1　**Influenza A virus defective viral genomes are inefficiently packaged into virions**
2　**relative to wild-type genomic RNAs**
3
4　Fadi G. Alnaji[a], William K. Reiser[a], Aartjan te Velthuis[b], Christopher B. Brooke[a,c] #
5
6　[a] Department of Microbiology, University of Illinois at Urbana-Champaign
7　[b] Lewis Thomas Laboratory, Department of Molecular Biology, Princeton University
8　[c] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-
9　Champaign
10
11　#Corresponding author (cbrooke@illinois.edu)
12
13　**Abstract**
14　Deletion-containing viral genomes (DelVGs) are commonly produced during influenza A
15　virus infection and have been implicated in influencing clinical infection outcomes. Despite
16　their ubiquity, the specific molecular mechanisms that govern DelVG formation and their
17　packaging into defective interfering particles (DIPs) remain poorly understood. Here, we
18　utilized next-generation sequencing to analyze DelVGs that form *de novo* early during
19　infection, prior to packaging. Analysis of these early DelVGs revealed that deletion
20　formation occurs in clearly defined hotspots and is significantly associated with both direct
21　sequence repeats and enrichment of adenosine and uridine bases. By comparing
22　intracellular DelVGs with those packaged into extracellular virions, we discovered that
23　DelVGs face a significant bottleneck during genome packaging relative to wild type
24　genomic RNAs. Surprisingly, packaged DelVGs exhibited no signs of enrichment for
25　specific deletion characteristics suggesting that all DelVGs are equally limited in
26　packaging efficiency. Our data provide the first unbiased, high-resolution portrait of the
27　diversity of DelVGs that are generated by the IAV replication machinery and shed light on
28　the mechanisms that underly DelVG formation.
29
30　**Importance**
31　Defective interfering particles (DIPs) are commonly produced by RNA viruses and have
32　been implicated in modulating clinical infection outcomes, hence, there is increasing
33　interest in the potential of DIPs as antiviral therapeutics. For influenza viruses, DIPs are
34　formed by the packaging of genomic RNAs harboring internal deletions. Despite decades
35　of study, the mechanisms that drive the formation of these deletion-containing viral
36　genomes (DelVGs) remain elusive. Here, we used a specialized sequencing pipeline to
37　characterize the first wave of DelVGs that form during influenza virus infection. This
38　dataset provides an unbiased profile of the deletion-forming preferences of the influenza
39　virus replicase. Additionally, by comparing the early intracellular DelVGs with those that
40　get packaged into extracellular virions, we described a significant segment-specific
41　bottleneck that limits DelVG packaging relative to wild type viral RNAs. Altogether, these
42　findings reveal factors that govern the production of both DelVGs and DIPs during
43　influenza virus infection.
44
45
46

1

INTRODUCTION

47 Influenza A virus (IAV) populations are highly heterogeneous and largely consist of virions
48 that lack functional copies of one or more gene segments (1, 2). A major contributor to
49 this heterogeneity is the common presence of defective interfering particles (DIPs) within
50 viral populations. DIPs are virions that harbor a large deletion in one or more genome
51 segments, resulting in an inability to express the full set of viral proteins required for
52 productive replication. DIPs have been demonstrated in numerous contexts to interfere
53 with wild-type virus replication (hence the name), ostensibly either by outcompeting wild
54 type genomes for replication and packaging and/or by triggering innate immune activation
55 (3–7). Recent studies have correlated the abundance of DIPs within clinical samples with
56 severity of both IAV and RSV infection, suggesting a role for DIPs in modulating viral
57 pathogenicity (8, 9). Despite being discovered over 60 years ago, the specific molecular
58 processes that drive DIP formation, as well as the effects of DIPs on the collective
59 behavior and pathogenicity of viral populations remain mysterious (10, 11).

60

61 The deletion-containing genomic RNAs carried by DIPs are commonly referred to as
62 defective viral genomes (DVGs) (3). This terminology is complicated for influenza viruses,
63 as a variety of distinct defective viral genome species have been described, including
64 hyper-mutated segments (12) and mini viral RNAs (mvRNAs), which carry enormous
65 deletions and do not get packaged into virions (13). In addition, it is not yet clear that the
66 production of some deletion-containing segments is actually detrimental to viral
67 population fitness (11). Thus, to minimize confusion, we use the term DelVG (Deletion-
68 containing Viral Genome) here to refer to any viral gene segments that contain deletions
69 greater than 10 nucleotides, yet retain the classical viral packaging signals.

70

71 Next generation sequencing (NGS) represents a powerful new tool for revealing the
72 specific processes and molecular determinants that underly DelVG formation (14–17).
73 The analysis of large numbers of individual DelVGs can reveal specific patterns that yield
74 mechanistic insight into the formation process (18). IAV DelVGs are typically studied in
75 the context of extracellular DIPs. A potential limitation of this approach for investigating
76 the DelVG formation process is that the requirements of intracellular trafficking and
77 packaging may specifically select for a subset of the total repertoire of DelVGs produced
78 within the cell (19). As a result, DelVGs isolated from extracellular DIPs may not be
79 representative of the full range of DelVG products produced within the infected cell. Such
80 a discrepancy between the DelVGs present within an infected cell and those that get
81 packaged into DIPs was recently described for chikungunya virus (20).

82

83 To gain a more accurate, comprehensive understanding of DelVG and DIP formation, we
84 specifically examined the DelVGs that formed *de novo* during the first hours of IAV
85 infection. Careful analysis of hundreds of intracellular and extracellular DelVGs revealed
86 the enrichment of specific sequence elements at deletion breakpoints. We also observed
87 that DelVGs represent a much larger fraction of viral RNAs within the cell compared with
88 what gets packaged, suggesting that IAV DelVGs are packaged much less efficiently than
89 wild type genomic RNAs. While the magnitude of this effect varied somewhat between
90 genome segments, we observed no significant differences in the sizes or locations of

