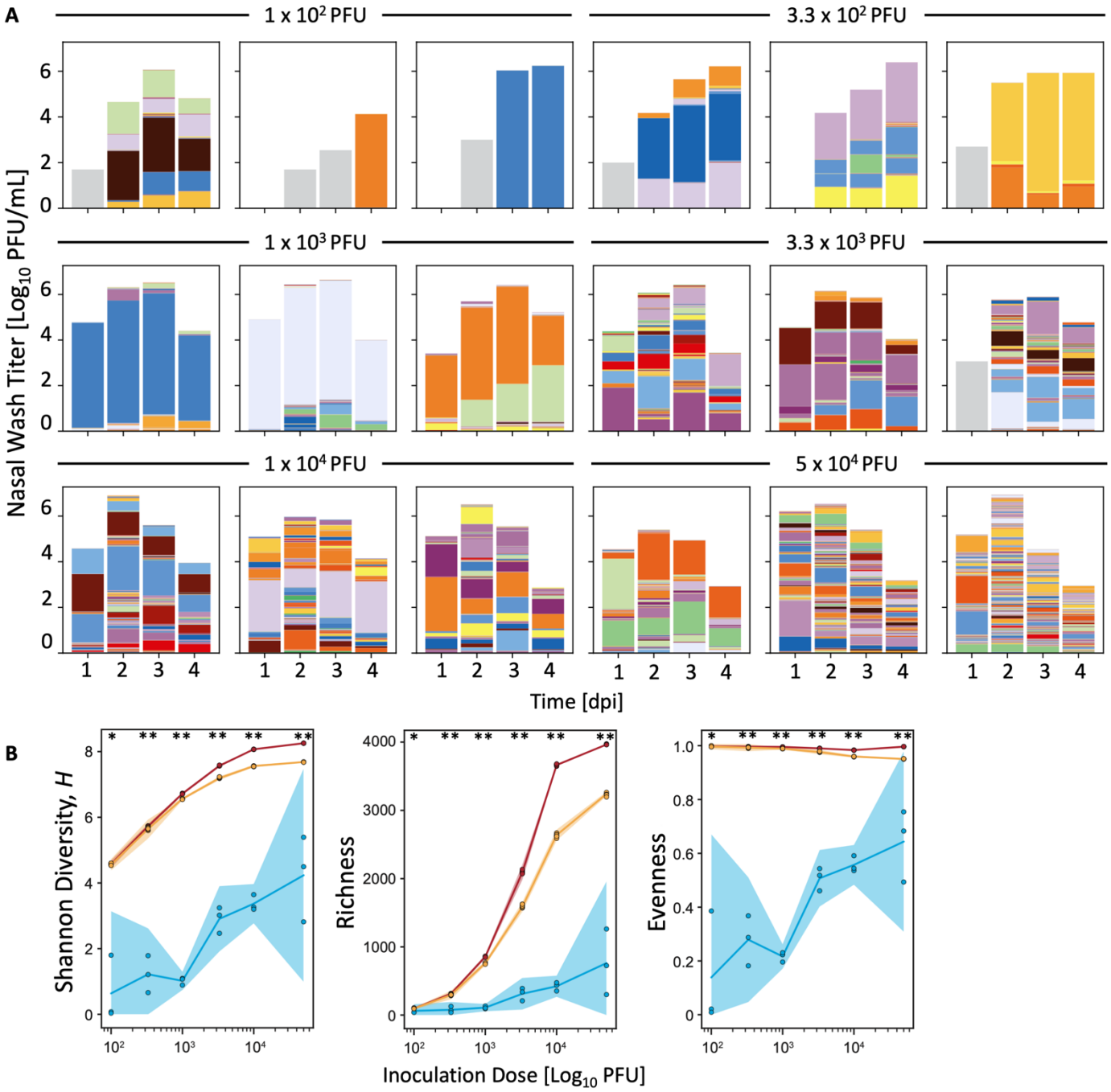


699
700
701
702
703
704





Supplemental Figure 4. Targeted amplicon sequencing of the barcode region is reliable and reproducible. A) Whole-genome sequencing (WGS) was performed on a subset of samples from experimental replicates 1 and 2. WGS reads that spanned the length of the barcode region were compared to targeted barcode sequencing (amplicon or Amp) results for each sample. Each graph is labeled with the experiment number (E1, E2, or E2), the guinea pig number (GP01, GP03, etc.), and the day post-inoculation (D1, D4, etc.). The number of reads mapping to the barcode region is indicated above the stacked bar plots. **B)** A subset of positive nasal wash samples from two animals (GP02, GP27) were sequenced in triplicate to evaluate reproducibility of sequencing results. These were compared to an animal that was plaque-negative throughout the experiment (GP25). **C)** Bray-Curtis Dissimilarity values between replicate sequencing runs from each sample are plotted for the plaque-negative animal and the plaque-positive animals (*** p-value < 0.0005, Wilcoxon Sign-Rank test).



