bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Dynamics of Influenza A H3N2 Defective Viral Genomes from a Human Challenge Study

Michael A. Martin[a,b,#], Drishti Kaul[c], Gene S. Tan[c,d], Christopher W. Woods[e], Katia Koelle[a]

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

[b]Population Biology, Ecology, and Evolution Graduate Program, Laney Graduate School, Emory University, Atlanta, GA, USA

[c]Infectious Diseases, The J. Craig Venter Institute, La Jolla, California, USA

[d]Division of Infectious Diseases, Department of Medicine, University California San Diego, La Jolla, California, USA

[d]Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, NC, USA

Running Head: Within-Host Influenza A Defective Viral Genomes

#Address correspondence to Michael A. Martin, michael.martin2@emory.edu

Abstract: 248 words

Importance: 135 words

Main text: 4,747 words

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

The rapid evolution of influenza is an important contributing factor to its high worldwide incidence. The emergence and spread of genetic point mutations has been thoroughly studied both within populations and within individual hosts. In addition, influenza viruses are also known to generate genomic variation during their replication in the form of defective viral genomes (DVGs). These DVGs are formed by internal deletions in at least one gene segment that render them incapable of replication without the presence of wild-type virus. DVGs have previously been identified in natural human infections and may be associated with less severe clinical outcomes. These studies have not been able to address how DVG populations evolve *in vivo* in individual infections due to their cross-sectional design. Here we present an analysis of DVGs present in samples from two longitudinal influenza A H3N2 human challenge studies. We observe the generation of DVGs in almost all subjects. Although the genetic composition of DVG populations was highly variable, identical DVGs were observed both between multiple samples within single hosts as well as between hosts. Most likely due to stochastic effects, we did not observe clear instances of selection for specific DVGs or for shorter DVGs over the course of infection. Furthermore, DVG presence was not found to be associated with peak viral titer or peak symptom scores. Our analyses highlight the diversity of DVG populations within a host over the course of infection and the apparent role that genetic drift plays in their population dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Importance

The evolution of influenza virus, in terms of single nucleotide variants and the reassortment of gene segments, has been studied in detail. However, influenza is known to generate defective viral genomes (DVGs) during replication, and little is known about how these genomes evolve both within hosts and at the population level. Studies in animal models have indicated that prophylactically or therapeutically administered DVGs can impact patterns of disease progression. However, the formation of naturally-occurring DVGs, their evolutionary dynamics, and their contribution to disease severity in human hosts is not well understood. Here, we identify the formation of *de novo* DVGs in samples from human challenge studies throughout the course of infection. We analyze their evolutionary trajectories, revealing the important role of genetic drift in shaping DVG populations during acute infections with well-adapted viral strains.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

Influenza defective viral genomes (DVGs) were first reported by von Magnus (1) and have since been characterized *in vivo* during high multiplicity of infection (MOI) passage studies (2–7) as well as from clinical human samples (8, 9). DVGs are classified as viral genomes harboring mutations which render them incapable of self-replication. Their propagation depends on replication by wild-type helper virus (10). Influenza DVGs are formed by large internal deletions (11), which retain the 3' and 5' untranslated regions that are necessary for replication (12–15) and virion packaging (16–18). Although DVGs have been observed in all eight influenza gene segments (8, 19, 20), they have been most commonly found in the three polymerase genes (PB2, PB1, PA) (19, 21), the longest gene segments of the influenza virus genome.

DVGs that interfere with the replication of wild-type virus have been termed defective interfering particles, or DIPs. It is thought that DIPs are either preferentially replicated (22) and/or packaged (23, 24) given their shorter length. This is thought to lead to the characteristic oscillations in the relative populations of DIP and wild-type virus during passage studies (6) as DIP populations outcompete wild-type virus initially but ultimately crash when the quantity of wild-type virus drops below that necessary to maintain DIP populations. DIPs may also contribute to immune system activation (25, 26). The presence of DVGs during the course of an infection also appears to be associated with less severe clinical outcomes (9). The use of exogenous DIPs has been proposed as a potential therapeutic for influenza, with recent animal studies demonstrating that DIPs administered prophylactically and/or therapeutically can reduce the severity of clinical disease outcomes (27–30).

Influenza A virus (IAV) DVGs have been previously detected in natural human infections from deep sequencing data (8). In this cohort study, DVGs were present in about half

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of the samples analyzed and were most common in the PB2, PB1, and PA gene segments. A limitation of this study, however, is that it offered only a cross-sectional view of IAV DVG populations. While it has been shown that Sendai virus DVG populations expand during the first 12 hours of infection in a mouse model (25), the evolution of IAV DVG populations within a human host over the course of an infection has not been well characterized.

Here, we report an analysis of IAV DVG populations identified from deep sequencing data taken over the course of infection during two longitudinal human challenge studies with different treatment cohorts. We observe the generation of *de novo* DVGs in nearly all subjects, primarily in the polymerase gene segments (PB2, PB1, and PA). DVG populations were highly variable over time in DVG species composition as well as in DVG species relative abundance. Over the course of infection, individual DVG species were observed to arise, fluctuate in abundance, as well as disappear from the DVG population. Overall, we found no trend towards decreasing diversity of DVG populations or towards shorter DVG species during the five days post challenge, likely due to the dominance of stochastic effects. Furthermore, we were unable to detect an association between DVG levels and peak viral titers, potentially due to the negative feedback between DVG and wild-type virus. Similarly, higher DVG levels were not associated with more severe symptoms. This study helps to illustrate the stochastic dynamics of DVG populations within a host during acute infection with a well-adapted viral strain, a scenario under which fitness variation in the wild-type virus population is expected to be relatively small.

Materials and Methods

Ethics statement. The procedures followed in the human challenge studies were in accordance with the Declaration of Helsinki. The studies were approved by the institutional review boards (IRBs) of Duke University Medical Center (Durham, NC), the Space and Naval Warfare

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Systems Center San Diego (SSD-SD) of the US Department of Defense (Washington, DC), the East London and City Research Ethics Committee 1 (London, UK), and the Independent Western Institutional Review Board (Olympia, WA). All participants provided written consent.

Subject enrollment and challenge study protocol. Data analyzed in this study were from two previously described human challenge studies ("study 1" indicated by three digit sample IDs and "study 2" indicated by four digit sample IDs beginning with a 5) (31–38). These studies were originally designed to assess changes in host gene expression during the course of influenza infection. Subjects were intranasally inoculated with 3.08 – 6.41 $\log_{10}(TCID_{50}/mL)$ of the challenge virus ("reference strain"). The reference strain was produced by passaging a human isolate of A/Wisconsin/67/2005 (H3N2) [GenBank accession numbers CY114381 to CY114388] three times in avian primary chicken kidney cells, 4 times in embryonated chicken eggs, and twice in GMP Vero cells.

A subset of subjects in study 2 were treated with oseltamivir on the evening of the first day post challenge ("early treatment cohort"). All study one and remaining study two received oseltamivir on the evening of the fifth day post challenge ("standard treatment cohort"). Nasal wash samples were taken at various time-points post-challenge (study 1: 0, 24, 48, 72, 96, 120, 144, 168 hours; study 2: 23, 29, 42, 53, 70, 76.5, 95, 100.5, 118, 124.5, 141.5, 148.5, 165 hours).

Time of peak viral titer was defined as the time from challenge to the earliest time point at which the maximum viral titer was reached. Duration of infection was defined as the time from challenge to the latest positive viral titer.

Modified Jackson symptom scores (41) were also collected throughout the seven days post challenge (study 1: 0, 12, 21, 36, 45, 60, 69, 84, 93, 108, 117, 132, 141, 156, and 164 hours;

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

119 study 2: 0, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96, 104, 112, 120, 132, 144, 156, 168
120 hours). Time to peak symptom score was defined as the time from challenge to the earliest time
121 point at which the maximum symptom score was reached. Duration of symptoms was defined as
122 the time from challenge to the last non-zero symptom score or to the end of follow-up,
123 whichever occurred sooner. The association between treatment cohort and clinical data was
124 assessed with Mann-Whitney *U* tests in RStudio v1.1.447 (42).
125 Previous analyses found no association between inoculum dose and probability of
126 infection. Given infection, inoculum dose was not associated with disease outcome or the
127 amount of viral shedding (34). We thus did not stratify any of our analyses by subject inoculum
128 dose.
129 *Generation of sequence data.* Samples that were IAV positive by cell culture or quantitative PCR
130 were further processed for whole genome sequencing. In brief, the eight genomic RNA segments
131 of IAV were reverse-transcribed and PCR amplified using a multi-segment RT-PCR (39) from
132 whole RNA extracted from nasopharyngeal samples. Individual samples were then barcoded
133 twice using the sequence independent single primer amplification (SISPA) method (40), which
134 involves a primer extension step with a Klenow fragment (37ºC for 60 minutes, 75ºC for 10
135 minutes and 4ºC hold) [New England Biolabs] and PCR amplification with a DNA polymerase
136 (Preheat at 94ºC for 2 minutes followed by 45 cycles of 94ºC for 30 seconds, 55ºC for 30
137 seconds and 68ºC for 30 seconds with a final extension time of 68ºC for 10 minutes and a 4ºC
138 hold) [Gotaq, Promega]. To reduce chimerism, PCR products were treated with exonuclease I
139 (37ºC for 60 minutes). Separately, a parallel SISPA was performed from the same sample set,
140 but was not treated with exonuclease I. Samples treated with or without exonuclease I were
141 pooled separately and sequenced on an Illumina HiSeq 2000 instrument (Paired-end sequencing;

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

142 2 × 100 bp read). SISPA barcoded reads were then demultiplexed and merged based on the
143 barcode sequence, followed by primer and barcode removal and quality trimming using an in-
144 house script at the JCVI. Sequencing runs with or without exonuclease I were used as technical
145 replicates.
146 *Sequence data analysis*. PCR chimeras were removed using a python v2.7 (43) script by
147 identifying forward and reverse reads from the same DNA fragment with conflicting barcodes.
148 Following the removal of chimeric reads, FastQC v0.11.3 (44) was performed on all samples to
149 ensure sequencing quality. Kraken2 v2.0.8-beta (45) with a complete RefSeq viral database was
150 used to identify reads assigned to influenza A, which were then further quality trimmed with
151 Trimmomatic v0.38 (46). Leading or trailing bases with quality < 3 were removed. Reads were
152 cut when the average quality per base in 4-base wide sliding windows was < 15, and reads with
153 less than 50 bases were excluded. Reads were aligned to the reference strain (GenBank
154 CY114381 - CY114388) using STAR v2.7.0e (47). A STAR pre-indexing string of length six
155 was used to generate genome indexing files. SAM files including only uniquely mapped reads
156 were converted to BAM files which were sorted and indexed using SAMtools v1.9 and HTSlib
157 v1.9 (48). Single-nucleotide polymorphisms (SNPs) were called using the BCFtools v1.9 (48)
158 “mpileup,” “call,” and “norm” commands. Only reads with mapping quality ≥ 255 (uniquely
159 mapped) and bases with quality ≥ 20 were used. BCFtools “consensus” was used to generate
160 sample-specific reference genomes including SNPs present in more than 50% of the high-quality
161 reads at a given position. Reads were aligned to sample-specific reference genomes using STAR
162 in basic two pass mode. BAM files including only primary alignments were generated using
163 SAMtools. PCR duplicates were marked and removed using Picard Tools v2.20.02 (49).

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Read depth and read length statistics of the final BAM files were calculated using the SAMtools "depth" and "view" commands along with a simple bash script. Read depth and length statistics were combined for both sequencing runs (with and without exonuclease) for the final analysis.

DVG identification. Split reads (reads with segments mapping to unique locations in the gene implying the presence of large internal deletions) were identified using a Python v2.7 script and pysam 0.15.2 (https://github.com/pysam-developers/pysam). Split reads with at least 15 alignment matches to the reference, a minimum of five consecutive alignment reference matches, no more than three small indels, and a minimum of 100 consecutive deleted reference bases were used to generate a filtered BAM file. Junction sites for individual DVGs were identified from the "jI" SAM tag, tabulated, and normalized to the total number of reads aligned to that gene segment (norm. DVG reads) using a bash script. Split read depth was calculated using the SAMtools "depth" command.

In order to reduce the number of spurious DVGs, we included only DVGs on a per sample-day basis which were identified in both sequencing runs (with and without exonuclease I) in the final analysis. Raw and normalized DVG read support measurements were combined between the technical replicates. All bioinformatic analyses were performed at the Pittsburgh Supercomputing Center using the Bridges resources.

We define a "DVG species" as DVGs with identical deletion breakpoints and "DVG populations" as all of the observed DVG species within a sample. DVG species are identified by the first and last reference bases deleted (first_last). DVG load for a given gene segment is defined as the sum of the normalized DVG read count for all observed DVGs. DVG load for all gene segments is the average of the normalized DVG read count over the eight genes. The effect

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of treatment cohort on peak DVG load was assessed using a Mann-Whitney *U* test calculated in RStudio. The association between DVG presence on different gene segments was assessed using Fisher's exact test calculated in RStudio.

DVG diversity calculation. To evaluate the degree of DVG diversity within a sample while adjusting for the variable number of observed DVG species, we utilized Pielou's evenness index (50), given by $\frac{-\sum_{i=1}^{s} p_i \ln p_i}{\ln s}$ where s is the number of DVG species and p_i is the proportion of DVG reads which support that DVG species. An evenness of 1 corresponds to a population in which all observed species are present at the same frequency. This metric was calculated in RStudio.