2

92   packaged DelVG deletions compared with intracellular DelVGs, indicating that the
93   bottleneck on DelVG packaging is not sequence-specific.
94
95   **RESULTS**
96   **Intracellular DelVGs generated early during infection are primarily derived from the**
97   **polymerase and HA segments**
98   Previous examinations of influenza DelVGs have looked specifically at the RNAs that get
99   successfully packaged into virions (DIPs). It is not actually clear how well the packaged
100   DelVGs observed within extracellular DIPs represent the total population of DelVGs
101   produced by the IAV replication machinery. To better understand the full array of deletions
102   commonly generated during IAV infection, and by extension the DelVG formation
103   process, we examined the distributions of deletion breakpoints found within intracellular
104   viral RNAs isolated early during infection.
105
106   We infected MDCK-SIAT1 cells at a multiplicity of infection (MOI) of 10 (based on Tissue
107   Culture Infectious Dose (TCID50)) with a recombinant stock of A/Puerto Rico/8/1934
108   (PR8) grown under conditions that minimize DIP content. We then harvested total cellular
109   RNA at 3 and 6 hours post infection (hpi) in order to capture the intracellular DelVGs
110   produced early during infection. We also extracted viral RNA from supernatants collected
111   at 24 hpi to capture DelVGs that were successfully packaged into DIPs. These RNAs,
112   along with genomic RNA extracted from the infecting viral stock, were used as templates
113   for whole genome RT-PCR and next generation sequencing as we have previously
114   described (17).
115
116   To focus our analysis on DelVGs that formed *de novo* during the experiment, we first
117   defined the specific DelVGs present within the inoculum and excluded them from
118   subsequent analyses **(Fig 1A)**. We also only considered deletion junctions if they were
119   represented by >5 reads within a given sample, a cutoff threshold that maximized junction
120   detection albeit with reduced correlation between technical replicates **(Fig 1B)**. We chose
121   to prioritize junction detection sensitivity in this study due to the low copy number and
122   read coverage of *de novo* generated DelVGs early during infection.
123
124   Using this approach, we identified hundreds of distinct DelVGs across all segments
125   except M and NS, in both extracellular and intracellular samples. For each time point, we
126   observed only partial overlap in the specific junctions shared between the three replicates,
127   suggesting significant stochasticity in the specific locations at which deletions form **(Fig**
128   **1C)**. When we examined the proportional distribution of junctions across the eight
129   genome segments, we found that these proportions were consistent across replicates,
130   indicating significant and reproducible variation in the intrinsic potential of the individual
131   genome segments to form DelVGs **(Fig 1D)**. Numerous previous studies have shown that
132   the majority of extracellular DelVGs found within DIPs are derived from the three
133   polymerase segments, PB2, PB1, and PA (17, 21, 22). Interestingly, we observed that
134   HA-derived DelVGs were roughly as numerous as polymerase-derived DelVGs at early
135   timepoints within infected cells but only constituted a small fraction of extracellular
136   DelVGs at 24 hpi, suggesting that population of DelVGs packaged into virions may not
137   accurately reflect the relative abundances of DelVGs produced within infected cells.
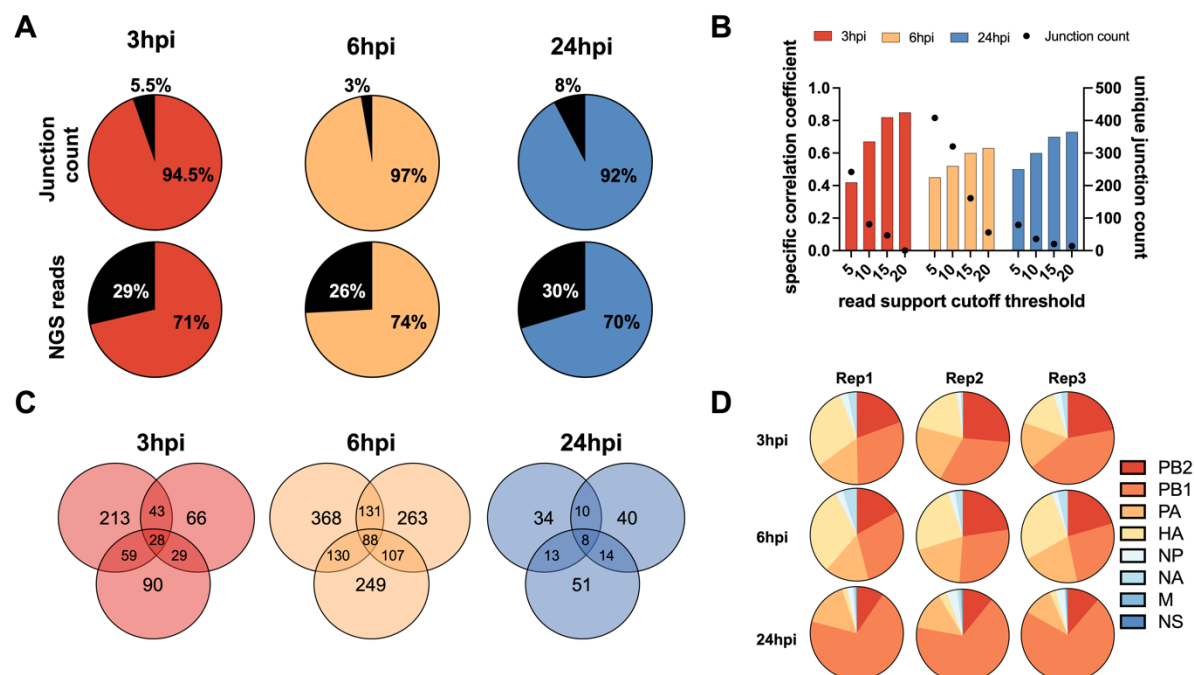
3

138
139

Fig 1. DelVG formation is partially stochastic and biased towards the polymerase and HA segments. (A) Pie charts show the percentages of the total normalized junction counts of all segments between the inoculum and viral populations from all time points at the levels of unique junction site (upper panel) and NGS reads (lower panel). The black fractions represent the junctions that were detected in the inoculum. **(B)** Correlation of distinct junctions in the PB2 segment between two technical replicate samples (generated from same RNA sample) using different NGS read cutoff values. **(C)** Venn diagram showing the overlap in specific junctions (for the PB2 segment) between three replicate samples collected at each time point. **(D)** Pie charts show the proportional abundance of normalized junction counts from each genome segment across replicates.

140
141

DelVG formation is not simply a function of segment length

142 It is not clear why some segments are more prone to DelVG formation than others, though
143 it has been postulated that this is simply a function of the segment length (23). To directly
144 test this hypothesis, we compared the normalized junction count between DelVGs
145 detected at 6 hpi and a perfect model that assumes a positive correlation between junction
146 count and the proportional length of each segment (**Fig 2A**). While the PB2 and PA
147 segments matched up well with the model predictions, the other segments deviated
148 significantly. These data demonstrate that rates of DelVG formation are not proportional
149 to segment length.

151

Direct sequence repeats and A/U bases are enriched at DelVG deletion junctions

152 By examining newly produced DelVGs isolated early during infection, we generated an
153 unbiased view of DelVG deletion formation by the IAV replicase. Previous reports have
154 described an enrichment of repeated sequences flanking DelVG deletions (termed "direct
155 repeats") and have hypothesized these direct repeats may promote DelVG formation by
156 facilitating RNA-dependent RNA polymerase (RdRp) re-engagement during the
157 replication process (15, 24). In support of these previous studies, we observed that
158 deletions lacking direct repeat sequences were significantly less abundant, and a subset

160   of direct repeat sequence lengths were more abundant in early intracellular DelVGs than
161   would be predicted if deletions were completely random **(Fig 2B)**.
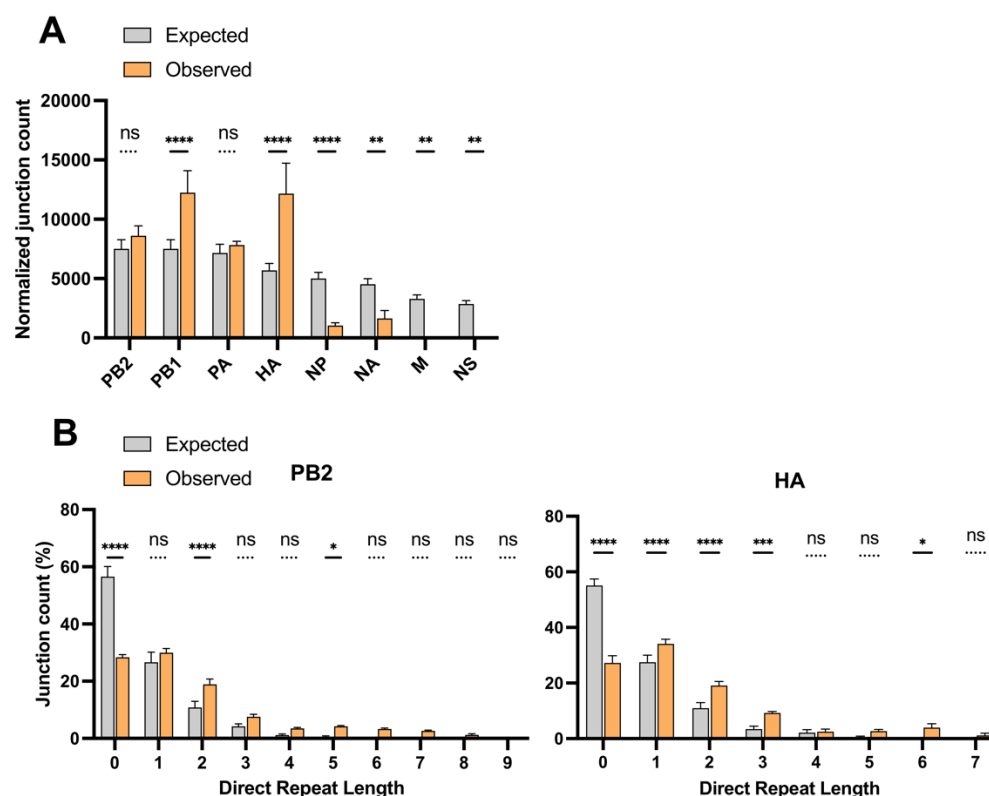162

Fig 2. Roles of segment length and direct repeat sequences in DelVG deletion formation. (A) Observed normalized junction counts per segment were compared with junction counts predicted by a model (expected) that assumes a simple positive correlation between segment length and DelVG junction count (see materials and methods). **(B)** The numbers of DelVG junctions detected at 6 hpi with no sequence repetition flanking the deletion breakpoints (Direct repeat length = 0), or with repeated sequences of varying length (Direct repeat length = 1-9) for the indicated segments (observed). Expected values plot the numbers of junctions with the indicated repeat lengths predicted from model simulations in which junction formation is random. Junction counts are plotted as a percentage of the total number of DelVGs detected for a given segment. Data are presented as means (n = 3 cell culture wells) ± standard deviations. * $P<0.05$, ** $P<0.01$, ***, $P < 0.001$; ****, $P < 0.0001$, ns = not significant (Two-way ANOVA).