Supplemental Figure 5. Growth-induced bottlenecks are not detected in inoculated animals, even at low inoculation doses. At each inoculation dose indicated above the plots, three guinea pigs were inoculated intranasally with Pan/99 NA-BC. **A)** Nasal lavage titers and barcode compositions over time. The height of the bars indicates viral titer in samples above the limit of detection (50 PFU/mL). Colors indicate barcodes detected, with barcode frequency shown by the height of the color. **B)** Maximum measured Shannon diversity, richness, and evenness of barcode

713 compositions in each animal are plotted by inoculation dose (blue). For comparison, theoretical
714 data are plotted to show the expected characteristics of inocula of each size derived from a
715 perfectly even population of 4096 barcodes (red) and from the Pan/99 NA-BC passage 1 stock
716 (orange). 95% confidence intervals are shown with shading. Statistical significance was
717 determined at each inoculation dose by Kruskal-Wallis Test (* p-value < 0.01, ** p-value < 0.001)
718 comparing the medians of the experimental data with those of the perfectly even population and
719 that of the passage 1 viral stock.
720
721 **References**

722 1. Nobusawa E, Sato K. Comparison of the mutation rates of human influenza A and B
723 viruses. J Virol. 2006;80(7):3675-8. doi: 10.1128/JVI.80.7.3675-3678.2006. PubMed PMID:
724 16537638; PubMed Central PMCID: PMCPMC1440390.
725 2. Smith DB, Inglis SC. The mutation rate and variability of eukaryotic viruses: an analytical
726 review. J Gen Virol. 1987;68 (Pt 11):2729-40. doi: 10.1099/0022-1317-68-11-2729. PubMed
727 PMID: 3316486.
728 3. Boni MF, Zhou Y, Taubenberger JK, Holmes EC. Homologous recombination is very rare
729 or absent in human influenza A virus. J Virol. 2008;82(10):4807-11. Epub 20080319. doi:
730 10.1128/JVI.02683-07. PubMed PMID: 18353939; PubMed Central PMCID: PMCPMC2346757.
731 4. Suarez DL, Senne DA, Banks J, Brown IH, Essen SC, Lee CW, et al. Recombination
732 resulting in virulence shift in avian influenza outbreak, Chile. Emerg Infect Dis. 2004;10(4):693-
733 9. doi: 10.3201/eid1004.030396. PubMed PMID: 15200862; PubMed Central PMCID:
734 PMCPMC3323103.
735 5. Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. Influenza virus reassortment occurs
736 with high frequency in the absence of segment mismatch. PLoS Pathog. 2013;9(6):e1003421.
