Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Nicolas Berg April 12th, 2021

Transmission of Influenza A Virus Genomic Diversity Between Human Hosts

by

Nicolas Berg

Katia Koelle
Adviser

Department of Biology

Katia Koelle

Adviser

Rustom Antia

Committee Member

Nic Vega

Committee Member

Anice Lowen

Committee Member

2021

Transmission of Influenza A Virus Genomic Diversity Between Human Hosts

By

Nicolas Berg

Katia Koelle

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2021

Abstract

Transmission of Influenza A Virus Genomic Diversity Between Human Hosts
By Nicolas Berg

Defective interfering particles (DIPs) are viruses containing a genetic abnormality wherein a large portion of the genome is deleted, rendering them incapable of independent reproduction. However, during a coinfection event between a wild-type virus and a DIP, the defective virus can hijack functional components from its counterpart and reproduce. This 'viral parasitism' has important effects on the within-host dynamics of many infections and has been shown to reduce symptom severity and infectivity. Exogenous administration of DIPs into a host is a current area of medicinal research as a promising anti-viral therapeutic. In this study, the possibility of transmission of defective interfering influenza virus is explored using whole-genome sequencing data of virus samples contained within verified household transmission pairs. Deletions characterizing DIP genomes were documented using a custom software pipeline, and various properties such as location and length were compared across groups. It was found that there was no significant relationship between transmission linkage and increased occurrence of shared deletions, lending support to the stochastic generation of identical DIPs across hosts rather than transmission.

Transmission of Influenza A Virus Genomic Diversity Between Human Hosts

By

Nicolas Berg

Katia Koelle

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2021

TABLE OF CONTENTS

Introduction:

A notable characteristic of many RNA viruses is their tendency to accumulate mutations due to the error-prone nature of their replication. These mutations include small point mutations that affect only a single nucleotide as well as large-scale deletions that fundamentally change the structure of the genome. A defective viral genome (DVG) is a type of virus genome that has been rendered inviable by a large loss-of-function mutation, often covering hundreds of nucleotides in length[1]. Viruses carrying DVGs are unable to replicate alone, but in the presence of a wild-type counterpart during a cellular coinfection event, the functional virus can serve to replicate defective segments[2]. This relationship can be parasitic in nature, and a DVG that replicates via gene products from another virus is said to become "interfering"[1] when its replication comes at the expense of the wild-type virus's replication. Virions that reproduce in this manner are called defective interfering particles (DIPs). DIPs are thought to be ubiquitously present in many infections and have been observed *in vivo* across a number of different viral species[3,4,5,6,7,8,9,10]. However, research of DIPs in human hosts is sparse, despite their potential to constrain the dynamics of wild-type virions within-host and even curtail symptoms[10].

Influenza DIPs were the first defective interfering viruses to be discovered and characterized[11]. Influenza is a negative-sense RNA virus with an eight-segmented genome. A complete copy of every segment is needed to assemble a functional progeny virus[12], so if even one segment in a virion contains a major deletion, the virus is considered defective for replication. Mutations constituting canonical DIPs generally delete a large portion of the entire

segment, leaving a stunted copy only a fraction of the original length. These deletions are often located in the central regions of the segments, away from the short packaging signals at both the 5' and 3' ends[13,14]. Prior studies have shown that although all eight segments can be rendered defective, the three segments coding for the viral polymerase complex (PB2, PB1, and PA) are the most prone to segment-spanning deletions that are characteristic of DIPs[14]. This phenomenon is not entirely understood but is thought to be tied to the length of the segments themselves[15]. As the three longest segments, PB1, PB2, and PA have the most opportunities for the viral polymerase to make an error during replication. Another possible explanation for the prevalence of defective polymerase segments is intracellular resource competition[15]. As the largest segments, the polymerase subunits require the most nucleotides to copy. Therefore, large deletions in these segments may be more prominent simply because the resulting defective segments are much faster to copy and use fewer nucleotides.

An aspect of DIP replication that is important to persistence in a host during infection is the preferential packaging of defective segments over functional ones[16]. During assembly, influenza capsids have been shown to have a greater affinity to take up incomplete genomes with intact packaging signals rather than their unaltered counterparts. This leads to favored production of defective virions during coinfection, and effectively entails that DIPs have greater fitness than the functional virus in these particular scenarios. However, because DIPs require coinfection to have any potential for reproduction, when wild-type virus is present at only low levels, DIP fitness is low. In *in vitro* studies, this leads to an oscillatory population dynamic between defective virus and wild-type virus that is reminiscent of host-parasite interactions[17].

WT virus must be present in certain levels in order to sustain growth and spread of DIPs within-host, but the rise of DIPs ultimately causes a downturn in reproduction of WT individuals and therefore slows the subsequent production of interfering virus. DIPs also are known to stimulate an immune response despite being non-infectious by themselves[18]. These factors can reduce the viral load during an infection, mitigating virulence potential and symptomatic response by the host[19]. As a result, use of inoculations of DIPs during an infection is a current topic in viral therapeutic research, and has shown great promise in ferret and mouse models[19,20].

Despite the notable impact of DIPs on disease pathology, little is known about their source during a natural infection. In the past, DIPs have been shown to arise via spontaneous errors in replication, such as non-homologous recombination or template-switching. However, the possibility of co-transmission of wild-type virus and defective interfering particles has been considered in the past, when Saira et al found an identical defective genome in a human host and another individual that person infected[9]. It was conjectured that this identical defective segment occurring across a transmission pair may indicate that influenza virus can have a coinfection of DIPs and WT variants at the very start of an infection, facilitating an event where a DIP is passed from one host to another. However, transmission bottleneck analyses of influenza have found that a typical infection results from the propagation of only a couple viral genomes[21]. Therefore, it would be counterintuitive for such a small initial exposure to result in a coinfection event. In spite of this small bottleneck, the possibility of a mechanism by which

virions remain in the same environment until host-cell uptake may raise the likelihood of co-transmission.

In this study, the prospect of host-to-host transmission of influenza defective interfering particles is investigated using retrospective analysis of deep-sequencing data collected from individuals within a cohort of household transmission pairs over the course of 2011 to 2015. Major deletions were found across all segments, with a higher incidence in the polymerase genes, consistent with prior research. DIPs were characterized by length, location, and support of deletions, and transmission pairs were probed for identical defective genomes. In addition, longitudinal samples were examined to investigate the persistence of DIPs within individuals over the course of infection. This study serves to shed light on the primary mechanisms of DIP production during natural infection of humans, and further describes the viral dynamics of DIPs between-hosts.