163
164

165   We also asked whether specific sequence motifs were enriched at DelVG deletion
166   junctions. Using the 6 hpi samples, we extracted the sequences flanking each deletion
167   breakpoint in the pre-deletion, wild type sequence and calculated the proportion of each
168   nucleotide at each position **(Fig 3)**. To determine whether the observed nucleotide
169   frequencies deviated from what would be expected if deletion formation was sequence
170   independent, we performed the same analysis on three replicate sets of deletions
171   randomly generated *in silico*. Comparison of the observed nucleotide frequencies with the
172   null model predictions revealed clear enrichment of adenosines or uridines at the 5'

5

173 deletion breakpoint position (labeled "J" in figure 3). We also detected enrichment of either
174 adenosines or uridines at the position immediately upstream of both the 5' and 3'
175 breakpoints (position 4 in both R1 and R3). Finally, we observed enrichment of either
176 cytidines or guanosines within the R2 region downstream of the 5' breakpoint. Altogether,
177 these results suggest that DelVG formation is facilitated in part by the presence of direct
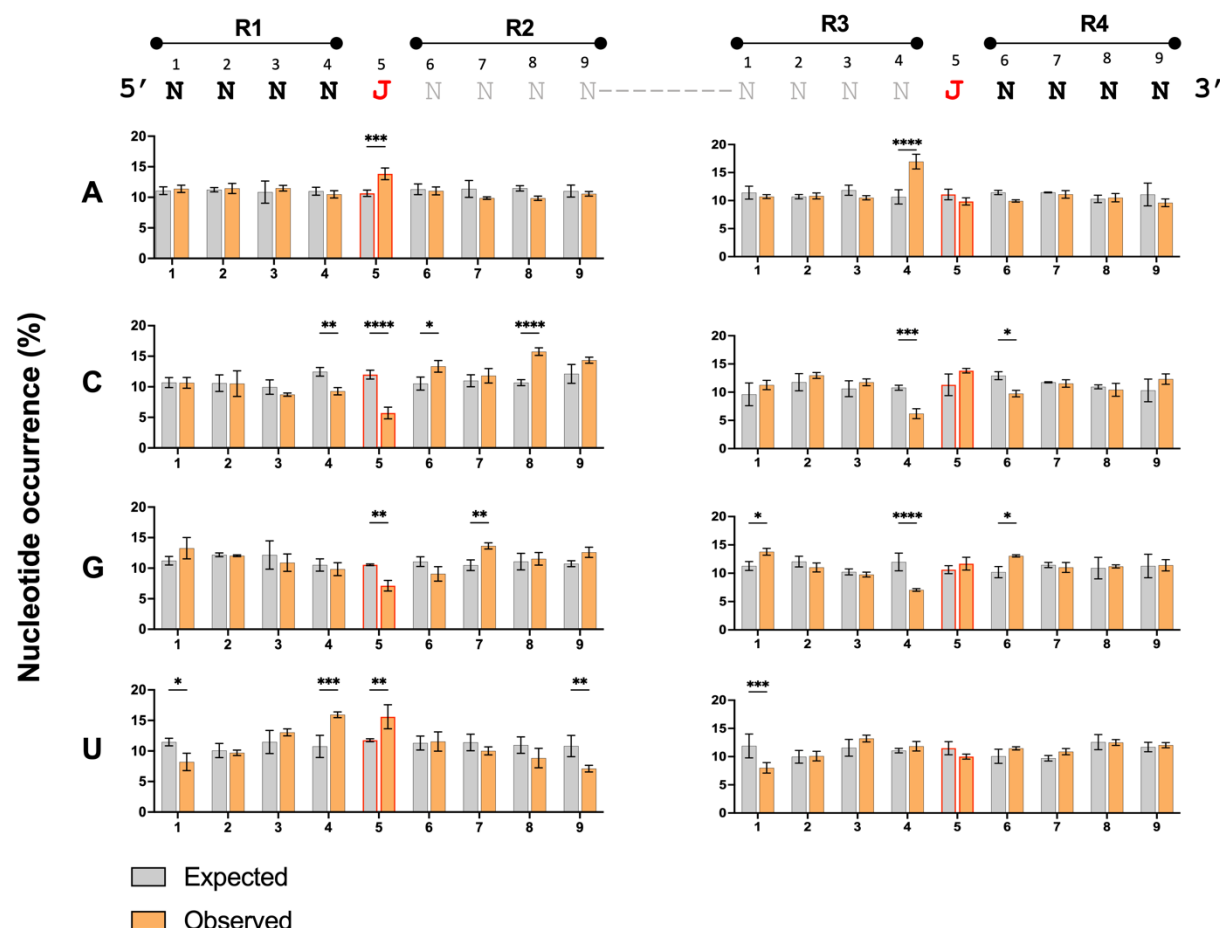178 sequence repeats and nucleotide composition surrounding the junction sites.
179

Fig 3. Enrichment of specific nucleotides at positions flanking DelVG deletions. The 4 nucleotides flanking DelVG junctions were numbered and divided into four regions (R1-R4) and the percentage occurrence of each nucleotide was calculated at each site within each region. The junction nucleotides immediately flanking the deletion are indicated by the red "J"'s at position 5 on the left and right. The grey nucleotides flank the junctions in the progenitor sequence but are lost through deletion in the actual DelVG sequence. The percentage of nucleotide occurrence (observed in PB2 6 hpi) at each site was plotted against a random control (expected). Data are presented as means (n = 3 cell culture wells) ± standard deviations. * P<0.05, ** P<0.01 ***, P < 0.001; ****, P < 0.0001 (Two-way ANOVA), ns = not significant.

180
181
182 **Intracellular DelVGs are inefficiently packaged**
183 Our finding that HA-derived DelVGs make up a large fraction of intracellular but not
184 extracellular DelVGs suggested that some DelVGs might be packaged into virions more
185 efficiently than others. To further investigate this, we compared both the normalized

6

186   numbers of unique junctions and read support for DelVGs from each segment between
187   intracellular (3 and 6 hpi) and extracellular (24 hpi) populations **(Fig 4A)**.
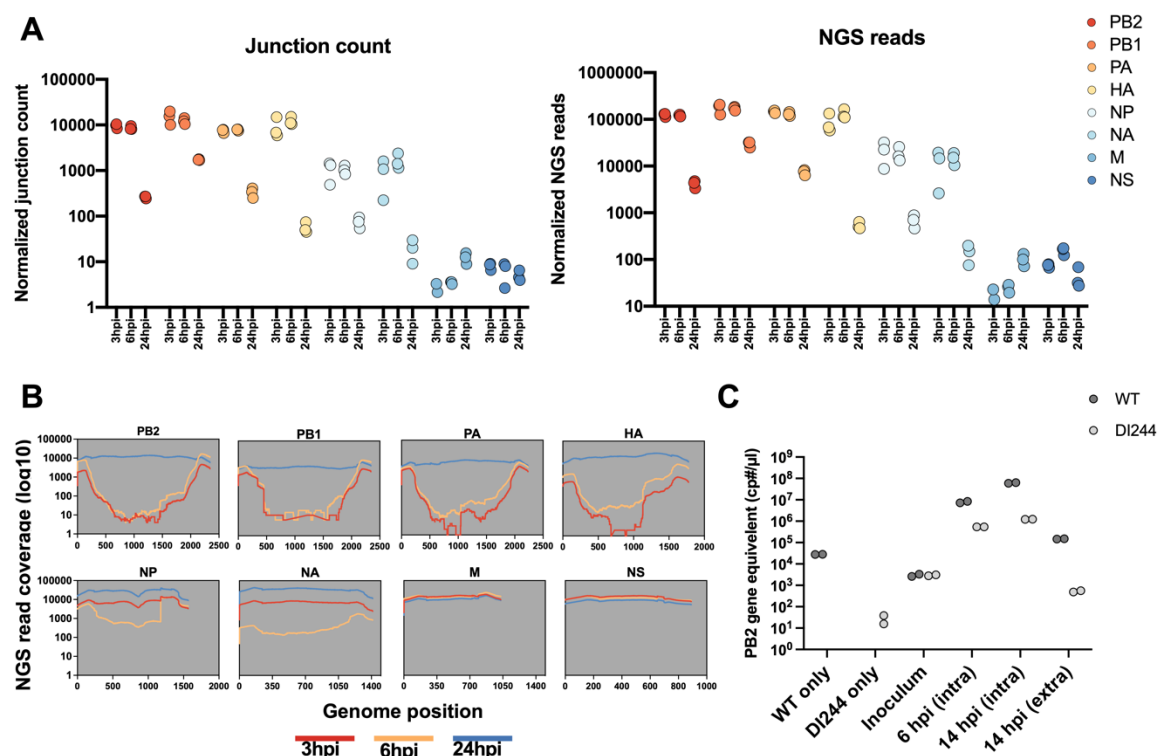188

