Title

Long title: Weak selection and stochastic processes limit the emergence of antigenic variants during household transmission of influenza A viruses

Short title: Limited antigenic evolution of influenza during human transmission

Authors

Hunter J. Ries[1], Joseph Lalli[1], Kelsey R. Florek[2], Shari Barlow[3], Maureen Goss[3], Richard Griesser[2], Tonya Danz[2], Amra Uzicanin[4], Jonathan Temte[3], Thomas C. Friedrich[1,5,6*]

Affiliations

[1] Department of Pathobiological Sciences, University of Wisconsin–Madison, Madison, Wisconsin, USA

[2] Wisconsin State Laboratory of Hygiene, Madison, Wisconsin, USA

[3] Department of Family Medicine and Community Health, University of Wisconsin–Madison, Madison, Wisconsin, USA

[4] National Center for Emerging and Zoonotic Infectious Diseases, US Centers for Disease Control and Prevention, Atlanta, Georgia, USA (retired)

[5] Wisconsin National Primate Research Center, University of Wisconsin–Madison, Madison, Wisconsin, USA

[6*] Correspondence should be addressed to T.C.F. (email: tfriedri@wisc.edu).

Keywords: Influenza A virus; Antigenic drift; Hemagglutinin antigenic sites; Transmission bottleneck; Intra-host single-nucleotide variants; Purifying selection; Household transmission; Whole-genome sequencing; A(H1N1)pdm09; A(H3N2).

Disclaimer: The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention.

Funding: This study was funded by cooperative agreement 5U01CK000542-02-00 from the Centers for Disease Control and Prevention. H.J.R. was supported by the NIH NIGMS T32 training grant to the Cellular and Molecular Pathology Training Program (5T32GM135119) and the Parasitology and Vector Biology Training Program (5T32AI007414). The funders had no role in the study design, data analysis, data interpretation, and this report's writing. The publication's contents are solely the authors' responsibility and do not necessarily represent the official view of the CDC or NIH. All authors had full access to the data in the study and accepted the responsibility to submit it for publication.

Author contributions: H.J.R. contributed to data curation, methodology, formal analysis, project administration, bioinformatics software, data visualization, writing—draft preparation, and draft review and editing. J.L. contributed to data curation, methodology, bioinformatic software, and writing—draft preparation. K.F. contributed to data curation and methodology. S.B. contributed to data curation and methodology. M.G. contributed to data curation and methodology. R.G. contributed to data curation and methodology. T.D. contributed to data curation and methodology. K.R. contributed to data curation and methodology. A.U.

contributed to conceptualization, data curation, methodology, patient recruitment, and project administration. J.T. contributed to conceptualization, data curation, methodology, patient recruitment, and project administration. T.C.F. contributed to conceptualization, data curation, methodology, project administration, writing—draft review and editing, and supervision.

Competing interests: The authors declare no competing financial interests.

Additional information: Correspondence and requests for materials should be emailed to T.C.F.

Abstract

Influenza viruses undergo antigenic drift, the gradual accumulation of mutations that cause antigenic changes in the viral surface proteins hemagglutinin (HA) and neuraminidase (NA). Although selection for antigenic variants is detectable on the global scale, the processes by which antigenic variants are generated and selected in individual hosts remain unclear. It has been hypothesized that selection for antigenic variants may occur during the establishment of a new infection, rather than over time in a single host. Here, we leveraged a large household cohort study to assess whether selection was detectable between acutely infected hosts. We investigated influenza A virus evolution using specimens from 384 children and household contacts with RT-PCR-confirmed influenza A infection, representing infections with A(H1N1)pdm09 and A(H3N2) viruses from 2017–19. In agreement with prior studies, we found that acute infections involved weak purifying selection across the viral genome. In addition, we identified 40 transmission events occurring in 31 households. During transmission, evolution between hosts was characterized by tight transmission bottlenecks and weak purifying selection. We found variability in the strength and direction of selection on antigenic regions of HA, but no clear evidence for selection of antigenic variants during transmission. Together, our results indicate that stochastic processes and weak natural selection dominate most acute influenza A virus infections and transmission events, and that selection of antigenic variants during transmission between acutely infected hosts is likely to be exceedingly rare.

Author Summary

Influenza viruses clearly evolve under selective pressure from immune responses in human populations, but recent work suggests that within individual infections random effects are stronger than selection. New viral variants that spread globally must nonetheless emerge in one person and be transmitted onwards—how does this happen? We characterized viral genomes collected over two influenza seasons from 384 children and their household contacts. We detected 40 transmissions among 31 of the households, allowing us to examine how selection acts during infection and transmission. We found that influenza virus genetic diversity is low in infected individuals, and mutations arising in one person are rarely transmitted to their household contacts, consistent with prior reports that influenza virus evolution is tightly constrained within hosts. We further examined all transmission events for evidence of selection between hosts, finding only one mutation that could plausibly affect antibody recognition. However, we found no evidence that this mutation was favored by natural selection. Our results suggest that chance events, together with weak selection, are the main forces affecting influenza virus evolution within and between hosts during typical acute infections. Selection for new variants may be more likely to occur over longer transmission chains and/or during prolonged infections.

Introduction

Like other rapidly evolving RNA viruses, influenza A virus causes significant global morbidity and mortality [1]. Influenza vaccine effectiveness is often reduced by antigenic drift, the emergence and preferential propagation of antibody-escape mutations in the antigenic regions of the hemagglutinin (HA) and neuraminidase (NA) proteins. Influenza vaccine antigens must be frequently updated to reflect these changes in the HA protein of circulating influenza strains. The National Institute for Allergy and Infectious Diseases has thus identified "accurately predicting how influenza viruses will evolve" as critical to developing an effective universal influenza vaccine [2]. Although influenza A virus evolution is detectable on the global scale, it remains unclear how new antigenic variants arise in one host and are propagated through transmission chains and among populations. An improved understanding of these processes will enhance our ability to forecast trends in viral evolution.

Variant viruses encoding one or more amino acid substitutions quickly arise within hosts after infection [3–7]. These intra-host single-nucleotide variants (iSNVs) cause nonsynonymous (amino-acid changing) or synonymous ("silent") mutations, which may impact pathogenesis [8]. Purifying selection purges deleterious mutations from viral populations, while diversifying selection favors changes away from the within-host consensus sequence. Globally, seasonal influenza viruses show an excess of synonymous variants throughout most of the genome, suggesting that most nonsynonymous mutations are deleterious and are quickly purged from circulation [8]. Classically defined antigenic regions of HA are the exception and are characterized by an excess of nonsynonymous substitutions. This excess suggests that HA sequences are subject to positive selection at a global scale, which is reflected in the “ladder-like” HA phylogeny and the continual emergence of new HA clades [8]. However, the action of natural selection appears to be relatively weak within individual hosts, even in HA antigenic

sites [8]. Studies by our group [3,4] and others [7–14] have thus far failed to detect evidence of selection for antigenic variants during acute infections, where even potential immune-escape mutations remain at low frequency. Interestingly, influenza viruses appear capable of accumulating antigenic changes during chronic infections in immunocompromised individuals [8,15,16], suggesting that acute infection may not allow enough time for antigenic variants to emerge [14].

Every positively selected variant must originate from a mutation in a single host [17,18]. What evolutionary processes allow some mutations to "escape" from the forces of randomness and weak selection in individual hosts to become dominant in the global population? Recently, Morris *et al.* [19] suggested that immune selection is strongest at the time a new infection is being established at a respiratory site in the presence of pre-existing virus-specific antibodies. In this model, incoming viruses are subject to "sieving" by pre-existing antibodies in the recipient, such that viruses capable of evading antibody detection are more likely to successfully found a new infection. Antigenic variants that are present at low frequency in a donor may therefore become fixed in a new host during a selective transmission bottleneck. In contrast, because antibodies specific for the incoming virus become detectable only near the end of a typical acute infection, there is a temporal disconnect between the peak of virus replication and the appearance of either recall or de-novo antibody responses capable of selecting for new antigenic variants. Transmission of these variants is further impacted by the transmission bottleneck, which involves a sharp reduction in viral diversity—according to most estimates, only 1–2 unique viral genomes from the donor establish infection in the recipient [7,20]. In their model, Morris *et al.* suggest that minor iSNVs under positive selection could reach 100% within-host frequency within the first 24 hours of infection [19]. In support of this, we have previously shown in a ferret model system that influenza A virus variants present at

less than 10% frequency in the donor can survive the transmission bottleneck and reach fixation in a new host [21].

Detecting potential selection on viral genomes during the establishment of a new infection requires characterizing virus populations in donor-recipient pairs and cannot be done through studies of individual infections alone. Our objective was to determine whether immune selection for antigenic variants is detectable during acute, person-to-person transmission of influenza A virus. We hypothesized that if antibody-mediated sieving acts at transmission, it might be possible to detect events in which low-frequency antigenic variants in donors would become enriched or fixed in recipients. Here, we sequenced seasonal influenza A viruses from 384 influenza-positive individuals enrolled in a school-associated, community-based influenza surveillance study, the Oregon Child Absenteeism Due to Respiratory Disease Study (ORCHARDS; Oregon, Wisconsin, USA) [22,23]. We then quantified within-host diversity and selection, linked donors and recipients using epidemiologic and genetic criteria, and tested whether variants—particularly in HA antigenic sites—were preferentially transmitted to recipients.

Results

Study population and viral genome sequencing

For this study, we used specimens collected from participants in the Oregon Child Absenteeism Due to Respiratory Disease Study (ORCHARDS) study. Full methodology was reported previously [22,23], and the protocol was approved by the University of Wisconsin–Madison Health Sciences Institutional Review Board, with written consent from adults and parental consent plus assent for minors. Briefly, the study enrolled school-aged children from

the Oregon School District (Oregon, Wisconsin, USA) who reported at least two acute respiratory symptoms (e.g., cough, sore throat, nasal congestion, runny nose, sneezing, or fever). Study staff performed home visits, collected specimens from the children, and offered to enroll other household members in a sub-study assessing within-household influenza transmission. Study staff then collected nasal swabs from any other participating household members. Respiratory specimens that tested RT-PCR-positive for influenza A virus were used for the work described here. In total, we sequenced 384 samples from the 2017–18 and 2018–19 influenza seasons, generating 283 complete viral genomes that passed our quality-control thresholds (see Methods) [**Table 1**]. The average Ct value for all samples included in our study was 27.2, with a standard deviation of 3.9, and a range of 18.52–35.37.

We first used deep sequencing to assess viral diversity within all samples. We mapped sequence reads using each season's A(H1N1)pdm09 or A(H3N2) vaccine strain as a reference and identified intra-host single-nucleotide variants (iSNVs) in each sample using conservative frequency thresholds (3%) and quality control criteria (see Methods). Then, we aligned all high-quality participant consensus sequences for each season and subtype and used the majority nucleotide at each position in this alignment to define a "within-season" reference sequence. Raw sequencing reads were then re-processed and aligned against the within-season reference sequence to contextualize the variants relative to the strains in local circulation at the time of sampling. We included the major genes in the 6 largest viral gene segments in our alignment and variant-calling pipeline. We excluded M1/M2 and NS1/NEP genes because their overlapping reading frames can confound site-based diversity metrics [8]. Some samples contained very high numbers of iSNVs, which could reflect truly high levels of within-host diversity, but may also indicate co-infection with distinct strains or possibly method errors. To remove such outliers, as in [8], we quantified the number of iSNVs (>0.5% allele frequency) for

each sample, ranked them in descending order within their season-subtype group, and then excluded samples that were in the top 10%.

Low-frequency variation and purifying selection dominate within-host viral evolution

To assess within-host diversity in our cohort, we plotted minority iSNV frequencies (3–50%) for each gene and annotated them by mutation type: nonsynonymous (amino-acid changing), synonymous ("silent"), or stop gained (**Fig 1A**). We detected nonsynonymous and synonymous iSNVs across the genome, with a large proportion (77.3%) of iSNVs detected below 10% allele frequency in all samples. The pattern of mostly low-frequency iSNVs found throughout the viral genome was consistent across influenza seasons and viral subtypes (**S1 Fig**). We then quantified the total number of iSNVs per sample: 68 samples contained no detectable iSNVs, and the median count was 4 (**S1 Fig**). We found no statistically significant difference in iSNV counts between individuals reporting recent influenza vaccination and those reporting no recent influenza vaccination (**S1B Fig**). Notably, 2 of 283 samples were clear outliers, each with more than 20 iSNVs. One sample (Ct 28.67) from the 2017–18 A(H3N2) season contained 28 iSNVs (3–12.5%) across HA, NA, NP, PA, PB1, and PB2 (**S1 Table**). Notable among these was an iSNV in HA causing an M168I change in antigenic site A at 6.73% (H3 numbering; see Methods). A sample from the 2018–19 A(H3N2) season (Ct 29.66) contained 62 iSNVs (3–5.6%) across HA, PA, PB1, and PB2 (**S1 Table**), of which 54 were present in HA (26 non-antigenic, 28 antigenic). Fifteen of the 28 antigenic iSNVs in HA were nonsynonymous, including antigenic site B mutations S159Y and K160T. Of the 1226 iSNVs detected in our 283 samples, we identified just 64 iSNVs in classical HA antigenic sites (40 nonsynonymous, 24 synonymous) in

31 samples. Taken together, our data support the current consensus that most within-host variants detected during acute influenza virus infections are present at low frequencies, and most samples contain only a few iSNVs—although occasional outliers exhibit substantially greater counts.

We next asked which evolutionary force(s) were most likely to be shaping viral evolution in these acutely infected individuals, as we have done previously [24,25]. Most minority iSNVs are detected below 10% frequency, as shown for A(H3N2) in the 2017–18 (**Fig 1B**) and 2018–19 (**Fig 1C**) seasons. Indeed, iSNVs detected below 10% frequency accounted for a greater proportion of all iSNVs than would be expected under neutral evolution (genetic drift), while, as expected, iSNVs that created stop codons were detected at very low frequencies (15 iSNVs at 3.0–7.5% and one iSNV in PB1 at 13.3%). Such stop-gained iSNVs were detected in PB2, PB1, NP, PA, and NA, but not HA. We next counted the cumulative number of iSNVs for all samples from 3–100% to assess the distribution of iSNVs across various allele frequency bins. Cumulative iSNV counts similarly show that most mutations are present at either very low or very high frequencies within hosts (**S1 Fig**). Together, these results suggest that influenza virus diversity is limited within acutely infected hosts and characterized by weak purifying selection.

Most iSNVs within hosts are synonymous

To quantify the synonymous and nonsynonymous nucleotide diversity in viral populations within hosts, we calculated π, the mean pairwise differences per nucleotide site for a set of sequences [26]. We use π here because its value is less sensitive to sequencing read depth than other common diversity metrics [27]. Our analysis revealed that synonymous diversity (π_S) was significantly greater than nonsynonymous diversity (π_N) in every viral gene, both when we considered all seasons and subtypes together (**Fig 2A**; $p < 0.001$, Wilcoxon signed-rank tests)

and when we analyzed each influenza season and subtype independently (**S2A–D Fig**; $p < 0.015$). We next sought to quantify the magnitude and directionality of selection occurring within hosts by calculating the difference between π_N and π_S for each gene. The distribution of $\pi_N - \pi_S$ values was slightly, but significantly, less than zero in all viral genes, both when we considered all data together across season and subtype (**Fig 2B**; $p < 0.001$, one-sample Wilcoxon signed-rank tests) and also when we analyzed each season and subtype individually (**S2E–H Fig**; $p < 0.015$). Together, these findings suggest that influenza A viruses are generally subjected to purifying selection within acutely infected hosts across influenza seasons, subtypes, and host vaccination status. This is also consistent with the finding that the preponderance of iSNVs are present below 10% frequency (**Fig 1**).

Calculating nucleotide diversity for an entire gene could mask selection directed at a small number of specific sites. We focused on HA to address the hypothesis that pre-existing antibody responses may exert selective pressure on viruses that could be detected as elevated $\pi_N - \pi_S$ values in certain antigenic regions. We calculated π_N and π_S (**Fig 2C**), as well as the difference between π_N and π_S (**Fig 2D**), using 30-codon sliding windows in HA for each sample. Our analysis of antigenic and non-antigenic regions (defined in H3 as antigenic regions A-E [28,29] and in H1 as sites Ca, Cb, Sa, Sb [30]) revealed that, while values for both π_N and π_S fluctuate along the length of the HA gene, π_S exhibits larger fluctuations. Local spikes in π_S lead to larger (i.e., more negative) $\pi_N - \pi_S$ differences in locations that include both classical antigenic sites and non-antigenic regions. This trend is present, with somewhat different patterns, in each influenza season and subtype (**S3 Fig**). Together, these findings show that, although HA genes are under overall purifying selection, the magnitude of purifying selection may vary over the length of HA as well as with season and viral subtype.

Tight bottlenecks and weak purifying selection shape between-host influenza virus evolution

Our study includes 36 households comprising 46 putative transmission events, which enables us to compare differences in virus population diversity between donors and recipients and test for evidence of selection during transmission. The first individual to report symptoms was identified as the donor case for each household in our dataset. Every other household member who had RT-PCR-confirmed influenza infection within ten days of the donor case was identified as a potential recipient case. Prior work has noted that household members might coincidentally acquire independent infections from outside the household [31–33], potentially confounding analyses of transmission events. To address this, we implemented a genetic criterion for household pairing, as described previously [24]. Briefly, we calculated the proportion of all iSNVs shared between each donor-recipient pair and excluded pairs with shared proportions below the 95th percentile of those observed in random community pairs (**Fig 3A**). The random community pairing followed the expected distribution: 95% of random community pairs shared fewer than 11.2% of variants (vertical dashed line). That is, 40 of the 46 potential household pairs shared a higher proportion of variants than 95% of randomly paired community samples. As a complementary approach to determine the relatedness of viruses infecting members of the same household, we generated maximum likelihood phylogenies for the HA consensus sequences of each influenza season and subtype (**S4 Fig**). Viruses from members of the same household clustered closely together on all trees, with identical or near-identical consensus sequences. Together, our epidemiological and genetic criteria enable us to rigorously identify cases of within-household transmission.

We next determined how transmission affected iSNV frequencies in the donor-recipient pairs. If transmission bottlenecks are loose, we expect iSNV frequencies in the recipient to resemble those in the donor. In total, we detected 1330 iSNVs in either donors or recipients. We categorized these iSNVs into three groups: present in both donor and recipient, present only in the donor, or present only in the recipient. The majority (72%) of iSNVs were detected in both the donor and recipient; of these, 99% exceeded 50% frequency (consensus level) and 95% were present above 90% frequency in both the donor and recipient. In contrast, only eight iSNVs (five synonymous, three nonsynonymous) below 90% were observed in both the donors and recipients at similar frequencies. These results indicate that highly abundant donor iSNVs compose most of the diversity shared between donors and recipients.

In one household pair, a synonymous substitution in PB1 codon 346 (PB1: N346N) was observed at 57% frequency in the donor and 76% in the recipient (**Fig 3B**, **S2 Table**). In a separate household pair, we detected a synonymous PB1 substitution (PB1: P627P) at 10% frequency in the donor and 29% in the recipient (**S2 Table**). In a third household cluster, a nonsynonymous PB2 substitution (PB2: A270S) was present at 60% frequency in the donor and 0%, 6%, and 15% in the three recipients (**S2 Table**). In a fourth household cluster, three intermediate-frequency iSNVs were potentially transmitted from the donor to three recipients (**S2 Table**). The donor sample harbored two NP iSNVs: T22A at 24% and G187G at 10%. Both mutations were lost in Recipient 1, while only T22A persisted at 30% in Recipient 2, and only G187G was maintained at 15% in Recipient 3. Similarly, in the same household, a synonymous HA mutation (HA: C471C, H1 numbering) was found at 32% in the donor, but was present at 100%, 71%, and 14% in the three recipients. Conversely, a nearby synonymous HA mutation (HA: G463G, H3 numbering) was present at 14% in the donor but was undetected in all recipients. The fact that intermediate-frequency donor iSNVs are lost, maintained, and/or fixed

in different patterns in different recipients of the same donor is consistent with the idea that transmission of seasonal influenza A viruses from acutely infected humans typically involves a very tight, stochastic bottleneck in agreement with previous reports [7,20].

Twelve percent of iSNVs were only present in the donor, with 91% of these detected below 50% frequency. Meanwhile, 16% of iSNVs were detected only in the recipient, and 89% of these recipient-only iSNVs were minority variants. We cannot determine whether such iSNVs were present in the donor below our detection limits or whether they emerged *de novo* in the recipient. Notably, 7% of the recipient-only iSNVs (representing 1% of all iSNVs in donors and recipients) were fixed in the recipient despite being undetected in the donor. Of these fixed recipient-only iSNVs, 4 were synonymous and 7 were nonsynonymous. Together, our results suggest that transmission of seasonal influenza A viruses generally involves very tight bottlenecks in which the likelihood that an iSNV is transmitted depends largely on its frequency in the donor.

While substantial within-host diversity is lost during transmission, the extent to which selection shapes this reduction remains unclear. We therefore next compared π_N and π_S between donors and recipients to assess whether transmission bottlenecks preferentially restrict nonsynonymous mutations. Overall, π_N and π_S values did not change significantly during transmission in any viral genes (**Fig 3CD**), except in PB2, which showed slightly but significantly higher π_N values in recipients relative to donors ($p = 0.025$). This significant increase was driven by higher π_N values specifically in the 2017–18 A(H3N2) group ($p = 0.003$; **S5A Fig**). In the 2017–18 A(H3N2) group, we also observed several additional genes in which recipients had slightly but significantly higher π_N values compared to donors (PA and HA: $p < 0.05$), along with a modest increase in π_S values for PB2 and NA ($p < 0.05$). These patterns were not observed in other groups (**S5 Fig**, **S6 Fig**). Across all genes when aggregating all

seasons and subtypes in this study, the distribution of $\pi_N - \pi_S$ values was significantly below zero ($p < 0.05$), further suggesting that purifying selection was the dominant selective force in donors and recipients (**Fig 3E**). Statistical comparisons between donors and recipients in **Fig 3E** yielded no significant differences (**S3 Table**). Taken together, our results indicate that, on balance, purifying selection appears to be equally strong in both donors and recipients.

