The coronavirus proofreading exoribonuclease mediates extensive viral recombination

Jennifer Gribble[1,3], Andrea J. Pruijssers[2], Maria L. Agostini[2], Jordan Anderson-Daniels[2], James D. Chappell[2], Xiaotao Lu[2], Laura J. Stevens[2], Andrew L. Routh[4,5]*, Mark R. Denison[1,2,3,#]*

[1] Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

[2] Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA

[3] Vanderbilt Institute for Infection, Immunology, and Inflammation (VI4), Vanderbilt University Medical Center, Nashville, TN, USA

[4] Department of Biochemistry and Molecular Biology, University of Texas – Medical Branch, Galveston, TX, USA

[5] Sealy Center for Structural Biology and Molecular Biophysics, University of Texas – Medical Branch, Galveston, TX, USA

[#]Lead Contact: M.R.D: mark.denison@vumc.org

*Correspondence: mark.denison@vumc.org; alrouth@utmb.edu

SUMMARY

Coronaviruses (CoVs) emerge as zoonoses and cause severe disease in humans, demonstrated by the SARS-CoV-2 (COVID-19) pandemic. RNA recombination is required during normal CoV replication for subgenomic mRNA (sgmRNA) synthesis and

generates defective viral genomes (DVGs) of unknown function. However, the determinants and patterns of CoV recombination are unknown. Here, we show that divergent β-CoVs SARS-CoV-2, MERS-CoV, and murine hepatitis virus (MHV) perform extensive RNA recombination in culture, generating similar patterns of recombination junctions and diverse populations of DVGs and sgmRNAs. We demonstrate that the CoV proofreading nonstructural protein (nsp14) 3'-to-5' exoribonuclease (nsp14-ExoN) is required for normal CoV recombination and that its genetic inactivation causes significantly decreased frequency and altered patterns of recombination in both infected cells and released virions. Thus, nsp14-ExoN is a key determinant of both high fidelity CoV replication and recombination, and thereby represents a highly-conserved and vulnerable target for virus inhibition and attenuation.

Keywords:
coronavirus, exoribonuclease, recombination, proofreading, direct RNA sequencing, nanopore, defective genome, SARS-CoV-2, MERS-CoV

INTRODUCTION

Coronaviruses (CoVs) are a family of positive-sense, single-stranded RNA viruses with genomes ranging in size between 26 and 32 kb (Figure 1A and 3A). CoV recombination has been reported to be associated with increased spread, severe disease, and vaccine failure in livestock CoV epidemics (Chen et al., 2017; Feng et al., 2018) and proposed to be important in the emergence of human CoVs including the beta-CoVs (β-CoVs) severe acute respiratory syndrome coronavirus (SARS-CoV) (Anthony et al., 2017;

Drosten et al., 2003; Hon et al., 2008; Ksiazek et al., 2003; Li et al., 2005) and Middle East respiratory syndrome coronavirus (MERS-CoV) (Yusof et al., 2017; Zaki et al., 2012). The ongoing severe global pandemic of SARS-CoV-2, the etiological agent of coronavirus disease 2019 (COVID-19) (Wu et al., 2020; Zhou et al., 2020) underlines the importance of defining the determinants of CoV evolution and emergence into human populations (Paraskevis et al., 2020; Randhawa et al., 2020; Wahba et al., 2020; Wong et al., 2020; Zhou et al., 2020). Studies comparing CoV strains that are closely related to SARS-CoV-2 have proposed that SARS-CoV-2 acquired the ability to infect human cells through recombination within the spike protein sequence (Huang et al., 2020; Li et al., 2020; Patiño-Galindo et al., 2020). Further, a study of genetic variation in patient SARS-CoV-2 samples has suggested that recombination may be occurring during infections in humans (Yi, 2020). Together, these data support the hypothesis that generation of novel CoVs, cross-species movement, and adaptation may be driven by recombination events in nature (Lau et al., 2015), and that CoV recombination can result in increased virulence, pathogenicity, and potential pandemic spread.

Experimental recombination has been most studied in the model β-CoV murine hepatitis virus (MHV). During mixed infections of related MHV strains, ~25% of progeny virions are generated by recombination (Keck et al., 1988; Kottier et al., 1995; Lai et al., 1985; Makino et al., 1986). During normal replication, the putative CoV replication-transcription complex (RTC) performs discontinuous transcription at virus-specific transcription regulatory sequences (TRSs) (Figure 1A, 3A) to generate a set of subgenomic mRNAs (sgmRNAs) with common 5’ leader sequences and 3’ ends, which

70 are subsequently translated into structural and accessory proteins (Dufour et al., 2011;
71 Kirchdoerfer and Ward, 2019; Subissi et al., 2014; Weiss et al., 1994). CoV replication
72 also generates defective viral genomes (DVGs) that contain multiple deletions of
73 genomic sequence. DVGs range in size from <1kb to >20kb, but retain intact 5' and 3'
74 genomic untranslated regions (5' and 3' UTRs) and are amplified by RTC machinery
75 supplied by co-infecting full-length helper CoVs (Brian and Spaan, 1997; Makino et al.,
76 1985; Schaad and Baric, 1994; Wu and Brian, 2010). Therefore, CoVs perform
77 recombination as a normal part of their replication, producing complex populations of
78 recombined RNA molecules. Prior to the advent of next generation sequencing, direct
79 analysis of recombined CoV RNAs was not possible and the determinants of
80 recombination could not be identified.
81
82 In other RNA virus families including picornaviruses and alphaviruses, regulation of
83 recombination has been mapped to replication fidelity determinants in the viral RNA-
84 dependent RNA polymerase (RdRp) (Kempf et al., 2016; Li et al., 2019; Poirier et al.,
85 2015; Woodman et al., 2018). In contrast to these viruses, CoV replication fidelity is
86 primarily determined by the 3'-5' exoribonuclease encoded in nonstructural protein 14
87 (nsp14-ExoN) that proofreads RNA during replication through excision of mismatched
88 incorporated nucleotides (Ferron et al., 2018a; Ma et al., 2015). Viral exonucleases are
89 essential for recombination in DNA viruses, including vaccinia virus (Gammon and
90 Evans, 2009) and herpes simplex virus 1 (Grady et al., 2017; Schumacher et al., 2012).
91 In contrast, the role of the CoV nsp14-ExoN in RNA recombination has not been
92 defined. Catalytic inactivation of nsp14-ExoN resulted in qualitatively reduced

93 abundance of MHV sgmRNA2 (Eckerle et al., 2007) and altered human CoV 229E
94 (HCoV-229E) sgmRNA detection during viral recovery (Minskaia et al., 2006). These
95 studies suggest that nsp14-ExoN RNA proofreading activity of CoVs may play a key
96 role in RNA recombination in addition to known functions in replication fidelity, viral
97 fitness and virulence, resistance to nucleoside analogs, and immune antagonism (Case
98 et al., 2017; Ferron et al., 2018a; Graepel et al., 2019).
99
100 In this study, we sought to define the frequency and patterns of recombination of
101 divergent β-CoVs SARS-CoV-2, MERS-CoV, and MHV. We show that all three viruses
102 perform extensive recombination during replication *in vitro,* with broadly similar patterns
103 of recombination, generating diverse populations of recombined molecules. We further
104 demonstrate that genetically engineered inactivation of MHV nsp14-ExoN activity
105 results in a significant decrease in recombination frequency, altered recombination
106 junction patterns across the genome, and altered junction site selection. These defects
107 and alterations result in a marked change in MHV-ExoN(-) recombined RNA
108 populations, including defective viral genomes (DVGs). Previous studies our lab and
109 others confirm the requirement of ExoN activity in high-fidelity replication, exclusion of
110 aberrant nucleotides, viral fitness, and virulence (Eckerle et al., 2007, 2007; Ferron et
111 al., 2018a; Graham et al., 2012). Combined with the multiple critical integrated functions
112 of nsp14-ExoN, the demonstration in this study that nsp14-ExoN activity is required for
113 WT-like recombination further defines it as an exceptionally conserved, vulnerable, and
114 highly specific target for inhibition by antiviral treatments and viral attenuation. These
115 results also support future studies aimed at illuminating the role of SARS-CoV-2 nsp14-

ExoN activity in RNA recombination, the regulation of sgmRNA expression, and its contribution to novel CoV zoonotic emergence.

RESULTS

SARS-CoV-2 and MERS-CoV undergo extensive RNA recombination to generate populations of recombination junctions. We first sought to quantify recombination frequency and identify recombination patterns in zoonotic CoVs by sequencing both MERS-CoV and SARS-CoV-2 RNA.

Vero CCL-81 cells were infected with MERS-CoV derived from an infectious clone at passage 1 with an MOI of 0.3 pfu/mL for 72 hours until >70% cytopathic effect was observed. Vero E6 cells were infected with SARS-CoV-2 derived from a patient isolate passaged in culture (passage 5) at an MOI of 0.45 pfu/mL for 60 hours until >70% cytopathic effect was observed. Total RNA from infected cells was isolated and poly(A)-selected to capture all genomic and subgenomic RNA, as well as defective viral genomes (DVGs). Equal amounts of total cell RNA from both viruses was sequenced by short-read Illumina RNA-sequencing (RNA-seq) and long-read direct RNA Nanopore sequencing. The depth and low error rate of RNA-seq facilitated the quantification and detection of both high- and low-abundance unique junctions, but did not allow for detection of junctions in the context of a full-length RNA molecule. Long-read direct RNA sequencing on the Oxford Nanopore Technologies MinION platform was used to sequence complete RNA molecules. By combining both short- and long-read RNA sequencing, we accomplished high-confidence quantification and detection of

recombination junctions as well as description of the genetic architectures of molecules formed by the junctions.

For RNA-seq, we generated libraries in triplicate from 2 μg samples of both MERS-CoV and SARS-CoV-2 infected cells. Reads were aligned to the respective viral genomes using a recombination-aware mapper, *ViReMa* (*Virus Re*combination *Ma*pper) (Routh and Johnson, 2014). *ViReMa* detects non-homologous recombination events resulting in subgenomic mRNAs (sgmRNAs) and defective viral genomes (DVGs) by identifying recombination junctions that generate a deletion greater than 5 base-pairs flanked on both sides by a 25 base-pair alignment. *ViReMa* aligned both recombined and non-recombined reads in the library and reported the total number of nucleotides aligned to the genome and all detected recombination junctions (Figure S1A-B).

MERS-CoV RNA contained an average of 19,367 detected junctions per dataset and SARS-CoV-2 libraries had an average of 56,082 detected junctions per dataset. To normalize for variations in library size between samples, recombination junction frequency (J_{freq}) was calculated for MERS-CoV and SARS-CoV-2 (Figure 1B). J_{freq} refers to the number of nucleotides in all detected junctions normalized to total mapped nucleotides in a library. MERS-CoV had a mean J_{freq} of 37.80 junctions detected per 10,000 mapped nucleotides. SARS-CoV-2 had a mean J_{freq} of 475.7 junctions per 10,000 mapped nucleotides (Figure 1B). To define the patterns of these detected recombination junctions, we mapped forward (5' → 3') recombination junctions according to their genomic position (Figure 1C-D, Figure S1C-D). Both MERS-CoV and

SARS-CoV-2 displayed clusters of junctions between the 5' and 3' ends of the genome, intermediate genomic positions and the 3' end of the genome, and within the 3' end of the genome (Figure 1B-C). SARS-CoV-2 also had many low-frequency junctions distributed across the genome, a major cluster of junctions representing local deletions and between the 5' UTR and the rest of the genome, and horizontal clusters of low-frequency junctions between common start sites at position ~2000 and ~8000 and the rest of the genome (Figure 1D). Overall, these data demonstrate that extensive RNA recombination in both MERS-CoV and SARS-CoV-2 generates diverse populations of junctions with similar high-abundance clusters.