Methods:

Influenza Virus Sequencing Data:

The data presented here were previously analyzed by McCrone et al in 2018[21]. The samples sequenced in this study were derived from the University of Michigan School of Public Health's Household Influenza Vaccine Evaluation Study (HIVE), containing 300 households of Southeastern Michigan with two or more children monitored for respiratory illnesses of varying severity and subject to sequencing using RT-PCR upon detection of influenza. Subjects reported weekly on the potential onset of respiratory illness based on the HIVE protocol, wherein two or more characteristic symptoms of respiratory infection were required to be considered 'symptomatic'. Symptomatic individuals then gave nasal and throat swabs at an onsite clinic. Based on epidemiological data and the timing of infection/symptom onset, McCrone et al identified 37 within-household influenza A subtype H3N2 transmission pairs[21]. These 37 pairs serve as the dataset in this study. The complete dataset can be found on NCBI GenBank under the project ID PRJNA412631.

Specimens from symptomatic patients were subject to broad testing to identify influenza infections after collection. Viral RNA was tested using RT-PCR for IAV. Samples that tested positive underwent further analysis for subtyping. Primers and probes were developed by the Centers for Disease Control and Prevention Influenza Division. Samples were quality-checked using treatment with RNase P. vRNA was extracted from nasal and throat specimens and amplified across all eight segments using RT-PCR. Successful amplification was verified using gel electrophoresis. The concentrations of barcoded cDNA libraries were established by

fluorescence-based quantification. Primer dimers were isolated via a 300–500 bp band and removed. Sequencing was performed on purified libraries using an Illumina HiSeq 2500. For use as a control, artificial amplicons from a mix of eight clonal plasmids containing the different genomic segments of IAV were subject to the same sample preparation and sequenced on the same flow cell as the patient samples.

Identification of DVGs:

The following procedure was adapted from the software pipeline presented by Alnaji et al[22] for the detection of influenza virus DIPs. Modifications were made to remove PCR duplicates from the dataset.

Preprocessing and read preparation steps:

Reference files (GenBank Accession numbers: 2010-2011 Perth CY121496-503, 2012-2013 Victoria KJ942681-8, 2014-2015 New York CY207231-8) were first indexed using BowTie2 v.2.4.1 and Bowtie v1.2.1.1[23]. Read quality was then checked using FastQC v0.11.9. Adapter sequences were removed, and reads were quality-filtered using Trimmomatic v0.39[24]. Trimmomatic sheared the leading and trailing 28 nucleotides on each read and discarded any reads shorter than 75 nucleotides in total length. Reads were then subject to a local alignment step to the appropriate reference using BowTie2[23], and a new BAM file was populated with the successfully aligned reads. This BAM was sorted using SamTools v1.10[25] and duplicate reads were removed using Picard v2.25.0. The modified BAM file was then converted to FASTQ format using BedTools v2.30.0[26]. BowTie2 was then used to globally align reads to the

reference, creating separate outputs containing reads that successfully aligned and reads that did not align. BowTie2 only considered reads with an alignment score of at least -0.3 · read length.

References were chosen based on the season in which the sample was collected. For the 2010-2011 infections, a Perth/2010 H3N2 reference was used. 2011-2012 and 2012-2013 samples were aligned to a Victoria/2011 H3N2 sequence. 2014-2015 seasonal infections were compared to the New York/2015 H3N2 strain.

Mapping of deletions:

Since DVGs that typically constitute DIPs contain very large deletions, a way to identify reads that support a DVG segment is to probe for 'split reads'; reads with two fragments that map to separate sites on the genome. A single read that spans a major gap in the reference is likely due to a defective segment. In order to identify split reads, the program ViReMa v0.15 (Viral Recombination Mapper) was employed[27]. ViReMa functions by identifying individual reads that support major structural changes in a viral genome. Only the reads that failed to align in the final BowTie2 step are entered into ViReMa, as reads that successfully align in their entirety during global alignment cannot be split reads. ViReMa disregarded 'microindels', or indels less than 20 nucleotides in length, and permitted only one nucleotide mismatch. Mismatches were not tolerated at all if within 8 nucleotides of a junction site. All deletions were logged, along with total read support values, in a separate output file[22].

Data filtration:

The dataset contained samples both with and without technical replicates depending on viral titer. Samples with low amounts of viral RNA (10^3-10^5 viral genomes/mL) were sequenced in duplicate to correct for RT-PCR errors.

For samples with technical replicates, reported deletions were filtered by exclusion of those that did not appear in both replicates. For samples without technical replicates, deletions supported by only one read were excluded from analysis.

Analysis of data:

Read depth was determined on the BAM files outputted after deduplication using the depth function within SamTools. Number of viral genome reads were found using the SamTools idxstats function on the deduplicated BAM files. Normalized DVG read levels were calculated by first obtaining the number of DVG reads in each sample using ViReMa, followed by dividing total DVG reads by viral genome reads for each segment.

BLAST analysis was performed on reads suspicious of being sourced from contamination. BLAST aligns sequences to the entire NCBI database and returns the closest matches. The tool can be accessed at https://blast.ncbi.nlm.nih.gov/Blast.cgi.

Shared deletion analysis was performed on a transmission pair and non-transmission pair basis. All documented transmission pairs were searched for deletions that both pass the filtering protocol and appear in the donor and recipient samples. Occurrence of identical DVGs in transmission pairs was compared to every possible combination of unrelated infections, or

infections that were in no way epidemiologically linked. In total, 37 H3N2 transmission pairs and 1,658 non-transmission pair combinations were analyzed. Conclusions were reached on the basis of frequency of occurrence of shared DVGs.

All code involved in results analysis was written in the Python and R programming languages.