Antigenic sites on hemagglutinin are under distinct selective pressures during transmission

The model of influenza virus antigenic evolution proposed by Morris *et al.* [19] posits that selection operates mainly at the time when a new infection is established. Because this is expected to be a rare event, detecting it would require access to many well-defined transmission pairs. To ascertain whether this phenomenon was detected in our donor-recipient pairs (**Fig 3B**), we investigated whether specific nonsynonymous mutations in antigenic regions of HA were acquired or increased in frequency during transmission. We detected the transmission of a nonsynonymous antigenic site variant in just one transmission pair, in which both the donor and recipient were vaccinated. A nonsynonymous mutation encoding HA I185T (H1 numbering, antigenic site Sb of 2018–19 A(H1N1)pdm09) became fixed across transmission (from 38% in the donor to 100% in the recipient; **S2 Table**). This event is consistent with selection during transmission; however, it could also have become fixed because of the tight transmission bottleneck and not because it conferred a fitness benefit. Four other transmission pairs maintained HA I185T at 100% frequency, and this mutation was present in 19 other samples from our cohort at 100% frequency, suggesting that it was circulating in the community during the study period. The four donor samples contained an

additional 23 mutations in HA, which were fixed relative to the within-season reference, including 2 synonymous mutations in antigenic region Ca1 (HA: E235E and HA: G237G). All of these mutations were present in the recipients, indicating that viruses infecting these donor-recipient pairs were divergent relative to most viruses sampled from the community during this time period (**S4D Fig**, taxon boxed in red). In all other donor-recipient pairs, we saw no stark increase in the frequency of iSNVs in classical antigenic sites, consistent with the expectation that this phenomenon is infrequent.

Although there was only one instance of an iSNV with potential antigenic impact increasing in frequency during transmission, we reasoned that patterns of nucleotide diversity could indicate whether selection acted specifically on antigenic sites of HA. To test this, we quantified the difference between π_N and π_S values for each codon in antigenic and non-antigenic regions (defined in H3 as antigenic regions A-E [28,29] and in H1 as sites Ca, Cb, Sa, Sb [30,34]) in each donor and recipient sample, then we normalized per-sample differences by the number of sites in each region. Across donors, recipients, and the pooled “all” group (i.e., all samples in the dataset considered together), antigenic and non-antigenic regions displayed significantly negative per-site $\pi_N - \pi_S$ values ($p < 0.0231$), consistent with purifying selection (**Fig 4**). These values remained stable during transmission, with no significant differences between donors and recipients for either region—a pattern that held across all seasons and subtypes analyzed in this study (**S7 Fig**). We detected a significant difference between antigenic and non-antigenic regions only in the "all" group ($p = 3.69e-11$), where non-antigenic regions showed stronger purifying selection. In this group, the median per-site $\pi_N - \pi_S$ value for antigenic regions was zero, with 32% of values exactly zero and 40% falling below zero. These results indicate that purifying selection is generally weaker in antigenic regions of HA than in non-antigenic regions. Still, it remains unclear whether this reflects reduced constraint,

increased selective pressure, or both. If distinct selection pressures do act on antigenic sites during transmission, they appear too weak or rare to detect with this approach.

Discussion

On the scale of the human population, influenza viruses clearly evolve under immune selection pressure. However, we are confronted by a paradox of scale: new antigenic variants must arise within individual infected hosts, but viral evolution within and between infected hosts appears to be dominated by randomness and purifying selection, making it difficult for new variants to emerge and transmit. Defining the evolutionary mechanisms connecting the individual to the global scale will enhance our ability to understand and forecast trends in influenza virus antigenic evolution [8]. Morris *et al.* predict that selection occurs not over time within hosts, but during the establishment of a new infection [19]. Here, we used 283 high-quality influenza A-positive samples from a community cohort study. Of these, 71 samples formed 40 validated donor–recipient pairs across 31 households for transmission analyses, and the remaining 212 single infections informed within-host analyses. We used these pairs to assess whether there was evidence for selection during transmission. Our study supports previous findings suggesting that the short duration of acute influenza virus infection, coupled with purifying selection and anatomical compartmentalization, dampens the generation of substantial diversity within hosts [6–8,17,27,35–37]. We also find that narrow transmission bottlenecks between hosts result in a substantial loss in low-frequency diversity and dampen the ability of selection to act efficiently during transmission.

Mutations arise during acute influenza virus infection and are culled or amplified by purifying or diversifying selective pressures, respectively. Consistent with previous studies, we observe low-frequency iSNVs across the viral genome (**Fig 1**) and signals of purifying selection

on every influenza virus gene (**Fig 2B**) [7]. To become dominant in human populations, variants must survive transmission bottlenecks and propagate in additional hosts. If transmission bottlenecks were extremely loose (allowing many viruses to be transferred from donor to recipient) and under neutral selection, we would expect iSNVs to be present in the donor and recipient at similar frequencies. Conversely, under a tight bottleneck, low-frequency iSNVs are rarely transmitted, but if they are, the limited number of unique founding genomes often causes these variants to be present at high frequency in the recipient. In the 31 households and 40 transmission events we analyzed, low-frequency donor iSNVs were generally lost, and the vast majority of transmitted iSNVs were ≥90% frequency in the recipient. This pattern is consistent with a narrow, stochastic bottleneck, as others have reported previously for acute transmission of influenza A and SARS-CoV-2 [7,24,25,37,38].

Importantly, selection for specific variants during transmission may also result in reduced viral genetic diversity in recipients in a manner that may be difficult to distinguish from a stochastic bottleneck [39]. We have observed such selective bottlenecks ourselves in experimental systems of influenza A virus transmission in ferrets [4,21]. However, in the present work, our data are mostly consistent with the stochastic model: we see dramatic shifts in iSNV frequency, in which iSNVs are more likely to transmit if they are at high frequency in the donor. There is no single iSNV that is consistently transmitted. In several households, we capture transmission from a single donor to multiple recipients, and in these cases, we see that each recipient acquired a different combination of iSNVs. This suggests that the iSNVs are present in different viruses (i.e., not genetically or physically linked) and that each recipient received a random sample of viruses present in their donor.

We expect selective bottlenecks to be rare in acute human influenza transmission, but it is interesting that we did observe one nonsynonymous mutation in an HA antigenic region, HA: I185T, which is in antigenic site Sb. This mutation was a minority variant in the donor and became fixed in the recipient. Fixation of a potential antigenic variant during transmission could be consistent with the Morris *et al.* hypothesis [19], but it is also important to consider the broader context of contemporaneous viral diversity: A(H1N1)pdm09 viruses circulating both globally and in the sampled community were polymorphic at HA:185 site during the study period. Threonine (T185) had been dominant at this position, but isoleucine (I185) was becoming more frequent in the 2018–19 season, present in 51% of our samples and 66% of global sequences (**S8 Fig**) [40,41]. The donor in our transmission pair carried 185T at 38% frequency and 185I at 62%, yet the recipient became productively infected with 185T alone, the minority variant of the donor and the globe at the time. Thus, given the cocirculation of both T185 and I185 in our community and our lack of information about immune responses to influenza A viruses in our cohort, we cannot determine whether transmission of this mutation was favored by natural selection in this donor-recipient pair.

Our findings are consistent with previous studies suggesting that tight transmission bottlenecks reduce the ability of selection to fix beneficial iSNVs during transmission [7,20,21,37,38]. When selection is weak or absent, the probability that an iSNV is transmitted will be a function of its frequency rather than fitness. To test whether selection might still act preferentially on certain regions, we compared levels of purifying selection across antigenic and non-antigenic sites of HA. As a whole, both antigenic and non-antigenic regions of HA were under purifying selection, and our analysis detected no significant difference in purifying selection on antigenic or non-antigenic regions between donors and recipients (**Fig 4**). Non-antigenic regions showed slightly stronger purifying selection than antigenic regions, but this

difference was not significant across most comparisons. Overall, within hosts, antigenic regions of HA appear to be under very weak purifying selection, while non-antigenic regions experience slightly stronger, yet still weak, purifying selection. This pattern was not clearly observed during transmission. While we did not observe any striking differences between the antigenic evolution of A(H1N1)pdm09 and A(H3N2), we are limited by the relatively small number of annotated antigenic sites for these viruses (50 and 130, respectively). Our results further support the idea that stochastic processes dominate influenza A virus evolutionary dynamics during transmission and that selective transmission leading to the global emergence of HA antigenic variants is an extremely rare, high-consequence event that will be difficult to observe in nature.

Our study has several limitations. We were unable to perform deep sequencing in technical replicates for this work, which could reduce the accuracy of iSNV frequencies, particularly near the 3% frequency threshold. To account for this deficit, we implemented stringent quality control criteria to eliminate the inclusion of samples of insufficient quality to detect low-frequency variants. Further, we only included samples with high whole-genome coverage and depth to facilitate comparison between genes. Our findings are also partially limited in that we have excluded M1/M2 and NS1/NEP genes from diversity analyses due to their overlapping reading frames and small size, which can confound site-based diversity metrics [8]. Our study design has additional limitations in both demographic and sampling structure. By nature of the study design, school-aged children reporting respiratory symptoms are the first household members enrolled in each household, and as a result, are likely to be recognized as donors in our transmission study. Indeed, all donors in our cohort's transmission networks were children. This raises a potential bias, as children may have had more recent or limited exposure to influenza viruses compared to adult household members, potentially

affecting their within-host diversity [31,42,43]. The study design also means that our cohort may be biased toward symptomatic infections in the donors. Importantly, we must caution that we do not know precisely when transmission occurred between putative donors and recipients. As such, donor samples may be derived later in the donor's acute infection than recipient samples. Additionally, misclassification of donor-recipient direction remains possible, particularly when household members have shared community exposure or separate introductions of genetically similar viruses. The key limitation of our study is our relatively few validated transmission events. For future studies to rigorously evaluate the possibility that selection for antigenic variants occurs during transmission, larger cohorts will be required that focus on enrolling households or people in other settings where transmissions are likely to be captured. Future studies should employ similar prospective cohort study designs and analyze deep-sequencing efforts alongside immunological and epidemiological data to expand upon our findings.

Materials and Methods

ORCHARDS participant recruitment and sampling

The full Oregon Child Absenteeism Due to Respiratory Disease Study (ORCHARDS; Oregon, Wisconsin, USA) methodology can be found in Temte *et al.* 2022 [23]. Briefly, ORCHARDS was a community cohort study designed to evaluate usability of cause-specific student absenteeism as a method for early detection of increased influenza activity in schools and surrounding community; ORCHARDS also served as a platform to study the characteristics and impact of different respiratory pathogens affecting school-aged children. A subset of participants in this cohort was recruited into a sub-study to assess within-household

transmission of respiratory pathogens. Household members were responsible for collecting specimens on Day 0 (within 24 hours of the home visit by study coordinators) and Day 7 (seven days after the initial collection). Each household participant collected the specimen without staff observation on the day of the home visit (Day 0) and again seven days later (Day 7). Two participants deviated from this protocol: one collected their first sample on Day 1 and another collected their second sample on Day 8. Both samples were retained for analysis (**S9 Fig**). In total, this study provided us with access to specimens from 384 children that tested positive for influenza A virus by RT-PCR (IVD CDC Human Influenza Virus RT-PCR Diagnostic Panel [44]) using the Applied Biosystems™ 7500 Real-Time PCR instrument (Thermo Fisher) at the Wisconsin State Laboratory of Hygiene (WSLH).

Illumina library preparation and sequencing

Following the manufacturer's protocol, influenza virus RNA was extracted from nasal swabs using the QIAcube HT (Qiagen) with the QIAamp 96 Virus QIAcube HT Kit (Qiagen). RT-PCR was then performed on extracted viral RNA using the SuperScript™ III One-Step RT-PCR System with Platinum™ Taq High Fidelity DNA Polymerase (Thermo Fisher) using the primers described by Zhou *et al.* [45], which amplify all of the eight influenza A virus gene segments in a single RT-PCR reaction, with the following thermal cycling conditions: 42˚C for 50 min, 50˚C for 10 min, 94˚C for 2 min, four cycles of the following: [94˚C for 30 sec, 43˚C for 30 sec, 68˚C for 3 min 50 sec], 30 cycles of the following: [94˚C for 30 sec, 57˚C for 30 sec, 68˚C for 3 min 30 sec increasing by 10 sec every cycle], then finally 68˚C for 10 min. The resulting PCR amplicons are then treated with Exonuclease I (Thermo Fisher) and incubated at 37˚C for 15 min and 80˚C for

15 min. The DNA is then quantified using Victor X2 (Perkin Elmer) using the Quant-iT High-Sensitivity dsDNA kit (Thermo Fisher) and normalized to 0.2 ng/μL. The DNA library is prepared for sequencing using the Nextera XT DNA Library Prep Kit (Illumina) following the manufacturer's protocol and sequenced using the MiSeq platform (Illumina).

iSNV calling pipeline

Intra-host single-nucleotide variants (iSNVs) were called using bespoke scripts through the Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison. All read processing scripts were executed through the rieshunter/tcflab:v2.04 Docker image (https://hub.docker.com/r/rieshunter/tcflab/tags) on a CHTC server node. This Docker image consists of a Linux GNU OS (18.04.6 LTS "Bionic Beaver") with all necessary commands and program files (Docker version 20.10.13).

Raw, paired FASTQ files from samples positive for the subtype of interest were adapter-trimmed and quality-trimmed to an average 5-base sliding-window quality of Q20 using Trimmomatic (version 0.39). Additionally, all reads less than 100 bases in length were discarded. Trimmed, paired reads were then merged using BBMerge.sh (BBMap version 38.96), outputting merged pairs and unmerged files. Merged and unmerged reads were normalized to 2000 reads for all 31-mers above 200 depth. Normalized, merged, and unmerged reads were aligned to the season-appropriate vaccine influenza strain, determined by sample collection time and PCR subtyping, using bwa mem (BWA version 0.7.17). Each sample was aligned to its respective season- and subtype-specific reference vaccine strain: A/Hong Kong/4801/2014 (2017–18 A(H3N2)), A/Singapore/INFIMH-16-0019/2016 (2018–19 A(H3N2)), or A/Michigan/45/2015 (2017–19 A(H1N1)pdm09) [44,46]. Alignments were concatenated with samtools merge to produce a single sample alignment file from the merged

file and the R1 and R2 unmerged files (Samtools version 1.15). iSNVs were called using callvariants.sh (BBMap version 38.96) to a minimum Phred quality score of Q30 and minimum position coverage of 100. A minimum frequency of 3% was applied after post-processing analysis in RStudio.

To generate consensus sequences for each sample, variants above 50% were applied to each sample's respective season- and subtype-specific reference vaccine strain sequence using BCFtools (version 1.15). Samples that failed to map to one or more gene segments, had <100x coverage at one or more segments, or <99% genome coverage were discarded. High-quality sample consensus genomes were split by segment and aligned to one another using ClustalO (version 1.2.4), and these alignments were visualized with JALview (version 2.11.2.6) to generate a single within-season consensus sequence for each season-subtype combination.

We then realigned all normalized reads to their respective within-season reference sequence to contextualize iSNVs relative to circulating strains in the community. Alignment and variant-calling were performed using the same quality metrics above to produce "realigned" alignment files and "recalled" iSNV files. Samples failing whole-genome coverage thresholds (described above) after realignment were excluded from downstream analyses. To detect and remove potential outliers with excess within-host diversity, we quantified the number of iSNVs >0.5% allele frequency for each sample, ranked them in descending order within their season-subtype group, and excluded samples that were in the top 10%, as in Xue *et al.* [8].

Diversity statistics

Variants were annotated using snpEff (version 5.1) relative to the within-season reference sequence to categorize mutation type (synonymous, nonsynonymous, stop-gained, etc.). Antigenic regions were defined in H3 as antigenic regions A-E [28,29] and in H1 as sites Ca,

Cb, Sa, Sb [30,34]. Amino acids composing antigenic sites H3 and H1 HAs are listed in **Supplemental Tables 4 and 5**, respectively. The per-segment, per-gene, and per-codon π diversity were calculated with SNPGenie (version 2019.10.31). Sample π diversity was computed as a per-nucleotide weighted average of each segment's π diversity.

Diversity statistics, data processing, and figure generation were performed with GNU bash (version 3.2.57(1)-release) and R (version 4.4.1 "Race for Your Life") through macOS Sequoia 15.5. Multiple packages were used in RStudio (version 2023.12.1+402 “Eye Holes”): tidyverse 2.0.0 (which loads tibble 3.2.1, dplyr 1.1.4, purrr 1.0.2, readr 2.1.5, tidyr 1.3.1, stringr 1.5.1, forcats 1.0.0), vcfR 1.15.0 for parsing VCF files, reshape2 1.4.4 for reshaping data, lubridate 1.9.3 for date-time manipulation, ggplot2 3.5.1 for graphics, plus cowplot 1.1.3, gridExtra 2.3, ggpubr 0.6.0, scales 1.3.0, and ggbeeswarm 0.7.2 for figure layout and annotation. All scripts are publicly available via GitHub (https://github.com/RiesHunter/ORCHARDS).

Statistical comparisons

Statistical comparisons were performed in R using the base-R wilcox.test function. We first assessed normality with the Shapiro–Wilk test: (shapiro.test). For non-normally distributed data, we employed non-parametric statistical tests. Paired Wilcoxon signed-rank tests (wilcox.test(x, y, paired = TRUE)) were used for matched (dependent) samples. The null hypothesis is that the median of the paired differences equals zero. One-sample Wilcoxon signed-rank tests (wilcox.test(x, mu = 0)) were used to test whether the sample median differs from zero. Unpaired two-sample Wilcoxon rank-sum tests (wilcox.test(x, y, paired = FALSE)), also known as Mann–Whitney U tests, were used for unmatched (independent) samples. The null hypothesis is that the two groups have the same distribution. All tests were two-sided, and

$p < 0.05$ was considered significant. Statistical test p-values for each figure can be found in **Supplemental Table 3**.

Donor-recipient pairing

Every ORCHARDS sample that passed quality control was paired based on household status and timing between infections. Individuals must have been within the same household to be considered pairs. Further, "donors" must have tested positive on the same day or before the "recipient's" sample was collected. To improve confidence in household pairing status, all samples within the same season and influenza subtype were randomly paired, and a proportion of shared variants was calculated for each random pair, as we have done previously [24]. This proportion was also calculated for potential household pairs. The proportion of shared variants was calculated as twice the number of shared variants between the two samples divided by the total number of variants between the two samples. Household pairs sharing a higher proportion of variants than 95% of random pairings within the season were considered valid donor-recipient pairs. Validated transmission pairs are shown in **Supplemental Figure 9**.

Maximum likelihood tree generation

All hemagglutinin consensus sequences were stratified by season and subtype and aligned with ClustalW (version 2.1) [47]. Aligned sequences were the input for RAxML-NG (version 1.2.2) [48] using the GTR+Gamma substitution model. RAxML-NG bestTree files were analyzed in RStudio and R using APE 5.8 [49], phangorn 2.12.1 [50], cowplot 1.1.3, RColorBrewer 1.1-3, ggplot 3.5.1 [51], and ggtree 3.12.0 [52–54]. Briefly, trees were imported with read.tree, roots

were resolved against the respective vaccine strain reference sequence, and tips were colored by household.

Data availability

The consensus genome sequences are available from GISAID (www.gisaid.org). Data, analysis pipelines, and figures are available for replication of these results on GitHub (https://github.com/RiesHunter/ORCHARDS).

Acknowledgments

We would like to thank the participants from this study as well as the Oregon School District staff, ORCHARDS team, and Wisconsin State Laboratory of Hygiene colleagues for their contribution to this study.

Disclaimers: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention (CDC).