To determine whether high-frequency recombination occurs randomly across MERS-CoV and SARS-CoV-2 genomes, we calculated mean positional recombination frequency (Figure 1E-F) by comparing the number of nucleotides in detected junctions (both start and stop sites) at that position normalized to nucleotide depth at that position. The first and last 5000 nucleotides of both MERS-CoV and SARS-CoV-2 genomes showed clusters of recombination (Figure 1E-F). SARS-CoV-2 had many peaks of recombination frequency of greater than 0.50 (50%) across the genome (Figure 1F), suggesting that SARS-CoV-2 generates a diverse set of RNA molecules with distinct genetic architectures through recombination.

Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 defines the architecture of full-length genome, subgenomic RNAs, and defective viral genomes.

We performed direct RNA Nanopore sequencing on the same RNA used for short-read RNA-seq. For both viruses, three independent sequencing experiments were performed on triplicate samples of 2 μg RNA for each virus. In order to remove prematurely truncated sequences, we bioinformatically filtered for Nanopore reads containing both genomic termini. Capture and sequencing of sgmRNAs and DVGs resulted in high sequencing coverage at both ends of the genome, while the relatively poorer capture of full-length genomes was limited due to their extreme length and resulting in depleted coverage in the middle of the genome (Figure 2A-B). We sequenced 132,493 MERS-CoV RNA molecules and 1,725,862 SARS-CoV-2 RNA molecules that had 85.6% and 82.2% identity to the parental genome, respectively (Table S1). We obtained 3 full-length direct RNA sequences of the SARS-CoV-2 genome containing over 29,850 consecutive nucleotides that aligned to the SARS-CoV-2 genome (Table S2). In MERS-CoV RNA, we detected 451 unique junction-containing full-length molecules (Figure 2A, Table S3). SARS-CoV-2 RNA contained distinct 172,191 junction-containing molecules (Figure 2B, Table S2).

To define the architectures of detected molecules, we filtered for junctions with at least 3 supporting Nanopore reads. In MERS-CoV, we defined 5 distinct species, including 3 sgmRNAs (6, 7, and 8) and 2 DVGs (Figure 2C). In SARS-CoV-2, there were 1166 isoforms with a single junction and 227 containing 2 junctions. Because the TRS sequences and sgmRNA species have not previously been experimentally confirmed in SARS-CoV-2, a Nanopore read was categorized as a predicted sgmRNA transcript if it had a junction between the predicted 5' TRS-L and the predicted structural and

accessory open reading frames to include both canonical and non-canonical sgmRNAs. The 15 most abundant isoforms in SARS-CoV-2 included 7 predicted sgmRNA transcripts and 8 DVGs (Figure 2D). We also identified potential alternative transcripts corresponding to a single gene (ORF8) (Figure 2D). In summary, direct RNA Nanopore sequencing defined a diverse set of recombined RNAs generated by both MERS-CoV and SARS-CoV-2. The results confirmed and extended the short-read RNA-seq datasets. Our results are likely an underestimate of total numbers and diversity based on the conservative bioinformatic limitation of analysis to isoforms with at least 3 supporting reads. Thus, both MERS-CoV and SARS-CoV-2 engage in extensive RNA recombination during replication in cell culture, producing diverse junctions across the viral genomes and many recombined RNA isoforms, with the indication that SARS-CoV-2 recombination exceeds that of MERS-CoV in Vero cells. These findings underline the importance of defining the determinants of CoV recombination.

The β-coronavirus murine hepatitis virus (MHV) lacking nsp14-ExoN activity has significantly decreased and altered distribution of RNA recombination events. In picornaviruses and alphaviruses, recombination is regulated by determinants of replication fidelity (Kempf et al., 2016; Li et al., 2019; Poirier et al., 2015; Woodman et al., 2018). CoVs encode a 3'-5' exoribonuclease (ExoN) in non-structural protein 14 (nsp14) that regulates replication fidelity by RNA proofreading (Eckerle et al., 2007; Ferron et al., 2018b; Minskaia et al., 2006). No proofreading-deficient nsp14-ExoN catalytic mutant has been rescued in MERS-CoV or SARS-CoV-2. We have rescued and extensively studied ExoN catalytic inactivation mutants (ExoN(-)) in β-CoVs murine

hepatitis virus (MHV) and SARS-CoV. We used MHV and compared an engineered ExoN catalytic inactivation mutant (MHV-ExoN(-)) to wild-type virus (MHV-WT). Murine delayed brain tumor cells (DBT-9) were infected in triplicate with either MHV-WT or MHV-ExoN(-) at an MOI of 0.01 pfu/mL and total infected cell monolayer RNA was extracted when the monolayer was 95% involved in cytopathic effect and >75% intact. To isolate virions, DBT-9 cell were infected in triplicate with either MHV-WT or MHV-ExoN(-) at an MOI of 0.01 pfu/mL and virions were isolated by ultracentrifugation from supernatant collected when the monolayer was 95% involved in cytopathic effect and >90% intact. RNA-seq datasets were aligned to the MHV genome using *ViReMa*, reporting the genomic coverage in infected cells (Figure S2A-B) and virions (Figure S3A-B). To determine whether loss of nsp14-ExoN activity alters recombination, we quantified recombination junction frequency (J_{freq}) in MHV-WT and MHV-ExoN(-). We found that MHV-ExoN(-) had significantly decreased J_{freq} relative to MHV-WT in both infected cells (Figure 3B) and virions (Figure 3D).

We next tested whether loss of nsp14-ExoN activity altered the population of junction sites. Recombination junctions detected by *ViReMa* were plotted according to their start (5') and stop (3') sites in infected cell RNA (Figure 3C, Figure S2C-D) and virion RNA (Figure 3E, Figure S3C-D). MHV-WT displayed clusters of junctions (i) between the 5' and 3' ends of the genome, (ii) between intermediate genomic positions and the 3' end of the genome, (iii) between the 5' UTR and the rest of the genome, (iv) in local deletions across the genome, and (v) within the 3' end of the genome (Figure 3C and 3E). Similarly, MHV-ExoN(-) accumulated junction clusters between the 5' and 3' ends

of the genome and within the 3' end of the genome (Figure 3C and 3E). However, MHV-ExoN(-) had fewer junctions between the 5' UTR and the rest of the genome and fewer junctions forming local deletions (Figure 3C and 3E). Together, these findings show that loss of MHV nsp14-ExoN activity resulted in decreased recombination frequency and altered junction patterns across the genome.

Loss of nsp14-ExoN activity alters recombination in distinct genomic regions. To determine whether the engineered defect in nsp14-ExoN alters recombination across the genome, we calculated and compared mean positional recombination frequencies of MHV-WT (blue) and MHV-ExoN(-) (orange) in both infected cell monolayers (Figure 4A) and virions (Figure 4B). MHV-WT has high recombination frequency at the 5' and 3' ends of the genome as well as at distinct sites across the genome. MHV-ExoN(-) has a similar distribution in both infected cells and virus particles (Figure 4A-B). There were 765 positions in infected cell RNA and 499 positions in virion RNA (Table S4) with significantly altered recombination frequency in MHV-ExoN(-) compared to MHV-WT. These differences are best illustrated when the distinct genomic regions are separately visualized (Figure S4).

In the 5' UTR, nucleotides within the TRS-leader (TRS-L, yellow) participate in recombination junctions in both MHV-WT and MHV-ExoN(-) as expected, although MHV-ExoN(-) has significantly decreased recombination frequency at all sites within the TRS-L (Figure S4A, Table S4). Both MHV-WT and MHV-ExoN(-) display recombination across the nonstructural proteins, although MHV-ExoN(-) demonstrates overall

277 decreased recombination in these regions compared with MHV-WT (Figure S4B-C,
278 Table S4). In virions, MHV-ExoN(-) has significantly increased recombination in nsp3,
279 nsp4, and two positions in nsp12 (Figure S4B-C, Table S4). In addition, MHV-ExoN(-)
280 has significantly increased recombination at one position within the 3' untranslated
281 region (3'-UTR) (Figure S4E, Table S4).
282
283 **MHV-ExoN(-) has decreased subgenomic mRNA populations and increased**
284 **abundance of defective viral genomes.** We next determined whether loss of nsp14-
285 ExoN activity influences sgmRNA patterns and abundance, as proposed in earlier
286 studies (Eckerle et al., 2007; Minskaia et al., 2006). We quantified recombination
287 junction frequencies (J_{freq}) of each of the known junctions formed between the common
288 TRS-L and 6 sgmRNA-specific body TRSs. sgmRNA J_{freq} refers to the number of
289 nucleotides corresponding to the sgmRNA leader-body junction normalized to mapped
290 library size. sgmRNAs 3, 4, 5, 6, and 7 were significantly decreased in MHV-ExoN(-)
291 infected cell RNA (Figure 4C). sgmRNA 4 and 6 were present at significantly decreased
292 frequencies in MHV-ExoN(-) virions compared to MHV-WT (Figure 4D). Thus, MHV-
293 ExoN(-) has altered frequency of multiple sgmRNAs both in infected cells and in virus
294 particles.
295
296 We further defined alterations to different populations of recombined RNA molecules in
297 MHV-ExoN(-) compared to MHV-WT by quantifying the percentages of sgmRNA and
298 DVG junctions in the total population of detected junctions. Any junction that did not
299 form a canonical leader-body TRS junction was considered a DVG junction, likely

300 including a subset of previously undefined non-canonical or alternative sgmRNA
301 junctions. The percentage of sgmRNA junctions were significantly decreased and the
302 percentage of DVG junctions significantly increased in both infected cell RNA (Figure
303 4E) and virion RNA (Figure 4F). These findings demonstrate that loss of nsp14-ExoN
304 activity alters the locations at which the replicase machinery recombines across the
305 genome, resulting in both decreased and altered sgmRNAs and relatively increased
306 percentage of DVGs in MHV-ExoN(-) compared to MHV-WT. Thus, catalytic inactivation
307 of nsp14-ExoN results in a defect in recombination site selection.
308
309 **Differential abundance of recombination junctions in MHV-ExoN(-) reveals altered**
310 **recombination junction site selection.** We next defined the distribution of
311 recombination junctions with altered abundance in MHV-ExoN(-) compared to MHV-WT.
312 We quantified the abundance junctions identified over three independent experiments
313 and identified junction populations with altered abundances in MHV-ExoN(-) compared
314 to MHV-WT using *DESeq2* (Love et al., 2014). MHV-ExoN(-) RNA contained distinct
315 junction populations with both significantly increased and decreased abundance relative
316 to MHV-WT in infected cells (Figure S5A, Table S5) and in virus particles (Figure S5B,
317 Table S5). We mapped significantly altered junctions by their genomic position (Figure
318 4G-H). Recombination junctions enriched in MHV-ExoN(-) were mainly found between
319 the 5' and 3' ends of the genome (Figure 4G-H). This observation suggests that MHV-
320 ExoN(-) is accumulating altered populations of DVGs that are formed by these
321 deletions.
322

Junctions with decreased abundance in MHV-ExoN(-) compared to MHV-WT in both infected cells (Figure 4G) and virions (Figures 4H) mainly clustered between the 5' UTR and the rest of the genome. Further, infected cell RNA demonstrates a cluster of junctions with decreased abundance in MHV-ExoN(-) in the 3' end of the genome that represent local deletions in this region (Figure 4G). This finding was complemented by the cluster of decreased abundance of local deletions of 10 – 50 bp in length across the genome in MHV-ExoN(-) virion RNA (Figure 4H). Together, these findings demonstrate that the populations of recombination junctions that were differentially abundant in MHV-ExoN(-) were not randomly distributed across the genome. This suggests that loss of nsp14-ExoN activity results in altered recombination site selection in distinct regions of the genome.