Results:

Read Depth:

The dataset consisted of 37 H3N2 influenza A virus transmission pairs, of which 3 were collected during the 2010-2011 season, 1 was collected during the 2012-2013 season, and 33 were collected during the 2014-2015 season. Deduplicated BAM files from the pipeline were probed for depth across all segments. Average depth on the genome scale was about 500 reads (**Figure 1)**, with depth tapering at the ends of the segments**.**

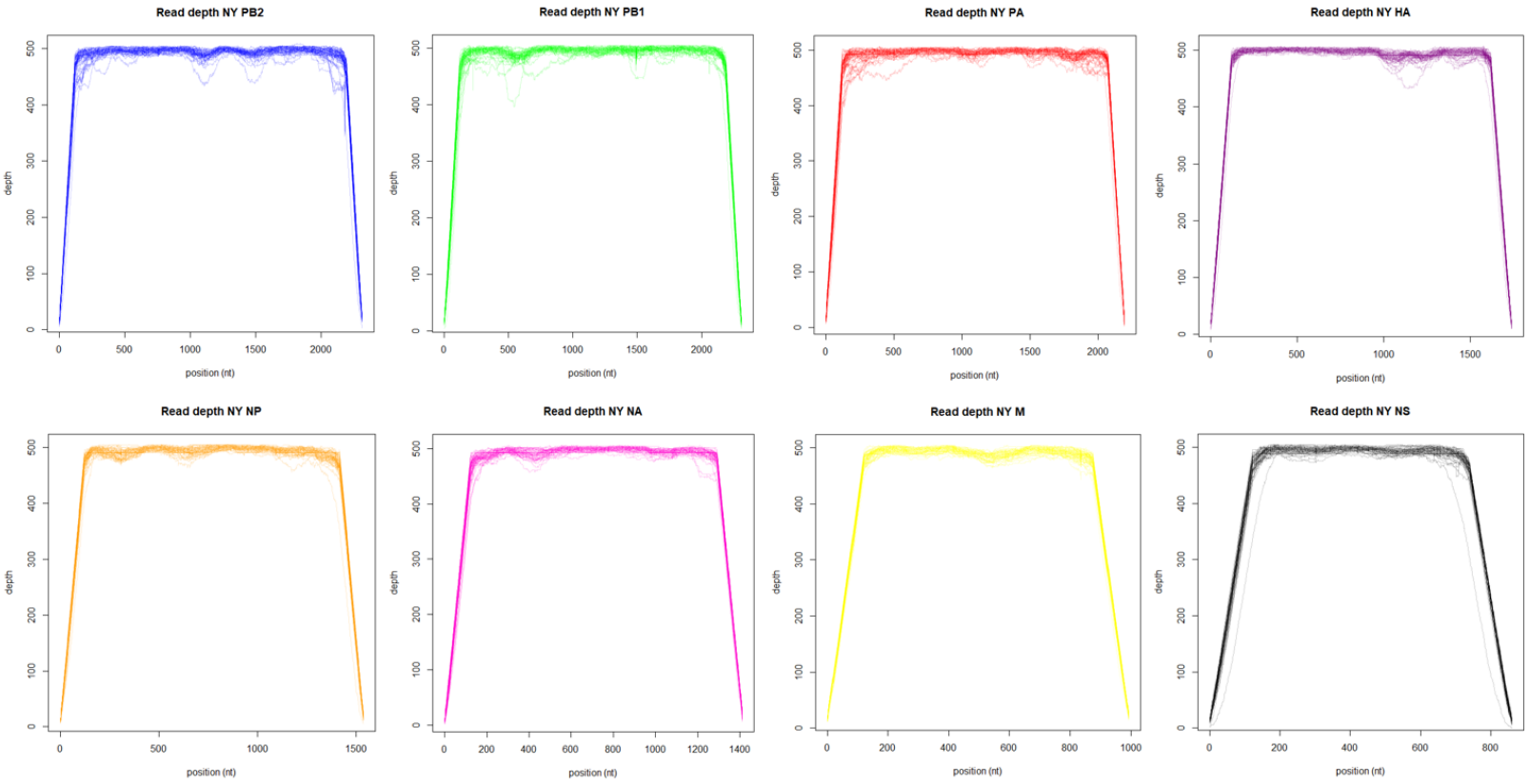


Figure 1: Read depth at each position across each gene segment for the 2014-2015 samples. Depth was calculated after local alignment step and removal of duplicate reads to better reflect results of downstream analyses.

Volume of Detected Deletions:

DVGs were located and characterized in every sample. To measure the amount of DVG reads, a normalized read count was used. As expected based on prior studies[14], levels of DVG reads were higher in the polymerase segments (PB2, PB1, PA) compared to structural and export genes (**Figure 2)**.