737 Epub 20130613. doi: 10.1371/journal.ppat.1003421. PubMed PMID: 23785286; PubMed
738 Central PMCID: PMCPMC3681746.
739 6. Barbezange C, Jones L, Blanc H, Isakov O, Celniker G, Enouf V, et al. Seasonal Genetic
740 Drift of Human Influenza A Virus Quasispecies Revealed by Deep Sequencing. Front Microbiol.
741 2018;9:2596. Epub 20181031. doi: 10.3389/fmicb.2018.02596. PubMed PMID: 30429836;
742 PubMed Central PMCID: PMCPMC6220372.
743 7. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. The Mutational
744 Robustness of Influenza A Virus. PLoS Pathog. 2016;12(8):e1005856. Epub 20160829. doi:
745 10.1371/journal.ppat.1005856. PubMed PMID: 27571422; PubMed Central PMCID:
746 PMCPMC5003363.
747 8. Simon B, Pichon M, Valette M, Burfin G, Richard M, Lina B, et al. Whole Genome
748 Sequencing of A(H3N2) Influenza Viruses Reveals Variants Associated with Severity during the
749 2016(-)2017 Season. Viruses. 2019;11(2). Epub 20190128. doi: 10.3390/v11020108. PubMed
750 PMID: 30695992; PubMed Central PMCID: PMCPMC6410005.
751 9. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Rev Genet.
752 2007;8(3):196-205. Epub 20070130. doi: 10.1038/nrg2053. PubMed PMID: 17262054.
753 10. Spielman SJ, Weaver S, Shank SD, Magalis BR, Li M, Kosakovsky Pond SL. Evolution of
754 Viral Genomes: Interplay Between Selection, Recombination, and Other Forces. Methods Mol
755 Biol. 2019;1910:427-68. doi: 10.1007/978-1-4939-9074-0_14. PubMed PMID: 31278673.

11. Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 2016;17(11):704-14. doi: 10.1038/nrg.2016.104. PubMed PMID: 27739533.
12. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008;453(7195):615-9. Epub 20080416. doi: 10.1038/nature06945. PubMed PMID: 18418375; PubMed Central PMCID: PMCPMC2441973.
13. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. 2018;7. Epub 20180503. doi: 10.7554/eLife.35962. PubMed PMID: 29683424; PubMed Central PMCID: PMCPMC5933925.
14. Han AX, Felix Garza ZC, Welkers MR, Vigeveno RM, Tran ND, Le TQM, et al. Within-host evolutionary dynamics of seasonal and pandemic human influenza A viruses in young children. Elife. 2021;10. Epub 20210803. doi: 10.7554/eLife.68917. PubMed PMID: 34342576; PubMed Central PMCID: PMCPMC8382297.
15. Sobel Leonard A, McClain MT, Smith GJ, Wentworth DE, Halpin RA, Lin X, et al. The effective rate of influenza reassortment is limited during human infection. PLoS Pathog. 2017;13(2):e1006203. Epub 20170207. doi: 10.1371/journal.ppat.1006203. PubMed PMID: 28170438; PubMed Central PMCID: PMCPMC5315410.
16. McCrone JT, Lauring AS. Genetic bottlenecks in intraspecies virus transmission. Curr Opin Virol. 2018;28:20-5. Epub 20171103. doi: 10.1016/j.coviro.2017.10.008. PubMed PMID: 29107838; PubMed Central PMCID: PMCPMC5835166.
17. Valesano AL, Taniuchi M, Fitzsimmons WJ, Islam MO, Ahmed T, Zaman K, et al. The Early Evolution of Oral Poliovirus Vaccine Is Shaped by Strong Positive Selection and Tight Transmission Bottlenecks. Cell Host Microbe. 2021;29(1):32-43 e4. Epub 20201118. doi: 10.1016/j.chom.2020.10.011. PubMed PMID: 33212020; PubMed Central PMCID: PMCPMC7815045.
18. Wang GP, Sherrill-Mix SA, Chang KM, Quince C, Bushman FD. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J Virol. 2010;84(12):6218-28. Epub 20100407. doi: 10.1128/JVI.02271-09. PubMed PMID: 20375170; PubMed Central PMCID: PMCPMC2876626.
19. Edwards CT, Holmes EC, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, et al. Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. BMC Evol Biol. 2006;6:28. Epub 20060323. doi: 10.1186/1471-2148-6-28. PubMed PMID: 16556318; PubMed Central PMCID: PMCPMC1444934.