Fig 4. DelVGs are inefficiently packaged compared with wild type vRNAs. (A) Numbers of DelVG junction counts (left panel) and NGS reads (right panel) were counted for three replicates in each segment at each indicated time point, normalized to 106 mapped reads. Each dot represents a replicate. **(B)** The NGS read coverage per nucleotide was plotted for each segment at each time point (the average coverage of the three replicates per position). **(C)** Competition between WT and DI244 PB2 segments during infection and packaging. MDCK-SIAT1 cells were co-infected with WT PR8 and DI244 at a 1:1 ratio to a final MOI of 20 PB2 gene equivalents/cell under single cycle conditions. Absolute copy numbers of WT and DI244 PB2 segments were quantified in by RTqPCR within either cell lysates (intracellular), supernatant (extracellular), or inoculum mixture at the indicated timepoints post-infection. PR8-only and DI244 only controls are also shown. Data represent absolute values from individual cell culture wells and are representative of two independent experiments.

189
190
191   While normalized DelVG junction count numbers were highly similar for individual
192   segments between 3 hpi and 6 hpi, we observed a 10-100 fold decrease in normalized
193   junction counts by 24 hpi for segments 1-6 **(Fig 4A, left panel)**. The same trend was also
194   observed when we compared DelVG read support across timepoints **(Fig 4A, right**
195   **panel)**. These data suggested that DelVGs constitute a much smaller proportion of viral
196   RNAs packaged into virions, compared with what is present within the infected cell.
197
198   To confirm this observation, we examined the per-nucleotide read coverage for each
199   segment. Since DelVGs are missing large internal regions of the genomic RNA sequence
200   but still retain the segment termini, they produce a characteristic "devil horns" read
201   coverage pattern, where read coverage is substantially lower in the middle of gene
202   segment compared with the termini (25). This pattern was obvious in intracellular samples
203   from 3 hpi and 6 hpi for segments 1-4, to a lesser extent for segments 5 and 6, and absent
204   from segments 7 and 8, all consistent with the observed variation in abundances of

205   DelVGs associated with each segment **(Fig 4B)**. In contrast, read coverage was
206   consistent across the length of all eight segments in extracellular samples from 24 hpi,
207   suggesting that DelVGs are much less abundant within this population of viral RNA.
208
209   To directly quantify the relative packaging efficiencies of DelVGs and WT vRNAs, we
210   performed an *in vitro* competition assay between WT PR8 and a recombinant stock of
211   DI244, a well characterized DIP/DelVG that harbors a 1946 nucleotide internal deletion
212   in the PB2 segment (26, 27), using qPCR primer/probe-sets specific for either the WT or
213   DI244 versions of the PB2 gene segment **(Fig 4C)**. We co-infected MDCK-SIAT1 cells
214   with a 1:1 ratio (based on PB2 gene equivalents) of WT PR8 and DI244 at a combined
215   MOI of 20 PB2 gene equivalents/cell under single cycle conditions where secondary
216   spread was not permitted. We confirmed that the absolute copy numbers of both WT and
217   DI244 PB2 gene equivalents present in the inoculum were equivalent using RTqPCR.
218
219   Quantification of both WT- and DI244-derived PB2 genome equivalents within infected
220   cells revealed that WT PB2 was ~50-fold more abundant than DI244 PB2 by 14 hpi,
221   indicating that the WT PB2 segment enjoyed a significant replicative advantage over
222   DI244. Quantification of extracellular RNA at the same timepoint (14 hpi) revealed an
223   even more pronounced, ~280-fold advantage for the WT PB2 segment compared with
224   DI244 **(Fig 4C)**. Collectively, these data suggest that DelVGs make up a much smaller
225   fraction of viral RNAs that get packaged into virions compared with what is present within
226   the infected cell, consistent with a packaging defect relative to WT vRNAs.
227
228   **No significant differences in deletion breakpoints between intracellular and**
229   **packaged DelVGs**
230   The discrepancies in both proportional abundance and distribution amongst the segments
231   that we observed between intracellular and extracellular DelVGs suggested the existence
232   of a significant bottleneck limiting packaging of DelVGs relative to wild type vRNAs. We
233   hypothesized that this might be due to the potential loss of sequences required for efficient
234   packaging during the formation of some DelVGs. If true, we expected that specific regions
235   of gene segments required for maximal packaging efficiency would be retained within
236   packaged extracellular DelVGs, but largely missing from intracellular DelVGs.
237
238   To test this, we first examined the positions of deletion junctions from intracellular (6hpi)
239   and extracellular (24 hpi) DelVGs **(Fig 5A)**. As we and others have previously reported
240   for packaged extracellular DelVGs, deletion junctions from both populations clustered
241   within clear hot spots near the segment termini (25, 28, 29). The extracellular DelVG
242   junction distribution appeared narrower than that of the 6 hpi intracellular DelVGs, but
243   that could be an artifact of the greater overall abundance of intracellular DelVGs. To more
244   rigorously evaluate whether the locations of deletion junctions differed between
245   intracellular and extracellular DelVGs, we used a cumulative score method that allowed
246   us to examine the proportional distribution of deletion junctions as a function of nucleotide
247   position **(Fig 5B)**. This approach clearly demonstrated that the distribution of deletions is
248   statistically indistinguishable between intracellular and extracellular DelVGs (P=0.9; t-
249   test), suggesting that DelVGs that retain specific features or sequences are not selected
250   for during the packaging process. Finally, we asked whether the sizes of DelVGs varied

251 between the three time points tested and again found no significant difference (p=0.57; t-
252 test) **(Fig 5C)**, indicating that DelVG size is not a determinant of packaging efficiency.
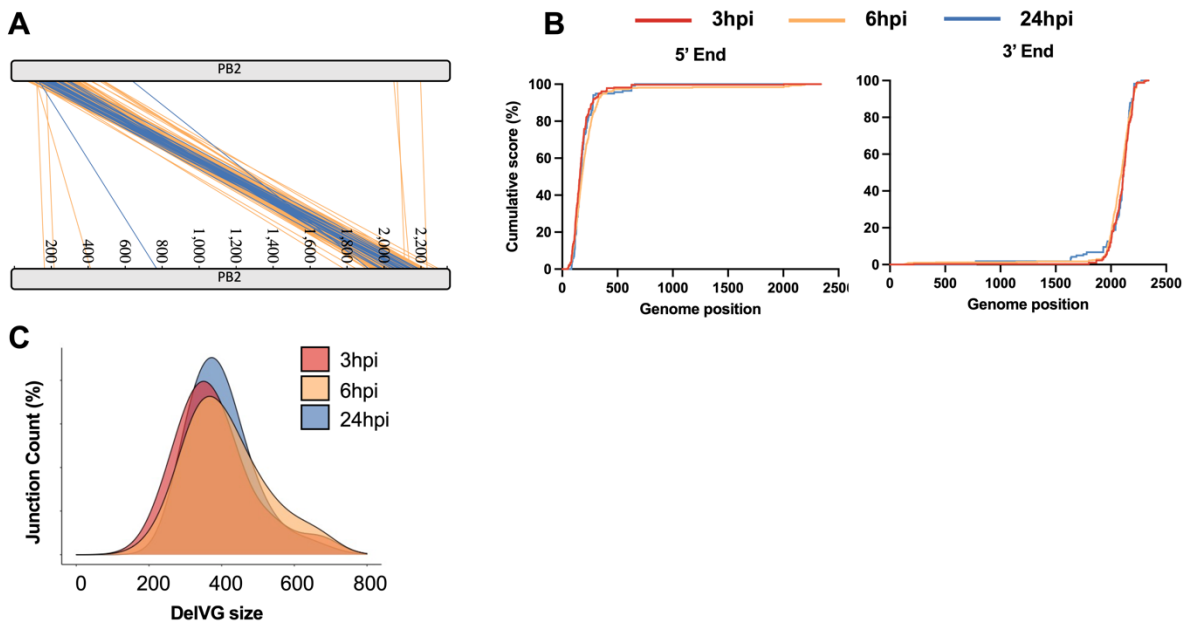253