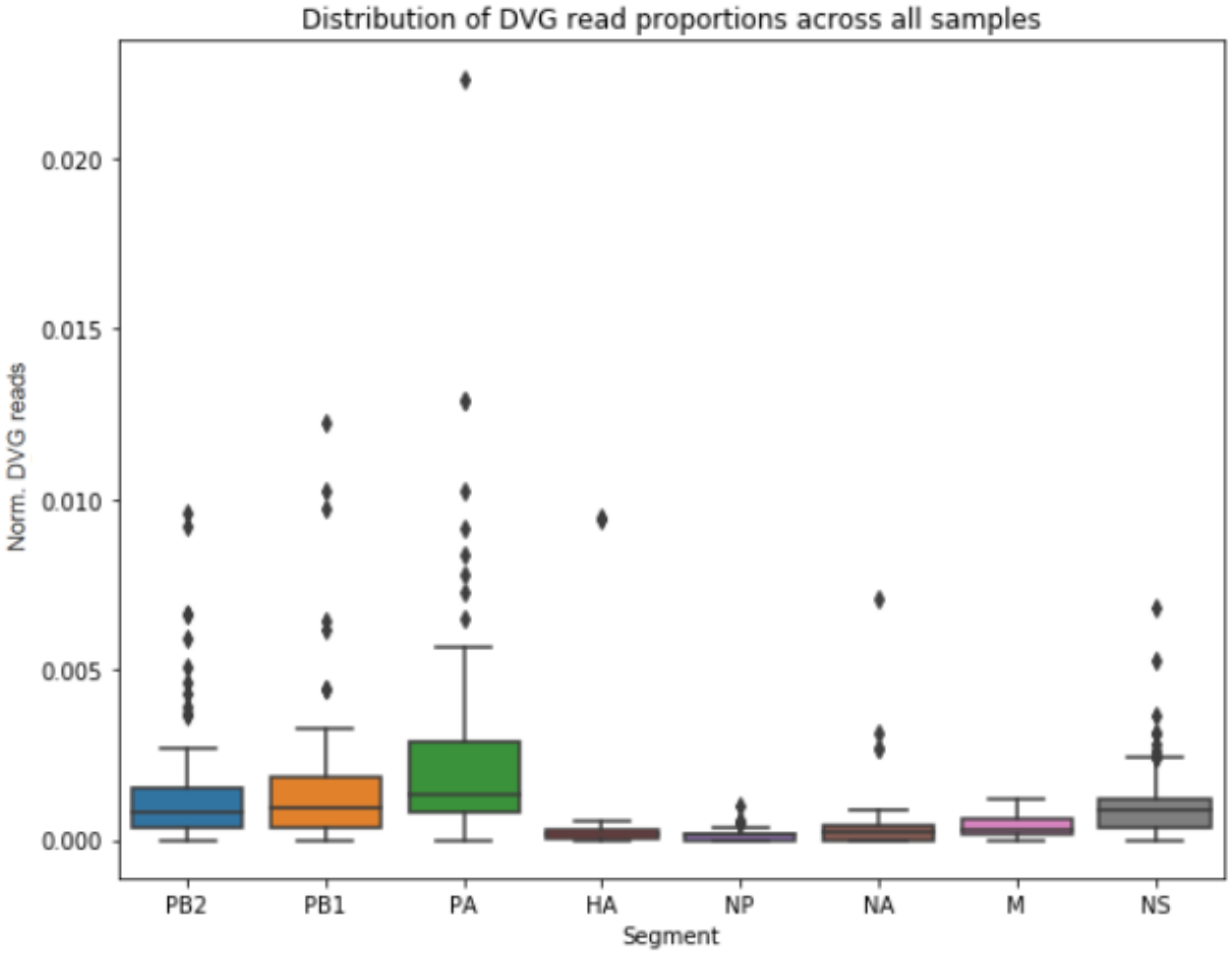


Figure 2: Average proportion of viral genome reads supporting DVGs after local alignment and removal of duplicates across all samples. Deletions were most prominent in the PA, PB2, and PB1 gene segments.

Typical proportions of DVG reads were on the order of approximately 1.5 per 1,000 in polymerase segments and 0.25 per 1,000 in other segments. Gene segment 8, however, which codes for two peptides involved in viral export from the cell nucleus and suppression of host gene expression, had a disproportionately greater amount of DVG reads compared to the other 4 non-polymerase segments, about 1 per 1,000 reads (**Figure 2,3)**.

In **Figure 3,** a sample-specific visualization of normalized DVG count can be seen. Levels of DVG reads in each segment appear to differ greatly across samples, but their occurrence in every sample is a testament to the pervasive nature of DIPs in natural infections. The large amount of variation likely is caused by the stochastic nature of within-host generation of DIPs.

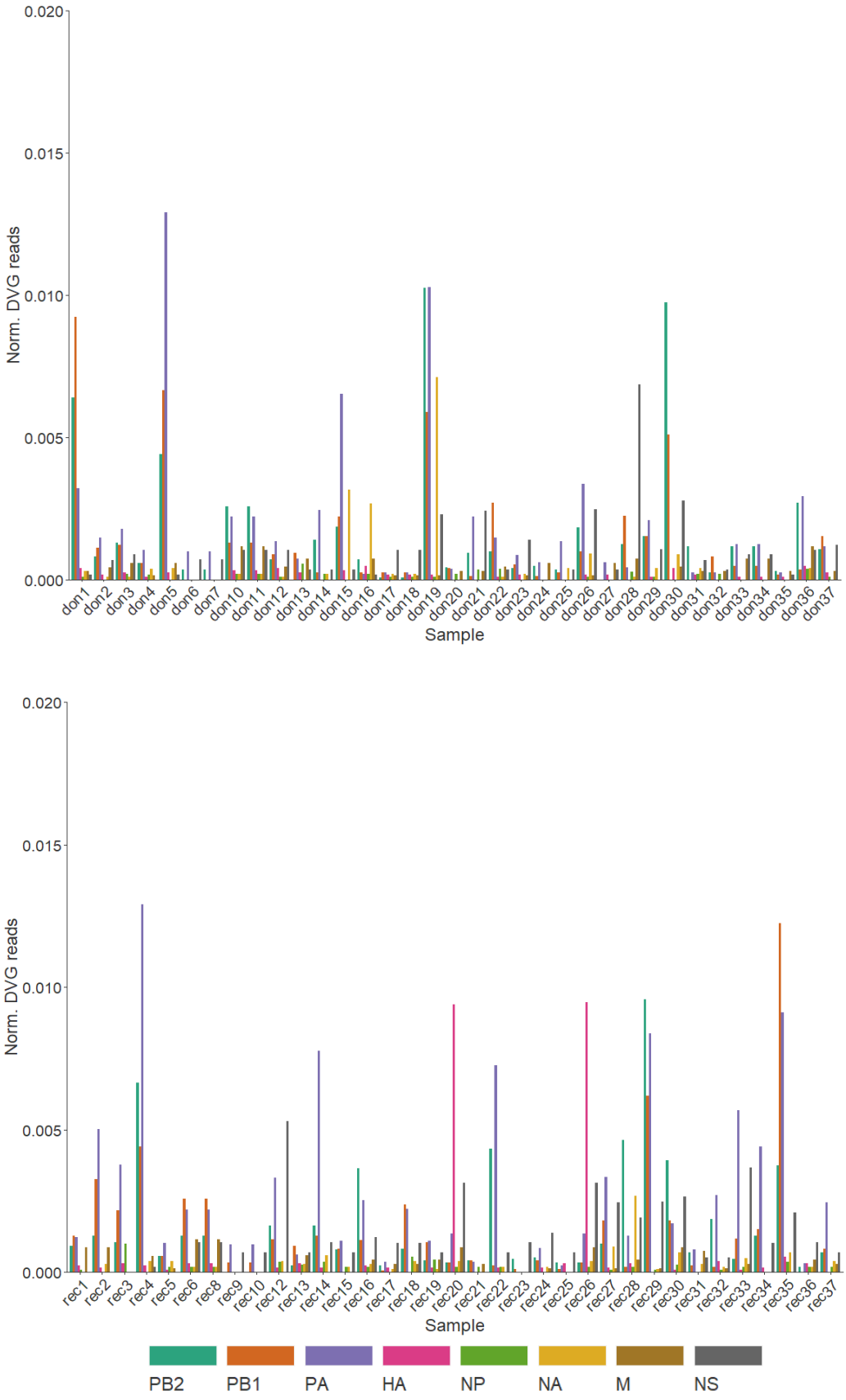
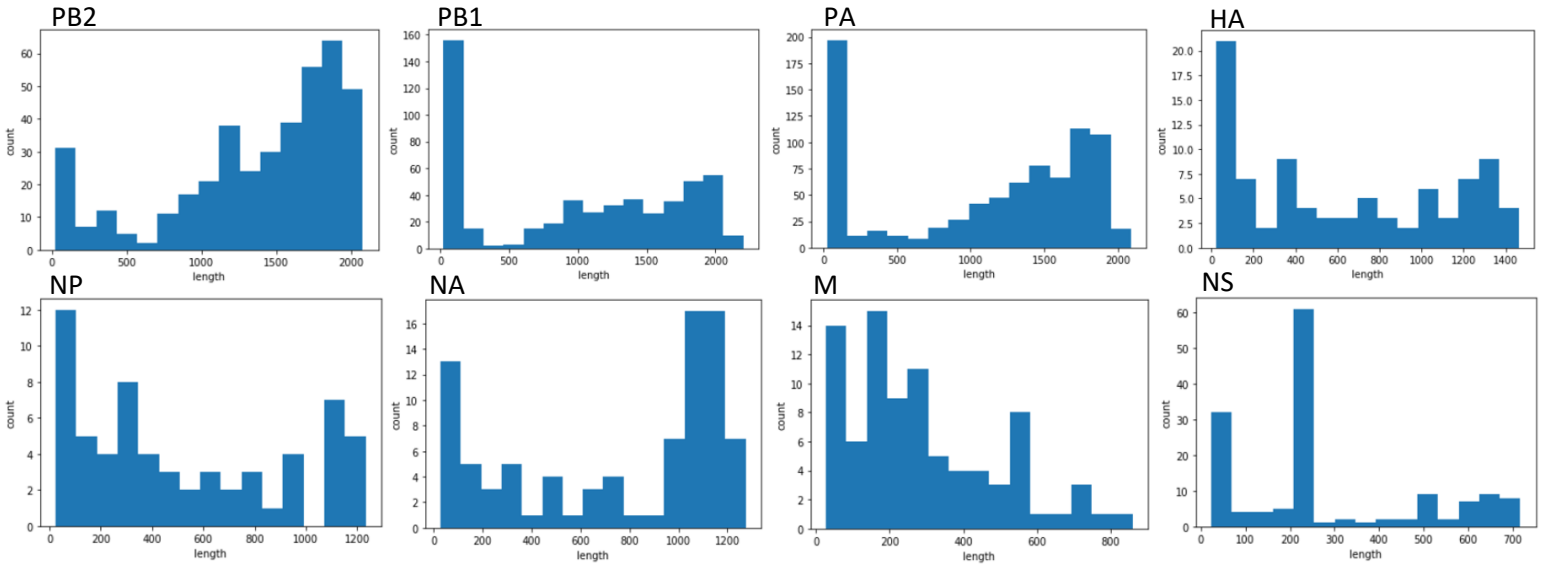