All figures were generated in RStudio using ggplot2 v3.1.0 (51) and cowplot v0.9.3 (52).

Raw sequencing data are accessible under NCBI BioProject PRJNA577644. Scripts used for the generation of data and figures in this report are available at https://github.com/koellelab/IAV_human_challenge_study_code.

Results

Data summary. Of the 37 participants in the human challenge studies, 17 were successfully infected and had at least one sample successfully sequenced. Seven of these 17 individuals belonged to the early treatment cohort; the remaining ten belonged to the standard treatment cohort. Peak viral titers ranged from 1.75 to 6.25 $\log_{10}$(TCID$_{50}$/mL) (mean [population standard deviation (sd)]: 4.5 [1.2] $\log_{10}$(TCID$_{50}$/mL)) and occurred 24 to 120 hours post-challenge (mean [sd]: 58 [26] hours)). Peak viral titer did not appear to differ between the early and standard treatment group (mean [sd]: 4.3 [1.5] v. 4.6 [0.8] $\log_{10}$(TCID$_{50}$/mL); p-value = 0.922). However, those in the early treatment cohort tended to reach peak viral titer faster (cohort (mean [sd]: 48

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[20] v. 65 [27] hours; p-value = 0.080) and tended to have shorter durations of infection (mean [sd]: 74 [22] v. 118 [37] hours; p-value = 0.035; Figure S1a, Table S1, Table S2).

Peak total symptom scores ranged from 1 to 16 (mean [sd]: 6.5 [4.8]) and did not differ significantly between treatment cohorts (p-value = 0.922). Time to peak symptom score was shorter in the early treatment cohort (mean [sd]: 29 [14] v. 59 [18]; p-value = 0.003) as was the duration of symptoms (mean [sd]: 83 [37] v. 134 [17]; p-value: 0.005). Cumulative symptom scores were highly variable between subjects and no difference was observed between the two cohorts (mean [sd]: 43 [50] v. 35 [29]; p-value = 0.922; Figure S1b, Table S1, Table S2).

A total of 43 samples, including the inoculum, were successfully deep-sequenced. The number of successfully sequenced samples per subject ranged from one to five (Figure 1A). Following read trimming, the per-sample, per-gene average read length ranged from 70 to 72 nucleotides (nt). The average genome-wide read depth was 118 reads (range: 63 to 166) (Table S3, Figure S2A, Figure S2B).

Widely observed de novo DVGs. DVGs in the challenge stock were largely absent, identified at only low levels in the NP, NA, and NS gene segments and absent from the other gene segments (Figure S3). On the contrary, DVGs were observed in all but one successfully infected subject (Figure 1). For the subject in which no DVGs were detected (subject 5017), only a single sample (day two) was sequenced. This subject was in the early treatment cohort and was also positive for influenza on day three, however, this sample was not successfully sequenced.

Amongst the other subjects, we observed DVGs in at least one of IAV's eight gene segments in 38/41 successfully sequenced samples. As expected, DVGs were more commonly observed in the polymerase genes (PB2 (n = 31), PB1 (n = 28), and PA (n = 26) v. HA (n=7), NP (n = 7), NA (n = 12), M (5), and NS (10)). DVGs were observed as early as day one and as late

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

as day six post challenge. We did not observe a difference in peak DVG load between treatment cohorts (p-value: 0.713).

Normalized DVG read counts, a proxy for the number of DVGs relative to wild-type virus *in vivo*, ranged from 0.0004 to 0.064 (Table S4, Figure S2C). Among samples with DVGs, deletions in polymerase genes tended to have higher normalized DVG counts, indicating the presence of more DVGs generated from these segments within a given host. Presence of DVGs in the PB2, PB1, and PA gene segments were positively associated with one another (PB2 × PB1 p-value: 2.5×10^{-6}; PB2 x PA p-value: 5.2×10^{-3}; PB1 x PA p-value: 2.5×10^{-3}; Figure S4).

To determine whether certain junction locations were favored in identified DVG species, we tabulated the most commonly observed 3' and 5' junction sites (Figure S5, Table S5). Unsurprisingly, most 3' junction locations were located in the first 500 nt of each gene and the 5' junction locations in the last 500 nt of each gene. We observed no junction locations within 40 nt of either end of the three polymerase genes, consistent with the theory that the sequences at either end are necessary for replication (12–15) and virion packaging (16–18). The mean [sd] (weighted by normalized DVG read support) number of deleted nucleotides was comparable between gene segments (PB2: 1646 [392]; PB1: 1625 [397], PA: 1593 [332]). A small number of DVGs were observed with 3' junction sites located towards the 5' end of a given gene segment. With the exception of a single PB2 DVG, these tended to be found in a small number of samples with low normalized DVG read support. However, 1482_2101 in PB2 was observed in 0.0045 and 0.0068 normalized reads in subject 5020 on days two and four, respectively. This was the dominant DVG present at day two and one of a number of codominant DVGs present at day four.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We observed the presence of identical DVG species across both samples and subjects in the PB2 (n = 16), PB1 (n = 13), PA (n = 13), NP (n = 2), NA (n = 2), and NS (n = 1) genes (Table S6). Repeat DVGs were most commonly present in multiple samples from the same subject (n = 37). For example, the PB2 deletion from nt 356 to nt 1937 in the reference sequence was observed in subject 5021 at four consecutive time points (day two through day five). However, a number were also present in multiple subjects (n = 10). For example, DVG 476_703 in the PB2 gene segment was observed in subject 5006 at day one, subjects 5019 and 5020 at day 2, and subject 5021 at day three. DVG 696_1378 in the NP gene segment was observed in the challenge stock as well as in subject 5002 (day two) and subject 5019 (day two and three, but not one). Given its identification in only these 4 samples, it is unclear whether this DVG was transmitted during challenge or whether it appeared *de novo* in these two subjects (with lack of detection on day one in subject 5019).

Dynamic within-host DVG populations. Due to the longitudinal nature of these data, we wished to analyze whether systematic changes in the DVG populations within individual hosts were evident. Specifically, we looked at the population composition within hosts to determine whether there was evidence of positive selection for specific DVG species or for changes in the composite characteristics of DVG populations. However, given the acute nature of the infections in this study, any positive selection may be overwhelmed by stochastic effects, as has been previously described for point mutations in acute influenza infections (53).

Our analyses revealed that the composition of DVG populations changed rapidly within-hosts. Individual DVG species were found to rise and fall in their relative read support, and DVG species arose and disappeared throughout the course of infection (Figure S6). For example, in subject 013 the number of individual DVG species increased notably between days two and

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

three, more than doubling in the PB2 and PB1 genes (Figure 2A). However, in general the DVGs observed at day two were still present in the sample at day three. Considerably different dynamics were observed in subject 5021 (Figure S6), from which we have data for day one through four. DVG populations in this subject underwent considerable turnover on day three. Amongst the two PB2 DVGs observed at day two, neither were observed at day three. However, one of these two was observed again at day four. Similar dynamics were observed in both the PB1 gene (in which no DVGs were observed at day three) and the PA gene (in which a unique DVG was observed only at day three).

The generation of novel DVG species between sampling timepoints was very common. However, in most cases at any given time point a small number of DVG species accounted for the majority of the relative read support. In certain cases, these dominant DVG species were consistent between time points, however in others the dominant DVG species varied considerably between time points. This suggests that while many different DVG species can be formed during viral replication, stochastic effects during DVG generation or a selective advantage of certain DVG species leads to the observed DVG species unevenness within hosts at any given time point.

To quantitatively assess whether there was evidence for selection for specific DVG species, we determined the trajectory of DVG diversity, measured by Pielou's evenness (J'), over the course of infection in individual subjects (Figure 2B). We did not observe a trend towards decreasing diversity over the course of infection. For example, amongst PB2 DVGs, subject 5004 as well as subject 5021 witnessed net decreases in DVG diversity overtime whereas there was limited change in J' for subjects 012 and 5006. Subject 5019 experienced only a transient reduction in diversity on day two when the DVG population was dominated by a single DVG

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(172_2079). In contrast, subject 5020 experienced a transient increase in diversity on days two and three. Similar stochastic patterns of diversity were also observed in PB1 and PA DVGs.

It has been proposed that DVG species are preferentially replicated over wild-type virus due to their shorter length (22). Therefore, it is reasonable to hypothesize that selection might lead to the evolution of shorter DVG species over the course of infection. To address this hypothesis, we analyzed the number of reference bases deleted over the course of infection amongst those subjects with DVGs in a specific gene at multiple time points (Figure S7). Our data indicated no systematic change in DVG length over the course of infection (Figure 2C, Figure S8). In certain subjects, DVGs tended to get shorter (subject 5004 PB2, subject 5019 PB1, subject 5020 PB2), however in others there was no discernible change in DVG length (subject 012 PB2 and PB1, subject 5002 PA). These results suggest that either genetic drift overwhelms selective forces or that factors other than DVG length more strongly affect DVG fitness.

Correlation between DVG presence and clinical data. While this study was not powered to detect statistically significant associations between the presence of DVGs and clinical data, we wished to see if there were any qualitative correlations. We first analyzed the relationship between peak viral titer and the peak DVG load within a subject (Figure 3A), expecting a positive correlation as wild-type virus is necessary for the replication of DVGs. We were unable to detect an association between the two measures in this study. The cyclical nature of relative DVG abundance is well established *in vitro* (6) and it is possible that while wild-type virus is necessary for DVG replication, the inhibitory effect of DVG replication on the amount of wild-type virus which is replicated (thereby reducing peak viral titers) obscures any obvious correlation between the two.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Previous reports have suggested an associated between DVG presence and IAV clinical outcomes (9). While none of the subjects in this controlled challenge study experienced severe clinical outcomes, we analyzed the association between self-reported symptom scores and DVG presence. As DVGs are thought to dampen the clinical manifestation of IAV infection, we expected high DVG levels to be associated with less severe symptom scores. However, we were unable to detect an association between peak DVG load and peak symptom score in this study (Figure 3B). This lack of association may be due to the contrasting effects of DVGs interfering with wild-type virus replication and packaging and their activation of the innate immune response (25, 26), which is known to be responsible for influenza symptom manifestation. Furthermore, our inability to find an association between DVG load and symptoms may be because the seasonal IAV strain used in this study is relatively avirulent and all subjects were healthy, leading to relatively mild clinical presentations.

Discussion

The presence of defective influenza genomes has been well characterized in cell cultures (2–7) and animal models (25). Influenza DVGs have also been observed in clinical human H1N1 samples (8). However, to date, no studies have performed longitudinal analyses of naturally occurring DVG populations within humans. Here we presented an analysis of DVG populations in samples from two H3N2 human challenge studies with different treatment protocols for up to seven days post-challenge.

Our analysis supports prior findings that DVG presence is nearly universal and most commonly found on the polymerase gene segments (8), both in terms of presence/absence as well as abundance. Furthermore, we observed that DVGs are often found on multiple polymerase genes within the same subject. The timing of oseltamivir treatment was not found to affect the

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

peak viral titer, peak symptom score, or cumulative symptom scores, nor did it effect the accumulation of DVGs within a host. We observed identical DVG species within single hosts at multiple time points as well as across multiple hosts. As identical DVG species were more likely to be observed within, as opposed to between hosts, this implies that ongoing within-host replication of DVG species following their stochastic generation is likely driving this phenomenon.

DVG populations were shown to be to be highly dynamic in terms of both the DVG species they comprised and the abundance levels of these species. There was no evidence for decreased diversity of DVG populations within a host over the course of infection. Furthermore, we saw no trend towards DVG species becoming on average shorter over time. These results imply that *in vivo* genetic drift may be overwhelming selective forces in shaping the evolutionary dynamics of DVG species in this study. IAV genetic drift playing a strong role in these human challenge studies is not unanticipated, given that the challenge reference strain was a seasonal influenza strain that was relatively well adapted to human hosts and that egg- and cell culture-adapted variants were quickly excluded from the *in vivo* viral populations (31). The effect of spatial structure within the host respiratory system (54) may further augment the effects of genetic drift on DVG populations.

We did not observe an association between peak viral titer or peak symptom score and peak DVG loads. This points to the complex feedback mechanisms which govern the amount of DVG and wild-type virus within a host as well as between the replicative inhibitory effect of DVG generation on wild-type virus replication and the interaction between DVGs and the host immune response.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

This analysis has several limitations, largely due to the nature of the available data. The sequencing reads were generated in 2013 and therefore read lengths are shorter and mean read coverage is lower than in more recently generated viral deep sequencing datasets. Furthermore, we did not confirm the presence of DVG species using PCR, as has been done in other studies (8) because samples from these human challenge studies are no longer available. Furthermore, a certain level of noise in the bioinformatics pipeline used to identify DVGs is to be expected. In order to reduce this noise we analyzed only DVG species present in both technical replicates (with and without exonuclease I), however, we opted not to remove DVG species with low supporting read counts as has been previously proposed (55) in order to maintain sensitivity in our measure of DVG diversity. The very low number of split reads observed in the non-polymerase genes, which are known to rarely form DVGs, implies that the level of noise in our analysis is relatively low.

Furthermore, sequencing data were only available at most once per day for each subject and therefore we were unable to assess the fine-scale evolution of DVG species. This sparse sampling is likely why observed DVG populations were so variable between time points. With more frequent sampling we predict it would be possible to observe more gradual transitions between DVG population compositions within a host.