Fig 5. No differences in deletion size or breakpoint locations between intracellular and packaged DelVGs. (A) PB2 DelVG deletion junction sites at 6 and 24 hpi were mapped to the genome position with each line representing a distinct deletion and the colors indicating the time point (orange = 6 hpi, blue = 24 hpi). **(B)** Plots show the cumulative occurrence of DelVG deletions as a function of gene segment position. Cumulative score was calculated by starting at zero at the end of the segment and then adding a score of 1 at each nucleotide where a unique junction breakpoint occurred. Scores were normalized by calculating the percentage of final value reached at each position. **(C)** DelVG size distributions (PB2 segment only shown) from samples collected at the indicated time points were plotted using R density plot (ggplot package). The area under the curve is equal to 100 percent of all probabilities.

254
255
256 **DISCUSSION**
257 Despite recent improvements in our fundamental understanding of the structure and
258 function of the IAV polymerase complex, we still don't know how or why DelVGs and DIPs
259 form during IAV replication (11, 30–33). Here, we used a robust combined NGS/analysis
260 pipeline to analyze the first wave of DelVGs that form within cells during infection, thus
261 providing the first unbiased view of *de novo* DelVG production by the IAV replicase. We
262 compared these intracellular DelVGs to the population of DelVGs that get packaged into
263 virions, revealing a significant bottleneck in DelVG packaging relative to wild type vRNAs.
264 Our data contradict to the dogma that DelVGs outcompete wild type vRNAs for packaging
265 and suggest that the accumulation of DIPs within IAV populations must occur through
266 other mechanisms.
267
268 In agreement with previous studies, we found that the majority of extracellular DelVGs
269 junctions were derived from the three polymerase segments (24). Within intracellular
270 DelVG populations however, the abundance of HA-derived DelVGs was comparable to
271 that of the polymerase-derived DelVGs, suggesting that DelVGs from segments 1-4 are
272 generated at similar rates but that the HA segment packaging efficiency is much more
273 sensitive to deletions than the polymerase segments. It has been suggested that this bias
274 in DelVG formation across the segments is a function of segment length, potentially

9

275 because DelVGs derived from longer segments have a greater length differential
276 compared with their wild type parents, resulting in a more pronounced replication
277 advantage (3, 34). Our findings that the enrichment of DelVGs from segments 1-4 is
278 already apparent by 3 hpi, suggests that this bias is emerging from the formation process
279 rather than replication. More work is needed to identify the specific determinant(s) that
280 influence the uneven distribution of DelVG formation across genome segments.
281
282 DelVG formation is thought to occur when the viral replicase pauses synthesis of the
283 daughter vRNA or cRNA but continues processing along the template and re-initiates
284 synthesis at a downstream site on the same template (3, 35). This process is not
285 completely random, as the vast majority of deletions start and stop within hotspots near
286 the segment termini, and individual segments vary greatly in DelVG formation (24, 25,
287 36). We observed the same distribution in intracellular DelVGs, indicating that these
288 hotspots reflect that what is produced by the viral RdRp and are not biased by selection
289 in the packaging process. Altogether, our data suggest that specific regions of the viral
290 genome are uniquely prone to DelVG formation, for reasons that are still not understood.
291
292 It has been suggested that the polymerase translocation is promoted by the presence of
293 a direct sequence repeat and a specific nucleotide composition at the junction site (6, 15,
294 24). We demonstrate significant enrichment of direct sequence repeats and A/U
295 nucleotides at DelVG deletion junctions, along with C/G nucleotides downstream of the
296 5' deletion breakpoint. Interestingly some of these nucleotides are located within the
297 regions that are not retained in DelVG final product (R2 and R3 in Fig 3), suggesting
298 possible roles for the sequence composition both upstream and downstream of junction
299 sites at both ends of the viral genome. These specific template sequence features likely
300 enhance the probability of RdRp translocation occurring, however these features are not
301 absolutely required, as large numbers of DelVG junctions that lack flanking direct repeats
302 or A/U bases can easily be observed. There also appears to be a significant degree of
303 stochasticity in the specific nucleotides at which deletions form, based on the limited
304 degree of overlap in breakpoint locations between replicates.
305
306 DelVGs/DIPs are known for their ability to inhibit the replication of WT virus, and it is
307 widely believed that this effect is partially driven by DelVGs outcompeting WT vRNAs for
308 packaging (37). Our data strongly suggest that the opposite is true: that DelVGs are
309 inefficiently packaged into virions compared with WT vRNAs. This finding complicates our
310 understanding of how DIPs outcompete WT virus at the population level, suggesting that
311 other advantages must help DelVGs/DIPs offset their packaging deficiencies.
312
313 Given the similar characteristics of deletions between intracellular and extracellular
314 DelVG populations, it appears that the effects of DelVG-associated deletions on
315 packaging efficiency are independent of deletion size or location. Our data suggest that
316 all DelVGs generated by the IAV RdRp are lacking some determinant required for
317 maximal packaging. IAV genome packaging is governed by multiple, discontinuous
318 regions that act in *cis* and in *trans* to facilitate the efficient and selective incorporation of
319 a single copy of each genome segment into the vast majority of virions (29). Each
320 segment contains packaging and bundling sequences that span both coding and non-

10

321   coding regions of the segment termini (38–42). Beyond the well-described packaging
322   signals in the segment termini, additional packaging determinants exist within the interiors
323   of some segments (43–46). We observed no enrichment of DelVGs that retained longer
324   stretches of the segment termini, suggesting the decreased efficiency of DelVG
325   packaging is due to the loss of internal packaging determinants.
326
327   Our analysis of the initial wave of DelVGs produced *de novo* during IAV infection provides
328   critical insights into the formation and packaging of both DelVGs and WT RNAs. This
329   approach generated a detailed portrait of the full range of DelVG produced by the IAV
330   replicase, allowing us to show that DelVGs are inefficiently packaged into particles
331   relative to WT, contrary to dogma. Additionally, we showed that DelVG formation is not
332   influenced by the segment length and partially influenced by sequence context. Several
333   fundamental questions remain, however, including the specific mechanism that triggers
334   DelVG formation and the factors that cause DelVGs to only form at defined hotspots and
335   in higher frequency within some segments than the others.
336

337 **MATERIALS AND METHODS**
338 **Virus and cells**
339   MDCK-SIAT1 and HEK293T cells were grown in minimal essential medium (MEM) plus
340   GlutaMax (Gibco), supplemented with 8.3% fetal bovine serum (Seradigm), at 37°C and
341   5% CO_2. Recombinant A/Puerto Rico/8/1934 (PR8) was generated from HEK293T cells
342   through standard influenza virus 8-plasmid reverse genetics transfection. Undiluted
343   transfection supernatants were directly inoculated onto MDCK-SIAT1 cells, and
344   supernatants were harvested at first signs of cytopathic effect (CPE) to generate seed
345   stocks. Working stocks of virus were generated by infecting MDCK-SIAT1 cells with seed
346   stock at an MOI of 0.0001 TCID50/cell and harvesting 48 hpi. Supernatant was clarified
347   at 3000 RPM for 5 minutes and 500 µL aliquots were stored at -70°C.
348