20. Smith DR, Adams AP, Kenney JL, Wang E, Weaver SC. Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. Virology. 2008;372(1):176-86. Epub 20071119. doi: 10.1016/j.virol.2007.10.011. PubMed PMID: 18023837; PubMed Central PMCID: PMCPMC2291444.
21. Forrester NL, Coffey LL, Weaver SC. Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. Viruses. 2014;6(10):3991-4004. Epub 20141023. doi: 10.3390/v6103991. PubMed PMID: 25341663; PubMed Central PMCID: PMCPMC4213574.

800 22. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, et al. HIV
801 transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science.
802 2014;345(6193):1254031. Epub 20140710. doi: 10.1126/science.1254031. PubMed PMID:
803 25013080; PubMed Central PMCID: PMCPMC4289910.
804 23. Sugrue E, Wickenhagen A, Mollentze N, Aziz MA, Sreenu VB, Truxa S, et al. The apparent
805 interferon resistance of transmitted HIV-1 is possibly a consequence of enhanced replicative
806 fitness. PLoS Pathog. 2022;18(11):e1010973. Epub 20221118. doi:
807 10.1371/journal.ppat.1010973. PubMed PMID: 36399512; PubMed Central PMCID:
808 PMCPMC9718408.
809 24. Itell HL, Guenthoer J, Humes D, Baumgarten NE, Overbaugh J. Host cell glycosylation
810 selects for infection with CCR5- versus CXCR4-tropic HIV-1. Nat Microbiol. 2024;9(11):2985-96.
811 Epub 20241003. doi: 10.1038/s41564-024-01806-7. PubMed PMID: 39363105.
812 25. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza A
813 virus transmission bottlenecks are defined by infection route and recipient host. Cell Host
814 Microbe. 2014;16(5):691-700. Epub 20141023. doi: 10.1016/j.chom.2014.09.020. PubMed
815 PMID: 25456074; PubMed Central PMCID: PMCPMC4272616.
816 26. Frise R, Bradley K, van Doremalen N, Galiano M, Elderfield RA, Stilwell P, et al. Contact
817 transmission of influenza virus between ferrets imposes a looser bottleneck than respiratory
818 droplet transmission allowing propagation of antiviral resistance. Sci Rep. 2016;6:29793. Epub
819 20160719. doi: 10.1038/srep29793. PubMed PMID: 27430528; PubMed Central PMCID:
820 PMCPMC4949428.
821 27. Hughes J, Allen RC, Baguelin M, Hampson K, Baillie GJ, Elton D, et al. Transmission of
822 equine influenza virus during an outbreak is characterized by frequent mixed infections and
823 loose transmission bottlenecks. PLoS Pathog. 2012;8(12):e1003081. Epub 20121220. doi:
824 10.1371/journal.ppat.1003081. PubMed PMID: 23308065; PubMed Central PMCID:
825 PMCPMC3534375.
826 28. Ferreri LM, Geiger G, Seibert B, Obadan A, Rajao D, Lowen AC, et al. Intra- and inter-host
827 evolution of H9N2 influenza A virus in Japanese quail. Virus Evol. 2022;8(1):veac001. Epub
828 20220108. doi: 10.1093/ve/veac001. PubMed PMID: 35223084; PubMed Central PMCID:
829 PMCPMC8865083.
830 29. Braun KM, Haddock Iii LA, Crooks CM, Barry GL, Lalli J, Neumann G, et al. Avian H7N9
831 influenza viruses are evolutionarily constrained by stochastic processes during replication and
832 transmission in mammals. Virus Evol. 2023;9(1):vead004. Epub 20230119. doi:
833 10.1093/ve/vead004. PubMed PMID: 36814938; PubMed Central PMCID: PMCPMC9939568.
834 30. Zaraket H, Baranovich T, Kaplan BS, Carter R, Song MS, Paulson JC, et al. Mammalian
835 adaptation of influenza A(H7N9) virus is limited by a narrow genetic bottleneck. Nat Commun.