Despite these limitations, this study adds to the growing body of evidence that influenza DVGs are present during human infections and evolve over the course of infection. While the expansion of DVGs in individual infections are surely impacted by wild-type viral dynamics, whether DVGs in turn play a role in shaping infection dynamics and determining disease progression remains an open question.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Further studies with greater temporal resolution and sequencing to a higher read depth may help to more precisely characterize the evolutionary trajectory of DVG populations within individual hosts. Analysis of the most common DVG species observed in future studies may reveal factors that impact DVG stability over the course of an infection. A thorough understanding of the interaction between wild-type virus, DVGs, and the host immune response may ultimately aid in the development of therapeutics based on exogeneous DIPs.

Acknowledgements

We thank Ashley Sobel Leonard for guidance in accessing and interpreting the available sequence data; Katherine E. E. Johnson, Tim Song, and Elodie Ghedin for explaining their DVG identification pipeline, which was adapted for this analysis; Fadi G. Alnaji and Christopher B. Brooke for their advice on the identification of DVGs from Illumina sequencing data; and David E. Wentworth for his assistance in interpreting the JCVI sequencing methods used to generate the data analyzed here. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges Regular Memory resource at the Pittsburgh Supercomputing Center (56) and was funded by the US Defense Advanced Research Projects Agency (DARPA) INTERCEPT W911NF-17-2-0034 contract. The funding for the challenge studies was provided by DARPA contract NC66001-07-C-2024. Work was partially funded by the National Institute of Allergy and Infectious Diseases, National Institute of Health award numbers HHSN272200900007C and U19AI110819).

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. von Magnus P. 1954. Incomplete forms of influenza virus. Adv Virus Res 2:59–79.

2. Davis AR, Nayak DP. 1979. Sequence relationships among defective interfering influenza viral RNAs. Proc Natl Acad Sci U S A 76:3092–6.

3. 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.

4. Marcus PI, Ngunjiri JM, Sekellick MJ. 2009. Dynamics of Biologically Active Subpopulations of Influenza Virus: Plaque-Forming, Noninfectious Cell-Killing, and Defective Interfering Particles. J Virol 83:8122–8130.

5. Janda JM, Davis AR, Nayak DP, De BK. 1979. Diversity and generation of defective interfering influenza virus particles. Virology 95:48–58.

6. 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.

7. Nayak DP, Tobita K, Janda JM, Davis a R, De BK. 1978. Homologous interference mediated by defective interfering influenza virus derived from a temperature-sensitive mutant of influenza virus. J Virol 28:375–86.

8. 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.

9. Vasilijevic J, Zamarreño N, Oliveros JC, Rodriguez-Frandsen A, Gómez G, Rodriguez G,

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pérez-Ruiz M, Rey S, Barba I, Pozo F, Casas I, Nieto A, Falcón A. 2017. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. PLoS Pathog 13:1–29.

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

11. 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.

12. Li X, Palese P. 1992. Mutational analysis of the promoter required for influenza virus virion RNA synthesis. J Virol 66:4331–4338.

13. Piccone ME, Fernandez-Sesma A, Palese P. 1993. Mutational analysis of the influenza virus vRNA promoter. Virus Res 28:99–112.

14. Flick R, Gabriele N, Hoffmann E, Neumeier E, Hobom G. 1996. Promoter elements in the influenza vRNA terminal structure. RNA 2:1046–1057.

15. Crow M, Crow M, Deng T, Deng T, Addley M, Addley M, Brownlee GG, Brownlee GG. 2004. Mutational Analysis of the Influenza Virus cRNA Promoter and Identi cation of Nucleotides Critical for Replication. Society 78:6263–6270.

16. Watanabe T, Watanabe S, Noda T, Fujii Y, Kawaoka Y. 2003. Exploitation of Nucleic Acid Packaging Signals To Generate a Novel Influenza Virus-Based Vector Stably Expressing Two Foreign Genes. J Virol 77:10575–10583.

17. Fujii K, Fujii Y, Noda T, Muramoto Y, Watanabe T, Takada A, Goto H, Horimoto T, Kawaoka Y. 2005. Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions. J Virol 79:3766–3774.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

454 18. Liang Y, Hong Y, Parslow TG, Liang Y, Hong Y, Parslow TG. 2005. cis -Acting
455 Packaging Signals in the Influenza Virus PB1 , PB2 , and PA Genomic RNA Segments cis
456 -Acting Packaging Signals in the Influenza Virus PB1 , PB2 , and PA Genomic RNA
457 Segments. J Virol 79:10348–10355.
458 19. Jennings PA, Finch JT, Winter G, Robertson JS. 1983. Does the higher order structure of
459 the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral
460 RNA? Cell 34:619–627.
461 20. Noble S, Dimmock NJ. 1995. Characterization of putative defective interfering (DI)
462 A/WSN RNAs isolated from the lungs of mice protected from an otherwise lethal
463 respiratory infection with influenza virus A/WSN (H1N1): A subset of the inoculum DI
464 RNAs. Virology.
465 21. Brooke CB. 2014. Biological activities of "noninfectious" influenza A virus particles.
466 Future Virol 9:41–51.
467 22. Wu C a, Harper L, Ben-Porat T. 1986. Molecular basis for interference of defective
468 interfering particles of pseudorabies virus with replication of standard virus. J Virol
469 59:308–17.
470 23. Duhaut SD, McCauley JW. 1996. Defective RNAs inhibit the assembly of influenza virus
471 genome segments in a segment-specific manner. Virology 216:326–337.
472 24. Odagiri T, Tashiro M. 1997. Segment-specific noncoding sequences of the influenza virus
473 genome RNA are involved in the specific competition between defective interfering RNA
474 and its progenitor RNA segment at the virion assembly step. J Virol 71:2138–45.
475 25. Tapia K, Kim W keun, Sun Y, Mercado-López X, Dunay E, Wise M, Adu M, López CB.
476 2013. Defective Viral Genomes Arising In Vivo Provide Critical Danger Signals for the

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Triggering of Lung Antiviral Immunity. PLoS Pathog 9.

26. Scott PD, Meng B, Marriott AC, Easton AJ, Dimmock NJ. 2011. Defective interfering influenza virus confers only short-lived protection against influenza virus disease: Evidence for a role for adaptive immunity in DI virus-mediated protection in vivo. Vaccine 29:6584–6591.

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

28. 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.

29. Noble S, Dimmock NJ. 1994. Defective interfering type A equine influenza virus (H3N8) protects mice from morbidity and mortality caused by homologous and heterologous subtypes of influenza A virus. J Gen Virol 75:3485–3491.

30. Noble S, McLain L, Dimmock NJ. 2004. Interfering vaccine: A novel antiviral that converts a potentially virulent infection into one that is subclinical and immunizing. Vaccine 22:3018–3025.

31. Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, Ransier A, Stockwell TB, Das SR, Gilbert AS, Lambkin-Williams R, Ginsburg GS, Woods CW, Koelle K. 2016. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. J Virol 90:11247–11258.

32. Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, Ransier

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A, Stockwell TB, Das SR, Gilbert AS, Lambkin-Williams R, Ginsburg GS, Woods CW, Koelle K, Illingworth CJR. 2017. The effective rate of influenza reassortment is limited during human infection. PLoS Pathog 13:1–26.

33. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO, Lucas J, Huang Y, Turner R, Gilbert A, Lambkin-Williams R, Øien NC, Nicholson B, Kingsmore S, Carin L, Woods CW, Ginsburg GS. 2009. Gene Expression Signatures Diagnose Influenza and Other Symptomatic Respiratory Viral Infections in Humans. Cell Host Microbe 6:207–217.

34. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, Øien NC, McClain MT, Varkey JB, Nicholson B, Carin L, Kingsmore S, Woods CW, Ginsburg GS, Hero AO. 2011. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. PLoS Genet 7.

35. Moody MA, Zhang R, Walter EB, Woods CW, Ginsburg GS, McClain MT, Denny TN, Chen X, Munshaw S, Marshall DJ, Whitesides JF, Drinker MS, Amos JD, Gurley TC, Eudailey JA, Foulger A, DeRosa KR, Parks R, Meyerhoff RR, Yu JS, Kozink DM, Barefoot BE, Ramsburg EA, Khurana S, Golding H, Vandergrift NA, Alam SM, Tomaras GD, Kepler TB, Kelsoe G, Liao HX, Haynes BF. 2011. H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. PLoS One 6.

36. Zaas AK, Burke T, Chen M, McClain M, Nicholson B, Veldman T, Tsalik EL, Fowler V, Rivers EP, Otero R, Kingsmore SF, Voora D, Lucas J, Hero AO, Carin L, Woods CW, Ginsburg GS. 2013. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. Sci Transl Med 5:1–10.

37. Woods CW, McClain MT, Chen M, Zaas AK, Nicholson BP, Varkey J, Veldman T,

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Kingsmore SF, Huang Y, Lambkin-Williams R, Gilbert AG, Hero AO, Ramsburg E, Glickman S, Lucas JE, Carin L, Ginsburg GS. 2013. A Host Transcriptional Signature for Presymptomatic Detection of Infection in Humans Exposed to Influenza H1N1 or H3N2. PLoS One 8.

38. Wilkinson TM, Li CKF, Chui CSC, Huang AKY, Perkins M, Liebner JC, Lambkin-Williams R, Gilbert A, Oxford J, Nicholas B, Staples KJ, Dong T, Douek DC, McMichael AJ, Xu X-N. 2012. Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. Nat Med 18:274.

39. Zhou B, Donnelly ME, Scholes DT, St. George K, Hatta M, Kawaoka Y, Wentworth DE. 2009. Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses. J Virol 83:10309–10313.

40. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, Spiro DJ. 2008. Viral genome sequencing by random priming methods. BMC Genomics 9:1–9.

41. JACKSON G, HF D, IG S, AV B. 1958. Transmission of the common cold to volunteers under controlled conditions: I. the common cold as a clinical entity. AMA Arch Intern Med 101:267–278.

42. RStudio Team. 2015. RStudio: Integrated Development for R.tle. RStudio, Inc., Boston, MA.

43. Python Software Foundation. 2013. Python Language Reference, version 2.7. Python Softw Found Version 3.03., http://www.python.org.

44. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

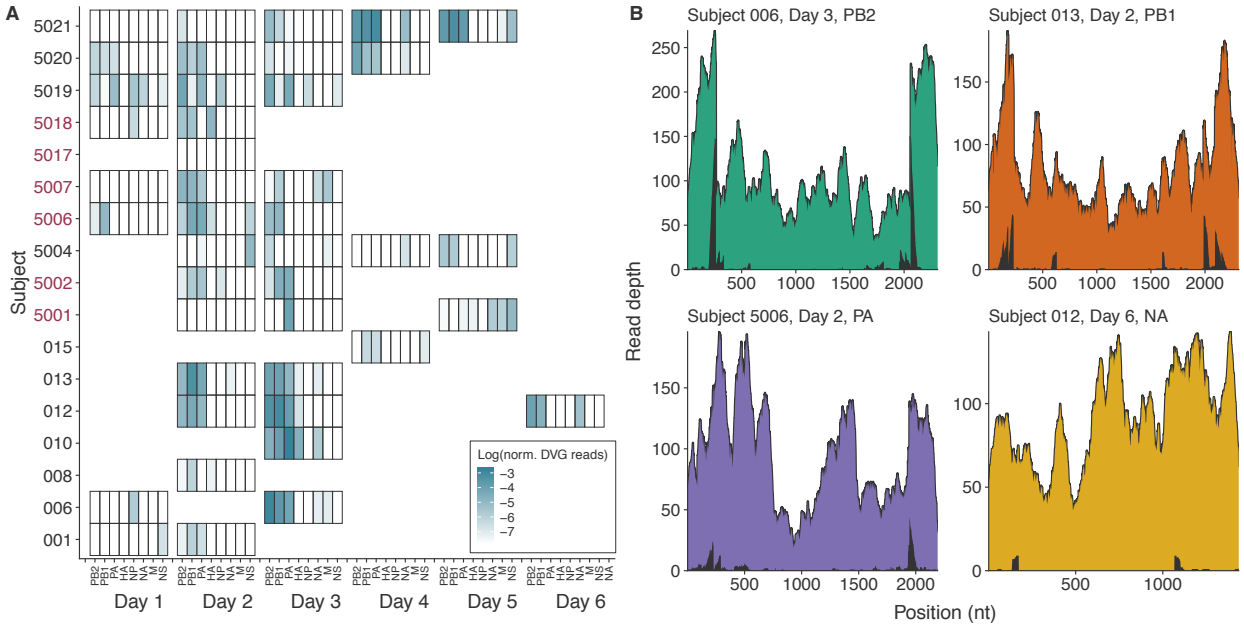
546 45. Wood DE, Salzberg SL. 2014. Kraken: Ultrafast metagenomic sequence classification
547 using exact alignments. Genome Biol 15.
548 46. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
549 sequence data. Bioinformatics 30:2114–2120.
550 47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
551 Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–
552 21.
553 48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
554 Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics
555 25:2078–2079.
556 49. Broad Institute. 2016. Picard tools. Https://BroadinstituteGithubIo/Picard/. Broad Institute.
557 50. Pielou EC. 1966. The measurement of diversity in different types of biological collections.
558 J Theor Biol 13:131–144.
559 51. Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
560 York.
561 52. Wilke CO. 2018. cowplot.
562 53. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018.
563 Stochastic processes constrain the within and between host evolution of influenza virus.
564 Elife 7:1–19.
565 54. Gallagher ME, Brooke CB, Ke R, Koelle K. 2018. Causes and Consequences of Spatial
566 Within-Host Viral Spread. Viruses 10:1–23.
567 55. Alnaji FG, Holmes JR, Rendon G, Vera JC, Fields, Christopher J.Martin BE, Brooke CB.
568 2018. Illumina-based sequencing framework for accurate detection and mapping of

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

influenza virus defective interfering particle-associated RNAs. bioRxiv.

56. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazelwood V, Lathrop S, Lifka D, Peterson GD, Rosies R, Scott JR, Wilkins-Diehr N. 2014. XSEDE: Accelerating Scientific Discovery. Comput Sci Eng 16:62–74.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



574

575 Figure 1. Graphical summary of the data used in this study. A) Heatmap showing the number of
576 sequencing reads indicating the presence of defective viral genomes in each gene segment
577 normalized by the total number of sequencing reads aligned to that gene segment. Rows
578 represent individual subjects (red text indicates early treatment (oseltamivir on the evening of the
579 first day post challenge) cohort). Columns represent day of sampling; sub columns indicate gene
580 segment. White space indicates lack of sequencing data. B) Representative coverage plots in
581 various gene segments. Background colored area shows the total read depth at a given nucleotide
582 (nt) position. Black color in the foreground represents the coverage depth of split reads,
583 indicative of DVGs.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

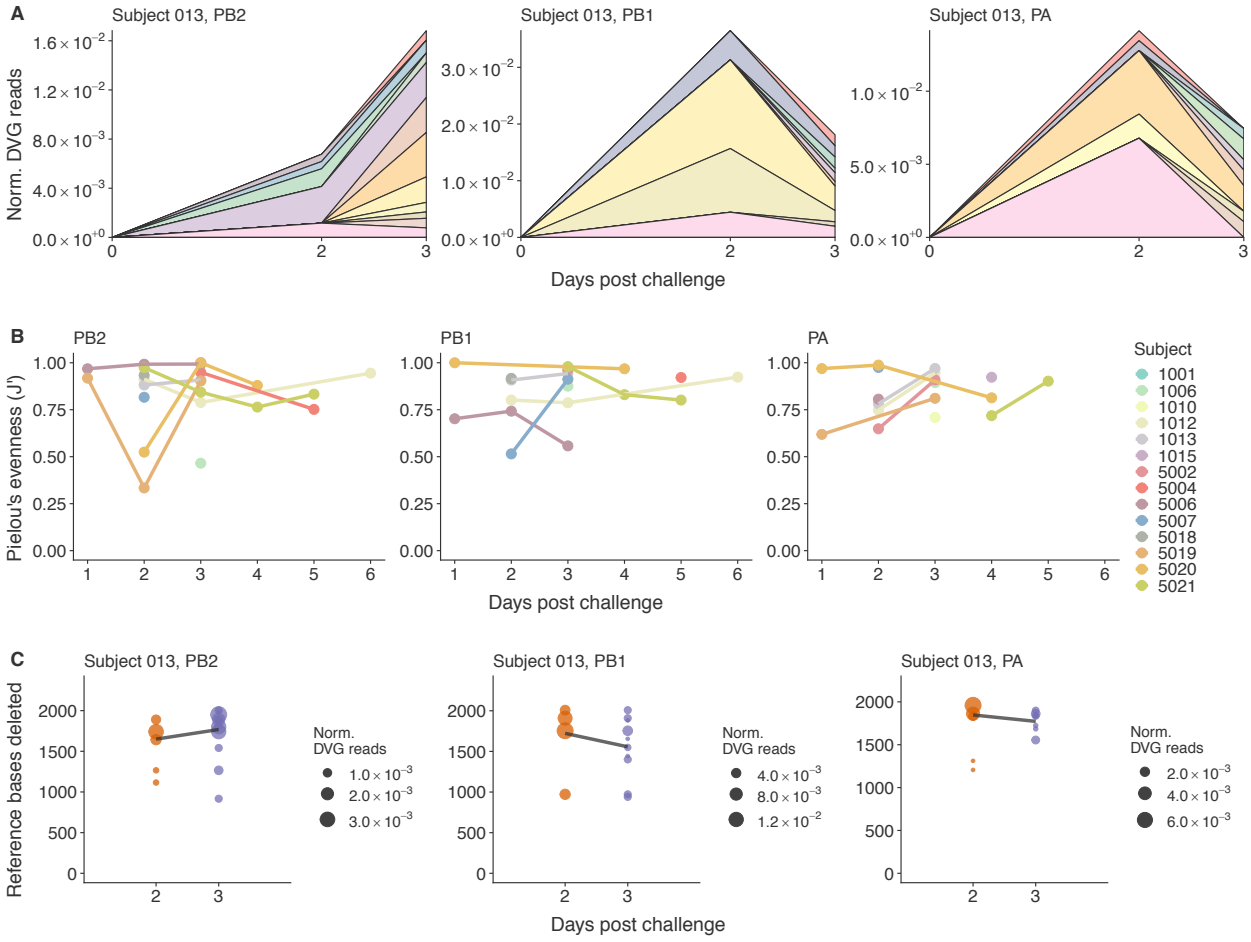


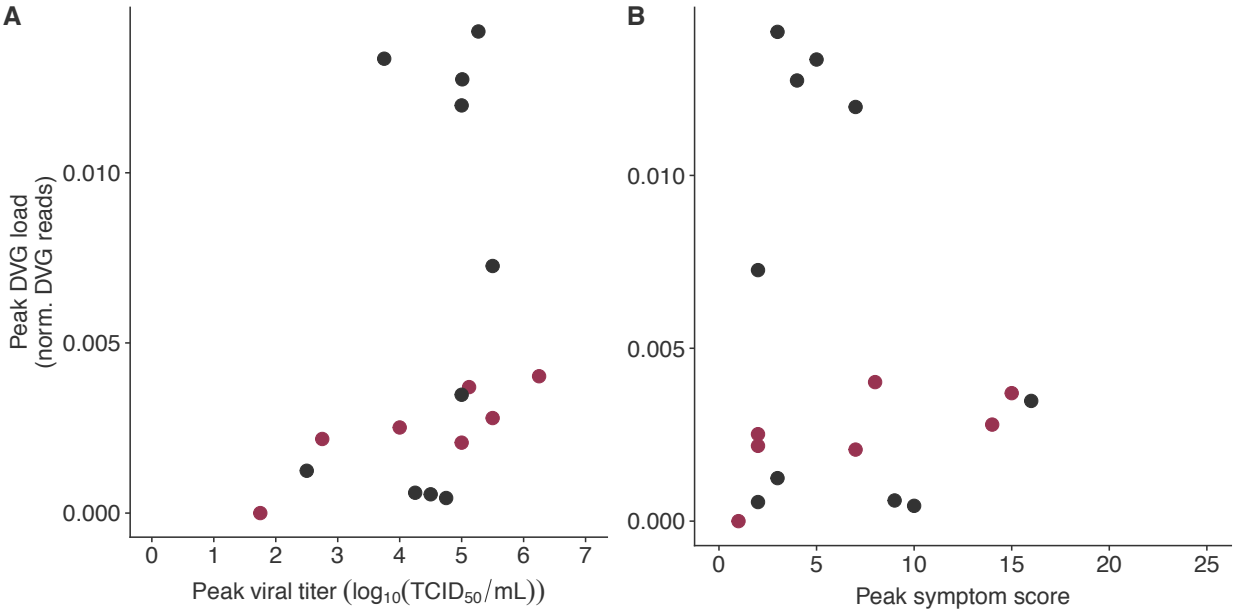
Figure 2. Population dynamics of observed defective viral genome (DVG) populations. A) Stacked area plots showing the DVG species observed at two and three days post challenge in subject 013 in the PB2, PB1, and PA gene segments. Each color represents an individual DVG species. The height of each region represents the normalized number of DVG reads supporting that DVG. B) Diversity of DVG populations in the PB2, PB1, and PA gene segments for each subject in the study. Lines connect data points from the same subject at multiple time points. Diversity is measured my Pielou's evenness (J'), which is given by $\frac{-\sum_{i=1}^{s} p_i \ln p_i}{\ln s}$ where s is the number of DVG species and p_i is the proportion of DVG reads which support that DVG species. C) Distribution of the number of deleted references bases in each observed DVG species in the PB2, PB1, and PA gene segments of subjects 013. Dot size represents the normalized number of

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

595 DVG reads supporting a specific DVG species. Trend lines connect the mean number of

596 reference bases deleted at each day, weighted by the normalized DVG read support.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



597

598 Figure 3. Dot plots of clinical data against the amount of defective viral genomes (DVGs). Each

599 dot represents a subject. A) Dot plot showing the lack of association between peak viral titer

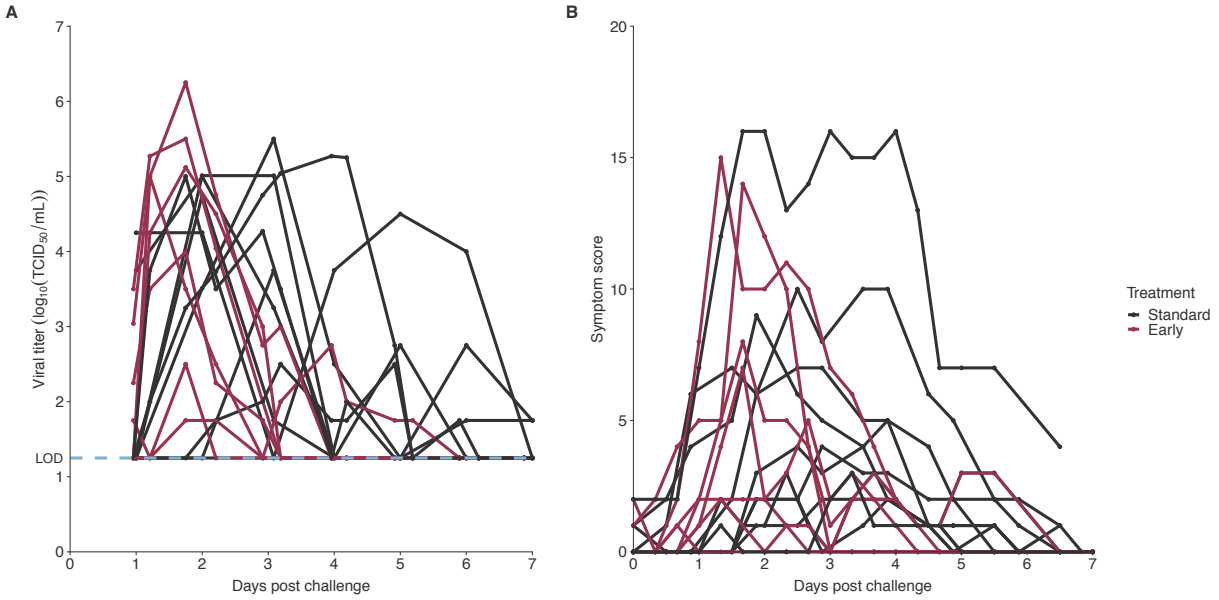
600 (log10(TCID50/mL) on the x-axis and the peak number of normalized DVG reads on the y-axis.

601 B) Dot plot showing the lack of association between peak Modified Jackson symptom score on

602 the x-axis and the peak number of normalized DVG reads on the y-axis.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

603 **Supplemental Material**



604

605 Figure S1. Clinical data included in the study. Red lines represent subjects in the early treatment
606 (oseltamivir on the evening of the first day post challenge) cohort, grey lines represent subjects
607 in the standard treatment (oseltamivir on the evening of the fifth day post challenge) cohort. A)
608 Viral titer ($log_{10}(TCID_{50}/mL)$) measurements for each subject at various time-points post-
609 challenge. Blue dotted line at 1.25 $log_{10}(TCID_{50}/mL)$ represents the limit of detection of the
610 assay used. B) Modified Jackson symptom scores for each subject at various time-points post-
611 challenge.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