Figure 3: Proportion of viral genome reads supporting DVGs after local alignment and removal of duplicates on a sample-by-sample basis. Samples are separated by donor hosts and recipient hosts in their respective household pairs.

Lengths of Deletions:

Located deletions were characterized by start site, stop site, and read support. Canonical influenza DIPs normally have a substantial portion of the genome deleted. To visualize the lengths of detected deletions in this dataset, all deletions that were located by ViReMa and passed the filtering process can be seen in **Figure 4a**. Most segments have a bimodal distribution of lengths, with many prominent spikes around 50-100 nucleotides and peaks indicating deletions that span nearly the entire segment. **Figure 4b** displays an example junction plot from a 2010-2011 sample, illustrating the start and stop sites for all deletions that were detected.

A



B

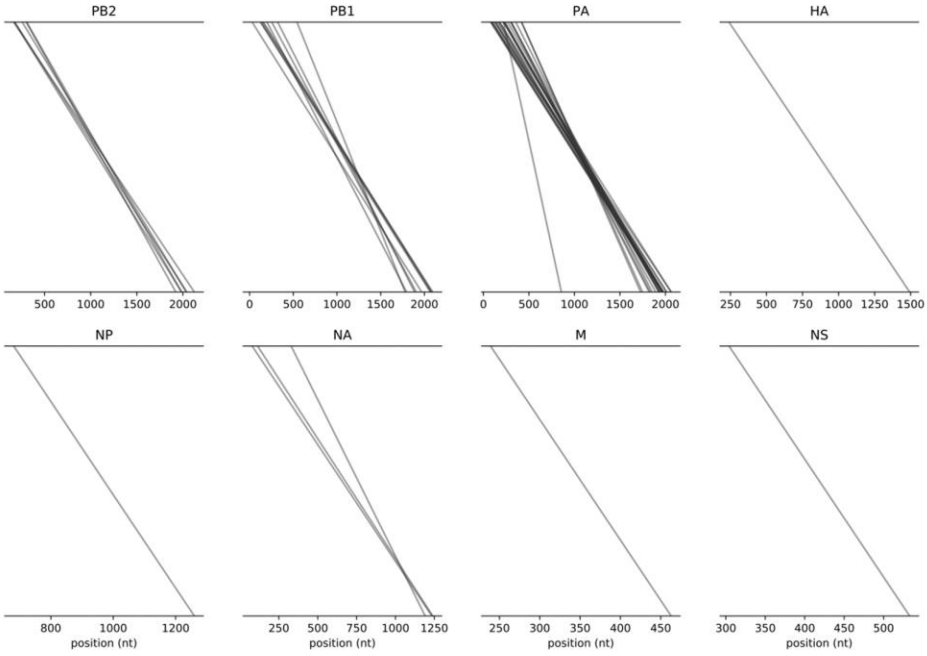


Figure 4: Summary visuals for deletion lengths in the household transmission cohort. A) Deletion length histograms across all samples. Bin width is 1/15th the length of the segment. B) Representative sample junction plot (ID SRR6121209), illustrating the individual start and stop sites for major deletions in a 2014 season natural infection. Additional junction plots can be found in the supplementary materials.

NS Segment Recurring Deletion Profile:

Based on the deletion length distributions, an anomalous pattern was noted in the NS segment. While the other segments had a large concentration of deletions in the left and rightmost bins, segment eight had an unusual peak in the middle of the distribution. Upon closer examination of the dataset, it was found that a specific deletion occurred across numerous samples in segment 8 at the same location and with higher read support than most other documented DVGs. Since other major deletions have been found to play major roles in the within-host dynamics of influenza[28], this deletion was given special focus to characterize its potential effects on resulting peptide products (**Figure 5)**.

