Reovirus recombination is highly selective, and its profiles are primarily dictated by viral gene segment identity

Alejandra Flores[a], Andrew Routh[b], Ryan D. Xavier[a], Kristen M. Ogden[a,c,#]

[a] Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[b] ClickSeq Technologies LLC, Davis, California, USA

[c] Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Running head: Reovirus recombination patterns rely on viral segment identity

[#]Address correspondence to Kristen M. Ogden, kristen.ogden@vumc.org

ABSTRACT

Recombination facilitates the generation of defective viral genomes (DVGs), truncated derivatives of the parental genome that require a helper virus to replicate. Recombination mechanisms are poorly understood for viruses with double-stranded RNA genomes. Two strains of the double-stranded RNA virus reovirus differ in the pattern of packaged DVGs during serial passage. To determine whether the polymerase complex or gene segment sequence contributes to these differences, we exchanged polymerase complexes between the two reovirus strains. We identified DVG patterns using RT-PCR and recombination junction profiles using ClickSeq. Reoviruses synthesized DVGs that maintained the 5′ and 3′ termini and contained large central deletions. The polymerase complex did not detectably affect DVG pattern following reovirus serial passage or viral recombination junction profiles. Instead, recombination junction profiles correlated with the identity of viral RNA gene segments, even in the presence of a non-native polymerase complex or virus background. Reovirus recombination junction start and stop sites often occur in regions of sequence microhomology. While we observed many instances of short stretches of identical nucleotides within a viral gene segment, only a select few positions were incorporated into recombination junctions. Overall, these data suggest that recombination events that can mediate reovirus DVG formation are highly selective, and properties of viral gene segments primarily dictate where recombination occurs. These observations suggest a model for double-stranded RNA virus recombination in which the polymerase pauses RNA synthesis and re-initiates further along the same template at a specific junction stop site that has sequence homology to the junction start site.

39

40 ## IMPORTANCE

41 Viral infection gives rise to defective viral genomes, which cannot complete a full
42 replication cycle. Recombination facilitates defective viral genome generation but is
43 understudied for RNA viruses with double-stranded genomes. We found that for a
44 double-stranded RNA virus, reovirus, recombination occurs preferentially at specific
45 sites in the genome that correspond with the identity of the gene segment.
46 Recombination tends to initiate and terminate at sites sharing identical sequences.
47 However, even if these short nucleotide sequences appear multiple times in a gene
48 segment, only specific sites are used for recombination. Our results indicate that
49 reovirus recombination is a highly orchestrated event in which individual gene segments
50 contain the characteristics that drive recombination. These findings suggest that RNA
51 properties, such as sequence and structure, drive recombination for double-stranded
52 RNA viruses, likely through reinitiation after re-hybridization of the newly formed RNA
53 product with the same RNA template molecule at a different location.

54

55 ## INTRODUCTION

56 Most RNA viruses encode RNA polymerases that lack proofreading activities and can
57 form defective viral genomes (DVGs). DVGs contain mutations, deletions, or changes in
58 their genetic structure that prevent them from replicating without the assistance of a
59 helper virus (1). DVGs can be packaged inside virions to form defective interfering
60 particles (DIPs), which can interfere with viral replication. DIPs can reorient resources,
61 outcompete wild-type viral replication, or activate the immune system (1-3). Viruses with

62 relatively less DVG accumulation have been associated with more severe or fatal
63 disease outcomes in humans and mice for negative-sense RNA virus influenza A virus,
64 and a critical role for DVGs in modulating host interferon responses and symptom
65 development has been suggested for positive-sense RNA virus SARS-CoV-2 (4, 5).
66 Since they are replication-incompetent and can interfere with viral replication, DVGs and
67 DIPs have been considered as potential therapeutics or vaccines for multiple viruses (6,
68 7). Although they have been well studied for positive-sense and negative-sense RNA
69 viruses, relatively little is known about DVGs and DIPs of double-stranded RNA
70 (dsRNA) viruses.

71 Mammalian orthoreovirus (reovirus) is a dsRNA virus that can package DVGs
72 (8). Reovirus belongs to the order Reovirales and has garnered interest for its oncolytic
73 therapy potential (9-11). Reovirus has a segmented dsRNA genome containing three
74 large (L, 3.8-3.9 kb), three medium (M, 1.9-2.2 kb), and four small (S, 1.2-1.6 kb)
75 segments (12). Reovirus gene segments contain an open reading frame (ORF) flanked
76 by 5′ (12–32 nt) and 3′ (35-83 nt) untranslated regions (UTRs) (13). The ten reovirus
77 dsRNA gene segments are present in equimolar proportions in virions, and particles
78 tend to contain either all ten segments or none of them, indicating that packaging is
79 highly orchestrated (14, 15). Numerous studies suggest the reovirus gene segment
80 termini are important for packaging (16-22). Consistent with this observation,
81 sequenced reovirus DVGs contain one or two large internal deletions but retain the 5′
82 and 3′ termini (8). However, the RT-PCR profiling used to obtain these sequences
83 preferentially amplifies DVGs that conserve the termini and limits the detection of other
84 types of DVGs. Two prototype human reovirus strains, T1L and T3D, differ in apoptosis

85 induction capacity, viral factory morphology, and other characteristics (12). Serial
86 passage of T1L and T3D in distinct evolutionary lineages in murine L929 fibroblasts (L
87 cells) yielded viruses that packaged distinct subsets of DVGs, which we refer to as
88 'patterns' (8). While T1L had a consistent DVG pattern across evolutionary lineages,
89 T3D packaged more diverse and changing subsets of DVGs. These observations
90 suggest an inherent difference in the specificity with which T1L and T3D synthesize or
91 package DVGs. Much work is needed to quantify the differences between T1L and T3D
92 DVGs, elucidate the factors that contribute to these differences, and identify features of
93 DVGs that promote their packaging.

94 For RNA viruses, recombination has been proposed as a major driver of DVG
95 formation (2, 23, 24). For positive-sense RNA viruses and influenza viruses, deletion
96 DVGs are detected that contain large internal deletions but often retain promoter
97 regions and terminal sequences necessary for their replication and packaging (25).
98 Deletion DVGs are thought to form when the viral RdRp dissociates from the template
99 and reassociates either intra- or intermolecularly with the same or a homologous
100 template (2, 23, 24). Copy-back DVGs, which are thought to occur when the viral
101 polymerase dissociates from the template strand and reinitiates on the newly
102 synthesized strand, are detected for non-segmented negative-sense RNA viruses (1,
103 25, 26). Recombination mechanisms for viruses containing segmented dsRNA
104 genomes are largely undescribed. Inside reovirus particles, the RNA-dependent RNA
105 polymerase (RdRp), λ3, forms a complex with its cofactor, μ2, which likely guides
106 template RNA in and out of the RdRp, provides energy through its nucleoside
107 triphosphatase activity, and acts as a helicase (27-31). Unlike simpler clamp-like

108 RdRps, λ3 contains tunnels for RNA template and product entry and exit and cannot
109 readily release and switch RNA templates (32-34). Following reovirus entry into cells,
110 RNA transcription occurs inside the viral core, where the minus strands of genomic
111 dsRNA segments are utilized as templates to synthesize positive-sense single-stranded
112 RNA (+ssRNA) transcripts that are extruded into the cytoplasm. During particle
113 assembly, the ten +ssRNA segments are used as templates for the synthesis of
114 complementary minus-strand RNA to form genomic dsRNA (33). Like positive-sense
115 RNA viruses, reovirus can form deletion DVGs (8). For rotavirus, another member of the
116 order Reovirales, segment rearrangements have been associated with the RdRp; these
117 rearrangements frequently involve partial head-to-tail RNA duplications that conserve
118 the gene segment termini, and it is hypothesized that their synthesis involves RdRp
119 dissociation and reassociation on the same template (35). Illumina RNA-seq analysis of
120 reovirus strain T1L revealed recombination hot spots across the genome, and at least
121 some junction sites appeared to occur near short regions of identical sequence, or
122 microhomology (8). These observations suggested that reovirus recombination might
123 occur at select sites. However, the studies were conducted for only two samples of a
124 single reovirus strain. Furthermore, while unbiased, Illumina RNA-seq library
125 preparation can introduce artifactual recombination events that account for a noteworthy
126 portion of output sequence reads and influence computational analyses of DVGs (36).
127 Thus, additional studies are needed to validate selective recombination events for
128 dsRNA viruses like reovirus.
129 Since recombination events leading to DVG generation involve viral polymerase
130 activity, we hypothesized that the observed differences between T1L and T3D in DVG

131 patterns among evolutionary lineages (8) were due to differences in the viral
132 polymerase complexes. Thus, we generated reoviruses in which the polymerase
133 complex was exchanged between T1L and T3D. We exchanged segments L1 and M1,
134 which respectively encode polymerase complex proteins λ3 and μ2, between T1L and
135 T3D, rescued the recombinant reoviruses, and characterized them. These exchanges
136 provided the opportunity to distinguish between the contributions of the RNA gene
137 segments and the polymerase complex proteins they encoded to differences in DVG
138 patterns. We serially passaged the viruses to determine DVG patterns in evolutionary
139 lineages. We also sequenced packaged RNA in wild-type and polymerase complex-
140 exchanged T1L and T3D reoviruses using an approach that minimizes artifactual
141 recombination and analyzed the recombination junction sites. We found that TIL and
142 T3D differ in their recombination junction site preferences, but the phenotype
143 segregates genetically with the identity of the gene segment rather than the polymerase
144 complex. At recombination junction start and stop sites, we detected conservation of
145 positional identity but not of a specific sequence motif directing recombination. Although
146 sequences exhibiting microhomology were found to appear multiple times in a gene
147 segment, preferred sites were used for recombination. Together, these observations
148 suggest that RNA primary sequence is required for DVG synthesis and selection.
149

150 ## RESULTS

151 **Polymerase complex-exchanged reoviruses exhibit distinct replication**
152 **phenotypes.** To investigate how the polymerase complex influences packaged DVG
153 populations, we engineered T1L and T3D reoviruses with exchanged polymerase

154 complexes. We exchanged the L1 and M1 gene segments of reovirus strains T1L and
155 T3D using reverse genetics to generate T1L-T3pol and T3D-T1pol (**Fig. 1A**) (37, 38).
156 To confirm that both L1 and M1 segments were swapped between the strains, we
157 utilized electropherotyping to separate viral gene segments based on size and electrical
158 charge (**Fig. 1B**). For T1L-T3pol, gene segments L1 and M1 migrated more slowly in
159 the gel compared to the T1L gene segments, instead resembling those of T3D. In
160 contrast, for T3D-T1pol, gene segments L1 and M1 migrated faster and resembled
161 those of T1L. These results indicate that polymerase complex-exchanged reoviruses
162 were rescued using reverse genetics.

163 We analyzed the replication characteristics of polymerase-exchanged reoviruses
164 using two different cell lines. In previous studies, T1L and T3D replicated with similar
165 efficiency in L cells, but T1L replicated more efficiently in canine epithelial MDCK cells
166 compared to T3D; these differences in growth segregated with the viral L1 and M1
167 segments (39, 40). To determine the effects of polymerase complex exchange on viral
168 replication, we adsorbed L cells or MDCK cells with three clones each of T1L, T3D,
169 T1L-T3pol, or T3D-T1pol at an MOI of 0.1 PFU/cell, washed, and incubated. Then, we
170 quantified viral titers in total cellular lysate. In L cells, T1L-T3pol had similar replication
171 kinetics to T1L, and T3D-T1pol had similar replication kinetics to T3D (**Fig. 1C**).
172 However, spread of the four viruses differed in a plaque assay, in which viruses
173 undergo multiple rounds of replication. The plaque size of T1L-T3pol was smaller than
174 that of T1L, while T3D-T1pol had the largest plaque size among the four viruses,
175 suggesting that the polymerase complex can confer a multi-cycle replication
176 disadvantage or advantage in its non-native context (**Fig. 1D**). Consistent with previous

177 results, T1L and T3D-T1pol were capable of replicating in MDCK cells, while T3D and
178 T1L-T3pol largely failed to replicate (**Fig. 1E**) (39, 40). These results confirm the
179 identities of reoviruses with exchanged polymerase complexes and suggest that their
180 replication phenotypes can be influenced by the polymerase complex.
181
182 **Exchanging the polymerase complex fails to exchange lineage-specific DVG**
183 **segment patterns.** We previously found that DVG patterns differed for serially-
184 passaged T1L and T3D reoviruses, with T1L accumulating similar DVGs and T3D
185 accumulating more diverse DVGs in distinct evolutionary lineages (8). To determine
186 whether the polymerase complex contributes to lineage-specific differences in packaged
187 DVG patterns, we serially passaged T1L, T3D, and the polymerase complex-exchanged
188 reoviruses in L cells. We initiated the passage series by adsorbing L cells at an MOI of
189 0.3 PFU/cell in triplicate to compare independent evolutionary lineages. For passage 1
190 (P1), the inoculum was recovered using reverse genetics and expanded once in L cells
191 to minimize the presence of DVGs and generate sufficient virus for the infections. After
192 adsorption, we incubated the cells for 48 h prior to lysis. Then, we infected a new round
193 of cells with a fixed volume of the lysate to allow a high MOI for DVG formation. We
194 repeated this procedure for 10 serial passages (**Fig. S1A**). We also conducted mock
195 infections in triplicate, using PBS for the inoculum, to use as controls. We determined
196 viral titers for each passage and lineage. Each analyzed virus showed distinct titer
197 trends (**Fig. S1B-E**). In most cases, titers of the three independent lineages of a given
198 virus exhibited similar increases or decreases across the passages.