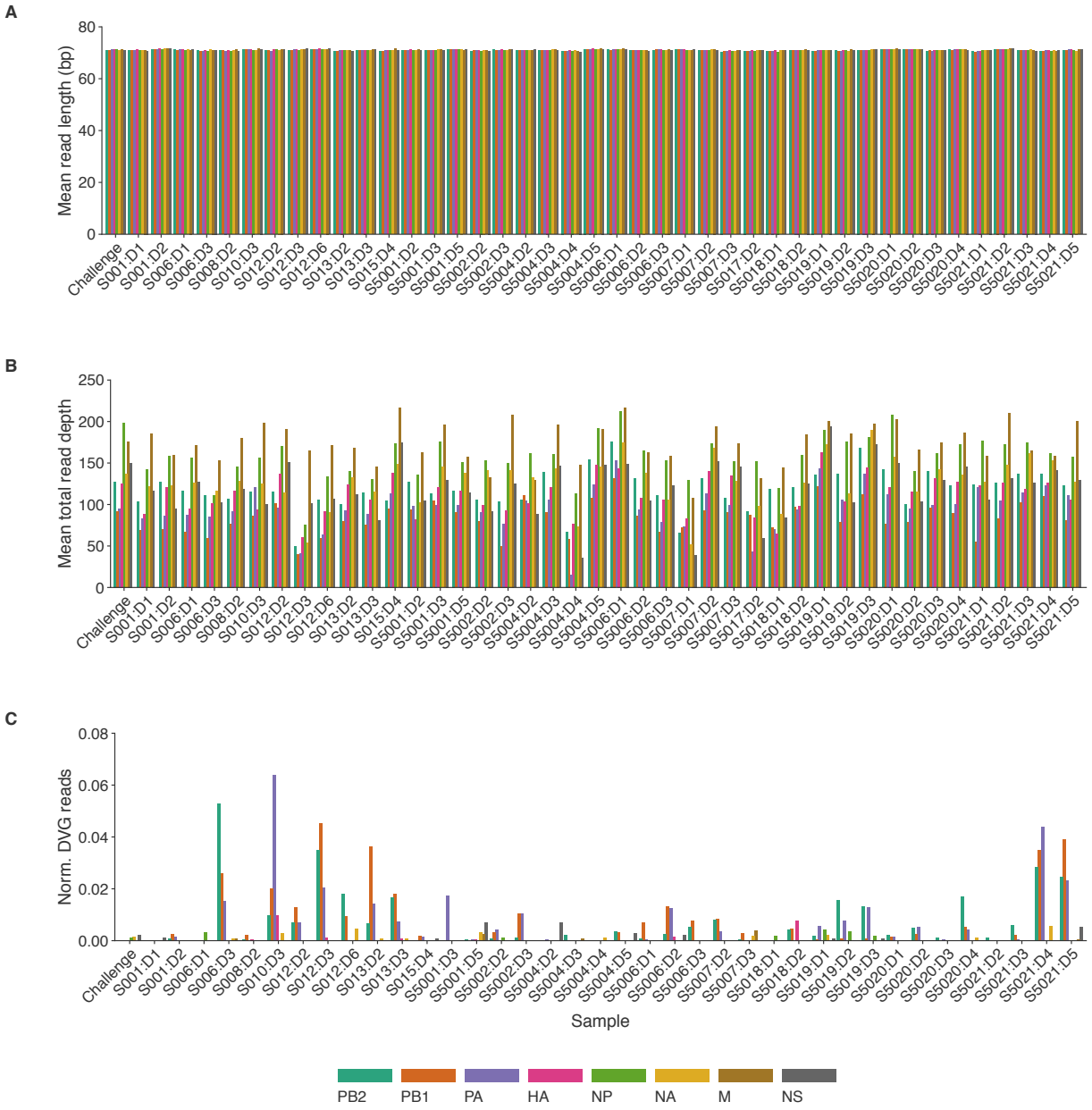


Figure S2. Summary of sequencing data. Bars are organized first by subject, then by sampling day, and finally by gene. Genes are also represented by colors. Data are summed across technical replicates. A) Mean length of mapped reads following read trimming, reference mapping, and removal of PCR duplicates. B) Mean total read depth following read trimming, reference mapping, and removal of PCR duplicates. C) Mean depth of split reads following read trimming, reference mapping, removal of PCR duplicates, and filtering of split reads to exclude those with

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

619 less than 15 alignment matches, less than 5 consecutive alignment matches, more than three

620 small indels, and with junction locations less than 100 bases apart.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

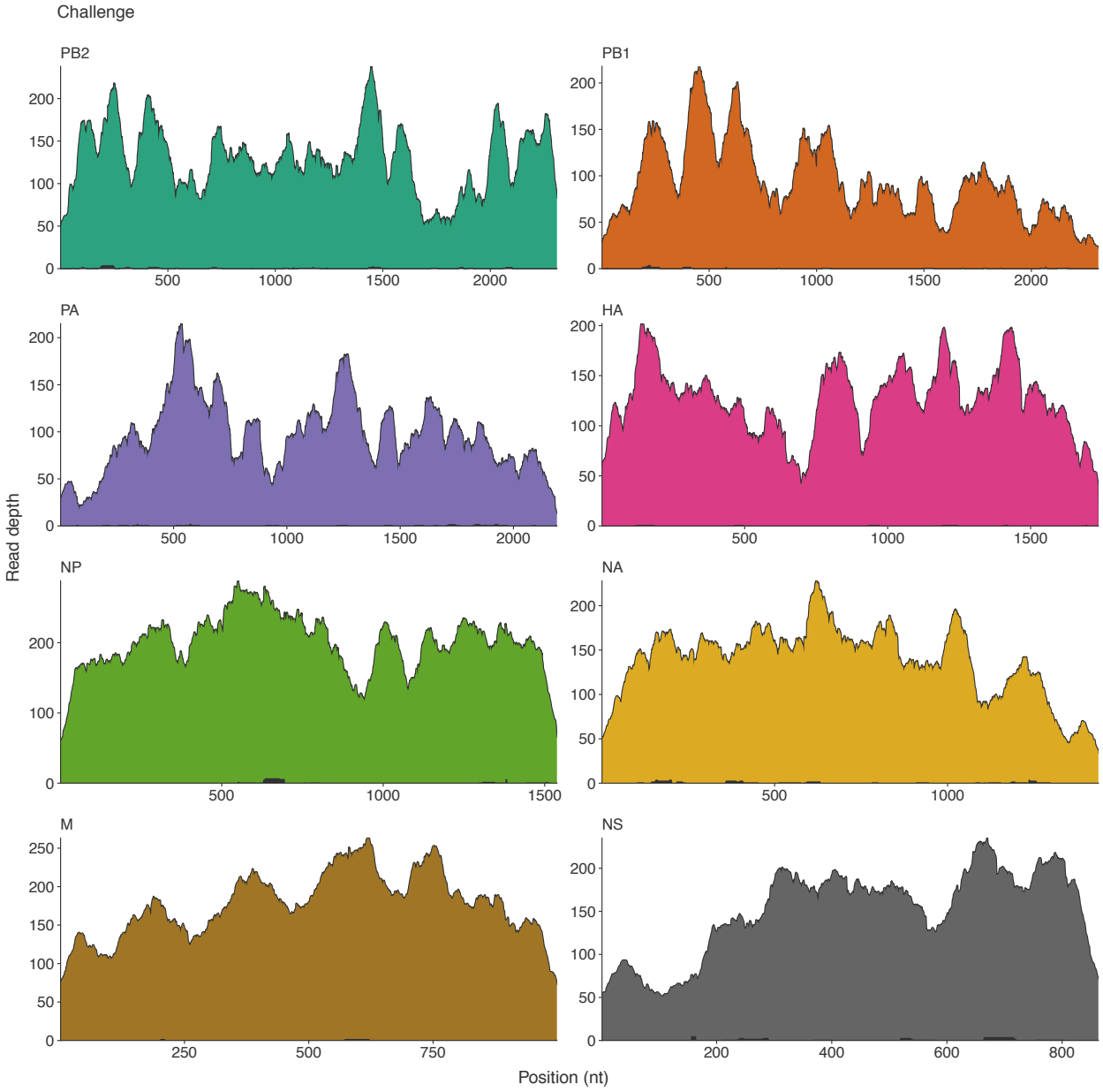
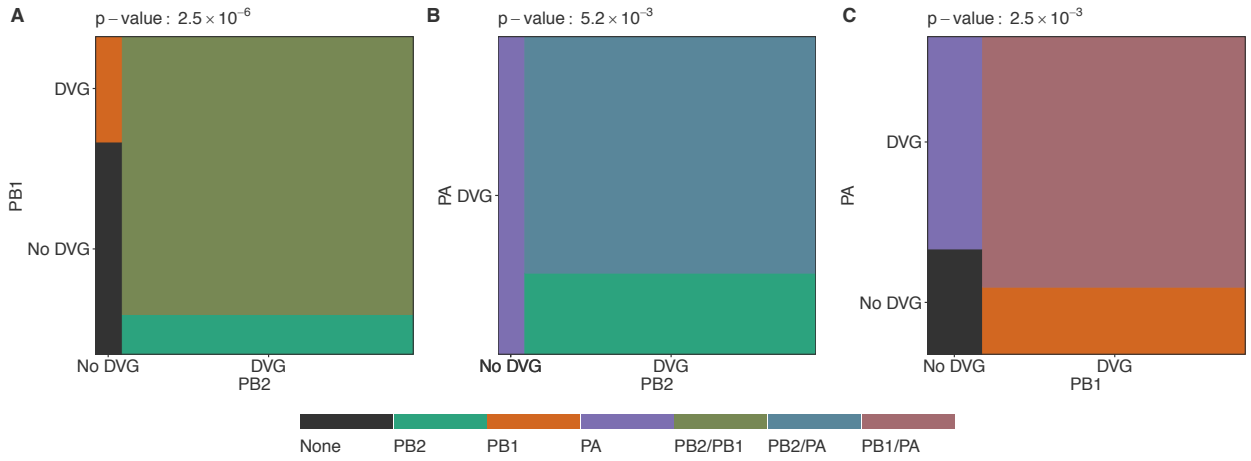


Figure S3. Coverage plot for each gene segment in the challenge stock used to infect all patients. Colored portions represent total read depth and black overlays represent read depth of split reads. Lack of appreciable split read depth indicates a lack of defective viral genomes (DVGs). The NP, NA, and NS gene segments each harbored a single DVG species, while no DVG species were identified in the other 5 gene segments.

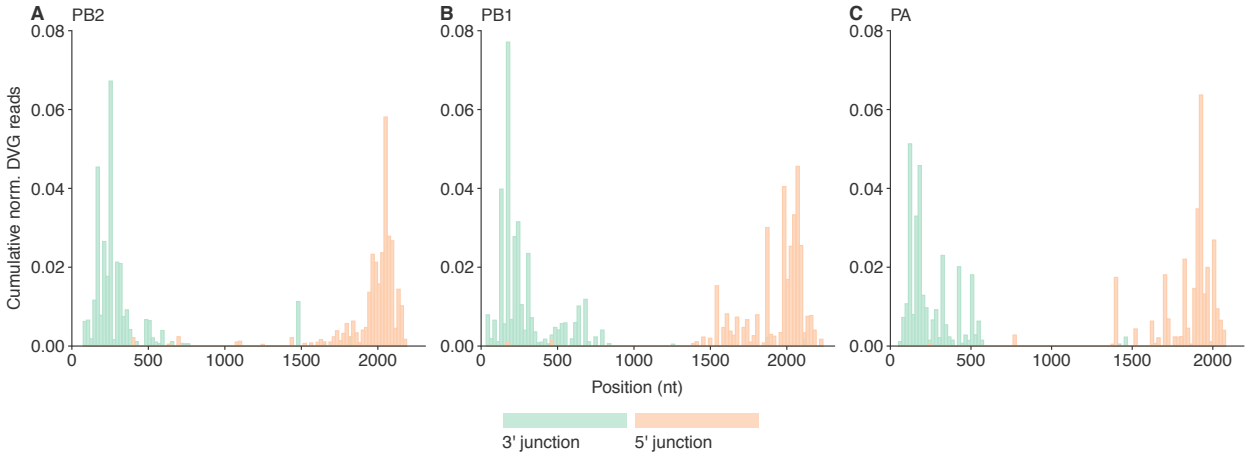
bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



627

628 Figure S4. Mosaic plots representing the per-sample dependence of defective viral genome
629 (DVG) presence in pairwise combinations of the PB2, PB1, and PA gene segments. Axis
630 represent the proportion of all samples. Grey area indicates lack of DVGs in either gene segment
631 and colored areas represent DVG presence in either or both gene segments. Fisher's exact test p-
632 values are listed above each plot. A) Association between DVG presence in the PB2 and PB1
633 genes. Green area represents samples with DVGs only in the PB2 gene, orange area represents
634 samples with DVGs only in the PB1 gene and forest green area represents samples with DVGs in
635 both genes. B) Association between DVG presence in the PB2 and PA gene segments. Green
636 area represents samples with DVGs only in the PB2 gene, purple area represents samples with
637 DVGs only in the PA gene, and blue area represents samples with DVGs in both. C) Association
638 between DVG presence in the PB1 and PA gene segments. Orange area represents samples with
639 DVGs only in the PB1 gene, purple area represents samples with DVGs only in the PA gene, and
640 mauve area represents samples with DVGs in both genes.