References

1. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. Nat Rev Dis Primer. 2018;4: 3. doi:10.1038/s41572-018-0002-y

2. Paules CI, Sullivan SG, Subbarao K, Fauci AS. Chasing Seasonal Influenza - The Need for a Universal Influenza Vaccine. N Engl J Med. 2018;378: 7–9. doi:10.1056/NEJMp1714916

3. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, et al. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. J Virol. 2016;90: 3355–3365. doi:10.1128/JVI.03248-15

4. Moncla LH, Zhong G, Nelson CW, Dinis JM, Mutschler J, Hughes AL, et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. Cell Host Microbe. 2016;19: 169–180. doi:10.1016/j.chom.2016.01.011

5. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, et al. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLoS Pathog. 2017;13: e1006194. doi:10.1371/journal.ppat.1006194

6. Valesano AL, Fitzsimmons WJ, McCrone JT, Petrie JG, Monto AS, Martin ET, et al. Influenza B Viruses Exhibit Lower Within-Host Diversity than Influenza A Viruses in Human Hosts. J Virol. 2020;94: e01710-19. doi:10.1128/JVI.01710-19

7. 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: e35962. doi:10.7554/eLife.35962

8. Xue KS, Bloom JD. Linking influenza virus evolution within and between human hosts. Virus Evol. 2020;6: veaa010. doi:10.1093/ve/veaa010

9. Zhao L, Abbasi AB, Illingworth CJR. Mutational load causes stochastic evolutionary outcomes in acute RNA viral infection. Virus Evol. 2019;5: vez008. doi:10.1093/ve/vez008

10. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, George KS, et al. Stochastic

processes are key determinants of short-term evolution in influenza a virus. PLoS Pathog. 2006;2: e125. doi:10.1371/journal.ppat.0020125

11. Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. Mol Ecol. 2018;27: 4041–4051. doi:10.1111/mec.14726

12. Han AX, Maurer-Stroh S, Russell CA. Individual immune selection pressure has limited impact on seasonal influenza virus evolution. Nat Ecol Evol. 2019;3: 302–311. doi:10.1038/s41559-018-0741-x

13. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Rev Genet. 2007;8: 196–205. doi:10.1038/nrg2053

14. Petrova VN, Russell CA. The evolution of seasonal influenza viruses. Nat Rev Microbiol. 2018;16: 47–60. doi:10.1038/nrmicro.2017.118

15. McMinn P, Carrello A, Cole C, Baker D, Hampson A. Antigenic drift of influenza A (H3N2) virus in a persistently infected immunocompromised host is similar to that occurring in the community. Clin Infect Dis Off Publ Infect Dis Soc Am. 1999;29: 456–458. doi:10.1086/520243

16. Xue KS, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, et al. Parallel evolution of influenza across multiple spatiotemporal scales. eLife. 2017;6: e26875. doi:10.7554/eLife.26875

17. Xue KS, Moncla LH, Bedford T, Bloom JD. Within-Host Evolution of Human Influenza Virus. Trends Microbiol. 2018;26: 781–793. doi:10.1016/j.tim.2018.02.007

18. Pauly MD, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife. 2017;6: e26437. doi:10.7554/eLife.26437

19. Morris DH, Petrova VN, Rossine FW, Parker E, Grenfell BT, Neher RA, et al. Asynchrony between virus diversity and antibody selection limits influenza virus evolution. eLife. 2020;9: e62105. doi:10.7554/eLife.62105

20. Xue KS, Bloom JD. Reconciling disparate estimates of viral genetic diversity during human influenza infections. Nat Genet. 2019;51: 1298–1301. doi:10.1038/s41588-019-0349-3

21. Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, et al. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. Nat Commun. 2013;4: 2636. doi:10.1038/ncomms3636

22. Temte J, Barlow S, Landsverk M, Temte E, Schemmel A, Maerz B, et al. Cause-Specific School Absenteeism Monitoring Identifies Community Influenza Outbreaks. Online J Public Health Inform. 2018;10: e103. doi:10.5210/ojphi.v10i1.8757

23. Temte JL, Barlow S, Goss M, Temte E, Bell C, He C, et al. The Oregon Child Absenteeism Due to Respiratory Disease Study (ORCHARDS): Rationale, objectives, and design. Influenza Other Respir Viruses. 2022;16: 340–350. doi:10.1111/irv.12920

24. Braun KM, Moreno GK, Wagner C, Accola MA, Rehrauer WM, Baker DA, et al. Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. PLoS Pathog. 2021;17: e1009849. doi:10.1371/journal.ppat.1009849

25. Braun KM, Haddock Iii LA, Crooks CM, Barry GL, Lalli J, Neumann G, et al. Avian H7N9 influenza viruses are evolutionarily constrained by stochastic processes during replication and transmission in mammals. Virus Evol. 2023;9: vead004. doi:10.1093/ve/vead004

26. Nelson CW, Moncla LH, Hughes AL. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinforma Oxf Engl. 2015;31: 3709–3711. doi:10.1093/bioinformatics/btv449

27. Zhao L, Illingworth CJR. Measurements of intrahost viral diversity require an unbiased diversity metric. Virus Evol. 2019;5: vey041. doi:10.1093/ve/vey041

28. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. Biol Direct. 2006;1: 34. doi:10.1186/1745-6150-1-34

29. Alymova IV, York IA, Air GM, Cipollo JF, Gulati S, Baranovich T, et al. Glycosylation changes in the globular head of H3N2 influenza hemagglutinin modulate receptor binding without affecting virus virulence. Sci Rep. 2016;6: 36216. doi:10.1038/srep36216

30. Matsuzaki Y, Sugawara K, Nakauchi M, Takahashi Y, Onodera T, Tsunetsugu-Yokota Y, et al. Epitope mapping of the hemagglutinin molecule of A/(H1N1)pdm09 influenza virus by using monoclonal antibody escape mutants. J Virol. 2014;88: 12364–12373. doi:10.1128/JVI.01381-14

31. Tsang TK, Lau LLH, Cauchemez S, Cowling BJ. Household Transmission of Influenza Virus. Trends Microbiol. 2016;24: 123–133. doi:10.1016/j.tim.2015.10.012

32. Gubareva LV, Novikov DV, Hayden FG. Assessment of hemagglutinin sequence

heterogeneity during influenza virus transmission in families. J Infect Dis. 2002;186: 1575–1581. doi:10.1086/345372

33. Tsang TK, Cowling BJ, Fang VJ, Chan K-H, Ip DKM, Leung GM, et al. Influenza A Virus Shedding and Infectivity in Households. J Infect Dis. 2015;212: 1420–1428. doi:10.1093/infdis/jiv225

34. Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE, Wilson IA. Structural basis of pre-existing immunity to the 2009 H1N1 pandemic influenza virus. Science. 2010;328: 357–360. doi:10.1126/science.1186430

35. Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, et al. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. J Virol. 2016;90: 11247–11258. doi:10.1128/jvi.01657-16

36. Moncla LH, Bedford T, Dussart P, Horm SV, Rith S, Buchy P, et al. Quantifying within-host diversity of H5N1 influenza viruses in humans and poultry in Cambodia. PLoS Pathog. 2020;16: e1008191. doi:10.1371/journal.ppat.1008191

37. Amato KA, Haddock LA, Braun KM, Meliopoulos V, Livingston B, Honce R, et al. Influenza A virus undergoes compartmentalized replication in vivo dominated by stochastic bottlenecks. Nat Commun. 2022;13: 3416. doi:10.1038/s41467-022-31147-0

38. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe. 2014;16: 691–700. doi:10.1016/j.chom.2014.09.020

39. Lumby CK, Nene NR, Illingworth CJR. A novel framework for inferring parameters of transmission from viral sequence data. PLoS Genet. 2018;14: e1007718. doi:10.1371/journal.pgen.1007718

40. Ketklao S, Boonarkart C, Phakaratsakul S, Auewarakul P, Suptawiwat O. Responses to the Sb epitope contributed to antigenic drift of the influenza A 2009 H1N1 virus. Arch Virol. 2020;165: 2503–2512. doi:10.1007/s00705-020-04758-5

41. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinforma Oxf Engl. 2018;34: 4121–4123. doi:10.1093/bioinformatics/bty407

42. Bodewes R, de Mutsert G, van der Klis FRM, Ventresca M, Wilks S, Smith DJ, et al. Prevalence of antibodies against seasonal influenza A and B viruses in children in Netherlands. Clin Vaccine Immunol CVI. 2011;18: 469–476. doi:10.1128/CVI.00396-10

43. Longini IM, Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. Am J Epidemiol. 1982;115: 736–751. doi:10.1093/oxfordjournals.aje.a113356

44. Garten R. Update: Influenza Activity in the United States During the 2017–18 Season and Composition of the 2018–19 Influenza Vaccine. MMWR Morb Mortal Wkly Rep. 2018;67. doi:10.15585/mmwr.mm6722a4

45. Zhou B, Deng Y-M, Barnes JR, Sessions OM, Chou T-W, Wilson M, et al. Multiplex Reverse Transcription-PCR for Simultaneous Surveillance of Influenza A and B Viruses. J Clin Microbiol. 2017;55: 3492–3501. doi:10.1128/jcm.00957-17

46. Xu X. Update: Influenza Activity in the United States During the 2018–19 Season and Composition of the 2019–20 Influenza Vaccine. MMWR Morb Mortal Wkly Rep. 2019;68. doi:10.15585/mmwr.mm6824a3

47. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22: 4673–4680.

48. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35: 4453–4455. doi:10.1093/bioinformatics/btz305

49. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20: 289–290. doi:10.1093/bioinformatics/btg412

50. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27: 592–593. doi:10.1093/bioinformatics/btq706

51. Wickham H. ggplot2: Elegant Graphics for Data Analysis. In: Create Elegant Data Visualisations Using the Grammar of Graphics [Internet]. 2016 [cited 5 May 2023]. Available: https://ggplot2.tidyverse.org/

52. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8: 28–36. doi:10.1111/2041-210X.12628

53. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol.

2004;5: R80. doi:10.1186/gb-2004-5-10-r80

54. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12: 115–121. doi:10.1038/nmeth.3252

Figures

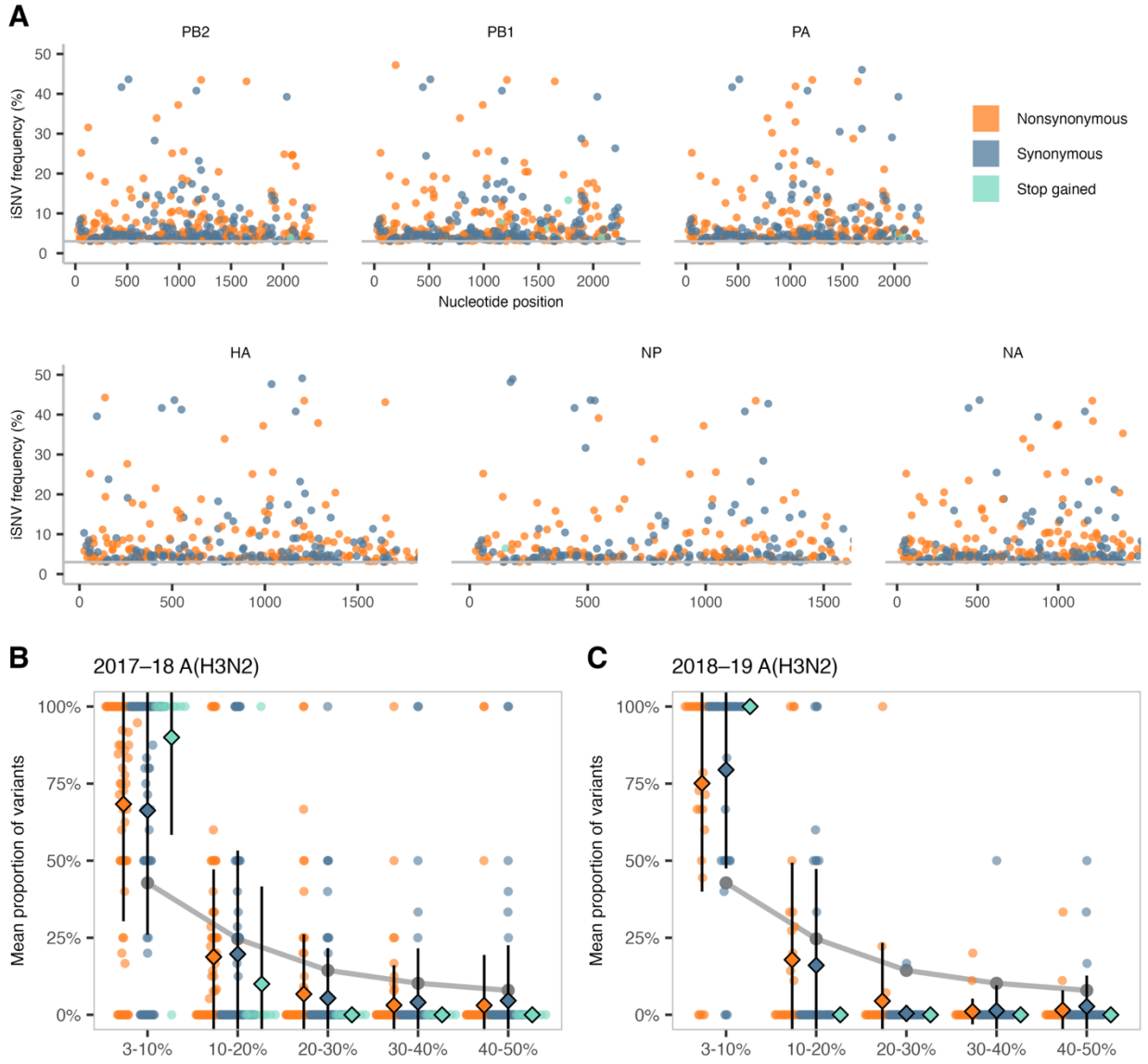


Fig 1. Low-frequency within-host variants predominate across influenza A/H3N2 genes and seasons. (A) Intra-host single-nucleotide variants (iSNVs) in at least 3% of sequencing reads (horizontal grey line) in a given sample were plotted for 2017–19 A(H3N2) according to their nucleotide position and frequency. Each dot represents an iSNV colored by mutation type: nonsynonymous (orange), synonymous (blue), or stop gained (green). Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin),

NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. **(B** and **C)** iSNV frequency spectra showing the mean proportion of each mutation type for each bin for 2017–18 A(H3N2) and 2018–19 A(H3N2) samples. Error bars denote the standard deviation for each bin's mutation type. The grey line indicates the proportion of iSNVs in each frequency bin expected under genetic drift. Per-sample proportions are shown as dots for each frequency bin.

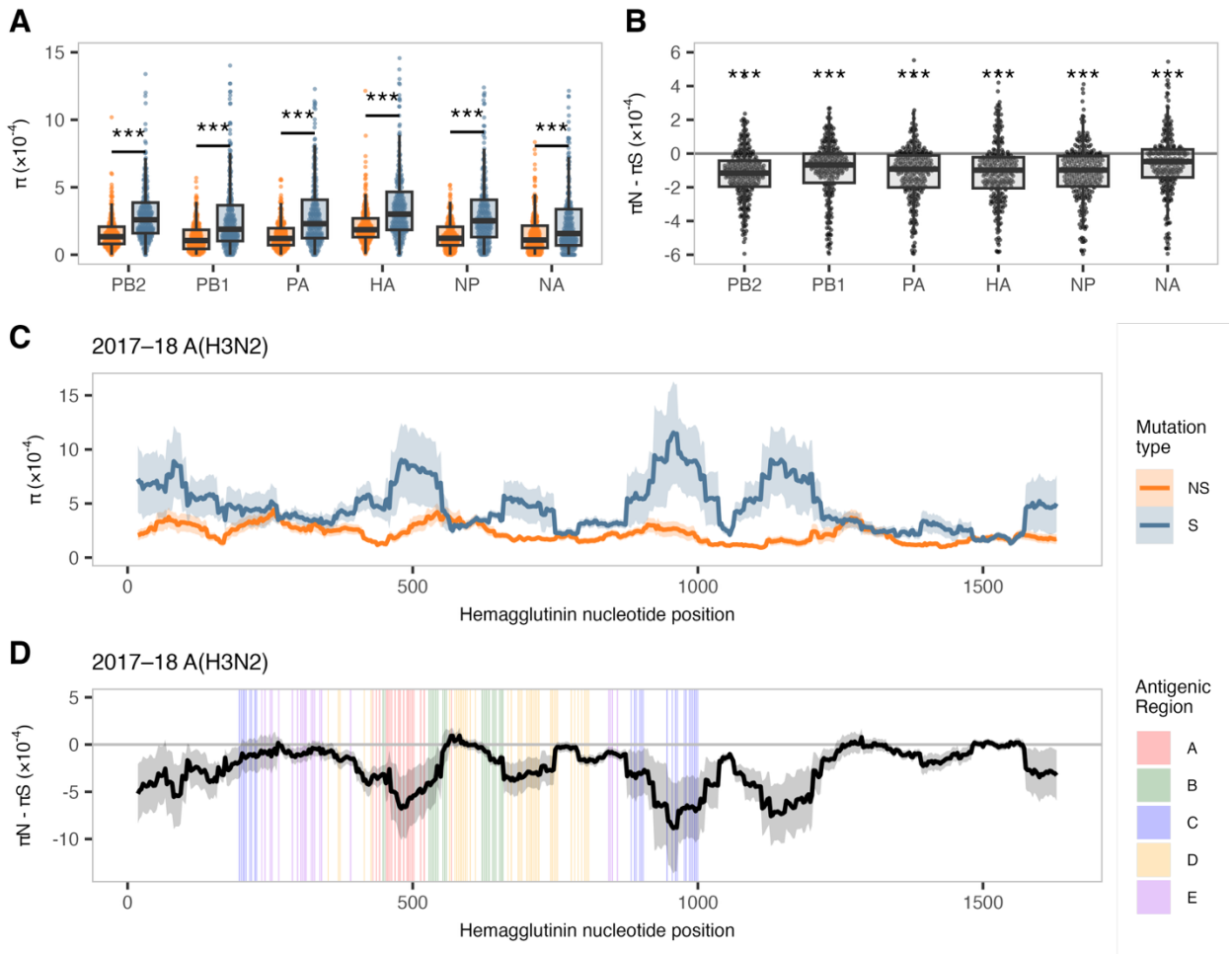


Fig 2. Synonymous diversity exceeds nonsynonymous diversity with evidence of purifying selection. (A) The average number of pairwise nonsynonymous differences per nonsynonymous site (π_N, orange) and synonymous differences per synonymous site (π_S, blue) across all analyzed viral genes, subtypes, and seasons. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range. **(B)** The difference between π_N and π_S across all viral genes, subtypes, and seasons. Boxplots are overlaid as described in A. The horizontal grey line represents $\pi_N = \pi_S$, with values above zero suggesting diversifying selection and values below zero suggesting purifying selection. **(C)** Mean π_N (orange) and π_S (blue) values for 30-codon sliding windows in the hemagglutinin gene for 2017–18 A(H3N2) viruses. **(D)** Mean $\pi_N - \pi_S$ values for 30-codon sliding windows in the hemagglutinin

gene for 2017–18 A(H3N2) viruses, colored by antigenic region. Shaded areas denote the standard error of the mean for each codon position. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. Paired comparisons of π_N and π_S were tested in A using Wilcoxon signed-rank tests. Median differences between π_N and π_S were tested in B against zero using one-sample Wilcoxon signed-rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

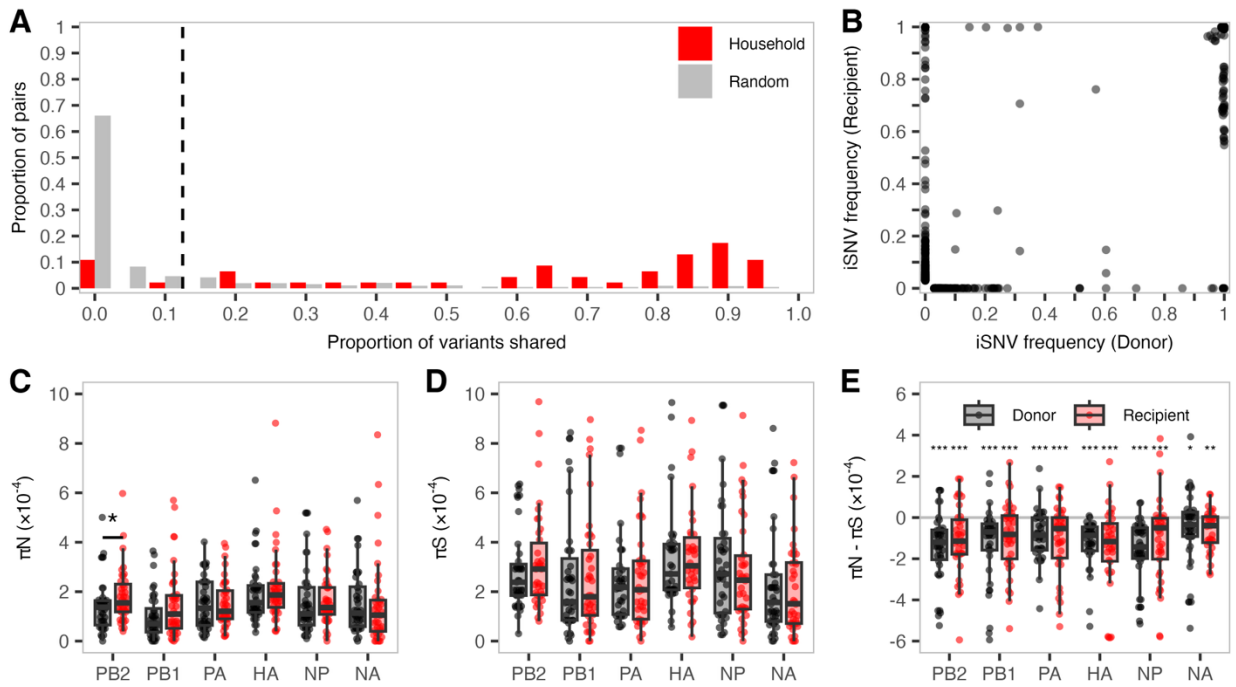


Fig 3. Donor–recipient variant frequencies and nucleotide diversity across genes. (A) The proportion of variants shared between putative household transmission pairs (red) or randomly paired virus sequences from a given season and subtype (grey) for all samples included in this study. The vertical line denotes the threshold below which 95% of random community pairs fell. **(B)** iSNV frequency in all donors and recipients as defined in A. Frequencies are shown as 0% in the paired individual if the iSNV was absent or below the 3% quality-control threshold. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range for **(C)** π_N, **(D)** π_S, and **(E)** $\pi_N - \pi_S$ values for all donor (black) and recipient (red) samples across all viral genes analyzed. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. The horizontal grey line in E represents zero, where $\pi_N = \pi_S$. Paired comparisons of nucleotide diversity between donors and recipients were tested in C and D using Wilcoxon signed-rank tests. Median differences between π_N and π_S were tested in E

against zero using one-sample Wilcoxon signed-rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