199 To detect DVGs generated during serial passage and their patterns, we used an
200 RT-PCR approach, in which RNA molecules containing the 5′ and 3′ termini of a specific
201 viral gene segment were amplified and visualized for individual serial passages and
202 lineages. The assay conditions permitted detection of small RNA molecules that
203 preserved the segment termini, which have previously been shown to include DVGs (8).
204 The selected segments vary in size and represent both exchanged and non-exchanged
205 segments. Using this approach, full-length reovirus S and M segments, which are < 2.5
206 kb in length, were consistently detected, although the longer L segments were less
207 consistently detected (**Figs. 2, S2-S4**). To determine whether our primers amplify host
208 RNA, we conducted RT-PCR reactions using RNA extracted from the three mock-
209 infected passages and for each set of primers. We found no evidence of off-target
210 amplification for any set of primers (**Fig. S5**). For segment L1, a DVG of ~ 300 bp was
211 identified for all passages and lineages (**Fig. 2**). For T1L and T1L-T3pol, the ~ 300-bp
212 DVG was predominant (**Fig. 2A-B**). For T3D-T1pol and T3D, additional L1 DVGs of
213 varying size were detected in specific passages and lineages (**Fig. 2C-D**). Thus, as
214 previously observed, for the L1 segment, the DVGs detected were relatively consistent
215 among T1L lineages and more variable among T3D lineages (8). However, L1 DVG
216 variability does not appear to have been exchanged with the polymerase complex, as
217 the T1L-T3pol lineages remained consistent, with a few additional DVGs detected in P9
218 and P10 of T1L-T3pol lin3, and the T3D-T1pol lineages were more variable (**Fig. 2**). For
219 segment L3, most DVGs ranged from ~ 100 bp to 800 bp in length, and some DVGs
220 became more prominent in later passages (**Fig.S2)**. For each passaged reovirus, one to
221 few L3 DVGs of similar size were detected across most passages for the three lineages

222 of a virus, while the others varied in size, although L3 DVG sizes were more consistent
223 across lineages for T1L. For segment M1, one or a small number of DVGs of the same
224 size were detected in multiple passages for all lineages of a given virus (**Fig. S3**). For
225 segment S4, a DVG of ~ 150 bp was consistently detected for each passage and
226 lineage of all four viruses (**Fig. S4**). For segments M1 and S4, products higher in size
227 than the full-length segment were occasionally detected, which might have arisen from
228 duplication events (**Figs. S3-S4**). These studies support the previous observation that
229 DVG patterns revealed by RT-PCR generally tend to be less variable among lineages
230 for T1L than T3D, at least for segments L1 and L3 (**Figs. 2 and S2**) (8). All the viruses
231 in the study formed DVGs that can be detected by RT-PCR. However, the DVG patterns
232 observed for T1L-T3pol and T3D-T1pol fail to conclusively support a role for the
233 polymerase complex in the variety of DVGs synthesized and packaged during serial
234 virus passage.
235
236 **DVGs from polymerase-exchanged viruses contain large internal deletions.** Our
237 previous study determined that reovirus DVGs from segments L1, M1, and S4 contain
238 relatively short intact termini and one or two large internal deletions (8). Resolved RT-
239 PCR products from the current study revealed that reoviruses package DVGs derived
240 from segments L1 and S4 that persist across the passage series (**Figs. 2, S4**). To
241 determine whether the persistent DVGs synthesized and packaged by different viruses
242 are identical, we extracted and sequenced RT-PCR-amplified products from segments
243 L1 and S4 using the Sanger method for polymerase-exchanged reoviruses. Analysis of
244 the sequences identified DVGs with internal deletions (**Fig. 3**). These DVGs are 276-

245 282 bp (L1) and 147 bp (S4) in total length. The L1 DVGs each contain an internal
246 deletion of 3579 bp in slightly different positions in terms of nucleotide number but at an
247 identical junction in terms of sequence. The L1 5′ UTR is completely conserved along
248 with 242 nt of the ORF and the last 16 nt of the 32-nt 3′ UTR, ~ 7% of the full-length
249 segment. The sequenced T3D-T1pol L1 DVG was identical to a T1L L1 DVG
250 sequenced in our previous study (8). The S4 DVGs from T1L-T3pol and T3D-T1pol
251 contained an identical internal 1050 bp deletion. The S4 5′ UTR is completely conserved
252 along with 101 nt of the ORF and the last 14 nt of the 66-nt 3′ UTR, ~ 12% of the full-
253 length segment. These observations confirm that reovirus DVGs can contain large
254 internal deletions and suggest that while both termini may be important for packaging,
255 only a portion of the 3′ UTR is required. Detecting DVGs that are identical across
256 passages, lineages, and virus backgrounds also suggests that DVGs may be selectively
257 synthesized or packaged.
258
259 **Exchanging the polymerase complex fails to exchange reovirus recombination**
260 **profiles.** Although informative, DVG profiling using RT-PCR is biased by primers that
261 bind reovirus gene segment termini and fails to provide sequence information. This
262 approach can also selectively amplify small products relative to longer ones,
263 misrepresenting their relative quantities. Thus, to elucidate sequence features of
264 packaged DVGs and contributions of the polymerase complex to DVG synthesis and/or
265 packaging, we deep-sequenced RNA encapsidated within T1L, T1L-T3pol, T3D-T1pol,
266 and T3D reovirus particles using ClickSeq (36). ClickSeq is an RNAseq approach that
267 uses bioconjugation as an alternative to fragmentation and enzymatic adaptor ligation

268 during library preparation. This process reduces artifactual recombination to ~2.5 events
269 per million reads, even without careful optimization of PCR conditions (36). Thus,
270 identified recombination junctions are likely to represent biologically relevant RNA
271 products. Although ClickSeq has many benefits, a caveat is that the 3'-azido nucleotide
272 (AzNTP):dNTP ratio used in library preparation yields cDNA fragments ~ 300-500 nt in
273 length (36). Thus, the capacity to detect the smaller products that are readily amplified
274 using RT-PCR, including small DVGs (**Figs. 2 and S2-S4**), is limited.
275 Following multiple rounds of replication in L cells, we independently purified
276 reovirus particles from three clones of each virus and treated them with benzonase to
277 remove non-encapsidated RNA. Then, we extracted encapsidated RNA, prepared
278 ClickSeq libraries from each RNA sample, and sequenced the products on an Illumina
279 NovaSeq platform. Sequence coverage depth was high for every virus preparation and
280 segment (**Fig. S6** and data not shown). Coverage depth varied positionally by ~10-100-
281 fold, with variation highly similar among replicates. Coverage biases may result from
282 differences in the efficiency of random priming, reverse transcription, and PCR
283 amplification and may be influenced by sequence-dependent factors, such as GC-
284 content or RNA structure. Overall coverage was within normal limits, and the profiles
285 generated were even. Greater than 99.9% of reads mapped to viral gene segments,
286 demonstrating that reovirus virions had been sufficiently purified (**Table 1**).
287 We used a viral recombination mapping algorithm (*ViReMa*) to identify recombination
288 junction sites in the sequencing data, which were defined as deletions or insertions
289 greater than five bp in length flanked both upstream and downstream by a minimum 25-
290 bp high-quality alignment (**Fig. 4A**) (41, 42). We detected an average viral genome-wide

291 recombination frequency (JFreq = reads with mapped junctions per 10,000 mapped
292 reads) of 1.02-2.15, which is substantially lower than reported for betacoronaviruses
293 (43) for example, and many noncanonical recombination junctions (leading to non-full-
294 length products) in individual viral gene segments were identified for each virus (**Table**
295 **1**). Although the full-length sequences of DVGs cannot be determined from these data
296 due to the short reads inherent to the Illumina sequencing platform, in at least one
297 replicate for each of the four viruses, we detected the L1 recombination start site shown
298 in **Fig. 3**, and for T3D, we detected the L1 recombination stop site. For T1L-T3pol, we
299 detected the S4 recombination start site that is shown in **Fig. 3** with a stop site at
300 nucleotide position 1146. Thus, some of the same recombination sites we found in
301 sequenced DVGs were identified using *ViReMa*. These recombination sites were not
302 identified in all samples, likely due to the reduced capacity of ClickSeq to capture very
303 small products (36), including the smaller DVGs efficiently amplified by RT-PCR (**Figs.**
304 **2 and S2-S4**).
305 To visualize the locations and frequencies of recombination junctions for
306 individual segments, we generated junction scatter plots (**Figs. 4 and S7-S10**). These
307 plots revealed noncanonical junctions within each viral gene segment that suggest the
308 presence of DVGs containing deletions or duplications (**Fig. 4B**), which have previously
309 been detected for dsRNA viruses (8, 44-47). For segment L1, the identified junction
310 profiles were visually different between the T1L and T3D strains (**Figs. 4C, S7**). We use
311 the term "profile" to refer to the distribution of recombination junction identities for a
312 given viral gene segment and their frequencies. The T1L-T3pol junction profiles for
313 segment L1 partially resembled those of T3D, with which it shares a T3D L1 segment,

314 while T3D-T1pol junction profiles for segment L1 partially resembled those of T1L, with
315 which it shares a T1L L1 segment. For segment L3, the recombination profiles between
316 T1L and T3D also visually differed (**Figs. 4D, S8**). However, the T1L-T3pol L3 junction
317 profiles visually resembled those of T1L, with which it shares the T1L L3 segment. The
318 T3D-T1pol junction profiles visually resembled those of T3D, which both contain the
319 T3D L3 segment. For segment M1, visual similarities among junction profiles were not
320 as evident compared to other segments (**Figs. 4E, S9**). For segment M3, the junction
321 profiles resembled those of L3, with those of T1L and T1L-T3pol or T3D and T3D-T1pol
322 more visually similar to one another (**Figs. 4F, S10**). Overall, viruses that share the
323 same gene segment presented visually similar recombination profiles for that given
324 segment, even in the presence of a different polymerase complex. Furthermore, the
325 junction profiles were retained when segments were encapsidated in a different virus
326 background, indicating that some property of the RNA segment dictates the
327 recombination junction profile.
328 Previous research in our lab suggested that T1L recombination might occur
329 preferentially at distinct sites, but only two samples were investigated for a single virus
330 strain (8). To determine whether the genome contains ‘recombination hotspots’, we
331 calculated the recombination frequency for segments L1, L3, M1, and M3 for three
332 clones of each of our four viruses. The frequency of the start or stop position of each
333 junction was evaluated by dividing the frequency of reads mapping to a recombination
334 junction by the total mapped reads. The total reads were calculated as the sum of the
335 depth. Results are shown as the mean of the three clones (**Figs. 5 and S11**) to facilitate
336 comparison between the viruses and segments, and they are shown individually (**Figs.**

337 **S12-S19**) to compare reproducibility among clones. Individual peaks representing
338 regions of high frequency were not identical in all clones, but distinguishing
339 recombination frequencies were typical for each virus. For segment L1, recombination
340 start and stop sites occurring with the highest frequencies for T1L tended to be located
341 at identical places in the gene segment as those of T3D-T1pol, whereas recombination
342 start and stop sites occurring with the highest frequencies for T3D tended to be located
343 at identical places in the gene segment as those of T1L-T3pol (**Figs. 5A, S11A, and**
344 **S12-S13**). In contrast, for segment L3, recombination start and stop sites occurring with
345 the highest frequencies for T1L tended to be located at identical places in the gene
346 segment as those of T1L-T3pol, while recombination start and stop sites occurring with
347 the highest frequencies for T3D tended to be located at identical places in the gene
348 segment as those of T3D-T1pol (**Figs. 5B, S11B, and S14-S15**). For segment M1 start
349 and stop site frequencies across the segment, T1L appeared most visually similar with
350 T3D-T1pol, whereas T3D resembled T1L-T3pol (**Figs. 5C, S11C, and S16-S17**).
351 Finally, for segment M3 start and stop sites, T1L and T1L-T3pol both tended to share
352 positions of high frequency recombination sites, while T3D and T3D-T1pol shared such
353 positions (**Figs. 5D, S11D, and S18-S19**). In general, like the recombination junction
354 patterns (**Figs. 4 and S7-S10**), the recombination frequency peak patterns of T1L-T3pol
355 and T3D-T1pol visually resembled those of the strain of origin for each given segment
356 investigated, regardless of whether the segment was associated with a different
357 polymerase complex or a new virus background. Though qualitative, these observations
358 suggest that reovirus recombination profiles segregate with the reovirus gene segment
359 identity independent of the source of the polymerase or other viral proteins.
360

361 **Principal component analysis reveals that recombination profiles segregate with**
362 **the reovirus gene segment independent of the polymerase complex.**
363 Recombination frequencies and junction scatter plot data visually suggest similar
364 profiles between reoviruses containing segments derived from the same virus
365 regardless of the origin of other viral proteins. To quantitatively assess similarities in
366 junction profiles among the four viruses, we conducted a principal component analysis
367 (PCA) for segments L1, L3, M1, and M3. For each virus and replicate, we analyzed the
368 collection of unique recombination junctions and their frequencies. The loadings table
369 was utilized to compare the similarity of viral recombination junction profiles for each
370 given segment. For segment L1, T1L clustered tightly with T3D-T1pol, T3D clustered
371 tightly with T1L-T3pol, and the proportion of variance for the two first principal
372 components was 92.93, indicating that nearly all variance in the dataset is explained by
373 the first two components (**Fig. 6A**). Similar clustering was observed for segment M1,
374 although the proportion of variance for the two first principal components was only 50.81
375 (**Fig. 6B**). On the other hand, for segments L3 and M3, T1L clustered tightly with T1L-
376 T3pol, T3D clustered tightly with T3D-T1pol, and the proportions of variance for the two
377 first principal components were 92.7 and 75.74, respectively (**Fig. 6C-D**). The PCA
378 analyses quantitatively support more qualitative observations from the junction scatter
379 plots and recombination frequency analyses and indicate that recombination profiles
380 correspond with reovirus gene segment identity rather than that of the polymerase
381 complex or other viral proteins.
382
383 **Recombination sites are conserved based on segment identity.** Although PCA
384 revealed that recombination junction profiles are similar for viruses that share a gene

385 segment, a small number of overrepresented junctions may drive these similarities.
386 Therefore, we determined the number of unique recombination junctions shared
387 between reoviruses T1L, T1L-T3pol, T3D-T1pol, and T3D. We compared recombination
388 junctions among all viruses for segments L1, L3, M1, and M3 (**Fig. 7**). For segments L1
389 and M1, T1L and T3D-T1pol or T3D and T1L-T3pol, in which these gene segments are
390 identical, share many unique recombination junction sites. In contrast, T1L and T1L-
391 T3pol or T3D and T3D-T1pol, in which these gene segments are of different origins,
392 share few recombination junctions. For segments L3 and M3, the same is true of the
393 reverse pairings: T1L and T1L-T3pol or T3D and T3D-T1pol share many junctions,
394 while T3D and T1L-T3pol or T1L and T3D-T1pol share few junctions. Some common
395 recombination junctions are anticipated among all strains for segments that share high
396 sequence identity. However, like the results for recombination junction frequency and
397 location, these observations suggest that recombination events resulting in DVG
398 synthesis or packaging selection are predominantly determined by the identity of a
399 reovirus gene segment rather than that of the polymerase complex or other viral
400 proteins. Importantly, many recombination junctions, not just a small number of
401 overrepresented junctions, are shared among reoviruses with similar recombination
402 patterns.
403
404 **Recombination events match nucleotides positionally between junction start and**
405 **stop sites, but no specific motif is detected.** For multiple virus families, the start and
406 stop sites of recombination junctions have been associated with a specific nucleotide
407 composition or with regions of microhomology (41, 48-51). For reovirus, regions of

408 sequence microhomology, typically identical sequences about four to five nucleotides in
409 length, appear to surround recombination junction sites (8). However, only two samples
410 were investigated for the T1L strain using a sequencing method that can introduce
411 artifactual recombination events. To validate the prior finding and determine whether
412 regions of microhomology are also associated with recombination junction sites for T3D
413 and polymerase-exchanged reoviruses, we analyzed all noncanonical junctions in
414 segments L1 and L3. In each case, 50 nucleotides surrounding junction start sites and
415 stop sites were analyzed. Every nucleotide surrounding the junction start site was
416 matched to the equivalent position relative to the junction stop site. Positionally identical
417 nucleotides were counted and graphed. For all analyzed segments and viruses, the
418 nucleotides surrounding the junction start were frequently identical to the nucleotides
419 surrounding the junction stop site (**Figs. 4A and 8A).** Most frequently, identical
420 nucleotides were identified in about six positions surrounding recombination junction
421 sites. For segment L1, substantially higher numbers of recombination junctions were
422 detected in the negative strand than the positive strand, with high identity at the same
423 relative positions (**Fig. S20A**). For segment L3, modestly more recombination junctions
424 were detected for the positive strand than the negative strand (**Fig. S21A**). The
425 percentages of junctions containing four, six, or eight identical nucleotides at junction
426 sites were calculated for segments L1 and L3. For both segments, ~ 10% of junctions
427 for each virus contained eight identical nucleotides at the junction site, ~ 40% contained
428 six identical nucleotides, and ~ 70-75% contained four identical nucleotides at the
429 junction site (**Table. 2**). These findings suggest that recombination for all tested
430 reoviruses, regardless of gene segment or polymerase complex identity, preferentially