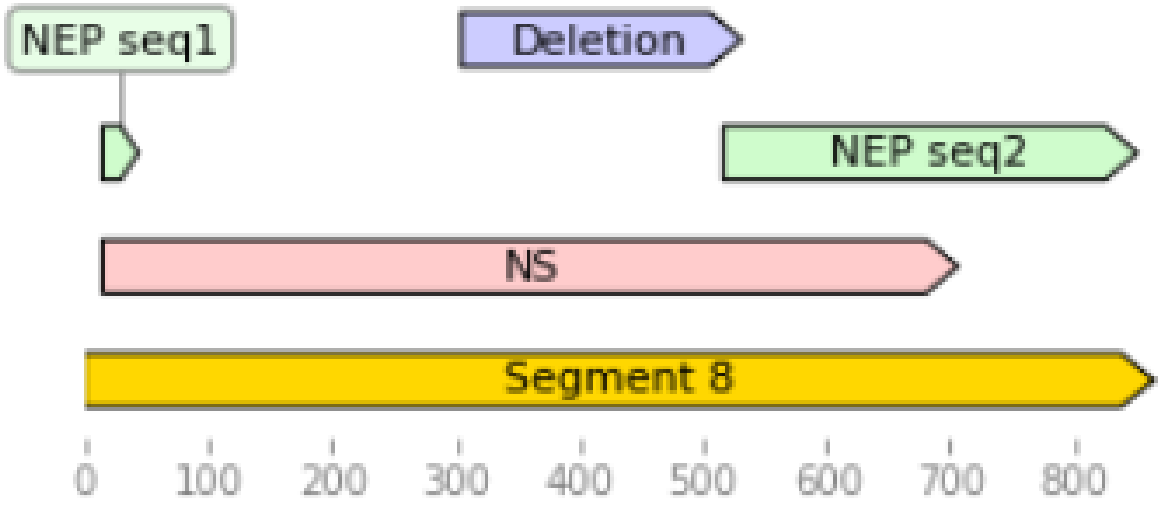


Figure 5: Schematic of deletion location relative to the reading frames of NS (a host gene expression inhibitor) and NEP (a protein enabling export of viral components from the cell nucleus) on the Perth reference.

After data filtration, the segment eight recurring deletion was found to occur in about 34% of samples in the household transmission cohort. It is 228 nucleotides in length (**Table 1**), and it located in the middle of the reading frame for NS1. It also slightly overlaps with the 5' splice site for NEP. The deletion location is in-frame with NS1. Based on a BLAST search of reads supporting this deletion, the only matches in the NCBI database were Influenza A samples, dismissing the possibility of contamination.

Reference Strain	Start Site	Stop Site
Perth	304	533
NY	301	530

Table 1: Specific locations of the segment 8 recurring deletion in the Perth and New York references. Deletion did not occur in either of the two samples using the Victoria strain reference.

Longitudinal Analysis:

The HIVE study consisted of 40 longitudinally sampled individuals. Time lag in longitudinal samples varied on a case-by-case basis from 1 to 4 days. In order to evaluate the stability of DIP populations within-host over a period of time, 17 samples from the household transmission cohort that also were part of longitudinal sampling were compared for matching deletions (**Figure 6**).

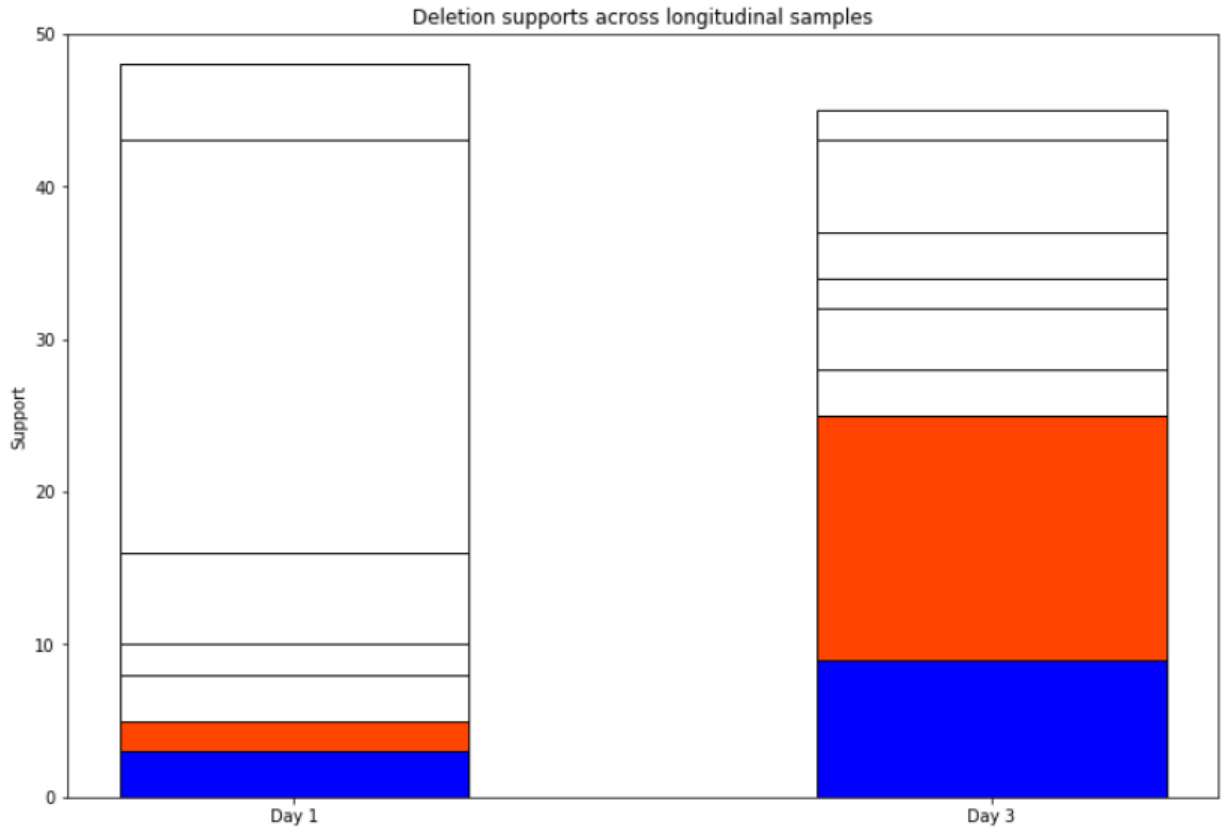


Figure 6: Representative comparison of a single individual who was sampled at symptom onset and again after a 3-day lag period. Matching colors indicated shared deletions across time points, white areas are deletions that were not shared. Additional longitudinal samples can be seen in the supplemental materials.