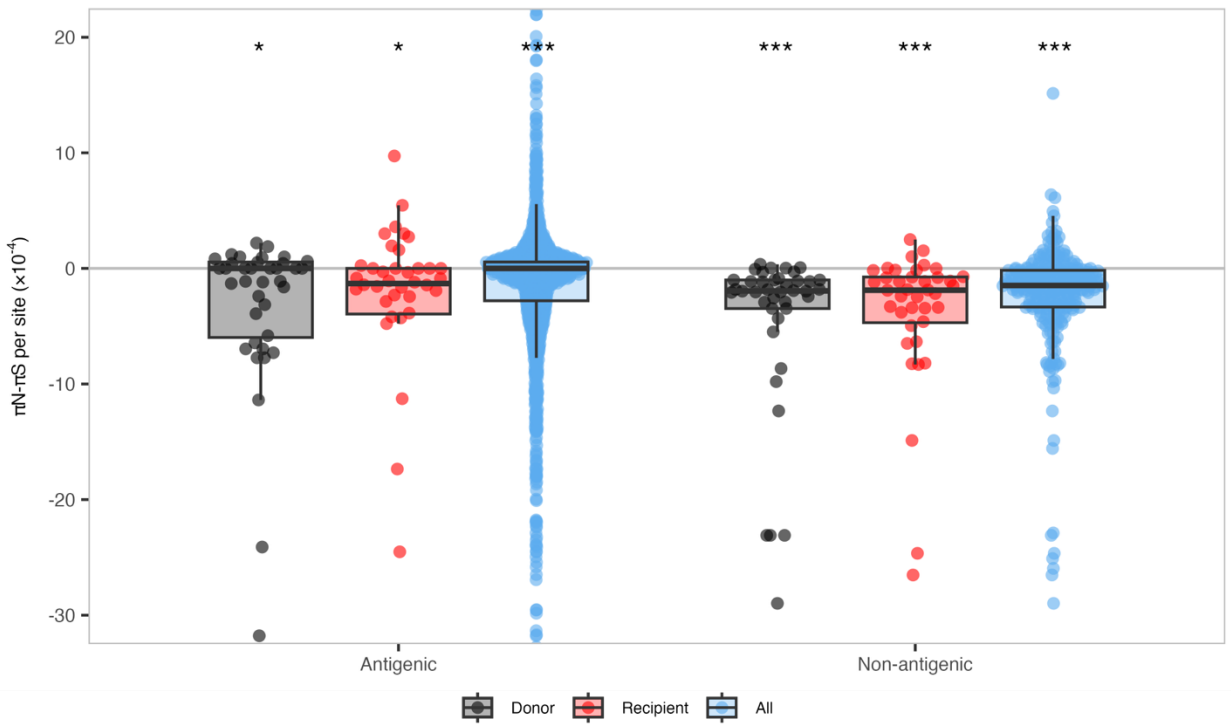


Fig 4. Nucleotide diversity across antigenic and non-antigenic regions of HA. Per-site $\pi_N - \pi_S$ values in antigenic and non-antigenic regions of HA for all samples in this study. $\pi_N - \pi_S$ values were normalized to account for differences in the amount of antigenic and non-antigenic sites, denoted $\pi_N - \pi_S$ per site. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range for each group: Donor (black), Recipient (red), or All (blue). The "All" group includes values for all viral genomes sampled (donors, recipients, and also those without a linked household contact). The horizontal grey line represents zero, where $\pi_N = \pi_S$. Median differences between π_N and π_S were tested against the null hypothesis ($\pi_N - \pi_S$ = zero) using one-sample Wilcoxon signed-rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

Table

Table 1. Study demographics

	2017–18		2018–19		2017–19
	A(H1N1) pdm09	A(H3N2)	A(H1N1) pdm09	A(H3N2)	Total
Samples sequenced	27	163	79	115	384
Passing quality control[1]	24 (89%)	119 (73%)	70 (89%)	70 (61%)	283 (74%)
Female	11 (46%)	55 (47%)	34 (49%)	39 (56%)	139 (49%)
Under 18 years of age	19 (79%)	102 (86%)	54 (77%)	65 (93%)	240 (85%)
Recent flu vaccination[2]	13 (54%)	47 (39%)	46 (66%)	47 (67%)	153 (54%)
Ct value (+/- SD)	27.2 (3.6)	28.4 (4.1)	26.4 (3.6)	26.2 (3.6)	27.2 (3.9)
Households with ≥2 flu-positive members[3]	3	16	12	5	36
Samples from households passing genetic criterion[4]	6	28	27	10	71
Two individuals	1	8	10	5	24
Three individuals	0	4	1	0	5
Four individuals	1	0	1	0	2

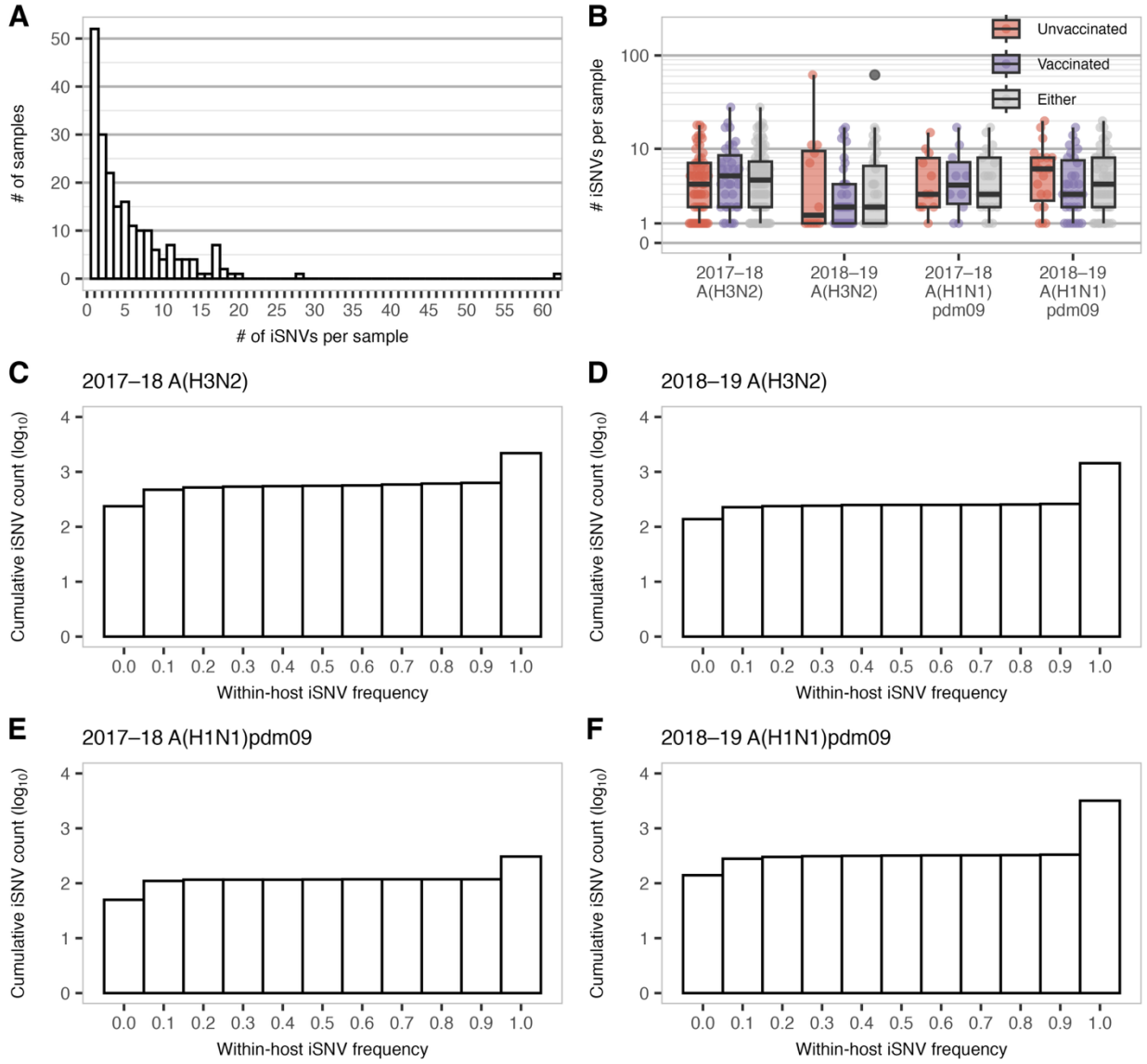
[1]All segments mapped; ≥99% genome coverage and ≥100× depth (see Methods).

[2]Current-season influenza vaccination at specimen collection, confirmed in the Wisconsin Immunization Registry [23].

[3]Households with at least two influenza-positive members, based on epidemiologic data alone. These 36 households contributed 46 potential transmission pairs (36 donors, 82 individuals).

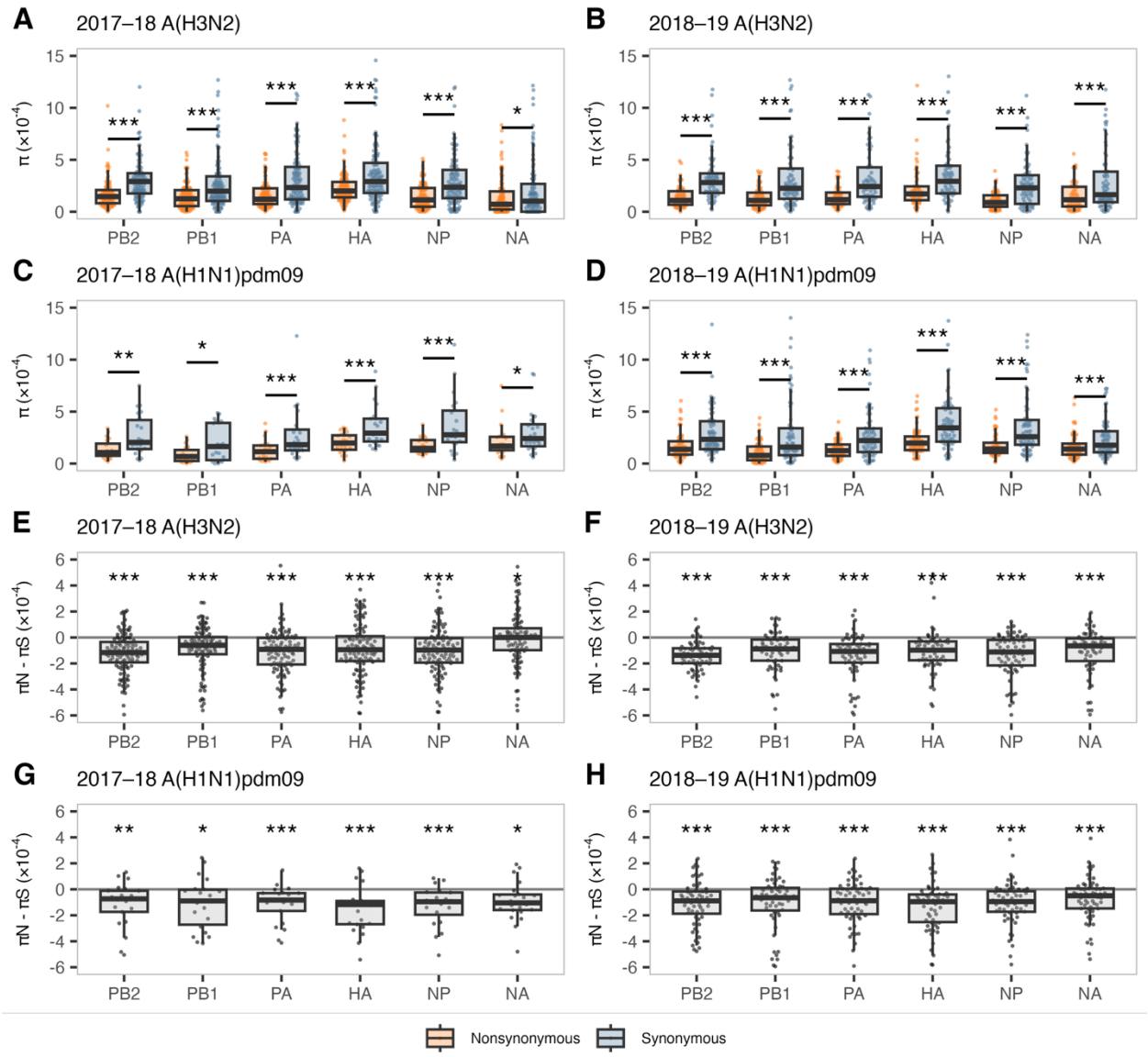
[4]Samples from households meeting both epidemiologic and genetic criteria for transmission. These 71 samples represent 40 transmission pairs involving 31 unique donors.

Supporting Information



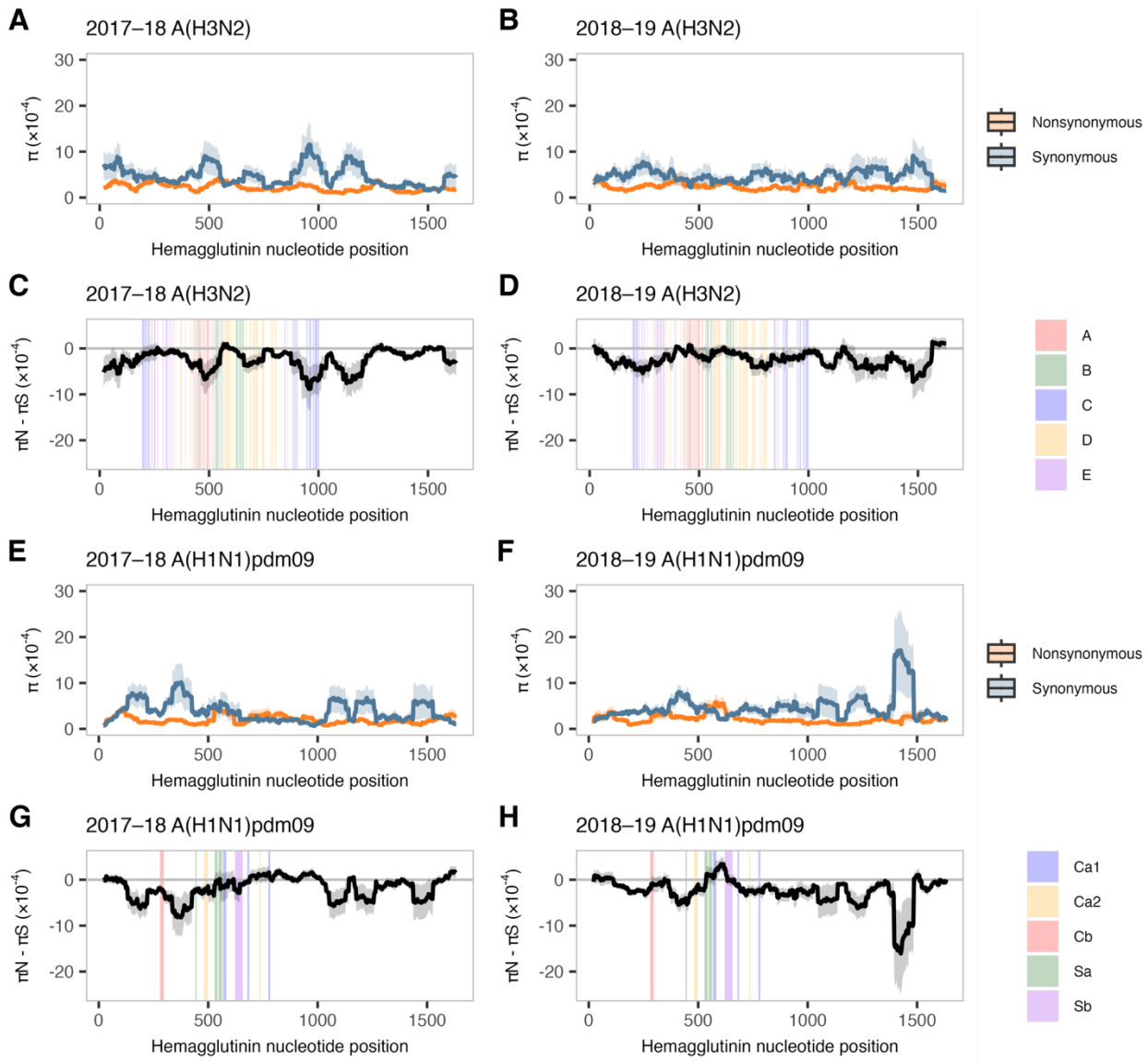
S1 Fig. Within-host iSNV counts and cumulative frequency by influenza season and subtype. (A) Histogram depicting the number of samples containing the indicated numbers of intra-host single-nucleotide variants (iSNVs) across all genes included in this study. **(B)** Number of iSNVs detected per sample (dots), stratified by reported recent influenza vaccination status: unvaccinated (red), vaccinated (purple). Vaccination refers to current-season influenza vaccination at specimen collection, confirmed in the Wisconsin Immunization Registry [23].

Either (grey) includes all samples regardless of vaccination status. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range for iSNV counts. Histograms of the cumulative number of iSNVs by frequency for **(C)** 2017–18 A(H3N2), **(D)** 2018–19 A(H3N2), **(E)** 2017–18 A(H1N1)pdm09, and **(F)** 2018–19 A(H1N1)pdm09. Each bin spans a 10% frequency interval of the form [a, b), where the lower bound is included and the upper bound is excluded—except for the final bin, which includes values equal to 100%. Variants below 3% frequency are excluded due to quality filtering. Paired comparisons of iSNV counts between vaccinated and unvaccinated groups within each season-subtype were tested in B using Wilcoxon rank-sum tests (AKA Mann-Whitney U tests). Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

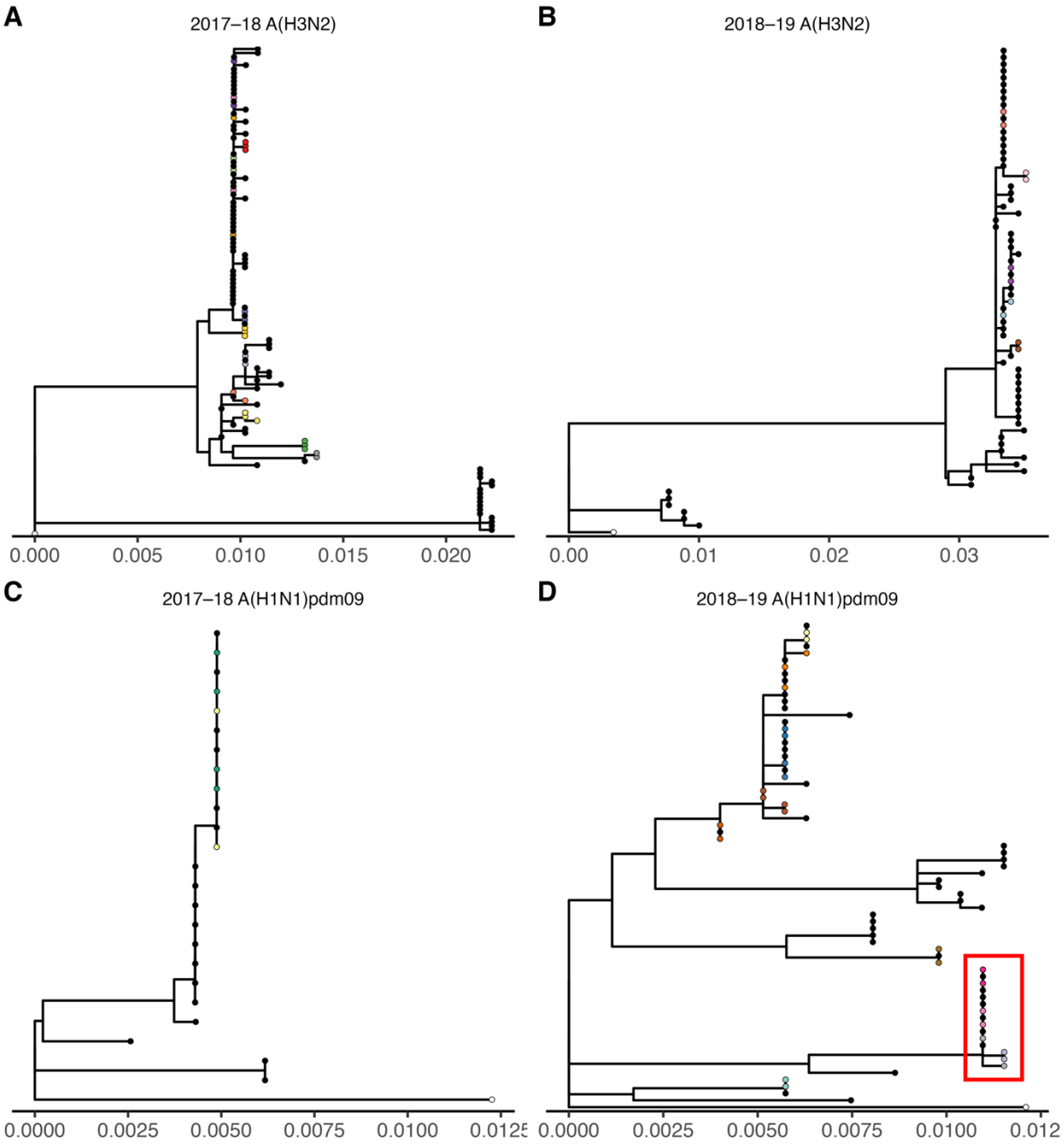


S2 Fig. Gene-wise nucleotide diversity across influenza seasons and subtypes. (A–D) The average number of pairwise nonsynonymous differences per nonsynonymous site (π_N, orange) and the average number of pairwise synonymous differences per synonymous site (π_S, blue) across all genes included in this study for **(A)** 2017–18 A(H3N2), **(B)** 2018–19 A(H3N2), **(C)** 2017–18 A(H1N1)pdm09, and **(D)** 2018–19 A(H1N1)pdm09. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range. **(E–H)** The

difference between π_N and π_S for **(E)** 2017–18 A(H3N2), **(F)** 2018–19 A(H3N2), **(G)** 2017–18 A(H1N1)pdm09, and **(H)** 2018–19 A(H1N1)pdm09. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. Boxplots are overlaid as described in above. The horizontal grey line represents $\pi_N = \pi_S$. Paired comparisons of π_N and π_S were tested using Wilcoxon signed-rank tests. Median differences between π_N and π_S were tested against the null hypothesis (i.e., $\pi_N - \pi_S$ = zero) using one-sample Wilcoxon signed-rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