431 occurs at sites of microhomology. Thus, reovirus recombination is likely sequence-
432 directed.
433 As nucleotides tend to match between recombination junction start and stop
434 sites, it is possible that specific sequence motifs direct recombination. Therefore,
435 nucleotide frequency relative to recombination junction start and stop sites was
436 calculated for noncanonical junctions in segments L1 and L3. For all generated
437 logoplots, no specific nucleotides were observed at a substantially higher frequency
438 relative to the recombination start or stop site for any virus on either the positive or
439 negative strand for either analyzed segment (**Figs. 8B and S20B-S21B**). Thus, there
440 was no indication of a specific sequence motif or absolute nucleotide identity driving
441 recombination. Instead, only positional identity relative to the junction start and stop site
442 for a given recombination event is conserved.
443
444 **Recombination site selection is specific.** After observing that recombination
445 frequently occurs at start and stop sites containing a stretch of four identical nucleotides
446 (**Fig. 8A and Table 2**), we wondered how frequently these stretches of identical
447 sequence could be found in a given reovirus gene segment. We identified all the unique
448 four-nucleotide sequences with microhomology at start and stop sites from
449 recombination events in the L1 and L3 genes. Then, we identified all the occurrences of
450 these sequences in L1 and L3. We found that in both segments, such four-nucleotide
451 sequences typically appeared between 10 to 20 times (**Fig. 9A**). Some four-nucleotide
452 sequences appeared much more frequently; ATGG appeared 41 times in the shared L1
453 genes of T1L and T3D-T1pol and 44 times in L1 of T3D and T1L-T3pol. Others had

454 fewer occurrences, such as CCCC, which appeared six times in the L1 genes of T1L
455 and T3D-T1pol. We next wondered how many occurrences of each of these four-
456 nucleotide sequences were part of a recombination junction in our data set, either at the
457 start or stop of the junction. Most occurrences of the four-nucleotide sequences (79-
458 87%) were not part of any junctions in our data set (**Fig. 9B**). We found that for
459 segments L1 and L3, ~ 15% and ~ 13% of occurrences of a four-nucleotide sequence
460 were used in a junction once, indicating their incorporation into a single unique junction.
461 Approximately 4% of occurrences of a four-nucleotide sequence for L1 and 2% of
462 occurrences for L3 were selected for recombination events more than once, indicating
463 their incorporation into more than one unique junction. The L1 sequences most
464 frequently used for recombination were GCGG and GGGA in T1L, which were
465 associated with six unique junction start sites. For the L1 stop position, the TGGG
466 sequence was part of nine unique junctions. In segment L3, TAAT in T3D-T1pol was
467 part of seven unique junction start sites, while TAAG was part of eleven unique junction
468 stop sites. Overall, while a single four-nucleotide sequence can be frequently detected
469 in a given reovirus gene segment, typically, only rare and specific occurrences of this
470 sequence are utilized during recombination. These observations suggest that reovirus
471 recombination is highly selective.
472

473 **DISCUSSION**

474 We sought to distinguish between contributions of RNA gene segments and the
475 polymerase complex to virus strain-specific differences in DVG patterns. Although the
476 presence of DVGs had been reported for reovirus (8, 44), the recombination

477 mechanisms leading to DVG formation and their strain specificity remained unclear. We
478 exchanged the polymerase complexes between reovirus strains T1L and T3D, which
479 were previously shown to differ in their DVG patterns (**Fig. 1**) (8). RT-PCR profiling of
480 DVGs for wild-type and polymerase-exchanged reoviruses failed to indicate a strict
481 correlation between the polymerase complex and the diversity of DVG patterns among
482 evolutionary lineages of serially passaged viruses (**Figs. 2 and S2-S4**). At least some
483 DVGs carried in the virus population throughout the passage series shared identical
484 sequences or internal deletions, suggesting a non-random recombination or packaging
485 selection process (**Figs. 2-3, and S4**). Based on ClickSeq sequencing and
486 computational analyses of packaged viral RNA, we found that T1L and T3D reoviruses
487 have distinct recombination sites, which are conserved based on the identity of the gene
488 segment rather than that of the polymerase complex (**Figs. 4-7 and S7-S19**).
489 Regardless of segment or polymerase complex identity, we identified positional identity
490 at recombination junction start and stop sites, where regions of microhomology were
491 detected (**Figs. 8 and S20-S21**, **Table 2**). Although the conserved sequence at a
492 recombination junction start and stop site were identified in many locations across a
493 given gene segment, we found that recombination preferentially occurs only at a small
494 subset of such locations (**Fig. 9**). Together these findings indicate that, like reovirus
495 packaging, the recombination events that can yield DVGs may be highly orchestrated
496 and primarily dictated by properties of viral RNA gene segments, which may include
497 sequence and RNA secondary structure.
498 DVGs that arise during serial passage may affect the virus population and
499 provide information about viral RNA synthesis and/or packaging requirements. In our

500 analyzed reovirus series, some L1 and S4 DVGs persisted across all passages (**Figs. 2**
501 **and S4**). Other DVGs were transient, detected solely in a few passages, or prominent
502 only in later passages. Variations in DVG diversity throughout passage series have
503 been observed for other viruses such as SARS-CoV-2, herpesvirus, and Flock House
504 virus (52-55). The persistence of a DVG in a virus population suggests that it is
505 preferentially replicated and/or packaged (56). In most cases, there was little evidence
506 of a substantial (>100-fold) decrease in viral titer across the series (**Fig. S1**), which
507 suggests that most persistent reovirus DVGs detected are unlikely to be DIPs and may
508 have neutral effects on the virus population. However, for segment L3, T3D-T1pol
509 DVGs became more prominent as detection of the full-length segment decreased,
510 which correlated with larger decreases in viral titers (**Figs. S1-S2**). DIPs have been
511 identified for avian reovirus, for which S1 DVGs correlate with a reduction of full-length
512 S1 and a decrease in infectious titers, suggesting a competitive relationship during
513 replication and/or packaging (44). For the persistent DVGs of reovirus segments L1 and
514 S4, the analyzed sequences revealed that these DVGs were identical between viruses
515 or contained identical deletions (**Fig. 3**), which suggests that DVGs with specific
516 characteristics are synthesized and/or packaged. Reovirus sequences located at the
517 segment termini, which encompass the UTR and extend into the adjacent ORF, are
518 thought to aid in assortment, packaging, and other replication processes (17-22, 57,
519 58). The DVGs that were sequenced retained the signals located at the 5′ end, but only
520 a partial 3′ UTR. In a prior study, two of six sequenced DVGs from various segments
521 also contained partial 3′ UTRs, although another two contained longer stretches of

522 sequence at the 3′ end than the 5′ end (8). Further analysis of packaged DVGs could
523 reveal the minimal components and features required for successful packaging.
524 In this study, a broad spectrum of recombination junctions was observed for T1L,
525 T1L-T3pol, T3D-T1pol, and T3D reoviruses (**Figs. 4, S7-S10**). Based on RT-PCR
526 profiling, DVG sequencing, and prior results, we expected to detect many small DVGs
527 with large internal deletions (**Figs. 2, S2-S4**) (8, 44). Relatively few recombination
528 events indicative of such DVGs were detected (**Figs. 4, S7-S10**). However, ClickSeq
529 inefficiently captures products of ~ 100-300 bp, limiting our capability to detect deletion
530 DVGs that are missing considerable portions of a viral gene segment and to make
531 correlations with results obtained from RT-PCR analyses of serially passaged viruses
532 (36). Overall, our results identified junctions that indicate the existence of various
533 packaged DVGs, including small and large internal deletions and duplications.
534 Recombination events in the upper left corner of junction pattern plots, which represent
535 the merger of a junction start site near the 3′ end of a segment with a stop site near the
536 5′ end of the same segment, are frequently discarded as artifacts. However,
537 recombination events that generate duplications in viral gene segments have been
538 identified for several dsRNA viruses [44-46, 60-64]. Based on ClickSeq data analyses
539 alone, we cannot determine whether all RNA gene segments packaged by reovirus that
540 contain recombination junctions also retain the segment termini, and we cannot
541 completely define the packaged reovirus DVG catalog.
542 Several factors may influence the location of recombination sites in packaged
543 reovirus RNA gene segments. In the analyses of ClickSeq data, distinct junctions and
544 recombination sites were identified for T1L and T3D (**Figs. 4-5 and S7-S10**). These

545 strains have high sequence identities in segments L1 (95.9%), L3 (97.7%), M1 (97.9%),
546 and M3 (88.0%). Nonetheless, strain-specific recombination preferences were
547 maintained in the presence of the polymerase complex of a different virus strain or upon
548 encapsidation of the viral segment inside a different virus background. Recombination
549 events leading to DVG generation have been observed to be enriched at specific
550 regions of the viral genome for other RNA viruses, indicating preferred recombination
551 sites (5, 59-61). Consistent with our results, rearrangements identified from rotavirus
552 infections in independent immunocompetent children repeatedly occurred at a preferred
553 location or hotspot (45-47). We also provide evidence that reovirus recombination
554 occurs preferentially at specific sites of microhomology, short stretches of conserved
555 sequence that are positionally identical between the recombination start and stop sites
556 but it is not directed by a specific sequence motif (**Figs. 8 and S20-S21**). For
557 rotaviruses, short regions of microhomology have been identified around the junctions
558 of gene segment rearrangements (45-47). For SARS-CoV-2 and MERS-CoV, an
559 enrichment of specific nucleotides upstream and downstream of the junction has been
560 detected (41). Influenza A virus and Sindbis virus have been associated with specific
561 sequence motifs in the deletion junctions that give rise to DVGs (48, 49). Thus,
562 unrelated viruses also exhibit recombination specificity, although the overall
563 mechanisms may differ. Additional factors that could influence recombination locations
564 are RNA secondary structures or other RNA-RNA interactions. In SARS-CoV-2, it was
565 found that secondary structures within the viral genome might guide recombination, as
566 RNA structural distance between DVG junction positions was shorter than any two
567 random genomic positions, which could facilitate the breaking and rejoining of those

568 close pairs (5). It has also been shown that RNA secondary structure can influence
569 chikungunya virus DVG production during replication (62). It has been predicted that the
570 segment termini of +ssRNA segments of dsRNA viruses, such as reovirus and rotavirus,
571 interact to form panhandle structures and contain stem-loops (58, 63-67). These
572 predicted folds and structures are thought to be important for segment selection and
573 packaging, but they may also be important for directing recombination. There is also
574 evidence for intersegment RNA-RNA interactions between +ssRNAs during the
575 assembly of dsRNA viruses including rotavirus and bluetongue virus (68-72). RNA
576 secondary structure or other RNA-RNA interactions could influence the selection of the
577 specific microhomology-containing regions that are used for recombination (**Fig. 9**).
578 Determining the contributions of sequence and structure to DVG production will require
579 the development of experimentally informed RNA secondary-structure and interaction
580 models for +ssRNA reovirus segments to interrogate or model their influence on
581 recombination events that might lead to DVG generation. Finally, packaging selection
582 might influence our observations of recombination specificity. It is likely that a more
583 diverse set of DVGs and recombination junctions is formed in infected cells than is
584 selectively packaged into assembling viral particles.

585 Based on our new findings, we propose a model for the mechanism of RNA
586 recombination in reoviruses (**Fig. 10**). The catalytic domain of reovirus RdRp λ3, like the
587 RdRps of other dsRNA viruses, is sandwiched by N- and C-terminal domains that
588 create a cage-like structure containing four tunnels for RNA template and product entry
589 and exit (33, 34). This structure could not readily facilitate dissociation and reassociation
590 of the polymerase on a distinct template. Therefore, a model of recombination that

591 involves pausing and reinitiation at a different point along the same template is more
592 likely. This model fits well with the detection of deletion DVGs, rearrangements, and
593 recombination events anticipated to yield these types of noncanonical viral RNAs for
594 dsRNA viruses (**Figs. 3, 4, and S7-S10**) (35, 44-47, 73, 74). The observation that ~
595 70% of the detected recombination junction sites occur at regions of microhomology,
596 where positional identity was identified (**Table 2**), suggests that re-hybridization with the
597 template might trigger the reinitiation of RNA synthesis. As dsRNA viruses undergo two
598 RNA synthesis events, it is unknown whether the recombination events that lead to
599 synthesis of reovirus DVGs and other noncanonical RNA molecules occur during
600 transcription, in which +ssRNA is generated, or during genome replication, which
601 produces the minus strand. For a segmented dsRNA ϕ6 bacteriophage, it has been
602 shown that recombination can occur during minus-strand synthesis (75). It is highly
603 unlikely that the dsRNA templates of transcription form substantial secondary structures
604 or RNA-RNA interactions that might induce an RdRp pause. On the other hand, the
605 +ssRNA templates of minus-strand synthesis (genome synthesis) are predicted to form
606 complex secondary structures and RNA-RNA interactions that could influence RdRp
607 activity or pausing, allowing the rehybridization of the template and reinitiation of RNA
608 synthesis that could yield recombination sites in the viral genome (58, 63-67, 70, 76,
609 77).

610 Observations in this study suggest that reovirus recombination occurs at
611 preferred sites across the genome, which differ by reovirus strain. Rather than the
612 polymerase complex, the identity of the gene segment seems to be a primary driver in
613 selecting sites of recombination or the populations of DVGs that are packaged into

614 reovirus particles. Together with properties of the reovirus RdRp and the types of
615 reovirus DVGs detected, microhomology at recombination sites suggests a model of
616 recombination in which the RdRp pauses synthesis and reinitiates after re-hybridization
617 of the nascent product with the same template molecule at a different location (**Fig. 10**).
618 Among the dsRNA viruses of the order Reovirales, many features are shared, including
619 replication strategy, RdRp structure, packaging determinants, and noncanonical RNA
620 properties. Thus, acquired knowledge for reovirus also may apply to the recombination
621 mechanisms of other dsRNA viruses that affect human health (rotavirus), livestock
622 (bluetongue), and agriculture (rice dwarf virus). Future studies should be aimed at
623 determining associations of DVGs with pathogenesis, the influence of DVGs that are
624 carried transiently or continuously in virus populations, discriminating between DVG
625 synthesis and packaging in selecting the DVG population encapsidated by virus
626 particles, defining the contributions of RNA sequence and structure to recombination,
627 and determining precise recombination mechanisms for dsRNA viruses.
628

629 ## MATERIALS AND METHODS

630 **Cells.** L cells were maintained in suspension culture in Joklik's minimum essential
631 medium (JMEM; U.S. Biological) supplemented with 5% fetal bovine serum (FBS;
632 Gibco), 2mM L-glutamine (Corning), and 100 U/ml penicillin plus 100 µg/ml
633 streptomycin (Corning). Baby hamster kidney cells expressing T7 RNA polymerase
634 controlled by a cytomegalovirus promoter (BHK-T7) cells were grown in Dulbecco's
635 minimum essential medium (DMEM; Corning) supplemented with 10% tryptose-
636 phosphate broth, 1% nonessential amino acids, 5% FBS, 2 mM L-glutamine, and 100