349 **Generation of DelVG through high MOI infection**
350   Confluent MDCK-SIAT1 were infected in triplicate with PR8 at an MOI of 10 TCID50/cell.
351   To harvest intracellular viral RNA at 3 and 6 hpi, cells were washed twice with PBS, and
352   RNA was extracted using the QIAgen RNeasy kit following the manufacturer's
353   instructions. To extract extracellular RNA from packaged virions, supernatant was
354   collected from infected cells at 24 hpi. After clarification, 140µl of supernatant was used
355   for RNA extraction using the Qiagen QIAamp viral RNA mini kit per the manufacturer's
356   instructions. All RNA was stored at -70°C.
357

358 **Viral cDNA amplification and sequencing**
359   Universal RT-PCR was performed on all the samples using a previously described
360   method (17, 47) . Briefly, 3µL of RNA was mixed with 1µL (2µM) MBTUni-12 primer + 1µL
361   (10µM) dNTPs + 8µL dH2O. Reactions were incubated as previously described, and
362   finally 1µL SuperScript III RT (Invitrogen), 4µL of 5X First-Strand Buffer, 1µL of DTT, 1µL
363   RNase-in (Invitrogen) were added before the final incubation. From this, 5µl was used
364   with 2.5µL (10µM) of each universal primer, 0.5µL Phusion polymerase (NEB), 10µL - 5x
365   HF buffer, 1µL (10mM dNTPs mix), and 28.5µL dH2O. The PCR conditions used were
366   98°C (30 s) followed by 25 cycles of 98°C (10 s), 57°C (30 s), and 72°C (90 s); a terminal

367 extension of 72°C (5 min); and a final 10°C hold. Finally, PCR amplicons were purified
368 using the Qiagen QIAquick PCR Purification Kit and sequenced with paired-ends 2x250nt
369 reads on an Illumina MiSeq using V2 chemistry (17).
370

371 **Generation of recombinant DI244**
372 We synthesized (Integrated DNA technologies, Inc.) and cloned the full-length DI244
373 sequence (NCBI # L41510.1) into the pDZ vector and transfected it along with 7 plasmids
374 encoding WT versions of segments 2-8 plasmids into PB2-expressing HEK293 cells
375 (HEK293-PB2), using the standard 8-plasmid reverse genetics approach. Briefly, 60 to
376 90% confluent PB2-expressing HEK293 cells were transfected with 500 ng of various
377 plasmids (pDZ::PB2-DI244, pDZ::PB1, pDZ::PA, pDZ::HA, pDZ::NP, pDZ::NA, pDZ::M,
378 and pDZ::NS) by using JetPrime (Polyplus) according to the manufacturer's instructions.
379 Under these conditions, no WT virus is generated due to the absence of WT PB2
380 segment. Transfection supernatant was used to infect PB2-expressing MDCK cells
381 (MDCK-PB2) for 48 hrs to generate a seed stock. HEK293-PB2 and MDCK-PB2 cells
382 were kindly provided by Dr. Stefan Pöhlmann and are described in this reference
383 (5). DI244 seed stock was titered on MDCK-PB2 cells and used to infect MDCK-PB2 cells
384 at an MOI of 0.001 TCID50/cell for 72 hrs to generate a working stock. Deep sequencing
385 of the DI244 working stock confirmed the presence of the expected DI244 deletion
386 junction and the absence of WT PB2 segment (data not shown).
387

388 **RTqPCR quantification of DI244 and WT PB2 gene segments**
389 We designed and optimized specific primer/probesets (**Table S1**) specific for either DI244
390 or WT PB2 using the IDT PrimerQuest webtool and validated efficiency and specificity
391 using serial dilutions of plasmids encoding either WT PB2 or DI244. Viral RNA was
392 extracted from cells or virions as described above and used to synthesize cDNA using
393 the universal primer described above and *Verso* cDNA synthesis Kit (Thermo Fisher). 3µl
394 of RNA was mixed with 8µl H20, 4µl 5X cDNA Synthesis Buffer, 2µl dNTP Mix (5 mM
395 each), 1µl universal primer (10µM), 1µl RT Enhancer, 1µl Verso Enzyme Mix, before
396 incubating for 50 minutes at 45°C. After this, 1µl of the of cDNA product was mixed with
397 7µl H20, 1µl forward primers (18µM), 1µl reverse primer (18µM), 1µl specific probe (5µM),
398 and 10µl TaqMan Fast Advanced Master Mix (Thermo Fisher). The qPCR conditions used
399 were: 50°C (2min), 95°(2 min), followed by 40 cycles of 95°C (1 s) and 61°C (20 s) using
400 a qPCR QuantStudio 3 thermocycler.
401

402 **Table S1: qPCR primers and probes used for the competition assay**
403

Primer/probe names	Sequence	Dye
PB2-WT FW-primer (920-940)	TTAGGCAGAACCCAACAGAAG	
PB2-WT reverse-primer (1040-1020)	ACTGATGATCCGCTTGTTCTC	
PB2-WT Probe (956-979)	TATGCAAGGCTGCAATGGGACTGA	5'Yakima yellow/ZEN/3'IBFQ
PB2-DI244 forward-primer (92-109)	AAACCACCGTGGACCATA	
PB2-DI244 Reverse-primer (2169-2148)	GCTCAGTTCATTGATGCTTAGT	
PB2-DI244 Probe (135-151 and 2098-2103)	AGGAAGACAGGAGAAGACTGAGG	5'FAM/ZEN/3'IBFQ

404

405 *In vitro* **competition assay**
406 MDCK-SIAT1 cells were coinfected in duplicate at an MOI of 10 PB2 gene
407 equivalents/cell with a 1:1 ratio of WT PR8 and DI244 (ratio based on PB2 gene

408 equivalents). Fractions of the inoculum mixture both before (0 hpi) and after the 1 hour
409 adsorption phase were set aside for RTqPCR. After adsorption for 1 hour at 4°C, inoculum
410 was removed, cells were washed, and MEM+FBS was added to cells. At 3 hpi,
411 neutralizing anti-HA monoclonal antibody H36-26 was added to each well to a final
412 concentration of 25ug/mL to block secondary spread. Cellular RNA was extracted from
413 duplicate wells of cells at 6 hpi and at 14hpi using the QIAgen RNeasy kit following the
414 manufacturer's instructions. At the 14hpi timepoint, viral RNA was also collected from
415 supernatants using the Qiagen QIAamp viral RNA mini kit per the manufacturer's
416 instructions. Control infections with only WT PR8 or only DI244 were also performed in
417 parallel.
418
419 **Sequencing analysis of deletion junctions**
420 Raw sequencing reads were fed into our DI-detection pipeline for junction detection and
421 characterization (17). Briefly, the pipeline uses set of algorithms including Bowtie (48) and
422 ViReMa (49) to perform two major steps. First, filtering the NGS reads based on quality
423 and origin (i.e. cellular or viral). Second, mapping and detecting the deletion junction and
424 reporting the start and end sites for each junction in each gene segment, along with the
425 number of NGS reads detected for each junction. The pipeline is accessible at:
426 https://github.com/BROOKELAB/Influenza-virus-DI-identification-pipeline.   NCBI
427 accession numbers for the PR8 reference genome used in this study: PB2=AF389115.1,
428 PB1=AF389116.1,   PA=AF389117.1,   HA=AF389118.1,   NP=AF389119.1,
429 NA=AF389120.1, M=AF389121.1, NS= AF389122.1
430
431 To account for sequencing read coverage variations between libraries, we normalized the
432 junction count and their NGS reads to 10^6 mapped reads per library per segment. For the
433 junction count, the number of junctions per segment was multiplied by 10^6 and divided by
434 the number of NGS reads aligned to the WT gene of the given segment in a given library.
435 The same was done to normalize for the NGS reads, where the junctions' read number
436 per segment was multiplied by 10^6 and divided by the number of NGS reads aligned to
437 the WT genome of the given segment plus the number of reads mapped to the junctions.
438
439 The direct repeat sequence lengths were extracted from the output file
440 "Virus_Recombination_Results.txt" generated by ViReMa algorithm. For the random
441 control, the junction sites were randomized using Excel function "=RANDBETWEEN()"
442 based on the actual sequence range and number detected in the 6 hpi population. Next,
443 a custom Perl code was used to extract their sequences from the corresponding PR8
444 gene segment (PB2 and HA). Finally, the direct repeat lengths were extracted and
445 compared to the real samples.
446
447 To analyze the nucleotide composition at the junction site, we analyzed the sequences
448 flanking the junction sites for enrichment of specific nucleotides. There are four possible
449 sequence regions that possibly involved in promoting the polymerase translocation, two
450 regions flanking the junction from each site, we numbered them region 1-4 **(Fig 3)**. From
451 these regions only, R 1& 4 are retained within the DelVG final product, while R 2 & 3 are
452 not, and their potential importance stems from their physical proximity to the junction. A
453 custom Perl code was used to extract 4 nucleotides plus the junction from each region
454 from all the detected DelVG in the three replicates of the 6 hpi population. Next the