836 2015;6:6553. Epub 20150408. doi: 10.1038/ncomms7553. PubMed PMID: 25850788; PubMed
837 Central PMCID: PMCPMC4403340.
838 31. Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, et al. Selection on
839 haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1
840 influenza viruses. Nat Commun. 2013;4:2636. doi: 10.1038/ncomms3636. PubMed PMID:
841 24149915; PubMed Central PMCID: PMCPMC3845350.
842 32. Magurran AE. Ecological Diversity and its Measurement. Princeton, NJ: Princeton
843 University Press; 1988.

844 33. Tao H, Steel J, Lowen AC. Intrahost dynamics of influenza virus reassortment. J Virol.
845 2014;88(13):7485-92. Epub 20140416. doi: 10.1128/JVI.00715-14. PubMed PMID: 24741099;
846 PubMed Central PMCID: PMCPMC4054463.
847 34. Xue KS, Bloom JD. Linking influenza virus evolution within and between human hosts.
848 Virus Evol. 2020;6(1):veaa010. Epub 20200217. doi: 10.1093/ve/veaa010. PubMed PMID:
849 32082616; PubMed Central PMCID: PMCPMC7025719.
850 35. McCune BT, Lanahan MR, tenOever BR, Pfeiffer JK. Rapid Dissemination and
851 Monopolization of Viral Populations in Mice Revealed Using a Panel of Barcoded Viruses. J Virol.
852 2020;94(2). Epub 20200106. doi: 10.1128/JVI.01590-19. PubMed PMID: 31666382; PubMed
853 Central PMCID: PMCPMC6955244.
854 36. Bacsik DJ, Dadonaite B, Butler A, Greaney AJ, Heaton NS, Bloom JD. Influenza virus
855 transcription and progeny production are poorly correlated in single cells. Elife. 2023;12. Epub
856 20230907. doi: 10.7554/eLife.86852. PubMed PMID: 37675839; PubMed Central PMCID:
857 PMCPMC10484525.
858 37. Heldt FS, Kupke SY, Dorl S, Reichl U, Frensing T. Single-cell analysis and stochastic
859 modelling unveil large cell-to-cell variability in influenza A virus infection. Nat Commun.
860 2015;6:8938. Epub 20151120. doi: 10.1038/ncomms9938. PubMed PMID: 26586423; PubMed
861 Central PMCID: PMCPMC4673863.
862 38. Jacobs NT, Onuoha NO, Antia A, Steel J, Antia R, Lowen AC. Incomplete influenza A virus
863 genomes occur frequently but are readily complemented during localized viral spread. Nat
864 Commun. 2019;10(1):3526. Epub 20190806. doi: 10.1038/s41467-019-11428-x. PubMed PMID:
865 31387995; PubMed Central PMCID: PMCPMC6684657.
866 39. Zath GK, Thomas MM, Loveday EK, Bikos DA, Sanche S, Ke R, et al. Influenza A viral burst
867 size from thousands of infected single cells using droplet quantitative PCR (dqPCR). PLoS
868 Pathog. 2024;20(7):e1012257. Epub 20240701. doi: 10.1371/journal.ppat.1012257. PubMed
869 PMID: 38950082; PubMed Central PMCID: PMCPMC11244780.
870 40. Sims A, Tornaletti LB, Jasim S, Pirillo C, Devlin R, Hirst JC, et al. Superinfection exclusion
871 creates spatially distinct influenza virus populations. PLoS Biol. 2023;21(2):e3001941. Epub
872 20230209. doi: 10.1371/journal.pbio.3001941. PubMed PMID: 36757937; PubMed Central
873 PMCID: PMCPMC9910727.
874 41. Delima GK, Ganti K, Holmes KE, Shartouny JR, Lowen AC. Influenza A virus coinfection
875 dynamics are shaped by distinct virus-virus interactions within and between cells. PLoS Pathog.
876 2023;19(3):e1010978. Epub 20230302. doi: 10.1371/journal.ppat.1010978. PubMed PMID:
877 36862762; PubMed Central PMCID: PMCPMC10013887.