637 U/mL penicillin plus 100 µg/mL streptomycin, with 1 mg/ml Geneticin (Gibco) added
638 during alternate passages. MDCK cells were maintained in minimum essential medium
639 with Earle's salts (EMEM; Corning) supplemented with 10% FBS, 2 mM L-glutamine,
640 and 100 U/ml penicillin plus 100 µg/ml streptomycin. All cells were cultivated at 37°C
641 with 5% CO_2.
642
643 **Viruses.** Reovirus strains T1L, T3D, T1L-L1M1-T3D, and T3D-L1M1-T1L were
644 engineered using plasmid-based reverse genetics (37, 38). T3D contained a T249I
645 mutation that protects σ1, the attachment protein, from proteolytic cleavage (37). A
646 pBacT7-S1T3D plasmid was previously engineered via 'round the horn PCR to include
647 the T249I mutation (78). T1L and T3D were rescued with a plasmid expressing C3P3 to
648 increase efficiency (79). Briefly, monolayers of BHK-T7 cells at approximately 50%
649 confluency in 12-well plates were co-transfected with 0.8 µg of pT7 plasmids encoding
650 each of the ten T1L or T3D reovirus positive-sense viral RNAs and TransIT-LT1
651 transfection reagent (Mirus Bio LLC) (37, 38). For the polymerase complex-exchanged
652 viruses, plasmids encoding T1L or T3D L1 and M1 positive-sense viral RNAs were
653 introduced into the T3D or T1L reovirus background, respectively, in place of the native
654 plasmids. Following five days of incubation, cells were subjected to two rounds of
655 freezing and thawing. For viruses used in the passage series, lysates were amplified
656 once in L cells, and viral titers were calculated by plaque assay in L cells (80). Viruses
657 used for recombination analysis were plaque purified (80). Individual plaques were
658 selected and amplified twice in L cells to generate stocks.
659

660 **Electropherotyping of recombinant reoviruses.** 4.8 × 10^6 L cells were inoculated
661 with three independently plaque-purified virus clones of each recombinant reovirus.
662 Cells were incubated at 37°C for about five days until cytopathic effects were detected.
663 Cells were collected and lysed in IGEPAL buffer (140 mM NaCl, 1.5 mM $MgCl_2$, 100 mM
664 Tris, pH 7.4) plus 5% IGEPAL CA-630 (Alfa Aesar). RNA from the supernatant was
665 precipitated and resuspended in nuclease-free water. RNA samples were mixed with 4×
666 sample buffer and heated for 10 min at 65°C. Samples were resolved by SDS-10%
667 PAGE and ethidium bromide staining and visualized using a ChemiDoc MP imaging
668 system (BIO-RAD).
669
670 **Replication kinetics analysis.** To compare the replication kinetics of wild-type and
671 polymerase-exchanged reoviruses, 8 × 10^5 L or MDCK cells in 12-well plates were
672 absorbed with three clones of each recombinant reovirus at an MOI of 0.1 PFU/cell. The
673 unbound virus was removed by washing, and cells were incubated at 37°C for 0, 6, 12,
674 18, 24, or 48 h. After incubation, cells were lysed, and viral titer in lysates was quantified
675 by plaque assay (80).
676
677 **Reovirus serial passages in L cells.** To initiate the passage series, 5 × 10^7 L cells
678 were pelleted and adsorbed with T1L, T1L-T3pol, T3D-T1pol, T3D, or PBS (mock) in
679 10 ml JMEM at an MOI of 0.3 PFU/cell for 1 h at 37°C in triplicate from the same stock.
680 Cell-virus mixtures were transferred into a suspension culture using 1 L bottles
681 containing a magnetic stirring rod, and JMEM was added to a total volume of 100 ml.
682 Suspension cultures were incubated for 48 h at 37°C, and the incubation was

683 terminated by completing two freeze-and-thaw cycles and clearing debris from lysate by
684 centrifugation at 1,000 × *g* for 10 min. For subsequent passages (P2 to P10), an
685 identical protocol was followed, except that 5×10^7 L cells were pelleted and
686 resuspended in an adsorption inoculum of 10 ml of cleared lysate from the previous
687 passage. For each virus, triplicate lineages were maintained throughout the series.
688 Virus titer in each passage for each lineage was determined by plaque assay (80). One
689 mock-infected passage was completed in triplicate.
690
691 **RT-PCR and electrophoretic analysis of DVG profiles.** RNA was extracted from
692 serial passage lysates using TRIzol LS Reagent (Invitrogen) according to manufacturer
693 instructions. The RNA was used as a template with the OneStep RT-PCR kit (Qiagen)
694 using manufacturer-recommended reaction conditions and gene-specific primers
695 targeting the 5′ and 3′ termini to identify gene segments and DVGs. Primer sequences
696 are available upon request. The extension time for segments >2 kb was 3 min and for
697 those <2 kb was 1 min. RT-PCR products were resolved in 1.2% agarose stained with
698 ethidium bromide and visualized using a ChemiDoc MP imaging system (BIO-RAD).
699 RT-PCR products were resolved alongside a 100 bp DNA ladder and 1 kb DNA ladder
700 (New England Biolabs).
701
702 **DVG sequencing.** RT-PCR products from P6 and P9 present in multiple passages and
703 lineages were gel extracted using a QIAquick gel extraction kit (Qiagen). The extracted
704 DNA was submitted for Sanger sequencing (Azenta) with the primers used for RT-PCR

705 amplification. Sequences were aligned to the corresponding reference viral genome
706 using CLC sequence viewer 8 (Qiagen).
707
708 **Reovirus particle enrichment and RNA collection.** 4×10^8 L cells were absorbed with
709 three independent clones of stock for viruses T1L, T1L-T3pol, T3D-T1pol, and T3D at
710 an MOI of 1 PFU/cell at 37°C for 72 h. Virus particles were purified from infected L cells
711 by deoxycholate permeabilization, Vertrel XF (DuPont) extraction, and CsCl gradient
712 centrifugation, as described previously(14). Reovirus particle concentration was
713 determined from the equivalence of one unit of optical density at 260 nm to 2.1×10^{12}
714 particles. 2.2×10^{12} viral particles for each individual virus and each replicate were
715 treated with benzonase (250U) to degrade extra-particle nucleic acids, and the enzyme
716 was inactivated with 0.5 M EDTA (pH 8.0). RNA was extracted from benzonase-treated
717 particles using Trizol LS Reagent (Invitrogen), according to the manufacturer's protocol.
718 The quality of the RNA was determined using a Bioanalyzer (Agilent), and the RNA
719 concentration was determined by Qubit (Invitrogen).
720
721 **ClickSeq library preparation and Next-Generation Sequencing (NGS).** ClickSeq
722 libraries were generated from 250 ng of extracted RNA for each virus replicate. To
723 initiate library preparation, the RNA underwent reverse transcription using SuperScript
724 III reverse transcriptase (Invitrogen) with the addition of 3′-azido-nucleotides at a 1:35
725 AzNTPs:dNTPs ratio to induce chain termination and to release 3′-azido blocked cDNA
726 fragments, which are click-ligated to Illumina i5 sequencing adaptors, as previously
727 described (36). The products were PCR amplified with barcoded Illumina p7 adaptors,

728 and libraries were sequenced by 150-bp paired-end sequencing on a NovaSeq X Plus
729 sequencing system (Illumina). Assistance with quality control and next-generation
730 sequencing was provided by ClickSeq Technologies LLC.
731
732 **Illumina RNA-sequencing processing and alignment.** Raw reads were processed by
733 first removing the Illumina TruSeq adapter using Cutadapt (81) with default settings.
734 Reads shorter than 50 bp were discarded and reads containing any base with a PHRED
735 score < 20 were removed using the FASTX toolkit. The raw FASTQ files were aligned to
736 the T1L and T3D reovirus genome segments (NCBI or GenBank accession numbers
737 AF378003.1, AF129820.1, AF461682.1, AF490617.1, AF174382.1, M14779.1,
738 L19774.1, M14325.1, M13139.1, EF494435.1, EF494436.1, EF494437.1, EF494438.1,
739 EF494439.1, EF494440.1, EF494441.1, EF494442.1, EF494443.1, EF494444.1; the
740 T1L L1 sequence does not perfectly match those in GenBank and is available upon
741 request) using the Python3 script *ViReMa* (Viral Recombination Mapper, version 0.32)
742 (42) using the command line parameters python3 ViReMa.py reference_index
743 input.fastq output.sam –Ouput_Dir sample_virema/--Output_Tag sample_virema -BED
744 –MicroIndel_Length 5. The sequence alignment map (SAM) file was processed using
745 the samtools (82) suite to calculate nucleotide depth at each position in a sorted binary
746 alignment map (BAM) file (using command line parameters samtools depth -a -m 0
747 sample_virema.sorted.bam > sample_virema.coverage).
748
749 **Recombination junction analysis.** Recombination junctions that contained micro
750 insertions or deletions of 5 nt or less were filtered into separate output BED files. The

751 frequency of the start and stop position of each junction was evaluated by dividing the
752 frequency of reads mapping to a recombination junction by the total mapped reads.
753 Junctions were plotted according to the genomic position and colored according to $\log_{10}$
754 of the frequency using *ggplot2* in RStudio (83). Recombination frequency was
755 calculated at each genomic position by dividing the number of nucleotides in any
756 junction mapping to the position by the total number of nucleotides sequenced at the
757 position.
758
759 **Principal component analysis.** A unique ID was created for each recombination event
760 using the *dplyr* package in RStudio (84). All unique recombination events were identified
761 for each virus, and the frequency value was added to a matrix. NA values were
762 substituted with zero, and the data was normalized. Principal component analysis (PCA)
763 was conducted and visualized for segments L1, L3, M1, and M3 using GraphPad Prism
764 10.3.1.
765
766 **Nucleotide composition analysis and sequence microhomology quantification.**
767 Nucleotide composition at each position surrounding recombination junction start and
768 stop sites was determined. To avoid bias of highly replicated sites and to more closely
769 reflect the stochastic nature of RNA recombination, each unique detected junction was
770 counted equally rather than weighting by read count (54). Sequences were extracted
771 from a sorted BED file that was converted into a text file to facilitate its manipulation.
772 Start site and stop site sequences were separated in RStudio and read sequences
773 containing less than 51 nucleotides in the stop site were filtered out. The nucleotide

774 frequency at each position was calculated using the *Biostrings* package in RStudio (85).
775 The nucleotide frequency was plotted as a logo plot using the *ggseqlogo* package in
776 RStudio (86). For microhomology quantification, the nucleotides surrounding the
777 junction sites were compared for positional identity relative to the start and stop sites.
778 Positionally identical sequences were quantified in RStudio and plotted in GraphPad
779 Prism 10.
780
781 **Microhomology sequence occurrence and usage.** To calculate the number and
782 percentage of recombination events containing microhomology regions of specific
783 lengths, even numbers of nucleotides surrounding the junction sites (eight, six, or four)
784 were extracted using *dplyr* (84). Extracted eight-, six-, or four-nucleotide sequences
785 were compared between junction start and stop sites. Matches between the start and
786 stop site were counted, and the percentage of junctions with microhomology were
787 calculated using the *Biostrings* package in RStudio. To determine the frequency with
788 which conserved four-nucleotide sequences found at recombination junctions recur in a
789 given segment, unique four-nucleotide sequences with microhomology at start and stop
790 sites from recombination events in the L1 and L3 genes were extracted using *dplyr* (84).
791 For sequences originating from a negative strand, the reverse complement was
792 obtained using *Biostrings (85).* Using the T1L or T3D L1 or L3 segment sequence as a
793 reference, occurrences of each four-nucleotide sequence in the reference were
794 quantified using *Biostrings (85)*. The location of each short sequence in the viral
795 genome was extracted using *Biostrings* (85). The location of short sequences from the
796 negative strand and positive strand were adjusted to match the location of the start and

797 stop sites of the recombination junctions detected by *ViReMa*. Then, the short sequence
798 location number was counted for the number of occurrences in the column associated
799 with the start and the stop sites separately from the dataset.
800
801 **Shared junction analysis.** Individual recombination junctions were identified, and
802 unique IDs were created for each recombination event using the *dplyr* (84) package in
803 RStudio. Recombination junctions for all viruses within individual segments were
804 compared to each other using the *ggvenn* package in RStudio (87), and numbers were
805 plotted.
806

807 ## ACKNOWLEDGEMENTS

808 We thank the Vanderbilt High Throughput Screening Facility for training and access to
809 equipment. We thank Dr. James C. Slaughter for biostatistics expertise. Research
810 reported in this work was supported by the National Institutes of Health (R01AI155646
811 to KMO). Alejandra Flores was supported by T32GM139800 and F31AI186483. The
812 contents of this publication are solely the responsibility of the authors and do not
813 necessarily represent the views of the National Institutes of Health.
814

815 ## DATA AVAILABILITY STATEMENT

816 Data generated by ClickSeq sequencing can be accessed at NCBI Sequence Read
817 Archive (SRA) under the BioProject accession PRJNA1259514. Code utilized in this
818 report can be accessed at (https://github.com/ogdenlab1/ReoPolSwap.git).
819

820 **REFERENCES**

821 1. Vignuzzi M, Lopez CB. 2019. Defective viral genomes are key drivers of the
822 virus-host interaction. Nat Microbiol 4:1075-1087.
823 2. Marriott AC, Dimmock NJ. 2010. Defective interfering viruses and their potential
824 as antiviral agents. Rev Med Virol 20:51-62.
825 3. Rezelj VV, Levi LI, Vignuzzi M. 2018. The defective component of viral
826 populations. Curr Opin Virol 33:74-80.
827 4. Vasilijevic J, Zamarreño N, Oliveros JC, Rodriguez-Frandsen A, Gómez G,
828 Rodriguez G, Pérez-Ruiz M, Rey S, Barba I, Pozo F, Casas I, Nieto A, Falcón A.
829 2017. Reduced accumulation of defective viral genomes contributes to severe
830 outcome in influenza virus infected patients. PLoS Pathog 13:e1006650.
831 5. Zhou T, Gilliam NJ, Li S, Spandau S, Osborn RM, Connor S, Anderson CS,
832 Mariani TJ, Thakar J, Dewhurst S, Mathews DH, Huang L, Sun Y. 2023.
833 Generation and Functional Analysis of Defective Viral Genomes during SARS-
834 CoV-2 Infection. mBio 14:e0025023.
835 6. Xiao Y, Lidsky PV, Shirogane Y, Aviner R, Wu CT, Li W, Zheng W, Talbot D,
836 Catching A, Doitsh G, Su W, Gekko CE, Nayak A, Ernst JD, Brodsky L, Brodsky
837 E, Rousseau E, Capponi S, Bianco S, Nakamura R, Jackson PK, Frydman J,
838 Andino R. 2021. A defective viral genome strategy elicits broad protective
839 immunity against respiratory viruses. Cell 184:6037-6051.e14.
840 7. Kalamvoki M, Norris V. 2022. A Defective Viral Particle Approach to COVID-19.
841 Cells 11.

842 8. Smith SC, Gribble J, Diller JR, Wiebe MA, Thoner TW, Jr., Denison MR, Ogden
843 KM. 2021. Reovirus RNA recombination is sequence directed and generates
844 internally deleted defective genome segments during passage. J Virol 95.
845 9. Comins C, Simpson GR, Relph K, Harrington KJ, Melcher A, Pandha H. 2014.
846 Reoviral Therapy for Cancer, p 185-198, Gene Therapy of Cancer
847 doi:10.1016/b978-0-12-394295-1.00013-5.
848 10. Russell L, Peng K-W. 2018. The emerging role of oncolytic virus therapy against
849 cancer. Chinese Clinical Oncology 7:16.
850 11. Thirukkumaran C, Morris DG. 2009. Oncolytic Viral Therapy Using Reovirus, p
851 607-634. *In* Walther W, Stein US (ed), Gene Therapy of Cancer: Methods and
852 Protocols doi:10.1007/978-1-59745-561-9_31. Humana Press, Totowa, NJ.
853 12. Dermody T, Parker J, Sherry B. 2013. Orthoreovirus, p 1304–46. *In* Knipe D,
854 Howley P (ed), Fields virology, 6th ed. Lippincott, Williams, & Wilkins,
855 Philadelphia.
856 13. Guo Y, Parker JSL. 2021. The Paradoxes of Viral mRNA Translation during
857 Mammalian Orthoreovirus Infection. Viruses 13.
858 14. Thoner TW, Jr., Ye X, Karijolich J, Ogden KM. 2021. Reovirus Low-Density
859 Particles Package Cellular RNA. Viruses 13.
860 15. Matthijnssens J, Attoui H, Bányai K, Brussaard CPD, Danthi P, Del Vas M,
861 Dermody TS, Duncan R, Fāng 方勤 Q, Johne R, Mertens PPC, Mohd Jaafar F,
862 Patton JT, Sasaya 笹谷孝英 T, Suzuki 鈴木信弘 N, Wei 魏太云 T. 2022. ICTV
863 Virus Taxonomy Profile: Sedoreoviridae 2022. J Gen Virol 103.