13

455 WebLogo platform (50) was used to measure the percentage occurrences of each
456 nucleotide at each position. To decide whether the percentage occurrence differed from
457 what would be expected in the absence of nucleotide enrichment, we generated a random
458 control samples the same way as in the direct repeats. Finally, the two samples:
459 observed/experimental and expected/computational were compared using Anova Two-
460 way. To confirm the validity of this approach we found the same results when repeated
461 the analysis for 3 hpi time point and different segments. Additionally, we found no
462 significant difference at all nucleotide position upon comparing two random samples,
463 three replicates each (data not shown).

465 The percentage length of each segment was calculated based on the total genome length
466 13585nts (*e.g.* PB2=17.2% and NS=6.5%). Next, the percentage length of each segment
467 was used to calculate the number of junctions per segment based on the total normalized
468 number of junctions of each sample (expected). Finally, these values were compared to
469 the observed values from the actual experiments.

471 **Data availability**
472 All NGS data sets generated in this study can be found under BioProject accession
473 number PRJNA725907.

475 **Acknowledgment**

484 **REFERENCES**
485 1. Brooke CB. 2017. Population Diversity and Collective Interactions during Influenza
486   Virus Infection. J Virol 91.

487 2. Brooke CB, Ince WL, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW.
488   2013. Most influenza a virions fail to express at least one essential viral protein. J
489   Virol 87:3155–3162.

490 3. Vignuzzi M, López CB. 2019. Defective viral genomes are key drivers of the virus-
491   host interaction. Nat Microbiol 4:1075–1087.

492 4. Dimmock NJ, Easton AJ. 2014. Defective interfering influenza virus RNAs: time to
493   reevaluate their clinical potential as broad-spectrum antivirals? J Virol 88:5217–
494   5227.

495 5. Bdeir N, Arora P, Gärtner S, Hoffmann M, Reichl U, Pöhlmann S, Winkler M. 2019.
496   A system for production of defective interfering particles in the absence of
497   infectious influenza A virus. PLoS ONE 14:e0212757.

14

6. Nayak DP, Chambers TM, Akkina RK. 1985. Defective-Interfering (DI) RNAs of Influenza Viruses: Origin, Structure, Expression, and Interference, p. 103–151. *In* Cooper, M, Eisen, H, Goebel, W, Hofschneider, PH, Koprowski, H, Melchers, F, Oldstone, M, Rott, R, Schweiger, HG, Vogt, PK, Wilson, I (eds.), Current Topics in Microbiology and Immunology. Springer, Berlin, Heidelberg.

7. Laske T, Heldt FS, Hoffmann H, Frensing T, Reichl U. 2016. Modeling the intracellular replication of influenza A virus in the presence of defective interfering RNAs. Virus Research 213:90–99.

8. Felt SA, Sun Y, Jozwik A, Paras A, Habibi MS, Nickle D, Anderson L, Achouri E, Feemster KA, Cárdenas AM, Turi KN, Chang M, Hartert TV, Sengupta S, Chiu C, López CB. 2021. Detection of respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes. Nat Microbiol https://doi.org/10.1038/s41564-021-00882-3.

9. Vasilijevic J, Zamarreño N, Oliveros JC, Rodriguez-Frandsen A, Gómez G, Rodriguez G, 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 Pathogens 13:e1006650.

10. von Magnus P. 1954. Incomplete Forms of Influenza Virus, p. 59–79. *In* Smith, KM, Lauffer, MA (eds.), Advances in Virus Research. Academic Press.

11. Alnaji FG, Brooke CB. 2020. Influenza virus DI particles: Defective interfering or delightfully interesting? PLoS Pathog 16:e1008436.

12. Kupke SY, Riedel D, Frensing T, Zmora P, Reichl U. 2019. A Novel Type of Influenza A Virus-Derived Defective Interfering Particle with Nucleotide Substitutions in Its Genome. Journal of Virology 93.

13. Te Velthuis AJW, Long JC, Bauer DLV, Fan RLY, Yen H-L, Sharps J, Siegers JY, Killip MJ, French H, Oliva-Martín MJ, Randall RE, de Wit E, van Riel D, Poon LLM, Fodor E. 2018. Mini viral RNAs act as innate immune agonists during influenza virus infection. Nat Microbiol 3:1234–1242.

14. Boussier J, Munier S, Achouri E, Meyer B, Crescenzo-Chaigne B, Behillil S, Enouf V, Vignuzzi M, van der Werf S, Naffakh N. 2020. RNA-seq accuracy and reproducibility for the mapping and quantification of influenza defective viral genomes. RNA 26:1905–1918.

15. Lui W-Y, Yuen C-K, Li C, Wong WM, Lui P-Y, Lin C-H, Chan K-H, Zhao H, Chen H, To KKW, Zhang AJX, Yuen K-Y, Kok K-H. 2019. SMRT sequencing revealed the diversity and characteristics of defective interfering RNAs in influenza A (H7N9) virus infection. Emerg Microbes Infect 8:662–674.

16. Timm C, Akpinar F, Yin J. 2014. Quantitative characterization of defective virus emergence by deep sequencing. J Virol 88:2623–2632.

17. 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.

18. Sun Y, Kim EJ, Felt SA, Taylor LJ, Agarwal D, Grant GR, López CB. 2019. A specific sequence in the genome of respiratory syncytial virus regulates the generation of copy-back defective viral genomes. PLoS Pathog 15:e1007707.

19. Genoyer E, López CB. 2019. Defective Viral Genomes Alter How Sendai Virus Interacts with Cellular Trafficking Machinery, Leading to Heterogeneity in the Production of Viral Particles among Infected Cells. J Virol 93.

20. Langsjoen RM, Muruato AE, Kunkel SR, Jaworski E, Routh A. 2020. Differential Alphavirus Defective RNA Diversity between Intracellular and Extracellular Compartments Is Driven by Subgenomic Recombination Events. mBio 11:e00731-20, /mbio/11/4/mBio.00731-20.atom.

21. Heterologous Protection of Mice from a Lethal Human H1N1 Influenza A Virus Infection by H3N8 Equine Defective Interfering Virus: Comparison of Defective RNA Sequences Isolated from the DI Inoculum and Mouse Lung | Elsevier Enhanced Reader.

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

23. 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.

24. Saira K, Lin X, DePasse JV, 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. Journal of Virology 87:8064–8074.

25. Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses | Journal of Virology.

26. Noble S, Dimmock NJ. 1995. Characterization of putative defective interfering (DI) A/WSN RNAs isolated from the lungs of mice protected from an otherwise lethal respiratory infection with influenza virus A/WSN (H1N1): a subset of the inoculum DI RNAs. Virology 210:9–19.

27. Dimmock NJ, Easton AJ. 2015. Cloned Defective Interfering Influenza RNA and a Possible Pan-Specific Treatment of Respiratory Virus Diseases. Viruses 7:3768–3788.

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

29. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. 2010. Genome packaging in influenza A virus. Journal of General Virology, 91:313–328.

30. te Velthuis AJW, Fodor E. 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nature Reviews Microbiology 14:479–493.

31. Fodor E, te Velthuis AJW. 2020. Structure and Function of the Influenza Virus Transcription and Replication Machinery. Cold Spring Harb Perspect Med 10:a038398.

32. Fan H, Walker AP, Carrique L, Keown JR, Serna Martin I, Karia D, Sharps J, Hengrung N, Pardon E, Steyaert J, Grimes JM, Fodor E. 2019. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature 573:287–290.