Direct RNA Nanopore sequencing identifies defects in MHV-ExoN(-) full-length recombined RNA populations. To test the alterations of full-length recombined RNAs due to loss of nsp14-ExoN proofreading activity, we sequenced MHV-WT and MHV-ExoN(-) virion RNA by direct RNA Nanopore sequencing. Reads were mapped to the MHV genome using minimap2 and mapped reads containing both genomic termini were identified. MHV-WT virion RNA contained 101,714 complete molecules and MHV-ExoN(-) virion RNA yielded 19,334 (Table S1). Reads mapping to the full-length parental genome were not observed, likely due to the length (31.4 kb) of the MHV genome, and datasets yielded only sgmRNAs and DVGs. MHV-ExoN(-) consistently demonstrated ~1 $\log_{10}$ reduction in nucleotide depth at positions across the genome relative to MHV-WT (Figure 5A). This finding directly corroborated the decreased

346 recombination junction frequency of MHV-ExoN(-) compared to MHV-WT demonstrated
347 in short-read Illumina RNA-seq datasets (Figure 3B).
348
349 Consistent with fewer full-length molecules sequenced, MHV-ExoN(-) had a global
350 decrease in the number of junctions across the genome (Figure 5B). Because
351 molecules with multiple junctions in a single molecule were detected but not supported
352 by more than one read, they were excluded from downstream analyses. We next
353 mapped junctions detected in full-length RNA molecules according to their genomic
354 position (Figure 5C). While MHV-WT contains 852 unique junctions present in full-length
355 molecules, MHV-ExoN(-) had only 43 (Table S3). MHV-WT recombination junctions
356 exclusively connect positions in the first ~8,000 nucleotides of the genome to the 3' UTR
357 (Figure 5C). MHV-ExoN(-) had fewer junction sites, none of which represented distinct
358 clusters (Figure 5D). We next grouped junctions within 10-base pair windows to define
359 distinct families of molecules supported by at least three reads. Nine such architectures
360 were identified in MHV-WT (Figure 5E). These populations contain both DVGs and
361 sgmRNAs. The four most abundant isoforms were also detected in MHV-ExoN(-) virion
362 RNA, which corresponded to a DVG and sgmRNAs 4,6 and 7 (Figure 5E). We did not
363 detect MHV-ExoN(-)-unique variants with at least 3 supporting reads, potentially due to
364 their low frequency in the population and the limited depth of Nanopore sequencing.
365 Additionally, all junctions generating RNA molecules found in MHV-WT and MHV-
366 ExoN(-) were present in short-read RNA-seq datasets. These data demonstrate that
367 loss of nsp14-ExoN activity drives the accumulation altered recombined RNA
368 populations and DVG species diversity.

DISCUSSION

In this study, we define the diversity of the CoV recombination landscape in MHV, MERS-CoV, and SARS-CoV-2. We also demonstrate that the loss of nsp14-ExoN activity in MHV results in decreased recombination and alters site selection of recombination junctions. Our results support a model in which nsp14-ExoN activity functions during replication to drive normal recombination, leading to the generation of populations of discrete molecules, including sgmRNAs and DVGs.

MERS-CoV and SARS-CoV-2 have extensive recombination networks producing diverse populations of RNA species. We show that both MERS-CoV and SARS-CoV-2, the etiological agent of the ongoing global pandemic of COVID-19, performs extensive recombination and generates diverse populations of RNA molecules. SARS-CoV-2 demonstrated relatively increased recombination frequency compared to MERS-CoV and displayed unique clusters of recombination junctions across the SARS-CoV-2 genome. We analyzed SARS-CoV-2 propagated from a clinical isolate that had been passaged in cell culture, while MERS-CoV was generated from an isogenic clone of low passage. It is possible that the differences could be the result of the diversity and composition of the original sample or propagation in different cell types. Alternatively, these differences may be due to alterations in transcription and RNA synthesis kinetics in SARS-CoV-2 compared to MERS-CoV, which have not been previously defined. It will be important for future studies to determine the role of the diversity of the viral population, cell environment, and virus-specific RNA synthesis kinetics in viral

recombination. Nevertheless, both viruses perform extensive, high-frequency recombination and generate diverse species of recombined RNAs.

Direct RNA Nanopore sequencing on the Oxford Nanopore Technologies MinION platform has sequenced sgmRNAs and DVGs in the alpha-coronavirus human coronavirus 229E (HCoV-229E) (Viehweger et al., 2019) and has been utilized to identify putative sgmRNAs in SARS-CoV-2 (Kim et al., 2020; Taiaroa et al., 2020). Direct RNA Nanopore sequencing further revealed the profound diversity of RNA species produced by SARS-CoV-2 during infection and yielded the first reported full-length, continuous direct RNA reads of SARS-CoV-2. The full-length genome sequences are currently the longest and most complete direct RNA Nanopore sequences reported for any CoV. Detection of multiple RNA molecules with similar but distinct genetic architectures corresponding to a single open reading frame suggests that SARS-CoV-2 produces alternative transcripts. Tandem genetic and proteomic studies of sgmRNA species in SARS-CoV-2 will be valuable to experimentally define populations of translated SARS-CoV-2 sgmRNAs their role in viral replication.

Murine hepatitis virus (MHV) incorporates diverse recombinant RNA in its virions. Comparison of recombination in MHV-infected cells and virus particles (virions) revealed that the observed recombination patterns and species generated in infected cells are largely retained in virions. Specifically, we observed three previously unexpected recombination junction clusters: (1) junctions connecting the canonical 5' TRS-L to non-canonical positions across the genome; (2) junctions representing small

415 local deletions distributed evenly across the genome; and (3) local recombination within
416 the 3' end of the genome, including structural proteins and the 3' UTR. This result was
417 surprising based on earlier studies reporting the requirement of specific packaging
418 signals for incorporation of select RNA species into the virions (Kuo and Masters, 2013).
419 Our direct RNA Nanopore sequencing studies show that MHV virions contain a diverse
420 population of recombined RNA molecules which exclude the canonical MHV packaging
421 signal at position 20,273 – 20,367 (Kuo and Masters, 2013). Thus, MHV may have
422 much lower barriers for packaging a diverse population of RNA species than previously
423 known, which may aid in the transmission and accumulation of DVGs even under low-
424 MOI conditions.
425
426 **MHV lacking nsp14-ExoN activity has globally decreased and altered**
427 **recombination.** MHV-ExoN(-) demonstrated a global decrease in recombination
428 junction frequency in both infected cells and virions while also exhibiting skewed
429 populations of sgmRNAs, an increased proportion of DVGs, and altered populations of
430 recombinant isoforms. Our findings demonstrate a novel role for nsp14-ExoN in CoV
431 RNA recombination and further suggest that loss of nsp14-ExoN activity results in
432 altered recombination site selection. Decreased recombination frequency at canonical
433 TRSs in MHV-ExoN(-) suggest that there is decreased accumulation of sgmRNAs. This
434 finding is supported by previous studies (Eckerle et al., 2007; Minskaia et al., 2006) that
435 report a defect in MHV-ExoN(-) sgmRNA synthesis. Perturbation of the abundance of
436 each sgmRNA will likely have a profound influence on the fitness and virulence of CoVs
437 as sgmRNAs encode critical viral proteins including nucleocapsid (N), spike (S), and

438 many accessory ORFs that impact replication efficiency and counteract host innate
439 immune responses (Cruz et al., 2011; Muth et al., 2018).
440
441 MHV-ExoN(-) virion RNA demonstrates a significant increase in recombination in five
442 distinct genomic regions: 1) nsp3, 2) nsp4, 3) nsp12, 4) the spike (S) protein, and 5) the
443 3' UTR. Previous studies show that CoV DVGs frequently contain sequences from the
444 first replicase proteins after the 5' UTR (Kim et al.; Makino; Makino et al., 1984, 1988;
445 Penzes et al., 1996). Increased recombination in nsp3 and nsp4 may be the result of
446 altered DVG formation and increased accumulation in virus particles. Some of the
447 detected DVGs may be non-canonical sgmRNAs, containing junctions near the TRSs
448 and potentially encode novel translation reading frames (Di et al., 2017; Viehweger et
449 al., 2019).
450
451 **Loss of MHV nsp14-ExoN activity results in altered recombination site selection.**
452 Differential abundance analysis, mapping of junction detected in multiple experiments,
453 and direct RNA Nanopore sequencing suggest that defective nsp14-ExoN activity drives
454 altered recombination site selection resulting in different populations of accumulated
455 molecules. In this model, abrogation of nsp14-ExoN activity would cause altered
456 replicase activity, resulting in selection of aberrant sites, potentially based on sequence
457 or structural elements. In both picornaviruses and alphaviruses, low fidelity mutant
458 viruses have altered polymerase speed and processivity (Campagnola et al., 2015) and
459 these properties contribute to recombination and the generation of defective interfering
460 particles (Kim et al., 2019; Langsjoen et al., 2020; Poirier et al., 2015). The MHV-ExoN(-

461) replicase may have altered protein-protein interactions that change the stability of the
462 complex and drive altered site selection. Alternatively, loss of nsp14-ExoN activity may
463 result in defective replicase-RNA interactions and therefore cause altered RNA
464 recombination site selection. In both models, nsp14-ExoN activity is an important
465 determinant of replicase activity during RNA synthesis and recombination. However,
466 there are likely other determinants of recombination in the CoV replicase that will be
467 important to define through future genetic and structure-function studies.
468
469 In this study, we identify key positions across the genome that are involved in
470 recombination junctions in the model β-CoV, MHV and we establish nsp14-ExoN
471 proofreading activity as a key determinant of recombination in β-CoVs. Our previous
472 studies of MHV nsp14-ExoN findings genetics and functions have been directly
473 reproduced in SARS-CoV (Agostini et al., 2018; Eckerle et al., 2010; Smith et al., 2013).
474 Thus, we would predict that SARS-CoV-2, will demonstrate the requirement of nsp14-
475 ExoN activity for recombination. Further, treatment of SARS-CoV-2 infection with small
476 molecules that target the replicase proteins including nsp14-ExoN may result in an
477 altered recombination landscape and perturb viral fitness. It will be important for future
478 studies to assess the role of RNA proofreading, replication fidelity, and small molecule
479 antivirals in recombination of SARS-CoV-2. Understanding the relationship between
480 protein determinants in RNA recombination will provide new targets for countermeasure
481 development to combat emergence of novel CoVs from animal reservoirs with
482 pandemic potential.
483

ACKNOWLEDGMENTS

We thank members of the Denison laboratory for insightful discussions regarding this study. This work was supported by National Institutes of Health grants R01AI108197 (M.R.D.), T32GM065086 (J.G.), F31AI133952 (M.L.A). We are grateful for support from the Dolly Parton COVID-19 Research Fund, the Vanderbilt University Office of Research, and the Elizabeth B. Lamb Center for Pediatric Research at Vanderbilt University.

AUTHOR CONTRIBUTIONS

Conceptualization (M.R.D, J.D.C., J.G.); supervision (M.R.D., A.L.R.), project administration (M.R.D., A.L.R.), funding acquisition (M.R.D., J.D.C., A.J.P.), MHV infections and deep sequencing of MHV, MERS-CoV, and SARS-CoV (J.G.), MERS-CoV infections (L.J.S.), SARS-CoV-2 infections (L.J.S), SARS-CoV-2 RNA isolation (X.L.), Nanopore sequencing (J.G.), bioinformatic analysis and visualization (J.G.), software development (J.G., A.L.R.), data curation (J.G.), writing (J.G.), review and editing (all authors).