864 16. Joklik WK, Roner MR. 1995. What reassorts when reovirus genome segments
865 reassort? J Biol Chem 270:4181-4.
866 17. Roner MR, Joklik WK. 2001. Reovirus reverse genetics: Incorporation of the CAT
867 gene into the reovirus genome. Proc Natl Acad Sci U S A 98:8036-41.
868 18. Roner MR, Bassett K, Roehr J. 2004. Identification of the 5' sequences required
869 for incorporation of an engineered ssRNA into the Reovirus genome. Virology
870 329:348-60.
871 19. Roner MR, Roehr J. 2006. The 3' sequences required for incorporation of an
872 engineered ssRNA into the Reovirus genome. Virol J 3:1.
873 20. Roner MR, Steele BG. 2007. Features of the mammalian orthoreovirus 3 Dearing
874 l1 single-stranded RNA that direct packaging and serotype restriction. J Gen Virol
875 88:3401-3412.
876 21. Roner MR, Steele BG. 2007. Localizing the reovirus packaging signals using an
877 engineered m1 and s2 ssRNA. Virology 358:89-97.
878 22. Demidenko AA, Blattman JN, Blattman NN, Greenberg PD, Nibert ML. 2013.
879 Engineering recombinant reoviruses with tandem repeats and a tetravirus 2A-like
880 element for exogenous polypeptide expression. Proc Natl Acad Sci U S A
881 110:E1867-76.
882 23. Lazzarini RA, Keene JD, Schubert M. 1981. The origins of defective interfering
883 particles of the negative-strand RNA viruses. Cell 26:145-54.
884 24. Xiao Y, Rouzine IM, Bianco S, Acevedo A, Goldstein EF, Farkov M, Brodsky L,
885 Andino R. 2016. RNA Recombination Enhances Adaptability and Is Required for
886 Virus Spread and Virulence. Cell Host Microbe 19:493-503.

887 25. Brennan JW, Sun Y. 2024. Defective viral genomes: advances in understanding
888 their generation, function, and impact on infection outcomes. mBio 15:e0069224.
889 26. Sun Y, Kim EJ, Felt SA, Taylor LJ, Agarwal D, Grant GR, López CB. 2019. A
890 specific sequence in the genome of respiratory syncytial virus regulates the
891 generation of copy-back defective viral genomes. PLOS Pathogens
892 15:e1007707.
893 27. Bao K, Zhang X, Li D, Sun W, Sun Z, Wang J, Zhu P. 2022. In situ structures of
894 polymerase complex of mammalian reovirus illuminate RdRp activation and
895 transcription regulation. Proc Natl Acad Sci U S A 119:e2203054119.
896 28. Kim J, Parker JS, Murray KE, Nibert ML. 2004. Nucleoside and RNA
897 triphosphatase activities of orthoreovirus transcriptase cofactor mu2. J Biol Chem
898 279:4394-403.
899 29. Noble S, Nibert ML. 1997. Core protein mu2 is a second determinant of
900 nucleoside triphosphatase activities by reovirus cores. J Virol 71:7728-35.
901 30. Yin P, Keirstead ND, Broering TJ, Arnold MM, Parker JS, Nibert ML, Coombs
902 KM. 2004. Comparisons of the M1 genome segments and encoded mu2 proteins
903 of different reovirus isolates. Virol J 1:6.
904 31. Pan M, Alvarez-Cabrera AL, Kang JS, Wang L, Fan C, Zhou ZH. 2021.
905 Asymmetric reconstruction of mammalian reovirus reveals interactions among
906 RNA, transcriptional factor μ2 and capsid proteins. Nat Commun 12:4176.
907 32. McDonald SM, Tao YJ, Patton JT. 2009. The ins and outs of four-tunneled
908 Reoviridae RNA-dependent RNA polymerases. Curr Opin Struct Biol 19:775-82.

909 33. Tao Y, Farsetta DL, Nibert ML, Harrison SC. 2002. RNA synthesis in a cage--
910 structural studies of reovirus polymerase lambda3. Cell 111:733-45.
911 34. Zhang X, Walker SB, Chipman PR, Nibert ML, Baker TS. 2003. Reovirus
912 polymerase lambda 3 localized by cryo-electron microscopy of virions at a
913 resolution of 7.6 A. Nat Struct Biol 10:1011-8.
914 35. Desselberger U. 1996. Genome rearrangements of rotaviruses. Adv Virus Res
915 46:69-95.
916 36. Routh A, Head SR, Ordoukhanian P, Johnson JE. 2015. ClickSeq:
917 Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors
918 to Stochastically Terminated 3'-Azido cDNAs. J Mol Biol 427:2610-6.
919 37. Kobayashi T, Antar AA, Boehme KW, Danthi P, Eby EA, Guglielmi KM, Holm
920 GH, Johnson EM, Maginnis MS, Naik S, Skelton WB, Wetzel JD, Wilson GJ,
921 Chappell JD, Dermody TS. 2007. A plasmid-based reverse genetics system for
922 animal double-stranded RNA viruses. Cell Host Microbe 1:147-57.
923 38. Kobayashi T, Ooms LS, Ikizler M, Chappell JD, Dermody TS. 2010. An improved
924 reverse genetics system for mammalian orthoreoviruses. Virology 398:194-200.
925 39. Ooms LS, Kobayashi T, Dermody TS, Chappell JD. 2010. A post-entry step in
926 the mammalian orthoreovirus replication cycle is a determinant of cell tropism. J
927 Biol Chem 285:41604-13.
928 40. Rodgers SE, Barton ES, Oberhaus SM, Pike B, Gibson CA, Tyler KL, Dermody
929 TS. 1997. Reovirus-induced apoptosis of MDCK cells is not linked to viral yield
930 and is blocked by Bcl-2. J Virol 71:2540-6.

931 41. Gribble J, Stevens LJ, Agostini ML, Anderson-Daniels J, Chappell JD, Lu X,
932 Pruijssers AJ, Routh AL, Denison MR. 2021. The coronavirus proofreading
933 exoribonuclease mediates extensive viral recombination. PLoS Pathog
934 17:e1009226.
935 42. Routh A, Johnson JE. 2014. Discovery of functional genomic motifs in viruses
936 with ViReMa-a Virus Recombination Mapper-for analysis of next-generation
937 sequencing data. Nucleic Acids Res 42:e11.
938 43. Zhou Y, Ahearn YP, Lokugamage KG, Alvarado RE, Estes LK, Meyers WM,
939 McLeland AM, Morgan AL, Murray JT, Walker DH, Johnson BA, Routh AL,
940 Menachery VD. 2024. SARS-CoV-2 EndoU-ribonuclease regulates RNA
941 recombination and impacts viral fitness. bioRxiv doi:10.1101/2024.11.11.622995.
942 44. Ni Y, Kemp MC. 1994. Subgenomic S1 segments are packaged by avian
943 reovirus defective interfering particles having an S1 segment deletion. Virus Res
944 32:329-42.
945 45. Alam MM, Kobayashi N, Ishino M, Nagashima S, Paul SK, Chawla-Sarkar M,
946 Krishnan T, Naik TN. 2008. Identical rearrangement of NSP3 genes found in
947 three independently isolated virus clones derived from mixed infection and
948 multiple passages of Rotaviruses. Arch Virol 153:555-9.
949 46. Schnepf N, Deback C, Dehee A, Gault E, Parez N, Garbarg-Chenon A. 2008.
950 Rearrangements of rotavirus genomic segment 11 are generated during acute
951 infection of immunocompetent children and do not occur at random. J Virol
952 82:3689-96.

953 47. Kojima K, Taniguchi K, Kawagishi-Kobayashi M, Matsuno S, Urasawa S. 2000.
954 Rearrangement generated in double genes, NSP1 and NSP3, of viable
955 progenies from a human rotavirus strain. Virus Res 67:163-71.
956 48. Poirier EZ, Mounce BC, Rozen-Gagnon K, Hooikaas PJ, Stapleford KA,
957 Moratorio G, Vignuzzi M. 2016. Low-Fidelity Polymerases of Alphaviruses
958 Recombine at Higher Rates To Overproduce Defective Interfering Particles.
959 Journal of Virology 90:2446-2454.
960 49. Jennings PA, Finch JT, Winter G, Robertson JS. 1983. Does the higher order
961 structure of the influenza virus ribonucleoprotein guide sequence rearrangements
962 in influenza viral RNA? Cell 34:619-27.
963 50. Li Y, Ball LA. 1993. Nonhomologous RNA recombination during negative-strand
964 synthesis of flock house virus RNA. J Virol 67:3854-60.
965 51. Peccoud J, Lequime S, Moltini-Conclois I, Giraud I, Lambrechts L, Gilbert C.
966 2018. A Survey of Virus Recombination Uncovers Canonical Features of Artificial
967 Chimeras Generated During Deep Sequencing Library Preparation. G3
968 (Bethesda) 8:1129-1138.
969 52. Girgis S, Xu Z, Oikonomopoulos S, Fedorova AD, Tchesnokov EP, Gordon CJ,
970 Schmeing TM, Götte M, Sonenberg N, Baranov PV, Ragoussis J, Hobman TC,
971 Pelletier J. 2022. Evolution of naturally arising SARS-CoV-2 defective interfering
972 particles. Communications Biology 5:1140.
973 53. Shitrit A, Nisnevich V, Rozenshtein N, Kobo H, Phan HV, Tay S, Szpara M,
974 Weitzman MD, Drayman N, Kobiler O. 2023. Shared sequence characteristics

975 identified in non-canonical rearrangements of HSV-1 genomes. J Virol
976 97:e0095523.
977 54. Jaworski E, Routh A. 2017. Parallel ClickSeq and Nanopore sequencing
978 elucidates the rapid evolution of defective-interfering RNAs in Flock House virus.
979 PLoS Pathog 13:e1006365.
980 55. Jaworski E, Routh A. 2018. ClickSeq: Replacing Fragmentation and Enzymatic
981 Ligation with Click-Chemistry to Prevent Sequence Chimeras. Methods Mol Biol
982 1712:71-85.
983 56. Hillung J, Olmo-Uceda MJ, Muñoz-Sánchez JC, Elena SF. 2024. Accumulation
984 Dynamics of Defective Genomes during Experimental Evolution of Two
985 Betacoronaviruses. Viruses 16.
986 57. Roner MR, Lin PN, Nepluev I, Kong LJ, Joklik WK. 1995. Identification of signals
987 required for the insertion of heterologous genome segments into the reovirus
988 genome. Proc Natl Acad Sci U S A 92:12362-6.
989 58. Dhar D, Mehanovic S, Moss W, Miller CL. 2024. Sequences at gene segment
990 termini inclusive of untranslated regions and partial open reading frames play a
991 critical role in mammalian orthoreovirus S gene packaging. PLoS Pathog
992 20:e1012037.
993 59. Killip MJ, Young DF, Gatherer D, Ross CS, Short JA, Davison AJ, Goodbourn S,
994 Randall RE. 2013. Deep sequencing analysis of defective genomes of
995 parainfluenza virus 5 and their role in interferon induction. J Virol 87:4798-807.

996 60. Ghorbani A, Ngunjiri JM, Rendon G, Brooke CB, Kenney SP, Lee CW. 2023.
997 Diversity and Complexity of Internally Deleted Viral Genomes in Influenza A Virus
998 Subpopulations with Enhanced Interferon-Inducing Phenotypes. Viruses 15.
999 61. Aguilar Rangel M, Dolan PT, Taguwa S, Xiao Y, Andino R, Frydman J. 2023.
1000 High-resolution mapping reveals the mechanism and contribution of genome
1001 insertions and deletions to RNA virus evolution. Proc Natl Acad Sci U S A
1002 120:e2304667120.
1003 62. Levi LI, Madden EA, Boussier J, Erazo D, Sanders W, Vallet T, Bernhauerova V,
1004 Moorman NJ, Heise MT, Vignuzzi M. 2024. Chikungunya Virus RNA Secondary
1005 Structures Impact Defective Viral Genome Production. Microorganisms 12.
1006 63. Chapell JD, Goral MI, Rodgers SE, dePamphilis CW, Dermody TS. 1994.
1007 Sequence diversity within the reovirus S2 gene: reovirus genes reassort in
1008 nature, and their termini are predicted to form a panhandle motif. J Virol 68:750-
1009 6.
1010 64. Chen D, Patton JT. 1998. Rotavirus RNA replication requires a single-stranded 3'
1011 end for efficient minus-strand synthesis. J Virol 72:7387-96.
1012 65. Tortorici MA, Shapiro BA, Patton JT. 2006. A base-specific recognition signal in
1013 the 5' consensus sequence of rotavirus plus-strand RNAs promotes replication of
1014 the double-stranded RNA genome segments. RNA 12:133-46.
1015 66. Li W, Manktelow E, von Kirchbach JC, Gog JR, Desselberger U, Lever AM.
1016 2010. Genomic analysis of codon, sequence and structural conservation with
1017 selective biochemical-structure mapping reveals highly conserved and dynamic

1018 structures in rotavirus RNAs with potential cis-acting functions. Nucleic Acids Res
1019 38:7718-35.
1020 67. Biswas S, Li W, Manktelow E, Lever J, Easton LE, Lukavsky PJ, Desselberger U,
1021 Lever AM. 2014. Physicochemical analysis of rotavirus segment 11 supports a
1022 'modified panhandle' structure and not the predicted alternative tRNA-like
1023 structure (TRLS). Arch Virol 159:235-48.
1024 68. Sung PY, Phelan JE, Luo D, Kulasegaran-Shylini R, Bohn P, Smyth RP, Roy P.
1025 2024. Recruitment of multi-segment genomic RNAs by Bluetongue virus requires
1026 a preformed RNA network. Nucleic Acids Res 52:8500-8514.
1027 69. Sung PY, Roy P. 2021. RNA Origami: Packaging a Segmented Genome in
1028 Orbivirus Assembly and Replication. Viruses 13.
1029 70. Sung PY, Roy P. 2014. Sequential packaging of RNA genomic segments during
1030 the assembly of Bluetongue virus. Nucleic Acids Res 42:13824-38.
1031 71. Borodavka A, Dykeman EC, Schrimpf W, Lamb DC. 2017. Protein-mediated
1032 RNA folding governs sequence-specific interactions between rotavirus genome
1033 segments. Elife 6.
1034 72. Kobayashi Y, Dadonaite B, van Doremalen N, Suzuki Y, Barclay WS, Pybus OG.
1035 2016. Computational and molecular analysis of conserved influenza A virus RNA
1036 secondary structures involved in infectious virion production. RNA Biol 13:883-
1037 94.
1038 73. Kennedy JC, Macdonald RD. 1982. Persistent infection with infectious pancreatic
1039 necrosis virus mediated by defective-interfering (DI) virus particles in a cell line
1040 showing strong interference but little DI replication. J Gen Virol 58:361-71.