33. te Velthuis AJW, Grimes JM, Fodor E. 2021. Structural insights into RNA polymerases of negative-sense RNA viruses. Nat Rev Microbiol https://doi.org/10.1038/s41579-020-00501-8.

34. Frensing T. 2015. Defective interfering viruses and their impact on vaccines and viral vectors. Biotechnology Journal 10:681–689.

35. Furusawa Y, Yamada S, da Silva Lopes TJ, Dutta J, Khan Z, Kriti D, van Bakel H, Kawaoka Y. 2019. Influenza Virus Polymerase Mutation Stabilizes a Foreign Gene Inserted into the Virus Genome by Enhancing the Transcription/Replication Efficiency of the Modified Segment. MBio 10.

36. Jennings PA, Finch JT, Winter G, Robertson JS. 1983. Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? Cell 34:619–627.

37. Duhaut SD, McCauley JW. 1996. Defective RNAs inhibit the assembly of influenza virus genome segments in a segment-specific manner. Virology 216:326–337.

38. Muramoto Y, Takada A, Fujii K, Noda T, Iwatsuki-Horimoto K, Watanabe S, Horimoto T, Kida H, Kawaoka Y. 2006. Hierarchy among Viral RNA (vRNA) Segments in Their Role in vRNA Incorporation into Influenza A Virions. J Virol 80:2318–2325.

39. Goto H, Muramoto Y, Noda T, Kawaoka Y. 2013. The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. J Virol 87:11316–11322.

608 40. Fujii Y, Goto H, Watanabe T, Yoshida T, Kawaoka Y. 2003. Selective incorporation
609   of influenza virus RNA segments into virions. Proceedings of the National Academy
610   of Sciences 100:2002–2007.

611 41. Marsh GA, Hatami R, Palese P. 2007. Specific residues of the influenza A virus
612   hemagglutinin viral RNA are important for efficient packaging into budding virions. J
613   Virol 81:9727–9736.

614 42. Gao Q, Palese P. 2009. Rewiring the RNAs of influenza virus to prevent
615   reassortment. Proc Natl Acad Sci U S A 106:15891–15896.

616 43. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, Laederach A,
617   Brown LE, Fodor E, Bauer DLV. 2019. The structure of the influenza A virus
618   genome. Nat Microbiol 4:1781–1789.

619 44. Gavazzi C, Yver M, Isel C, Smyth RP, Rosa-Calatrava M, Lina B, Moules V,
620   Marquet R. 2013. A functional sequence-specific interaction between influenza A
621   virus genomic RNA segments. Proceedings of the National Academy of Sciences
622   110:16604–16609.

623 45. Gilbertson B, Zheng T, Gerber M, Printz-Schweigert A, Ong C, Marquet R, Isel C,
624   Rockman S, Brown L. 2016. Influenza NA and PB1 Gene Segments Interact during
625   the Formation of Viral Progeny: Localization of the Binding Region within the PB1
626   Gene. Viruses 8.

627 46. Noda T, Sugita Y, Aoyama K, Hirase A, Kawakami E, Miyazawa A, Sagara H,
628   Kawaoka Y. 2012. Three-dimensional analysis of ribonucleoprotein complexes in
629   influenza A virus. Nature Communications 3:639.

630 47. Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine
631   Production for Classical and Swine Origin Human Influenza A Viruses | Journal of
632   Virology.

633 48. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. 4.
634   Nature Methods 9:357–359.

635 49. Routh A, Johnson JE. 2014. Discovery of functional genomic motifs in viruses with
636   ViReMa–a Virus Recombination Mapper–for analysis of next-generation
637   sequencing data. Nucleic Acids Res 42:e11.

638 50. Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display
639   consensus sequences. Nucleic Acids Res 18:6097–6100.

640 **FIGURE LEGENDS**
641 **Fig 1. DelVG formation is partially stochastic and biased towards the polymerase**
642 **and HA segments. (A)** Pie charts show the percentages of the total normalized junction
643 counts of all segments between the inoculum and viral populations from all time points at
644 the levels of unique junction site (upper panel) and NGS reads (lower panel). The black

645   fractions represent the junctions that were detected in the inoculum. **(B)** Correlation of
646   distinct junctions in the PB2 segment between two technical replicate samples (generated
647   from same RNA sample) using different NGS read cutoff values. **(C)** Venn diagram
648   showing the overlap in specific junctions (for the PB2 segment) between three replicate
649   samples collected at each time point. **(D)** Pie charts show the proportional abundance of
650   normalized junction counts from each genome segment across replicates.

651

652   **Fig 2. Roles of segment length and direct repeat sequences in DelVG deletion**
653   **formation. (A)** Observed normalized junction counts per segment were compared with
654   junction counts predicted by a model (expected) that assumes a simple positive
655   correlation between segment length and DelVG junction count (see materials and
656   methods). **(B)** The numbers of DelVG junctions detected at 6 hpi with no sequence
657   repetition flanking the deletion breakpoints (Direct repeat length = 0), or with repeated
658   sequences of varying length (Direct repeat length = 1-9) for the indicated segments
659   (observed). Expected values plot the numbers of junctions with the indicated repeat
660   lengths predicted from model simulations in which junction formation is random. Junction
661   counts are plotted as a percentage of the total number of DelVGs detected for a given
662   segment. Data are presented as means (n = 3 cell culture wells) ± standard deviations. *
663   P<0.05, ** P<0.01, ***, P < 0.001; ****, P < 0.0001, ns = not significant (Two-way
664   ANOVA).

665

666   **Fig 3. Enrichment of specific nucleotides at positions flanking DelVG deletions.** The
667   4 nucleotides flanking DelVG junctions were numbered and divided into four regions (R1-
668   R4) and the percentage occurrence of each nucleotide was calculated at each site within
669   each region. The junction nucleotides immediately flanking the deletion are indicated by
670   the red "J"'s at position 5 on the left and right. The grey nucleotides flank the junctions in
671   the progenitor sequence but are lost through deletion in the actual DelVG sequence. The
672   percentage of nucleotide occurrence (observed in PB2 6 hpi) at each site was plotted
673   against a random control (expected). Data are presented as means (n = 3 cell culture
674   wells) ± standard deviations. * P<0.05, ** P<0.01 ***, P < 0.001; ****, P < 0.0001 (Two-
675   way ANOVA), ns = not significant.

676

677   **Fig 4. DelVGs are inefficiently packaged compared with wild type vRNAs. (A)**
678   Numbers of DelVG junction counts (left panel) and NGS reads (right panel) were counted
679   for three replicates in each segment at each indicated time point, normalized to 10^6
680   mapped reads. Each dot represents a replicate. **(B)** The NGS read coverage per
681   nucleotide was plotted for each segment at each time point (the average coverage of the
682   three replicates per position). **(C)** Competition between WT and DI244 PB2 segments
683   during infection and packaging. MDCK-SIAT1 cells were co-infected with WT PR8 and
684   DI244 at a 1:1 ratio to a final MOI of 20 PB2 gene equivalents/cell under single cycle
685   conditions. Absolute copy numbers of WT and DI244 PB2 segments were quantified in
686   by RTqPCR within either cell lysates (intracellular), supernatant (extracellular), or
687   inoculum mixture at the indicated timepoints post-infection. PR8-only and DI244 only
688   controls are also shown. Data represent absolute values from individual cell culture wells
689   and are representative of two independent experiments.

690

691   **Fig 5. No differences in deletion size or breakpoint locations between intracellular**
692   **and packaged DelVGs. (A)** PB2 DelVG deletion junction sites at 6 and 24 hpi were
693   mapped to the genome position with each line representing a distinct deletion and the
694   colors indicating the time point (orange = 6 hpi, blue = 24 hpi). **(B)** Plots show the
695   cumulative occurrence of DelVG deletions as a function of gene segment position.
696   Cumulative score was calculated by starting at zero at the end of the segment and then
697   adding a score of 1 at each nucleotide where a unique junction breakpoint occurred.
698   Scores were normalized by calculating the percentage of final value reached at each
699   position. **(C)** DelVG size distributions (PB2 segment only shown) from samples collected
700   at the indicated time points were plotted using R density plot (ggplot package). The area
701   under the curve is equal to 100 percent of all probabilities.