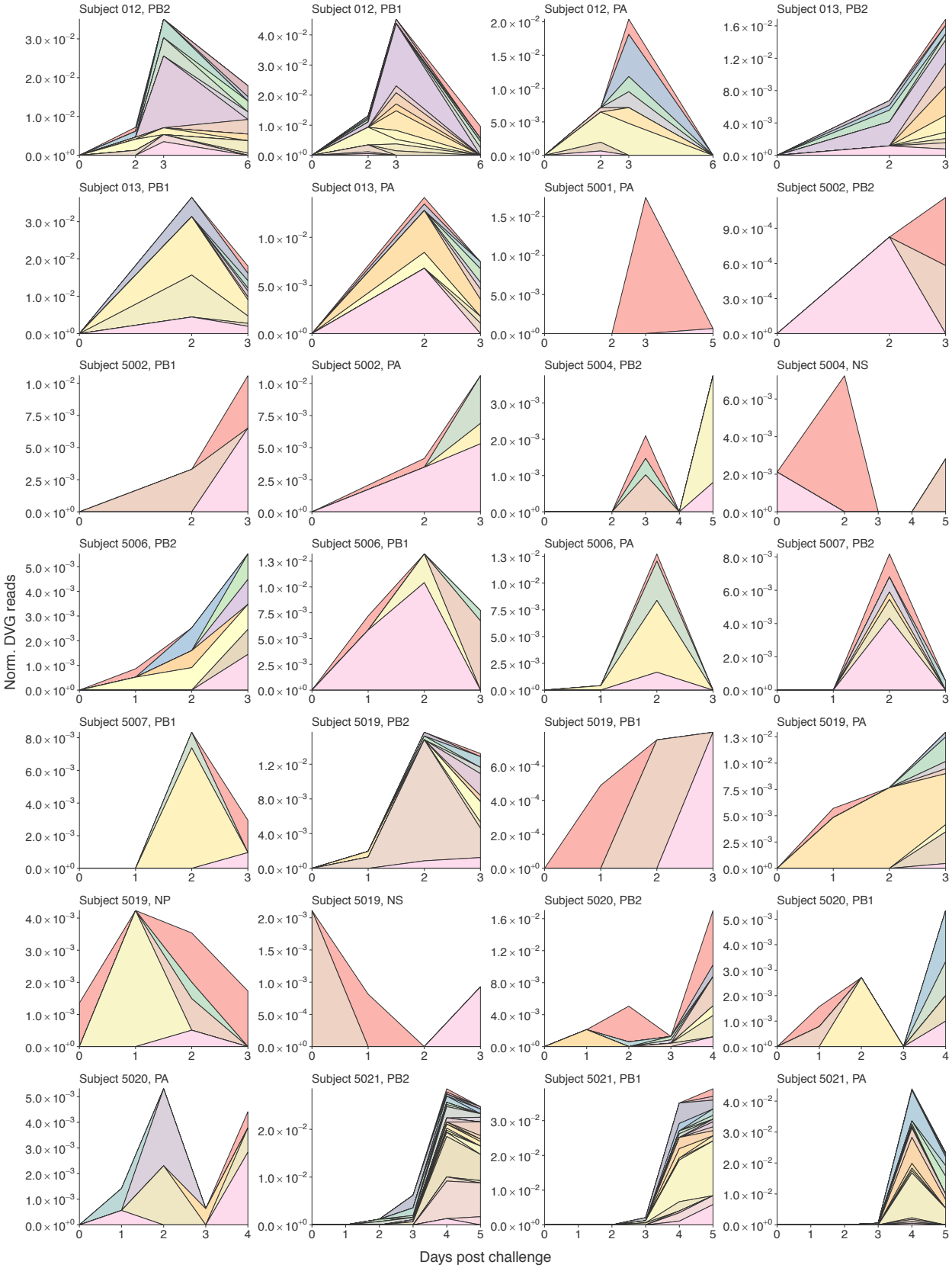
bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



641

642 Figure S5. Histograms (100 bins) showing the cumulative relative read support values for the 3'
643 (green) and 5' (orange) junction locations (compared to the reference) of the deletions generating
644 individual DVG species. The relative read support for DVGs observed within individual samples
645 was first normalized to the total number of reads mapped to that gene segment and then summed
646 across subjects. Data are shown for DVGs observed in the A) PB2, B) PB1, and C) PA genes.

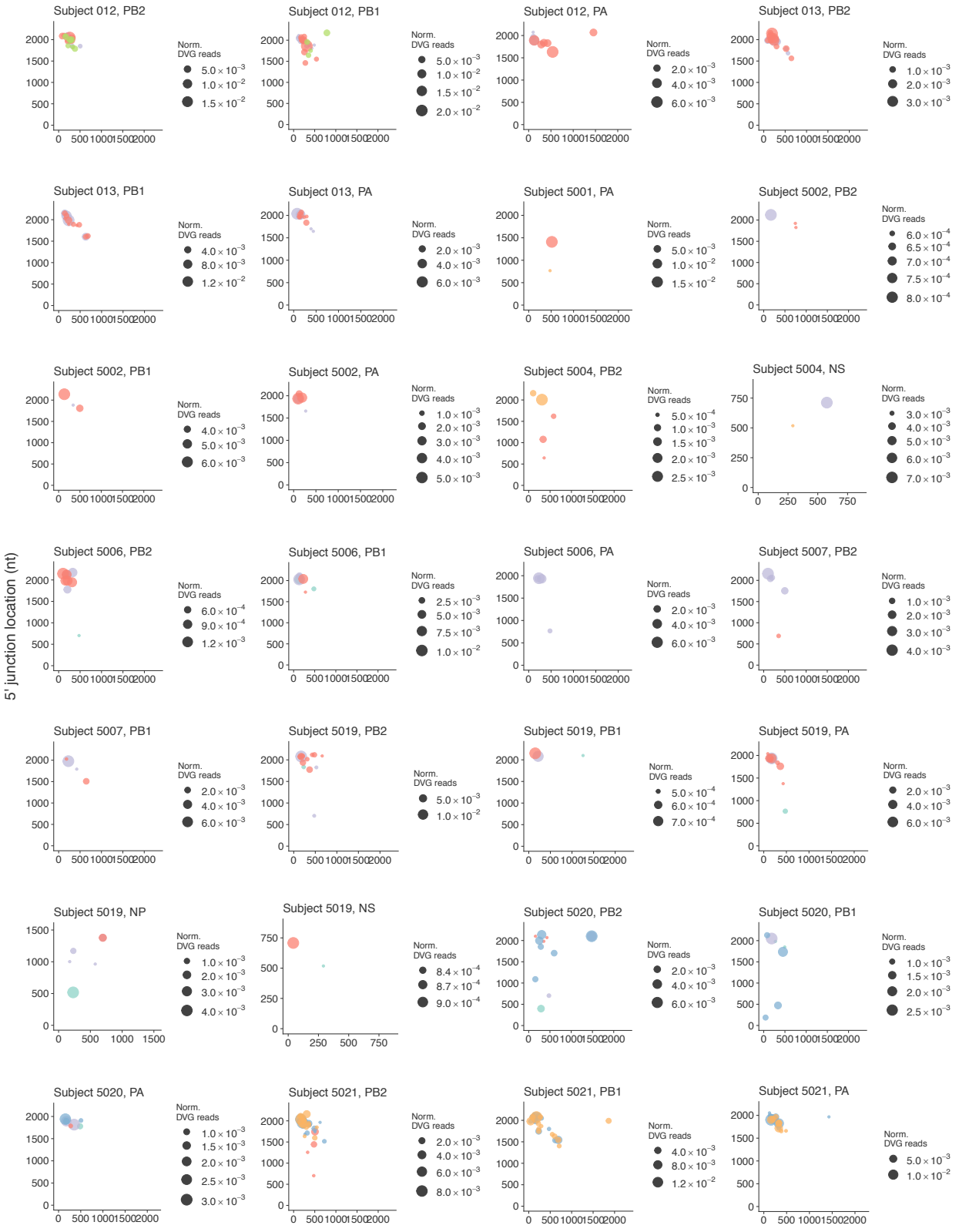
bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S6. Stacked area plots representing the defective viral genome (DVG) species within each gene for subjects with DVGs observed on multiple days. Each color represents an individual DVG species. The height of each region represents the normalized number of DVG reads supporting that DVG. DVG colors are not consistent between subjects.

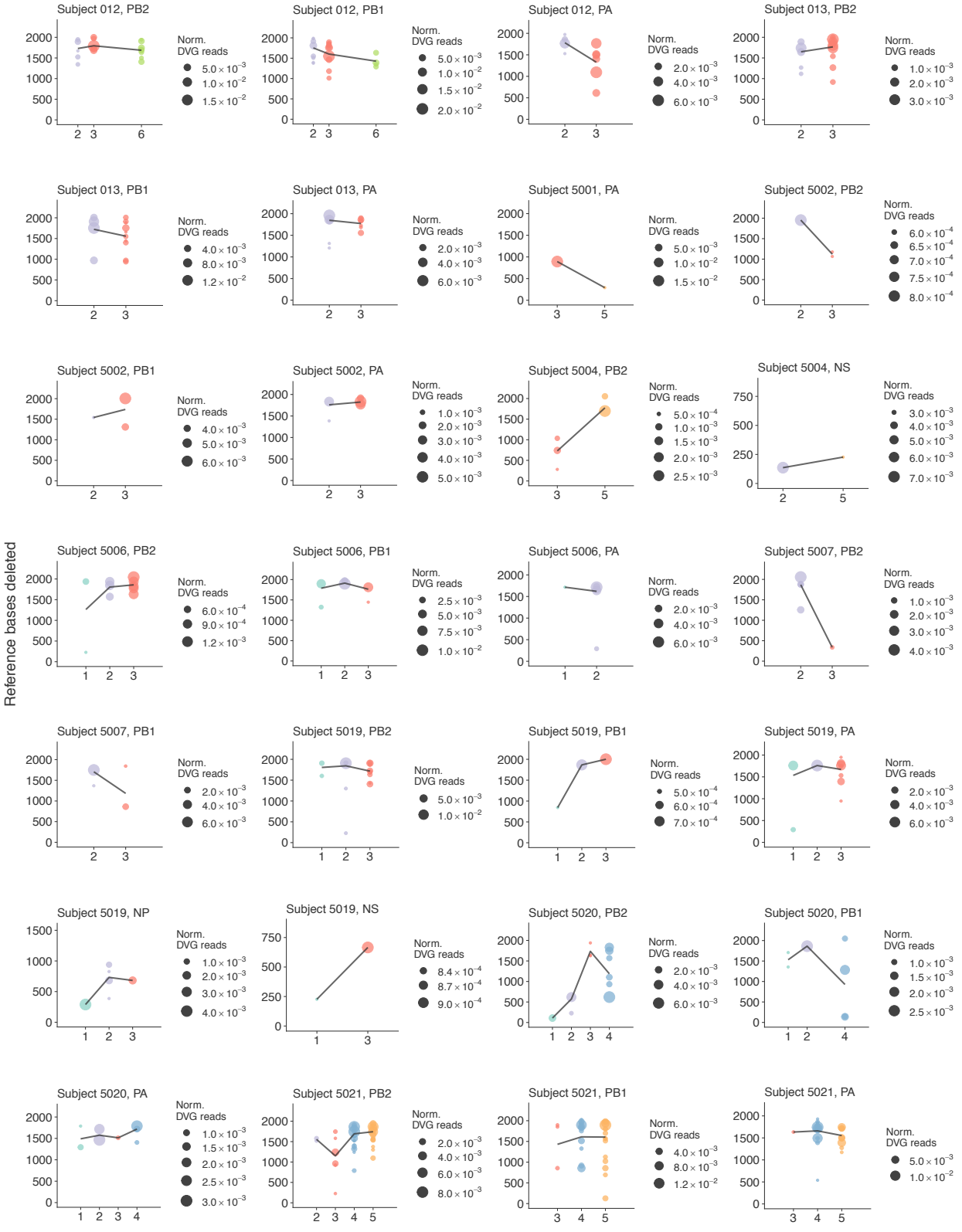
bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S7. Dot plots representing the defective viral genome (DVG) species within each gene for subjects with DVGs observed on multiple days. The x-axis represents the 3' junction location and the y-axis represents the 5' junction location. Dot size is dependent on the number of reads supporting a given DVG species, normalized by the number of reads mapped to that gene segment. Color represents sampling day.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S8. Dot plots representing the number of reference bases deleted in the observed defective viral genomes (DVG) species with each gene for subjects with DVGs observed on multiple days. Dot size is dependent on the number of reads supporting a given DVG species, normalized by the number of reads mapped to that gene segment. Color represents sampling day. Trend lines connect the mean number of reference bases deleted on each given day, weighted by the normalized number of supporting reads.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TABLE S1 Viral load and symptom scores

	Cohort			
	All (mean [sd])	Standard (mean [sd])	Early (mean [sd])	p-value
Peak viral titer (log_{10}(TCID$_{50}$/mL)	4.46 [1.15]	4.55 [0.84]	4.34 [1.48]	0.922
Time to peak viral titer (hours)	57.85 [25.75]	64.95 [27.01]	47.71 [19.82]	0.080
Duration of infection (hours)	99.71 [38.57]	117.80 [37.23]	73.86 [22.32]	0.035
Peak SS	6.47 [4.75]	6.10 [4.25]	7.00 [5.35]	0.922
Time to peak SS (hours)	46.65 [22.18]	59.30 [17.57]	28.57 [14.09]	0.003
Duration of symptoms (hours)	113.06 [36.98]	134.20 [17.21]	82.86 [36.71]	0.005
Cumulative SS	39.88 [42.94]	43.30 [49.99]	35.00 [29.43]	0.887

Table S1. Summary statistics for peak viral titer (log10(TCID50/mL), time to peak viral titer (hours), duration of infection, peak Modified Jackson symptom score, time to peak symptom score (hours), duration of symptoms (hours), and cumulative symptom score. Mean and population standard deviation are presented for all subjects, those in the early (treatment with oseltamivir on the evening of the first day post challenge) cohort, and those in the standard (treatment with oseltamivir on the evening of the fifth day post challenge) cohort. P-values comparing the early and standard treatment cohorts resultant from Mann-Whitney *U* tests are shown at right.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TABLE S2 Clinical data by subject

Subject	Inoculum (TCID50/ml)	Treatment cohort	Peak viral titer (TCID50/ml)	Time to peak viral titer (hours)	Duration of infection (hours)	Peak symptom score	Time to peak symptom score (hours)	Cumulative symptom score
001	6.41	Standard	4.25	24.0	48.0	9	45.0	40
006	5.25	Standard	5.00	48.0	168.0	7	36.0	55
008	5.25	Standard	4.75	48.0	74.0	10	60.0	60
010	4.41	Standard	3.75	74.0	120.0	5	93.0	22
012	4.41	Standard	5.01	48.0	168.0	4	69.0	18
013	3.08	Standard	5.50	74.0	96.0	2	69.0	12
015	3.08	Standard	4.50	120.0	144.0	2	45.0	9
5001	5.50	Early	2.75	95.0	124.5	2	0.0	11
5002	5.50	Early	5.50	42.0	70.0	14	40.0	63
5004	5.50	Standard	2.50	76.5	118.0	3	56.0	23
5006	5.50	Early	6.25	42.0	76.5	8	40.0	45
5007	5.50	Early	4.00	42.0	70.0	2	32.0	6
5017	5.50	Early	1.75	42.0	53.0	1	16.0	1
5018	5.50	Early	5.00	29.0	53.0	7	40.0	33
5019	5.50	Early	5.12	42.0	70.0	15	32.0	86
5020	5.50	Standard	5.00	42.0	100.5	16	40.0	184
5021	5.50	Standard	5.27	95.0	141.5	3	80.0	10