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Genome-wide recombination generates populations of diverse RNA molecules in MERS-CoV and SARS-CoV-2. (A) Genome organization of MERS-CoV (violet) and SARS-CoV-2 (gray). Nonstructural (nsps 1 – 16) and structural (S, E, M, N) and accessory open reading frames (ORFs) are labelled. The common 5' leader transcription leader sequence (TRS-L) is denoted with an unfilled red star. Body TRSs are labelled with filled red stars. MERS-CoV total cell lysates (black) and SARS-CoV-2 infected cell monolayers (violet) were sequenced by RNA-seq. (B) Junction frequency (J_{freq}) was calculated by comparing the number of nucleotides in *ViReMa*-detected junctions to all mapped nucleotides. Error bars represent standard errors of the mean for three independent sequencing libraries (N = 3). Recombination junctions are mapped according to their genomic position (5' junction site, Start Position; 3' junction site, Stop Position) and colored according to their frequency in the population of all junctions in MERS-CoV (C) and SARS-CoV-2 (D). The highest frequency junctions are magenta and completely opaque. The lowest frequency junctions are red and the most transparent. Dashed boxes represent clusters of junctions: (i) 5' → 3'; (ii) mid-genome → 3' UTR; (iii) 3' → 3'; (iv) local deletions; (v) 5' UTR → rest of genome. Recombination frequency is quantified across the MERS-CoV (E) and SARS-CoV-2 (F) genomes. Recombination frequency is represented as a mean of three independent sequencing libraries (N = 3). See also Figure S1.

Figure 2. Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 reveals accumulation of distinct recombined RNA populations. Direct RNA Nanopore sequencing of poly-adenylated MERS-CoV total cell lysates and SARS-CoV-

2 infected cell monolayer RNA. Three sequencing experiments were performed for each virus. Nanopore reads passing quality control were combined and mapped to the viral genome using *minimap2*. Full-length reads containing the genomic termini were identified by bioinformatic filtering. Genome coverage maps and Sashimi plots visualizing junctions (arcs) in full-length (A) MERS-CoV (black) and (B) SARS-CoV-2 (violet) RNA reads. (C) Distinct RNA molecules identified in MERS-CoV (black) with at least 3 supporting reads are visualized. The number of sequenced reads containing the junction is listed (Count). Genetic sequences of each RNA molecule are represented by filled boxes and deleted regions are noted (Deleted Region(s)) and represented by dashed lines. Molecules containing junctions between the MERS-CoV TRS sequences are listed as sgmRNAs. (D) The 15 most abundant SARS-CoV-2 (violet) recombined RNA molecules and 3 full-genome reads are visualized. Molecules containing junctions linking the 5' UTR and the predicted open reading frames of structural and accessory proteins are listed as predicted sgmRNAs. See also Table S1, Table S2, and Table S3.

Figure 3. Loss of nsp14-ExoN activity decreases recombination frequency and alters recombination junction patterns across the genome. (A) Genome organization of MHV. Nonstructural (nsps 1 – 16) and structural (S, E, M, N) and accessory open reading frames (ORFs) are labelled. The common 5' leader transcription leader sequence (TRS-L) is denoted with an unfilled red star. Body TRSs are labelled with filled red stars. Infected monolayer and virion RNA from independent experiments were poly(A) selected, sequenced by RNA-seq, and aligned to the MHV genome using *ViReMa*. Junction frequency (J_{freq}) in infected monolayer RNA (B) and

552 virion RNA (D) was calculated by dividing the number of sequenced nucleotides in all
553 junctions by the total number of nucleotides sequenced in the library. Error bars
554 represent standard error of the means (SEM) (N = 3). Statistical significance was
555 determined by the unpaired student's t-test. *, $p < 0.05$. Unique forward (5' → 3')
556 recombination junctions detected in infected monolayers (C) and virions (E) were
557 mapped in MHV-WT and MHV-ExoN(-) according to their genomic position. Junctions
558 are colored according to their frequency in the population (high frequency = magenta;
559 low frequency = red). Clusters are marked by dashed boxes: (i) 5' → 3'; (ii) mid-genome
560 → 3'; (iii) 3' → 3'; (iv) local deletions; (v) 5' UTR → rest of genome. See also Figure S2,
561 Figure S3.
562
563 **Figure 4. Loss of nsp14-ExoN alters recombination at multiple genomic loci and**
564 **skews recombined RNA populations.** Mean recombination frequency at each position
565 across the MHV genome was compared in MHV-WT (blue) and MHV-ExoN(-) (orange)
566 infected monolayer (A) and virion RNA (B). Positions with significantly altered
567 recombination frequencies were identified by a 2-way ANOVA with multiple
568 comparisons (N=3). The junction frequencies (J_{freq}) of each sgmRNA were compared in
569 MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (C) and virions (D).
570 Error bars represent standard errors of the mean (SEM) (N = 3) and statistical
571 significance was determined by a 2-way ANOVA with multiple comparisons, * $p < 0.05$,
572 **** $p < 0.0001$. (D) The percentages of sgmRNA and DVG junctions were compared in
573 MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (E) and virions (F).
574 Error bars represent SEM (N = 3). Statistical significance was determined by a 2-way

575 ANOVA with multiple comparisons, ** $p < 0.01$, *** $p < 0.001$. The abundance of
576 junctions in MHV-ExoN(-) was compared to MHV-WT in infected monolayers (G) and
577 virions (H) by *DESeq2*. Junctions with statistically significant altered abundance ($p <$
578 0.05, N = 3) in MHV-ExoN(-) are mapped across the genome and colored according to
579 their fold-change (red squares = decreased abundance, green circles = increased
580 abundance). See also Figures S4 and S5B, Tables S4-S5.
581
582 **Figure 5. Direct RNA Nanopore sequencing of MHV full-length recombined RNA**
583 **molecules.** Direct RNA Nanopore sequencing of MHV virion RNA. Full-length reads
584 were identified by bioinformatic filtering for reads containing the genomic termini. (A)
585 Genome coverage maps of full-length MHV-WT (blue) and MHV-ExoN(-) (orange)
586 Nanopore reads aligned to the MHV-A59 genome using *minimap2*. (B) Sashimi plot
587 visualizing junctions (arcs) in MHV-WT (blue) and MHV-ExoN(-) (orange). (C) Junctions
588 in reads containing only 2 discontinuous regions (1 junction) are mapped according to
589 genome position. The junction 5' site is mapped as the Start Position on the y-axis and
590 the 3' site is mapped as the Stop Position on the x-axis. Junctions are colored according
591 to their frequency in the population ($\log_{10}$(Frequency)). MHV-WT (low = blue, high =
592 cyan) and MHV-ExoN(-) (low = orange, high = yellow). (D) RNA molecule genetic
593 architectures with at least 3 supporting reads identified in both MHV-WT and MHV-
594 ExoN(-) (yellow) and unique to MHV-WT (blue). There were no reads unique to MHV-
595 ExoN(-) supported by at least 3 reads. Genetic sequences of the RNA molecule are
596 represented by filled boxes. Deleted regions are reported (Deleted Region) and

represented by dashed lined. Canonical sgmRNA species are noted. The number of reads supporting each species are noted (Count). See also Table S1 and Table S2.

RESOURCE AVAILABILITY

Lead Contact. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mark R. Denison (mark.denison@vumc.org).

Materials Availability. This study did not generate new unique reagents.

Data and Code Availability. The datasets generated during this study are available at the Sequence Read Archive (SRA) under BioProject accession numbers PRJNA623001, PRNJA623016, PRJNA623285, PRJNA623325, PRJNA623312, PRNJA623282, PRJNA623323, PRJNA623314, PRJNA623580, PRJNA623578. The in-house scripts utilized in this study are publicly available at https://github.com/DenisonLabVU/rna-seq-pipeline.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines. DBT-9 (delayed brain tumor, murine astrocytoma clone 9) cells were maintained at 37°C as described previously (Chen and Baric, 1996). DBT-9 cells were originally obtained from Ralph Baric at University of North Carolina-Chapel Hill and were maintained within 50 passages of this progenitor stock. Cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal clone serum (FCS) (Invitrogen), 100 U/mL penicillin and streptomycin (Gibco), and 0.25 μg/mL amphotericin B (Corning). *Cercopithecus aethiops* Vero CCL-81 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final

concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for MERS-CoV-2 infection. Vero CCL-81 cells were obtained from ATCC. Vero E6 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for SARS-CoV-2 infections. Vero E6 cells were obtained from ATCC.

METHOD DETAILS

Viruses. All MHV work was performed using the recombinant WT strain MHV-A59 (GenBank accession number AY910861.1 (Yount et al., 2002)) at passage 4 and an engineered ExoN(-) strain of MHV-A59 at passage 2. Experiments involving MERS-CoV were conducted using the human EMC/2012 strain recovered from an infectious clone (GenBank accession number JX869059.2) (Scobey et al., 2013). Experiments involving SARS-CoV-2 were conducted with a passage 5 virus inoculum generated from a Seattle, WA, USA COVID-19 patient (GenBank accession number MT020881.1). All virus manipulations were performed under stringent BSL-3 laboratory conditions according to strict protocols designed for safe and controlled handling of MERS-CoV and SARS-CoV-2.

MHV isolation and virion purification. Subconfluent 150-cm^2 flasks were infected with either MHV-A59 or MHV-ExoN(-) at an MOI of 0.01 PFU/cell. Supernatant was harvested at either 16 hours post infection (MHV-A59) or 24 hours post infection (MHV-ExoN(-)) when the monolayer was >95% fused and remained intact. Infection supernatant was clarified by centrifugation at 1500 x *g* for 5 minutes at 4°C. Viral

643 particles were purified on a 30% sucrose cushion by ultracentrifugation at 25,000 RPM
644 at 4°C for 16 hours. The viral pellet was resuspended in MSE buffer (10mM MOPS, pH
645 6.8; 150mM NaCl; 1 mM EDTA). Viral RNA was extracted using the TRIzol-LS reagent
646 according to manufacturer's protocols. RNA was quantified using the Qubit RNA HS
647 assay. Virion data in this paper is the result of three independent experiments
648 sequenced independently.

649 **MHV isolation from infected monolayers.** Three subconfluent 150-cm^2 flasks of DBT-
650 9 cells were infected with either MHV-WT or MHV-ExoN(-) at an MOI or 0.01 PFU/cell.
651 Monolayer was harvested at either 16 hpi (MHV-WT) or 24 hpi (MHV-ExoN(-)) when the
652 monolayer was >95% fused and >75% intact. RNA was extracted with TRIzol according
653 to manufacturer's protocols. Infected monolayer data in this paper is the result of three
654 independent experiments sequenced independently.

655 **MERS-CoV infection.** Three nearly confluent 25-cm^2 flasks of Vero CCL-81 cells were
656 infected with MERS-CoV at an MOI of 0.3 pfu/cell. Total infected cell lysates were
657 collected at 72 hpi with the monolayer $\leq$ 50% intact. RNA was extracted in TRIzol
658 according to manufacturer's protocols.

659 **SARS-CoV-2 infection.** Five subconfluent 25-cm^2 flasks of Vero E6 cells were infected
660 at an MOI = 0.45 pfu/cell and cellular monolayers were harvested 60 hpi when the
661 monolayer was >90% fused. RNA was extracted in TRIzol according to manufacturer's
662 protocols.

663 **Short-read Illumina RNA-sequencing of viral RNA.** Next generation sequencing
664 (NGS) libraries were generated using 2 μg of RNA. RNA was submitted to Genewiz for
665 library preparation and sequencing. Briefly, after quality control, polyadenylated RNA

was selected during library preparation. Isolated RNA was heat fragmented and libraries were prepared for 2 x 250 nucleotide paired-end sequencing performed (Illumina). Genewiz performed basecalling and read demultiplexing.