878 42. Pearson JE, Krapivsky P, Perelson AS. Stochastic theory of early viral infection:
879 continuous versus burst production of virions. PLoS Comput Biol. 2011;7(2):e1001058. Epub
880 20110203. doi: 10.1371/journal.pcbi.1001058. PubMed PMID: 21304934; PubMed Central
881 PMCID: PMCPMC3033366.
882 43. Ramos I, Smith G, Ruf-Zamojski F, Martinez-Romero C, Fribourg M, Carbajal EA, et al.
883 Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells
884 Revealed Paracrine Induction of Interferon Lambda 1. J Virol. 2019;93(20). Epub 20190930. doi:
885 10.1128/JVI.00559-19. PubMed PMID: 31375585; PubMed Central PMCID: PMCPMC6798124.
886 44. Heaton NS, Langlois RA, Sachs D, Lim JK, Palese P, tenOever BR. Long-term survival of
887 influenza virus infected club cells drives immunopathology. J Exp Med. 2014;211(9):1707-14.

888 Epub 20140818. doi: 10.1084/jem.20140488. PubMed PMID: 25135297; PubMed Central
889 PMCID: PMCPMC4144728.
890 45. Fiege JK, Stone IA, Dumm RE, Waring BM, Fife BT, Agudo J, et al. Long-term surviving
891 influenza infected cells evade CD8+ T cell mediated clearance. PLoS Pathog.
892 2019;15(9):e1008077. Epub 20190926. doi: 10.1371/journal.ppat.1008077. PubMed PMID:
893 31557273; PubMed Central PMCID: PMCPMC6782110.
894 46. Ferreri LM, Seibert B, Caceres CJ, Patatanian K, Holmes KE, Gay LC, et al. Dispersal of
895 influenza virus populations within the respiratory tract shapes their evolutionary potential. Proc
896 Natl Acad Sci U S A. 2025;122(4):e2419985122. Epub 20250121. doi:
897 10.1073/pnas.2419985122. PubMed PMID: 39835898; PubMed Central PMCID:
898 PMCPMC11789087.
899 47. Ganti K, Bagga A, Carnaccini S, Ferreri LM, Geiger G, Joaquin Caceres C, et al. Influenza A
900 virus reassortment in mammals gives rise to genetically distinct within-host subpopulations. Nat
901 Commun. 2022;13(1):6846. Epub 20221111. doi: 10.1038/s41467-022-34611-z. PubMed PMID:
902 36369504; PubMed Central PMCID: PMCPMC9652339.
903 48. Amato KA, Haddock LA, 3rd, Braun KM, Meliopoulos V, Livingston B, Honce R, et al.
904 Influenza A virus undergoes compartmentalized replication in vivo dominated by stochastic
905 bottlenecks. Nat Commun. 2022;13(1):3416. Epub 20220614. doi: 10.1038/s41467-022-31147-
906 0. PubMed PMID: 35701424; PubMed Central PMCID: PMCPMC9197827.
907 49. Shartouny JR, Lee CY, Delima GK, Lowen AC. Beneficial effects of cellular coinfection
908 resolve inefficiency in influenza A virus transcription. PLoS Pathog. 2022;18(9):e1010865. Epub
909 20220919. doi: 10.1371/journal.ppat.1010865. PubMed PMID: 36121893; PubMed Central
910 PMCID: PMCPMC9521904.
911 50. Ferguson NM, Galvani AP, Bush RM. Ecological and immunological determinants of
912 influenza evolution. Nature. 2003;422(6930):428-33. doi: 10.1038/nature01509. PubMed PMID:
913 12660783.
914 51. Koelle K, Cobey S, Grenfell B, Pascual M. Epochal evolution shapes the phylodynamics of
915 interpandemic influenza A (H3N2) in humans. Science. 2006;314(5807):1898-903. doi:
916 10.1126/science.1132745. PubMed PMID: 17185596.
917 52. Fitch WM, Leiter JM, Li XQ, Palese P. Positive Darwinian evolution in human influenza A
918 viruses. Proc Natl Acad Sci U S A. 1991;88(10):4270-4. doi: 10.1073/pnas.88.10.4270. PubMed
919 PMID: 1840695; PubMed Central PMCID: PMCPMC51640.
920 53. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, et al. Vaccination
921 has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLoS Pathog.