1041 74. Chiba S, Lin YH, Kondo H, Kanematsu S, Suzuki N. 2016. A novel
1042 betapartitivirus RnPV6 from Rosellinia necatrix tolerates host RNA silencing but
1043 is interfered by its defective RNAs. Virus Res 219:62-72.
1044 75. Qiao X, Qiao J, Mindich L. 1997. Anin VitroSystem for the Investigation of
1045 Heterologous RNA Recombination. Virology 227:103-110.
1046 76. AlShaikhahmed K, Leonov G, Sung PY, Bingham RJ, Twarock R, Roy P. 2018.
1047 Dynamic network approach for the modelling of genomic sub-complexes in multi-
1048 segmented viruses. Nucleic Acids Res 46:12087-12098.
1049 77. Fajardo T, Jr., Sung PY, Roy P. 2015. Disruption of Specific RNA-RNA
1050 Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and
1051 Virus Infectivity. PLoS Pathog 11:e1005321.
1052 78. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ. 1989. A simple
1053 method for site-directed mutagenesis using the polymerase chain reaction.
1054 Nucleic Acids Res 17:6545-51.
1055 79. Eaton HE, Kobayashi T, Dermody TS, Johnston RN, Jais PH, Shmulevitz M.
1056 2017. African Swine Fever Virus NP868R Capping Enzyme Promotes Reovirus
1057 Rescue during Reverse Genetics by Promoting Reovirus Protein Expression,
1058 Virion Assembly, and RNA Incorporation into Infectious Virions. J Virol 91.
1059 80. Berard A, Coombs KM. 2009. Mammalian reoviruses: propagation, quantification,
1060 and storage. Curr Protoc Microbiol Chapter 15:Unit15C.1.
1061 81. Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput
1062 sequencing reads. EMBnetjournal 17.

82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

83. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, https://ggplot2.tidyverse.org.

84. Vaughan HWaRFaLHaKMaD. 2023. dplyr: A Grammar of Data Manipulation, https://CRAN.R-project.org/package=dplyr.

85. Pagès H AP, Gentleman R, DebRoy S. 2023. Biostrings: Efficient manipulation of biological strings. R package version 2.74.1. https://bioconductor.org/packages/Biostrings. Accessed

86. Wagih O. 2024. ggseqlogo: A 'ggplot2' Extension for Drawing Publication-Ready Sequence Logos, https://CRAN.R-project.org/package=ggseqlogo.

87. Yan L. 2024. ggvenn: Draw Venn Diagram by 'ggplot2, https://yanlinlin82.github.io/ggvenn/.

1080 ## FIGURE LEGENDS

1081 **Figure 1. Polymerase complex-exchanged reovirus characterization.** (A) Schematic
1082 of polymerase-exchanged viruses and their polymerase complexes. The origin of viral
1083 gene segments, represented as parallel lines inside the particle, the virus background,
1084 represented by the particle, and polymerase complex are indicated by color (red, T1L;
1085 blue, T3D). Created using Biorender.com (B) Electropherotypes of T1L, T3D, and
1086 polymerase-exchanged reoviruses. RNA extracted from cells infected with the indicated
1087 virus was resolved by SDS-PAGE. Blue circles indicate T3D L1 or M1 gene segments.
1088 Red circles indicate T1L L1 or M1 gene segments. (C) Replication kinetics of parental
1089 and polymerase-exchanged reoviruses in L cells. Viral titers were determined by
1090 fluorescent focus assay. Error bars = SD, n = 3 clones per virus. Not statistically significant
1091 by two-way ANOVA. (D) Plaque size of parental and polymerase-exchanged reoviruses
1092 in L cells. ns, not significant; **, $P < 0.01$; ****, $P < 0.0001$ by one-way ANOVA. (E)
1093 Replication kinetics of parental and polymerase-exchanged reoviruses in MDCK cells.
1094 Viral titers were determined by plaque assay. Error bars = SD, n = 3 clones per virus. Not
1095 statistically significant by two-way ANOVA.

1096

1097 **Figure 2. Serial passage lineage 1 to 3 reovirus gene segment L1 patterns.** RNA
1098 extracted from serial passages was used as a template for RT-PCR reactions in which
1099 primers targeting terminal sequences in the L1 segments of both T1L and T3D were
1100 included. Products from T1L (A), T1L-T3pol (B), T3D-T1pol (C), or T3D (D) P1-P10 were
1101 resolved in 1.2% ethidium bromide-stained agarose gels. The evolutionary lineage is

1102 indicated. Black asterisks indicate the position of the full-length segment. Triangles
1103 indicate products smaller than the full-length segment.
1104
1105 **Figure 3. Sequences of isolated DVGs from serial passages.** Schematics, drawn to
1106 scale, of full-length reovirus L1 and S4 segments and of sequenced T1L-T3pol and T3D-
1107 T1pol reovirus DVGs. The ORF is colored purple (L1) or green (S4). The 5′ and 3′ UTRs
1108 are colored gray. Central deletions were identified for each DVG. Dotted lines represent
1109 the missing sequences. Below the schematics are the DVG sequences, with DVG
1110 junctions indicated by "/".
1111
1112 **Figure 4. Reovirus recombination junction site location and frequency.** (A)
1113 Schematic of a recombination junction site in the context of the parental genome
1114 segment, in which the 5' and 3' termini are conserved (bottom). The junction is labeled
1115 with a red caret (^), and example sequences are shown. Sequences upstream of the
1116 junction start site and downstream of the junction stop site may form a final product that
1117 is missing the internal genome sequence (top). (B) Schematic of junction scatter plots
1118 indicating the different cluster within the plot marked by dashed boxes: (i) 5' → 3' (ii) mid-
1119 genome → 3' (iii) 3' → 3' (iv) local deletions/microindels (v) 5' UTR → rest of genome (vi)
1120 duplications. (C-F) Recombination junction site location and frequency in sequenced T1L,
1121 T1L-T3pol, T3D-T1pol, and T3D virion RNA for gene segments L1 (C), L3 (D), M1 (E),
1122 and M3 (F). Junction sites are indicated by dots whose position corresponds to upstream
1123 and downstream sequences that are merged to form a novel junction. Junction frequency

1124 is indicated by dot color, according to the legend at the right of each image. Results from
1125 a representative clone for each virus are shown.
1126
1127 **Figure 5. Reovirus recombination start-site frequency across gene segments.**
1128 Recombination start-site frequency at each nucleotide position across L1 (A), L3 (B), M1
1129 (C), and M3 (D) segments from T1L, T1L-T3pol, T3D-T1pol, and T3D. Mean positional
1130 recombination frequency for sequenced virion RNA from three clones per virus is shown.
1131
1132 **Figure 6. Principal component analysis of reovirus recombination profiles.** Score
1133 plots for principal component analyses (PCA) of gene segments L1 (A), M1 (B), L3 (C),
1134 and M3 (D) applied to junctional recombination frequency data for three independent
1135 samples of reovirus strains T1L, T1L-T3pol, T3D-T1pol, and T3D.
1136
1137 **Figure 7. Recombination junctions shared among wild-type and polymerase-**
1138 **exchanged viruses.** All the unique recombination junctions that are common between
1139 sequenced viral RNA from T1L, T1L-T3pol, T3D-T1pol, and T3D were identified using
1140 the “ggvenn” package in R. Analyses were conducted for segments L1 (A), L3 (B), M1
1141 (C), and M3 (D). Black cells show total unique junctions for each virus within the
1142 indicated gene segment. Combined data from three independent samples of each virus
1143 were included in the analysis.
1144
1145 **Figure 8. Microhomology and logo plots.** (A) Sequence microhomology for T1L, T1L-
1146 T3pol, T3D-T1pol, and T3D. The nucleotides that overlap at each position in 25-bp

1147 regions flanking the start and stop recombination junction sites were quantified for
1148 segments L1 and L3. The junction is labeled zero and denotated with a solid red line. (B)
1149 Nucleotide composition was calculated as the percent adenosine (A), cytosine (C),
1150 guanine (G), and Thymine (T) at each position in a 10-bp region surrounding the start and
1151 stop recombination junction sites. Positions upstream (−10 to −1) and downstream (+1 to
1152 +10) of the junction position are indicated. Each analyzed virus contains combined data
1153 from three independent samples.
1154
1155 **Figure 9. Recombination site selection.** (A) Frequency of microhomology sequences
1156 in segments. Four-nucleotide sequences present in unique recombination junctions in
1157 the L1 or L3 segment from T1L, T1L-T3pol, T3D-T1pol, and T3D were identified. Then,
1158 the number of times each four-nucleotide sequence is present in segment L1 or L3 was
1159 computationally quantified. For L1, n = 176-240. For L3, n = 145-171. Dashed lines
1160 represent the median and quartiles. (B) Usage of microhomology sites for
1161 recombination. The number of times each occurrence of a four-nucleotide sequence
1162 from a unique start or stop site analyzed in (A) was incorporated into a recombination
1163 site in the L1 or L3 segment was computationally quantified for T1L, T1L-T3pol, T3D-
1164 T1pol, and T3D. For L1, n= 2915-3180. For L3, n= 2502-2851. Dashed lines represent
1165 the mean.
1166
1167 **Figure 10. Recombination model for dsRNA viruses.** (1) The RdRp pauses RNA
1168 synthesis at the junction start site during minus-strand synthesis. (2) The plus-strand
1169 template moves through the RdRp while synthesis is paused. Regions of sequence

1170 homology to junction start-site residues are bypassed. (3) The RdRp re-initiates
1171 synthesis at a specific junction stop site that has sequence homology to the junction
1172 start site. Annealing of the newly synthesized strand and other features, such as RNA
1173 secondary structure, interactions, or surrounding sequences, may influence re-initiation.
1174 (4) The RdRp continues synthesis to form a DVG or DVG template.
1175
1176

Table 1 Alignment statistics

Alignment statistic	Segment	T3D 1	T3D 2	T3D 3	T3D-T1pol 1	T3D-T1pol 2	T3D-T1pol 3	T1L-T3pol 1	T1L-T3pol 2	T1L-T3pol 3	T1L 1	T1L 2	T1L 3
% reads mapped to virus		>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
Total reads		25260410	28604063	30095851	26658380	27516061	29361973	30308219	21458634	25077376	26946402	29957691	28476911
Jfreq [a]		1.98	2.15	1.24	1.82	1.64	1.91	1.41	1.02	1.28	1.42	1.51	1.05
# reads with recombination events													
	L1	1158	1513	1321	1390	1531	1802	1164	583	1018	973	1885	1081
	L2	1197	1424	591	1016	1026	1166	578	212	249	343	414	242
	L3	599	749	681	752	595	743	926	499	412	542	848	667
	M1	184	237	140	188	159	252	210	108	139	207	184	94
	M2	190	234	180	189	126	182	390	202	302	503	368	258
	M3	1043	1385	441	769	720	1009	207	120	181	270	158	115
	S1	262	135	47	164	49	76	310	214	581	707	322	234
	S2	107	137	69	90	63	80	138	61	95	96	86	59
	S3	178	210	202	154	170	177	75	67	142	110	67	63
	S4	94	129	71	139	83	118	284	122	102	67	200	183
# reads mapped to viral gene segment													
	L1	5,106,848	5,824,267	7,563,379	5,830,947	6,058,222	6,264,212	5,801,679	3,638,187	3,704,550	4,417,355	5,891,865	5,463,064
	L2	6,662,826	6,978,216	7,265,286	6,848,569	7,341,564	7,322,106	4,720,881	2,995,701	2,889,015	3,371,773	4,430,912	4,339,775
	L3	3,162,471	3,729,886	4,220,554	3,152,933	3,516,584	3,631,513	4,330,594	3,031,626	2,864,871	3,379,424	4,377,796	4,097,063
	M1	2,010,161	2,401,261	2,171,824	2,366,934	2,368,327	2,690,442	2,186,178	1,630,315	1,863,444	2,278,138	2,344,993	2,167,978
	M2	1,988,216	2,285,659	2,066,003	1,943,737	1,936,186	2,209,843	5,631,920	3,996,113	3,903,288	4,333,452	5,549,127	5,224,089
	M3	2,331,392	2,867,077	2,901,518	2,395,419	2,473,377	2,921,266	1,715,465	1,352,975	1,606,108	1,774,066	1,645,633	1,639,310
	S1	1,321,199	1,452,403	1,090,694	1,336,338	1,258,269	1,428,736	2,284,308	2,002,678	3,496,513	3,331,144	2,410,284	2,278,743
	S2	1,059,981	1,048,569	824,693	1,122,301	956,053	1,121,603	1,550,675	1,161,403	1,783,005	1,500,017	1,435,369	1,241,740
	S3	500,594	590,814	767,147	549,583	524,739	570,384	816,523	675,750	1,445,408	1,328,825	799,117	921,942
	S4	1,112,954	1,342,805	1,219,969	1,099,702	1,079,295	1,198,661	1,266,340	971,054	1,519,289	1,230,540	1,069,488	1,100,512

[a] Number of reads containing junctions per 10,000 mapped reads

Table 2 Microhomology surrounding junctions within the L1 and L3 segments

L1 Segment				L3 Segment			
Virus	Total Junctions	Number conserved	Percent Conserved	Virus	Total Junctions	Number Conserved	Percent Conserved
Eight nucleotide positions conserved							
T1L	663	65	10	T1L	525	41	8
T1L-T3pol	735	73	10	T1L-T3pol	430	48	11
T3D-T1pol	916	74	8	T3D-T1pol	709	63	9
T3D	800	54	7	T3D	420	57	14
Six nucleotide positions conserved							
T1L	663	294	44	T1L	525	201	38
T1L-T3pol	735	297	40	T1L-T3pol	430	188	44
T3D-T1pol	916	330	36	T3D-T1pol	709	217	31
T3D	800	288	36	T3D	420	183	44
Four nucleotide positions conserved							
T1L	663	506	76	T1L	525	333	63
T1L-T3pol	735	568	77	T1L-T3pol	430	316	73
T3D-T1pol	916	611	67	T3D-T1pol	709	394	56
T3D	800	600	75	T3D	420	323	77

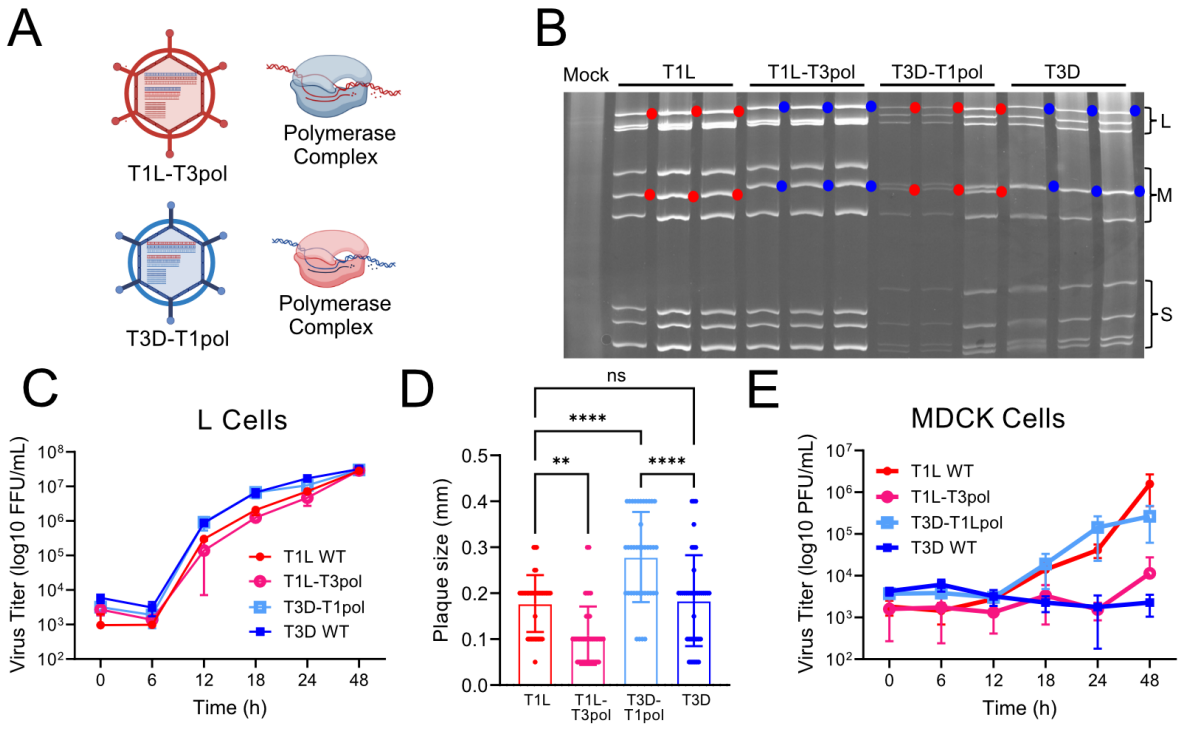

Figure 1. Polymerase complex-exchanged reovirus characterization. (A) Schematic of polymerase-exchanged viruses and their polymerase complexes. The origin of viral gene segments, represented as parallel lines inside the particle, the virus background, represented by the particle, and polymerase complex are indicated by color (red, T1L; blue, T3D). Created using Biorender.com (B) Electropherotypes of T1L, T3D, and polymerase-exchanged reoviruses. RNA extracted from cells infected with the indicated virus was resolved by SDS-PAGE. Blue circles indicate T3D L1 or M1 gene segments. Red circles indicate T1L L1 or M1 gene segments. (C) Replication kinetics of parental and polymerase-exchanged reoviruses in L cells. Viral titers were determined by fluorescent focus assay. Error bars = SD, n = 3 clones per virus. Not statistically significant by two-way ANOVA. (D) Plaque size of parental and polymerase-exchanged reoviruses in L cells. ns, not significant; **, $P < 0.01$; ****, $P < 0.0001$ by one-way ANOVA. (E) Replication kinetics of parental and polymerase-exchanged reoviruses in MDCK cells. Viral titers were determined by plaque assay. Error bars = SD, n = 3 clones per virus. Not statistically significant by two-way ANOVA.