Direct RNA Nanopore sequencing. RNA from ultracentrifuge-purified virions was prepared for direct RNA Nanopore sequencing on the Oxford Nanopore Technologies MinION platform according to the manufacturer's protocols. Libraries were sequenced on fresh MinION R9.4 flow-cells for 24 hours, or until the pore occupancy was under 20%. Virion RNA from three independent experiments was sequenced on three separate flow cells for both MHV-WT and MHV-ExoN(-). MERS-CoV RNA from three independent cultures was sequenced on three separate flow cells. SARS-CoV-2 RNA isolated from three independent infections was sequenced on three separate flow cells.

QUANTIFICATION AND STATISTICAL ANALYSIS

Illumina RNA-seq processing and alignment. Raw reads were processed by first removing the Illumina TruSeq adapter using *Trimmomatic* (Bolger et al., 2014) default settings (command line parameters `java –jar trimmomatic.jar PE sample_R1.fastq.gz sample_R2.fastq.gz output_paired_R1.fastq output_unpaired_R1.fastq output_paired_R2.fastq output_unpaired_R2_unpaired.fastq ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGIWINDOW:4:15 MINLEN:`36). Reads shorter than 36 bp were removed and low-quality bases (Q score < 30) were trimmed from read ends. The raw FASTQ files were aligned to the MHV-A59 genome (AY910861.1), the MERS-CoV genome (JX869059.2), and the SARS-CoV-2 genome (MT020881.1) using the Python2 script *ViReMa* (Viral Recombination Mapper)

(Routh and Johnson, 2014) using the command line parameters `python2 ViReMa.py reference_index input.fastq output.sam --OuputDir sample_virema/ --OutputTag sample_virema -BED --MicroIndelLength 5`. The sequence alignment map (SAM) file was processed using the `samtools` (Li et al., 2009) suite to calculate nucleotide depth at each position in a sorted binary alignment map (BAM) file (using command line parameters `samtools depth -a -m 0 sample_virema.sorted.bam > sample_virema.coverage`).

Recombination junction analysis. Recombination junction frequency (J_{freq}) was calculated by comparing the number of nucleotides in detected recombination junctions to the total number of mapped nucleotides in a library. Nucleotides in detected recombination junctions were quantified as a sum of nucleotide depth reported at each junction in the BED file generated by *ViReMa*. Total nucleotides mapped to the MHV-A59 genome were quantified as a sum of nucleotide depth at each position across the genome in the tab-delineated text file generated by the *samtools*. J_{freq} was reported as junctions per 10^4 nucleotides sequenced. Mean J_{freq} values were compared between MHV-WT and MHV-ExoN(-) and statistical significance determined by an unpaired student's t-test. Junctions were mapped across the genome according to their start (5') and stop (3') positions. These junctions were first filtered in the forward (5' → 3') direction using the *dpylr* package (RStudio). The frequency of each junction was calculated by comparing the depth of the unique junction to the total number of nucleotides in all detected junctions in a library. Junctions were plotted according to the genomic position and colored according to $\log_{10}$ of the frequency using *ggplot2* in RStudio.

712 Recombination frequency was calculated at each genomic position by dividing the
713 number of nucleotides in any junction mapping to the position divided by the total
714 number of nucleotides sequenced at the position. Mean recombination frequencies
715 were compared between MHV-WT and MHV-ExoN(-) for three independent sequencing
716 experiments by a two-way ANOVA statistical analysis with multiple comparisons.
717 **Identification of sgmRNA and DVG junctions.** Forward recombination junctions were
718 classified as either canonical sgmRNA junctions or DVG junctions based on the position
719 of their junction sites and filtered in Microsoft Excel. Briefly, junction start sites were
720 filtered to those positioned between nucleotide 62 – 72 in the genome. The stop sites
721 were then filtered for those positioned within each respective sgmRNA TRS. sgmRNA
722 frequency was calculated by dividing the sum of the depth of all junctions corresponding
723 to an individual sgmRNA by the total number of mapped nucleotides at each position.
724 The filtered sgmRNA junctions were compiled and DVG junctions were filtered in
725 RStudio by performing an exclusionary `anti_join()` using *dplyr*. DVG junctions
726 were filtered in the forward direction. The percentage of sgmRNA and DVG junctions
727 was calculated by comparing the depth of all filtered sgmRNA or DVG junctions to the
728 sum of all detected forward junctions. Mean percent sgmRNA and DVG was compared
729 between three independent sequencing experiments in virion RNA. Statistical
730 significance was determined by a 2-way ANOVA test with multiple comparisons.
731 **Differential abundance of junctions.** To compare the abundance of junctions in MHV-
732 A59 and MHV-ExoN(-), the ViReMa output list of junctions was analyzed by in-house
733 scripts (https://github.com/DenisonLabVU) and the R package *DESeq2* (Love et al.,
734 2014). Junctions significantly up- or down-regulated in MHV-ExoN(-) were visualized

using *bioinfokit* (Bedre, 2020) and further mapped according to their genomic positions. Statistical significance was determined by the p-value of each junction calculated by the DESeq2 package in RStudio and junctions with a significant alteration of abundance in MHV-ExoN(-) compared to MHV-WT were visualized as either red or green in the graph generated by *bioinfokit*.

Direct RNA Nanopore alignment and analysis. Live basecalling was performed by *Guppy* in *MinKNOW*. Run statistics were generated from each sequencing experiment by *NanoPlot* (De Coster et al., 2018). Pass reads from all three experiments were concatenated for each virus and aligned to the genome using minimap2 (Li, 2018) and *FLAIR* (Full Length Alternative Isoforms of RNA) (Tang et al., 2018) to generate alignment files and BED files listing deletions detected in each sequenced RNA molecule. Both BAM and BED files were filtered for full length molecules using *samtools* and Microsoft Excel, respectively. Full-length MHV molecules were defined as encoding coverage at positions 71 and 31034, representing the canonical TRS-L and beginning of the 3' UTR required for replication. Because no TRS sequences have been experimentally confirmed in either MERS-CoV or SARS-CoV-2, we filtered both datasets for those that included any sequence of the 5' UTR and 3' UTR of the respective viruses. BED files generated by the flair align module were parsed based on the number of junctions were identified. Nanopore reads containing only 1 junction were identified using Microsoft Excel and unique junctions were quantified in RStudio using base–R functions. Sequencing coverage maps were generated from *samtools* depth analysis of filtered BAM files. All junctions present in sequenced libraries were mapped in Sashimi plots generated by the Integrated Genome Viewer (IGV) (Robinson et al.,

2011). Junctions present in full-length MHV RNA molecules with a single deletion were mapped according to their genomic positions as previously described. The genetic architectures of full-length RNA molecules sequenced by direct RNA Nanopore sequencing were visualized by filtering RNA molecules for at least 3 supporting reads. Low frequency variants were removed from this analysis.

Table S3. Direct RNA Nanopore read isoforms, related to Figures 2 and 5. Direct RNA Nanopore reads aligning to viral genome by minimap2. Individual reads are listed by read name. Genomic positions of read alignment are listed ("Read Start", "Read Stop"). Read segments aligning to the genome are noted ("# Segments") and start positions and aligned segment legnths listed ("Segment Start", "Segment Length").

Table S4. Genomic positions with significantly altered positional recombination frequency in MHV-ExoN(-) infected monolayer and virion RNA compared to MHV-WT, related to Figure 4. Positions with significantly altered recombination frequency in MHV-ExoN(-) infected monolayer RNA compared to MHV-WT and in MHV-ExoN(-) virion RNA compared to MHV-WT as determined by a 2-way ANOVA with multiple comparisons are listed. Genomic regions are noted. (N = 3 for each infected cell and virion RNA samples)

Table S5. Differential abundance of recombination junctions in MHV-ExoN(-) infected monolayer compared to MHV-WT, related to Figure 4. Junctions with altered abundance in MHV-ExoN(-) infected monolayer RNA compared to MHV-WT and in MHV-ExoN(-) virion RNA compared to MHV-WT are listed. P-values calculated by DESeq2. (N = 3 for each infected monolayer and virion RNA samples)

781

782 **REFERENCES**

783

784 Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., Smith,
785 E.C., Case, J.B., Feng, J.Y., Jordan, R., et al. (2018). Coronavirus Susceptibility to the
786 Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the
787 Proofreading Exoribonuclease. MBio *9*.

788 Anthony, S.J., Gilardi, K., Menachery, V.D., Goldstein, T., Ssebide, B., Mbabazi, R.,
789 Navarrete-Macias, I., Liang, E., Wells, H., Hicks, A., et al. (2017). Further Evidence for
790 Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus.
791 MBio *8*, e00373-17, /mbio/8/2/e00373-17.atom.

792 Bedre, R. (2020). reneshbedre/bioinfokit.

793 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for
794 Illumina sequence data. Bioinformatics *30*, 2114–2120.

795 Brian, D.A., and Spaan, W.J.M. (1997). Recombination and coronavirus defective
796 interfering RNAs. Seminars in Virology *8*, 101–111.

797 Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M., and Peersen, O.B. (2015).
798 Structure-function relationships underlying the replication fidelity of viral RNA-dependent
799 RNA polymerases. Journal of Virology *89*, 275–286.

800 Case, J.B., Li, Y., Elliott, R., Lu, X., Graepel, K.W., Sexton, N.R., Smith, E.C., Weiss,
801 S.R., and Denison, M.R. (2017). Murine Hepatitis Virus nsp14 Exoribonuclease Activity
802 Is Required for Resistance to Innate Immunity. J Virol *92*, e01531-17.

803 Chen, W., and Baric, R.S. (1996). Molecular anatomy of mouse hepatitis virus
804 persistence: coevolution of increased host cell resistance and virus virulence. Journal of
805 Virology *70*, 3947–3960.

806 Chen, N., Li, S., Zhou, R., Zhu, M., He, S., Ye, M., Huang, Y., Li, S., Zhu, C., Xia, P., et
807 al. (2017). Two novel porcine epidemic diarrhea virus (PEDV) recombinants from a
808 natural recombinant and distinct subtypes of PEDV variants. Virus Research *242*, 90–
809 95.

810 Cruz, J.L.G., Sola, I., Becares, M., Alberca, B., Plana, J., Enjuanes, L., and Zuñiga, S.
811 (2011). Coronavirus Gene 7 Counteracts Host Defenses and Modulates Virus
812 Virulence. PLOS Pathogens *7*, e1002090.

813 De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018).
814 NanoPack: visualizing and processing long-read sequencing data. Bioinformatics *34*,
815 2666–2669.

816 Di, H., Madden, J.C., Morantz, E.K., Tang, H.-Y., Graham, R.L., Baric, R.S., and
817 Brinton, M.A. (2017). Expanded subgenomic mRNA transcriptome and coding capacity
818 of a nidovirus. Proc. Natl. Acad. Sci. U.S.A. *114*, E8895–E8904.

819 Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S.,
820 Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A.M., et al. (2003). Identification
821 of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J.
822 Med. *348*, 1967–1976.

823 Dufour, D., Mateos-Gomez, P.A., Enjuanes, L., Gallego, J., and Sola, I. (2011).
824 Structure and functional relevance of a transcription-regulating sequence involved in
825 coronavirus discontinuous RNA synthesis. Journal of Virology *85*, 4963–4973.

826 Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L., and Denison, M.R. (2007). High fidelity of
827 murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants.
828 Journal of Virology *81*, 12135–12144.

829 Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S.,
830 Graham, R.L., Baric, R.S., Stockwell, T.B., et al. (2010). Infidelity of SARS-CoV Nsp14-
831 exonuclease mutant virus replication is revealed by complete genome sequencing.
832 PLoS Pathogens *6*, e1000896.

833 Feng, K.Y., Chen, T., Zhang, X., Shao, G.M., Cao, Y., Chen, D.K., Lin, W.C., Chen, F.,
834 and Xie, Q.M. (2018). Molecular characteristic and pathogenicity analysis of a virulent
835 recombinant avain infectious bronchitis virus isolated in China. Poultry Science *97*,
836 3519–3531.