Longitudinal samples tended to share more identical deletions than sequences taken from differing individuals, with all but 2 samples containing at least 1 shared deletion. Of the 17 longitudinal sample sets, 9 contained 3 or more, with one individual sharing 15 deletions between time points. None of the comparisons made between unrelated individuals yielded greater than 2 shared deletions. Occurrence of shared deletions in longitudinal samples was compared to samples from different individuals using a two-sample Kolmogorov-Smirnov (K-S) test. The difference was found to be statistically significant ($p \approx 0$), indicating that shared deletions are more common in the same individual over time than in different persons. The persistence of DIP variants over extended time periods in the same host is a demonstration of the successful nature of these mutants in the presence of an adequate level of WT virus, and further validates the pipeline's fidelity through consistent detection of genomic patterns.

DVG Transmission Analysis:

To evaluate for transmission of DIPs, a comparison was conducted between the prevalence of shared deletions among verified household transmission pairs and unrelated infections within the cohort (**Figure 7**). In the non-epidemiologically linked pairs, about 78% did not have any identical deletions that passed the filtering step. Approximately 21.75% did share a single matching DVG, and 0.25% shared two deletions. In the verified transmission pairs, 74.5% had no matching deletions, 23.5% shared one deletion, and 2% matched for two DVGs. A two-sample K–S test was used to evaluate for differences in the probability mass functions of the groups and no significant relationship was found between epidemiological linkage and likelihood to have more identical DVGs ($p = 0.99$).

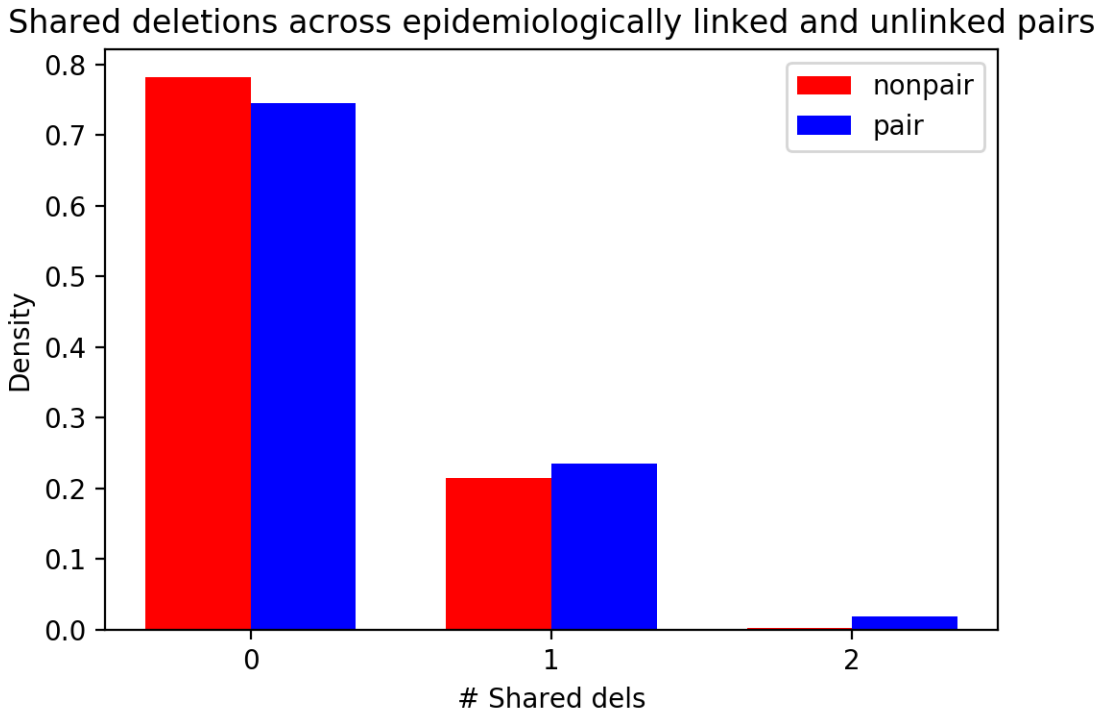


Figure 7: Proportion of transmission pairs vs epidemiologically unlinked pairs sharing a specified number of identical deletions.

These data show evidence for *de novo* generation of identical DVGs within-host after transmission. The exact mechanism by which such a substantial fraction of the tested pairs contained at least one analogous DVG is not certain but may be due to an intrinsic property of the junction sites themselves. Particular locations along the viral gene segment may be prone to polymerase errors and therefore formation of deletion junctions at these sites could be more likely.

An additional consideration that was factored into this analysis was the effect of the recurring segment 8 deletion on the proportion of samples carrying shared DVGs. Among the 37 linked pairs, 14 shared deletions were located. Of these 14, only 4 were unique while 10 were a result of the segment 8 deletion. In the 1,658 combinations of unlinked pairs, 362 shared deletions were found by the pipeline. 31 out of the 362 were unique while 331 were derived from the recurring deletion. To verify that the insignificant results from the initial test were not due to artificial inflation of shared deletions in unlinked pairs by the segment 8 DVG, the recurring deletion was filtered from the data and analyses were repeated. The differences in the two groups were again insignificant ($p = 0.96$), indicating that these results were not overwhelmingly influenced by the pervasive nature of a single mutation.

Discussion:

The existence of defective interfering particles during natural influenza A virus infections of human hosts has been documented in the past[9], and matching deletions have been noted in prior work as well. Despite this, investigation into the means by which DIPs arise during infections has yet to be performed. This work clarified the general means by which defective viral genomes arise in human hosts, as well as presenting a general overview of the DIP populations within a large dataset of human hosts.

The findings of this research support prior work on the nature and dynamics of influenza virus DIPs within single hosts. Every sample tested contained reads consistent with the presence of a DVG as reported by the detection pipeline, lending further support to the ubiquitous nature of DIPs during viral infections. Additionally, the polymerase subunit genes contained an elevated proportion of defective reads compared to other segments as previously documented. The eighth segment, often referred to as NS, contained a higher level of deletions relative to other non-polymerase genes. This is the result of a common recurring deletion with higher than average read support.

Distributions of deletion lengths across gene segments tended to follow a bimodal distribution. The majority of deletions tended to be either under 100 nucleotides in length or covered most of the entire segment. As would be expected, no deletions fell within the frame of the packaging signals at the ends of the segments as these mutations would not be viable. The prevalence of the recurring deletion in NS caused an abnormal peak in deletions of length 228, leading to further inquiries into the nature of this DVG. Based on examination of the

location of the deletion, it likely would completely disrupt expression of the nuclear export protein by cutting into a vital splicing site. However, its central location in NS and its in-frame positioning may result in a truncated version of this peptide. The potential role of this deletion during an infection is presently unknown, but its consistent appearance across individuals and even strains merits more inquiry in the future.