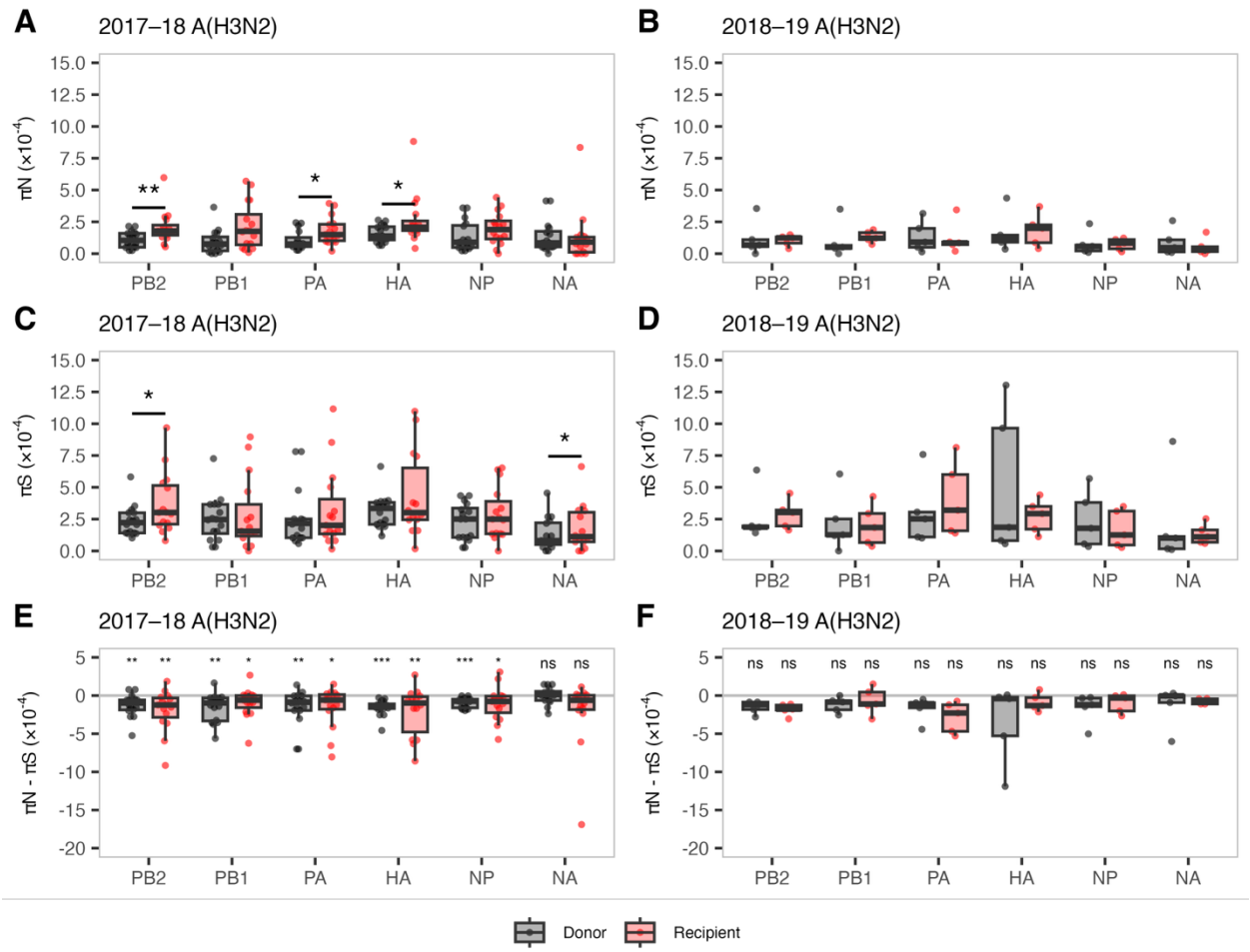


S3 Fig. Sliding-window hemagglutinin nucleotide diversity across influenza seasons and subtypes. (A, **B**, **E**, **F)** Mean π_N (orange) and π_S (blue) values for 30-codon sliding windows in the hemagglutinin gene for **(A)** 2017–18 A(H3N2), **(B)** 2018–19 A(H3N2), **(E)** 2017–18 A(H1N1)pdm09, and **(F)** 2018–19 A(H1N1)pdm09. **(C**, **D**, **G**, **H)** Mean $\pi_N - \pi_S$ values (black) for 30-codon sliding windows in the hemagglutinin gene for **(C)** 2017–18 A(H3N2), **(D)** 2018–19 A(H3N2), **(G)** 2017–18 A(H1N1)pdm09, and **(H)** 2018–19 A(H1N1)pdm09, colored by respective antigenic region. Shaded areas denote the standard error for each codon position.



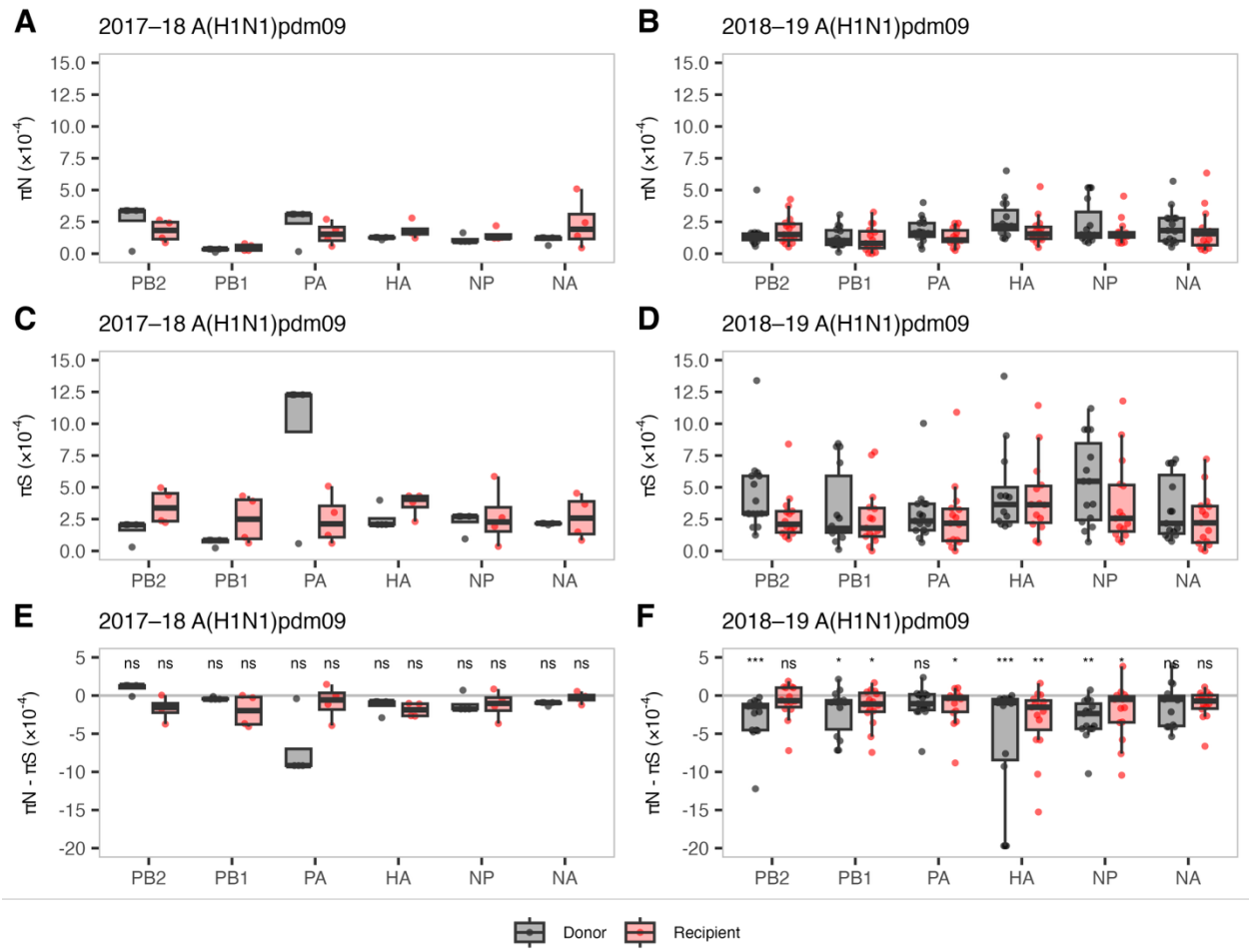
S4 Fig. Maximum-likelihood hemagglutinin phylogenies by influenza season and subtype. **(A)** 2017–18 A(H3N2), **(B)** 2018–19 A(H3N2), **(C)** 2017–18 A(H1N1)pdm09, and **(D)** 2018–19 A(H1N1)pdm09 (divergent taxon boxed in red). Tips for validated donor–recipient pairs are colored by household; all other tips are black. The season-appropriate vaccine strain reference

used as the outgroup is shown in white. Trees were inferred with RAxML-NG under GTR+Gamma and rooted to the vaccine reference; scale bars indicate substitutions per site.



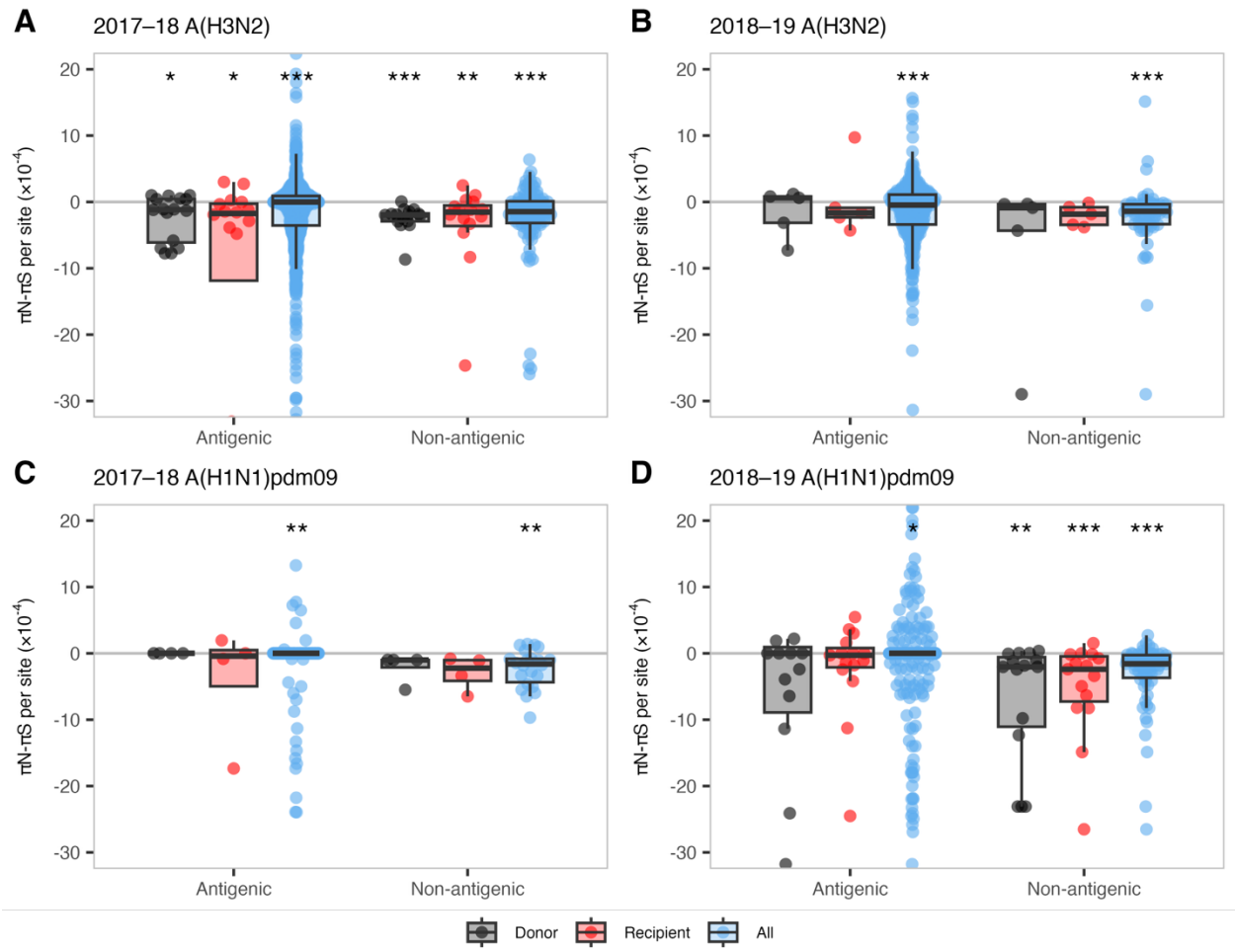
S5 Fig. Gene-wise π_N, π_S, and $\pi_N - \pi_S$ values in A(H3N2) donors and recipients. (A and **B)** π_N, **(C** and **D)** π_S, and **(E** and **F)** $\pi_N - \pi_S$ values for all donor (black) and recipient (red) samples across all genes included in this study, for 2017–18 A(H3N2) and 2018–19 A(H3N2), respectively. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. The horizontal grey line in E and F represents zero, where $\pi_N = \pi_S$. Paired comparisons of π_N and π_S between donors and recipients using Wilcoxon signed-rank tests. Median differences between π_N and π_S were tested in against zero using one-sample Wilcoxon signed-

rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

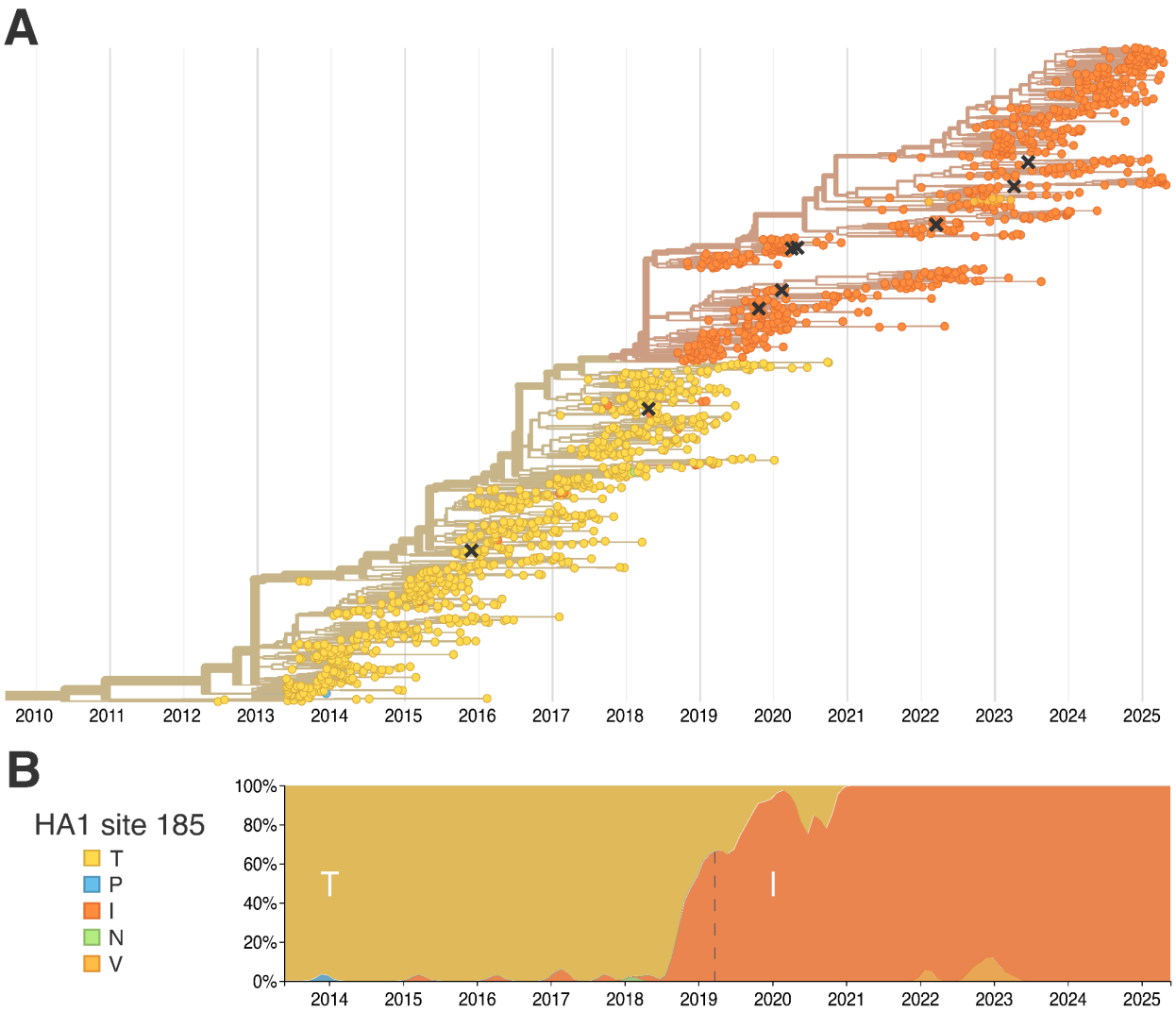


S6 Fig. Gene-wise π_N, π_S, and $\pi_N - \pi_S$ values in A(H1N1)pdm09 donors and recipients. (A and **B)** π_N, **(C** and **D)** π_S, and **(E** and **F)** $\pi_N - \pi_S$ values for all donor (black) and recipient (red) samples across all genes included in this study, for 2017–18 A(H1N1)pdm09 and 2018–19 A(H1N1)pdm09, respectively. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames. The horizontal grey line in E and F represents 0, where $\pi_N = \pi_S$. Paired comparisons of π_N and π_S between donors and recipients using Wilcoxon signed-rank tests. Median differences between π_N and π_S were tested in against zero using one-sample Wilcoxon signed-

rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.

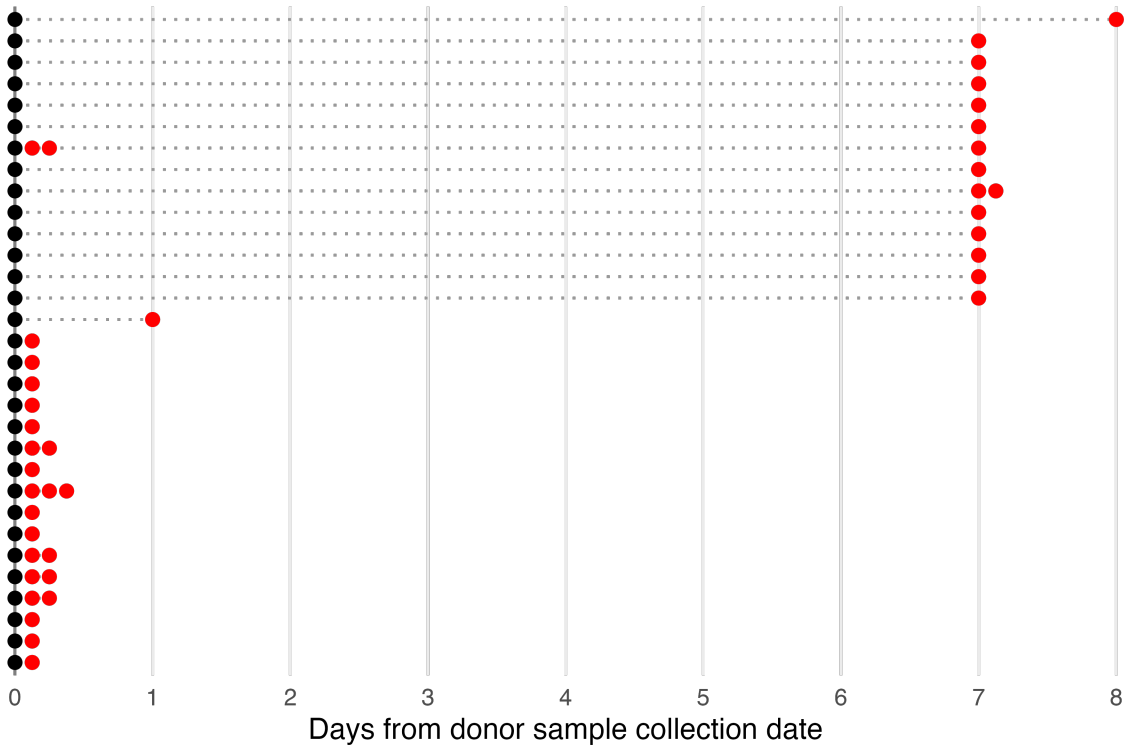


S7 Fig. Per-site $\pi_N - \pi_S$ values in antigenic and non-antigenic regions of HA. (A) 2017–18 A(H3N2), **(B)** 2018–19 A(H3N2), **(C)** 2017–18 A(H1N1)pdm09, and **(D)** 2018–19 A(H1N1)pdm09. $\pi_N - \pi_S$ values were normalized to account for differences in the amount of antigenic and non-antigenic sites, denoted $\pi_N - \pi_S$ per site. Boxplots are overlaid to indicate the median, interquartile range, and values within 1.5× the interquartile range for each group: Donor (black), Recipient (red), or All (blue). The "All" group includes values for all viral genomes sampled (donors, recipients, and also those without a linked household contact). The horizontal grey line represents zero, where $\pi_N = \pi_S$. Median differences between π_N and π_S were tested against zero using one-sample Wilcoxon signed-rank tests. Asterisks indicate significance thresholds: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns = not significant.



S8 Fig. Global dynamics of HA1 site 185 amino acid identities. (A) Time-resolved phylogeny of H1 consensus sequences (n = 1,461), constructed using Nextstrain and GISAID data. Branches and tips are colored by amino acid identity at HA1 site 185: threonine (T, yellow), proline (P, blue), isoleucine (I, dark orange), asparagine (N, green), and valine (V, light orange). **(B)** Temporal dynamics of HA1 185 genotypes among global sequences, showing the proportion of each amino acid over time. The dashed vertical line marks the approximate timing of I185T fixation in a specific donor-recipient pair, during a period when 185I was

present in 66% of global sequences. Both panels were adapted from Nextstrain [41] and modified to align x-axes and improve label clarity.



S9 Fig. Household sample collection timelines relative to donor sample collection date. Sample collection timelines for the 31 households meeting both epidemiologic and genetic criteria for influenza transmission. Each row represents a single household, with the donor sample shown as a black circle and recipient sample(s) shown as red circles. Dotted gray lines connect each donor to their corresponding recipient(s). When multiple samples within a household were collected on the same day, recipient points are offset by +0.125 days along the x-axis to avoid overlap.

S1 Table. Within-host variants from two outlier samples. Intra-host single-nucleotide variants (iSNVs) from two samples with greater than twenty iSNVs. The table was generated from a Variant Call Format file from each sample. Individual sample iSNVs were annotated with snpEff, and antigenicity was annotated with RStudio (see Methods).