837 Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L.,
838 Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., et al. (2018a). Structural and
839 molecular basis of mismatch correction and ribavirin excision from coronavirus RNA.
840 Proc Natl Acad Sci USA *115*, E162–E171.

841 Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L.,
842 Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., et al. (2018b). Structural and
843 molecular basis of mismatch correction and ribavirin excision from coronavirus RNA.
844 Proceedings of the National Academy of Sciences of the United States of America *115*,
845 E162–E171.

846 Gammon, D.B., and Evans, D.H. (2009). The 3′-to-5′ Exonuclease Activity of Vaccinia
847 Virus DNA Polymerase Is Essential and Plays a Role in Promoting Virus Genetic
848 Recombination. J Virol *83*, 4236–4250.

849 Grady, L.M., Szczepaniak, R., Murelli, R.P., Masaoka, T., Le Grice, S.F.J., Wright, D.L.,
850 and Weller, S.K. (2017). The Exonuclease Activity of Herpes Simplex Virus 1 UL12 Is
851 Required for Production of Viral DNA That Can Be Packaged To Produce Infectious
852 Virus. J. Virol. *91*.

Graepel, K.W., Agostini, M.L., Lu, X., Sexton, N.R., and Denison, M.R. (2019). Fitness Barriers Limit Reversion of a Proofreading-Deficient Coronavirus. J Virol *93*, e00711-19, /jvi/93/20/JVI.00711-19.atom.

Graham, R.L., Becker, M.M., Eckerle, L.D., Bolles, M., Denison, M.R., and Baric, R.S. (2012). A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. Nat. Med. *18*, 1820–1826.

Hon, C.-C., Lam, T.-Y., Shi, Z.-L., Drummond, A.J., Yip, C.-W., Zeng, F., Lam, P.-Y., and Leung, F.C.-C. (2008). Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus. Journal of Virology *82*, 1819–1826.

Huang, J.-M., Jan, S.S., Wei, X., Wan, Y., and Ouyang, S. (2020). Evidence of the Recombinant Origin and Ongoing Mutations in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). BioRxiv 2020.03.16.993816.

Keck, J.G., Matsushima, G.K., Makino, S., Fleming, J.O., Vannier, D.M., Stohlman, S.A., and Lai, M.M. (1988). In vivo RNA-RNA recombination of coronavirus in mouse brain. Journal of Virology *62*, 1810–1813.

Kempf, B.J., Peersen, O.B., and Barton, D.J. (2016). Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. J. Virol. *90*, 8410–8421.

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. BioRxiv 2020.03.12.988865.

Kim, H., Ellis, V.D., Woodman, A., Zhao, Y., Arnold, J.J., and Cameron, C.E. (2019). RNA-Dependent RNA Polymerase Speed and Fidelity are not the Only Determinants of the Mechanism or Efficiency of Recombination. Genes (Basel) *10*.

Kim, Y.-N., Lai, M.M.C., and Makino, S. Generation and selection of coronavirus defective interfering RNA with large open reading frame by RNA recombination and possible editing. Virology *194*, 244–253.

Kirchdoerfer, R.N., and Ward, A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun *10*, 2342.

Kottier, S.A., Cavanagh, D., and Britton, P. (1995). Experimental Evidence of Recombination in Coronavirus Infectious Bronchitis Virus. Virology *213*, 569–580.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al. (2003). A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. *348*, 1953–1966.

Kuo, L., and Masters, P.S. (2013). Functional Analysis of the Murine Coronavirus Genomic RNA Packaging Signal. J Virol *87*, 5182–5192.

888 Lai, M.M., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L., and Stohlman,
889 S.A. (1985). Recombination between nonsegmented RNA genomes of murine
890 coronaviruses. Journal of Virology *56*, 449–456.

891 Langsjoen, R., Muruato, A., Kunkel, S., Jaworski, E., and Routh, A. (2020). Differential
892 alphavirus defective RNA diversity between intracellular and encapsidated
893 compartments is driven by subgenomic recombination events (Molecular Biology).

894 Lau, S.K.P., Feng, Y., Chen, H., Luk, H.K.H., Yang, W.-H., Li, K.S.M., Zhang, Y.-Z.,
895 Huang, Y., Song, Z.-Z., Chow, W.-N., et al. (2015). Severe Acute Respiratory Syndrome
896 (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from
897 Greater Horseshoe Bats through Recombination. J. Virol. *89*, 10532–10547.

898 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics
899 *34*, 3094–3100.

900 Li, C., Wang, H., Shi, J., Yang, D., Zhou, G., Chang, J., Cameron, C.E., Woodman, A.,
901 and Yu, L. (2019). Senecavirus-Specific Recombination Assays Reveal the Intimate
902 Link between Polymerase Fidelity and RNA Recombination. J Virol *93*, e00576-19,
903 /jvi/93/13/JVI.00576-19.atom.

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

907 Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z.,
908 Zhang, H., et al. (2005). Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *310*,
909 5.

910 Li, X., Giorgi, E.E., Marichann, M.H., Foley, B., Xiao, C., Kong, X., Chen, Y., Korber, B.,
911 and Gao, F. (2020). Emergence of SARS-CoV-2 through Recombination and Strong
912 Purifying Selection. BioRxiv 2020.03.20.000885.

913 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
914 dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

915 Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., and
916 Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus
917 nsp14–nsp10 complex. Proc Natl Acad Sci USA *112*, 9436–9441.

918 Makino, S. Structure of the Intracellular Defective Viral RNAs of Defective Interfering
919 Particles of Mouse Hepatitis Virus. J. VIROL. 8.

920 Makino, S., Taguchi, Fumihiro, and Fujiwara, Kosaku (1984). Defective interfering
921 particles of mouse hepatitis virus. Virology *133*, 9–17.

Makino, S., Fujioka, N., and Fujiwara, K. (1985). Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. Journal of Virology *54*, 329–336.

Makino, S., Keck, J.G., Stohlman, S.A., and Lai, M.M. (1986). High-frequency RNA recombination of murine coronaviruses. Journal of Virology *57*, 729–737.

Makino, S., Shieh, C.-K., Soe, L.H., Baker, S.C., and Lai, M.M.C. (1988). Primary structure and translation of a defective interfering RNA of murine coronavirus. Virology *166*, 550–560.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., and Ziebuhr, J. (2006). Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proceedings of the National Academy of Sciences of the United States of America *103*, 5108–5113.

Muth, D., Corman, V.M., Roth, H., Binger, T., Dijkman, R., Gottula, L.T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtarič, D., et al. (2018). Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep *8*, 1–11.

Paraskevis, D., Kostaki, E.G., Magiorkinis, G., Panayiotakopoulos, G., and Tsiodras, S. (2020). Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. BioRxiv 2020.01.26.920249.

Patiño-Galindo, J.Á., Filip, I., AlQuraishi, M., and Rabadan, R. (2020). Recombination and lineage-specific mutations led to the emergence of SARS-CoV-2. BioRxiv 2020.02.10.942748.

Penzes, Z., Wroe, C., Brown, T.D.K., Britton, P., and Cavanagh, D. (1996). Replication and Packaging of Coronavirus Infectious Bronchitis Virus Defective RNAs Lacking a Long Open Reading Frame. J. VIROL. *70*, 9.

Poirier, E.Z., Mounce, B.C., Rozen-Gagnon, K., Hooikaas, P.J., Stapleford, K.A., Moratorio, G., and Vignuzzi, M. (2015). Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. J. Virol. *90*, 2446–2454.

Randhawa, G.S., Soltysiak, M.P.M., Roz, H.E., Souza, C.P.E. de, Hill, K.A., and Kari, L. (2020). Machine learning analysis of genomic signatures provides evidence of associations between Wuhan 2019-nCoV and bat betacoronaviruses. BioRxiv 2020.02.03.932350.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat Biotechnol *29*, 24–26.

Routh, A., and Johnson, J.E. (2014). Discovery of functional genomic motifs in viruses with ViReMa–a Virus Recombination Mapper–for analysis of next-generation sequencing data. Nucleic Acids Res *42*, e11.

Schaad, M.C., and Baric, R.S. (1994). Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. Journal of Virology *68*, 8169–8179.

Schumacher, A.J., Mohni, K.N., Kan, Y., Hendrickson, E.A., Stark, J.M., and Weller, S.K. (2012). The HSV-1 Exonuclease, UL12, Stimulates Recombination by a Single Strand Annealing Mechanism. PLOS Pathogens *8*, e1002862.

Scobey, T., Yount, B.L., Sims, A.C., Donaldson, E.F., Agnihothram, S.S., Menachery, V.D., Graham, R.L., Swanstrom, J., Bove, P.F., Kim, J.D., et al. (2013). Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. PNAS *110*, 16157–16162.

Smith, E.C., Blanc, H., Vignuzzi, M., and Denison, M.R. (2013). Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics. PLoS Pathog *9*, e1003565.

Subissi, L., Imbert, I., Ferron, F., Collet, A., Coutard, B., Decroly, E., and Canard, B. (2014). SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Research *101*, 122–130.

Taiaroa, G., Rawlinson, D., Featherstone, L., Pitt, M., Caly, L., Druce, J., Purcell, D., Harty, L., Tran, T., Roberts, J., et al. (2020). Direct RNA sequencing and early evolution of SARS-CoV-2 (Microbiology).

Tang, A.D., Soulette, C.M., Baren, M.J. van, Hart, K., Hrabeta-Robinson, E., Wu, C.J., and Brooks, A.N. (2018). Full-length transcript characterization of *SF3B1* mutation in chronic lymphocytic leukemia reveals downregulation of retained introns (Genomics).

Viehweger, A., Krautwurst, S., Lamkiewicz, K., Madhugiri, R., Ziebuhr, J., Hölzer, M., and Marz, M. (2019). Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. Genome Res. *29*, 1545–1554.

Wahba, L., Jain, N., Fire, A.Z., Shoura, M.J., Artiles, K.L., McCoy, M.J., and Jeong, D.E. (2020). Identification of a pangolin niche for a 2019-nCoV-like coronavirus through an extensive meta-metagenomic search. BioRxiv 2020.02.08.939660.

Weiss, S.R., Hughes, S.A., Bonilla, P.J., Turner, J.D., Leibowitz, J.L., and Denison, M.R. (1994). Coronavirus polyprotein processing. In Positive-Strand RNA Viruses, (Vienna: Springer Vienna), pp. 349–358.

993 Wong, M.C., Javornik Cregeen, S.J., Ajami, N.J., and Petrosino, J.F. (2020). Evidence
994 of recombination in coronaviruses implicating pangolin origins of nCoV-2019
995 (Microbiology).

996 Woodman, A., Lee, K.-M., Janissen, R., Gong, Y.-N., Dekker, N.H., Shih, S.-R., and
997 Cameron, C.E. (2018). Predicting Intraserotypic Recombination in Enterovirus 71. J
998 Virol *93*, e02057-18, /jvi/93/4/JVI.02057-18.atom.

999 Wu, H.-Y., and Brian, D.A. (2010). Subgenomic messenger RNA amplification in
1000 coronaviruses. Proceedings of the National Academy of Sciences of the United States
1001 of America *107*, 12257–12262.

1002 Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian,
1003 J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory
1004 disease in China. Nature 1–5.