922 2017;13(1):e1006194. Epub 20170131. doi: 10.1371/journal.ppat.1006194. PubMed PMID:
923 28141862; PubMed Central PMCID: PMCPMC5302840.
924 54. Valesano AL, Fitzsimmons WJ, McCrone JT, Petrie JG, Monto AS, Martin ET, et al.
925 Influenza B Viruses Exhibit Lower Within-Host Diversity than Influenza A Viruses in Human
926 Hosts. J Virol. 2020;94(5). Epub 20200214. doi: 10.1128/JVI.01710-19. PubMed PMID:
927 31801858; PubMed Central PMCID: PMCPMC7022338.
928 55. Valesano AL, Rumfelt KE, Dimcheff DE, Blair CN, Fitzsimmons WJ, Petrie JG, et al.
929 Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts.
930 PLoS Pathog. 2021;17(4):e1009499. Epub 20210407. doi: 10.1371/journal.ppat.1009499.
931 PubMed PMID: 33826681; PubMed Central PMCID: PMCPMC8055005.

932 56. Gu H, Quadeer AA, Krishnan P, Ng DYM, Chang LDJ, Liu GYZ, et al. Within-host genetic
933 diversity of SARS-CoV-2 lineages in unvaccinated and vaccinated individuals. Nat Commun.
934 2023;14(1):1793. Epub 20230331. doi: 10.1038/s41467-023-37468-y. PubMed PMID:
935 37002233; PubMed Central PMCID: PMCPMC10063955.
936 57. Braun KM, Moreno GK, Wagner C, Accola MA, Rehrauer WM, Baker DA, et al. Acute
937 SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission
938 bottlenecks. PLoS Pathog. 2021;17(8):e1009849. Epub 20210823. doi:
939 10.1371/journal.ppat.1009849. PubMed PMID: 34424945; PubMed Central PMCID:
940 PMCPMC8412271.
941 58. Kistler KE, Huddleston J, Bedford T. Rapid and parallel adaptive mutations in spike S1
942 drive clade success in SARS-CoV-2. Cell Host Microbe. 2022;30(4):545-55 e4. Epub 20220322.
943 doi: 10.1016/j.chom.2022.03.018. PubMed PMID: 35364015; PubMed Central PMCID:
944 PMCPMC8938189.
945 59. Chen H, Ye J, Xu K, Angel M, Shao H, Ferrero A, et al. Partial and full PCR-based reverse
946 genetics strategy for influenza viruses. PLoS One. 2012;7(9):e46378. Epub 20120928. doi:
947 10.1371/journal.pone.0046378. PubMed PMID: 23029501; PubMed Central PMCID:
948 PMCPMC3460856.
949 60. Zhou B, Wentworth DE. Influenza A virus molecular virology techniques. Methods Mol
950 Biol. 2012;865:175-92. doi: 10.1007/978-1-61779-621-0_11. PubMed PMID: 22528160.
951 61. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-
952 resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3. Epub
953 20160523. doi: 10.1038/nmeth.3869. PubMed PMID: 27214047; PubMed Central PMCID:
954 PMCPMC4927377.
955 62. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12(4):656-64. doi:
956 10.1101/gr.229202. PubMed PMID: 11932250; PubMed Central PMCID: PMCPMC187518.
957 63. Chung NC, Miasojedow B, Startek M, Gambin A. Jaccard/Tanimoto similarity test and
958 estimation methods for biological presence-absence data. BMC Bioinformatics. 2019;20(Suppl
959 15):644. Epub 20191224. doi: 10.1186/s12859-019-3118-5. PubMed PMID: 31874610; PubMed
960 Central PMCID: PMCPMC6929325.
961 64. J. Roger Bray JTC. An Ordination of the Upland Forest Communities of Southern
962 Wisconsin. Ecological Monographs. 1957;27(4):325-49. doi: 10.2307/1942268.
963