CHROM	POS	ID	REF	ALT	QUAL	FILTER	STA	STO	R1P	R1M	AD	DP	AF	NVC	FLG	HMP	Allele	Ann	Ann_Impact	Gene_Name	HGVS.c	HGVS.p	WSLH_ID	Antigenicity
A/Hong_Kong/4801/2014|EPI_ISL_233740|	978	NA	A	T	36.6	Sample_1	977	978	32	28	60	1959	0.0306	0	0	0	ANN=T	missense_variant	MODERATE	NA	c.971A>T	p.Asp324Val	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	118	NA	A	G	34.23	Sample_1	117	118	45	13	58	1795	0.0323	0	0	3	ANN=G	missense_variant	MODERATE	PB1	c.106A>G	p.Thr36Ala	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NP	357	NA	C	T	37.07	Sample_1	356	357	74	47	121	3711	0.0326	0	0	1	ANN=T	synonymous_variant	LOW	NP	c.324C>T	p.Leu108Leu	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	1539	NA	G	T	36.83	Sample_1	1538	1539	40	42	82	2365	0.0347	0	0	1	ANN=T	synonymous_variant	LOW	PB1	c.1527G>T	p.Leu509Leu	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	2114	NA	C	T	34.14	Sample_1	2113	2114	22	68	90	2580	0.0349	0	0	1	ANN=T	missense_variant	MODERATE	PB1	c.2102C>T	p.Pro701Leu	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NP	1071	NA	C	T	35.85	Sample_1	1070	1071	88	95	183	4766	0.0382	2	0	2	ANN=T	synonymous_variant	LOW	NP	c.1038C>T	p.Phe346Phe	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NP	1064	NA	T	C	37.29	Sample_1	1063	1064	75	122	197	4768	0.0413	0	0	0	ANN=C	missense_variant	MODERATE	NP	c.1031T>C	p.Leu344Ser	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB2	1728	NA	G	C	36.93	Sample_1	1727	1728	54	65	119	2814	0.0423	2	0	0	ANN=C	missense_variant	MODERATE	PB2	c.1713G>C	p.Leu571Phe	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NP	334	NA	G	A	36.45	Sample_1	333	334	97	53	150	3534	0.0424	0	0	1	ANN=A	missense_variant	MODERATE	NP	c.301G>A	p.Asp101Asn	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB2	1729	NA	T	C	36.99	Sample_1	1728	1729	58	67	125	2778	0.045	2	0	0	ANN=C	missense_variant	MODERATE	PB2	c.1714T>C	p.Tyr572His	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|HA	68	NA	A	T	36.78	Sample_1	67	68	50	29	79	1693	0.0467	0	0	0	ANN=T	missense_variant	MODERATE	HA	c.51A>T	p.Gln17His	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	2008	NA	C	G	36.85	Sample_1	2007	2008	84	107	191	3974	0.0481	0	0	1	ANN=G	missense_variant	MODERATE	PA	c.1996C>G	p.Leu666Val	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	2035	NA	A	T	36.89	Sample_1	2034	2035	85	111	196	4038	0.0485	0	0	0	ANN=T	stop_gained	HIGH	PA	c.2023A>T	p.Lys675*	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	2013	NA	T	C	37.1	Sample_1	2012	2013	86	109	195	3900	0.05	0	0	0	ANN=C	synonymous_variant	LOW	PA	c.2001T>C	p.Leu667Leu	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	2020	NA	C	G	36.89	Sample_1	2019	2020	85	111	196	3915	0.0501	4	0	0	ANN=G	missense_variant	MODERATE	PA	c.2008C>G	p.Gln670Glu	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	2019	NA	T	A	36.89	Sample_1	2018	2019	85	111	196	3901	0.0502	4	0	0	ANN=A	synonymous_variant	LOW	PA	c.2007T>A	p.Val669Val	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB2	858	NA	G	C	37.15	Sample_1	857	858	65	52	117	2155	0.0543	0	0	0	ANN=C	missense_variant	MODERATE	PB2	c.843G>C	p.Leu281Phe	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PA	1912	NA	G	A	37.44	Sample_1	1911	1912	98	126	224	4014	0.0558	0	0	0	ANN=A	missense_variant	MODERATE	PA	c.1900G>A	p.Gly634Ser	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|HA	271	NA	C	A	37.17	Sample_1	270	271	98	73	171	2849	0.06	0	0	0	ANN=A	missense_variant	MODERATE	HA	c.254C>A	p.Ala85Asp	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|HA	569	NA	G	A	37.42	Sample_1	568	569	127	88	215	3197	0.0673	0	0	0	ANN=A	missense_variant	MODERATE	HA	c.552G>A	p.Met184Ile	Sample_1	Antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NA	1397	NA	C	T	33.86	Sample_1	1396	1397	13	41	54	770	0.0701	2	0	1	ANN=T	missense_variant	MODERATE	NA	c.1390C>T	p.Leu464Phe	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NA	1103	NA	A	G	37.34	Sample_1	1102	1103	48	77	125	1789	0.0702	0	0	0	ANN=G	missense_variant	MODERATE	NA	c.1096A>G	p.Ile366Val	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|NA	1264	NA	T	C	37.39	Sample_1	1263	1264	46	93	139	1933	0.0719	0	0	1	ANN=C	synonymous_variant	LOW	NA	c.1257T>C	p.Asn419Asn	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	1142	NA	A	G	37.44	Sample_1	1141	1142	74	78	152	1925	0.079	0	0	1	ANN=G	missense_variant	MODERATE	PB1	c.1130A>G	p.Asp377Gly	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	1211	NA	C	T	37.39	Sample_1	1210	1211	67	91	158	1999	0.079	0	0	0	ANN=T	missense_variant	MODERATE	PB1	c.1199C>T	p.Thr400Ile	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|HA	1481	NA	A	G	37.53	Sample_1	1480	1481	71	126	197	2473	0.0797	0	0	0	ANN=G	synonymous_variant	LOW	HA	c.1464A>G	p.Lys488Lys	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB1	694	NA	A	G	37.17	Sample_1	693	694	78	48	126	1560	0.0808	0	0	1	ANN=G	missense_variant	MODERATE	PB1	c.682A>G	p.Thr228Ala	Sample_1	Non-antigenic
A/Hong_Kong/4801/2014|EPI_ISL_233740|PB2	1404	NA	A	G	37.88	Sample_1	1403	1404	105	155	260	2079	0.1251	0	0	0	ANN=G	synonymous_variant	LOW	PB2	c.1389A>G	p.Val463Val	Sample_1	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB1	1401	NA	G	A	34.22	Sample_2	1400	1401	9	6	15	498	0.0301	1	0	0	ANN=A	synonymous_variant	LOW	PB1	c.1389G>A	p.Val463Val	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1650	NA	G	A	34.22	Sample_2	1649	1650	7	8	15	492	0.0305	0	0	0	ANN=A	missense_variant	MODERATE	HA	c.1633G>A	p.Val545Ile	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	782	NA	A	G	34.68	Sample_2	781	782	18	19	37	1209	0.0306	0	0	2	ANN=G	synonymous_variant	LOW	HA	c.765A>G	p.Pro255Pro	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	250	NA	A	G	35.38	Sample_2	249	250	18	12	30	976	0.0307	1	0	1	ANN=G	missense_variant	MODERATE	HA	c.233A>G	p.Glu78Gly	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB2	1110	NA	G	A	35.02	Sample_2	1109	1110	12	11	23	730	0.0315	0	0	0	ANN=A	missense_variant	MODERATE	PB2	c.1095G>A	p.Met365Ile	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1379	NA	T	C	35.44	Sample_2	1378	1379	9	16	25	782	0.032	0	0	1	ANN=C	synonymous_variant	LOW	HA	c.1362T>C	p.Asp454Asp	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	812	NA	A	G	34.81	Sample_2	811	812	18	20	38	1170	0.0325	0	0	2	ANN=G	synonymous_variant	LOW	HA	c.795A>G	p.Gly265Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	251	NA	A	G	35.7	Sample_2	250	251	20	14	34	999	0.034	1	0	0	ANN=G	synonymous_variant	LOW	HA	c.234A>G	p.Glu78Glu	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	998	NA	A	G	35.98	Sample_2	997	998	19	17	36	1042	0.0345	0	0	0	ANN=G	synonymous_variant	LOW	HA	c.981A>G	p.Gln327Gln	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	155	NA	G	A	34.47	Sample_2	154	155	21	8	29	821	0.0353	0	0	2	ANN=A	synonymous_variant	LOW	HA	c.138G>A	p.Thr46Thr	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	326	NA	C	T	33.57	Sample_2	325	326	24	18	42	1151	0.0365	0	0	4	ANN=T	synonymous_variant	LOW	HA	c.309C>T	p.Phe103Phe	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	710	NA	G	A	34.76	Sample_2	709	710	22	21	43	1173	0.0367	0	0	2	ANN=A	synonymous_variant	LOW	HA	c.693G>A	p.Pro231Pro	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	386	NA	C	A	36.12	Sample_2	385	386	23	15	38	1033	0.0368	1	0	0	ANN=A	synonymous_variant	LOW	HA	c.369C>A	p.Ser123Ser	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1286	NA	G	A	35.32	Sample_2	1285	1286	11	17	28	759	0.0369	0	0	1	ANN=A	synonymous_variant	LOW	HA	c.1269G>A	p.Gln423Gln	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	548	NA	C	T	35.38	Sample_2	547	548	14	22	36	968	0.0372	0	0	0	ANN=T	synonymous_variant	LOW	HA	c.531C>T	p.Tyr177Tyr	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	721	NA	C	T	34.21	Sample_2	720	721	21	22	43	1157	0.0372	0	0	3	ANN=T	missense_variant	MODERATE	HA	c.704C>T	p.Ser235Phe	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1104	NA	G	A	35.84	Sample_2	1103	1104	22	11	33	886	0.0372	1	0	0	ANN=A	missense_variant	MODERATE	HA	c.1087G>A	p.Val363Met	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	296	NA	T	C	35.95	Sample_2	295	296	28	15	43	1135	0.0379	0	0	0	ANN=C	synonymous_variant	LOW	HA	c.279T>C	p.Asp93Asp	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1121	NA	C	T	34.51	Sample_2	1120	1121	19	11	30	773	0.0388	0	0	2	ANN=T	synonymous_variant	LOW	HA	c.1104C>T	p.Gly368Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	231	NA	T	C	35.76	Sample_2	230	231	23	15	38	978	0.0389	0	0	0	ANN=C	missense_variant	MODERATE	HA	c.214T>C	p.Tyr72His	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	189	NA	C	T	35.68	Sample_2	188	189	23	10	33	839	0.0393	0	0	1	ANN=T	synonymous_variant	LOW	HA	c.172C>T	p.Leu58Leu	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	544	NA	A	C	35.45	Sample_2	543	544	16	22	38	950	0.04	1	0	0	ANN=C	missense_variant	MODERATE	HA	c.527A>C	p.Lys176Thr	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	200	NA	C	T	35.7	Sample_2	199	200	24	11	35	871	0.0402	0	0	1	ANN=T	synonymous_variant	LOW	HA	c.183C>T	p.Asn61Asn	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	541	NA	C	A	35.44	Sample_2	540	541	16	22	38	943	0.0403	1	0	0	ANN=A	missense_variant	MODERATE	HA	c.524C>A	p.Ser175Tyr	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	63	NA	A	G	35.33	Sample_2	62	63	17	7	24	592	0.0405	0	0	0	ANN=G	missense_variant	MODERATE	HA	c.46A>G	p.Thr16Ala	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	555	NA	T	C	35.85	Sample_2	554	555	16	23	39	962	0.0405	1	0	0	ANN=C	synonymous_variant	LOW	HA	c.538T>C	p.Leu180Leu	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1042	NA	G	A	34.3	Sample_2	1041	1042	23	17	40	985	0.0406	0	0	3	ANN=A	missense_variant	MODERATE	HA	c.1025G>A	p.Arg342Lys	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	340	NA	A	G	36.12	Sample_2	339	340	25	21	46	1127	0.0408	1	0	0	ANN=G	missense_variant	MODERATE	HA	c.323A>G	p.Lys108Arg	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	496	NA	A	G	35.43	Sample_2	495	496	17	21	38	921	0.0413	1	0	0	ANN=G	missense_variant	MODERATE	HA	c.479A>G	p.Lys160Arg	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1313	NA	A	T	35.44	Sample_2	1312	1313	10	22	32	775	0.0413	0	0	0	ANN=T	synonymous_variant	LOW	HA	c.1296A>T	p.Thr432Thr	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB1	1170	NA	G	A	33.93	Sample_2	1169	1170	9	10	19	459	0.0414	0	0	2	ANN=A	synonymous_variant	LOW	PB1	c.1158G>A	p.Arg386Arg	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	43	NA	G	A	34.61	Sample_2	42	43	15	6	21	506	0.0415	0	0	0	ANN=A	missense_variant	MODERATE	HA	c.26G>A	p.Cys9Tyr	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	497	NA	A	T	35.78	Sample_2	496	497	17	22	39	933	0.0418	1	0	0	ANN=T	missense_variant	MODERATE	HA	c.480A>T	p.Lys160Asn	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	578	NA	T	G	35.83	Sample_2	577	578	20	20	40	958	0.0418	0	0	1	ANN=G	missense_variant	MODERATE	HA	c.561T>G	p.Asn187Lys	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	337	NA	A	G	36.2	Sample_2	336	337	25	23	48	1142	0.042	1	0	0	ANN=G	missense_variant	MODERATE	HA	c.320A>G	p.Asn107Ser	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	281	NA	A	G	33.21	Sample_2	280	281	28	20	48	1141	0.0421	0	0	5	ANN=G	synonymous_variant	LOW	HA	c.264A>G	p.Gly88Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	557	NA	A	G	35.93	Sample_2	556	557	18	23	41	968	0.0424	1	0	0	ANN=G	synonymous_variant	LOW	HA	c.540A>G	p.Leu180Leu	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	643	NA	C	T	36.26	Sample_2	642	643	25	24	49	1139	0.043	0	0	1	ANN=T	missense_variant	MODERATE	HA	c.626C>T	p.Ser209Phe	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	300	NA	C	A	35.77	Sample_2	299	300	19	9	28	647	0.0433	0	0	1	ANN=A	missense_variant	MODERATE	HA	c.293C>A	p.Ala98Glu	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	653	NA	A	T	36.35	Sample_2	652	653	28	24	52	1171	0.0444	0	0	0	ANN=T	synonymous_variant	LOW	HA	c.636A>T	p.Ala212Ala	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	608	NA	T	G	32.66	Sample_2	607	608	29	20	49	1089	0.045	0	0	6	ANN=G	synonymous_variant	LOW	HA	c.591T>G	p.Gly197Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1277	NA	G	A	35.7	Sample_2	1276	1277	13	23	36	798	0.0451	0	0	1	ANN=A	synonymous_variant	LOW	HA	c.1260G>A	p.Gly420Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1166	NA	T	C	35.86	Sample_2	1165	1166	15	20	35	771	0.0454	0	0	0	ANN=C	synonymous_variant	LOW	HA	c.1149T>C	p.Leu383Leu	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	893	NA	G	A	35.25	Sample_2	892	893	27	27	54	1184	0.0456	0	0	2	ANN=A	synonymous_variant	LOW	HA	c.876G>A	p.Lys292Lys	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1281	NA	A	G	35.6	Sample_2	1280	1281	12	24	36	790	0.0456	0	0	0	ANN=G	missense_variant	MODERATE	HA	c.1264A>G	p.Ile422Val	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PA	1835	NA	C	T	36.36	Sample_2	1834	1835	15	29	44	951	0.0463	0	0	0	ANN=T	missense_variant	MODERATE	PA	c.1823C>T	p.Thr608Ile	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	458	NA	T	A	35.79	Sample_2	457	458	19	22	41	882	0.0465	1	0	0	ANN=A	synonymous_variant	LOW	HA	c.441T>A	p.Thr147Thr	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	477	NA	T	G	35.79	Sample_2	476	477	20	23	43	921	0.0467	0	0	0	ANN=G	missense_variant	MODERATE	HA	c.460T>G	p.Ser154Ala	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	447	NA	G	A	35.8	Sample_2	446	447	19	23	42	894	0.047	0	0	0	ANN=A	missense_variant	MODERATE	HA	c.430G>A	p.Ala144Thr	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	383	NA	A	C	36.08	Sample_2	382	383	25	23	48	1019	0.0471	1	0	1	ANN=C	synonymous_variant	LOW	HA	c.366A>C	p.Ala122Ala	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	392	NA	A	G	36.11	Sample_2	391	392	24	25	49	1041	0.0471	0	0	1	ANN=G	synonymous_variant	LOW	HA	c.375A>G	p.Arg125Arg	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	457	NA	C	A	35.82	Sample_2	456	457	19	23	42	887	0.0474	1	0	1	ANN=A	missense_variant	MODERATE	HA	c.440C>A	p.Thr147Asn	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	977	NA	A	C	36.3	Sample_2	976	977	24	28	52	1098	0.0474	0	0	1	ANN=C	synonymous_variant	LOW	HA	c.960A>C	p.Ala320Ala	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	452	NA	G	A	35.81	Sample_2	451	452	19	23	42	883	0.0476	1	0	0	ANN=A	synonymous_variant	LOW	HA	c.435G>A	p.Gly145Gly	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	932	NA	A	C	35.12	Sample_2	931	932	23	30	53	1111	0.0477	0	0	2	ANN=C	synonymous_variant	LOW	HA	c.915A>C	p.Pro305Pro	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	428	NA	C	A	33.52	Sample_2	427	428	20	23	43	888	0.0484	1	0	4	ANN=A	missense_variant	MODERATE	HA	c.411C>A	p.Asp137Glu	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1259	NA	G	A	35.96	Sample_2	1258	1259	12	29	41	844	0.0486	0	0	1	ANN=A	synonymous_variant	LOW	HA	c.1242G>A	p.Glu414Glu	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	426	NA	G	A	35.83	Sample_2	425	426	19	24	43	881	0.0488	1	0	1	ANN=A	missense_variant	MODERATE	HA	c.409G>A	p.Asp137Asn	Sample_2	Antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|HA	1223	NA	G	A	31.74	Sample_2	1222	1223	11	29	40	805	0.0497	0	0	4	ANN=A	synonymous_variant	LOW	HA	c.1206G>A	p.Gly402Gly	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB1	2206	NA	T	C	33.66	Sample_2	2205	2206	1	11	12	239	0.0502	1	0	1	ANN=C	missense_variant	MODERATE	PB1	c.2194T>C	p.Ser732Pro	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB2	1690	NA	G	A	35.16	Sample_2	1689	1690	22	22	44	805	0.0547	0	0	2	ANN=A	missense_variant	MODERATE	PB2	c.1675G>A	p.Ala559Thr	Sample_2	Non-antigenic
A/Singapore/INFIMH-16-0019/2016|EPI_ISL_296168|PB2	1831	NA	G	T	36.41	Sample_2	1830	1831	16	24	40	720	0.0556	0	0	1	ANN=T	missense_variant	MODERATE	PB2	c.1816G>T	p.Val606Leu	Sample_2	Non-antigenic

S2 Table. Notable iSNVs present in both donor and recipient samples. Merge_ID denotes the year and subtype of the sample. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames.

AA	iSNV	Donor_ID	Recipient_ID	Donor_AF	Recipient_AF	Merge_ID	Ann
G187G	A594G-NP	Donor 1	Recipient 1a	0.1003	0.1488	18-19_H1N1	synonymous_variant
T22A	A97G-NP	Donor 1	Recipient 1b	0.242	0.2977	18-19_H1N1	missense_variant
C471C	C1485T-HA	Donor 1	Recipient 1a	0.3164	0.1427	18-19_H1N1	synonymous_variant
C471C	C1485T-HA	Donor 1	Recipient 1b	0.3164	0.7063	18-19_H1N1	synonymous_variant
C471C	C1485T-HA	Donor 1	Recipient 1c	0.3164	0.9991	18-19_H1N1	synonymous_variant
P627P	C1893T-PB1	Donor 2	Recipient 2	0.1045	0.2876	17-18_H3N2	synonymous_variant
A270S	G823T-PB2	Donor 3	Recipient 3a	0.6047	0.1471	17-18_H1N1	missense_variant
A270S	G823T-PB2	Donor 3	Recipient 3b	0.6047	0.058	17-18_H1N1	missense_variant
N346N	T1050C-PB1	Donor 4	Recipient 4	0.5705	0.761	18-19_H1N1	synonymous_variant
I185T	T626C-HA	Donor 5	Recipient 5	0.3767	1	18-19_H1N1	missense_variant

S3 Table. Statistical tests and results from the study. Genes shown: PB2 (polymerase basic 2), PB1 (polymerase basic 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), and NA (neuraminidase). Not included: MP (matrix; M1/M2) and NS (nonstructural; NS1/NEP) due to overlapping reading frames.