Table S2. Clinical data, including inoculum dose (log_{10}(TCID$_{50}$/mL)), treatment cohort, peak viral titer (log_{10}(TCID$_{50}$/mL)), time to peak viral titer (hours), duration of infection (hours), peak Modified Jackson symptom score, time to peak symptom score (hours), and cumulative symptom score, by subject

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S3 Sequencing data

		Mean [sd] read length (bp)							
Subject	Day	PB2	PB1	PA	HA	NP	NA	M	NS
Challenge	0	70.86 [0.15]	70.90 [0.17]	71.14 [0.17]	71.18 [0.16]	71.10 [0.14]	70.98 [0.18]	71.23 [0.18]	70.99 [0.22]
001	1	70.91 [0.17]	70.83 [0.21]	70.82 [0.20]	71.14 [0.20]	70.81 [0.18]	70.99 [0.19]	71.05 [0.19]	70.70 [0.27]
	2	71.35 [0.14]	71.23 [0.18]	71.29 [0.17]	71.65 [0.15]	71.28 [0.15]	71.44 [0.17]	71.58 [0.17]	71.45 [0.25]
006	1	71.16 [0.15]	70.96 [0.21]	71.15 [0.18]	71.17 [0.19]	70.94 [0.16]	71.12 [0.18]	71.09 [0.19]	71.11 [0.24]
	3	70.75 [0.16]	70.50 [0.24]	70.66 [0.20]	70.96 [0.19]	70.60 [0.21]	71.15 [0.19]	71.05 [0.20]	71.01 [0.27]
008	2	70.78 [0.17]	70.83 [0.19]	70.73 [0.19]	70.96 [0.18]	70.47 [0.18]	70.88 [0.19]	71.13 [0.18]	70.45 [0.27]
010	3	71.11 [0.15]	71.10 [0.17]	71.32 [0.14]	71.25 [0.19]	71.06 [0.16]	70.90 [0.19]	71.54 [0.15]	71.43 [0.24]
	2	70.94 [0.16]	70.91 [0.17]	70.66 [0.19]	71.15 [0.16]	71.12 [0.15]	71.06 [0.19]	71.38 [0.17]	71.11 [0.21]
012	3	70.84 [0.25]	71.05 [0.26]	71.12 [0.26]	71.27 [0.22]	70.92 [0.24]	71.12 [0.28]	71.21 [0.18]	71.52 [0.23]
	6	71.13 [0.16]	71.10 [0.21]	71.14 [0.20]	71.51 [0.17]	71.10 [0.16]	71.29 [0.21]	71.41 [0.17]	71.46 [0.23]
013	2	70.62 [0.18]	70.54 [0.20]	70.79 [0.18]	71.01 [0.17]	70.81 [0.18]	70.97 [0.18]	71.00 [0.19]	70.65 [0.28]
	3	70.78 [0.16]	70.80 [0.20]	70.99 [0.19]	71.09 [0.18]	70.97 [0.18]	70.91 [0.20]	71.17 [0.20]	71.32 [0.28]
015	4	70.59 [0.18]	70.61 [0.18]	70.81 [0.16]	71.00 [0.16]	70.95 [0.15]	70.82 [0.18]	71.46 [0.15]	70.99 [0.21]
	2	70.95 [0.15]	70.96 [0.17]	70.94 [0.17]	71.13 [0.20]	70.86 [0.18]	70.88 [0.22]	71.19 [0.18]	71.00 [0.27]
5001	3	70.80 [0.16]	70.78 [0.17]	70.98 [0.17]	70.98 [0.18]	70.98 [0.16]	71.15 [0.17]	71.21 [0.17]	70.92 [0.24]
	5	71.10 [0.15]	71.13 [0.17]	71.33 [0.16]	71.24 [0.17]	71.17 [0.16]	71.37 [0.16]	70.90 [0.21]	71.13 [0.25]
5002	2	70.70 [0.17]	70.93 [0.19]	70.98 [0.18]	70.93 [0.20]	70.72 [0.17]	71.05 [0.17]	70.97 [0.22]	70.69 [0.31]
	3	71.12 [0.16]	70.82 [0.24]	70.97 [0.19]	71.21 [0.18]	70.81 [0.17]	70.97 [0.18]	71.22 [0.16]	71.16 [0.23]
5004	2	70.77 [0.17]	70.96 [0.16]	70.90 [0.17]	70.97 [0.19]	70.84 [0.16]	70.84 [0.19]	70.98 [0.22]	71.10 [0.28]
	3	70.79 [0.15]	70.77 [0.19]	70.92 [0.17]	71.02 [0.17]	70.94 [0.17]	71.10 [0.18]	71.30 [0.17]	70.80 [0.24]
	4	70.69 [0.22]	70.42 [0.24]	70.58 [0.47]	70.83 [0.23]	70.69 [0.20]	70.94 [0.25]	70.56 [0.22]	70.39 [0.50]
	5	71.36 [0.12]	71.35 [0.14]	71.36 [0.14]	71.61 [0.13]	71.38 [0.13]	71.34 [0.16]	71.49 [0.16]	71.38 [0.20]
5006	1	71.21 [0.12]	71.02 [0.14]	71.21 [0.13]	71.31 [0.15]	71.23 [0.13]	71.17 [0.15]	71.50 [0.15]	71.30 [0.20]
	2	70.88 [0.15]	70.80 [0.19]	70.76 [0.18]	71.07 [0.18]	70.82 [0.16]	71.02 [0.17]	70.80 [0.20]	70.53 [0.29]
	3	70.88 [0.17]	71.24 [0.20]	71.13 [0.20]	71.18 [0.19]	71.00 [0.17]	71.06 [0.21]	71.41 [0.19]	71.00 [0.26]
5007	1	71.14 [0.20]	71.17 [0.19]	71.18 [0.19]	71.20 [0.20]	71.00 [0.17]	70.96 [0.30]	71.00 [0.24]	71.27 [0.39]
	2	70.96 [0.14]	70.79 [0.18]	70.84 [0.16]	71.03 [0.16]	70.85 [0.16]	71.20 [0.15]	71.35 [0.17]	71.00 [0.21]
	3	70.39 [0.18]	70.48 [0.19]	70.46 [0.19]	70.82 [0.17]	70.65 [0.18]	70.54 [0.20]	70.99 [0.19]	70.77 [0.23]
5017	2	70.73 [0.18]	70.60 [0.19]	70.52 [0.29]	70.83 [0.22]	70.63 [0.18]	70.87 [0.23]	70.99 [0.22]	70.90 [0.35]
5018	1	70.71 [0.16]	70.65 [0.21]	70.74 [0.21]	70.97 [0.24]	70.37 [0.21]	70.86 [0.23]	70.94 [0.21]	70.87 [0.31]
	2	70.97 [0.15]	70.87 [0.17]	71.02 [0.17]	71.03 [0.19]	70.96 [0.16]	70.98 [0.19]	71.31 [0.17]	71.03 [0.24]
5019	1	70.70 [0.15]	70.44 [0.17]	70.82 [0.15]	70.98 [0.15]	70.85 [0.15]	70.94 [0.16]	70.97 [0.18]	71.06 [0.19]
	2	70.78 [0.15]	70.65 [0.20]	70.67 [0.18]	70.84 [0.19]	70.75 [0.17]	70.67 [0.21]	71.13 [0.18]	70.90 [0.28]
	3	71.06 [0.12]	70.75 [0.16]	71.04 [0.14]	71.09 [0.15]	71.00 [0.15]	71.22 [0.14]	71.42 [0.16]	71.12 [0.20]
5020	1	71.35 [0.13]	71.31 [0.18]	71.17 [0.16]	71.41 [0.16]	71.28 [0.13]	71.17 [0.16]	71.45 [0.16]	71.22 [0.22]
	2	71.14 [0.16]	71.24 [0.18]	71.32 [0.16]	71.13 [0.17]	71.11 [0.17]	71.29 [0.19]	71.27 [0.18]	71.39 [0.25]
	3	70.71 [0.15]	70.77 [0.18]	70.72 [0.18]	70.98 [0.17]	70.88 [0.16]	70.98 [0.18]	70.94 [0.19]	70.96 [0.24]
	4	71.14 [0.14]	70.82 [0.18]	71.12 [0.16]	71.23 [0.16]	71.12 [0.15]	71.10 [0.18]	71.17 [0.17]	71.05 [0.22]
5021	1	70.48 [0.17]	70.16 [0.26]	70.62 [0.17]	70.72 [0.19]	70.81 [0.15]	70.80 [0.19]	70.92 [0.21]	71.06 [0.25]
	2	71.21 [0.14]	71.36 [0.16]	71.25 [0.15]	71.42 [0.15]	71.41 [0.14]	71.32 [0.16]	71.62 [0.15]	71.55 [0.20]
	3	71.00 [0.14]	70.83 [0.17]	71.03 [0.16]	71.02 [0.18]	71.02 [0.15]	71.12 [0.16]	71.00 [0.20]	70.73 [0.26]
	4	70.62 [0.15]	70.43 [0.17]	70.82 [0.16]	70.79 [0.18]	70.62 [0.17]	70.83 [0.18]	70.66 [0.21]	70.81 [0.23]
	5	71.01 [0.15]	70.92 [0.19]	71.19 [0.15]	71.38 [0.17]	70.98 [0.16]	70.74 [0.20]	71.32 [0.17]	71.11 [0.23]