Longitudinal analysis of samples revealed some stability in the DIP populations within-host over time. DVGs appeared to be capable of persisting over extended periods of infection, further solidifying their reproductive viability *in vivo*. However, whether these enduring DVGs are the result of persistence or novel generation is unclear. Further testing could be done at shorter time intervals to shed light on the actual mechanism by which these DIPs appear in different stages of an infection.

Final examination of shared deletions across verified transmission pairs yielded no evidence to show that shared DVGs occur more often in transmission pairs than in unrelated infections. These results support the stochastic generation of identical DIPs rather than transmission. Although many shared deletions were found, they were equally prevalent across pairs and non-epidemiologically linked persons. However, about one quarter of the comparisons yielded at least one shared deletion, suggesting that *de novo* generation of DVGs is not entirely random, but may frequently occur at specific sites that are inclined to undergo replication errors or confer a fitness advantage to the resultant DIP compared to other deletions.

The primary limitation of this research is potential for spurious DVG reports in the output of the DIP detection pipeline. The filtering protocol was implemented in order to maximize reliability of results, but since this is a retrospective sequence analysis, the actual presence of these deletions cannot be verified via techniques like PCR. Despite this, the exclusion of reads that do not occur across replicates or have very low support should reduce computational artifacts to a minimum. In addition to potential for false readouts, there was a single sample in the transmission pair set that consistently failed to be processed by the pipeline (GenBank ID SRR6121463). Whether this is due to a software failure or something intrinsically awry with the raw data file itself is still under investigation.

Overall, this study contributes to the understanding of DIP dynamics during natural human IAV infections and lends further evidence to the universal prevalence of DVGs among wild-type counterparts. Dynamics of DIPs in humans remains an under-researched field, and further development of DVG-based antivirals could depend on knowledge of which DIPs naturally persist in an infection or generate commonly across unrelated hosts.

References:

[1] Huang a. S, Baltimore D. 1970. Defective viral particles and viral disease processes. *Nature* **226**:325–327.

[2] Nayak DP, Chambers TM, Akkina RK. 1985. Defective-interfering (DI) RNAs of influenza viruses: origin, structure, expression, and interference. *Curr. Top. Microbiol. Immunol.* **114**:103–151.

[3] Cattaneo R, Schmid A, Eschle D, Baczko K, ter Meulen V, Billeter MA. 1988. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55**:255–265.

[4] Lundquist RE, Sullivan M, Maizel JV. 1979. Characterization of a new isolate of poliovirus defective interfering particles. *Cell* **18**:759–769.

[5] Makino S, Yokomori K, Lai MM. 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *Journal of Virology* **64**:6045–6053.

[6] Petterson E, Guo T-C, Evensen Ø, Mikalsen AB. 2016. Experimental piscine alphavirus RNA recombination in vivo yields both viable virus and defective viral RNA. *Scientific Reports* **6**.

[7] Stauffer Thompson KA, Rempala GA, Yin J. 2009. Multiple-hit inhibition of infection by defective interfering particles. *Journal of General Virology* **90**:888–899.

[8] Sun Y, Jain D, Koziol-White CJ, Genoyer E, Gilbert M, Tapia K, Panettieri RA, Hodinka RL, López CB. 2015. Immunostimulatory Defective Viral Genomes from Respiratory Syncytial Virus Promote a Strong Innate Antiviral Response during Infection in Mice and Humans. *PLOS Pathogens* **11**.

[9] Saira K, Lin X, DePasse J V., Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth DE, Ghedin E. 2013. Sequence Analysis of In Vivo Defective Interfering-Like RNA of Influenza A H1N1 Pandemic Virus. *J Virol* **87**:8064–8074.

[10] Dimmock NJ, Rainsford EW, Scott PD, Marriott AC. 2008. Influenza Virus Protecting RNA: an Effective Prophylactic and Therapeutic Antiviral. *J Virol* **82**:8570–8578.

[11] Von Magnus P. 1954. Incomplete forms of influenza virus. *Adv. Virus Res*. **2**:59–79.

[12] Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. 2010. Genome packaging in influenza A virus. *J. Gen. Virol.* **91**:313–328.

[13] Davis AR, Hiti AL, Nayak DP. 1980. Influenza defective interfering viral RNA is formed by internal deletion of genomic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **77**:215–219.

[14] Davis AR, Nayak DP. 1979. Sequence relationships among defective interfering influenza viral RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **76**:3092–3096.

[15] Alnaji FG, Brooke CB. 2020. Influenza virus DI particles: Defective interfering or delightfully interesting? *PLOS Pathogens* **16**.

[16] Odagiri T, Tashiro M. 1997. Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. *J. Virol*. **71**:2138–2145.

[17] Kantorovich-Prokudina EN, Semyonova NP, Berezina ON, Zhdanov VM. 1980. Gradual changes of influenza virions during passage of undiluted material. *J Gen Virol* **50**:23–31.

[18] Tapia K, Kim W keun, Sun Y, Mercado-López X, Dunay E, Wise M, Adu M, López CB. 2013. Defective Viral Genomes Arising In Vivo Provide Critical Danger Signals for the Triggering of Lung Antiviral Immunity. *PLoS Pathog* **9**.

[19] Mann A, Marriott AC, Balasingam S, Lambkin R, Oxford JS, Dimmock NJ. 2006. Interfering vaccine (defective interfering influenza A virus) protects ferrets from influenza, and allows them to develop solid immunity to reinfection. *Vaccine* **24**:4290–4296.

[20] White, KA; Morris, TJ. 1994. "Nonhomologous RNA recombination in tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions". *Journal of Virology*. **68** (1): 14–24.

[21] McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* **7**:1–19.

[22] Alnaji FG, Holmes JR, Rendon G, Vera JC, Fields CJ, Martin BE, Brooke CB. 2019. Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. *Journal of Virology* **93**.

[23] Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

[24] Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.

[25] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.

[26] Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.

[27] Routh A, Johnson JE. 2013. Discovery of functional genomic motifs in viruses with ViReMa–a Virus Recombination Mapper–for analysis of next-generation sequencing data. *Nucleic Acids Research* **42**.

[28] Dou D, Revol R, Östbye H, Wang H, Daniels R. 2018. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology* **9**.

Supplementary Figures:

Figure 3b:

https://drive.google.com/drive/folders/1m7P65P9owNZ3r-lw-LjtnvRAfOwBaYF6?usp=sharing

Figure 5:
https://drive.google.com/drive/folders/10R38yr1ttJGQAdoiJastgG5x5EnQxoNP?usp=sharing