Figure 2A

Test: shapiro.test

Comparison: normality

product	Mutation	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Nonsynonymous	257	0.823	2.03E-16	FALSE
PB2	Synonymous	257	0.337	1.70E-29	FALSE
PB1	Nonsynonymous	257	0.895	2.10E-12	FALSE
PB1	Synonymous	257	0.619	1.58E-23	FALSE
PA	Nonsynonymous	257	0.893	1.74E-12	FALSE
PA	Synonymous	257	0.592	3.31E-24	FALSE
HA	Nonsynonymous	257	0.734	5.06E-20	FALSE
HA	Synonymous	257	0.399	2.20E-28	FALSE
NP	Nonsynonymous	257	0.92	1.48E-10	FALSE
NP	Synonymous	257	0.529	1.03E-25	FALSE
NA	Nonsynonymous	257	0.838	1.14E-15	FALSE
NA	Synonymous	257	0.621	1.79E-23	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: πN vs πS

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	257	3810	9.92E-27	***
PB1	257	5437	4.46E-20	***
PA	257	3942	3.26E-26	***
HA	257	5318	3.79E-21	***
NP	257	4048	1.38E-25	***
NA	257	7994	3.02E-10	***

Figure 2B

Test: shapiro.test

Comparison: normality

product	n	W	p_value	is_normal
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	257	0.354	3.40E-29	FALSE
PB1	257	0.653	1.40E-22	FALSE
PA	257	0.596	4.20E-24	FALSE
HA	257	0.471	5.65E-27	FALSE
NP	257	0.535	1.40E-25	FALSE
NA	257	0.654	1.51E-22	FALSE

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	257	3810	9.92E-27	***
PB1	257	5437	4.46E-20	***
PA	257	3942	3.26E-26	***
HA	257	5318	3.79E-21	***
NP	257	4048	1.38E-25	***
NA	257	7994	3.02E-10	***

Figure 3C

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	40	0.836	0.0000419	FALSE
HA	Recipient	40	0.764	0.00000133	FALSE
NA	Donor	40	0.873	0.00035	FALSE
NA	Recipient	40	0.737	0.000000421	FALSE
NP	Donor	40	0.823	0.0000207	FALSE
NP	Recipient	40	0.904	0.00251	FALSE
PA	Donor	40	0.926	0.0122	FALSE
PA	Recipient	40	0.922	0.00862	FALSE
PB1	Donor	40	0.856	0.000127	FALSE
PB1	Recipient	40	0.841	0.0000542	FALSE
PB2	Donor	40	0.862	0.000178	FALSE
PB2	Recipient	40	0.87	0.000288	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN vs Recipient πN

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	40	244	0.025	TRUE
PB1	40	288	0.103	FALSE
PA	40	414	0.963	FALSE
HA	40	356	0.476	FALSE
NP	40	367	0.572	FALSE
NA	40	474	0.393	FALSE

Figure 3D

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	40	0.659	2.22E-08	FALSE
HA	Recipient	40	0.444	3.98E-11	FALSE
NA	Donor	40	0.809	1.05E-05	FALSE
NA	Recipient	40	0.445	4.05E-11	FALSE
NP	Donor	40	0.874	3.58E-04	FALSE
NP	Recipient	40	0.378	8.00E-12	FALSE
PA	Donor	40	0.741	4.95E-07	FALSE
PA	Recipient	40	0.459	5.77E-11	FALSE
PB1	Donor	40	0.66	2.27E-08	FALSE
PB1	Recipient	40	0.424	2.40E-11	FALSE
PB2	Donor	40	0.748	6.68E-07	FALSE
PB2	Recipient	40	0.332	2.82E-12	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πS vs Recipient πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	40	352	0.444	FALSE
PB1	40	418	0.921	FALSE
PA	40	397	0.868	FALSE
HA	40	432	0.775	FALSE
NP	40	407	0.973	FALSE
NA	40	376	0.656	FALSE

Figure 3E

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Donor	40	0.792	4.71E-06	FALSE
PB2	Recipient	40	0.367	6.18E-12	FALSE
PB1	Donor	40	0.645	1.37E-08	FALSE
PB1	Recipient	40	0.437	3.33E-11	FALSE
PA	Donor	40	0.799	6.53E-06	FALSE
PA	Recipient	40	0.45	4.65E-11	FALSE
HA	Donor	40	0.582	1.80E-09	FALSE
HA	Recipient	40	0.471	7.87E-11	FALSE
NP	Donor	40	0.807	9.36E-06	FALSE
NP	Recipient	40	0.409	1.65E-11	FALSE
NA	Donor	40	0.917	6.26E-03	FALSE
NA	Recipient	40	0.492	1.36E-10	FALSE

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Gene	Group	n	statistic	p_value	label
<fct>	<chr>	<int>	<dbl>	<dbl>	<chr>
PB2	Donor	40	79	0.00000887	***
PB2	Recipient	40	120	0.0000381	***
PB1	Donor	40	75	0.0000114	***
PB1	Recipient	40	156	0.000402	***
PA	Donor	40	92	0.0000197	***
PA	Recipient	40	124	0.0000508	***
HA	Donor	40	3	4.64E-08	***
HA	Recipient	40	77	0.00000106	***
NP	Donor	40	27	0.000000272	***
NP	Recipient	40	125	0.000223	***
NA	Donor	40	249	0.0499	*
NA	Recipient	40	166	0.00309	**

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN – πS vs Recipient πN – πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	40	424	0.858	FALSE
PB1	40	354	0.46	FALSE
PA	40	420	0.9	FALSE
HA	40	372	0.618	FALSE
NP	40	382	0.715	FALSE
NA	40	431	0.785	FALSE

Figure 4

Test: shapiro.test

Comparison: normality

Region	Group	n	W	p_value	is_normal
<fct>	<fct>	<int>	<dbl>	<dbl>	<lgl>
Antigenic	Donor	40	0.567	1.14E-09	FALSE
Antigenic	Recipient	40	0.624	6.84E-09	FALSE
Antigenic	All	1542	0.237	7.62E-62	FALSE
Non-antigenic	Donor	40	0.611	4.44E-09	FALSE
Non-antigenic	Recipient	40	0.587	2.07E-09	FALSE
Non-antigenic	All	257	0.612	1.08E-23	FALSE

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons; wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Comparing πN – πS between various groups

Region	comparison	statistic	p_value	label
<fct>	<chr>	<dbl>	<dbl>	<chr>
Antigenic	Donor vs Recipient	351	0.783	ns
Antigenic	Donor vs All	28444	0.394	ns
Antigenic	Recipient vs All	25470	0.0559	ns
Non-antigenic	Donor vs Recipient	340	0.354	ns
Non-antigenic	Donor vs All	4301	0.097	ns
Non-antigenic	Recipient vs All	4425	0.157	ns

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Region	Group	n	statistic	p_value	label
<fct>	<fct>	<int>	<dbl>	<dbl>	<chr>
Antigenic	Donor	40	142	2.31E-02	*
Antigenic	Recipient	40	163	1.31E-02	*
Antigenic	All	1542	186681	5.35E-20	***
Non-antigenic	Donor	40	14	1.06E-07	***
Non-antigenic	Recipient	40	60	1.83E-07	***
Non-antigenic	All	257	4946	3.05E-22	***

Test: wilcox.test(x, y, paired = TRUE) for donor and recipient comparisons; wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Antigenic vs. non-antigenic πN – πS per group

Group	a	n	paired	statistic	p_value	label
<fct>	<list>	<list>	<lgl>	<dbl>	<dbl>	<chr>
Donor	40	40	TRUE	490	2.85E-01	ns
Recipient	40	40	TRUE	456	5.45E-01	ns
All	1542	257	FALSE	248636	3.69E-11	***

Median πN – πS values

Region	Group	median_per10k
<fct>	<fct>	<dbl>
Antigenic	Donor	0
Antigenic	Recipient	-1.31
Antigenic	All	0
Non-antigenic	Donor	-1.95
Non-antigenic	Recipient	-1.88
Non-antigenic	All	-1.48

Percent πN – πS values that are positive, negative, zero

Region	n_total	n_neg	n_zero	n_pos	pct_neg	pct_zero	pct_pos
<fct>	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>
Antigenic	1542	609	490	443	39.5	31.8	28.7
Non-antigenic	257	206	1	50	80.2	0.389	19.5

S1B Fig

Test: wilcox.test(x, y, paired = FALSE)

Comparison: iSNV counts between vaccinated and unvaccinated individuals

Merge_ID	n	statistic	p_value	label
<chr>	<dbl>	<dbl>	<chr>	<lgl>
2017–18 H3N2	92	957	0.748	ns
2018–19 H3N2	47	200	0.819	ns
2017–18 H1N1	21	64	0.557	ns
2018–19 H1N1	50	292	0.976	ns

Medians

Merge_ID	Group	median_Count
2017–18 H3N2	Vaccinated	5
2017–18 H3N2	Unvaccinated	4
2017–18 H3N2	Either	4.5
2017–18 H3N2	Either	4.5
2018–19 H3N2	Vaccinated	2
2018–19 H3N2	Unvaccinated	1.5
2018–19 H3N2	Either	2
2018–19 H3N2	Either	2
2017–18 H1N1	Vaccinated	4
2017–18 H1N1	Unvaccinated	3
2017–18 H1N1	Either	3
2017–18 H1N1	Either	3
2018–19 H1N1	Vaccinated	3
2018–19 H1N1	Unvaccinated	6
2018–19 H1N1	Either	4
2018–19 H1N1	Either	4
All seasons	Vaccinated	3
All seasons	Unvaccinated	4
All seasons	Either	4
All seasons	Either	4

S2A Fig

Test: shapiro.test

Comparison: normality

product	Mutation	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Nonsynonymous	108	0.774	1.34E-11	FALSE
PB2	Synonymous	108	0.247	4.64E-21	FALSE
PB1	Nonsynonymous	108	0.893	2.95E-07	FALSE
PB1	Synonymous	108	0.511	2.59E-17	FALSE
PA	Nonsynonymous	108	0.86	1.08E-08	FALSE
PA	Synonymous	108	0.561	1.95E-16	FALSE
HA	Nonsynonymous	108	0.867	2.17E-08	FALSE
HA	Synonymous	108	0.429	1.31E-18	FALSE
NP	Nonsynonymous	108	0.913	2.80E-06	FALSE
NP	Synonymous	108	0.482	8.67E-18	FALSE
NA	Nonsynonymous	108	0.769	9.54E-12	FALSE
NA	Synonymous	108	0.539	8.00E-17	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: πN vs πS

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	108	744	1.60E-11	***
PB1	108	1144	5.90E-08	***
PA	108	748	1.74E-11	***
HA	108	1255	2.31E-07	***
NP	108	945	1.54E-09	***
NA	108	1732	4.21E-02	*

S2B Fig

Test: shapiro.test

Comparison: normality

product	Mutation	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Nonsynonymous	64	0.884	2.08E-05	FALSE
PB2	Synonymous	64	0.634	2.30E-11	FALSE
PB1	Nonsynonymous	64	0.915	3.00E-04	FALSE
PB1	Synonymous	64	0.82	2.30E-07	FALSE
PA	Nonsynonymous	64	0.908	1.68E-04	FALSE
PA	Synonymous	64	0.851	1.81E-06	FALSE
HA	Nonsynonymous	64	0.629	1.92E-11	FALSE
HA	Synonymous	64	0.316	1.06E-15	FALSE
NP	Nonsynonymous	64	0.919	4.68E-04	FALSE
NP	Synonymous	64	0.876	1.14E-05	FALSE
NA	Nonsynonymous	64	0.889	3.15E-05	FALSE
NA	Synonymous	64	0.755	5.50E-09	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: πN vs πS

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	64	58	5.25E-11	***
PB1	64	214	5.56E-08	***
PA	64	172	6.58E-09	***
HA	64	278	3.53E-07	***
NP	64	175	7.41E-09	***
NA	64	265	2.23E-07	***

S2C Fig

Test: shapiro.test

Comparison: normality

product	Mutation	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Nonsynonymous	22	0.901	0.0304	FALSE
PB2	Synonymous	22	0.909	0.0447	FALSE
PB1	Nonsynonymous	22	0.88	0.0121	FALSE
PB1	Synonymous	22	0.861	0.00535	FALSE
PA	Nonsynonymous	22	0.886	0.0158	FALSE
PA	Synonymous	22	0.738	0.000063	FALSE
HA	Nonsynonymous	22	0.969	0.681	TRUE
HA	Synonymous	22	0.888	0.0172	FALSE
NP	Nonsynonymous	22	0.918	0.0684	TRUE
NP	Synonymous	22	0.886	0.0155	FALSE
NA	Nonsynonymous	22	0.79	0.000355	FALSE
NA	Synonymous	22	0.838	0.00207	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: πN vs πS

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	22	40	0.00367	**
PB1	22	45	0.015	*
PA	22	23	0.000305	***
HA	22	27	0.000593	***
NP	22	19	0.000146	***
NA	22	51	0.0127	*

S2D Fig

Test: shapiro.test

Comparison: normality

product	Mutation	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Nonsynonymous	63	0.856	2.90E-06	FALSE
PB2	Synonymous	63	0.535	7.86E-13	FALSE
PB1	Nonsynonymous	63	0.892	4.72E-05	FALSE
PB1	Synonymous	63	0.725	1.52E-09	FALSE
PA	Nonsynonymous	63	0.958	3.11E-02	FALSE
PA	Synonymous	63	0.843	1.15E-06	FALSE
HA	Nonsynonymous	63	0.881	1.88E-05	FALSE
HA	Synonymous	63	0.743	3.55E-09	FALSE
NP	Nonsynonymous	63	0.862	4.34E-06	FALSE
NP	Synonymous	63	0.826	3.94E-07	FALSE
NA	Nonsynonymous	63	0.842	1.09E-06	FALSE
NA	Synonymous	63	0.876	1.31E-05	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: πN vs πS

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	63	370	0.0000127	***
PB1	63	365	0.0000109	***
PA	63	320	0.00000252	***
HA	63	207	4.25E-08	***
NP	63	221	7.27E-08	***
NA	63	497	0.000474	***

S2E Fig

Test: shapiro.test
Comparison: normality

product	n	W	p_value	is_normal
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	108	0.272	9.36E-21	FALSE
PB1	108	0.536	7.03E-17	FALSE
PA	108	0.571	2.96E-16	FALSE
HA	108	0.476	6.98E-18	FALSE
NP	108	0.512	2.68E-17	FALSE
NA	108	0.639	6.55E-15	FALSE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	108	744	1.60E-11	***
PB1	108	1144	5.90E-08	***
PA	108	748	1.74E-11	***
HA	108	1255	2.31E-07	***
NP	108	945	1.54E-09	***
NA	108	1732	4.21E-02	*

S2F Fig

Test: shapiro.test
Comparison: normality

product	n	W	p_value	is_normal
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	64	0.515	3.21E-13	FALSE
PB1	64	0.796	5.37E-08	FALSE
PA	64	0.85	1.59E-06	FALSE
HA	64	0.409	1.28E-14	FALSE
NP	64	0.869	6.32E-06	FALSE
NA	64	0.719	9.48E-10	FALSE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	64	58	5.25E-11	***
PB1	64	214	5.56E-08	***
PA	64	172	6.58E-09	***
HA	64	278	3.53E-07	***
NP	64	175	7.41E-09	***
NA	64	265	2.23E-07	***

S2G Fig

Test: shapiro.test
Comparison: normality

product	n	W	p_value	is_normal
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	22	0.902	0.0319	FALSE
PB1	22	0.956	0.407	TRUE
PA	22	0.795	0.000415	FALSE
HA	22	0.966	0.609	TRUE
NP	22	0.809	0.000696	FALSE
NA	22	0.957	0.431	TRUE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	22	40	0.00367	**
PB1	22	45	0.015	*
PA	22	23	0.000305	***
HA	22	27	0.000593	***
NP	22	19	0.000146	***
NA	22	51	0.0127	*

S2H Fig

Test: shapiro.test
Comparison: normality

product	n	W	p_value	is_normal
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	63	0.576	3.29E-12	FALSE
PB1	63	0.753	5.98E-09	FALSE
PA	63	0.875	1.21E-05	FALSE
HA	63	0.727	1.66E-09	FALSE
NP	63	0.837	7.91E-07	FALSE
NA	63	0.927	1.13E-03	FALSE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

product	n	statistic	p_value	label
<fct>	<int>	<dbl>	<dbl>	<chr>
PB2	63	370	0.0000127	***
PB1	63	365	0.0000109	***
PA	63	320	0.00000252	***
HA	63	207	4.25E-08	***
NP	63	221	7.27E-08	***
NA	63	497	0.000474	***

S5A Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	16	0.925	0.2	TRUE
HA	Recipient	16	0.721	0.000292	FALSE
NA	Donor	16	0.796	0.00239	FALSE
NA	Recipient	16	0.588	0.0000124	FALSE
NP	Donor	16	0.821	0.0053	FALSE
NP	Recipient	16	0.974	0.895	TRUE
PA	Donor	16	0.805	0.00315	FALSE
PA	Recipient	16	0.916	0.148	TRUE
PB1	Donor	16	0.843	0.0108	FALSE
PB1	Recipient	16	0.881	0.0401	FALSE
PB2	Donor	16	0.913	0.131	TRUE
PB2	Recipient	16	0.79	0.00202	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN vs. Recipient πN

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	16	14	0.00336	TRUE
PB1	16	30	0.0507	FALSE
PA	16	28	0.0386	TRUE
HA	16	22	0.0155	TRUE
NP	16	39	0.144	FALSE
NA	16	81	0.518	FALSE

S5B Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	5	0.79	0.067	TRUE
HA	Recipient	5	0.957	0.789	TRUE
NA	Donor	5	0.829	0.137	TRUE
NA	Recipient	5	0.809	0.0954	TRUE
NP	Donor	5	0.78	0.0553	TRUE
NP	Recipient	5	0.925	0.566	TRUE
PA	Donor	5	0.924	0.554	TRUE
PA	Recipient	5	0.752	0.031	FALSE
PB1	Donor	5	0.713	0.013	FALSE
PB1	Recipient	5	0.969	0.87	TRUE
PB2	Donor	5	0.81	0.0975	TRUE
PB2	Recipient	5	0.913	0.485	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN vs. Recipient πN

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	5	7	1	FALSE
PB1	5	5	0.625	FALSE
PA	5	6	0.812	FALSE
HA	5	7	1	FALSE
NP	5	8	1	FALSE
NA	5	11	0.438	FALSE

S5C Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	16	0.879	0.038	FALSE
HA	Recipient	16	0.541	0.00000472	FALSE
NA	Donor	16	0.862	0.0203	FALSE
NA	Recipient	16	0.532	0.00000395	FALSE
NP	Donor	16	0.908	0.106	TRUE
NP	Recipient	16	0.406	0.00000039	FALSE
PA	Donor	16	0.744	0.000538	FALSE
PA	Recipient	16	0.505	0.00000233	FALSE
PB1	Donor	16	0.905	0.0965	TRUE
PB1	Recipient	16	0.493	0.00000185	FALSE
PB2	Donor	16	0.854	0.0154	FALSE
PB2	Recipient	16	0.377	0.000000239	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πS vs. Recipient πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	16	29	0.0443	TRUE
PB1	16	59	0.669	FALSE
PA	16	46	0.274	FALSE
HA	16	50	0.375	FALSE
NP	16	38	0.13	FALSE
NA	16	24	0.0214	TRUE

S5D Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	5	0.816	0.108	TRUE
HA	Recipient	5	0.968	0.861	TRUE
NA	Donor	5	0.664	0.00385	FALSE
NA	Recipient	5	0.905	0.437	TRUE
NP	Donor	5	0.903	0.43	TRUE
NP	Recipient	5	0.86	0.227	TRUE
PA	Donor	5	0.813	0.104	TRUE
PA	Recipient	5	0.895	0.382	TRUE
PB1	Donor	5	0.87	0.265	TRUE
PB1	Recipient	5	0.939	0.657	TRUE
PB2	Donor	5	0.645	0.00232	FALSE
PB2	Recipient	5	0.943	0.687	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πS vs. Recipient πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	5	7	1	FALSE
PB1	5	10	0.625	FALSE
PA	5	3	0.312	FALSE
HA	5	10	0.625	FALSE
NP	5	8	1	FALSE
NA	5	6	0.812	FALSE

S5E Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Donor	16	0.917	0.15	TRUE
PB2	Recipient	16	0.4	0.000000348	FALSE
PB1	Donor	16	0.919	0.164	TRUE
PB1	Recipient	16	0.444	0.000000752	FALSE
PA	Donor	16	0.809	0.00363	FALSE
PA	Recipient	16	0.488	0.00000168	FALSE
HA	Donor	16	0.853	0.0148	FALSE
HA	Recipient	16	0.563	0.00000742	FALSE
NP	Donor	16	0.902	0.0857	TRUE
NP	Recipient	16	0.429	0.000000577	FALSE
NA	Donor	16	0.936	0.302	TRUE
NA	Recipient	16	0.611	0.0000204	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN – πS vs. Recipient πN – πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	16	86	0.375	FALSE
PB1	16	56	0.562	FALSE
PA	16	81	0.528	FALSE
HA	16	82	0.495	FALSE
NP	16	81	0.528	FALSE
NA	16	105	0.0577	FALSE

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Gene	Group	n	statistic	p_value	label
<fct>	<chr>	<int>	<dbl>	<dbl>	<chr>
PB2	Donor	16	13	0.0048	**
PB2	Recipient	16	15	0.00418	**
PB1	Donor	16	11	0.00346	**
PB1	Recipient	16	27	0.0335	*
PA	Donor	16	14	0.00564	**
PA	Recipient	16	21	0.0131	*
HA	Donor	16	0	0.000478	***
HA	Recipient	16	16	0.00516	**
NP	Donor	16	0	0.000478	***
NP	Recipient	16	20	0.0249	*
NA	Donor	16	67	0.712	ns
NA	Recipient	16	25	0.0901	ns

S5F Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Donor	5	0.857	0.219	TRUE
PB2	Recipient	5	0.884	0.328	TRUE
PB1	Donor	5	0.957	0.788	TRUE
PB1	Recipient	5	0.972	0.888	TRUE
PA	Donor	5	0.777	0.0521	TRUE
PA	Recipient	5	0.889	0.353	TRUE
HA	Donor	5	0.79	0.0666	TRUE
HA	Recipient	5	0.947	0.714	TRUE
NP	Donor	5	0.775	0.0494	FALSE
NP	Recipient	5	0.838	0.159	TRUE
NA	Donor	5	0.693	0.00799	FALSE
NA	Recipient	5	0.83	0.138	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN – πS vs. Recipient πN – πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	5	11	0.438	FALSE
PB1	5	4	0.438	FALSE
PA	5	13	0.188	FALSE
HA	5	3	0.312	FALSE
NP	5	7	1	FALSE
NA	5	10	0.625	FALSE

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Gene	Group	n	statistic	p_value	label
<fct>	<chr>	<int>	<dbl>	<dbl>	<chr>
PB2	Donor	5	0	0.0625	ns
PB2	Recipient	5	0	0.0625	ns
PB1	Donor	5	0	0.1	ns
PB1	Recipient	5	5	0.625	ns
PA	Donor	5	0	0.0625	ns
PA	Recipient	5	0	0.0625	ns
HA	Donor	5	1	0.125	ns
HA	Recipient	5	2	0.188	ns
NP	Donor	5	0	0.0625	ns
NP	Recipient	5	2	0.188	ns
NA	Donor	5	4	0.438	ns
NA	Recipient	5	0	0.0625	ns