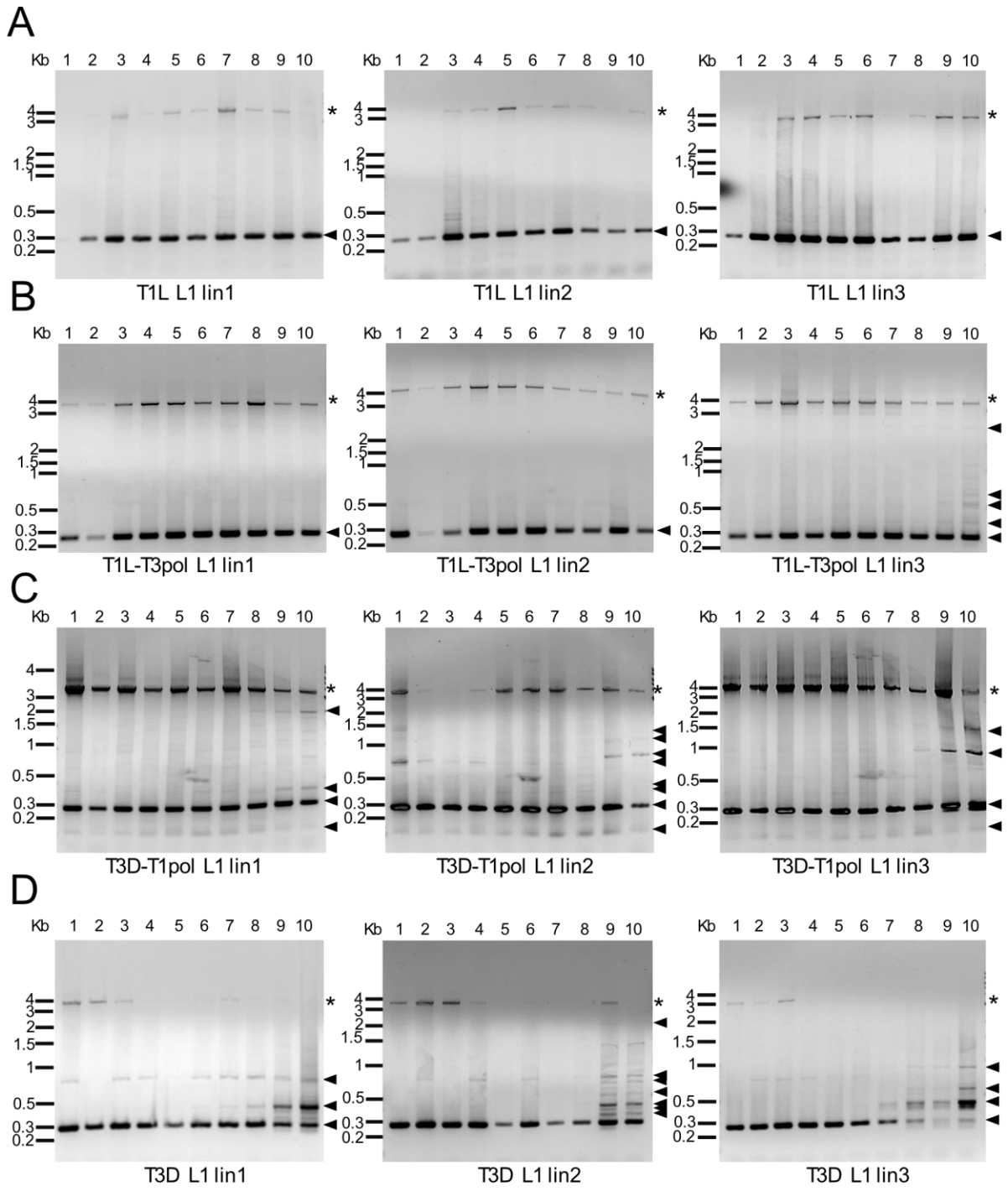

Figure 2. Serial passage lineage 1 to 3 reovirus gene segment L1 patterns. RNA extracted from serial passages was used as a template for RT-PCR reactions in which primers targeting terminal sequences in the L1 segments of both T1L and T3D were included. Products from T1L (A), T1L-T3pol (B), T3D-T1pol (C), or T3D (D) P1-P10 were resolved in 1.2% ethidium bromide-stained agarose gels. The evolutionary lineage is indicated. Black asterisks indicate the position of the full-length segment. Triangles indicate products smaller than the full-length segment.

T1L L1 DVG (Sequence from T3D-T1pol passages)
GCTACACGTTCCACGACAATGTCATCCATGATACTGACTCAGTTTAGACCGTTCATTGAAAGCATTTCAGGTATCACT
GACCAATCGAATGACGTGTTTGAAGATGCAGCAAAAGCATTCTCTATGTTTACTCGCAGCGACGTCTACAAGGCGCT
GGATGAAATACCTTTCTCTGATGATGCGATGCTTCCCATCCCTCCAACTATATATACTAAACCATCTCACGATTCATAT
TATTACATTGATGCTCTAAACCGTGT/GGTGCGTCAACTCATC

T3D L1 DVG (Sequence from T1L-T3pol passages)
GCTACACGTTCCACTTCCACGACAATGTCATCCATGATACTGACTCAGTTTGGACCGTTCATTGAGAGCATTTCAGGT
ATCACTGATCAATCGAATGACGTGTTTGAAGATGCAGCAAAAGCATTCTCTATGTTTACTCGCAGCGATGTCTACAAG
GCGCTGGATGAAATACCTTTCTCTGATGATGCGATGCTTCCAATCCCTCCAACTATATATACGAAACCATCTCACGAT
TCATATTATTACATTGATGCTCTAAACCGTGT/GGTGCGTCAACTCATC

T1L S4 DVG (Sequence from T1L-T3pol passages)
GCTATTTTTGCCTCTTCCCAAACGTTGTCGCAATGGAGGTGTGCTTGCCCAACGGTCATCAGATCGTGGACTTGATT
AACAACGCTTTTGAAGGTCGTGTATCAATCTACAGCGCGCAAGAGGGATGGGACAA/ACACCCCCATCTTC

T3D S4 DVG (Sequence from T3D-T1pol passages)
GCTATTTTTGCCTCTTCCCAGACGTTGTCGCAATGGAGGTGTGCTTGCCCAACGGTCATCAGGTCGTGGACTTGATT
AACAACGCTTTTGAAGGTCGTGTATCAATCTACAGCGCGCAAGAGGGATGGGACAA/AGGCTTCATTCATC

Figure 3. Sequences of isolated DVGs from serial passages. Schematics, drawn to scale, of full-length reovirus L1 and S4 segments and of sequenced T1L-T3pol and T3D-T1pol reovirus DVGs. The ORF is colored purple (L1) or green (S4). The 5′ and 3′ UTRs are colored gray. Central deletions were identified for each DVG. Dotted lines represent the missing sequences. Below the schematics are the DVG sequences, with DVG junctions indicated by “/”.

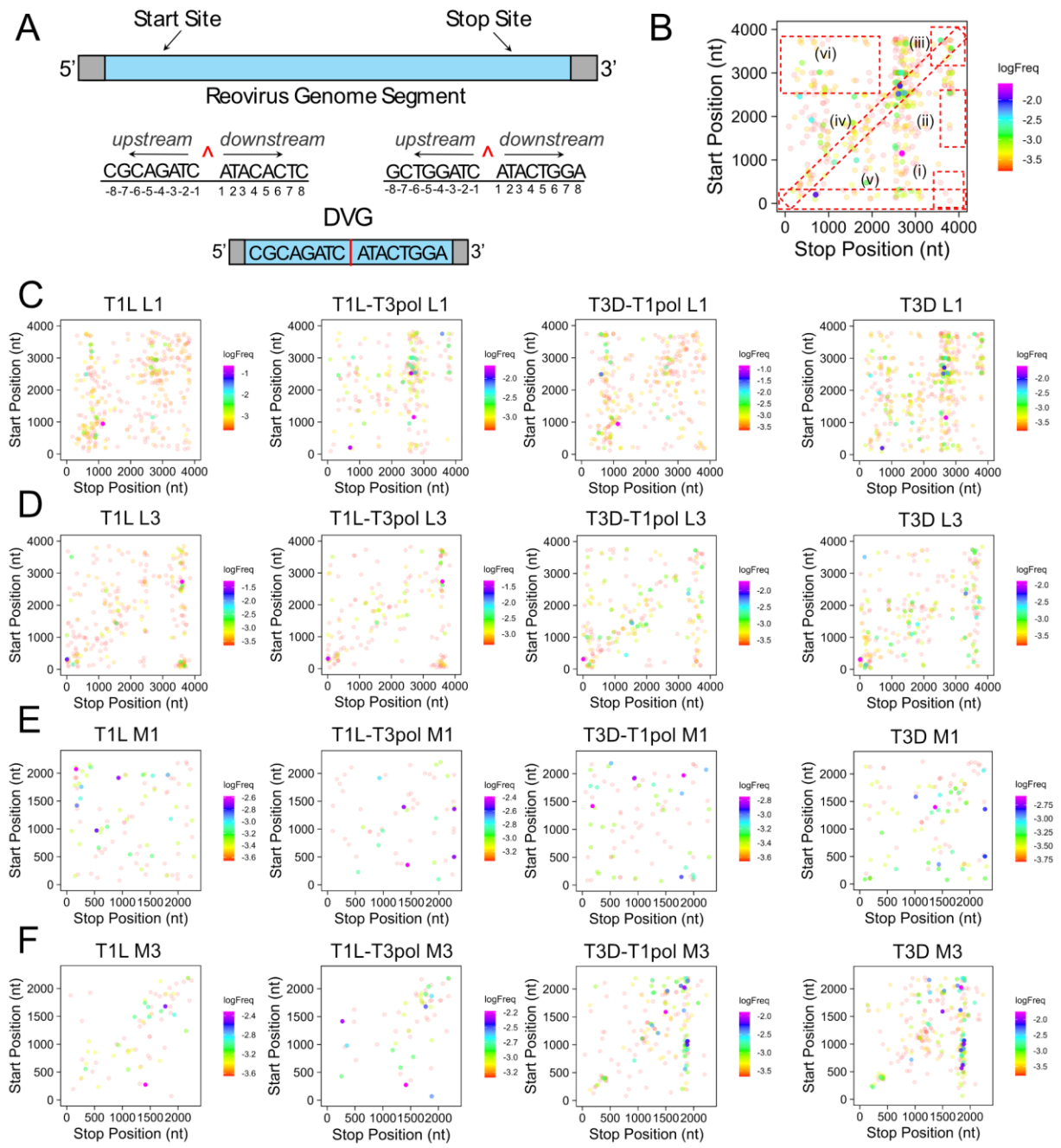

Figure 4. Reovirus recombination junction site location and frequency. (A) Schematic of a recombination junction site in the context of the parental genome segment, in which the 5' and 3' termini are conserved (bottom). The junction is labeled with a red caret (^), and example sequences are shown. Sequences upstream of the junction start site and downstream of the junction stop site may form a final product that is missing the internal genome sequence (top). (B) Schematic of junction scatter plots indicating the different cluster within the plot marked by dashed boxes: (i) 5' → 3' (ii) mid-genome → 3' (iii) 3' → 3' (iv) local deletions/microindels (v) 5' UTR → rest of genome (vi) duplications. (C-F) Recombination junction site location and frequency in sequenced T1L, T1L-T3pol, T3D-T1pol, and T3D virion RNA for gene segments L1 (C), L3 (D),

M1 (E), and M3 (F). Junction sites are indicated by dots whose position corresponds to upstream and downstream sequences that are merged to form a novel junction. Junction frequency is indicated by dot color, according to the legend at the right of each image. Results from a representative clone for each virus are shown.