1005 Yi, H. (2020). 2019 novel coronavirus is undergoing active recombination. Clin Infect
1006 Dis.

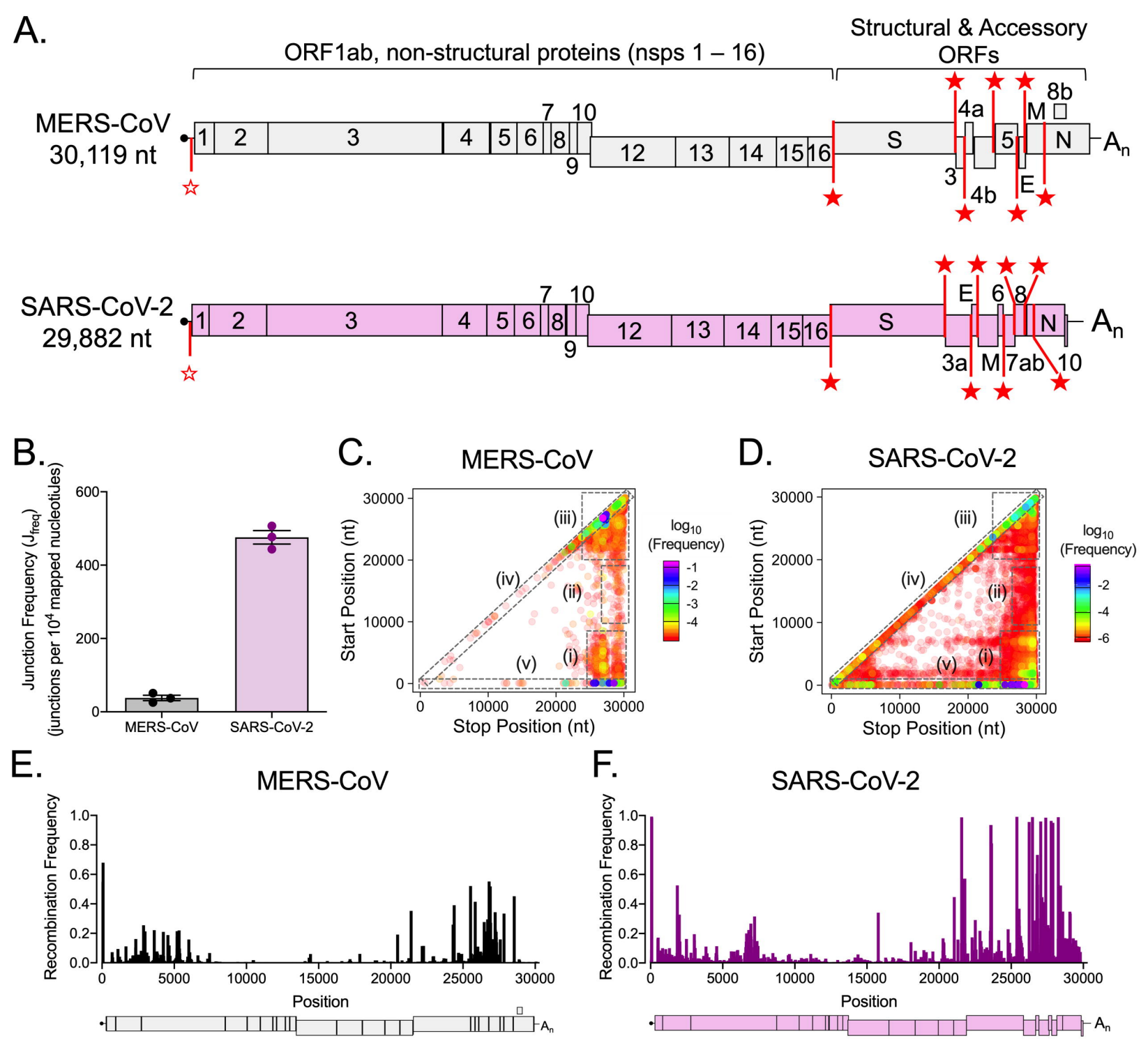
1007 Yount, B., Denison, M.R., Weiss, S.R., and Baric, R.S. (2002). Systematic Assembly of
1008 a Full-Length Infectious cDNA of Mouse Hepatitis Virus Strain A59. Journal of Virology
1009 *76*, 11065–11078.

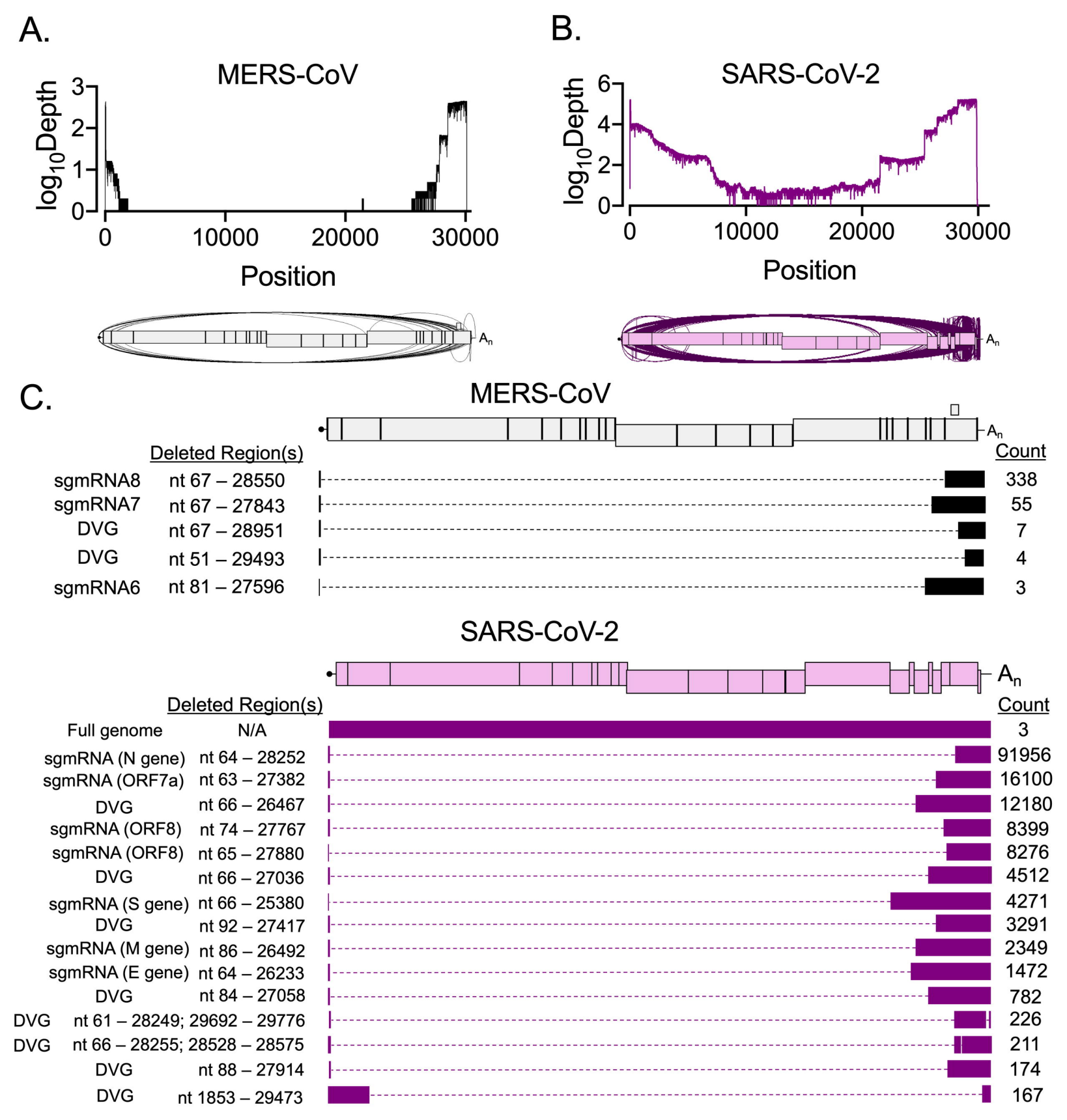
1010 Yusof, M.F., Queen, K., Eltahir, Y.M., Paden, C.R., Al Hammadi, Z.M.A.H., Tao, Y., Li,
1011 Y., Khalafalla, A.I., Shi, M., Zhang, J., et al. (2017). Diversity of Middle East respiratory
1012 syndrome coronaviruses in 109 dromedary camels based on full-genome sequencing,
1013 Abu Dhabi, United Arab Emirates. Emerging Microbes & Infections *6*, 1–10.

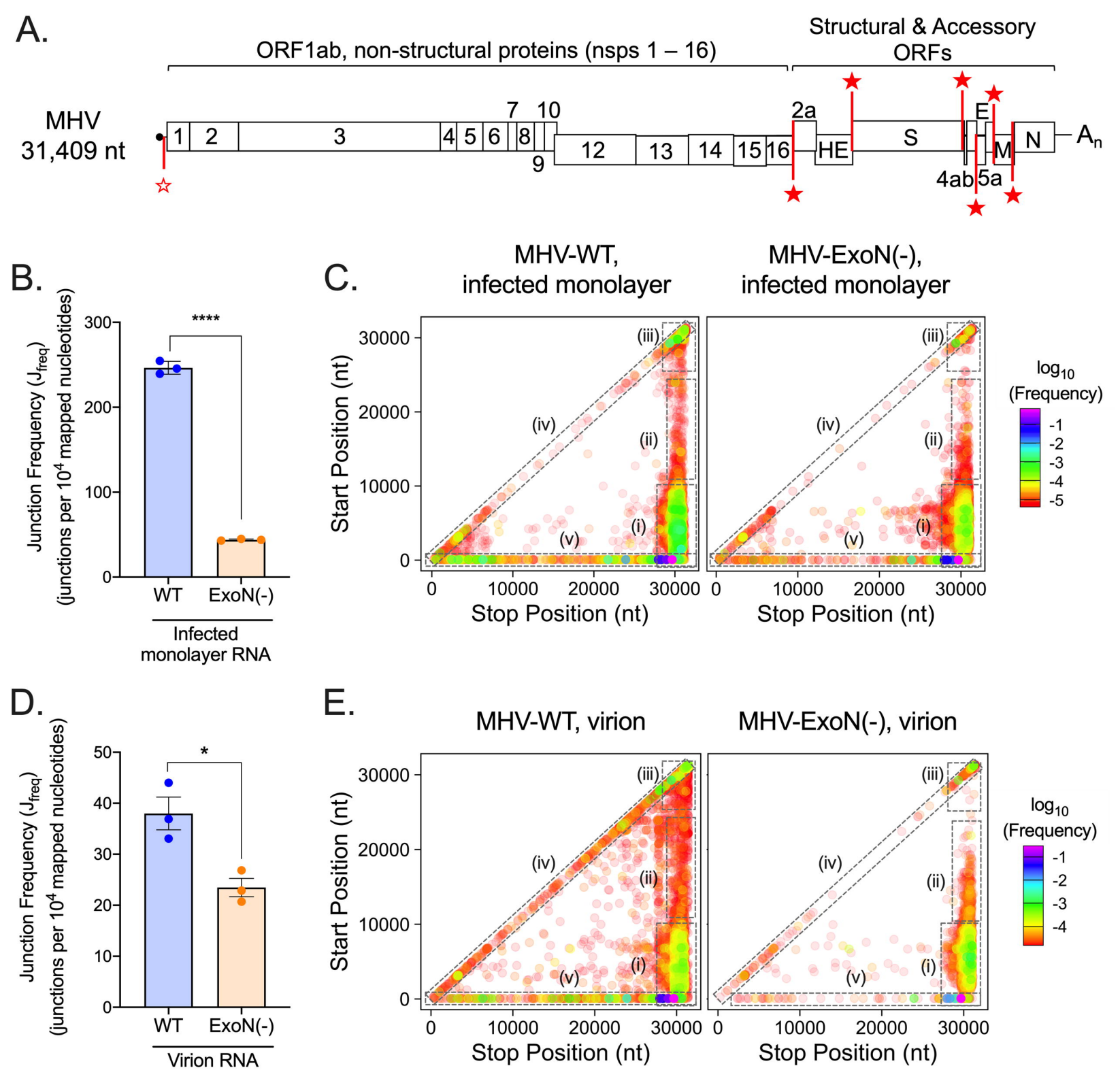
1014 Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., and Fouchier,
1015 R.A.M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi
1016 Arabia. The New England Journal of Medicine *367*, 1814–1820.