S6A Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	4	0.63	0.00124	FALSE
HA	Recipient	4	0.863	0.27	TRUE
NA	Donor	4	0.63	0.00124	FALSE
NA	Recipient	4	0.939	0.649	TRUE
NP	Donor	4	0.63	0.00124	FALSE
NP	Recipient	4	0.675	0.00576	FALSE
PA	Donor	4	0.63	0.00124	FALSE
PA	Recipient	4	0.987	0.94	TRUE
PB1	Donor	4	0.63	0.00124	FALSE
PB1	Recipient	4	0.841	0.199	TRUE
PB2	Donor	4	0.63	0.00124	FALSE
PB2	Recipient	4	0.881	0.344	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN vs. Recipient πN

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	4	7	0.625	FALSE
PB1	4	1	0.25	FALSE
PA	4	8	0.375	FALSE
HA	4	1	0.25	FALSE
NP	4	0	0.125	FALSE
NA	4	2	0.375	FALSE

S6B Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	15	0.846	0.0151	FALSE
HA	Recipient	15	0.81	0.00499	FALSE
NA	Donor	15	0.885	0.0565	TRUE
NA	Recipient	15	0.81	0.00501	FALSE
NP	Donor	15	0.738	0.000647	FALSE
NP	Recipient	15	0.753	0.000981	FALSE
PA	Donor	15	0.95	0.527	TRUE
PA	Recipient	15	0.936	0.332	TRUE
PB1	Donor	15	0.904	0.111	TRUE
PB1	Recipient	15	0.921	0.2	TRUE
PB2	Donor	15	0.592	0.0000213	FALSE
PB2	Recipient	15	0.91	0.138	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πN vs. Recipient πN

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	15	36	0.188	FALSE
PB1	15	66	0.762	FALSE
PA	15	87	0.135	FALSE
HA	15	88	0.121	FALSE
NP	15	80	0.277	FALSE
NA	15	73	0.489	FALSE

S6C Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	4	0.63	0.00124	FALSE
HA	Recipient	4	0.792	0.089	TRUE
NA	Donor	4	0.63	0.00124	FALSE
NA	Recipient	4	0.914	0.505	TRUE
NP	Donor	4	0.63	0.00124	FALSE
NP	Recipient	4	0.944	0.681	TRUE
PA	Donor	4	0.63	0.00124	FALSE
PA	Recipient	4	0.939	0.646	TRUE
PB1	Donor	4	0.63	0.00124	FALSE
PB1	Recipient	4	0.834	0.178	TRUE
PB2	Donor	4	0.63	0.00124	FALSE
PB2	Recipient	4	0.845	0.21	TRUE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πS vs. Recipient πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	4	0	0.125	FALSE
PB1	4	2	0.375	FALSE
PA	4	9	0.25	FALSE
HA	4	1	0.25	FALSE
NP	4	6	0.875	FALSE
NA	4	3	0.625	FALSE

S6D Fig

Test: shapiro.test

Comparison: normality

Gene	Group	n	W	p_value	is_normal
<chr>	<chr>	<int>	<dbl>	<dbl>	<lgl>
HA	Donor	15	0.746	0.000798	FALSE
HA	Recipient	15	0.822	0.00704	FALSE
NA	Donor	15	0.814	0.00555	FALSE
NA	Recipient	15	0.914	0.155	TRUE
NP	Donor	15	0.926	0.242	TRUE
NP	Recipient	15	0.831	0.00959	FALSE
PA	Donor	15	0.739	0.000665	FALSE
PA	Recipient	15	0.785	0.00236	FALSE
PB1	Donor	15	0.713	0.000346	FALSE
PB1	Recipient	15	0.848	0.0163	FALSE
PB2	Donor	15	0.79	0.00276	FALSE
PB2	Recipient	15	0.473	0.00000215	FALSE

Test: wilcox.test(x, y, paired = TRUE)

Comparison: Donor πS vs. Recipient πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	15	86	0.151	FALSE
PB1	15	77	0.359	FALSE
PA	15	68	0.679	FALSE
HA	15	85	0.169	FALSE
NP	15	79	0.303	FALSE
NA	15	84	0.188	FALSE

S6E Fig

Test: shapiro.test
Comparison: normality

Gene	Group	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Donor	4	0.63	0.00124	FALSE
PB2	Recipient	4	0.969	0.835	TRUE
PB1	Donor	4	0.63	0.00124	FALSE
PB1	Recipient	4	0.806	0.113	TRUE
PA	Donor	4	0.63	0.00124	FALSE
PA	Recipient	4	0.972	0.853	TRUE
HA	Donor	4	0.63	0.00124	FALSE
HA	Recipient	4	0.76	0.0475	FALSE
NP	Donor	4	0.63	0.00124	FALSE
NP	Recipient	4	0.981	0.91	TRUE
NA	Donor	4	0.63	0.00124	FALSE
NA	Recipient	4	0.988	0.945	TRUE

Test: wilcox.test(x, y, paired = TRUE)
Comparison: Donor πN – πS vs. Recipient πN – πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	4	10	0.125	FALSE
PB1	4	7	0.625	FALSE
PA	4	1	0.25	FALSE
HA	4	7	0.625	FALSE
NP	4	4	0.875	FALSE
NA	4	0	0.125	FALSE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

Gene	Group	n	statistic	p_value	label
<fct>	<chr>	<int>	<dbl>	<dbl>	<chr>
PB2	Donor	4	9	0.186	ns
PB2	Recipient	4	1	0.25	ns
PB1	Donor	4	0	0.089	ns
PB1	Recipient	4	1	0.25	ns
PA	Donor	4	0	0.089	ns
PA	Recipient	4	3	0.625	ns
HA	Donor	4	0	0.089	ns
HA	Recipient	4	0	0.125	ns
NP	Donor	4	1	0.186	ns
NP	Recipient	4	2	0.375	ns
NA	Donor	4	0	0.089	ns
NA	Recipient	4	3	0.625	ns

S6F Fig

Test: shapiro.test
Comparison: normality

Gene	Group	n	W	p_value	is_normal
<fct>	<chr>	<int>	<dbl>	<dbl>	<lgl>
PB2	Donor	15	0.728	0.000505	FALSE
PB2	Recipient	15	0.527	0.00000584	FALSE
PB1	Donor	15	0.727	0.000488	FALSE
PB1	Recipient	15	0.906	0.117	TRUE
PA	Donor	15	0.826	0.00798	FALSE
PA	Recipient	15	0.777	0.00188	FALSE
HA	Donor	15	0.671	0.000124	FALSE
HA	Recipient	15	0.815	0.00571	FALSE
NP	Donor	15	0.893	0.0742	TRUE
NP	Recipient	15	0.901	0.1	TRUE
NA	Donor	15	0.936	0.336	TRUE
NA	Recipient	15	0.884	0.0546	TRUE

Test: wilcox.test(x, y, paired = TRUE)
Comparison: Donor πN – πS vs. Recipient πN – πS

Gene	n	statistic	p_value	significant
<fct>	<int>	<dbl>	<dbl>	<lgl>
PB2	15	35	0.169	FALSE
PB1	15	40	0.277	FALSE
PA	15	56	0.847	FALSE
HA	15	44	0.389	FALSE
NP	15	43	0.359	FALSE
NA	15	41	0.303	FALSE

Test: wilcox.test(x, mu = 0)
Comparison: πN – πS vs. mu

Gene	Group	n	statistic	p_value	label
<fct>	<chr>	<int>	<dbl>	<dbl>	<chr>
PB2	Donor	15	0	0.000718	***
PB2	Recipient	15	42	0.33	ns
PB1	Donor	15	16	0.0134	*
PB1	Recipient	15	21	0.0256	*
PA	Donor	15	29	0.0829	ns
PA	Recipient	15	23	0.0353	*
HA	Donor	15	1	0.000881	***
HA	Recipient	15	11	0.00336	**
NP	Donor	15	4	0.0016	**
NP	Recipient	15	20	0.0215	*
NA	Donor	15	27	0.0646	ns
NA	Recipient	15	33	0.135	ns

S7A Fig

Test: shapiro.test

Comparison: normality

Region	Group	n	W	p_value	is_normal
<fct>	<fct>	<int>	<dbl>	<dbl>	<lgl>
Antigenic	Donor	16	0.8	2.76E-03	FALSE
Antigenic	Recipient	16	0.683	1.11E-04	FALSE
Antigenic	All	648	0.326	2.31E-43	FALSE
Non-antigenic	Donor	16	0.741	4.90E-04	FALSE
Non-antigenic	Recipient	16	0.574	9.35E-06	FALSE
Non-antigenic	All	108	0.583	5.03E-16	FALSE

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Comparing πN – πS between various groups

Region	comparison	statistic	p_value	label
<fct>	<chr>	<dbl>	<dbl>	<chr>
Antigenic	Donor vs Recipient	81	0.528	ns
Antigenic	Donor vs All	4544	0.398	ns
Antigenic	Recipient vs All	3775	0.0626	ns
Non-antigenic	Donor vs Recipient	64	0.86	ns
Non-antigenic	Donor vs All	662	0.133	ns
Non-antigenic	Recipient vs All	766	0.467	ns

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Region	Group	n	statistic	p_value	label
<fct>	<fct>	<int>	<dbl>	<dbl>	<chr>
Antigenic	Donor	16	21	1.61E-02	*
Antigenic	Recipient	16	17	1.58E-02	*
Antigenic	All	648	47333	9.33E-13	***
Non-antigenic	Donor	16	1	5.80E-04	***
Non-antigenic	Recipient	16	18	7.63E-03	**
Non-antigenic	All	108	1245	3.26E-07	***

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Antigenic vs. non-antigenic πN – πS per group

Group	a	n	paired	statistic	p_value	label
<fct>	<list>	<list>	<lgl>	<dbl>	<dbl>	<chr>
Donor	16	16	TRUE	74	0.776	ns
Recipient	16	16	TRUE	57	0.597	ns
All	648	108	FALSE	38746	0.0737	ns

Median πN – πS values

Region	Group	median_per10k
<fct>	<fct>	<dbl>
Antigenic	Donor	-1.16
Antigenic	Recipient	-1.74
Antigenic	All	0
Non-antigenic	Donor	-1.97
Non-antigenic	Recipient	-1.54
Non-antigenic	All	-1.46

S7B Fig

Test: shapiro.test

Comparison: normality

Region	Group	n	W	p_value	is_normal
<fct>	<fct>	<int>	<dbl>	<dbl>	<lgl>
Antigenic	Donor	5	0.816	1.09E-01	TRUE
Antigenic	Recipient	5	0.767	4.24E-02	FALSE
Antigenic	All	384	0.269	3.28E-36	FALSE
Non-antigenic	Donor	5	0.65	2.66E-03	FALSE
Non-antigenic	Recipient	5	0.92	5.28E-01	TRUE
Non-antigenic	All	64	0.562	1.56E-12	FALSE

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Comparing πN – πS between various groups

Region	comparison	statistic	p_value	label
<fct>	<chr>	<dbl>	<dbl>	<chr>
Antigenic	Donor vs Recipient	5	0.625	ns
Antigenic	Donor vs All	960	1	ns
Antigenic	Recipient vs All	898	0.807	ns
Non-antigenic	Donor vs Recipient	4	0.438	ns
Non-antigenic	Donor vs All	156	0.926	ns
Non-antigenic	Recipient vs All	146	0.763	ns

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Region	Group	n	statistic	p_value	label
<fct>	<fct>	<int>	<dbl>	<dbl>	<chr>
Antigenic	Donor	5	6	0.812	ns
Antigenic	Recipient	5	5	0.625	ns
Antigenic	All	384	20960	5.73E-08	***
Non-antigenic	Donor	5	0	0.0625	ns
Non-antigenic	Recipient	5	0	0.0625	ns
Non-antigenic	All	64	239	8.63E-08	***

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Antigenic vs. non-antigenic πN – πS per group

Group	a	n	paired	statistic	p_value	label
<fct>	<list>	<list>	<lgl>	<dbl>	<dbl>	<chr>
Donor	5	5	TRUE	11	0.438	ns
Recipient	5	5	TRUE	7	1	ns
All	384	64	FALSE	14782	0.0093	**

Median πN – πS values

Region	Group	median_per10k
<fct>	<fct>	<dbl>
Antigenic	Donor	0.621
Antigenic	Recipient	-1.64
Antigenic	All	-0.467
Non-antigenic	Donor	-0.872
Non-antigenic	Recipient	-1.83
Non-antigenic	All	-1.4

S7C Fig

Test: shapiro.test

Comparison: normality

Region	Group	n	W	p_value	is_normal
<fct>	<fct>	<int>	<dbl>	<dbl>	<lgl>
Antigenic	Donor	4	NA	NA	FALSE
Antigenic	Recipient	4	0.75	3.85E-02	FALSE
Antigenic	All	132	0.391	4.45E-21	FALSE
Non-antigenic	Donor	4	0.63	1.24E-03	FALSE
Non-antigenic	Recipient	4	0.884	3.56E-01	TRUE
Non-antigenic	All	22	0.933	1.44E-01	TRUE

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Comparing πN – πS between various groups

Region	comparison	statistic	p_value	label
<fct>	<chr>	<dbl>	<dbl>	<chr>
Antigenic	Donor vs Recipient	4	0.789	ns
Antigenic	Donor vs All	290	0.655	ns
Antigenic	Recipient vs All	228	0.548	ns
Non-antigenic	Donor vs Recipient	6	0.875	ns
Non-antigenic	Donor vs All	44	1	ns
Non-antigenic	Recipient vs All	37	0.644	ns

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Region	Group	n	statistic	p_value	label
<fct>	<fct>	<int>	<dbl>	<dbl>	<chr>
Antigenic	Donor	4	0	NaN	ns
Antigenic	Recipient	4	2	0.789	ns
Antigenic	All	132	108	0.00626	**
Non-antigenic	Donor	4	0	0.089	ns
Non-antigenic	Recipient	4	0	0.125	ns
Non-antigenic	All	22	32	0.00126	**

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Antigenic vs. non-antigenic πN – πS per group

Group	a	n	paired	statistic	p_value	label
<fct>	<list>	<list>	<lgl>	<dbl>	<dbl>	<chr>
Donor	4	4	TRUE	10	0.089	ns
Recipient	4	4	TRUE	6	0.875	ns
All	132	22	FALSE	1957	0.00211	**

Median πN – πS values

Region	Group	median_per10k
<fct>	<fct>	<dbl>
Antigenic	Donor	0
Antigenic	Recipient	-0.422
Antigenic	All	0
Non-antigenic	Donor	-1
Non-antigenic	Recipient	-2.23
Non-antigenic	All	-1.6

S7D Fig

Test: shapiro.test

Comparison: normality

Region	Group	n	W	p_value	is_normal
<fct>	<fct>	<int>	<dbl>	<dbl>	<lgl>
Antigenic	Donor	15	0.764	1.33E-03	FALSE
Antigenic	Recipient	15	0.737	6.31E-04	FALSE
Antigenic	All	378	0.166	1.05E-37	FALSE
Non-antigenic	Donor	15	0.718	3.86E-04	FALSE
Non-antigenic	Recipient	15	0.773	1.68E-03	FALSE
Non-antigenic	All	63	0.698	4.30E-10	FALSE

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Comparing πN – πS between various groups

Region	comparison	statistic	p_value	label
<fct>	<chr>	<dbl>	<dbl>	<chr>
Antigenic	Donor vs Recipient	36	0.315	ns
Antigenic	Donor vs All	2426	0.266	ns
Antigenic	Recipient vs All	2367	0.204	ns
Non-antigenic	Donor vs Recipient	43	0.359	ns
Non-antigenic	Donor vs All	412	0.451	ns
Non-antigenic	Recipient vs All	398	0.345	ns

Test: wilcox.test(x, mu = 0)

Comparison: πN – πS vs. mu

Region	Group	n	statistic	p_value	label
<fct>	<fct>	<int>	<dbl>	<dbl>	<chr>
Antigenic	Donor	15	24	0.45	ns
Antigenic	Recipient	15	29	0.456	ns
Antigenic	All	378	3151	0.0387	*
Non-antigenic	Donor	15	7	0.00284	**
Non-antigenic	Recipient	15	6	0.000854	***
Non-antigenic	All	63	153	4.92E-09	***

Test: wilcox.test(x, y, paired = TRUE) for donor-recipient comparisons

Test: wilcox.test(x, y, paired = FALSE) for "All" comparisons

Comparison: Antigenic vs. non-antigenic πN – πS per group

Group	a	n	paired	statistic	p_value	label
<fct>	<list>	<list>	<lgl>	<dbl>	<dbl>	<chr>
Donor	15	15	TRUE	67	7.12E-01	ns
Recipient	15	15	TRUE	81	2.52E-01	ns
All	378	63	FALSE	18087	2.56E-13	***

Median πN – πS values

Region	Group	median_per10k
<fct>	<fct>	<dbl>
Antigenic	Donor	0
Antigenic	Recipient	-0.295
Antigenic	All	0
Non-antigenic	Donor	-2.04
Non-antigenic	Recipient	-2.4
Non-antigenic	All	-1.58

S4 Table. Genomic positions for H3 antigenic sites. Genes shown: HA (hemagglutinin) only.

Start	Stop	ID	Antigenic_Site	HA_number
429	431	Antigenic	A	122
435	437	Antigenic	A	124
441	443	Antigenic	A	126
453	455	Antigenic	A	130
456	458	Antigenic	A	131
459	461	Antigenic	A	132
462	464	Antigenic	A	133
468	470	Antigenic	A	135
474	476	Antigenic	A	137
477	479	Antigenic	A	138
483	485	Antigenic	A	140
489	491	Antigenic	A	142
492	494	Antigenic	A	143
495	497	Antigenic	A	144
498	500	Antigenic	A	145
501	503	Antigenic	A	146
513	515	Antigenic	A	150
519	521	Antigenic	A	152
567	569	Antigenic	A	168
447	449	Antigenic	B	128
450	452	Antigenic	B	129
528	530	Antigenic	B	155
531	533	Antigenic	B	156
534	536	Antigenic	B	157
537	539	Antigenic	B	158
540	542	Antigenic	B	159
543	545	Antigenic	B	160
552	554	Antigenic	B	163
555	557	Antigenic	B	164
558	560	Antigenic	B	165
621	623	Antigenic	B	186
624	626	Antigenic	B	187
627	629	Antigenic	B	188
630	632	Antigenic	B	189
633	635	Antigenic	B	190
639	641	Antigenic	B	192
642	644	Antigenic	B	193
645	647	Antigenic	B	194
651	653	Antigenic	B	196
654	656	Antigenic	B	197
657	659	Antigenic	B	198
195	197	Antigenic	C	44
198	200	Antigenic	C	45
201	203	Antigenic	C	46

204	206	Antigenic	C	47
207	209	Antigenic	C	48
213	215	Antigenic	C	50
216	218	Antigenic	C	51
222	224	Antigenic	C	53
225	227	Antigenic	C	54
882	884	Antigenic	C	273
888	890	Antigenic	C	275
891	893	Antigenic	C	276
897	899	Antigenic	C	278
900	902	Antigenic	C	279
903	905	Antigenic	C	280
945	947	Antigenic	C	294
954	956	Antigenic	C	297
960	962	Antigenic	C	299
963	965	Antigenic	C	300
975	977	Antigenic	C	304
978	980	Antigenic	C	305
984	986	Antigenic	C	307
987	989	Antigenic	C	308
990	992	Antigenic	C	309
993	995	Antigenic	C	310
996	998	Antigenic	C	311
999	1001	Antigenic	C	312
351	353	Antigenic	D	96
369	371	Antigenic	D	102
372	374	Antigenic	D	103
414	416	Antigenic	D	117
426	428	Antigenic	D	121
564	566	Antigenic	D	167
573	575	Antigenic	D	170
576	578	Antigenic	D	171
579	581	Antigenic	D	172
582	584	Antigenic	D	173
585	587	Antigenic	D	174
588	590	Antigenic	D	175
591	593	Antigenic	D	176
594	596	Antigenic	D	177
600	602	Antigenic	D	179
609	611	Antigenic	D	182
666	668	Antigenic	D	201
672	674	Antigenic	D	203
684	686	Antigenic	D	207
687	689	Antigenic	D	208
690	692	Antigenic	D	209

699	701	Antigenic	D	212
702	704	Antigenic	D	213
705	707	Antigenic	D	214
708	710	Antigenic	D	215
711	713	Antigenic	D	216
714	716	Antigenic	D	217
717	719	Antigenic	D	218
720	722	Antigenic	D	219
741	743	Antigenic	D	226
744	746	Antigenic	D	227
747	749	Antigenic	D	228
750	752	Antigenic	D	229
753	755	Antigenic	D	230
777	779	Antigenic	D	238
783	785	Antigenic	D	240
789	791	Antigenic	D	242
795	797	Antigenic	D	244
801	803	Antigenic	D	246
804	806	Antigenic	D	247
807	809	Antigenic	D	248
234	236	Antigenic	E	57
240	242	Antigenic	E	59
249	251	Antigenic	E	62
252	254	Antigenic	E	63
264	266	Antigenic	E	67
288	290	Antigenic	E	75
297	299	Antigenic	E	78
303	305	Antigenic	E	80
306	308	Antigenic	E	81
309	311	Antigenic	E	82
312	314	Antigenic	E	83
321	323	Antigenic	E	86
324	326	Antigenic	E	87
327	329	Antigenic	E	88
336	338	Antigenic	E	91
339	341	Antigenic	E	92
390	392	Antigenic	E	109
843	845	Antigenic	E	260
846	848	Antigenic	E	261
849	851	Antigenic	E	262
858	860	Antigenic	E	265

S5 Table. Genomic positions for H1 antigenic sites. Genes shown: HA (hemagglutinin) only.

Start	Stop	ID	Antigenic_Site	HA_number
280	297	Antigenic	Cb	70
442	447	Antigenic	Sa	124
481	498	Antigenic	Ca2	137
529	543	Antigenic	Sa	153
547	564	Antigenic	Sa	159
568	582	Antigenic	Ca1	166
622	657	Antigenic	Sb	184
679	687	Antigenic	Ca1	203
733	738	Antigenic	Ca2	221
775	783	Antigenic	Ca1	235