679

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S3 (cont.) Sequencing data

		Mean [sd] read depth								
Subject	Day	Genome	PB2	PB1	PA	HA	NP	NA	M	NS
Challenge	0	130.06 [52.95]	127.70 [38.73]	91.83 [42.16]	95.24 [39.26]	125.10 [34.74]	198.38 [40.59]	137.17 [39.55]	176.27 [41.50]	150.11 [49.92]
001	1	105.56 [41.69]	103.17 [25.92]	68.62 [21.91]	83.20 [25.05]	88.15 [26.48]	142.68 [30.72]	122.12 [27.07]	185.26 [32.26]	116.76 [28.61]
	2	113.22 [41.24]	127.14 [29.16]	69.79 [18.45]	86.60 [25.43]	120.38 [34.46]	158.72 [30.33]	123.17 [27.81]	159.25 [34.73]	94.52 [26.69]
006	1	110.78 [42.77]	116.57 [26.32]	66.61 [24.71]	87.77 [28.21]	94.82 [30.87]	156.83 [26.89]	125.91 [23.33]	171.00 [36.92]	127.25 [37.10]
	3	99.82 [40.89]	110.84 [57.47]	59.49 [15.91]	85.41 [23.48]	100.90 [26.52]	111.04 [23.14]	116.61 [22.68]	153.25 [39.28]	102.94 [28.13]
008	2	113.41 [40.34]	106.35 [24.87]	76.83 [25.04]	92.26 [26.07]	115.97 [24.47]	145.97 [28.23]	128.02 [30.74]	180.31 [34.93]	119.10 [47.41]
010	3	119.43 [43.94]	115.34 [32.97]	86.64 [22.62]	120.70 [35.37]	93.73 [24.75]	156.55 [37.43]	124.88 [34.71]	198.01 [38.01]	100.53 [28.63]
	2	126.84 [43.61]	115.09 [32.53]	101.59 [29.02]	96.15 [26.89]	137.44 [39.56]	170.32 [36.08]	114.69 [17.74]	190.56 [34.17]	151.51 [40.19]
012	3	63.32 [43.06]	49.91 [39.48]	40.00 [23.27]	40.92 [14.73]	60.78 [26.14]	75.17 [19.73]	53.68 [13.11]	165.06 [40.19]	100.89 [36.52]
	6	95.41 [44.16]	105.38 [39.59]	59.09 [21.16]	63.29 [22.32]	91.49 [36.44]	133.23 [23.12]	90.93 [27.26]	171.92 [34.36]	107.00 [37.60]
013	2	112.48 [38.53]	100.57 [21.73]	79.72 [33.78]	92.37 [27.64]	124.28 [32.24]	140.01 [22.83]	133.02 [21.51]	168.21 [38.61]	111.76 [32.21]
	3	104.34 [34.72]	114.63 [40.05]	75.03 [18.78]	88.39 [24.34]	105.32 [26.22]	130.78 [17.15]	115.17 [25.38]	145.23 [33.25]	81.35 [23.64]
015	4	134.29 [47.29]	104.83 [28.54]	95.00 [25.56]	113.62 [27.78]	138.17 [40.59]	173.60 [24.57]	148.29 [23.56]	216.44 [32.05]	174.65 [46.81]
	2	110.54 [36.89]	127.36 [33.58]	93.87 [23.77]	98.40 [31.09]	82.29 [31.53]	136.26 [24.58]	102.39 [29.27]	162.95 [27.35]	104.86 [27.77]
5001	3	128.28 [40.79]	113.66 [24.05]	104.42 [25.65]	99.62 [25.44]	120.32 [27.06]	175.76 [33.23]	145.06 [24.48]	195.75 [33.56]	129.46 [32.22]
	5	118.42 [36.63]	117.02 [27.56]	90.22 [27.35]	98.79 [28.67]	117.00 [32.22]	151.29 [24.31]	137.83 [22.35]	157.81 [31.82]	113.92 [45.35]
5002	2	108.42 [37.70]	105.78 [30.46]	80.25 [22.42]	90.56 [29.29]	99.65 [26.66]	153.16 [28.78]	140.74 [36.07]	132.97 [27.18]	91.97 [26.91]
	3	107.12 [52.31]	103.33 [35.41]	50.00 [22.62]	76.93 [26.53]	93.27 [29.64]	150.14 [31.62]	141.28 [30.65]	207.92 [26.98]	124.62 [34.43]
5004	2	115.79 [37.03]	105.43 [27.83]	110.79 [33.77]	104.62 [31.04]	101.59 [26.24]	161.45 [29.79]	133.09 [24.60]	129.19 [40.40]	88.16 [39.56]
	3	130.44 [39.07]	138.70 [27.24]	90.92 [18.15]	105.39 [23.20]	120.72 [32.72]	160.42 [22.74]	143.18 [21.67]	196.52 [31.48]	146.21 [33.37]
	4	68.14 [42.28]	66.86 [30.54]	58.40 [18.18]	14.78 [9.39]	76.24 [24.65]	113.55 [29.63]	73.08 [18.56]	147.74 [26.75]	35.58 [14.47]
	5	146.08 [42.29]	154.27 [27.75]	108.22 [26.89]	123.64 [35.86]	147.62 [43.76]	191.72 [32.06]	145.11 [24.71]	190.59 [26.93]	148.25 [40.29]
5006	1	165.53 [43.03]	175.88 [29.02]	131.60 [25.84]	152.71 [35.73]	143.46 [36.89]	212.54 [33.32]	174.14 [31.23]	216.26 [31.28]	148.85 [43.29]
	2	119.50 [41.96]	131.30 [31.50]	86.17 [20.91]	93.71 [41.21]	108.39 [33.88]	164.76 [25.35]	137.73 [30.07]	162.43 [34.95]	104.45 [30.76]
	3	105.76 [39.33]	110.57 [28.05]	66.47 [21.51]	78.56 [21.71]	105.35 [29.63]	153.19 [31.31]	105.67 [17.87]	158.32 [23.16]	122.84 [34.83]
5007	1	77.68 [34.32]	66.21 [22.47]	72.25 [22.22]	72.98 [26.40]	83.51 [29.53]	129.63 [31.82]	51.46 [13.22]	107.91 [23.70]	39.23 [16.92]
	2	137.64 [47.75]	131.16 [36.17]	92.55 [33.66]	112.74 [29.08]	140.41 [46.79]	174.09 [25.26]	168.71 [36.77]	193.59 [37.90]	151.99 [47.76]
	3	121.60 [37.73]	107.81 [22.88]	90.21 [29.32]	99.59 [25.60]	135.14 [30.84]	151.99 [22.48]	128.28 [24.47]	173.69 [32.53]	145.66 [38.65]
5017	2	90.64 [44.18]	92.17 [34.62]	87.88 [35.07]	42.94 [27.10]	83.79 [28.90]	152.08 [30.09]	98.05 [24.10]	131.47 [35.36]	59.75 [24.48]
5018	1	92.19 [35.56]	118.53 [27.01]	72.54 [17.66]	69.80 [23.76]	64.42 [22.19]	119.32 [26.95]	88.38 [23.81]	144.81 [29.91]	84.16 [28.46]
	2	119.67 [41.05]	121.00 [36.26]	97.33 [29.51]	94.29 [26.36]	98.65 [31.48]	159.58 [29.00]	126.20 [24.09]	184.60 [31.24]	125.66 [35.19]
5019	1	156.68 [44.40]	135.82 [38.49]	121.54 [30.87]	143.33 [32.70]	162.31 [42.45]	189.70 [30.71]	172.60 [23.12]	200.09 [39.61]	193.60 [49.27]
	2	120.85 [45.35]	137.00 [35.58]	79.06 [22.11]	105.55 [29.20]	103.99 [34.07]	175.37 [35.49]	113.37 [23.06]	185.75 [31.28]	102.50 [35.94]
	3	156.41 [44.08]	167.84 [38.59]	111.86 [21.39]	136.82 [36.14]	144.16 [37.54]	181.09 [26.27]	189.70 [28.79]	197.68 [39.11]	172.45 [47.86]
5020	1	137.26 [52.20]	141.90 [38.88]	76.92 [18.67]	111.84 [33.70]	120.94 [38.13]	207.55 [29.35]	157.01 [25.94]	202.69 [29.29]	150.07 [35.68]
	2	109.16 [36.03]	100.37 [26.59]	78.86 [20.88]	95.27 [26.23]	115.49 [31.89]	140.46 [27.21]	115.58 [26.20]	166.18 [28.39]	103.89 [32.33]
	3	129.17 [41.06]	140.08 [34.40]	95.70 [23.80]	99.50 [34.00]	131.35 [38.03]	161.50 [26.56]	141.93 [24.15]	174.47 [44.14]	129.35 [32.58]
	4	127.35 [42.12]	123.43 [29.88]	89.79 [27.19]	99.87 [25.11]	127.09 [33.38]	172.44 [27.99]	135.91 [23.15]	186.18 [26.46]	146.02 [48.68]
5021	1	119.23 [48.62]	124.16 [40.50]	54.82 [22.68]	120.59 [30.38]	122.58 [38.86]	176.88 [33.80]	127.77 [27.01]	158.76 [29.81]	105.67 [44.61]
	2	129.49 [43.63]	126.45 [21.87]	82.78 [16.16]	104.53 [27.36]	126.65 [29.59]	172.31 [26.37]	147.77 [26.82]	210.60 [33.68]	131.21 [34.92]
	3	133.33 [39.74]	136.60 [33.58]	102.49 [22.82]	114.86 [32.83]	118.42 [30.18]	174.26 [32.23]	162.01 [28.46]	165.12 [37.60]	126.60 [34.94]
	4	135.21 [45.59]	136.99 [46.39]	110.48 [38.74]	122.80 [60.09]	126.64 [30.28]	161.63 [27.12]	153.02 [26.25]	158.88 [42.28]	141.37 [47.72]
	5	122.18 [45.43]	123.22 [39.92]	81.08 [33.16]	111.55 [36.85]	105.47 [28.74]	157.43 [24.94]	126.80 [25.93]	200.10 [31.82]	129.41 [39.81]

Table S3. Sequencing data, including mean [population standard deviation (sd)] read length (bp) and mean [sd] coverage (reads), for all samples. Data are calculated following read trimming (removal of Illumina adapters, leading or trailing bases with quality <3, portions of reads where the average quality per-base in 4-base wide sliding windows was <15 and reads with <50 bases)

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

685 (46), alignment to the reference genome, realignment to a consensus sequence (47), and removal
686 of PCR duplicates (49). Technical sequencing replicates were processed separately and summary
687 statistics calculated from the combination of the two samples.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TABLE S4 DVG read support

		Normalized DVG reads							
Subject	Day	PB2	PB1	PA	HA	NP	NA	M	NS
Challenge	0					0.0014	0.0014		0.0021
1001	1								0.0013
	2	0.0007	0.0026	0.0015					
1006	1					0.0031			
	3	0.0529	0.0259	0.0154			0.0008	0.0009	
1008	2	0.0006	0.0023		0.0007				
1010	3	0.0099	0.0200	0.0641	0.0098		0.0030		
1012	2	0.0072	0.0131	0.0071					
	3	0.0351	0.0452	0.0204	0.0013				
	6	0.0179	0.0096				0.0047		
1013	2	0.0068	0.0365	0.0141			0.0007		
	3	0.0168	0.0181	0.0074	0.0008		0.0008		
1015	4		0.0019	0.0016					0.0009
5001	3			0.0174					
	5	0.0005		0.0006	0.0007		0.0031	0.0026	0.0069
5002	2	0.0008	0.0033	0.0042		0.0011			
	3	0.0012	0.0106	0.0106					
5004	2			0.0006					0.0072
	3	0.0021						0.0007	
	4						0.0013		
	5	0.0038	0.0034						0.0028
5006	1	0.0009	0.0072	0.0004					
	2	0.0025	0.0132	0.0127	0.0015				0.0023
	3	0.0055	0.0077						
5007	2	0.0082	0.0083	0.0036					
	3	0.0006	0.0029				0.0018	0.0039	
5018	1					0.0018			
	2	0.0044	0.0045		0.0076				
5019	1	0.0020	0.0005	0.0057		0.0042	0.0023		0.0008
	2	0.0157	0.0008	0.0077		0.0035			
	3	0.0132	0.0008	0.0129		0.0017			0.0009
5020	1	0.0021	0.0016	0.0014					
	2	0.0051	0.0027	0.0053					
	3	0.0013		0.0006					
	4	0.0170	0.0053	0.0044			0.0010		
5021	2	0.0012							
	3	0.0062	0.0021	0.0006					
	4	0.0285	0.0351	0.0440			0.0057		
	5	0.0247	0.0392	0.0233				0.0007	0.0055

Table S4. Normalized DVG read count for all subjects for each gene segment. Values represent the proportion of the reads mapped to a given gene which support DVG species present in both

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

691 technical replicates. Blank cells represent instances where no DVG reads were observed.

692 Subjects highlighted in grey represent those in the early treatment (oseltamivir on the first day

693 post challenge) treatment cohort.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TABLE S5 DVG junction sites

		3' junction site			5' junction site			Mean [sd] bases deleted (nt)
	Reference length (nt)	Mean [sd] (nt)	Minimum (nt)	Maximum (nt)	Mean [sd] (nt)	Minimum (nt)	Maximum (nt)	
PB2	2310	325 [295]	84	1814	1971 [236]	398	2177	1646 [392]
PB1	2312	304 [259]	43	2113	1929 [225]	187	2224	1625 [397]
PA	2192	256 [180]	66	1455	1849 [204]	243	2074	1593 [332]

Table S5. Summary statistics (mean [population standard deviation (sd)], weighted by the normalized DVG read count) for the 3' and 5' junction sites (relative to the reference) of DVGs observed in the PB2, PB1, and PA genes as well as the mean [sd] number of reference bases deleted.

bioRxiv preprint doi: https://doi.org/10.1101/814673; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TABLE S6 DVG species observed in multiple samples

Gene	DVG ID	Samples			
PB2	93_1983	013, Day 2	013, Day 3		
PB2	154_2069	5019, Day 2	5019, Day 3		
PB2	170_2035	5021, Day 4	5021, Day 5		
PB2	171_2087	010, Day 3	5007, Day 2		
PB2	172_2079	5019, Day 1	5019, Day 2	5019, Day 3	
PB2	187_2125	5006, Day 1	5006, Day 2	5006, Day 3	
PB2	209_1983	010, Day 3	5006, Day 3		
PB2	211_1956	5021, Day 3	5021, Day 4	5021, Day 5	
PB2	241_1982	013, Day 2	013, Day 3		
PB2	249_2005	5021, Day 4	5021, Day 5		
PB2	356_1937	5021, Day 2	5021, Day 3	5021, Day 4	5021, Day 5
PB2	386_1903	5021, Day 2	5021, Day 4		
PB2	476_703	5006, Day 1	5019, Day 2	5020, Day 2	5021, Day 3
PB2	505_1743	5021, Day 3	5021, Day 4		
PB2	523_1789	013, Day 2	013, Day 3		
PB2	1482_2101	5020, Day 2	5020, Day 4		
PB1	43_1978	5021, Day 4	5021, Day 5		
PB1	90_2073	5021, Day 4	5021, Day 5		
PB1	131_2028	5006, Day 1	5006, Day 2		
PB1	142_2150	013, Day 2	013, Day 3		
PB1	179_2076	5021, Day 3	5021, Day 4	5021, Day 5	
PB1	183_1994	012, Day 2	012, Day 3		
PB1	183_2092	013, Day 2	013, Day 3		
PB1	226_1924	010, Day 3	5021, Day 5		
PB1	231_1746	5021, Day 4	5021, Day 5		
PB1	232_1986	013, Day 2	013, Day 3		
PB1	313_1863	012, Day 2	012, Day 3		
PB1	631_1603	013, Day 2	013, Day 3		
PB1	689_1544	5021, Day 3	5021, Day 4	5021, Day 5	
PA	102_1936	5002, Day 2	5002, Day 3		
PA	124_1891	012, Day 2	012, Day 3		
PA	153_1941	5020, Day 1	5020, Day 4		
PA	153_2015	013, Day 2	013, Day 3		
PA	160_1903	5021, Day 4	5021, Day 5		
PA	175_1892	5020, Day 2	5020, Day 4		
PA	175_1935	5019, Day 1	5019, Day 2	5019, Day 3	
PA	197_1916	5021, Day 4	5021, Day 5		
PA	229_1944	5006, Day 1	5006, Day 2		
PA	263_1968	5021, Day 4	5021, Day 5		
PA	328_1719	5021, Day 4	5021, Day 5		
PA	331_1824	5021, Day 4	5021, Day 5		
PA	476_767	5001, Day 5	5004, Day 2	5006, Day 2	5019, Day 1
NP	229_518	006, Day 1	5018, Day 1	5019, Day 1	
NP	696_1378	5002, Day 2	5019, Day 2	5019, Day 3	Challenge, Day 0
NA	421_803	006, Day 3	013, Day 2		
NA	633_737	010, Day 3	5019, Day 1	5021, Day 4	
NS	291_518	001, Day 1	5004, Day 5	5019, Day 1	5021, Day 5

Table S6. Defective viral genome (DVG) species observed in multiple samples. DVGs are organized by gene segment. DVG IDs represent the first_last deleted reference based as identified using the "jI" SAM tag.