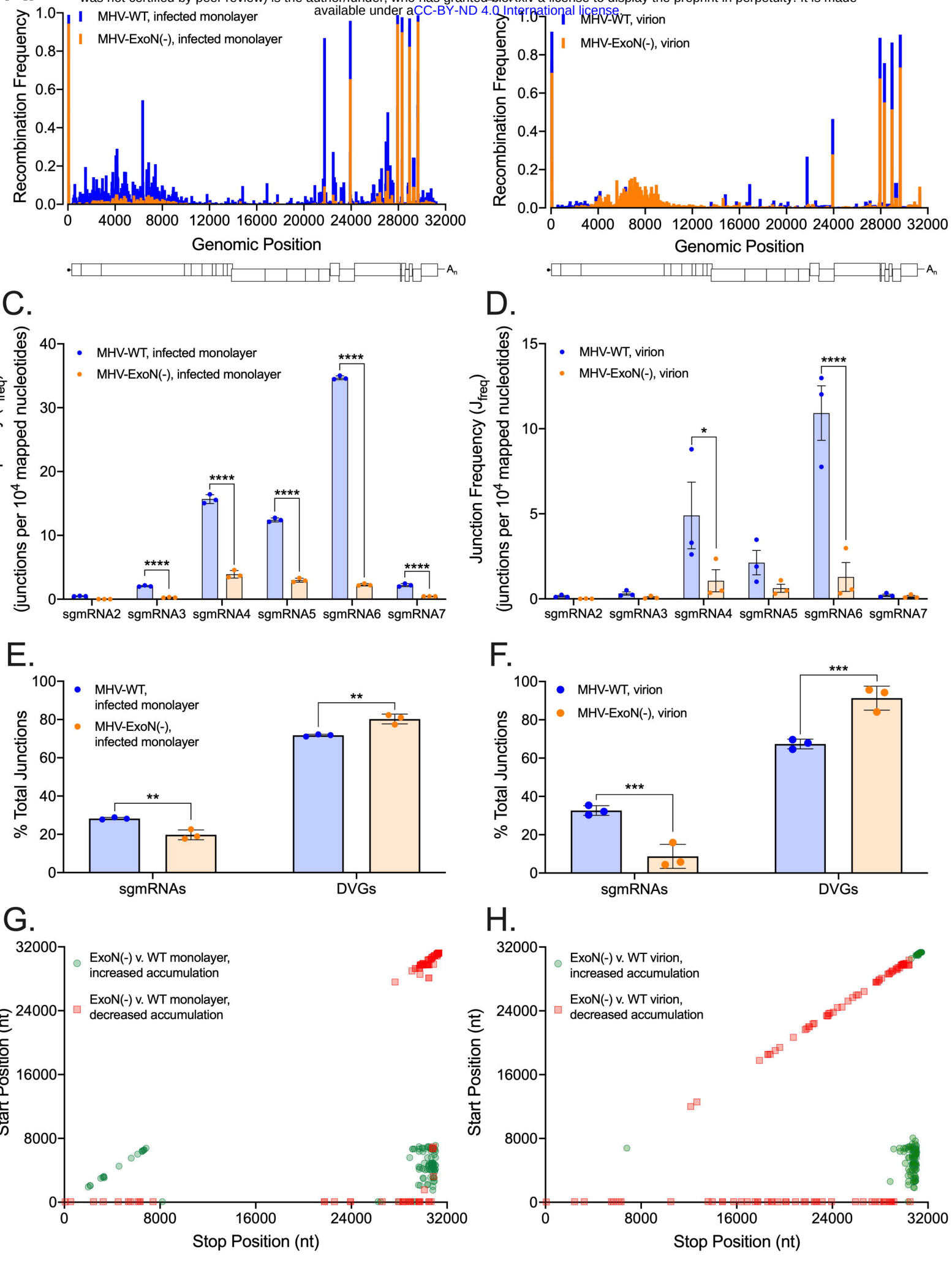
1017 Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li,
1018 B., Huang, C.-L., et al. (2020). Discovery of a novel coronavirus associated with the
1019 recent pneumonia outbreak in humans and its potential bat origin. BioRxiv
1020 2020.01.22.914952.

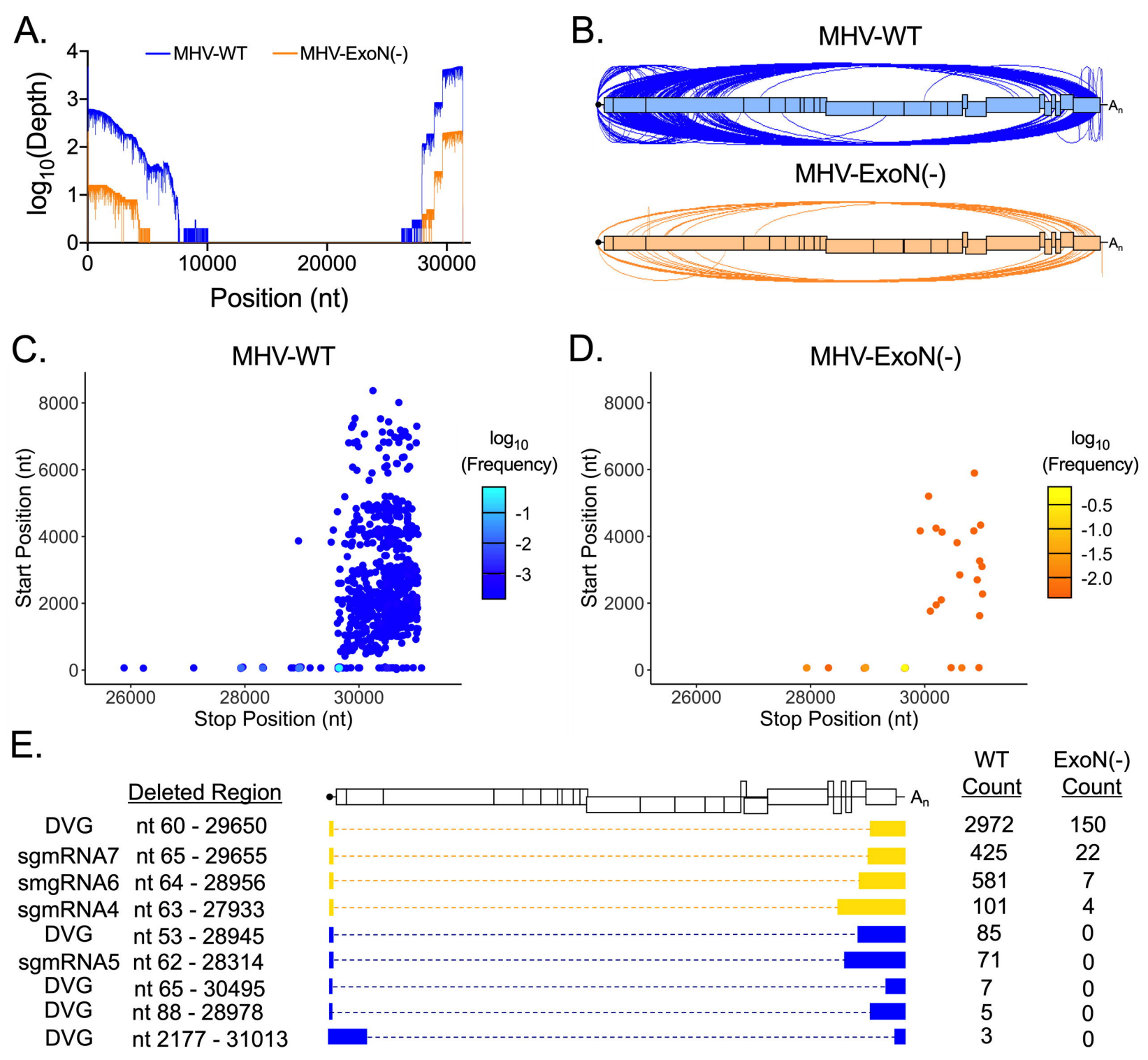
1021


A.
ORF1ab, non-structural proteins (nsps 1 – 16)
Structural & Accessory ORFs
MERS-CoV
30,119 nt
SARS-CoV-2
29,882 nt
B.
Junction Frequency (J_freq)
(junctions per 10^4 mapped nucleotides)
MERS-CoV
SARS-CoV-2
C.
MERS-CoV
Start Position (nt)
Stop Position (nt)
log10 (Frequency)
D.
SARS-CoV-2
E.
MERS-CoV
Recombination Frequency
Position
F.
SARS-CoV-2


A.
MERS-CoV
log10Depth
3
2
1
0
0
10000
20000
30000
Position
An
B.
SARS-CoV-2
log10Depth
6
4
2
0
0
10000
20000
30000
Position
An
C.
MERS-CoV
An
Deleted Region(s)
Count
sgmRNA8 nt 67 – 28550 338
sgmRNA7 nt 67 – 27843 55
DVG nt 67 – 28951 7
DVG nt 51 – 29493 4
sgmRNA6 nt 81 – 27596 3
SARS-CoV-2
An
Deleted Region(s)
Count
Full genome N/A 3
sgmRNA (N gene) nt 64 – 28252 91956
sgmRNA (ORF7a) nt 63 – 27382 16100
DVG nt 66 – 26467 12180
sgmRNA (ORF8) nt 74 – 27767 8399
sgmRNA (ORF8) nt 65 – 27880 8276
DVG nt 66 – 27036 4512
sgmRNA (S gene) nt 66 – 25380 4271
DVG nt 92 – 27417 3291
sgmRNA (M gene) nt 86 – 26492 2349
sgmRNA (E gene) nt 64 – 26233 1472
DVG nt 84 – 27058 782
DVG nt 61 – 28249; 29692 – 29776 226
DVG nt 66 – 28255; 28528 – 28575 211
DVG nt 88 – 27914 174
DVG nt 1853 – 29473 167


A.
ORF1ab, non-structural proteins (nsps 1 – 16)
Structural & Accessory ORFs
MHV
31,409 nt
1
2
3
4
5
6
7
8
9
10
12
13
14
15
16
2a
HE
S
E
M
N
4ab
5a
An
B.
Junction Frequency (Jfreq)
(junctions per 10^4 mapped nucleotides)
****
WT
ExoN(-)
Infected monolayer RNA
C.
MHV-WT, infected monolayer
MHV-ExoN(-), infected monolayer
Start Position (nt)
Stop Position (nt)
(i)
(ii)
(iii)
(iv)
(v)
log10 (Frequency)
D.
*
Virion RNA
E.
MHV-WT, virion
MHV-ExoN(-), virion




A.
MHV-WT
MHV-ExoN(-)
log10(Depth)
Position (nt)
B.
MHV-WT
MHV-ExoN(-)
An
C.
MHV-WT
Start Position (nt)
Stop Position (nt)
log10 (Frequency)
D.
MHV-ExoN(-)
Start Position (nt)
Stop Position (nt)
log10 (Frequency)
E.
Deleted Region
WT Count
ExoN(-) Count
DVG nt 60 - 29650 2972 150
sgmRNA7 nt 65 - 29655 425 22
smgRNA6 nt 64 - 28956 581 7
sgmRNA4 nt 63 - 27933 101 4
DVG nt 53 - 28945 85 0
sgmRNA5 nt 62 - 28314 71 0
DVG nt 65 - 30495 7 0
DVG nt 88 - 28978 5 0
DVG nt 2177 - 31013 3 0