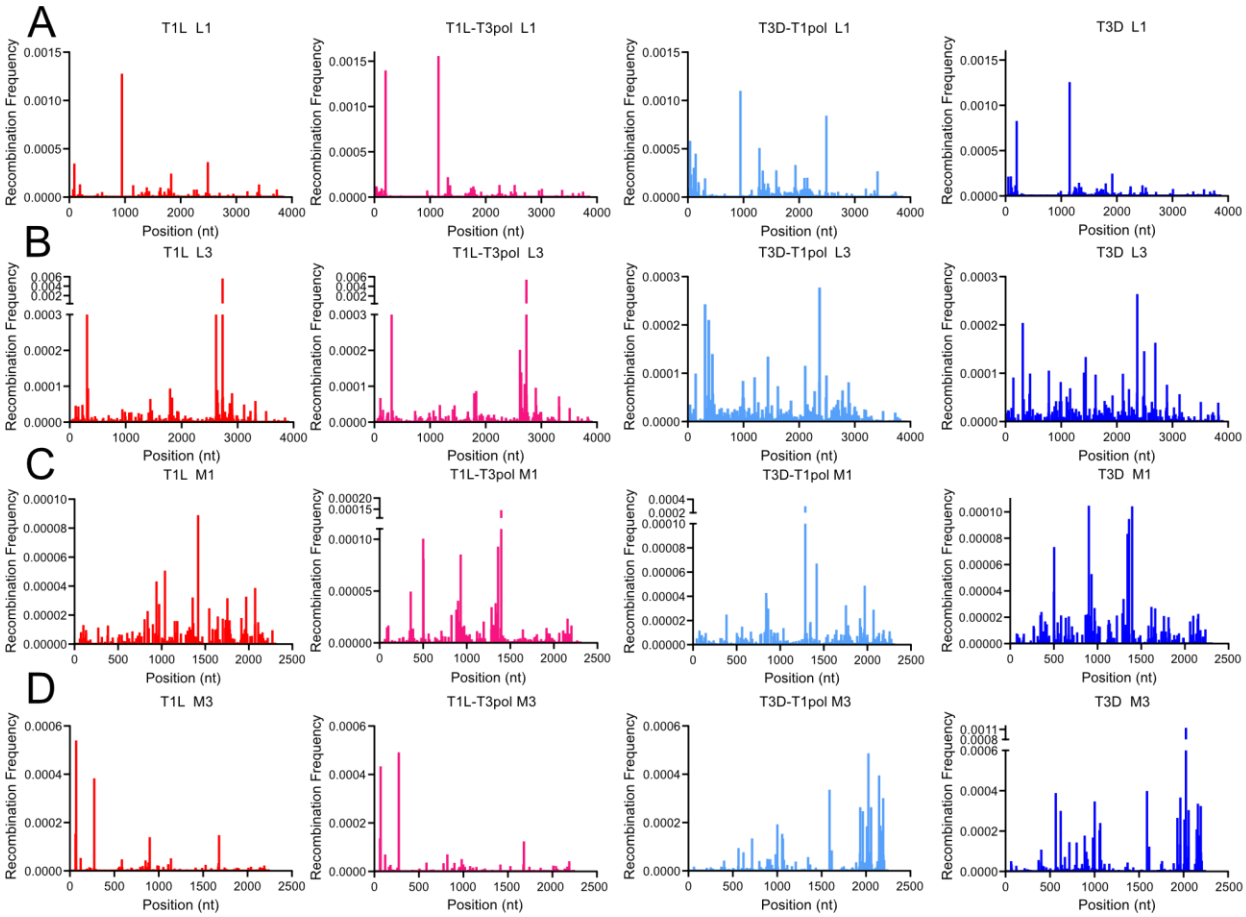

Figure 5. Reovirus recombination start-site frequency across gene segments. Recombination start-site frequency at each nucleotide position across L1 (A), L3 (B), M1 (C), and M3 (D) segments from T1L, T1L-T3pol, T3D-T1pol, and T3D. Mean positional recombination frequency for sequenced virion RNA from three clones per virus is shown.

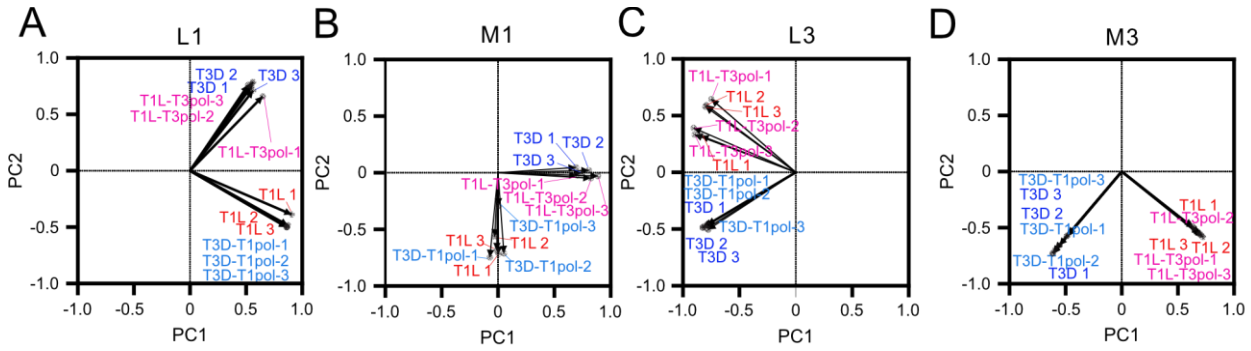

Figure 6. Principal component analysis of reovirus recombination profiles. Score plots for principal component analyses (PCA) of gene segments L1 (A), M1 (B), L3 (C), and M3 (D) applied to junctional recombination frequency data for three independent samples of reovirus strains T1L, T1L-T3pol, T3D-T1pol, and T3D.

A

T1L	385			
T1L-T3pol	13	387		
T3D-T1pol	144	21	615	
T3D	26	214	29	816
L1 Segment	T1L	T1L-T3pol	T3D-T1pol	T3D

B

T1L	341			
T1L-T3pol	91	245		
T3D-T1pol	15	12	523	
T3D	15	19	104	399
L3 Segment	T1L	T1L-T3pol	T3D-T1pol	T3D

C

T1L	131			
T1L-T3pol	2	128		
T3D-T1pol	26	4	197	
T3D	3	21	11	183
M1 Segment	T1L	T1L-T3pol	T3D-T1pol	T3D

D

T1L	114			
T1L-T3pol	41	118		
T3D-T1pol	1	2	453	
T3D	3	3	203	387
M3 Segment	T1L	T1L-T3pol	T3D-T1pol	T3D

Figure 7. Recombination junctions shared among wild-type and polymerase-exchanged viruses. All the unique recombination junctions that are common between sequenced viral RNA from T1L, T1L-T3pol, T3D-T1pol, and T3D were identified using the “ggvenn” package in R. Analyses were conducted for segments L1 (A), L3 (B), M1 (C), and M3 (D). Black cells show total unique junctions for each virus within the indicated gene segment. Combined data from three independent samples of each virus were included in the analysis.

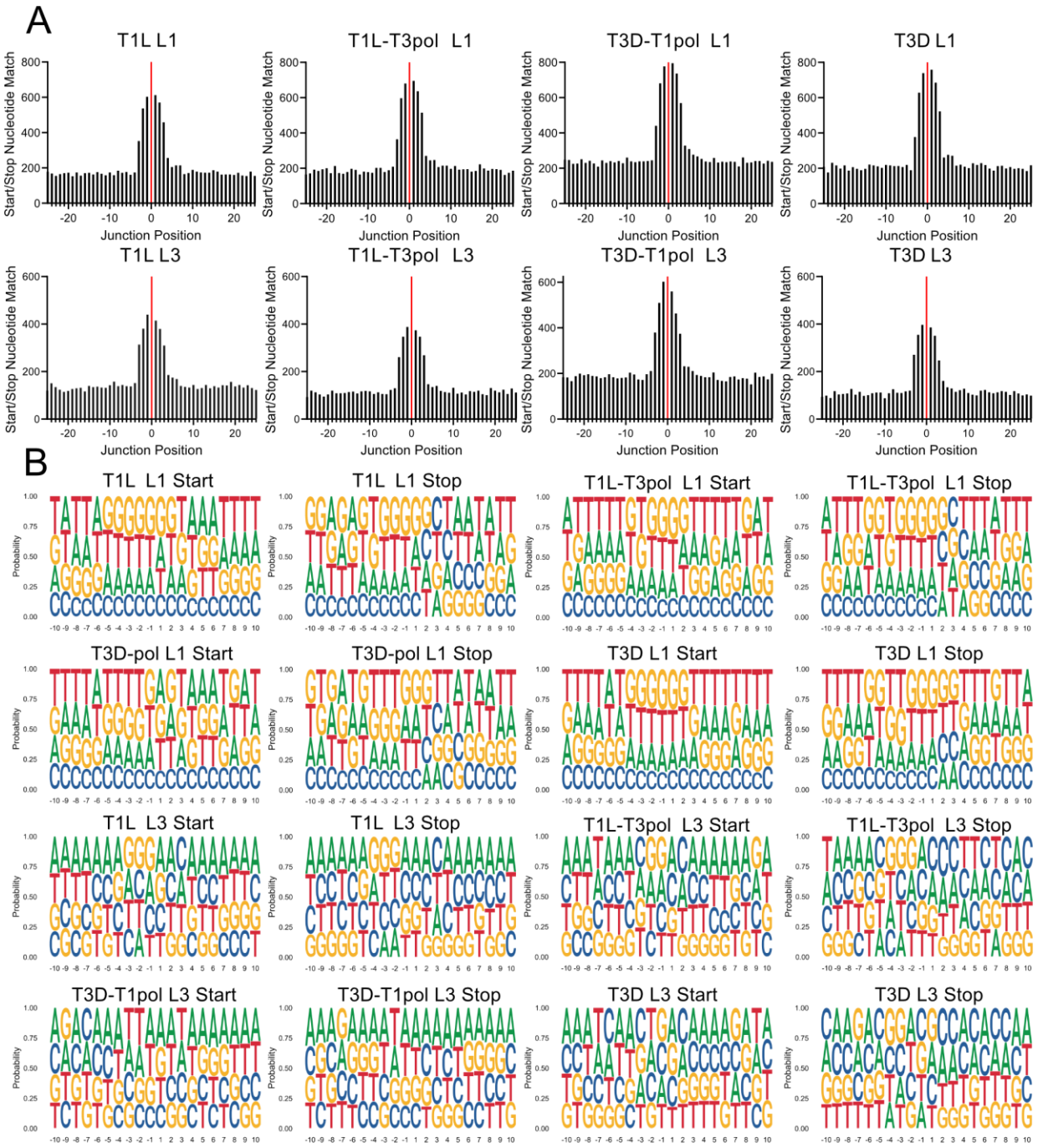

Figure 8. Microhomology and logo plots. (A) Sequence microhomology for T1L, T1L-T3pol, T3D-T1pol, and T3D. The nucleotides that overlap at each position in 25-bp regions flanking the start and stop recombination junction sites were quantified for segments L1 and L3. The junction is labeled zero and denotated with a solid red line. (B) Nucleotide composition was calculated as the percent adenosine (A), cytosine (C), guanine (G), and Thymine (T) at each position in a 10-bp region surrounding the start and stop recombination junction sites. Positions upstream (−10 to −1) and downstream (+1 to +10) of the junction position are indicated. Each analyzed virus contains combined data from three independent samples.

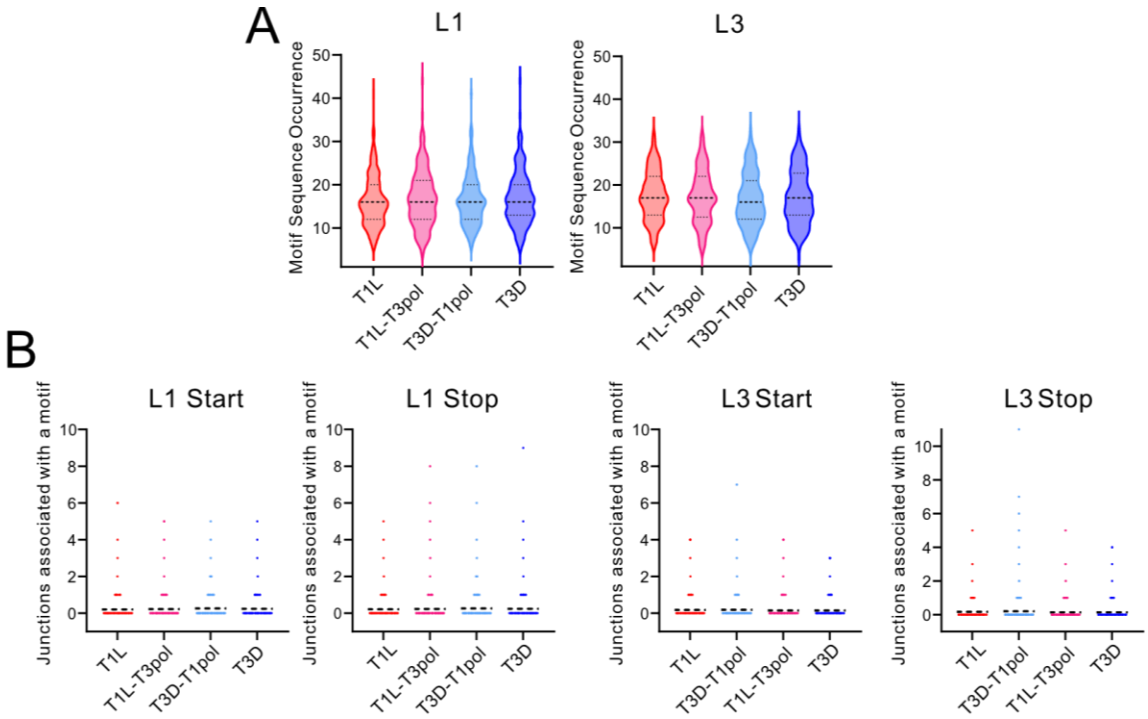

Figure 9. Recombination site selection. (A) Frequency of microhomology sequences in segments. Four-nucleotide sequences present in unique recombination junctions in the L1 or L3 segment from T1L, T1L-T3pol, T3D-T1pol, and T3D were identified. Then, the number of times each four-nucleotide sequence is present in segment L1 or L3 was computationally quantified. For L1, n = 176-240. For L3, n = 145-171. Dashed lines represent the median and quartiles. (B) Usage of microhomology sites for recombination. The number of times each occurrence of a four-nucleotide sequence from a unique start or stop site analyzed in (A) was incorporated into a recombination site in the L1 or L3 segment was computationally quantified for T1L, T1L-T3pol, T3D-T1pol, and T3D. For L1, n= 2915-3180. For L3, n= 2502-2851. Dashed lines represent the mean.

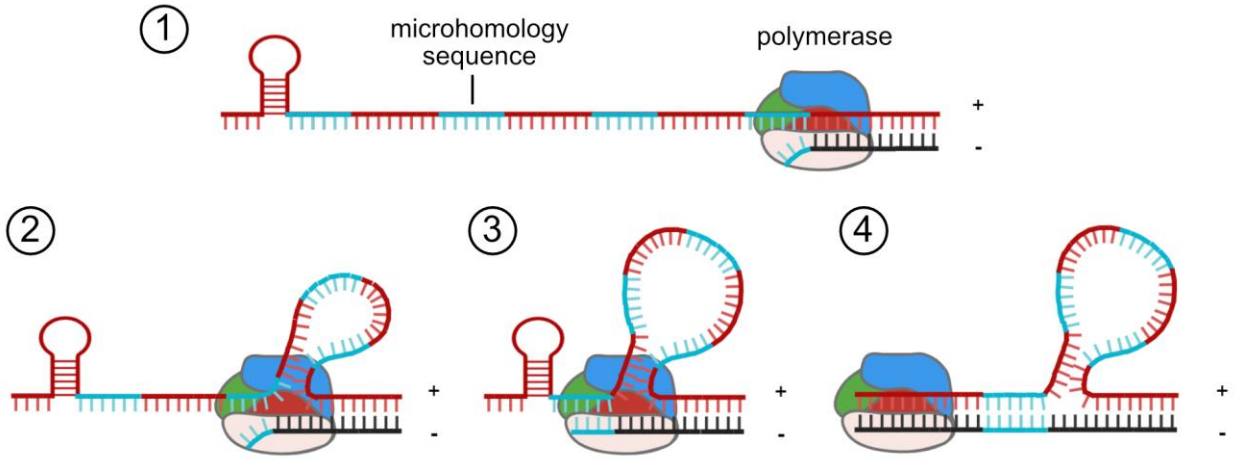

Figure 10. Recombination model for dsRNA viruses. (1) The RdRp pauses RNA synthesis at the junction start site during minus-strand synthesis. (2) The plus-strand template moves through the RdRp while synthesis is paused. Regions of sequence homology to junction start-site residues are bypassed. (3) The RdRp re-initiates synthesis at a specific junction stop site that has sequence homology to the junction start site. Annealing of the newly synthesized strand and other features, such as RNA secondary structure, interactions, or surrounding sequences, may influence re-initiation. (4) The RdRp continues synthesis to form a DVG or DVG template.